

CHAPTER 1

Predictive Analytics and Machine Learning for Medical Informatics

A Survey of Tasks and Techniques

Deepti Lamba^{a,*}, William H. Hsu^{*} and Majed Alsadhan^{*}

^{*} Kansas State University, Department of Computer Science

^a Corresponding: dlamba@ksu.edu

Abstract

In this article, we survey machine learning and predictive analytics methods for medical informatics. We begin by surveying the current state of practice, key task definitions, and open research problems related to predictive modeling in diagnostic medicine. This follows the traditional supervised, unsupervised, and reinforcement learning taxonomy. Next, we review current research on semi-supervised, active, and transfer learning, and on differentiable computing methods such as deep learning. The focus of this section is on deep neural networks with common use cases in computational medicine, including self-supervised learning scenarios: these include convolutional neural networks for image analysis, recurrent neural networks for time series, and generative adversarial models for correction of class imbalance in differential diagnosis and anomaly detection. We then continue by assessing salient connections between the current state of machine learning research and data-centric healthcare analytics, focusing specifically on diagnostic imaging and multisensor integration as crucial research topics within predictive analytics. This section includes synthesis experiments on analytics and multisensor data fusion within a diagnostic test bed. Finally, we conclude by relating open problems of machine learning for prediction-based medical informatics surveyed in this article to the impact of big data and its associated challenges, trends, and limitations of current work, including privacy and security of sensitive patient data.

Keywords: Predictive Analytics, Machine Learning, Deep Learning, Medical Informatics, Health Informatics, Prognosis, Diagnosis, Health Recommender Systems, Integrative medicine

Chapter points

- The chapter provides an introduction to predictive analytics in medical informatics.
- This is followed by a brief background on relevant literature in the field.
- The chapter also presents a practical demonstration of using Deep Learning on a publicly available image data set for classification.

1. Introduction: Predictive Analytics for Medical Informatics

Medical informatics is a broad domain at the intersection of technology and health care which aims to (1) make medical data of patients available to them and to health care providers, thus enabling them to make timely medical decisions; and (2) manage this data for educational and research purposes. According to Morris Collen, the first articles on medical informatics appeared in the 1950s [1]. However, it was first identified as a new specialty in the 1970s.[2]

This section surveys goals, the state of practice, and specific task definitions for machine learning in medical fields and the practice of health care. These sectors produce an enormous amount of data which is highly complex and comes from heterogeneous sources: electronic health records (EHRs) [3], medical equipment and devices, wearable technologies, handwritten notes, lab results, prescriptions, and clinical information. The application of predictive analytics to this data offers potential benefits such as improved standards of care for patients, lower medical costs, and higher resultant patient satisfaction with health care providers.

1.1. Overview: Goals of Machine Learning

Predictive analytics is a branch of data science that applies various techniques including statistical inference, machine learning, data mining, and information visualization towards the ultimate goal of forecasting, modeling, and understanding the future behavior of a system based on historical and/or real-time data. This chapter focuses on machine learning [4][5] algorithms for building predictive models. In addition, we will survey applications of machine learning to automation and computer vision, especially image classification, which in some medical domains has achieved accuracy comparable to that of a human expert [6]. Sidey-Gibbons and Sidey-Gibbons (2019) [7] provide an introduction to machine learning using a publicly available data set for cancer diagnosis. In recent years, deep learning [8][9] has attained technical success and scientific attention in application domains including medicine [10] and health care [11]. Deep neural networks such as convolutional neural nets (*ConvNets* or *CNNs*) have become the predominant state-of-the-art method for analysis of images such as MRI scans, to predict diseases such as Alzheimer’s dementia [12].

Deep learning models face several challenges in medical domains which hinder their acceptability to the medical community - temporality of data, domain complexity, and lack of interpretability [10]. According to Miotto et al. (2018) [10], the most used deep architectures in the health domain, briefly discussed in Section 3, include recurrent neural networks [13], ConvNets [14], restricted Boltzmann machines [15][16], autoencoders [17][18], and variations thereof. This chapter focuses on five tasks of medical informatics: differential diagnosis [19], prediction [20], therapy recommendation [21], automation of treatment [22], and analytics in integrative medicine

[23].

1.2. Current State of Practice

A trend forecast study [24] published by the Society of Actuaries indicates a growing usage of predictive analytics for healthcare. In 2019, 60% of healthcare organizations were already using predictive analytics, and 20% indicated that they would be using the same in the following year. Among those that currently use predictive analytics, 39% reported a decrease in healthcare costs and 42% improvement in patient satisfaction. These statistics demonstrate the interest of organizations in using predictive analytics in medical domain for improving their services.

1.3. Key Task Definitions

This section provides an overview of machine learning goals in health informatics. The goals of prediction are introduced above in Section 1.1.

1.3.1. Diagnosis

Differential diagnosis is defined as the process of differentiating between probability of one disease versus that of other diseases with similar symptoms that could possibly account for illness in a patient. A technical series published by World Health Organization in 2016 states that the most important task performed by primary care providers is diagnosis [25]. Machine learning tools have been used primarily for disease diagnosis throughout the history of medical informatics. Graber, Franklin, and Gordon (2005) [26] conducted a study to determine the causes of diagnostic errors and to develop a comprehensive taxonomy for the classification of these errors.

Miller (1994) [27] provides a representative bibliography of the state of the art and history of medical diagnostic decision support systems (MDDSS) at the time. These systems can be divided into several subcategories, among which expert systems have been used most often [28]. Many of the earliest rule-based expert systems [29] were developed for medical diagnosis. Shortliffe (1986) [30] gives insights into the design of expert systems for diagnostic medicine developed during the 1970s and 1980s, including: (1) MYCIN [31] [32], which focused on infectious diseases; (2) INTERNIST-1 [33]; (3) QMR [34][35]; (4) DXplain[36][37] [38], a diagnostic decision support system developed continuously between 1986 and the early 2000s.

The most prominent limitation of expert systems was the acquisition of knowledge [39] or building a knowledge base, which is both, time-consuming and a complex process that requires access to expert domain knowledge. Additionally, updating the knowledge base requires significant human effort. These systems were usually designed to support users with an expert level of medical knowledge. A more recent review of expert systems is presented by Abu-Nasser (2017) [40]. Expert systems are still around but their limitations led to advances in rule learning and classification for

differential diagnosis. Salman and Abu-Naser (2020) [41] developed a diagnostic system for COVID-19 using medical websites for the knowledge base. COVID-19 is a novel viral disease that has affected millions of people around the world. The system was tested by a group of doctors and they were satisfied with its performance and ease of use. Another expert system for COVID-19 was built by Almadhoun and Abu-Naser (2020) [42] for helping patients determine if they have been infected with COVID-19. The system gives instructions to the user based on the symptoms. The knowledge base was compiled using medical sites such as NHS Trust.

Kononenko (2001) [43] provides an historical overview of ML methods used in medical domain and a discussion about state-of-the art algorithms: Assistant-R and Assistant-I [44], lookahead feature construction (LFC) [45], naïve Bayesian classifier [46], semi-naïve Bayesian classifier [47], k-NN [48], and back propagation with weight elimination [49]. The paper’s experimental findings show that most classifiers have a comparable performance which makes model explainability a deciding factor behind the choice of classifier.

1.3.2. Predictive Analytics

Prognosis is defined as a forecast of the probable course and/or outcome of a disease. It is an important task in clinical patient management. Cruz and Wishart (2006) [50] outline the focal predictive tasks of prognosis/predictions for cancer. Based on these predictive tasks, the general definition of the prognosis task comprises these variants: (1) prediction of disease susceptibility (or likelihood of developing any disease prior to the actual occurrence of the disease); (2) prediction of disease recurrence (or predicting the likelihood of redeveloping the disease after its resolution), and (3) prediction of survivability (or predicting an outcome after the diagnosis of the disease in terms of life expectancy, survivability, disease progression, etc).

Ohno-Machado (2001) [51] define prognosis as “an estimate of cure, complication, recurrence of disease, level of function, length of stay in health care facilities, or survival for a patient.” The author focuses on techniques that are used to model prognosis - especially the *survival analysis methods*. A detailed discussion of survival analysis methods is beyond the scope of this article. We refer the interested readers to the book [52] and a review of survival analysis techniques [53]. Prognostic tasks are categorized as (1) prediction for a single point in time, and (2) time-related predictions. Methods used to build prognostic models include Cox proportional hazards [54], logistic regression [55], neural networks [56].

Mendez-Tellez and Dorman (2005) [57], published an article that states that ICUs have increased the critical care being provided to injured or critically ill patients. However, the costs for the ICU treatments are very high, which has given rise to prediction models, which are classified as disease-specific or generic models. These systems work by employing a scoring system that assigns points according to illness severity

and then generate a probability estimate as an outcome of the model. We do not discuss scoring systems in this article. We refer the interested reader to a compendium of scoring systems for outcome distributions [58]. A few of the outcome prediction models used for intensive care predictions include: Mortality Probability Model II [59], Simplified Acute Physiology Score (SAPS) II [60], Acute Physiology and Chronic Health Evaluation (APACHE) II [61] and APACHE III [62]. These systems build logistic regression models to predict hospital mortality by using a set of clinical and physiology variables.

Another important application of learning is cancer prognosis and prediction. Early diagnosis and prognosis of any life-threatening disease, especially cancer, presents crucial real-time requirements and poses research challenges. Machine learning is being used to build classification models for categorization of cases by risk level. This is essential for clinical management of cancer patients. Kourou et al. (2015) [63] review methods that have been used to model the progression of cancer. The methods used for this task include artificial neural networks (ANNs) [56] and decision trees [64], which have been used for three decades for cancer detection. The authors also noticed a growing trend of using methods like support vector machines (SVM) [65][66] and Bayesian networks (BN) [67] for cancer prediction and prognosis.

1.3.3. Therapy Recommendation

A classic example of a machine learning application is a recommender system [68][69]. *Recommender systems* are widely used to recommend items, services, merchandise and users to each other based on similarity. However, the use of recommender systems in health and medical domain has not been widespread. The earliest article on recommender system in health is from the year 2007 and by 2016 only seventeen articles were found for the query “recommender system health” in web of science [70].

Valdez et al. (2016) [70] argue that the lack of popularity of recommender systems in the medical domain is due to several reasons: (1) the benchmarking criteria in medical scenarios; (2) domain complexity; (3) the different end-user groups. The end-users or target users for recommender systems can be patients, medical professionals, or people who are healthy. Recommender systems can be designed to recommend therapies, sports or physical activities, medication, diagnosis or even food or other nutritional information. This article also outlines major challenges faced by recommender systems in the medical domain. Challenges include a lack of clear task definition for recommender systems in the health domain. The definition depends on the target user and the item being recommended.

Wiesner and Pfeifer (2014) [71] propose a health recommender system (HRS) that recommends relevant medical information to the patient by using the graphical user interface of the personal health record (PHR) [72]. The HRS uses the PHR to build a user-profile and the authors argue that collaborative filtering is an appropriate approach

for building such a system.

Gräßer et al. (2017) [21] propose two methods for recommending therapies for patients suffering from psoriasis: a Collaborative Recommender and a hybrid Demographic-Recommender. The two are compared and combined to form an ensemble of recommender systems in order to combat drawbacks of the individual systems. The data for the experiments was acquired from University Hospital Dresden. Collaborative filtering [73][74] is applied, where therapies are items and therapy responses are treated as user preferences.

Stark et al.(2019) [75] present a systematic literature review on recommender systems in medicine that covers existing systems and compares them on the basis of various features. Some interesting finds from the review include the following observations: (1) Most studies attempt to develop a general purpose recommender systems, i.e., one system for all diseases; (2) Disease specific systems focus on drug recommendation for diabetes. The review, points to several future research directions which include building a recommender system for recommending dosage of medicine and finding highly scalable solutions. Recommender systems can be used to suggest drugs for treatment. A popular commercial solution is IBM’s AI machine Watson Health [76] which is used by healthcare providers and researchers to make suitable decisions about providing treatment to patients based on insights from the system.

1.3.4. Automation of Treatment

In surgical area, research focus has been on automating tasks like surgical suturing, implantation, and biopsy procedures. Taylor et al. (2013) [77] present a broad overview of medical robot systems within the context of computer integrated surgery (CIS). This article also provides a high level-classification of such systems: (1) surgical CAD/CAM systems and (2) surgical assistants. The former refers to the process of pre-operative planning involving the analysis of medical images and other patient information to produce a model of the patient. This article presents examples of both kinds of robotic systems.

Mayer et al. (2008) [78] developed an experimental system for automating recurring tasks in minimal invasive surgery by extending the learning by demonstration paradigm [79] [80] [81]. The system consists of four robotic arms which can be equipped with minimally invasive instruments or a camera. The benchmark task selected for this work is minimally invasive knot-tying.

Moustris et al. (2011) [82] present a literature review of commercial medical systems and surgical procedures. This work solely focuses on systems that have been experimentally implemented on real robots. Automation has also been used for simulating treatment plans on virtual surrogates of patients called phantoms [83]. The phantoms represent the anatomy of a patient but they are too generic and hence cannot accurately represent individuals. These phantoms are especially used in pediatric

oncology to study the effects of radiation treatment and late adverse effects. Virgolin et al. (2020) [84] propose an approach to build automatic phantoms by combining machine learning with imaging data. The problem of structuring a pediatric phantom is divided into three prediction tasks: (1) prediction of a representative body segmentation, (2) prediction of center of mass of the organ at risk and (3) prediction of representative segmentation. Machine learning algorithms used for all three prediction tasks are least angle regression [85], least absolute shrinkage and selection operator [86], random forests (RF) [87], traditional genetic programming (GP-Trad)[88], and genetic programming - gene pool optimal mixing evolutionary algorithm [89].

1.3.5. Other Tasks in Integrative Medicine

The Consortium of Academic Health Centers for Integrative Medicine [90] defines the term *integrative medicine* as “an approach to the practice of medicine that makes use of the best-available evidence taking into account the whole person (body, mind, and spirit), including all aspects of lifestyle.” There are many definitions for integrative medicine in the literature, but all share the commonalities that reaffirm the importance of focusing on the whole person and lifestyle rather than just physical healing. According to Maizes, Rakel, and Niemiec (2009) [91], integrative medicine gained recognition due to the realization that people spend only a fraction of time on prevention of disease and maintaining good health. The authors present a data-driven example to promote the importance of integrative medicine - walking everyday for two hours for adults afflicted with diabetes reduces mortality by 39%. It is important to note that integrative medicine is not synonymous with complementary and alternative medicine (CAM) [92]. We refer interested readers to Baer (2004) [93] which chronicles the evolution of conventional and integrative medicine in the United States.

