Research Center
Finance & Information Management

Fraunhofer
FIT

Project Group
Business & Information
Systems Engineering

Machine Learning approaches along the Radiology Value Chain – Rethinking Value Propositions

by

Peter Hofmann, Severin Oesterle, Paul Rust[1], Nils Urbach

December 2018

to be presented at: 27th European Conference on Information Systems (ECIS),
Stockholm-Uppsala, Sweden, 2019

[1] Student

University of Augsburg, D-86135 Augsburg
Visitors: Universitätsstr. 12, 86159 Augsburg
Phone: +49 821 598-4801 (Fax: -4899)

University of Bayreuth, D-95440 Bayreuth
Visitors: Wittelsbacherring 10, 95444 Bayreuth
Phone: +49 921 55-4710 (Fax: -844710)

www.fim-rc.de

WI-845

MACHINE LEARNING APPROACHES ALONG THE RADIOLOGY VALUE CHAIN – RETHINKING VALUE PROPOSITIONS

Research paper

Hofmann, Peter, Project Group Business & Information Systems Engineering of the Fraunhofer FIT, Bayreuth, Germany, peter.hofmann@fit.fraunhofer.de

Oesterle, Severin, FIM Research Center, University of Bayreuth, Bayreuth, Germany, severin.oesterle@fim-rc.de

Rust, Paul, University of Bayreuth, Bayreuth, Germany, paul.rust@uni-bayreuth.de

Urbach, Nils, FIM Research Center, University of Bayreuth, Project Group Business & Information Systems Engineering of the Fraunhofer FIT, Bayreuth, Germany, nils.urbach@fim-rc.de

Abstract

Radiology is experiencing an increased interest in machine learning with its ability to use a large amount of available data. However, it remains unclear how and to what extent machine learning will affect radiology businesses. Conducting a systematic literature review and expert interviews, we compile the opportunities and challenges of machine learning along the radiology value chain to discuss their implications for the radiology business. Machine learning can improve diagnostic quality by reducing human errors, accurately analysing large amounts of data, quantifying reports, and integrating data. Hence, it strengthens radiology businesses seeking product or service leadership. Machine learning fosters efficiency by automating accompanying activities such as generating study protocols or reports, avoiding duplicate work due to low image quality, and supporting radiologists. These efficiency improvements advance the operational excellence strategy. By providing personnel and proactive medical solutions beyond the radiology silo, machine learning supports a customer intimacy strategy. However, the opportunities face challenges that are technical (i.e., lack of data, weak labelling, and generalisation), legal (i.e., regulatory approval and privacy laws), and persuasive (i.e., radiologists' resistance and patients' distrust). Our findings shed light on the strategic positioning of radiology businesses, contributing to academic discourse and practical decision-making.

Keywords: Artificial Intelligence, Machine Learning, Radiology, Health IT, Business Models.

1 Introduction

In recent years, artificial intelligence (AI) has entered business in various forms. According to the Gartner Hype Cycle for Emerging Technologies, AI is one of the three emerging technology trends that could potentially impact any industry (Panetta, 2018). In particular, machine learning, a branch of AI, has been deployed in various applications to learn underlying patterns and relationships in large data sets (Paliwal and Kumar, 2009). One application field of machine learning approaches in the health sector is radiology. Radiology has been one of the early adopters of health IT and has greatly benefited from various advances in imaging technology such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET). Consequently, the large amount of available image data enables applying machine learning in image interpretation, for example extracting valuable information for diagnosis and therapeutic decisions. Machine learning approaches with their ability to utilise large amounts of data and the abundance of data in radiology even leads to the controversy as to whether a machine can ultimately replace radiologists (Silverman, 2017). Academic literature addresses (specific) applications of machine learning in radiology (e.g., Litjens et al., 2017; Lakhani et al., 2018; Mazurowski et al., 2018; Wang and Summers, 2012), machine learning models for the detection or treatment of (specific) diseases (e.g., Dhungel et al., 2015; Armato et al., 2001), and concentrates on a selection of opportunities or challenges (e.g., Balthazar et al., 2018; Bruijne, 2016). However, it is still unclear how and to what extent machine learning will affect the radiology business.

Addressing the research gap identified above, we aim at identifying the challenges and opportunities of machine learning in radiology to discuss their impact on radiology business. Specifically, we address the following research question: *How does machine learning affect the value propositions of radiology businesses and which challenges and opportunities exist?* To answer our research question, we rely on the radiology business models of Enzmann and Schomer (2013), a value discipline approach based on Treacy and Wiersema (1997). According to this concept, a radiology business model combines the value propositions: operational excellence, product and service leadership, and customer intimacy to a certain extent. Pursuing operational excellence, a company offers a product at the lowest price possible while maintaining an acceptable level of quality. For example, a radiology practice uses economies of scale to achieve low costs per scan or report. Businesses that pursue product and service leadership offer innovative or high-performance products that yield high margins. A radiology practice may maintain a leadership role in diagnostic services, because of the development of imaging technologies. Companies following the customer intimacy strategy focus on providing a solution to an individual customer rather than an entire market segment. Confronted with these value propositions, radiology businesses tend to concentrate on one value proposition at the expense of the others (Enzmann and Schomer, 2013). We rely on a structured and holistic framework to analyse the opportunities and challenges associated with the application of machine learning in radiology, since the application of machine learning in radiology is not limited to image interpretation (Lakhani et al., 2018) and because of the fact that the value chain influences the choice of the business model (Enzmann, 2012). In this respect, we follow the concept of the radiology value chain proposed by Enzmann (2012). Therefore, we conducted a systematic literature review and complement and validate our findings through expert interviews. This proceeding allows us to contribute to the literature by compiling the opportunities and challenges of machine learning along the radiology value chain. This structuring not only serves as a basis for academic discourse but also supports decision-makers in radiology. We further contribute to the body of knowledge by discussing implications for the radiology business models introduced by Enzmann and Schomer (2013).

We structure the remainder of this paper as follows: In the second section, we contextualise the deployment of machine learning along the radiology value chain providing both potential and already established use cases. The third section includes a description of the research approach. In the fourth section, we introduce the opportunities and challenges of machine learning in radiology resulting from the literature review and expert interviews. We discuss our findings and their implications for radiology businesses in the fifth section. In the sixth section, we conclude by summarizing the findings of our work and discussing its limitations and future research opportunities.

2 Background

The key activities of the radiology value chain according to Enzmann (2012) include acquiring images, reading images, generating reports, and providing medical decisions, which we elaborate on in more detail in the following. Furthermore, we present potential machine learning approaches for each step. Figure 1 depicts the activities of the radiology value chain according to Enzmann (2012).

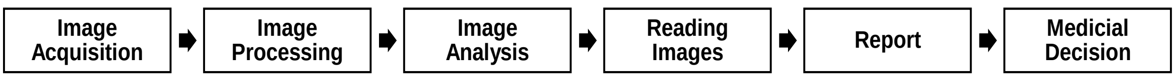

Figure 1. Radiology's value chain according to Enzmann (2012, p.246).

2.1 Image acquisition, processing, and analysis

Based on a request for a specific examination, the radiologist creates a study protocol before performing the actual examination. In this activity, a machine learning model may reduce human effort by predicting the correct MRI protocols, estimating the need for contrast agents, and determining the priority of the case (Brown and Marotta, 2018). After creating the study protocol, image acquisition aims at maintaining satisfactory image quality, which is indispensable for the subsequent activities. Thus, quality assessment helps to evaluate whether an image is suitable for diagnosis and consequently to avoid unnecessary work. In recent years, there have been a few attempts to automate this step (e.g., Abdi et al., 2017). For instance, monitoring image quality during the examination allows technologists to react in time, enabling them to make the necessary adjustments during the examination (Esses et al., 2018).

After acquiring the image, image processing (i.e., reconstruction, denoising, registration, and segmentation) takes place. Reconstruction refers to generating images from the acquired data (Levitan and Herman, 1987). In this context, machine learning can, for example, augment the reconstruction process to create images from weaker scanners (e.g., Bahrami et al., 2017) or reconstruct 'normal-dose' CT images from 'low-dose' images reducing the risk of radiation without causing a decline in image quality (e.g., Kang et al., 2017). Furthermore, denoising enables reducing image noise that causes inaccuracies in clinical images and consequently facilitating the detection of important findings such as the location of a lesion (Vaishali et al., 2015). For instance, Jiang et al. (2017) describe a deep learning framework that reduces the noise of MRI images and outperforms state of the art methods. Next, registration refers to aligning two different images (e.g., from pre- and post-surgery) in a common reference place (Bauer et al., 2013). In particular, unsupervised methods which do not need labelled training data lead to an improved registration performance (Liao et al., 2017b). While dividing an image into specific parts (i.e., segmentation) is important for diagnosis, surgery and treatment planning, manually outlining structures is time-consuming and prone to human error (Akkus et al., 2017). Machine learning algorithms automate this process and show high levels of accuracy (Mazurowski et al., 2018).

Subsequent to image processing, image analysis extracts additional information from the data. The extraction of quantitative features (e.g., size, volume) indicates, for example, the observable characteristics of a tumour's genetic makeup and its direct environment (Gillies et al., 2016). Moreover, cardiac imaging can reveal various cardiac pathologies such as coronary heart disease. Additional quantitative information can help to identify the specific disease (Babu-Narayan et al., 2016). Machine learning models have shown to be an effective method to provide quantitative information across different modalities such as CT, MRI, and echocardiography (Slomka et al., 2017). This includes estimating left ventricle volumes on MRI images in order to assess the ejection function of the heart (Liao et al., 2017a). Applications may use convolutional neural networks (CNN) to calculate coronary artery calcium scores, which serve as a predictor of cardiovascular events such as heart attacks (Wolterink et al., 2016).

2.2 Reading images

After acquiring the images, the radiologist opens them by using a picture archiving and communication system (PACS), searches for any abnormalities using, amongst others, computer-aided detection (CADe), and characterises the region (Enzmann, 2012). CADe systems use, for example, machine learning models to detect lung nodules on thoracic CT images (Armato et al., 2001) or breast cancer on mammography (Becker et al., 2017). When the radiologist finds a suspicious region on a medical image, it may still be challenging for the radiologist to characterise and classify it correctly. For example, in the case of nodule characterisation, there is no single feature to distinguish between malignant and benign nodules perfectly. Supporting the characterisation task, computer-aided diagnosis (CADx) systems help to determine the disease type, severity, stage, and progression (Chockley and Emanuel, 2016). For instance, Chen et al. (2010) describe a CADx system to classify lung nodules using an Artificial Neural Network, which performs only slightly worse than a senior radiologist. Wang et al.

(2016) indicate that deep learning models have higher discriminative power than traditional machine learning algorithms in classifying microcalcifications on mammograms. After detecting and characterizing a finding, the radiologist needs to interpret the finding to establish the diagnosis. Content-based image retrieval (CBIR) can support the radiologist in his/her interpretative task by providing similar images in databases (Akgül et al., 2011; Müller et al., 2004). Hence, machine learning models can support the CBIR in feature extraction and similarity matching. For example, CBIR systems using support vector machines retrieve mammograms with similar breast tissue density (Oliveira et al., 2011).

The integration phase combines image and non-image data such as histopathologic data, clinical findings, or demographic data, as isolated image data is often not sufficient to make an adequate diagnosis (Enzmann, 2012). Radiomics, a concept closely related to integration, describes the conversion of images into higher-dimensional data and its analysis to obtain predictive or prognostic information to infer about the underlying genomic and proteomics patterns (Lambin et al., 2012). Using a classifier model, radiomics involves acquiring the image, detecting and segmenting regions of interests, extracting features, and predicting outcomes (Gillies et al., 2016). For example, Kumar et al. (2015) extract 500 image-based radiomic sequences of lung CT images using a CNN architecture and a decision tree to distinguish between malignant and benign lesions of lung cancer patients. Another study uses a CNN to extract image-based features and predict the mutation status of isocitrate dehydrogenase 1 (Li et al., 2017). Predicting the mutation status is helpful, as it may serve as an indicator of the tumours' response to chemotherapy (Cohen et al., 2013).

2.3 Report and medical decision

Furthermore, machine learning may support automating report generation, analysing unstructured reports and facilitating the conversion of findings into clinical codes. Natural language processing (NLP) allows partially or fully automating report generation. For example, Jing et al. (2017) propose a deep learning framework analysing over 7,000 chest x-rays with their corresponding reports and then generating text descriptions for abnormal regions found on the test set images. However, the unstructured, free-text format impedes automated information extraction. In a recent study, Shin et al. (2017) use a deep learning model to analyse the radiology reports of intensive care unit patients. The model categorises the reports in normal and abnormal studies and indicates the presence of acute findings (e.g., acute intracranial bleed). Thus, prioritizing the most severe cases enables faster treatment. Trained professionals convert the findings to clinical codes such as the International Statistical Classification of Diseases (ICD) Code (Shi et al., 2017). Studies deploying deep learning algorithms demonstrate that not relying on explicitly designed features enables the automation of ICD coding (e.g., Xie and Xing, 2018).

As an additional contribution to the report, the predictive power of machine learning could help to estimate a patient's outcome by adjusting treatment and consequently the effectiveness of patient care. For example, Yoo et al. (2016) apply deep learning on MRI images to predict the future disease activity of patients with symptoms of multiple sclerosis. The proposed CNN architecture identifies relevant features to predict patients with a higher risk of disease-related attacks who therefore might benefit from a more aggressive treatment in the early disease stages. In another example, machine learning models predict the degree of post-stroke cognitive damage as well as the probable course of recovery over time by using demographic, behavioural, and imaging data (Hope et al., 2013).

3 Research Process and Method

To give a comprehensive overview of the relevant challenges and opportunities of machine learning in radiology, we first conduct a systematic literature review and second, complement and validate our findings with expert interviews. Systematic literature reviews are a fundamental technique in scientific research to efficiently aggregate existing information and answer a specific research question (Ressing et al., 2009). Since some scholars have introduced machine learning use cases in a prototype state or as a theoretical construct, it remains ambiguous how and to what extent machine learning will actually influence the radiology business. For this very reason, we validate and add the opportunities and challenges identified in the literature review from a practical perspective with expert interviews. Expert interviews – as a qualitative-empirical research approach – deem to gain a deeper understanding, generate new insights, and gather specific information (Bettis et al., 2015).

Our systematic literature review, based on a database search, aims at providing an overview of the challenges and opportunities of machine learning in radiology, discussed in scientific literature. Reducing the likelihood of missing relevant contributions, we opt for a broader search string consisting of the following three terms: Radiology, artificial intelligence, and machine learning. We incorporated the term radiology because it portrays our application field. In addition, we incorporated the term machine learning referring to the specific technology under investigation. The regular use of the term artificial intelligence in business, computer science, and medical research motivated us to include this term in our search terms as well. The frequent use is due, among other things, to the fact that artificial intelligence is used as an umbrella term for a number of areas and techniques (Balthazar et al., 2018). Our research addresses a variety of research streams and sub-streams such as healthcare (in particular health IT and radiology information systems), medicine, informatics, data analytics and many more. To identify all relevant contributions, we used three meta-databases (HSG metasearch, UB Catalogue, Primus). The initial search led to 3,072 results. Based on our inclusion criteria (date of publication: 2012 to 2018, academic peer-reviewed journals, English language) 1,188 publications remained. We proceeded with the title screening and excluded 920 publications based on subject focus. Afterwards, we screened the abstract of the remaining 268 manuscripts and excluded 239 non-relevant publications and duplicates. Hence, we obtained 29 relevant papers, which we included in our meta-analysis to answer which challenges and opportunities concerning machine learning in radiology exist.

Concerning interview preparation and execution, we proceeded as follows: First, we clarified whom we consider as experts in our research field. An expert needs to have privileged access to information about a specific topic or possesses extensive knowledge about a subject, gained from professional activities (Bogner et al., 2009). Thus, we define specific criteria (e.g., profession and personal involvement with the research subject) for the selection process (Bhattacherjee, 2012). In total, we identified 17 European experts from which six experts accepted our invitation to participate in our study. Three experts declined to participate without giving reasons for their denial and eight experts did not respond to our invitation. In Table 1, we provide an overview of the experts and their background.

	Expert A	**Expert B**	**Expert C**	**Expert D**	**Expert E**	**Expert F**
Industry	Medical Technology	IT	IT	IT	Consulting	Academia Research
Position	Chief Medical Officer	Chief Technology Officer	Machine Intelligence Engineer	Clinical Director	Managing Director	Section Chief of Thoracic Imaging
Type	Telephone	Personal	Personal	Telephone	Telephone	Telephone
Back-ground	Radiologist (M.D. and Professor)	AI Start-up Founder; PhD Comp. Sc.	PhD in Computer Science	Radiologist (M.D.)	Healthcare IT, Medical Imaging	Radiologist (M.D. and Professor)

Table 1. Overview of experts.