Complementary and alternative medicine (CAM) refers to medical products and practices that are not part of standard medical care. Ernst (2000) [94] presents examples of techniques used in complementary and alternative medicine which include, but are not limited to the following: acupuncture, aromatherapy, chiropractic, herbalism, homeopathy, massage, spiritual healing and traditional Chinese medicine.

Zhao et al. (2015) [95] present an overview of machine learning approaches used in traditional Chinese medicine (TCM). TCM specialists have established four diagnostic methods for TCM: observation, auscultation and olfaction, interrogation, and palpation. This article explains each of the four diagnostic methods and provides a list of machine learning methods used for each task. The most common methods are k-nearest neighbors (kNN) and support vector machines (SVM). Other methods include decision trees, Naïve Bayes (NB), and artificial neural networks (ANNs).

1.4. Open Research Problems

A recently published editorial by Suzanne Bakken [96] highlights five clinical informatics articles that reflect a consequentialist perspective. One of the articles that we discuss here focuses on a methodological concern, i.e., predictive model calibration [97]. Predictive models are an important research topic as discussed in Section 1.3.2., but many studies continue to focus on model discrimination rather than calibration. Ghassemi et al. (2020) [98] outline several promising research directions, specifically highlighting issues of data temporality, model interpretability, and learning appropriate representations. Machine learning models in most of the existing literature have been trained on large amount of historical data and fail to account for temporality of data in the medical domain, where patient symptoms and or treatment procedures change with time. The authors cite Google Flu Trends as an example of the need to update machine learning models to account for this data temporality, as it persistently overestimated flu [99]. Another promising research area is model interpretability [100] [101]. The authors suggest many directions towards the achievement of this goal: (1) model justification to justify the predictive path rather than just explaining a specific prediction; (2) building collaborative systems, where humans and machines work together. A final research topic is representation learning, which can improve predictive performance and account for conditional relationships of interest in the medical domain.

1.4.1. Learning for Classification and Regression

Classification is the identification of one or more categories or sub-populations to which a new observation belongs, on the basis of a training data set containing observations, or instances. In the data sciences of statistics and machine learning, classification may be supervised (where class labels are known) [102], unsupervised (where they are not and assignment is based on cohesion and similarity among instances) [103], or semi-supervised. Dreiseitl and Ohno-Machado (2002) [104] survey early work using logistic regression (LR, particularly the binomial logit model) and artificial neural networks (particularly multilayer perceptrons or MLP) for dichotomous classification, also known as binary classification or concept learning, on diagnostic and prognostic tasks from 72 papers in the existing literature. In parallel with this broad study of discriminative approaches to diagnosis and prognosis, Dybowski and Roberts (2005) [105] compiled a comprehensive anthology of probabilistic models primarily for generative classification.

In contrast with these broad surveys, which are included for completeness and historical breadth, application papers tend to focus on specific use cases for classification, such as prediction of mortality. Eftekhari et al. (2005) [106] is one such paper which addresses the task of predicting head trauma mortality rate based on initial clinical data, and focuses methodologically on logistic regression (LR) and multilayer percep-

trons (MLP), as do Dreiseitl and Ohno-Machado (2002) [104].

Regression is the problem of mapping an input instance to a real-valued scalar or tuple, which in data science is defined as an estimation task. In medical informatics, many predictive applications can be formulated as risk analysis tasks i.e. tasks requiring estimation of syndrome probability, given data from electronic medical records (EMR). Typical examples include estimating risk of a particular form of cancer, such as in a study by Ayer et al. (2010) [107], where they use LR and MLP to estimate risk of breast cancer. In some additional use cases, the predictive task requires estimation of a continuous value such as the size (widest diameter) of a cancer mass, rather than a probability of occurrence. Royston and Sauerbrei (2008) [108] is a methodological introduction to numerical estimation methods for such tasks.

1.4.2. *Learning to Act: Control and Planning*

Another general category of tasks falls under the rubric of *learning to act*, or intelligent control and planning in engineering terminology. This includes the application of machine learning to the overlapping subarea of *optimal control*, the branch of mathematical optimization that deals with maximizing an objective function such as cost-weighted proximity to a target.

One example of an optimal control task, which was investigated by Vogelzang, Zijlstra, and Nijsten (2005) [109] is maintaining a patient’s blood glucose level via automatic control of an insulin pump. The functional requirement of the system is to regulate the change in pump rate as a function of past pump rate, target glucose level, and past blood glucose measurements. The **G**lucose **R**egulation in **I**ntensive care **P**atients (**GRIP**) system developed by Vogelzang et al. used a fixed weighted optimal control function based on previous clinical studies [109]. Other optimal control tasks include prolonging the onset of drug resistance in treatment applications such as chemotherapy, a task studied by Ledzewicz and Schättler (2006) [110], who formulated a dynamical system for the development of drug resistance over time and applied ordinary differential equation (ODE) solvers to the task. Such numerical models can also be developed for therapeutic objectives such as minimizing tumor volume as a function of angiogenic inhibitors administered over time, an optimal control task studied by Ledzewicz et al. (2008) [111].

By formulating parametric models for problems such as maintaining a patient’s health care characteristics (e.g., blood glucose level, tumor size) within desired ranges while minimizing the total cost of doing so, optimization methods from industrial engineering and operations research, such as control charts, can be used. For example, Dobi and Zempléni (2019) apply Markov chain models and a variety of control charts to this cost-optimal control task [112].

Yet another family of intelligent control approaches originates from classical planning, particularly the inverse problem of *plan recognition* (mapping from observed

action sequences to individual plan steps, preconditions, and desired postconditions) and the problem of plan revision, which entails modifying a plan (such as a course of drug therapy) due to an identified complication (such as a toxic episode or other adverse reaction or interaction). Such systems are discussed by Shahar and Musen (1995) [113]. Plan revision may be necessitated as a consequence of historical observation (case studies), predictive simulation, or inference using a domain theory.

Finally, *enterprise resource planning* (ERP) is an integrative planning task of managing business processes (in healthcare, these include: sales and marketing, patient services, provider human resources, specialist referrals, procurement of equipment and materials, treatment, billing and insurance). Van Merode, Groothuis, and Hasman (2004) [114] survey ERP requirements and systems for hospitals.

1.4.3. Towards Greater Autonomy: Active Learning & Self-Supervision

Machine learning depends on availability of a training experience, but this experience need not come in the form of labeled data, which is expensive to acquire even with copious available resources such as expertise and non-expert annotator time, whose cost may be reduced by gamification or other means of crowd-sourcing to volunteers. In this section, we survey three species of learning without full supervision that help to free machine learning users from some aspects of these data requirements and other experiential requirements. These are:

1. *active learning* [115], the problem of developing a learning system that can seek out its own experiences
2. *transfer learning* [116], training on a set of experiences for one or more tasks or domains and using the resulting representations to facilitate learning, reasoning, and problem solving in new tasks and domains
3. *self-supervised learning* [117], the generalized task of learning with unlabeled data and inducing intrinsic labels by discovering relationships between sub-components of the training input (such as views or parts of an object in computer vision tasks)

Active learning in medical informatics spans a gamut of experiential domains from text to case studies, to controllers and policies. An example of active learning in text is the work of Druck, Settles, and McCallum (2009) [118], who apply categorical feature labels to words. This is a typical methodology in biomedical texts, where domain lexicons are organized into syndromic, pharmaceutical, and anatomical hierarchies, among others. Chen, Mani, and Xu (2012) [119] use active learning with labels on two text categorization tasks. The first of these is at the sentence level, on the ASSERTION data set, a clinical health care text corpus for the 2010 i2b2/VA natural language processing (NLP) challenge. The second is at the whole-document level, on the NOVA data set, an e-mail corpus with labels corresponding to a generic “religion vs. politics” topic classification task. In subsequent work, Chen et al. (2017) [120] applied a conditional random field (CRF)-based active learning system to the corpus of the 2010

i2b2/VA NLP challenge to show how annotation time could be reduced: by using latent Dirichlet allocation (LDA) for sentential topic modeling (sentence-level clustering) and by bootstrapping the process of set expansion for the named entity recognition (NER) task using active learning. Dligach, Miller, and Savova (2013) [121] also developed an active learning system for the i2b2 task, but focus on document-level phenotyping (pre-filtering of ICD-9 codes, CPT codes, laboratory results, medication orders, etc. followed by category labeling). In their work, a document consists of Electronic Health Records (EHR) and all associated data for a given patient, which may be generated at multiple stages of an admission and treatment workflow.

Transfer learning is typically defined as task-to-task or domain-to-domain but what constitutes a domain in medical informatics can vary. Wiens, Gutttag, and Horvitz (2014) [122] investigate inter-hospital transfer learning by training on subsets of 132853 admissions at three different hospitals among which 1348 positive cases of *Clostridium difficile* infection were diagnosed, to boost hospital-specific precision and recall as measured holistically using the area under the receiver operating characteristic curve (AUROC). As with active learning, transfer learning can be based on natural language features, typically at the word, sentence, or document level for medical informatics. For example, Lee, Dernoncourt, and Szolovits (2017) [123] outlined a neural network approach to de-identification in patient notes, which is an instance of NER and crucial for compliance with patient confidentiality laws such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States of America. They applied a long short-term memory (LSTM), a type of deep learning neural network for sequence modeling, to classify named entities that represented protected health information (PHI). The transfer learning task, training on a large labeled data set to a smaller one with fewer labels, is another case of inter-domain transfer. Yet another NER transfer learning problem for EHRs is defined and studied by Wang et al. (2018) [124], who apply a Bidirectional LSTM (Bi-LSTM) on a shared training corpus to create a shared representation (specifically, a word embedding) for text classification that is then fine-tuned for source and target domains by training a CRF model with labeled training data as available, to achieve label-aware NER. The experimental corpus, in this case, is a Chinese-language medical NER corpus (CM-NER) consisting of 1600 anonymized EHRs across four departments: Cardiology (500), Respiratory (500), Neurology (300) and Gastroenterology (300). Wang et al. demonstrate effective inter-departmental NER (domain-to-domain) transfer through experiments on CM-NER. [124] Similar deep learning approaches for natural language processing (NLP) are applied by Du et al. (2018) [125], who demonstrate transfer using recurrent neural networks (RNNs) from clinical notes on psychological stressors to tweets on Twitter, to detect posts by users at risk of suicide. New architectures for implementing transfer are demonstrated by Peng, Yan, and Lu (2019) [126], who use the deep bidirectional transformers BERT and ELMo to achieve cross-domain transfer among

ten benchmarking data sets from the Biomedical Language Understanding Evaluation (BLUE) compendium.

Self-supervised learning consists of generating labels by means of (1) comparing objects (e.g., clustering), (2) extracting relationships, or (3) designing experiments (especially model selection). The first approach applies similarity or distance metrics over entire instances (unlabeled examples) and has traditionally been similar to unsupervised learning for classification tasks in general, while the second involves using relational patterns and/or probabilistic inference over structured data models to capture new relationships, and the third involves generating multiple candidate models (by random sampling or parameter optimization methods such as identifying support vectors for large margin discriminative classifiers), and then applying model selection.

Hoffmann, Zhang, and Weld (2010) [127] take the second approach, introducing LUCHS, a self-supervised, relation-specific system for information extraction (IE) from text. Roller and Stevenson (2014) [128] also use a relation-specific, ontology-aware approach to relation extraction; rather than being based on lexicon expansion, however, it uses the curated Unified Medical Language System (UMLS), a biomedical knowledge base.

Stewart and Nejd1 (2011) [129] present an application of the third approach to the task of event detection in the domain of social media-based epidemic intelligence, where self-supervised learning consists of tokenizing text corpora, namely ProMED-Mail and World Health Organization (WHO) outbreak reports, to obtain bag of words (BOW or word vector space) embeddings. A support vector machine (SVM) classifier is then trained, to which the authors apply model selection, testing the result against an avian flu text corpus.

As Blendowski, Nickisich, and Heinrich (2019) [130] note, self-supervision in medical imaging applications is a necessity because of the comparatively high cost of supervision for medical images and video versus general computer vision and video. In medical domains, annotation may require specialization in radiology and other medical sub-disciplines, whereas generic images and videos may be annotated via micro work systems such as Amazon Mechanical Turk, or even via volunteer crowdsourcing. Blendowski et al. present a modern, deep learning-based approach to self-supervision, applying convolutional neural networks (CNNs) to capture 3-D context features.

2. Background

2.1. Diagnosis

2.1.1. *Diagnostic Classification and Regression Tasks*

Nadeem et al. (2020) [131] present a very comprehensive survey on classification of brain tumor. Jha et al. (2019) [132] evaluated 32 supervised learning methods across 17 classification data sets (in domains that include cancer, tumors, heart and liver

diseases) to determine that decision tree-based methods perform better than others on these data sets. Mostafa et al. (2018) [133] use three classification methods (decision trees, ANNs and Naïve Bayes) to determine the presence of Parkinson’s disease by using features extracted from human voice recordings, reaching a similar conclusion that decision trees performed best on this data set.

Polat and Güneş (2007) [134] present a binary classification task for categorizing breast cancer as malignant or benign. The authors use **Least Square Support Vector Machine (LS-SVM)** [65] for classification.

Another machine learning task used for diagnosis is regression. Kayaer and Yildirim (2003) [135] use general regression neural network (GRNN) [136] for diagnosing diabetes using the Pima Indian diabetes data set [137]. The results show that it performs better than standard multi-layer perceptron (MLP) and radial basis function (RBF) feedforward neural networks [138] that have been used by other studies using the same data set. Hannan, Manza, and Ramteke (2010) [139] have also used GRNN and RBF for heart disease diagnosis. Jeyaraj and Nadar (2019) [140] propose a regression based partitioned deep CNN for the classification of oral cancer as malignant or benign. The network obtains accuracy comparable to that of a human expert oncologist.

2.1.2. Diagnostic Policy-Learning Tasks

Yu, Liu, and Nemati (2019) [141] present the first comprehensive survey of reinforcement learning (RL) applications in healthcare. The aim of the survey is to provide the research community with an understanding of the foundations, methodologies, existing challenges, and recent applications of RL in healthcare domain. The range of applications vary from Dynamic Treatment Regimes (DTRs) in chronic diseases and critical care, automated clinical diagnosis, to other tasks like clinical resource allocation and scheduling.

Early RL systems for medical informatics were predominantly designed for medical image processing and analysis (as a specialized application of computer vision). Sabha, Tizhoosh, and Salama (2006) [142] developed such a system, based on Q-learning, for ultrasound image analysis, focusing on tasks such as local thresholding, feature extraction, and segmentation of organs (in this case, the prostate gland). In subsequent work, Sabha, Tizhoosh, and Salama (2007) [143] extended this Q-learning system for organ segmentation to an adversarial framework that they termed “Opposition-Based Learning”. This framework allowed for more flexible formulation of utility gradients for RL problems such as balancing exploration vs. exploitation.

Peng et al. (2018) [144] introduce REFUEL, a deep Q-network (DQN)-based system for reward shaping and feature construction (“rebuilding”) in differential diagnosis of diseases. Such systems are examples of reinforcement-based meta-learning and can potentially incorporate aspects of both self-supervision and active learning of representation. DQN has also been used by researchers such as Al and Yun (2019) [145]

to learn policies for recognizing anatomical landmarks in X-ray-based computerized tomography (CT) and magnetic resonance imaging (MRI) images. In addition, DQN has recently been applied to learn control policies for clinical decision support tasks such as guiding a healthcare professional or first responder, especially an emergency medical technician (EMT), in obtaining ultrasound images as a remote sensing step before administering treatment. Milletari, Birodkar, and Sofka (2019) present a novel application of this method to Point of Care Ultrasound (POCUS) for scanning the left ventricle of the heart. [146]

Utility functions for deep RL in medical informatics may be tied to anatomical mapping and other automation tasks of internal medicine. Examples of such mapping include the context maps of Tu and Bai (2009) [147], which are based on learning, i.e., parameter estimation, in Markov random fields (MRFs) and conditional random fields (CRFs). In this work, the authors train and apply inference using these models to solve high-level vision tasks such as image segmentation, configuration estimation (orientation), and region labeling of 3-D brain images.

A key family of RL applications is that of control policies for medical treatment. Weng et al. (2017) [148] describe a deep RL method using a sparse autoencoder for glycemic control in septic patients. While experimental validation of this system was performed using historical data, the RL framework presented can include medical devices and mixed-initiative systems.

Finally, RL can also be applied to interactive differential diagnosis using natural language, i.e., dialogue. Liu et al. (2018) [149] describe a dialogue-based system for disease phenotype identification (eliciting observable characteristics or traits of diseases, such as the presentation and development of symptoms, morphology, biochemical or physiological properties, or patient behavior). The dialogue policy formulated here is a Markov decision process (MDP). The RL problem is that of detecting symptoms of any or all four known pediatric diseases by simulating query-based dialogue using an annotated corpus. The authors show that DQN for dialogue outperforms SVM for supervised classification learning, random dialogue generation, and a rule-based dialogue agent.

2.1.3. Active, Transfer, and Self-Supervised Learning

Sánchez (2010) [150] proposes a Computer Aided Diagnosis (CAD) [151] system for Diabetic Retinopathy screening using active learning approaches. There are four components that constitute the DR screening process: quality verification, normal anatomy detection, bright lesion detection and red lesion detection [152]. The findings from these four components need to be fused in order to generate an outcome for a patient. The outcome is in the form of a likelihood that the patient will be referred to an ophthalmologist. The output from the four components of the DR screening are used to extract some features which are further used to train a kNN classifier. Active learning

is applied in the training phase to select an unlabeled sample from the pool of samples and pose a query to the expert in order to acquire a label for the sample. This is an iterative process which only stops when a stopping criteria has been reached. This work used two different query functions: uncertainty sampling [153] and query-by-bagging (QBB) sampling [154].

Apostolopoulos and Mpesiana (2020) [155] use a transfer learning approach for the classification of medical images to diagnose COVID-19. This study uses publicly available thoracic X-rays of healthy people as well as patients suffering from COVID-19 to build an automatic diagnostic system. The aim of their work is to evaluate the effectiveness of state-of-the-art pre-trained convolutional neural networks models for the diagnosis of COVID-19. The CNN used for this study include: VGG19 [156], MobileNetv2 [157], Inception [158], Xception [159], InceptionResNet v2 [160]. The study formulates the task as a multi-class classification problem with three classes: normal people, pneumonia patients, and COVID-19 patients. The study does accomplish its goals of establishing the benefits of transfer learning by using state-of-the-art CNN models.

Bai et al. (2019) [161] use a semi-supervised approach for learning features from unlabeled data for the task of cardiac MR image segmentation. This segmentation is important for characterising the function of the heart. The authors discuss the different angulated planes at which the MR images are acquired. For brevity, we refer the readers to [161]. Specific views of the scans and their labels have been traditionally used to train a network from scratch. The authors use a standard U-net architecture [162] with three variations to it. The results of their work show that by using self-supervised learning even a small data set is able to outperform a standard U-net that has been trained from scratch.

2.2. Predictive Analytics

2.2.1. Prediction by Classification and Regression

We start by discussing *prediction of disease susceptibility*: Kim and Kim (2018) [163] try to predict an individual’s susceptibility to cancer by using genomic data. The authors use kNN for building a multi-class classification model.

Next we move on to *prediction of survivability*: Choi, Han, and Park (2009) [164] propose a hybrid ANN and Bayesian network model to predict five year survival rates for breast cancer patients. The model combines the best of both worlds using black box ANN for their higher accuracy and Bayesian networks for their explainability.

The survivability of a cancer patient depends on the stage of cancer which is based on tumor size, location, spread and other factors. Machine learning models that predict survivability in breast cancer research usually use breast cancer stage as a feature for training the model. Kate and Nadig (2017) [165] refer to such a model as a joint model. Their work uses the SEER data set [166], which classifies cancer into four

stages: in-situ, localized, regional, and distant.

Mobadersany et al. (2018) [167] propose an approach called **Survival Convolutional Neural Network (SCNN)**, which uses a CNN integrated with cox survival analysis technique for the task of survivability prediction for patients with brain tumors. The CNN includes a Cox proportional hazards layer that models overall survival. The proposed approach surpasses the prognostic accuracy of human experts.

The final task that we discuss is *prediction of recurrence*: The task involves correctly predicting the recurrence of disease with a binary outcome. Abreu et al. (2016) [168] present a literature review to evaluate the performance of machine learning methods for the task of predicting breast cancer recurrence. The review covers literature during the years 2007 - 2014 and find that the key algorithms used for the task include: decision trees, logistic regression, ANN, NB, K-Means, Random Forests, and k-nearest neighbor. These algorithms are discussed in section 3. The authors outline a few challenges based on their review that make this task less popular: (1) Lack of publicly available data of reasonable size - Most of the studies have used local data sets usually with a small number of patients. (2) Data imbalance is not handled in most cases; (3) Feature selection was performed manually in most studies with the help of domain experts. There is no agreement on variables that are important for the study of breast cancer recurrence; (4) Accuracy as evaluation metric for classification performance which is not appropriate for imbalanced data sets, and (5) Lack of model interpretability of machine learning models. These challenges provide a scope for future research in the area. A more recent review was published by Zhu et al. (2020) [169] that covers the usage of deep learning for the task of cancer prognosis. The study also presents similar challenges encountered by deep learning models in the domain as were pointed by Abreu et al. (2016) [168]. Some of the challenges listed include: (1) availability of small data sets; (2) handling data imbalance; (3) handling sparse and missing data; (4) handling high dimensional sequencing data; and (5) need of researchers with expertise in both, machine learning and biomedical domain.

2.2.2. Learning to Predict from Reinforcements and by Supervision

The advent of institution-wide terascale to petascale data mining, considered “big data” as of 2020, has brought machine learning for predictive analytics to the fore in many clinical domains. Shah, Steyerberg, and Kent (2018) [170] is a commentary piece on the state of the field in data mining for predictive analytics in medical informatics. It cites the $CHA_2DS_2 - VASc$ score as an example of a predictive rule regarding the doubling of thromboembolic risk in atrial fibrillation as a function of congestive heart failure, hypertension, age, diabetes, and previous stroke or transient ischemic attack, citing it as a use case of predictive models with consequence for finely balanced treatment decisions, which the authors note are commonplace. Rajkumar, Dean, and Kohane (2019) [171] review a broader set of diagnostic and prognostic

applications, along with best practices for using machine learning as an augmentative technology for clinicians. Shameer et al. (2019) [172] survey the field of artificial intelligence (AI) in medicine even more broadly, discussing the species of machine learning surveyed in this chapter and the actionable products thereof, from classification rules and regression formulas to annotated images and policies from RL.

Many predictive analytics applications involve estimating risk of adverse effects, or detecting imminent adverse outcomes in time for preventative or anticipatory measures. For example, Kendale et al. (2018) [173] develop a logistic regression-based system for predicting hypotension in patients after surgical anesthesia.

Just as supervised learning finds precedent in rule-based expert systems for differential diagnosis, reinforcement learning finds precedent in both educational technology and early automation for the practice of internal medicine and general surgery. Examples are surveyed below in Section 4. One general use case of RL for predictive analytics is given by Khurana, Samulowitz, and Turaga (2018) [174], who cast feature engineering for predictive applications as an RL task where the policy is defined in terms of data transformations.

Taking such diverse use cases as a whole, a major consideration in medical predictive analytics, which overlaps with artificial intelligence, ethics, and society (AIES), as well as with AI safety and security, is how to determine appropriate regulatory standards of clinical benefit. Parikh, Overmeyer, and Navathe (2019) [175] discuss meaningful minimum standards of functionality and interoperability, in the context of endpoints (clinical outcomes), appropriate benchmarks, and the importance of associating predictive systems with interventions.

2.2.3. Transfer Learning in Prediction

Gliomas are a type of tumors that are found in the brain. They are of different types and are usually graded from I to IV with grade IV being the most aggressive type. MRI images are used for grading gliomas into two categories: lower grade glioma (LGG), which is grade II and III and higher grade glioma, which is a grade IV. Cabezas et al. (2018) [176] propose a model using CNN for classifying LGG and HGG by analyzing MRI images. Two CNN architectures were explored for this purpose, namely: AlexNet and GoogLeNet. These two architecture are discussed in section 3.3.2 of this article. The results indicate that with transfer learning and fine-tuning the performance improved for both deep learning architectures. Yang et al. (2018) [177] present a transfer learning approach to segment the gliomas and its sub-regions and use the results along with other clinical features to predict patient survival. The approach is divided into two main tasks, where the first task focuses on segmentation of glioma using a 3-D U-net[162]. The second task segments the tumor sub-regions using a small ensemble net [178]. Their work uses a VGG-16 network, which is discussed in section 3.3.2 of this article.

2.3. Therapy Recommendation

2.3.1. Supervised Therapy Recommender Systems

Therapy recommendation remains a common use case of supervised learning, beyond diagnosis. General-purpose models for classification such as nearest-neighbor and rule-based classification are often effective, particularly when explainability is desired. Zhang, Cao, Gross, and Zaiane (2013) [179] discuss supervised learning for therapy recommendation in the domain of physical therapy, where class imbalance is a frequent issue, and derive both a modified rule-learning algorithm (ARIPPER) and an application of the selective minority oversampling technique (SMOTE) for this task.

Mental health and well being are major health issues in the world. There is a growing research effort to alleviate the symptoms of depression. Rohani et al. (2020) [180] build a recommender system for the mental health domain. The authors cite the motivation behind such systems stemming from research that suggests that when participants regularly participate in a pleasant activity then it has a beneficial impact on the mental health of the individual [181] [182]. An effective treatment for depression has been the pleasant event scheduling (PES) system [183] that has been shown to be an effective treatment for depression. Wahle et al. (2016) [184] describe multi-user and personalized variants of an affective recommender system developed using data from two mobile health applications for both clinically depressed users and non-clinical users. Mood ratings for activities are predicted using trained Naïve Bayes (NB) and support vector machine (SVM) models. These classification-based predictions enable personalized treatment recommendations using a mobile app.

2.4. Automation of Treatment

2.4.1. Classification and Regression-Based Tasks

Learning in the presence of hybrid training data (consisting of continuous and discrete or nominal variables) is a frequent and typical necessity in medical AI applications. Schilling et al. [185] describe the use of Classification and Regression Trees (CART) to identify thresholds for cardiovascular disease. Such hybrid tasks can also be addressed using binary response models such as probit models, but as the authors note, three distinct strengths of CART are that: (1) its hierarchical structure makes it more human comprehensible than linear or logistic regression models; (2) it can capture non-linearity in response and some multivariate interactions; (3) it has more expressiveness than regression models.

For some predictive applications, accuracy, precision, and recall are of highest significance, and may be considered more important than explainability by diagnosticians. Taylor, Moore, Cheung, and Brandt (2018) [186] present a task where regression methods, gradient boosting improvements, and committee machines such as bagged random forests and Adaboost are highly effective. In some cases, association

rule mining and basic NLP methods suffice: De Silva et al. (2018) [187] describe a social media mining task, analytics of online social groups for cancer patients, where the technical objective is to extract information about user demographics in relation to terms indicating patient age, cancer stage, side effects reported, and sentiments.

2.4.2. RL for Automation

An essential task in automation of surgical tasks is optimal path planning to avoid collisions between tool and surrounding tissue before resection automation. Baek et al. (2018) [188] aim to create a global path to cut a tissue during surgery without colliding with the surrounding tissue. The generated path was simulated on APOLLON, which is a Single Incision Laparoscopic Surgery system developed by KAIST Telerobotics and Control Lab [189]. Their work uses a popular path planning method called probabilistic roadmap (RPM) [190], which creates a path from a static environment to a desired point and Q-learning [191], a RL technique.

2.4.3. Active Learning in Automation

In a hospital setting, hand hygiene is one of the most important factors for the prevention of infectious diseases. Monitoring hand hygiene could be vital in reducing any outbreak within an operating room (OR). Kim, Choi, and Kim (2020) [192] propose a fully augmented automatic hand hygiene monitoring tool for monitoring the anesthesiologists on operating room video. The aim is to identify the alcohol based hand rubbing actions of anesthesiologists in the OR pre-and post-surgery. The proposed approach uses a 3-D convolutional neural network for classification task with two classes: rubbing hands and other actions. The data was collected from a hospital over a span of four months from a single OR. Additional data was generated by simulating a situation in OR for synthetic data. The proposed approach uses I3D model, an Inception v1 architecture that inflates 2D convolutions into 3-D convolutions to train 3 models: I3D networks for RGB, I3D networks for optical flow inputs, and a joint model. Transfer learning approach is employed by pre-training the CNN on Kinetics-400 [193] data set. The I3D for RGB outperforms the other two models.