To conduct our interviews, we designed an interview guideline (Flick, 2014) consisting of three main sections: (1) introduction and state-of-the art of machine learning in radiology, (2) challenges of machine learning in radiology, and (3) opportunities of machine learning in radiology. To give the experts preparation time, we sent out the interview guideline to each expert in advance. For a thorough and rigorous data analysis, we have recorded and transcribed the interviews with the interviewees' consent. After having transcribed our 225 interview minutes, we evaluated them following Mayring's qualitative content analysis, which consists of paraphrasing, generalizing and arranging in categories of specific passages (Mayring, 2014). According to Mayring (2014), categories are defined using either a deductive or an inductive approach. Since we aim at evaluating categories, which are based on our literature review findings, we opt for a deductive method. In the subsequent section, we present the findings of our complementary approaches.

4 Findings

In the following sections, we present both the opportunities (Section 4.1) and challenges (Section 4.2) of machine learning along the radiology value chain.

4.1 Opportunities

Machine learning yields many opportunities along the radiology value chain. We characterise benefits in effectivity and efficiency improvements based on the findings of the literature review and expert interviews. The opportunities are illustrated in Figure 2.

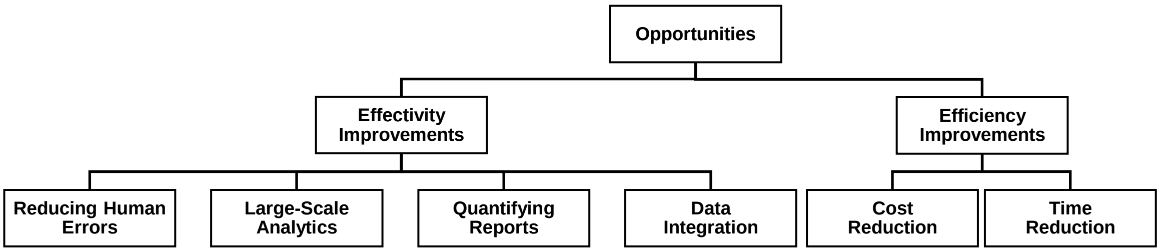

Figure 2. Overview of the opportunities of machine learning in radiology.

4.1.1 Effectivity improvements

The raison d'être of radiology is to answer medical questions, raised by a referring physician or the patients themselves (Enzmann, 2012). The need to reduce medical errors to save lives and costs is self-evident. A good place to start is improving the initial disease diagnosis, as it is the foundation for future treatment decisions. Diagnostic errors are considered to be as frequent as 3% to 5% of all daily diagnoses. In specific cases such as in screening mammography, there are even higher error rates that can be up to 31% (Nelson et al., 2016; Brady, 2017). Human errors such as cognitive biases, lack of knowledge, and faulty reasoning contribute to diagnostic error (Brady, 2017). Machine learning models are prone to bias from the input data, but they are robust to human bias and cognitive shortcomings such as fatigue (Obermeyer and Emanuel, 2016). Gichoya et al. (2018) list common biases in radiology and illustrate how machine learning can counteract them. For example, to overcome the satisfaction of search bias, which describes the tendency to stop checking for more abnormalities after an initial abnormal finding, machine learning models can provide smart checklists and flag potential blind spots of the radiologist. The experts also consider reduced human errors to be one of the most promising advantages; *"It's a recognised fact that radiologists and physicians in general make mistakes both in terms of getting the diagnosis wrong but also in terms of missing things while doing the diagnosis. AI can definitely play a role there [...]"* (Interview E). A constant, high-quality diagnosis that is unaffected by human conditions such as stress, lack of concentration, or fatigue will lay a more objective fundament for future treatment decisions (Interview C). Machine learning is able to extract information, visible and invisible to the human eye, from images and to constantly process large-scale datasets more accurately than humans (Obermeyer and Emanuel, 2016). Deep learning models have already surpassed human performance in everyday image object detection and demonstrate high levels of accuracy in analysing medical images (Ribli et al., 2018).

Including quantitative data to the report will facilitate the trend toward evidence-based medicine. Evidence-based medicine uses the most current and accurate evidence as the basis for decision-making for patient care (Masic et al., 2008). Machine learning models can extract quantitative data such as the heart chamber volume or the thickness of the myocardium with higher accuracy and in a fraction of time than performed by a human (Carneiro et al., 2017). Quantifying the report provides objective support for the radiologist's diagnosis and allows tracking patients over time, thus increasing the value of the report (Interview E). It also facilitates data analysis in clinical trials, since cohort features become more comparable (Brink et al., 2017).

Machine learning enables integrating information from various data sources into the report. Data extracted directly from electronic health records, imaging and knowledge databases can be correlated with downstream patient outcomes, so that the report not only contains the status quo but also includes prognoses (Obermeyer and Emanuel, 2016). Predictions about survival times or treatment responses could improve treatment decisions and thereby the value of the radiological answer (Yasaka et al., 2018). The concept of radiomics promises to improve prognoses even further by combining image data with omics-data. Radiomics allows, for example, predicting the survival time of brain cancer patients and treatment responses of kidney tumours (Kickingereder et al., 2016; Goh et al., 2011). Three experts confirm that integrating omics-data and non-imaging data would have a positive impact on the diagnosis (Interview A; Interview E; Interview F). One expert notes that he/she does not see an immediate future for radiomics, as too little data has been collected in this area and humans themselves have not yet fully decoded the human genome (Interview E). However, deep learning does not require hand-crafted features and can thus reduce human bias (Afshar et al., 2018). Moreover, by revealing correlations, radiomics provides additional information for diagnostic decisions such as the choice of biopsy sites (Gillies et al., 2016). Besides, the progress in understanding genomes or proteins and the reduction of bias in the training of machine learning models complement each other.

4.1.2 Efficiency improvements

Machine learning applications can support the activities along the radiology value chain and improve the efficiency of the entire chain by lowering costs and saving time. Cost and time reductions are interdependent, but a change in costs does not necessarily entail a change in time and vice versa.

Radiology businesses typically position themselves as low-cost service providers as the entire health care system aims to minimise costs (Enzmann and Schomer, 2013). Machine learning can support cost reduction, amongst others, by improving scanner utilisation, reducing double examinations and lowering CT radiation doses. The utilisation rate of the scanners is a major driver of the total examination costs due to the high investment and associated labour costs. One way to improve the utilisation of a scanner is a reduction in scanning time. For example, machine learning algorithms that reconstruct sparse data into complete images enable performing MRI scans twice as fast (Lakhani et al., 2018). Two of the experts confirm that reducing MRI scan times is one of the most beneficial applications of machine learning (Interview D; Interview F). Another approach to improve the utilisation rate is to estimate the length per time slot and then arrange MRI scans accordingly by incorporating various input factors such as demographic data or study protocol (Muelly et al., 2017). Optimizing the arrangement of exam slots could prevent scanner downtime significantly. Both approaches, shortening scan time and optimizing slot arrangement, increase the utilisation and result in lower costs per exam (Massat, 2018).

Avoiding unnecessary double examinations due to insufficient image quality also enables cost saving. Machine learning-based image reconstruction and denoising techniques have shown to improve image quality significantly and require less time than traditional methods (Yasaka et al., 2018; Gondara, 2016). Interviewee D and F emphasise this argument. Furthermore, the radiologist could monitor image quality during the examination with the help of deep learning models to inform technologists in time about possible quality losses, enabling them to make the necessary adjustments during the exam and therefore avoiding double examination (Esses et al., 2018). Lowering radiation doses in CT scans while maintaining or even improving the image quality is another opportunity (Yasaka et al., 2018; Lakhani et al., 2018; Massat, 2018). Machine learning models can recreate high-quality images from 'low-dose' scans, consequently reducing the amount of contrast media needed and limiting potentially harmful effects of radiation and improving cost-effectiveness. Three experts also confirm this opportunity (Interview A; Interview D; Interview F).

The perhaps biggest impact of machine learning applications on the value chain is task automation that results in significant time savings (Brink et al., 2017). Bearing in mind that the average radiologist has to interpret an image every three to four seconds to cope with the workloads, the need to facilitate image analysis becomes clear (McDonald et al., 2015). Radiologists using machine learning for the detection and interpretation tasks will not only be able to deal with the increasing workloads but also achieve higher diagnostic accuracies than radiologists without machine learning support (Gichoya et al., 2018). The use cases presented in the background section illustrate that machine learning can support almost every activity which can lead to full or semi-automation of various tasks. During image acquisition, machine learning models can automatically determine the correct study protocols and reduce scan time (Lakhani et al., 2018). Furthermore, image analysis tasks such as detection, segmentation, and classification can also be fully automated to decrease the time required by the radiologist (Shiraishi et al., 2011; Massat, 2018). For example, one approach for screening scenarios involving a large number of normal studies would be to eliminate cases that are definitely negative (Mayo and Leung, 2018; Mazurowski et al., 2018). Two of the experts regard this approach as very promising because even ruling out only a small percentage of all normal studies could result in significant time and labour savings (Interview B; Interview C). Nonetheless, Interviewee A states that image interpretation may not be fully automatic shortly, but current applications that pre-select the most relevant images of a scan already help to process the large amounts of data. CBIR, patient triaging (i.e., automatic case prioritisation), automatic reporting, and other applications that automate parts of the diagnostic workflow have already entered the market or are imminent (Massat, 2018). All six experts agree that machine learning will inevitably make the workflow more efficient by task automating. *"The acceleration of the examination is essential. I expect the radiologist to be supported in his/her diagnostic process, whether this is the diagnosis itself or steps in that direction, I think there is still a lot to be gained"* (Interview F). Automation results in higher productivity by an increased patient throughput.

4.2 Challenges

Although machine learning approaches can support many activities along the radiology value chain, they face a variety of challenges. Based on the findings of the literature review and expert interviews, we classify these into technical, legal, and persuasion challenges as illustrated in Figure 3.

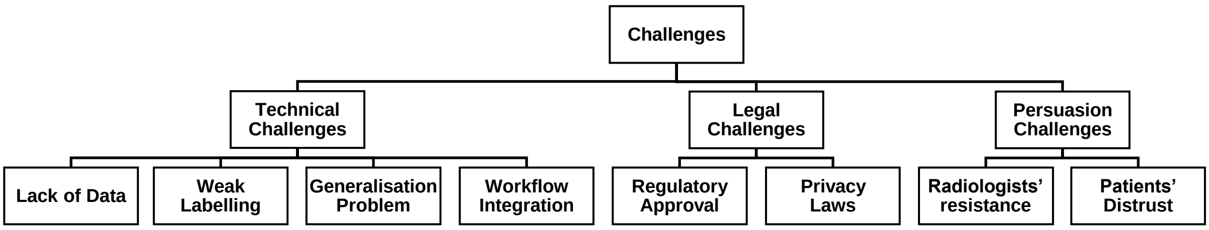

Figure 3. Overview of the challenges of machine learning in radiology.

4.2.1 Technical challenges

Academic literature frequently refers to technical challenges. We categorise them as *Lack of Data*, *Weak Labelling*, *Generalisation Problems,* and *Workflow Integration*. Other technical challenges that few publications only briefly mention include lack of necessary IT infrastructure (Massat, 2018), lack of computing power (Thrall et al., 2018), and failing to reach clinically acceptable accuracy levels (Weese and Lorenz, 2016).

Among the technical challenges, the lack of data is the most frequently cited challenge. Four of the interviewees also highlight this challenge. *"Getting all this data […], so that models work well on medical images is the most important challenge"* (Interview C). Since the performance of machine learning models typically improves with an increased dataset, there is a strong demand for large-scale, well-annotated datasets (Akkus et al. 2017). The limited availability of such datasets affects the reliability of machine learning models (Carneiro et al., 2017). For example, training deep learning models that often consist of more than 10^6 parameters becomes difficult if the medical data set only contains a few thousand samples (Suzuki, 2017). One expert mentions that even if a data set consists of high-quality image samples, the set fails to provide pathological results, which are crucial for the performance of the algorithm (Interview A). The lack of annotated datasets stems from the fact that before the rise of machine learning, there was no need to annotate the data in a particular fashion. *"Because radiologists don't annotate data to feed algorithms, we label data for clinical benefit. We don't sit there drawing around lesions for every slide of the image. [...]"* (Interview D). Nevertheless, the amount of unstructured and unannotated data steadily increases, although the creation of structured, annotated databases has just begun (Blum and Zins, 2017). Besides the sheer volume of data, two experts emphasise the importance of extremely infrequent cases in the dataset. *"It's really difficult to train an algorithm on things that only occur one time and are not present in your training data"* (Interview C). Interviewee B emphasises that failing to classify these edge cases correctly during clinical routine could have adverse effects. Another data-related problem revolves around image quality. Insufficient image quality (e.g., image artefacts, different resolutions, image noise) in combination with varying image protocols cause variations in data (Tang et al., 2018). These variations, in turn, affect the training process since the more heterogeneous the data, the more difficult it is to analyse underlying patterns.

Another frequently cited challenge is weak labelling. For supervised learning, the performance strongly depends on the quality of the labels. *"The quality of the data is essential. 'Garbage in – garbage out'. An algorithm can only be as good as the data you feed it with"* (Interview A). There are two main labelling challenges: the definition of ground truth (e.g., biopsy or clinical outcome) and the correct labelling process itself (Tang et al., 2018). Defining the ground truth for the labels is difficult in specific cases because medicine itself allows different interpretations. While a tissue biopsy can attain reliable ground truth in classification tasks (e.g., cancer vs. no cancer), expert variability heavily influences the ground truth for segmentation tasks (Cabitza et al., 2017; Akkus et al., 2017). Expert D notes that: *"[...] medicine is not an exact science. Medicine doesn't have hard results; there are rules of thumbs. So, it's quite hard to get what is known as the ground truth"* (Interview D). Even if medical experts are available for labelling, human error affects the quality of annotations. One expert stresses the missing standardisation of labelling, which further contributes to the problem (Interview B).

The generalisation problem refers to machine learning models that perform well in isolated test settings but fail to generalise their results once integrated into clinical practice (Lakhani et al., 2018). Machine learning models work under the premise that the data engineer randomly samples both training and test sets from the same distribution. However, aspects such as different imaging protocols or varying patient populations violate this assumption and affect the performance of the machine learning model (Bruijne, 2016). Once deployed in a clinical setting, the machine learning model needs to deal with heterogeneous data, resulting from demographic or ethnic factors differences as well as varying disease prevalence and organ sizes (Thrall et al., 2018). Tang et al. (2018) demonstrate that machine learning may show high performances in a US lung screening trial but perform considerably worse at Oxford University Hospitals. In order to achieve a widespread distribution in radiology, machine learning needs to overcome the generalisation problem (Burt et al., 2018). To overcome the generalisation hurdle, it is important to use heterogeneous data for training the machine learning model (Kim et al., 2019). Therefore, a multicentre approach or an active learning approach can support generating a heterogeneous data set.

A seamless workflow integration of the application in the existing PACS and the saving of the results to Digital Imaging and Communications in Medicine (DICOM) standard are necessary for clinical practice. Current machine learning products sometimes fail to utilise the existing interfaces and present themselves as isolated rather than integrated solutions (Tang et al., 2018). Additionally, machine learning models typically address a single case and are not transferable to perform other tasks. Thus, the integration and testing process would have to be repeated several times, which is not economically sustainable for the radiology business (Weese and Lorenz, 2016). One way to facilitate the integration would be to provide an algorithm marketplace that provides an interface to the existing PACS and allows AI companies to offer their algorithms there (Interview E).

4.2.2 Legal challenges

The development of machine learning models and their subsequent deployment in clinical routine faces two main legal challenges: regulatory approval and privacy laws. The market launch of machine learning products for clinical practice requires regulatory approval in some form. However, gaining regulatory approval remains challenging, because of the so-called black box problem and the missing validation framework. The black box problem describes the lack of transparency in the decision-making process of deep learning models (Yasaka et al., 2018). Due to the lack of transparency, regulatory bodies have great difficulty assessing the algorithms that use circumstantial factors rather than actual disease factors (Bruijne, 2016; Blum and Zins, 2017). Three of the experts also name the black box problem as one of the main barriers to receiving regulatory approval. For traditional machine learning this does not pose a problem, because statistical knowledge explains and visualises the derived conclusion (e.g., with a decision tree) (Dwyer et al., 2018). Even though the decision making process gains transparency, results might still be misleading (Cabitza et al., 2017). Clinical validations have to test the safety of a machine learning product to gain regulatory approval. However, no existing framework clearly outlines the necessary approval requirements such as accuracy scores or the size of the testing population (Thrall et al., 2018; Lakhani et al., 2018; Pesapane et al., 2018). The Food and Drug Administration (FDA), for example, recommends testing generalisability with different imaging devices but fails to define specific performance metrics (Burt et al., 2018). One expert points out that regulators still need to familiarise themselves with the new technology before they can develop clear guidelines (Interview D).