2.5. Integrating Medical Informatics and Health Informatics

This section surveys machine learning at the interface of medical informatics and health informatics, particularly clinical health and medical informatics (HMI).

2.5.1. Classification and Regression Tasks in HMI

Ralston et al. (2007) [194] discuss patient web services in integrated delivery system (IDS) portals, which provide access to EHRs as well as financing and delivery of other healthcare products such physician consultation, medical devices, prescription medicines. These comprise point of care (POC) and post-POC services, which serve

as both information retrieval and EHR compilation mechanisms that can produce data for subsequent machine learning and data science. Data integration is a key requirement of such systems [194]. As underscored by the work of Oztekin, Delen, and Kong (2009) [195] on predicting heart-lung transplant survival using a combination of case data, elicited subject matter expertise, and commonsense reasoning from a basic domain theory, the most effective learning strategy is sometimes to incorporate all available data sources and then use algorithms for feature selection and construction on these. Furthermore, Holzinger and Jurisca (2014) [196] advocate for an integrative and interactive approach to knowledge discovery and data mining in biomedical informatics (BMI) - specifically, one that is informed by objectives of both human-computer interaction (HCI) and knowledge discovery in databases (KDD).

2.5.2. Reinforcement Learning for HMI

Medical informatics is a field where medical practitioners have traditionally used mixed-initiative AI, especially human-in-the-loop systems. In such systems, performance elements of machine learning are used for recommendation and decision support rather than for full automation at the point of care (POC). Holzinger and Jurisca (2016) [197] discuss this general practice and its rationale in modern interactive POC systems, specifically when humans in the loop are beneficial. They note that there are specific usage contexts where humans are good at spotting irrelevant features, assessing novelty and anomaly in annotation, and behavioral modeling of adversaries in security and safety contexts (such as anonymity and privacy of patient data in HMI systems). This is an important consideration for mixed-initiative HMI because the interactive machine learning (iML) framework they advocate is rooted in RL.

RL is not limited to Q-learning and its dynamic programming relatives, the temporal differences and SARSA family of algorithms, but can include complex adaptive systems such as multi-armed bandits and contextual bandits. Yom-Tov et al. (2017) [198] explore the problem of generating physical activity reminders to diabetes patients using RL in a contextual bandit. They show a slightly positive slope in reduction of hemoglobin A1c (HbA1c) for experimental users of this system over a 6-month period, compared to a slightly negative slope for a control group.

Deep RL (DRL) systems for HMI often frame treatment tasks as optimal control problems as outlined in Section 1.4.2. Raghu et al. (2017) [199] give one example of such a DRL application: the administration of intravenous fluids and vasopressors in septic patients. Their DQN-based approach calculates discounted return from off-policy (historical) data rather than by directly controllable reinforcements. Liu et al. (2017) [200] apply DQN to a high-dimension space of (about 270) actions corresponding to medicines given to Acute Myeloid Leukemia (AML) patients who received hematopoietic cell transplantation (HCT), with the goal of preventing acute Graft Versus Host Disease (GVHD). The training data set consists of historical medical registry

data for 6021 AML patients who underwent HCT between 1995 and 2007; this data is sequential but asynchronous (with standard follow-up forms collected at 100 days, 6 months, 12 months, 2 years, and 4 years). The authors demonstrate value enhancement of up to 21.4% using DQN and note that a larger training data set compiled by the Center for International Blood and Marrow Transplant Research (CIBMTR) can potentially be used for DRL.

Inverse Reinforcement Learning (IRL) is the problem of inferring the reward function of an observed agent, given its behavior as reflecting a policy. Yu, Liu, and Zhao (2019) [201] address the IRL task of capturing the reward functions for mechanical ventilation and sedative dosing in intensive care units (ICUs) from the Medical Information Mart for Intensive Care (MIMIC III), an open data set containing records of demographics, vital signs, laboratory tests, diagnoses, and medications for nearly 40000 adult and 8000 neonatal ICU patients. The authors demonstrated that IRL from this critical care data is feasible using a Bayesian inverse Q-learning algorithm (fitted Q-iteration) applied to a Markov Decision Process (MDP) representation.

An MDP is also used by Yu et al. (2019) [202] in a direct RL application. Here, they incorporate causal factors between options of anti-HIV drugs and observed effects into a model-free policy gradient RL learning algorithm, to learn DTRs for human immunodeficiency virus (HIV). This approach is shown to facilitate direct learning of causal policy gradient (CPG) parameters without requiring a model-based intermediary, a finding which has potential relevance to DRL and IRL as well.

2.5.3. Self-Supervised, Transfer, and Active Learning in HMI

Qiu and Sun (2019) [203] apply this paradigm to RL in vision, for iterative refinement of tomographic images and in order to improve diagnostic image classification. The specific type of imaging is called *optical coherence tomography (OCT)*.

Self-supervision is also used in other deep learning representations. Among the earliest of these to be developed were convolutional neural networks, which in HMI are designed to perform visual phenotyping. Gildenblat and Klaiman (2019) [204] developed a siamese network for segmentation of pathological images (e.g., into regions corresponding to stromata, tumors infiltrating lymphocytes, blood vessels, fat, healthy tissue, necrosis, etc). Sarkar and Etemad (2020) [205] use a CNN to learn relevant patterns from electrocardiograms (ECG) for emotion recognition.

Another self-supervised learning representation is found in word embedding models in natural language processing (NLP). Meng et al. (2020) [206] use BERT to classify radiology reports by urgency, learning a contextual representation rather than using more traditional sentiment analysis or other methods requiring word-level grading or subjective annotation of the training corpus.

Yet another self-supervised learning architecture is the generative adversarial network (GAN), which has been applied to create artificial images (deepfakes) and achieve

style transfer from drawings and paintings to photographic images. Tachibana et al. (2020) [207] use a GAN to improve the electronic cleansing process to remove fecal artifacts in computerized tomography (CT) colonoscopy images.

3. Techniques for Machine Learning

3.1. Supervised, Unsupervised, and Semi-Supervised Learning

We provide a very brief description of each of the machine learning methods that have been used in the medical informatics literature.

3.1.1. Shallow

A *decision tree* [64] is a directed tree model that is used as a rule-induction system where each node represents a feature (uni-variate decision-tree) and following a path of features (represented as nodes in the tree) leads to a specific classification.

Naïve Bayes [46] is a probabilistic classifier used for binary or multi-class classification problems. It is based on the poor assumption of conditional independence among the features and uses Bayes’ theorem.

Artificial Neural Networks (ANNs) [56] are inspired by biological neural network and the brain’s ability to process massive amount of information in parallel that allows it to recognize and classify the world around it. Researchers introduced artificial neural network that can be created using a weighted directed graph to process information in parallel where nodes of the graph are connected together in a similar way the neurons in the brain are connected.

Support Vector Machines (SVMs) [65] use a mathematical function called kernel that transforms linearly inseparable data to linearly separable by finding the hyper-plane (decision boundary) that maximizes distance from data points on either side of the plane.

Random Forests [87] are initially a group of decision-trees where each tree randomly selects a sample from the data set. Once the trees (forest) have been built, each tree produces a class prediction, which is used in the vote for the most popular class. The majority vote is used as the final classification of an input example.

k Nearest Neighbor (kNN) [48] is a supervised, non-parametric and lazy learning algorithm that is used for classification and regression problems. A kNN implementation has 3 important components - training and test data set; an integer value K; and a distance metric such as Euclidean, Manhattan or Hamming.

3.1.2. Deep

Convolutional Neural Networks: Tajbakhsh et al. (2016) [208] explain CNNs as a special class of ANNs where each neuron from one layer does not fully-connect to all neurons in the next layer. CNNs work by using convolutional layers that are mainly for

detecting certain local features in all locations of their input images. A set of convolutional kernels is responsible for learning a set of local features within input images which results in a feature mapping. Anwar et al. (2018) [209] shows that CNNs are widely used in medical imaging which help to achieve the following tasks: Segmentation, Detection and Classification of Abnormality, Computer Aided Detection or Diagnosis, and Medical Image Retrieval.

A *U-Net* [162] is a CNN network that consists of a contracting path to capture context, and a symmetric expanding path that enables precise localization (class label is supposed to be assigned to each pixel). The architecture of the network resembles a U-shaped network of CNNs where the left-side of the U-shape consists of a contracting path and the right-side consists of an expansive path.

Stacked Denoising Auto Encoders (SDAE) [210] are multi-layered denoising auto-encoders that are stacked together for training purposes. These networks are trained by adding noise to the raw input, which is then fed to the first denoising auto-encoder in the stack. After minimizing the loss (convergence), the information in the hidden layer (latent features) is obtained and again noise is added, and the resulting features are fed to the next denoising auto-encoder in the stack. This process continues for all denoising auto-encoders in the stack.

Generative Adversarial Networks (GANs) [211] consist of two models - a generative model and a discriminative model, that are simultaneously trained. The generative model captures the data distribution by producing fake data, and the discriminative model determines if a sample comes from the real training data or from the generative model (fake data). The training goal is to maximize the probability of the discriminative model making a mistake.

Recurrent Neural Networks (RNNs) [212] are a special type of neural network where the output from previous step is fed as input to the current step. The most important feature of RNN is their internal states which allow the network to have the ability of remembering information about a sequence of inputs. This is especially useful when trying any kind of prediction that relies on information that can be viewed as a sequence, such as an EEG signal.

3.2. Reinforcement Learning

3.2.1. Traditional

Q-learning [191] is a simple reinforcement learning algorithm that given the current state, seeks to find the best action to take in that state. It is an off-policy algorithm because it learns from actions that are random, i.e., outside the policy. The algorithm works in three basic steps: (1) The agent starts in a state and takes an action and receives a reward; (2) For the next action, the agent has two choice - either reference the Q table and select an action with the highest value or take a random action; (3) Agent updates the Q values, i.e. $Q[\text{State}, \text{Action}]$. A Q-table is a reference table for

the agent to select the best action based on the Q-value.

3.2.2. Deep RL

The *Deep Q-Network (DQN) model* by Sorokin et al. [213] combine Q-learning (discussed above) with a deep CNN. The goal of this model is to train a network to approximate the value of the Q function which maps state-action pairs to their expected discounted return. The inputs to the neural network are the state variables and the outputs are the Q values.

3.3. Self-Supervised, Transfer, and Active Learning

3.3.1. Traditional

Although the term “self-supervised” dates back to learning systems that used shallow representations rather than deep neural networks and similar representations [127] [129] [128], we can see from the distinct purpose and methodology used to achieve self-supervision that deep learning has emerged as a paradigm that facilitates new forms of self-supervised learning. Transfer learning has similarly been studied extensively since the advent of deep learning, but compared to self-supervision, is better understood as an independent task. A broad and comprehensive overview of the problem is provided by Torrey and Shavlik (2010) [214].

3.3.2. Deep

AlexNet [215] is an 8-layered CNN network of 5 convolutional layers followed by 3 fully-connected layers with 60 million parameters and 650,000 neurons. The authors trained this network using a subset of the popular *ImageNet* image corpus.

VGG [156] is a very deep CNN for large-scale image recognition developed as part of the Large-Scale Visual Recognition Challenge (ILSVRC).

ResNets [216], or residual neural networks, are a type of deep neural net architecture modeled on pyramidal cells (high-connectivity cortical neurons). To alleviate the difficulty of training deeper neural networks, ResNets contain shortcut (skip) connections that propagate activations forward by one or more layers; these activation blocks are referred to as *residual blocks* and this structure reformulates a baseline plain network into its counterpart residual version. ResNets have a typical depth of 50 to over 150 layers (ResNet-152), which is very big compared to VGG.

SqueezeNet [217] is another CNN architecture which was mainly developed to achieve a similar accuracy to AlexNet, but with far less number of parameters and size. It maintains an accuracy on ImageNet data set that is comparable to that of AlexNet.

DenseNet [218] was developed based on the observation that CNNs can be deeper, more accurate, and very efficient to train if they contain shorter connections between layers close to the input and those close to the output. DenseNet consists of dense

blocks where each block connects each layer to every subsequent layer in that block. The advantages of DenseNet include avoiding the vanishing-gradient problem, strengthen feature propagation, feature reuse, and reduce the number of training parameters.

4. Applications

We now survey selected popular test beds for medical predictive analytics.

4.1. Test Beds for Diagnosis and Prognosis

The most used data set for breast cancer diagnosis is the Breast Cancer Wisconsin (Diagnostic) Data Set 2.1.1 [137]. This data set is publicly available from the University of California Irvine (UCI) Machine Learning Repository.

Data sets used for the task of predicting breast cancer recurrence [168] include: (1) The Wisconsin prognostic breast cancer (WPBC) data set from the UCI ML repository [137]; (2) the data set from the SEER database [166] [164].

A RNA-seq data set for cancer prediction [219], which is freely available from the The Cancer Genome Atlas (TCGA) program database [220]. TCGA is a cancer genomics program which has generated over 2.5 petabytes of genomic, epigenomic, transcriptomic, and proteomic data. This data has been used to facilitate improvements in the diagnosis, treatment, and prevention of cancer.

4.1.1. New Test Beds

We refer the interested reader to work by Deserno, Welter, and Horsch (2012) [221], Murphy et al. (2015) [222], Svensson-Ranallo, Adam, and Sainfort (2011) [223], and the guidelines of the National Heart, Lung, and Blood Institute (NHLBI) for preparation of clinical study data [224].

4.2. Test Beds for Therapy Recommendation and Automation

We refer the interested reader to relevant work by Calero Valdez et al. (2016) [70], Son and Thong (2015) [225], and Thong and Son (2015) [226].

4.2.1. Prescriptions

We refer the interested reader to work by Galeano and Paccanaro (2018) [227], Kushwaha et al. (2014) [228], Zhang et al. (2015) [229], Bao and Jiang (2016) [230], Bhat and Aishwarya (2013) [231], and Guo et al. (2016) [232].