Health and patients' genetic data classify as sensitive data under the Health Insurance Portability and Accountability Act (US) and the General Data Protection Regulation (Europe). These legal frameworks exist to ensure patients' privacy (Pesapane et al., 2018). Machine learning's massive demand for data may interfere with the privacy and confidentiality of patient data (Balthazar et al., 2018). The General Data Protection Regulation, for example, provides a useful framework that regulates the use of data by algorithms (e.g., anonymisation). Anonymisation of health data is a necessity for training purposes, but it is questionable whether full anonymisation can be achieved (Tang et al., 2018). For example, by using brain MRI images, it is possible to reconstruct the facial anatomy of a patient (Bischoff-Grethe et al., 2007). Also, the sharing of data between institutions would help to overcome the lack of data but is legally problematic due to the risks of data breaches in the data management systems of the institutions or their partners (Gibson et al., 2018). The discussion about whether privacy laws present a challenge to machine learning is controversial. Two of the experts note that privacy laws do not constitute a significant restriction for the usage of machine learning and the new data protection laws even facilitate the set-up of machine learning projects (Interview A; Interview F). In contrast, two of the other experts argue that laws restrict access to data and raise the issue of obtaining consent for data processing (Interview C; Interview D).

4.2.3 Persuasion challenges

The persuasion challenges of integrating machine learning approaches in the radiology value chain depend on the technical and legal challenges and consist of the missing acceptance by radiologists and patients' distrust. Radiology has always undergone technological change, and radiologists have benefitted from better imaging modalities and improved medical software. Concerns of the radiologists about machine learning could hinder widespread adoption of this technology (Thrall et al., 2018). Besides technical integration issues, the radiologists' resistance to change hinders the widespread adoption of machine learning. A lack of trust in technology, which is also a result of the limited technical understanding and the negative experiences with CADe systems, partly explains this reluctance (Mayo and Leung, 2018; Ganeshan et al., 2018). In order to avoid the pitfalls of CADe systems and overcome the intrinsic resistance, machine learning products need to provide concrete evidence of workflow improvements such as time reduction or higher diagnostic accuracy (Blum and Zins, 2017). Furthermore, the lack of transparency in decision-making places radiologists in a predicament, as they have to explain the results to the referring physician and the patient (Lakhani et al., 2018). The experts remark that adoption might depend on the radiologist's age, cultural background and the degree of automation of the specific use case (Interview A; Interview E; Interview F). For instance, radiologists' fear of a machine replacing them is one of the main adoption barriers when deep learning first gained popularity (Syeda-Mahmood, 2018). However, this fear may vanish as the proposed products focus on assisting radiologists rather than fully automating their work (Interview F). Interestingly, the fear of replacement does not exist in developing markets such as China, which has a severe shortage of radiologists (Interview E).

While patients tend to have a positive attitude towards precision medicine in general, they share concerns about data privacy and the predictive analysis of their health information (Balthazar et al., 2018). Patients become the co-creators of machine learning models by providing their data voluntarily or involuntarily. Thus, if machine learning approaches do not fully address the concerns of patients, patient interest groups could prevent data collection and the adoption of machine learning in radiology (Tang et al., 2018). Also, many ethical questions concerning AI, in general, have not yet been answered, which only further reinforces patients' distrust (Pesapane et al., 2018). If the algorithm factors in data additional to health information, it could lead to unethical outcomes. For example, if costs are considered a relevant factor for treatment recommendations, patients with basic insurance might be recommended a less effective but cheaper treatment option for the same disease than people with premium insurance (Kohli et al., 2017). Other ethical concerns are about the adoption of systematic or implicit bias, because ethnic or economic minorities could be underrepresented in the training of the algorithm (Balthazar et al., 2018; Caliskan et al., 2017). One expert also encourages the involvement of relevant stakeholders (e.g., ethicists and patient advocates) to gain patients' trust (Interview F).

5 Discussion

The results of the literature review and the expert interviews provide a structured presentation of the various challenges and opportunities of machine learning in radiology. However, the question remains as to which implications our findings have on the radiology business model.

For radiology businesses that follow the operational excellence strategy (i.e., providing diagnostic services at the lowest possible price), the efficiency improvements will be most relevant. Better scanner utilisation, standardised reporting, and workflow automation already help to achieve economies of scale. Machine learning may foster efficiency by, amongst others, automating accompanying activities such as generating study protocols or reports, avoiding duplicate work due to low image quality, and supporting the radiologist. For example, replacing the second read in screening programs could reduce the highest cost drivers (Enzmann and Schomer, 2013; Obermeyer and Emanuel, 2016). Machine learning applications that yield efficiency improvements support radiology businesses such as teleradiology that focus on volume rather than value. Considering that this a common position of radiology businesses today and machine learning workflow tools are the first to enter the market, this will most likely be the short-term development of machine learning in radiology.

Radiology businesses seeking product or service leadership benefit from effectivity improvements, primarily through large-scale analytics as well as quantitative information and data integration. Machine learning applications improve the effectiveness of patient care by, amongst others, superior image quality, better processing of images (e.g., registration, segmentation), additional information, and computer-aided detection, classification, interpretation, and integration. The trend towards evidence-based medicine creates an environment where radiology – as an information business – could strive to make better medical decisions by providing quantitative, actionable information (Enzmann and Schomer, 2013). Radiology businesses collaborating with pathology departments can develop machine learning-based radiomics applications to provide innovative and integrated solutions that ensure product leadership.

In recent years, the 'value-based healthcare' concept (i.e., focusing on health outcome rather than single diagnosis or treatment) has gained popularity (Putera, 2017). Radiology businesses concentrating on customer intimacy are likely to flourish in a value-based health care system. Integrating imaging as well as non-imaging data and using machine learning algorithms to predict patients' treatment responses enables accurate and individual treatment planning. Non-image data can even include health-related data from wearables such as average heart rate, sleeping behaviour, and activity levels. The reintegration of diagnostic and interventional radiology would be highly profitable as it encompasses the entire diagnostic and therapeutic value chain. In addition, sharing data between different medical disciplines and health information systems will blur the boundaries of radiology. Radiology businesses that focus on collecting standardised, high-quality, and well-annotated data may sell the datasets.

In Figure 4, we summarise the opportunities that are most likely to affect the value propositions of the radiology business model of Enzmann and Schomer (2013). All models will profit from machine learning applications either by effectivity or efficiency improvements, whereas the latter seems to have the most impact recently. Besides that, the opportunities of machine learning do not exclusively favour individual value propositions. For example, data integration benefits product and service leadership as well as customer intimacy.

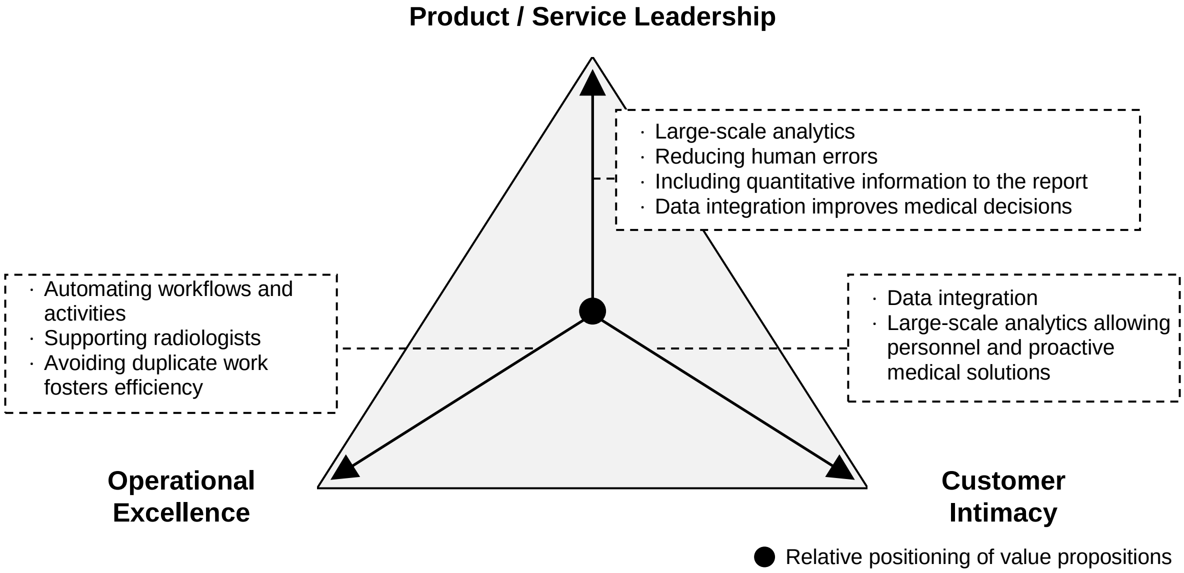

Figure 4. The impact of machine learning on the value propositions of the business model.