4.2.2. Surgery

We refer the interested reader to relevant work by Ciecierski, Raś, and Przybyszewski (2012) [233], Petscharnig and Schöffmann (2017) [234], Wang and Fey (2018) [235], Shvets et al. (2018) [236].

Table 1.1 Classification of Skin Cancer: accuracy, precision, recall, and F1 score for deep learning neural networks

Network	Acc wtd.	Precision benign	Precision malignant	Recall benign	Recall malignant	F1 benign	F1 malignant
AlexNet	0.87	0.91	0.85	0.87	0.89	0.89	0.87
VGG11_bn	0.85	0.91	0.81	0.82	0.9	0.86	0.85
Resnet18	0.83	0.92	0.77	0.78	0.91	0.84	0.84
SqueezeNet	0.85	0.92	0.8	0.81	0.92	0.86	0.85
DenseNet	0.84	0.9	0.8	0.81	0.89	0.85	0.84

5. Experimental Results

5.1. Test Bed

As an experimental example, we include a test bed introduced by Fanconi (2019) [237] in this chapter. This consists of a subset of 3600 pictures extracted from the data set in the international skin imaging collaboration archive (ISIC), which contains more than 23,000 mole pictures. The test bed contains balanced cases of malignant (1) and benign (0) skin moles. Here we present a simple performance comparison between different pre-trained deep learning architectures (DLAs) in the task of binary classifying a skin mole as being a malignant or a benign mole.

This data set has been used by several recent publications[238] [239] [240] [241].

5.2. Results & Discussion

We use weighted accuracy, precision, recall, and F1-score to test the performance of the DLAs. Table 1.1 shows the performance of the different DLAs used in this comparison. In terms of weighted accuracy, AlexNet have achieved the best accuracy. SqueezeNet and Resnet-18 achieved the best precision score when classifying a mole as benign, but the worst precision score when classifying a mole as malignant. AlexNet has the best precision score when classifying a mole as malignant and the best F1-score as well. We can also see that the difference in the performance of the DLAs is only marginal and we believe that different data set would yield different performance between the tested DLAs.

6. Conclusion: Machine Learning for Computational Medicine

This chapter has introduced machine learning tasks for diagnostic medicine and automation of treatment, comprising supervised, self-supervised, and reinforcement learning. After surveying task definitions, we outlined existing methods for diagnosis, pre-

dictive analytics, therapy recommendation, automation of treatment, and integrative health and medical informatics (HMI). We now look forward to current and continuing work in applied machine learning for medical applications.

6.1. Frontiers: Pre-Clinical, Translational, and Clinical

Machine learning has developed into a cross-cutting technology across areas of medical informatics, from theory, education, and training (*pre-clinical*) aspects to the observation and treatment of patients (*clinical* computational medicine), to *translational* medicine, which aims towards bridging experimental tools and treatments to deployed ones used in clinical practice. For pre-clinical surveys, we refer the interested reader to the following: Prashanth et al. (2016), who discuss experimental Parkinson’s treatments [242]; Kolachalama and Garg (2018) survey machine learning as a subarea of artificial intelligence (AI), particularly in medical education [243]; and Bannach-Brown et al. (2018) [244] review natural language learning and text mining in medical informatics. Ravi et al. (2016) [245] survey deep learning in translational health informatics; Weintraub and Rumsfeld (2018) [246] review translational text analytics; and Shah et al. (2019) [247] broadly examine machine learning and AI in translational medicine, from sensors and medical devices to drug discovery. Finally, Savage (2012) [248] provides an early survey of big data for improving clinical medicine, while Char, Shah, and Magnus (2018) [249] address ethical considerations.

6.2. Towards the Future: Learning and Medical Automation

As of this writing in 2020, the coronavirus disease SARS-Cov-2 is of pervasive interest as a highly time-critical use case of intelligent systems for improved diagnosis and epidemiological modeling, as discussed by Randhawa et al. (2020) [250], and for understanding etiology towards vaccine discovery, as discussed by Alimadadi et al. (2020) [251]. The Stanford HAI conference organized by Altman (2020) [252] surveys broad methods for combatting Covid-19 and future pandemics. More generally, the fields of diagnostic medicine and predictive analytics have grown significantly since the early work of Kononenko (2001) [43] and the health informatics applications of Pakhomov, Buntrock, and Chute (2006) [253], but there are deep technical gaps and significant challenges remaining before the “eDoctor” envisioned by Handelman et al. (2018) [254] becomes a reality. Meanwhile, AI safety concerns, as mapped out by Yampolskiy (2016) [255] in both historical and anticipatory contexts of risks and remedies, loom large.

REFERENCES

1. Collen M, Origins of medical informatics. *The Western journal of medicine* 1986; 145(6):778–785.
2. Hasman A, Mantas J, Zarubina T, An abridged history of medical informatics education in europe. *Acta Informatica Medica* 2014; 22(1):25, doi:10.5455/aim.2014.22.25-36.
3. Thakkar M, Davis DC, Risks, barriers, and benefits of ehr systems: a comparative study based on size of hospital. *Perspectives in Health Information Management/AHIMA*, American Health Information Management Association 2006; 3.
4. Samuel AL, Some studies in machine learning using the game of checkers. *IBM Journal of research and development* 1959; 3(3):210–229, doi:10.1147/rd.33.0210.
5. Jordan MI, Mitchell TM, Machine learning: Trends, perspectives, and prospects. *Science* 2015; 349(6245):255–260, doi:10.1126/science.aaa8415.
6. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S, Dermatologist-level classification of skin cancer with deep neural networks. *nature* 2017; 542(7639):115–118, doi: <https://doi.org/10.1038/nature21056>.
7. Sidey-Gibbons JA, Sidey-Gibbons CJ, Machine learning in medicine: a practical introduction. *BMC medical research methodology* 2019; 19(1):64, doi:<https://doi.org/10.1186/s12874-019-0681-4>.
8. LeCun Y, Bengio Y, Hinton G, Deep learning. *nature* 2015; 521(7553):436–444, doi:<https://doi.org/10.1038/nature14539>.
9. Goodfellow I, Bengio Y, Courville A, Deep learning. MIT press, 2016. <http://www.deeplearningbook.org>.
10. Miotto R, Wang F, Wang S, Jiang X, Dudley JT, Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics* 2018; 19(6):1236–1246, doi:10.1093/bib/bbx044.
11. Kwak GHJ, Hui P, Deephealth: Deep learning for health informatics. *arXiv preprint arXiv:190900384* 2019; .
12. Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D, Early diagnosis of alzheimer’s disease with deep learning. In: 2014 IEEE 11th international symposium on biomedical imaging (ISBI), IEEE, 2014, pp. 1015–1018, doi:10.1109/ISBI.2014.6868045.
13. Schuster M, Paliwal KK, Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing* 1997; 45(11):2673–2681, doi:10.1109/78.650093.
14. Lawrence S, Giles CL, Tsoi AC, Back AD, Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks* 1997; 8(1):98–113, doi:10.1109/72.554195.
15. Nair V, Hinton GE, Rectified linear units improve restricted boltzmann machines. In: *ICML*, 2010, pp. 807–814.
16. Fischer A, Igel C, An introduction to restricted boltzmann machines. In: *Iberoamerican congress on pattern recognition*, Springer, 2012, pp. 14–36, doi:https://doi.org/10.1007/978-3-642-33275-3_2.
17. Baldi P, Autoencoders, unsupervised learning, and deep architectures. In: *Proceedings of ICML workshop on unsupervised and transfer learning*, 2012, pp. 37–49.
18. Baxter J, Learning internal representations. In: *Proceedings of the eighth annual conference on Computational learning theory*, 1995, pp. 311–320.
19. Sajda P, Machine learning for detection and diagnosis of disease. *Annu Rev Biomed Eng* 2006; 8:537–565, doi:10.1146/annurev.bioeng.8.061505.095802.
20. Chen JH, Asch SM, Machine learning and prediction in medicine—beyond the peak of inflated expectations. *The New England journal of medicine* 2017; 376(26):2507, doi:10.1056/NEJMp1702071.
21. Gräßer F, Beckert S, Küster D, Schmitt J, Abraham S, Malberg H, Zaunseder S, Therapy decision support based on recommender system methods. *Journal of healthcare engineering* 2017; 2017, doi:10.1155/2017/8659460.
22. Mayer H, Gomez F, Wierstra D, Nagy I, Knoll A, Schmidhuber J, A system for robotic heart surgery that learns to tie knots using recurrent neural networks. *Advanced Robotics* 2008; 22(13-14):1521–1537.
23. Kawanabe T, Kamarudin ND, Ooi CY, Kobayashi F, Mi X, Sekine M, Wakasugi A, Odaguchi H, Hanawa T, Quantification of tongue colour using machine learning in kampo medicine. *European Journal of Integrative Medicine* 2016; 8(6):932–941, doi:<https://doi.org/10.1016/j.eujim.2016.04.002>.

24. Healthcare. URL: <https://healthcare-analytics.soa.org/>.
25. World Health Organization, Diagnostic errors. World Health Organization 2016; .
26. Graber ML, Franklin N, Gordon R, Diagnostic error in internal medicine. *Archives of internal medicine* 2005; 165(13):1493–1499, doi:10.1001/archinte.165.13.1493.
27. Miller RA, Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. *Journal of the American Medical Informatics Association* 1994; 1(1):8–27, doi:10.1136/jamia.1994.95236141.
28. Shortliffe EH, Buchanan BG, Feigenbaum EA, Knowledge engineering for medical decision making: A review of computer-based clinical decision aids. *Proceedings of the IEEE* 1979; 67(9):1207–1224, doi:10.1109/PROC.1979.11436.
29. Giarratano JC, Riley G, Expert systems. PWS publishing co., 1998.
30. Shortliffe EH, Medical expert systems—knowledge tools for physicians. *Western Journal of Medicine* 1986; 145(6):830.
31. Shortliffe EH, Buchanan BG, Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project. Addison-Wesley Publishing Company, 1985.
32. Shortliffe E, Computer-based medical consultations: MYCIN, vol. 2. Elsevier, 2012.
33. Miller RA, Pople Jr HE, Myers JD, Internist-i, an experimental computer-based diagnostic consultant for general internal medicine. *New England Journal of Medicine* 1982; 307(8):468–476, doi:10.1056/NEJM198208193070803.
34. Miller RA, Masarie Jr FE, Quick medical reference (qmr): An evolving, microcomputer-based diagnostic decision-support program for general internal medicine. In: *Proceedings of the Annual Symposium on Computer Application in Medical Care*, American Medical Informatics Association, 1989, p. 947.
35. Rassinoux AM, Miller RA, Baud RH, Scherrer JR, Modeling principles for qmr medical findings. In: *Proceedings of the AMIA Annual Fall Symposium*, American Medical Informatics Association, 1996, p. 264.
36. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP, Dxplain: an evolving diagnostic decision-support system. *Jama* 1987; 258(1):67–74, doi:10.1001/jama.1987.03400010071030.
37. Dxplain. URL: <http://www.mghlcs.org/projects/dxplain>.
38. Bartold SP, Hannigan GG, Dxplain. *Journal of the Medical Library Association* 2002; 90(2):267.
39. Gaines BR, Knowledge acquisition: Past, present and future. *International Journal of Human-Computer Studies* 2013; 71(2):135–156, doi:https://doi.org/10.1016/j.ijhcs.2012.10.010.
40. Abu-Nasser B, Medical expert systems survey. *International Journal of Engineering and Information Systems (IJEAIS)* 2017; 1(7):218–224.
41. Salman FM, Abu-Naser SS, Expert system for covid-19 diagnosis. *International Journal of Academic Information Systems Research (IJASIR)* 2020; .
42. Almadhoun HR, Abu-Naser SS, An expert system for diagnosing coronavirus (covid-19) using sl5. *International Journal of Academic Engineering Research (IJAEER)* 2020; .
43. Kononenko I, Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine* 2001; 23(1):89–109, doi:https://doi.org/10.1016/S0933-3657(01)00077-X.
44. Kononenko I, Simec E, Induction of decision trees using relieff. In: *Proceedings of the ISSEK94 workshop on mathematical and statistical methods in artificial intelligence*, Springer, 1995, pp. 199–220, doi:https://doi.org/10.1007/978-3-7091-2690-5_14.
45. Ragavan H, Rendell LA, Lookahead feature construction for learning hard concepts. In: *ICML*, 1993, pp. 252–259, doi:10.1016/b978-1-55860-307-3.50039-3.
46. Rish I, et al., An empirical study of the naive bayes classifier. In: *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, vol. 3, 2001, pp. 41–46.
47. Kononenko I, Semi-naive bayesian classifier. In: *European Working Session on Learning*, Springer, 1991, pp. 206–219, doi:https://doi.org/10.1007/BFb0017015.
48. Dudani SA, The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics* 1976; SMC-6(4):325–327, doi:10.1109/tsmc.1976.5408784.
49. Weigend AS, Rumelhart DE, Huberman BA, Back-propagation, weight-elimination and time series

- prediction. *Connectionist Models* 1991; :105–116doi:10.1016/b978-1-4832-1448-1.50016-0.
50. Cruz JA, Wishart DS, Applications of machine learning in cancer prediction and prognosis. *Cancer informatics* 2006; 2:117693510600200030.
51. Ohno-Machado L, Modeling medical prognosis: survival analysis techniques. *Journal of biomedical informatics* 2001; 34(6):428–439, doi:10.1006/jbin.2002.1038.
52. Cantor A, et al., SAS survival analysis techniques for medical research. SAS Institute, 2003.
53. Prinja S, Gupta N, Verma R, Censoring in clinical trials: review of survival analysis techniques. *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine* 2010; 35(2):217, doi:10.4103/0970-0218.66859.
54. Cox DR, Oakes D, Analysis of survival data, vol. 21. CRC Press, 1984.
55. Kleinbaum DG, Dietz K, Gail M, Klein M, Klein M, Logistic regression. Springer, 2002.
56. Hassoun MH, et al., Fundamentals of artificial neural networks. MIT press, 1995.
57. Mendez-Tellez PA, Dorman T, Predicting patient outcomes, futility, and resource utilization in the intensive care unit: the role of severity scoring systems and general outcome prediction models. In: *Mayo Clinic Proceedings*, vol. 80, vol. 80, Elsevier, 2005, pp. 161–163, doi:https://doi.org/10.4065/80.2.161.
58. Rapsang AG, Shyam DC, Scoring systems in the intensive care unit: a compendium. *Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine* 2014; 18(4):220, doi:10.4103/0972-5229.130573.
59. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J, Mortality probability models (mpm ii) based on an international cohort of intensive care unit patients. *Jama* 1993; 270(20):2478–2486, doi:10.1001/jama.1993.03510200084037.
60. Le Gall JR, Lemeshow S, Saulnier F, A new simplified acute physiology score (saps ii) based on a european/north american multicenter study. *Jama* 1993; 270(24):2957–2963, doi:10.1001/jama.270.24.2957.
61. Knaus WA, Draper EA, Wagner DP, Zimmerman JE, Apache ii: a severity of disease classification system. *Critical care medicine* 1985; 13(10):818–829.
62. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A, et al., The apache iii prognostic system: risk prediction of hospital mortality for critically iii hospitalized adults. *Chest* 1991; 100(6):1619–1636, doi:10.1378/chest.100.6.1619.
63. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI, Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal* 2015; 13:8–17, doi:10.1016/j.csbj.2014.11.005.
64. Brodley CE, Utgoff PE, Multivariate decision trees. *Machine learning* 1995; 19(1):45–77, doi:https://doi.org/10.1023/A:1022607123649.
65. Suykens JA, Vandewalle J, Least squares support vector machine classifiers. *Neural processing letters* 1999; 9(3):293–300, doi:https://doi.org/10.1023/A:1018628609742.
66. Vapnik V, The nature of statistical learning theory. Springer science & business media, 2013.
67. Friedman N, Geiger D, Goldszmidt M, Bayesian network classifiers. *Machine learning* 1997; 29(2-3):131–163, doi:https://doi.org/10.1023/A:1007465528199.
68. Portugal I, Alencar P, Cowan D, The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications* 2018; 97:205–227, doi:https://doi.org/10.1016/j.eswa.2017.12.020.
69. Melville P, Sindhvani V, Recommender systems. *Encyclopedia of machine learning* 2010; 1:829–838, doi:10.1007/978-0-387-30164-8_705.
70. Valdez AC, Zieffle M, Verbert K, Felfernig A, Holzinger A, Recommender systems for health informatics: state-of-the-art and future perspectives. In: *Machine Learning for Health Informatics*, Springer, 2016; pp. 391–414, doi:10.1007/978-3-319-50478-0_20.
71. Wiesner M, Pfeifer D, Health recommender systems: concepts, requirements, technical basics and challenges. *International journal of environmental research and public health* 2014; 11(3):2580–2607, doi:10.3390/ijerph110302580.
72. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ, Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *Journal of the American Medical Informatics*

- Association 2006; 13(2):121–126.
73. Sarwar B, Karypis G, Konstan J, Riedl J, Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th international conference on World Wide Web, 2001, pp. 285–295, doi:<https://doi.org/10.1145/371920.372071>.
74. Su X, Khoshgoftaar TM, A survey of collaborative filtering techniques. *Advances in artificial intelligence* 2009; 2009, doi:<https://doi.org/10.1155/2009/421425>.
75. Stark B, Knahl C, Aydin M, Elish K, A literature review on medicine recommender systems. *International Journal of Advanced Computer Science and Applications* 2019; 10(8):6–13, doi: 10.14569/IJACSA.2019.0100802.
76. Ai healthcare solutions. URL: <https://www.ibm.com/watson-health>.
77. Taylor RH, Menciassi A, Fichtinger G, Fiorini P, Dario P, Medical robotics and computer-integrated surgery. In: Springer handbook of robotics, Springer, 2016; pp. 1657–1684, doi:https://doi.org/10.1007/978-3-319-32552-1_63.
78. Mayer H, Nagy I, Burschka D, Knoll A, Braun EU, Lange R, Bauernschmitt R, Automation of manual tasks for minimally invasive surgery. In: Fourth International Conference on Autonomic and Autonomous Systems (ICAS’08), IEEE, 2008, pp. 260–265, doi:10.1109/ICAS.2008.16.
79. Schaal S, Learning from demonstration. In: *Advances in neural information processing systems*, 1997, pp. 1040–1046.
80. Atkeson CG, Schaal S, Robot learning from demonstration. In: *ICML*, vol. 97, vol. 97, Citeseer, 1997, pp. 12–20.
81. Argall BD, Chernova S, Veloso M, Browning B, A survey of robot learning from demonstration. *Robotics and autonomous systems* 2009; 57(5):469–483, doi:<https://doi.org/10.1016/j.robot.2008.10.024>.
82. Moustris GP, Hiridis SC, Deliparaschos K, Konstantinidis K, Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature. *The international journal of medical robotics and computer assisted surgery* 2011; 7(4):375–392, doi:10.1002/rcs.408.
83. Xu XG, An exponential growth of computational phantom research in radiation protection, imaging, and radiotherapy: a review of the fifty-year history. *Physics in Medicine & Biology* 2014; 59(18):R233.
84. Virgolin M, Wang Z, Alderliesten T, Bosman PA, Machine learning for automatic construction of pediatric abdominal phantoms for radiation dose reconstruction. In: *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*, vol. 11318, vol. 11318, International Society for Optics and Photonics, 2020, p. 1131815, doi:<https://doi.org/10.1117/12.2548969>.
85. Efron B, Hastie T, Johnstone I, Tibshirani R, et al., Least angle regression. *The Annals of statistics* 2004; 32(2):407–499.
86. Tibshirani R, Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 1996; 58(1):267–288, doi:<https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.
87. Breiman L, Random forests. *Machine learning* 2001; 45(1):5–32, doi:<https://doi.org/10.1023/A:1010933404324>.
88. Koza JR, Genetic programming as a means for programming computers by natural selection. *Statistics and computing* 1994; 4(2):87–112, doi:<https://doi.org/10.1007/BF00175355>.
89. Virgolin M, Alderliesten T, Witteveen C, Bosman PA, Scalable genetic programming by gene-pool optimal mixing and input-space entropy-based building-block learning. In: *Proceedings of the Genetic and Evolutionary Computation Conference*, 2017, pp. 1041–1048, doi:<https://doi.org/10.1145/3071178.3071287>.
90. Introduction. 2020. URL: <https://imconsortium.org/about/introduction/>.
91. Maizes V, Rakel D, Niemiec C, Integrative medicine and patient-centered care. *Explore* 2009; 5(5):277–289, doi:10.1016/j.explore.2009.06.008.
92. Snyderman R, Weil AT, Integrative medicine: bringing medicine back to its roots. *Archives of Internal medicine* 2002; 162(4):395–397, doi:10.1001/archinte.162.4.395.
93. Baer HA, Toward an integrative medicine: Merging alternative therapies with biomedicine. Rowman Altamira, 2004.

94. Ernst E, The role of complementary and alternative medicine. *Bmj* 2000; 321(7269):1133, doi: 10.1136/bmj.321.7269.1133.
95. Zhao C, Li GZ, Wang C, Niu J, Advances in patient classification for traditional chinese medicine: a machine learning perspective. *Evidence-based complementary and alternative medicine: eCAM* 2015; 2015, doi:10.1155/2015/376716.
96. Bakken S, Hot topics in clinical informatics. 2020, doi:10.1093/jamia/ocaa025. URL: <https://doi.org/10.1093/jamia/ocaa025>.
97. Vaicenavicius J, Widmann D, Andersson C, Lindsten F, Roll J, Schön TB, Evaluating model calibration in classification. *arXiv preprint arXiv:190206977* 2019; .
98. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R, A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings* 2020; 2020:191.
99. Lazer D, Kennedy R, King G, Vespignani A, The parable of google flu: traps in big data analysis. *Science* 2014; 343(6176):1203–1205, doi:10.1126/science.1248506.
100. Ahmad MA, Eckert C, Teredesai A, Interpretable machine learning in healthcare. In: *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*, 2018, pp. 559–560, doi:<https://doi.org/10.1145/3233547.3233667>.
101. Chakraborty S, Tomsett R, Raghavendra R, Harborne D, Alzantot M, Cerutti F, Srivastava M, Preece A, Julier S, Rao RM, et al., Interpretability of deep learning models: a survey of results. In: *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, IEEE, 2017, pp. 1–6, doi:10.1109/UIC-ATC.2017.8397411.
102. Caruana R, Niculescu-Mizil A, An empirical comparison of supervised learning algorithms. In: *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 161–168, doi: <https://doi.org/10.1145/1143844.1143865>.
103. Ghahramani Z, Unsupervised learning. In: *Summer School on Machine Learning*, Springer, 2003, pp. 72–112.
104. Dreiseitl S, Ohno-Machado L, Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics* 2002; 35(5-6):352–359, URL: [https://doi.org/10.1016/S1532-0464\(03\)00034-0](https://doi.org/10.1016/S1532-0464(03)00034-0), doi:10.1016/S1532-0464(03)00034-0. URL: [https://doi.org/10.1016/S1532-0464\(03\)00034-0](https://doi.org/10.1016/S1532-0464(03)00034-0).
105. Dybowski R, Roberts S, An anthology of probabilistic models for medical informatics. In: *Probabilistic Modeling in Bioinformatics and Medical Informatics*, Springer, 2005; pp. 297–349, doi: 10.1007/1-84628-119-9_10. https://doi.org/10.1007/1-84628-119-9_10.
106. Eftekhari B, Mohammad K, Ardebili HE, Ghodsi M, Ketabchi E, Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC medical informatics and decision making* 2005; 5(1):1–8, URL: <https://doi.org/10.1186/1472-6947-5-3>, doi:10.1186/1472-6947-5-3. URL: <https://doi.org/10.1186/1472-6947-5-3>.
107. Ayer T, Chhatwal J, Alagoz O, Kahn Jr CE, Woods RW, Burnside ES, Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Radiographics* 2010; 30(1):13–22, URL: <https://doi.org/10.1148/rg.301095057>, doi:10.1148/rg.301095057. URL: <https://doi.org/10.1148/rg.301095057>.
108. Royston P, Sauerbrei W, Multivariable model-building: a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables, vol. 777. John Wiley & Sons, 2008, doi:10.1002/9780470770771.
109. Vogelzang M, Zijlstra F, Nijsten MW, Design and implementation of grip: a computerized glucose control system at a surgical intensive care unit. *BMC medical informatics and decision making* 2005; 5(1):38, URL: <https://doi.org/10.1186/1472-6947-5-38>, doi:10.1186/1472-6947-5-38. URL: <https://doi.org/10.1186/1472-6947-5-38>.
110. Ledzewicz U, Schättler H, Drug resistance in cancer chemotherapy as an optimal control problem. *Discrete & Continuous Dynamical Systems-B* 2006; 6(1):129, doi:10.3934/dcdsb.2006.6.129.

111. Ledzewicz U, Marriott J, Maurer H, Schättler H, The scheduling of angiogenic inhibitors minimizing tumor volume. *Journal of Medical Informatics & Technologies* 2008; 12, doi:10.3934/dcdsb.2009.12.415.
112. Dobi B, Zempléni A, Markov chain-based cost-optimal control charts for health care data. *Quality and Reliability Engineering International* 2019; 35(5):1379–1395, URL: <https://doi.org/10.1002/qre.2518>, doi:doi.org/10.1002/qre.2518. URL: <https://doi.org/10.1002/qre.2518>.
113. Shahar Y, Musen MA, Plan recognition and revision in support of guideline-based care. In: *Working notes of the AAAI Spring Symposium on Representing Mental States and Mechanisms*, 1995, pp. 118–126.
114. van Merode GG, Groothuis S, Hasman A, Enterprise resource planning for hospitals. *International journal of medical informatics* 2004; 73(6):493–501, URL: <https://doi.org/10.1016/j.ijmedinf.2004.02.007>, doi:10.1016/j.ijmedinf.2004.02.007. URL: <https://doi.org/10.1016/j.ijmedinf.2004.02.007>.
115. Settles B. Active learning literature survey. tech. rep., University of Wisconsin-Madison Department of Computer Sciences, 2009.
116. Pan SJ, Yang Q, A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 2009; 22(10):1345–1359, doi:10.1109/TKDE.2009.191.
117. Ross T, Zimmerer D, Vemuri A, Isensee F, Wiesenfarth M, Bodenstedt S, Both F, Kessler P, Wagner M, Müller B, et al., Exploiting the potential of unlabeled endoscopic video data with self-supervised learning. *International journal of computer assisted radiology and surgery* 2018; 13(6):925–933, doi:<https://doi.org/10.1007/s11548-018-1772-0>.
118. Druck G, Settles B, McCallum A, Active learning by labeling features. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1, EMNLP '09*, USA: Association for Computational Linguistics, 2009, p. 81–90. ISBN 9781932432596.
119. Chen Y, Mani S, Xu H, Applying active learning to assertion classification of concepts in clinical text. *Journal of biomedical informatics* 2012; 45(2):265–272, URL: <https://doi.org/10.1016/j.jbi.2011.11.003>, doi:doi.org/10.1016/j.jbi.2011.11.003. URL: <https://doi.org/10.1016/j.jbi.2011.11.003>.
120. Chen Y, Lask TA, Mei Q, Chen Q, Moon S, Wang J, Nguyen K, Dawodu T, Cohen T, Denny JC, et al., An active learning-enabled annotation system for clinical named entity recognition. *BMC medical informatics and decision making* 2017; 17(2):35–44, URL: <https://doi.org/10.1186/s12911-017-0466-9>, URL: <https://doi.org/10.1186/s12911-017-0466-9>.
121. Dligach D, Miller T, Savova G, Active learning for phenotyping tasks. In: *Proceedings of the Workshop on NLP for Medicine and Biology associated with RANLP 2013*, Hissar, Bulgaria: INCOMA Ltd. Shoumen, BULGARIA, 2013, pp. 1–8. URL: <https://www.aclweb.org/anthology/W13-5101>.
122. Wiens J, Gutttag J, Horvitz E, A study in transfer learning: leveraging data from multiple hospitals to enhance hospital-specific predictions. *Journal of the American Medical Informatics Association* 2014; 21(4):699–706, URL: <https://doi.org/10.1136/amiajnl-2013-002162>, doi:10.1136/amiajnl-2013-002162. URL: <https://doi.org/10.1136/amiajnl-2013-002162>, <https://academic.oup.com/jamia/article-pdf/21/4/699/5992536/21-4-699.pdf>.
123. Lee JY, Dernoncourt F, Szolovits P, Transfer learning for named-entity recognition with neural networks. *arXiv preprint arXiv:170506273* 2017; .
124. Wang Z, Qu Y, Chen L, Shen J, Zhang W, Zhang S, Gao Y, Gu G, Chen K, Yu Y, Label-aware double transfer learning for cross-specialty medical named entity recognition. *arXiv preprint arXiv:180409021* 2018; .
125. Du J, Zhang Y, Luo J, Jia Y, Wei Q, Tao C, Xu H, Extracting psychiatric stressors for suicide from social media using deep learning. *BMC medical informatics and decision making* 2018; 18(2):43, URL: <https://doi.org/10.1186/s12911-018-0632-8>, doi:10.1186/s12911-018-0632-8. URL: <https://doi.org/10.1186/s12911-018-0632-8>.
126. Peng Y, Yan S, Lu Z, Transfer learning in biomedical natural language processing: An evaluation of bert and elmo on ten benchmarking datasets. *arXiv preprint arXiv:190605474* 2019; doi:10.18653/v1/w19-5006.