Effectivity improvements will play a major role once the radiology business overcomes the main challenges including the lack of data and the generalisation problem. Nonetheless, radiology businesses should consider all identified challenges as they affect every value proposition to a greater or lesser extent. The technical challenges pose the main hurdle and directly influence the legal and persuasion challenges. Our findings identify the lack of data as the bottleneck of the development and integration of machine learning in clinical practice. High-scale and high-quality data would improve the performance and therefore facilitate overcoming regulatory barriers as well as gaining the trust of the radiologists. While overcoming the lack of data appears to be only a matter of time, the generalisation problem presents a major obstacle and will require further research. This will determine whether machine learning applications enable the deployment in a more general context or remain limited to specific use cases. Legal challenges such as liability and intellectual property as well as ethical concerns are still a major obstacle due to the black box problem and a missing validation framework. The persuasion challenges partly result from the two previous challenges. Once machine learning can show its clinical value and receives approval by the regulatory bodies, the adoption rate is very likely to increase, while research and practice need to address the concerns of physicians and patients to achieve widespread distribution.

6 Conclusion

This research provides a comprehensive and systematic overview of the opportunities and challenges of machine learning in radiology by analysing primary and secondary data and discusses the implications on the radiology business models using the identified opportunities and challenges. Thereby, our findings do not only contribute to academic discourse in the emerging research field of health IT, but also supports decision-makers in healthcare in general and radiology in particular. Our study illustrates that information systems approaches, such as machine learning, are useful instruments in the field of healthcare and medicine as they contain the potential to enhance the effectivity of medicine and the efficiency of the radiology value chain. Almost every activity of the radiology value chain presents a potential use case for machine learning applications. Machine learning can improve the diagnostic quality by reducing human errors, analysing large amounts of data accurately, integrating quantitative information to the report, and integrating non-image data. Besides improving diagnostic quality, these opportunities enable better predictions, and thus enhance the value of the medical answer and consequently improve the effectiveness of patient care. In addition, the automation of the radiology workflow through machine learning reduces costs and time demand. However, the opportunities face technical, legal, and persuasive challenges that research, practicing radiologists, and regulators need to address.

Our study further contributes to the body of knowledge by discussing the implications of the identified opportunities and challenges for the radiology business model. Our findings shed light on the strategic positioning of radiology businesses regarding the value propositions operational excellence, product respectively service leadership, and customer intimacy. Academic literature and experts both agree: Machine learning will have a major impact on the radiology business and patient care. Radiologists who fail to recognise the opportunities of this technology and its value contribution to patient care will eventually face the risk of commoditisation and marginalisation. Therefore, our findings are not limited to radiology as other image-analysing industries and other medical application areas may face the same impact on their businesses.

Although rigorously following our research approach, our study has some limitations. First, our literature research is limited by the number of relevant and available publications. Most of the literature gives a general overview of machine learning in radiology rather than explicitly addressing the challenges and opportunities. Even though radiology has been using machine learning for a long time, there has only recently been an interest to evaluate the overall impact of machine learning in this field. Consequently, it remains ambiguous how and to what extent machine learning will have an impact in the future. To overcome this shortcoming, we validated our literature review findings with expert interviews to gain a deeper understanding of the challenges and opportunities. Secondly, relevant scholars often use small sample sizes in their – more or less isolated – experiments and are therefore far from regulatory approval. It is important to challenge whether trained models maintain their performance when deployed in clinical practice. Nevertheless, the sheer quantity and variety of applications indicate the impact of machine learning on the value chain. The third limitation lies in the qualitative research itself since we cannot generalise the findings with a similar certainty as quantitative analysis can (Atieno, 2009). According to Bertaux (1981), the "saturation of knowledge" depicts the necessary number of corresponding interviews. During the interviews, the experts could not mention new challenges and opportunities, indicating that our interviews reach the point of saturation. In addition, we do not raise the opinions of all stakeholders so that the opinion and expertise of patient advocates, regulators, experts from other countries (e.g., US and Asia) are missing. To overcome this limitation, the next step in our research endeavour will be to increase the number of interview participants and include participants across all relevant areas. Besides that, future research could quantitatively survey the perception of machine learning in radiology and the shift between business models or, more specifically, their value propositions. As far as this work is concerned, further research should provide more information for each of the opportunities and challenges. In particular, future research should discuss solutions to the identified challenges. Moreover, quantitative research to determine the economic impact of machine learning in radiology could help to convince stakeholders not only of its medical benefits but also of its economic effectivity.

References

Abdi, A. H., C. Luong, T. Tsang, G. Allan, S. Nouranian, J. Jue, D. Hawley, S. Fleming, K. Gin, J. Swift, R. Rohling and P. Abolmaesumi (2017). "Automatic Quality Assessment of Echocardiograms Using Convolutional Neural Networks: Feasibility on the Apical Four-Chamber View." *IEEE transactions on medical imaging* 36 (6), 1221–1230.

Afshar, P., A. Mohammadi, K. N. Plataniotis, A. Oikonomou and H. Benali (2018). "From Hand-Crafted to Deep Learning-based Cancer Radiomics: Challenges and Opportunities."

Akgül, C. B., D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan and B. Acar (2011). "Content-based image retrieval in radiology: current status and future directions." *Journal of digital imaging* 24 (2), 208–222.

Akkus, Z., A. Galimzianova, A. Hoogi, D. L. Rubin and B. J. Erickson (2017). "Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions." *Journal of digital imaging* 30 (4), 449–459.

Armato, S. G., M. L. Giger and H. MacMahon (2001). "Automated detection of lung nodules in CT scans: preliminary results." *Medical physics* 28 (8), 1552–1561.

Atieno, O. P. (2009). "An Analysis of the Strengths and Limitation of Qualitative and Quantitative Research Paradigms." *Problems of Education in the 21st century* 13 (1), 13–18.

Babu-Narayan, S. V., G. Giannakoulas, A. M. Valente, W. Li and M. A. Gatzoulis (2016). "Imaging of congenital heart disease in adults." *European heart journal* 37 (15), 1182–1195.

Bahrami, K., I. Rekik, F. Shi and D. Shen (2017). "Joint Reconstruction and Segmentation of 7T-like MR Images from 3T MRI Based on Cascaded Convolutional Neural Networks." In: *Medical Image Computing and Computer Assisted Intervention − MICCAI 2017*. Ed. by M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins and S. Duchesne. Springer, p. 764–772.

Balthazar, P., P. Harri, A. Prater and N. M. Safdar (2018). "Protecting Your Patients' Interests in the Era of Big Data, Artificial Intelligence, and Predictive Analytics." *Journal of the American College of Radiology* 15 (3), 580–586.

Bauer, S., R. Wiest, L.-P. Nolte and M. Reyes (2013). "A survey of MRI-based medical image analysis for brain tumor studies." *Physics in Medicine and Biology* 58 (13), R97-R129.

Becker, A. S., M. Marcon, S. Ghafoor, M. C. Wurnig, T. Frauenfelder and A. Boss (2017). "Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer." *Investigative radiology* 52 (7), 434–440.

Bertaux, D. (1981). *Biography and society:* The life history approach in the social sciences. Beverly Hills, USA: Sage Publications.

Bettis, R. A., A. Gambardella, C. Helfat and W. Mitchell (2015). "Qualitative empirical research in strategic management." *Strategic Management Journal* 36 (5), 637–639.

Bhattacherjee, A. (2012). *Social science research:* Principles, methods, and practices. North Charleston, USA: CreateSpace Independent Publishing Platform.

Bischoff-Grethe, A., I. B. Ozyurt, E. Busa, B. T. Quinn, C. Fennema-Notestine, C. P. Clark, S. Morris, M. W. Bondi, T. L. Jernigan, A. M. Dale, G. G. Brown and B. Fischl (2007). "A Technique for the Deidentification of Structural Brain MR Images." *Human brain mapping* 28 (9), 892–903.

Blum, A. and M. Zins (2017). "Radiology: Is its future bright?" *Diagnostic and interventional imaging* 98 (5), 369–371.

Bogner, A., B. Littig and W. Menz (2009). "Introduction: Expert Interviews — An Introduction to a New Methodological Debate." In: *Interviewing experts*. Ed. by A. Bogner, B. Littig and W. Menz. London, UK: PALGRAVE MACMILLAN, p. 1–13.

Brady, A. P. (2017). "Error and discrepancy in radiology: inevitable or avoidable?" *Insights Imaging* 8 (1), 171–182.

Brink, J. A., R. L. Arenson, T. M. Grist, J. S. Lewin and D. Enzmann (2017). "Bits and bytes: The future of radiology lies in informatics and information technology." *European radiology* 27 (9), 3647–3651.

Brown, A. D. and T. R. Marotta (2018). “Using machine learning for sequence-level automated MRI protocol selection in neuroradiology.” *Journal of the American Medical Informatics Association JAMIA* 25 (5), 568–571.

Bruijne, M. de (2016). “Machine learning approaches in medical image analysis: From detection to diagnosis.” *Medical image analysis* 33, 94–97.

Burt, J. R., N. Torosdagli, N. Khosravan, H. RaviPrakash, A. Mortazi, F. Tissavirasingham, S. Hussein and U. Bagci (2018). “Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks.” *The British journal of radiology* 91, 20170545.

Cabitza, F., R. Rasoini and G. F. Gensini (2017). “Unintended Consequences of Machine Learning in Medicine.” *JAMA* 318 (6), 517.

Caliskan, A., J. J. Bryson and A. Narayanan (2017). “Semantics derived automatically from language corpora contain human-like biases.” *Science* 356 (6334), 183–186.

Carneiro, G., Y. Zheng, F. Xing and L. Yang (2017). “Review of Deep Learning Methods in Mammography, Cardiovascular, and Microscopy Image Analysis.” In: *Deep Learning and Convolutional Neural Networks for Medical Image Computing: Precision Medicine, High Performance and Large-Scale Datasets*. Ed. by Le Lu, Y. Zheng, G. Carneiro and L. Yang. Cham, Switzerland: Springer International Publishing, p. 11–32.

Chen, H., Y. Xu, Y. Ma and B. Ma (2010). “Neural network ensemble-based computer-aided diagnosis for differentiation of lung nodules on CT images: clinical evaluation.” *Academic Radiology* 17 (5), 595–602.

Chockley, K. and E. Emanuel (2016). “The End of Radiology? Three Threats to the Future Practice of Radiology.” *Journal of the American College of Radiology* 13 (12), 1415–1420.

Cohen, A. L., S. L. Holmen and H. Colman (2013). “IDH1 and IDH2 mutations in gliomas.” *Current neurology and neuroscience reports* 13 (5), 345.

Dhungel, N., G. Carneiro and A. P. Bradley (2015). “Automated Mass Detection in Mammograms Using Cascaded Deep Learning and Random Forests.” In: *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA): Adelaide, Australia, 23-25 November 2015*. Ed. by I. C. o. D. I. C.: T. a. Applications. Piscataway, NJ: IEEE, p. 1–8.

Dwyer, D. B., P. Falkai and N. Koutsouleris (2018). “Machine Learning Approaches for Clinical Psychology and Psychiatry.” *Annual review of clinical psychology* 14, 91–118.

Enzmann, D. R. (2012). “Radiology's value chain.” *Radiology* 263 (1), 243–252.

Enzmann, D. R. and D. F. Schomer (2013). “Analysis of radiology business models.” *Journal of the American College of Radiology* 10 (3), 175–180.

Esses, S. J., X. Lu, T. Zhao, K. Shanbhogue, B. Dane, M. Bruno and H. Chandarana (2018). “Automated image quality evaluation of T2 -weighted liver MRI utilizing deep learning architecture.” *Journal of Magnetic Resonance Imaging* 47 (3), 723–728.

Flick, U. (2014). *An introduction to qualitative research.* 5th Edition. Los Angeles, USA: Sage.

Ganeshan, D., P.-A. T. Duong, L. Probyn, L. Lenchik, T. A. McArthur, M. Retrouvey, E. H. Ghobadi, S. L. Desouches, D. Pastel and I. R. Francis (2018). “Structured Reporting in Radiology.” *Academic Radiology* 25 (1), 66–73.

Gibson, E., W. Li, C. Sudre, L. Fidon, D. I. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, T. Whyntie, P. Nachev, M. Modat, D. C. Barratt, S. Ourselin, M. J. Cardoso and T. Vercauteren (2018). “NiftyNet: a deep-learning platform for medical imaging.” *Computer Methods and Programs in Biomedicine* (158), 113–122.

Gichoya, J. W., S. Nuthakki, P. G. Maity and S. Purkayastha (2018). *Phronesis of AI in radiology: Superhuman meets natural stupidity*. URL: http://arxiv.org/pdf/1803.11244v1 (visited on 11/27/2018).

Gillies, R. J., P. E. Kinahan and H. Hricak (2016). “Radiomics: Images Are More than Pictures, They Are Data.” *Radiology* 278 (2), 563–577.

Goh, V., B. Ganeshan, P. Nathan, J. K. Juttla, A. Vinayan and K. A. Miles (2011). “Assessment of response to tyrosine kinase inhibitors in metastatic renal cell cancer: CT texture as a predictive biomarker.” *Radiology* 261 (1), 165–171.

Gondara, L. (2016). *Medical image denoising using convolutional denoising autoencoders*. URL: https://arxiv.org/pdf/1608.04667.pdf (visited on 11/27/2018).

Hope, T. M. H., M. L. Seghier, A. P. Leff and C. J. Price (2013). “Predicting outcome and recovery after stroke with lesions extracted from MRI images.” *NeuroImage. Clinical* 2, 424–433.

Jiang, D., W. Dou, L. Vosters, X. Xu, Y. Sun and T. Tan (2017). *Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network*. URL: http://arxiv.org/pdf/1712.08726v2 (visited on 11/27/2018).

Jing, B., P. Xie and E. Xing (2017). *On the Automatic Generation of Medical Imaging Reports*. URL: http://arxiv.org/pdf/1711.08195v3 (visited on 11/27/2017).

Kang, E., J. Min and J. C. Ye (2017). “A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction.” *Medical physics* 44 (10), e360-e375.

Kickingereder, P., S. Burth, A. Wick, M. Gotz, O. Eidel, H.-P. Schlemmer, K. H. Maier-Hein, W. Wick, M. Bendszus, A. Radbruch and D. Bonekamp (2016). “Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models.” *Radiology* 280 (3), 880–889.

Kim, D. W., H. Y. Jang, K. W. Kim, Y. Shin and S. H. Park (2019). “Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers.” *Korean journal of radiology* 20 (3), 405–410.

Kohli, M., L. M. Prevedello, R. W. Filice and J. R. Geis (2017). “Implementing Machine Learning in Radiology Practice and Research.” *American journal of roentgenology* 208 (4), 754–760.

Kumar, D., M. J. Shafiee, A. G. Chung, F. Khalvati, M. A. Haider and A. Wong (2015). *Discovery Radiomics for Pathologically-Proven Computed Tomography Lung Cancer Prediction*. URL: http://arxiv.org/pdf/1509.00117v3 (visited on 11/27/2018).

Lakhani, P., A. B. Prater, R. K. Hutson, K. P. Andriole, K. J. Dreyer, J. Morey, L. M. Prevedello, T. J. Clark, J. R. Geis, J. N. Itri and C. M. Hawkins (2018). “Machine Learning in Radiology: Applications Beyond Image Interpretation.” *Journal of the American College of Radiology* 15 (2), 350–359.

Lambin, P., E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. P. M. van Stiphout, P. Granton, C. M. L. Zegers, R. Gillies, R. Boellard, A. Dekker and H. J. W. L. Aerts (2012). “Radiomics: extracting more information from medical images using advanced feature analysis.” *European journal of cancer* 48 (4), 441–446.