127. Hoffmann R, Zhang C, Weld DS, Learning 5000 relational extractors. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden: Association for Computational Linguistics, 2010, pp. 286–295. URL: <https://www.aclweb.org/anthology/P10-1030>.
128. Roller R, Stevenson M, Self-supervised relation extraction using umls. In: International Conference of the Cross-Language Evaluation Forum for European Languages, Springer, 2014, pp. 116–127. URL: https://doi.org/10.1007/978-3-319-11382-1_12.
129. Stewart A, Diaz-Aviles E, Nanopoulos A, Self-supervised detection of disease reporting events in outbreak reports. In: 2011 IEEE International Conference on Information Reuse Integration, 2011, pp. 416–421.
130. Blendowski M, Nickisch H, Heinrich MP, How to learn from unlabeled volume data: self-supervised 3d context feature learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 649–657. URL: https://doi.org/10.1007/978-3-030-32226-7_72.
131. Nadeem MW, Ghamdi MAA, Hussain M, Khan MA, Khan KM, Almotiri SH, Butt SA, Brain tumor analysis empowered with deep learning: A review, taxonomy, and future challenges. *Brain Sciences* 2020; 10(2):118, doi:10.3390/brainsci10020118.
132. Jha SK, Pan Z, Elahi E, Patel N, A comprehensive search for expert classification methods in disease diagnosis and prediction. *Expert Systems* 2019; 36(1):e12343, doi:<https://doi.org/10.1111/exsy.12343>.
133. Mostafa SA, Mustapha A, Khaleefah SH, Ahmad MS, Mohammed MA, Evaluating the performance of three classification methods in diagnosis of parkinson’s disease. In: International Conference on Soft Computing and Data Mining, Springer, 2018, pp. 43–52, doi:https://doi.org/10.1007/978-3-319-72550-5_5.
134. Polat K, Güneş S, Breast cancer diagnosis using least square support vector machine. *Digital signal processing* 2007; 17(4):694–701, doi:<https://doi.org/10.1016/j.dsp.2006.10.008>.
135. Kayaer K, Yildirim T, et al., Medical diagnosis on pima indian diabetes using general regression neural networks. In: Proceedings of the international conference on artificial neural networks and neural information processing (ICANN/ICONIP), vol. 181, vol. 181, 2003, p. 184.
136. Specht DF, et al., A general regression neural network. *IEEE transactions on neural networks* 1991; 2(6):568–576, doi:10.1109/72.97934.
137. [archive.ics.uci.edu](http://archive.ics.uci.edu/ml). URL: <http://archive.ics.uci.edu/ml>.
138. Broomhead D, Lowe D, Radial basis functions, multi-variable functional interpolation and adaptive networks. *ROYAL SIGNALS AND RADAR ESTABLISHMENT MALVERN (UNITED KINGDOM)* 1988; RSRE-MEMO-4148.
139. Hannan SA, Manza R, Ramteke R, Generalized regression neural network and radial basis function for heart disease diagnosis. *International Journal of Computer Applications* 2010; 7(13):7–13, doi:10.5120/1325-1799.
140. Jeyaraj PR, Nadar ERS, Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *Journal of cancer research and clinical oncology* 2019; 145(4):829–837.
141. Yu C, Liu J, Nemati S, Reinforcement learning in healthcare: A survey. *arXiv preprint arXiv:190808796* 2019; .
142. Sahba F, Tizhoosh HR, Salama MM, A reinforcement learning framework for medical image segmentation. In: The 2006 IEEE International Joint Conference on Neural Network Proceedings, IEEE, 2006, pp. 511–517.
143. Sahba F, Tizhoosh HR, Salama MM, Application of opposition-based reinforcement learning in image segmentation. In: 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing, IEEE, 2007, pp. 246–251.
144. Peng YS, Tang KF, Lin HT, Chang E, Refuel: Exploring sparse features in deep reinforcement learning for fast disease diagnosis. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors, *Advances in Neural Information Processing Systems* 31, Curran Associates, Inc., 2018; pp. 7322–7331. <http://papers.nips.cc/paper/7962-refuel-explo>

- ring-sparse-features-in-deep-reinforcement-learning-for-fast-disease-diagnosis.pdf.
145. Al WA, Yun ID, Partial policy-based reinforcement learning for anatomical landmark localization in 3d medical images. *IEEE Transactions on Medical Imaging* 2019; 39(4):1245–1255.
 146. Milletari F, Brodukar V, Sofka M, Straight to the point: reinforcement learning for user guidance in ultrasound. In: *Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis*, Springer, 2019; pp. 3–10.
 147. Tu Z, Bai X, Auto-context and its application to high-level vision tasks and 3d brain image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 2009; 32(10):1744–1757.
 148. Weng WH, Gao M, He Z, Yan S, Szolovits P, Representation and reinforcement learning for personalized glycemic control in septic patients. *arXiv preprint arXiv:171200654* 2017; .
 149. Liu Q, Wei Z, Peng B, Tou H, Chen T, Huang XJ, Wong KF, Dai X, Task-oriented dialogue system for automatic diagnosis. In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2018, pp. 201–207.
 150. Sánchez CI, Niemeijer M, Abramoff MD, van Ginneken B, Active learning for an efficient training strategy of computer-aided diagnosis systems: application to diabetic retinopathy screening. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2010, pp. 603–610, doi:https://doi.org/10.1007/978-3-642-15711-0_75.
 151. Castellino RA, Computer aided detection (cad): an overview. *Cancer Imaging* 2005; 5(1):17.
 152. Niemeijer M, Abramoff MD, Van Ginneken B, Information fusion for diabetic retinopathy cad in digital color fundus photographs. *IEEE transactions on medical imaging* 2009; 28(5):775–785, doi:10.1109/TMI.2008.2012029.
 153. Lewis DD, Gale WA, A sequential algorithm for training text classifiers. In: *SIGIR’94*, Springer, 1994, pp. 3–12, doi:https://doi.org/10.1007/978-1-4471-2099-5_1.
 154. Abe N, Query learning strategies using boosting and bagging. *Proc of 15th Int Conf on Machine Learning (ICML98)* 1998; :1–9.
 155. Apostolopoulos ID, Mpesiana TA, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine* 2020; :1doi:<https://doi.org/10.1007/s13246-020-00865-4>.
 156. Simonyan K, Zisserman A, Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556* 2014; .
 157. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC, Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520, doi:10.1109/CVPR.2018.00474.
 158. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A, Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9, doi:10.1109/CVPR.2015.7298594.
 159. Chollet F, Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258, doi:10.1109/CVPR.2017.195.
 160. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA, Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-first AAAI conference on artificial intelligence*, 2017, pp. 4278–4284.
 161. Bai W, Chen C, Tarroni G, Duan J, Guitton F, Petersen SE, Guo Y, Matthews PM, Rueckert D, Self-supervised learning for cardiac mr image segmentation by anatomical position prediction. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 541–549, doi:https://doi.org/10.1007/978-3-030-32245-8_60.
 162. Ronneberger O, Fischer P, Brox T, U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241, doi:https://doi.org/10.1007/978-3-319-24574-4_28.
 163. Kim BJ, Kim SH, Prediction of inherited genomic susceptibility to 20 common cancer types by a supervised machine-learning method. *Proceedings of the National Academy of Sciences* 2018; 115(6):1322–1327.

164. Choi JP, Han TH, Park RW, A hybrid bayesian network model for predicting breast cancer prognosis. *Journal of Korean Society of Medical Informatics* 2009; 15(1):49–57, doi:10.4258/jksmi.2009.15.1.49.
165. Kate RJ, Nadig R, Stage-specific predictive models for breast cancer survivability. *International journal of medical informatics* 2017; 97:304–311, doi:10.1016/j.ijmedinf.2016.11.001.
166. Seer incidence database - seer data & software. URL: <https://seer.cancer.gov/data/>.
167. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Vega JEV, Brat DJ, Cooper LA, Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences* 2018; 115(13):E2970–E2979, doi:<https://doi.org/10.1073/pnas.1717139115>.
168. Abreu PH, Santos MS, Abreu MH, Andrade B, Silva DC, Predicting breast cancer recurrence using machine learning techniques: a systematic review. *ACM Computing Surveys (CSUR)* 2016; 49(3):1–40, doi:<https://doi.org/10.1145/2988544>.
169. Zhu W, Xie L, Han J, Guo X, The application of deep learning in cancer prognosis prediction. *Cancers* 2020; 12(3):603, doi:10.3390/cancers12030603.
170. Shah ND, Steyerberg EW, Kent DM, Big data and predictive analytics: recalibrating expectations. *Jama* 2018; 320(1):27–28, doi:10.1001/jama.2018.5602.
171. Rajkomar A, Dean J, Kohane I, Machine learning in medicine. *New England Journal of Medicine* 2019; 380(14):1347–1358, doi:10.1056/NEJMr1814259.
172. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP, Machine learning in cardiovascular medicine: are we there yet? *Heart* 2018; 104(14):1156–1164, doi:10.1136/heartjnl-2017-311198.
173. Kendale S, Kulkarni P, Rosenberg AD, Wang J, Supervised machine-learning predictive analytics for prediction of postinduction hypotension. *Anesthesiology: The Journal of the American Society of Anesthesiologists* 2018; 129(4):675–688, doi:10.1097/ALN.0000000000002374.
174. Khurana U, Samulowitz H, Turaga D, Feature engineering for predictive modeling using reinforcement learning. *arXiv preprint arXiv:170907150* 2017; .
175. Parikh RB, Obermeyer Z, Navathe AS, Regulation of predictive analytics in medicine. *Science* 2019; 363(6429):810–812, URL: <https://science.sciencemag.org/content/363/6429/810>, doi:10.1126/science.aaw0029. URL: <https://science.sciencemag.org/content/363/6429/810>, full.pdf.
176. Cabezas M, Valverde S, González-Villà S, Clérigues A, Salem M, Kushibar K, Bernal J, Oliver A, Lladó X, Survival prediction using ensemble tumor segmentation and transfer learning. *arXiv preprint arXiv:181004274* 2018; .
177. Yang Y, Yan LF, Zhang X, Han Y, Nan HY, Hu YC, Hu B, Yan SL, Zhang J, Cheng DL, et al., Glioma grading on conventional mr images: a deep learning study with transfer learning. *Frontiers in neuroscience* 2018; 12:804, doi:10.3389/fnins.2018.00804.
178. Kamnitsas K, Bai W, Ferrante E, McDonagh S, Sinclair M, Pawlowski N, Rajchl M, Lee M, Kainz B, Rueckert D, et al., Ensembles of multiple models and architectures for robust brain tumour segmentation. In: *International MICCAI Brainlesion Workshop*, Springer, 2017, pp. 450–462, doi: https://doi.org/10.1007/978-3-319-75238-9_38.
179. Zhang J, Cao P, Gross DP, Zaiane OR, On the application of multi-class classification in physical therapy recommendation. *Health Information Science and Systems* 2013; 1(1):15, doi:10.1186/2047-2501-1-15.
180. Rohani DA, Springer A, Hollis V, Bardram JE, Whittaker S, Recommending activities for mental health and well-being: Insights from two user studies. *IEEE Transactions on Emerging Topics in Computing* 2020; doi:10.1109/TETC.2020.2972007.
181. MacPhillamy DJ, Lewinsohn PM, The pleasant events schedule: Studies on reliability, validity, and scale intercorrelation. *Journal of Consulting and Clinical Psychology* 1982; 50(3):363, doi: <https://doi.org/10.1037/0022-006X.50.3.363>.
182. Fredrickson BL, Cultivating positive emotions to optimize health and well-being. *Prevention & treatment* 2000; 3(1):1a, doi:<https://doi.org/10.1037/1522-3736.3.1.31a>.
183. Lewinsohn PM, Libet J, Pleasant events, activity schedules, and depressions. *Journal of abnormal*