Levitan, E. and G. T. Herman (1987). “A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography.” *IEEE transactions on medical imaging* 6 (3), 185–192.

Li, Z., Y. Wang, J. Yu, Y. Guo and W. Cao (2017). “Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma.” *Scientific reports* 7 (1), 1–11.

Liao, F., X. Chen, X. Hu and S. Song (2017a). “Estimation of the volume of the left ventricle from MRI images using deep neural networks.” *IEEE Transactions on Cybernetics*, 1–10.

Liao, R., S. Miao, P. d. Tournemire, S. Grbic, A. Kamen, T. Mansi and D. Comaniciu (2017b). *An Artificial Agent for Robust Image Registration*. URL: http://arxiv.org/pdf/1611.10336v1 (visited on 11/27/2018).

Litjens, G., T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken and C. I. Sánchez (2017). “A Survey on Deep Learning in Medical Image Analysis.” *Medical image analysis* 42, 60–88.

Masic, I., M. Miokovic and B. Muhamedagic (2008). “Evidence based medicine - new approaches and challenges.” *Acta Informatica Medica Journal of the Society for Medical Informatics of Bosnia & Herzegovina* 16 (4), 219–225.

Massat, M. B. (2018). “Artificial intelligence in radiology: Hype or hope?” *Applied Radiology* 2018, 22–26.

Mayo, R. C. and J. Leung (2018). “Artificial intelligence and deep learning – Radiology's next frontier?” *Clinical Imaging* 49, 87–88.

Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt: SSOAR.

Mazurowski, M. A., M. Buda, A. Saha and M. R. Bashir (2018). *Deep learning in radiology:* An overview of the concepts and a survey of the state of the art. URL: http://arxiv.org/pdf/1802.08717v1 (visited on 11/27/2018).

McDonald, R. J., K. M. Schwartz, L. J. Eckel, F. E. Diehn, C. H. Hunt, B. J. Bartholmai, B. J. Erickson and D. F. Kallmes (2015). "The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload." *Academic Radiology* 22 (9), 1191–1198.

Muelly, M., P. B Stoddard and S. S Vasanwala (2017). *Using machine learning with dynamic exam block lengths to decrease patient wait time and optimize MRI schedule fill rate*. URL: http://archive.ismrm.org/2017/4343.html (visited on 11/27/2018).

Müller, H., N. Michoux, D. Bandon and A. Geissbuhler (2004). "A review of content-based image retrieval systems in medical applications-clinical benefits and future directions." *International journal of medical informatics* 73 (1), 1–23.

Nelson, H. D., M. Pappas, A. Cantor, J. Griffin, M. Daeges and L. Humphrey (2016). "Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation." *Annals of internal medicine* 164 (4), 256–267.

Obermeyer, Z. and E. J. Emanuel (2016). "Predicting the Future - Big Data, Machine Learning, and Clinical Medicine." *The New England journal of medicine* 375 (13), 1216–1219.

Oliveira, J. E. E. de, A. de Albuquerque Araújo and T. M. Deserno (2011). "Content-based image retrieval applied to BI-RADS tissue classification in screening mammography." *World journal of radiology* 3 (1), 24–31.

Paliwal, M. and U. A. Kumar (2009). "Neural networks and statistical techniques: A review of applications." *Expert Systems with Applications* 36 (1), 2–17.

Panetta, K. (2018). *5 Trends Emerge in the Gartner Hype Cycle for Emerging Technologies, 2018*. URL: https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/ (visited on 11/27/2018).

Pesapane, F., C. Volonté, M. Codari and F. Sardanelli (2018). "Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States." *Insights Imaging* 9 (5), 745–753.

Putera, I. (2017). "Redefining Health: Implication for Value-Based Healthcare Reform." *Cureus* 9 (3), e1067.

Ressing, M., M. Blettner and S. J. Klug (2009). "Systematic literature reviews and meta-analyses: part 6 of a series on evaluation of scientific publications." *Deutsches Arzteblatt international* 106 (27), 456–463.

Ribli, D., A. Horváth, Z. Unger, P. Pollner and I. Csabai (2018). "Detecting and classifying lesions in mammograms with Deep Learning." *Scientific Reports* 8 (1), 14.

Shi, H., P. Xie, Z. Hu, M. Zhang and E. P. Xing (2017). *Towards Automated ICD Coding Using Deep Learning*. URL: http://arxiv.org/pdf/1711.04075v3 (visited on 11/27/2018).

Shin, B., F. H. Chokshi, T. Lee and J. D. Choi (2017). "Classification of Radiology Reports Using Neural Attention Models." *2017 International Joint Conference on Neural Networks*, 4363–4370.

Shiraishi, J., Q. Li, D. Appelbaum and K. Doi (2011). "Computer-Aided Diagnosis and Artificial Intelligence in Clinical Imaging." *Seminars in Nuclear Medicine* 41 (6), 449–462.

Silverman, L. (2017). *Scanning The Future, Radiologists See Their Jobs At Risk*. URL: https://www.npr.org/sections/alltechconsidered/2017/09/04/547882005/scanning-the-future-radiologists-see-their-jobs-at-risk?t=1535784563867 (visited on 11/27/2018).

Slomka, P. J., D. Dey, A. Sitek, M. Motwani, D. S. Berman and G. Germano (2017). "Cardiac imaging: working towards fully-automated machine analysis & interpretation." *Expert review of medical devices* 14 (3), 197–212.

Suzuki, K. (2017). "Overview of deep learning in medical imaging." *Radiological Physics and Technology* 10 (3), 257–273.

Syeda-Mahmood, T. (2018). “Role of Big Data and Machine Learning in Diagnostic Decision Support in Radiology.” *Journal of the American College of Radiology* 15 (3 Pt B), 569–576.

Tang, A., R. Tam, A. Cadrin-Chênevert, W. Guest, J. Chong, J. Barfett, L. Chepelev, R. Cairns, J. R. Mitchell, M. D. Cicero, M. G. Poudrette, J. L. Jaremko, C. Reinhold, B. Gallix, B. Gray and R. Geis (2018). “Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology.” *Canadian Association of Radiologists Journal* 69 (2), 120–135.

Thrall, J. H., X. Li, Q. Li, C. Cruz, S. Do, K. Dreyer and J. Brink (2018). “Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success.” *Journal of the American College of Radiology* 15 (3 Pt B), 504–508.

Treacy, M. and F. D. Wiersema (1997). *The discipline of market leaders:* Choose your customers, narrow your focus, dominate your market. [Expanded edition]. Reading, Mass: Addison-Wesley Pub. Co.

Vaishali, S., K. K. Rao and G. V. S. Rao (2015). “A review on noise reduction methods for brain MRI images.” In: *2015 International Conference on Signal Processing and Communication Engineering Systems*. IEEE, p. 363–365.

Wang, J., X. Yang, H. Cai, W. Tan, C. Jin and L. Li (2016). “Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning.” *Scientific reports* 6, 27327.

Wang, S. and R. M. Summers (2012). “Machine learning and radiology.” *Medical image analysis* 16 (5), 933–951.

Weese, J. and C. Lorenz (2016). “Four challenges in medical image analysis from an industrial perspective.” *Medical image analysis* (33), 44–49.

Wolterink, J. M., T. Leiner, B. D. de Vos, R. W. van Hamersvelt, M. A. Viergever and I. Isgum (2016). “Automatic coronary artery calcium scoring in cardiac CT angiography using paired convolutional neural networks.” *Medical image analysis* 34, 123–136.

Xie, P. and E. Xing (2018). “A Neural Architecture for Automated ICD Coding.” In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, p. 1066–1076.

Yasaka, K., H. Akai, A. Kunimatsu, S. Kiryu and O. Abe (2018). “Deep learning with convolutional neural network in radiology.” *Japanese journal of radiology* 36 (4), 257–272.

Yoo, Y., L. W. Tang, T. Brosch, D. K. B. Li, L. Metz, A. Traboulsee and R. Tam (2016). “Deep Learning of Brain Lesion Patterns for Predicting Future Disease Activity in Patients with Early Symptoms of Multiple Sclerosis.” In: *Deep Learning and Data Labeling for Medical Applications*. Ed. by G. Carneiro, D. Mateus, L. Peter, A. Bradley, J. M. R. S. Tavares, V. Belagiannis, J. P. Papa, J. C. Nascimento, M. Loog, Z. Lu, J. S. Cardoso and J. Cornebise. Cham, Switzerland: Springer International Publishing, p. 86–94.