- psychology 1972; 79(3):291, doi:10.1037/h0033207.
184. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S, Mobile sensing and support for people with depression: a pilot trial in the wild. *JMIR mHealth and uHealth* 2016; 4(3):e1111, doi:10.2196/mhealth.5960.
185. Schilling C, Mortimer D, Dalziel K, Heeley E, Chalmers J, Clarke P, Using classification and regression trees (cart) to identify prescribing thresholds for cardiovascular disease. *Pharmacoeconomics* 2016; 34(2):195–205, doi:10.1007/s40273-015-0342-3.
186. Taylor RA, Moore CL, Cheung KH, Brandt C, Predicting urinary tract infections in the emergency department with machine learning. *PloS one* 2018; 13(3):e0194085, doi:10.1371/journal.pone.0194085.
187. De Silva D, Ranasinghe W, Bandaragoda T, Adikari A, Mills N, Iddamalgoda L, Alahakoon D, Lawrentschuk N, Persad R, Osipov E, et al., Machine learning to support social media empowered patients in cancer care and cancer treatment decisions. *PloS one* 2018; 13(10):e0205855, doi:10.1371/journal.pone.0205855.
188. Baek D, Hwang M, Kim H, Kwon DS, Path planning for automation of surgery robot based on probabilistic roadmap and reinforcement learning. In: 2018 15th International Conference on Ubiquitous Robots (UR), IEEE, 2018, pp. 342–347, doi:10.1109/URAI.2018.8441801.
189. Medical robot — kaist telerobotics and control lab. URL: <http://robot.kaist.ac.kr/medical-robots/>.
190. Kavraki LE, Kolountzakis MN, Latombe JC, Analysis of probabilistic roadmaps for path planning. *IEEE Transactions on Robotics and Automation* 1998; 14(1):166–171, doi:10.1109/70.660866.
191. Watkins CJ, Dayan P, Q-learning. *Machine learning* 1992; 8(3-4):279–292.
192. Kim M, Choi J, Kim N, Fully automated hand hygiene monitoring\ in operating room using 3d convolutional neural network. *arXiv preprint arXiv:200309087* 2020; .
193. Carreira J, Zisserman A, Quo vadis, action recognition? a new model and the kinetics dataset. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6299–6308, doi:10.1109/CVPR.2017.502.
194. Ralston JD, Carrell D, Reid R, Anderson M, Moran M, Hereford J, Patient Web Services Integrated with a Shared Medical Record: Patient Use and Satisfaction. *Journal of the American Medical Informatics Association* 2007; 14(6):798–806, URL: <https://doi.org/10.1197/jamia.M2302>, doi:10.1197/jamia.M2302. URL: <https://doi.org/10.1197/jamia.M2302>, <https://academic.oup.com/jamia/article-pdf/14/6/798/2268319/14-6-798.pdf>.
195. Oztekin A, Delen D, Kong ZJ, Predicting the graft survival for heart–lung transplantation patients: An integrated data mining methodology. *International journal of medical informatics* 2009; 78(12):e84–e96, doi:10.1016/j.ijmedinf.2009.04.007.
196. Holzinger A, Jurisica I, Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions. In: *Interactive knowledge discovery and data mining in biomedical informatics*, Springer, 2014; pp. 1–18, doi:10.1007/978-3-662-43968-5_1. https://doi.org/10.1007/978-3-662-43968-5_1.
197. Holzinger A, Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics* 2016; 3(2):119–131, URL: <https://doi.org/10.1007/s40708-016-0042-6>. URL: <https://doi.org/10.1007/s40708-016-0042-6>.
198. Yom-Tov E, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Hochberg I, Encouraging physical activity in patients with diabetes: Intervention using a reinforcement learning system. *J Med Internet Res* 2017; 19(10):e338, URL: <https://doi.org/10.2196/jmir.7994>, doi:10.2196/jmir.7994. URL: <https://doi.org/10.2196/jmir.7994>.
199. Raghu A, Komorowski M, Celi LA, Szolovits P, Ghassemi M, Continuous state-space models for optimal sepsis treatment—a deep reinforcement learning approach. *arXiv preprint arXiv:170508422* 2017; .
200. Liu Y, Logan B, Liu N, Xu Z, Tang J, Wang Y, Deep reinforcement learning for dynamic treatment regimes on medical registry data. In: 2017 IEEE International Conference on Healthcare Informatics (ICHI), IEEE, 2017, pp. 380–385, doi:10.1109/ICHI.2017.45.
201. Yu C, Liu J, Zhao H, Inverse reinforcement learning for intelligent mechanical ventilation and seda-

- tive dosing in intensive care units. *BMC medical informatics and decision making* 2019; 19(2):57, URL: <https://doi.org/10.1186/s12911-019-0763-6>. URL: <https://doi.org/10.1186/s12911-019-0763-6>.
202. Yu C, Dong Y, Liu J, Ren G, Incorporating causal factors into reinforcement learning for dynamic treatment regimes in hiv. *BMC medical informatics and decision making* 2019; 19(2):19–29, URL: <https://doi.org/10.1186/s12911-019-0755-6>. URL: <https://doi.org/10.1186/s12911-019-0755-6>.
203. Qiu J, Sun Y, Self-supervised iterative refinement learning for macular oct volumetric data classification. *Computers in biology and medicine* 2019; 111:103327, URL: <https://doi.org/10.1016/j.compbimed.2019.103327>. URL: <https://doi.org/10.1016/j.compbimed.2019.103327>.
204. Gildenblat J, Klaiman E, Self-supervised similarity learning for digital pathology. *arXiv preprint arXiv:190508139* 2019; .
205. Sarkar P, Etemad A, Self-supervised ecg representation learning for emotion recognition. *arXiv preprint arXiv:200203898* 2020; .
206. Meng X, Ganoe CH, Sieberg RT, Cheung YY, Hassanpour S, Self-supervised contextual language representation of radiology reports to improve the identification of communication urgency. *AMIA Summits on Translational Science Proceedings* 2020; 2020:413.
207. Tachibana R, Näppi JJ, Hironaka T, Yoshida H, Self-supervised generative adversarial network for electronic cleansing in dual-energy ct colonography. In: *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*, vol. 11318, vol. 11318, International Society for Optics and Photonics, 2020, p. 113181E. URL: <https://doi.org/10.1117/12.2549234>.
208. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J, Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE transactions on medical imaging* 2016; 35(5):1299–1312, doi:10.1109/TMI.2016.2535302.
209. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK, Medical image analysis using convolutional neural networks: a review. *Journal of medical systems* 2018; 42(11):226, doi: <https://doi.org/10.1007/s10916-018-1088-1>.
210. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA, Bottou L, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research* 2010; 11(12).
211. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y, Generative adversarial nets. In: *Advances in neural information processing systems*, 2014, pp. 2672–2680.
212. Übeyli ED, Recurrent neural networks employing lyapunov exponents for analysis of ecg signals. *Expert systems with applications* 2010; 37(2):1192–1199, doi:<https://doi.org/10.1016/j.eswa.2009.06.022>.
213. Sorokin I, Seleznev A, Pavlov M, Fedorov A, Ignateva A, Deep attention recurrent q-network. *arXiv preprint arXiv:151201693* 2015; .
214. Torrey L, Shavlik J, Transfer learning. In: *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, IGI global, 2010; pp. 242–264, doi:10.4018/978-1-60566-766-9.ch011.
215. Krizhevsky A, Sutskever I, Hinton GE, Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, 2012, pp. 1097–1105.
216. He K, Zhang X, Ren S, Sun J, Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
217. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K, Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. *arXiv preprint arXiv:160207360* 2016; .
218. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708, doi:10.1109/CVPR.2017.243.
219. Xiao Y, Wu J, Lin Z, Zhao X, A deep learning-based multi-model ensemble method for cancer

- prediction. *Computer methods and programs in biomedicine* 2018; 153:1–9, doi:<https://doi.org/10.1016/j.cmpb.2017.09.005>.
220. The cancer genome atlas program. URL: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>.
221. Deserno TM, Welter P, Horsch A, Towards a repository for standardized medical image and signal case data annotated with ground truth. *Journal of digital imaging* 2012; 25(2):213–226.
222. Murphy SN, Herrick C, Wang Y, Wang TD, Sack D, Andriole KP, Wei J, Reynolds N, Plesniak W, Rosen BR, et al., High throughput tools to access images from clinical archives for research. *Journal of digital imaging* 2015; 28(2):194–204, doi:10.1007/s10278-014-9733-9.
223. Svensson-Ranallo PA, Adam TJ, Sainfort F, A framework and standardized methodology for developing minimum clinical datasets. *AMIA Summits on Translational Science Proceedings* 2011; 2011:54.
224. Guidelines for preparing clinical study data sets for submission to the nhlbi data repository. URL: <https://www.nhlbi.nih.gov/grants-and-training/policies-and-guidelines/guidelines-for-preparing-clinical-study-data-sets-for-submission-to-the-nhlbi-data-repository>.
225. Son LH, Thong NT, Intuitionistic fuzzy recommender systems: an effective tool for medical diagnosis. *Knowledge-Based Systems* 2015; 74:133–150, URL: <https://doi.org/10.1016/j.knosys.2014.11.012>. URL: <https://doi.org/10.1016/j.knosys.2014.11.012>.
226. Thong NT, et al., Hifcf: An effective hybrid model between picture fuzzy clustering and intuitionistic fuzzy recommender systems for medical diagnosis. *Expert Systems with Applications* 2015; 42(7):3682–3701, URL: <https://doi.org/10.1016/j.eswa.2014.12.042>. URL: <https://doi.org/10.1016/j.eswa.2014.12.042>.
227. Galeano D, Paccanaro A, A recommender system approach for predicting drug side effects. In: 2018 International Joint Conference on Neural Networks (IJCNN), IEEE, 2018, pp. 1–8.
228. Kushwaha N, Goyal R, Goel P, Singla S, Vyas OP, Lod cloud mining for prognosis model (case study: Native app for drug recommender system). *Advances in Internet of Things* 2014; 2014, doi: 10.4236/ait.2014.43004.
229. Zhang Q, Zhang G, Lu J, Wu D, A framework of hybrid recommender system for personalized clinical prescription. In: 2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), IEEE, 2015, pp. 189–195, doi:10.1109/ISKE.2015.98.
230. Bao Y, Jiang X, An intelligent medicine recommender system framework. In: 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), IEEE, 2016, pp. 1383–1388, doi: 10.1109/ICIEA.2016.7603801.
231. Bhat S, Aishwarya K, Item-based hybrid recommender system for newly marketed pharmaceutical drugs. In: 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2013, pp. 2107–2111, doi:10.1109/ICACCI.2013.6637506.
232. Guo L, Jin B, Yao C, Yang H, Huang D, Wang F, Which doctor to trust: a recommender system for identifying the right doctors. *Journal of medical Internet research* 2016; 18(7):e186, doi:10.2196/jmir.6015.
233. Ciecierski K, Raś ZW, Przybyszewski AW, Foundations of recommender system for stn localization during dbs surgery in parkinson’s patients. In: International Symposium on Methodologies for Intelligent Systems, Springer, 2012, pp. 234–243. URL: https://doi.org/10.1007/978-3-642-34624-8_28.
234. Petschmann S, Schöffmann K, Deep learning for shot classification in gynecologic surgery videos. In: International Conference on Multimedia Modeling, Springer, 2017, pp. 702–713. URL: https://doi.org/10.1007/978-3-319-51811-4_57.
235. Wang Z, Fey AM, Deep learning with convolutional neural network for objective skill evaluation in robot-assisted surgery. *International journal of computer assisted radiology and surgery* 2018; 13(12):1959–1970, URL: <https://doi.org/10.1007/s11548-018-1860-1>. URL: <https://doi.org/10.1007/s11548-018-1860-1>.
236. Shvets AA, Rakhlin A, Kalinin AA, Iglovikov VI, Automatic instrument segmentation in robot-assisted surgery using deep learning. In: 2018 17th IEEE International Conference on Machine

- Learning and Applications (ICMLA), IEEE, 2018, pp. 624–628, doi:10.1109/ICMLA.2018.00100.
237. Fanconi C, Skin cancer: Malignant vs. benign. 2019. URL: <https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign>.
238. Zhang J, Xie Y, Xia Y, Shen C, Attention residual learning for skin lesion classification. *IEEE transactions on medical imaging* 2019; 38(9):2092–2103.
239. Tschandl P, Sinz C, Kittler H, Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation. *Computers in biology and medicine* 2019; 104:111–116.
240. Nida N, Irtaza A, Javed A, Yousaf MH, Mahmood MT, Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy c-means clustering. *International journal of medical informatics* 2019; 124:37–48.
241. Mahbod A, Schaefer G, Ellinger I, Ecker R, Pitiot A, Wang C, Fusing fine-tuned deep features for skin lesion classification. *Computerized Medical Imaging and Graphics* 2019; 71:19–29.
242. Prashanth R, Roy SD, Mandal PK, Ghosh S, High-accuracy detection of early parkinson’s disease through multimodal features and machine learning. *International journal of medical informatics* 2016; 90:13–21, URL: <https://doi.org/10.1016/j.ijmedinf.2016.03.001>. URL: <https://doi.org/10.1016/j.ijmedinf.2016.03.001>.
243. Kolachalama VB, Garg PS, Machine learning and medical education. *NPJ digital medicine* 2018; 1(1):1–3, URL: <https://doi.org/10.1038/s41746-018-0061-1>. URL: <https://doi.org/10.1038/s41746-018-0061-1>.
244. Bannach-Brown A, Przybyła P, Thomas J, Rice AS, Ananiadou S, Liao J, Macleod MR, The use of text-mining and machine learning algorithms in systematic reviews: reducing workload in pre-clinical biomedical sciences and reducing human screening error. *BioRxiv* 2018; :255760URL: <https://doi.org/10.1101/255760>. URL: <https://doi.org/10.1101/255760>.
245. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, Yang GZ, Deep learning for health informatics. *IEEE journal of biomedical and health informatics* 2016; 21(1):4–21, doi:10.1109/JBHI.2016.2636665.
246. Weintraub WS, Fahed AC, Rumsfeld JS, Translational medicine in the era of big data and machine learning. *Circulation research* 2018; 123(11):1202–1204, URL: <https://doi.org/10.1161/CIRCRESAHA.118.313944>. URL: <https://doi.org/10.1161/CIRCRESAHA.118.313944>.
247. Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, Ringel M, Schork N, Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ digital medicine* 2019; 2(1):1–5, URL: <https://doi.org/10.1038/s41746-019-0148-3>. URL: <https://doi.org/10.1038/s41746-019-0148-3>.
248. Savage N, Better medicine through machine learning. *Communications of the ACM* 2012; 55(1):17–19, URL: <https://doi.org/10.1145/2063176.2063182>. URL: <https://doi.org/10.1145/2063176.2063182>.
249. Char DS, Shah NH, Magnus D, Implementing machine learning in health care—addressing ethical challenges. *The New England journal of medicine* 2018; 378(11):981.
250. Randhawa GS, Soltysiak MP, El Roz H, de Souza CP, Hill KA, Kari L, Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: Covid-19 case study. *Plos one* 2020; 15(4):e0232391.
251. Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X, Artificial intelligence and machine learning to fight covid-19. 2020.
252. Stanford institute for human-centered artificial intelligence (hai). COVID-19 and AI: A Virtual Conference. URL: <https://hai.stanford.edu/events/covid-19-and-ai-virtual-conference>.
253. Pakhomov SV, Buntrock JD, Chute CG, Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques. *Journal of the American Medical Informatics Association* 2006; 13(5):516–525, doi:10.1197/jamia.M2077.
254. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H, edocter: machine learning and the future of medicine. *Journal of Internal Medicine* 2018; 284(6):603–619.
255. Yampolskiy RV, Artificial intelligence safety and cybersecurity: A timeline of ai failures. *arXiv preprint arXiv:161007997* 2016; .