


REVIEW

Challenges and perspectives in use of artificial intelligence to support treatment recommendations in clinical oncology

Gregor Duwe¹  | Dominique Mercier² | Crispin Wiesmann¹ | Verena Kauth¹ | Kerstin Moench¹ | Markus Junker² | Christopher C. M. Neumann³ | Axel Haferkamp¹ | Andreas Dengel² | Thomas Höfner^{1,4}

¹Department of Urology and Pediatric Urology, University Medical Center, Johannes Gutenberg University, Mainz, Germany

²Research Unit Smart Data and Knowledge Services, German Research Center for Artificial Intelligence, Kaiserslautern, Germany

³Department of Hematology, Oncology and Tumor Immunology, Charité-Universitätsmedizin Berlin, Freie Universität Berlin, Humboldt-Universität zu Berlin, Berlin, Germany

⁴Department of Urology, Ordensklinikum Linz Elisabethinen, Linz, Austria

Correspondence

Gregor Duwe, Department of Urology and Pediatric Urology, University Medical Center, Johannes Gutenberg University, Langenbeckstrasse 1, 55131 Mainz, Germany.

Email: gregor.duwe@unimedizin-mainz.de

Funding information

Bundesministerium für Bildung und Forschung, Grant/Award Number: 16SV9053

Abstract

Artificial intelligence (AI) promises to be the next revolutionary step in modern society. Yet, its role in all fields of industry and science need to be determined. One very promising field is represented by AI-based decision-making tools in clinical oncology leading to more comprehensive, personalized therapy approaches. In this review, the authors provide an overview on all relevant technical applications of AI in oncology, which are required to understand the future challenges and realistic perspectives for decision-making tools. In recent years, various applications of AI in medicine have been developed focusing on the analysis of radiological and pathological images. AI applications encompass large amounts of complex data supporting clinical decision-making and reducing errors by objectively quantifying all aspects of the data collected. In clinical oncology, almost all patients receive a treatment recommendation in a multidisciplinary cancer conference at the beginning and during their treatment periods. These highly complex decisions are based on a large amount of information (of the patients and of the various treatment options), which need to be analyzed and correctly classified in a short time. In this review, the authors describe the technical and medical requirements of AI to address these scientific challenges in a multidisciplinary manner. Major challenges in the use of AI in oncology and decision-making tools are data security, data representation, and explainability of AI-based outcome predictions, in particular for decision-making processes in multidisciplinary cancer conferences. Finally, limitations and potential solutions are described and compared for current and future research attempts.

KEYWORDS

artificial intelligence, clinical oncology, genitourinary cancer, multidisciplinary cancer conferences, treatment recommendation

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Cancer Medicine* published by John Wiley & Sons Ltd.

1 | STATE OF RESEARCH AND POTENTIAL USE OF ARTIFICIAL INTELLIGENCE IN TREATMENT RECOMMENDATION IN CLINICAL ONCOLOGY

1.1 | Treatment recommendations in multidisciplinary cancer conferences

The field of clinical oncology has become more and more complex over the last decades. Recommending the best available treatment for patients with oncological diseases is a central and complex task. This is commonly conducted in multidisciplinary cancer conferences (MCCs). In these conferences, physicians from various different backgrounds sit together to make comprehensive and complex therapy decisions. The ever more emerging complexity in clinical oncology has made these decisions more difficult as regular updates change first-, second- or third-line treatments all across tumor entities.^{1,2} For this purpose, national and international medical guidelines offer regularly updated references of medical knowledge for every subspecialty and most common diseases, in order to provide evidence-based medicine.^{3,4} Medical guidelines represent the basis for treatment decisions made in MCC or by individual physicians.^{5–8}

Artificial intelligence (AI) has the potential to reduce this complexity for physicians while increasing the level of evidence-based treatment recommendations.⁹ We assume it will be possible over the next decade to support

physicians in their daily treatment recommendations by generating an objective, evidence-based AI-generated treatment recommendation (Figure 1).^{10,11} For instance, there is already the possibility to generate treatment options by using large language models, but their quality remains questionable.¹² Therefore, it is important that physicians obtain a medical and technical understanding of how AI treatment decision support systems (could) work, how to interpret these results and to understand the current limitations of this new, groundbreaking technology. Consequently, this review aims to give a broad overview for physicians on these computer science-specific aspects as well as an overview on current applications, future possibilities, and hurdles that need to be overcome. We want physicians to be able to independently understand the technical principles in order to critically scrutinize them before applications are implemented in their everyday lives in the future.

1.2 | Predicting the clinical success of treatments in clinical oncology

The decision for the right therapy for any cancer disease is always accompanied by the predicted success of a therapy. AI applications are already being used as predictive therapy platforms.¹³ Previous studies using AI to predict treatment response differ in two major outcomes: first, which treatment is being predicted, and second, which source of information is being used to train the AI for the prediction.

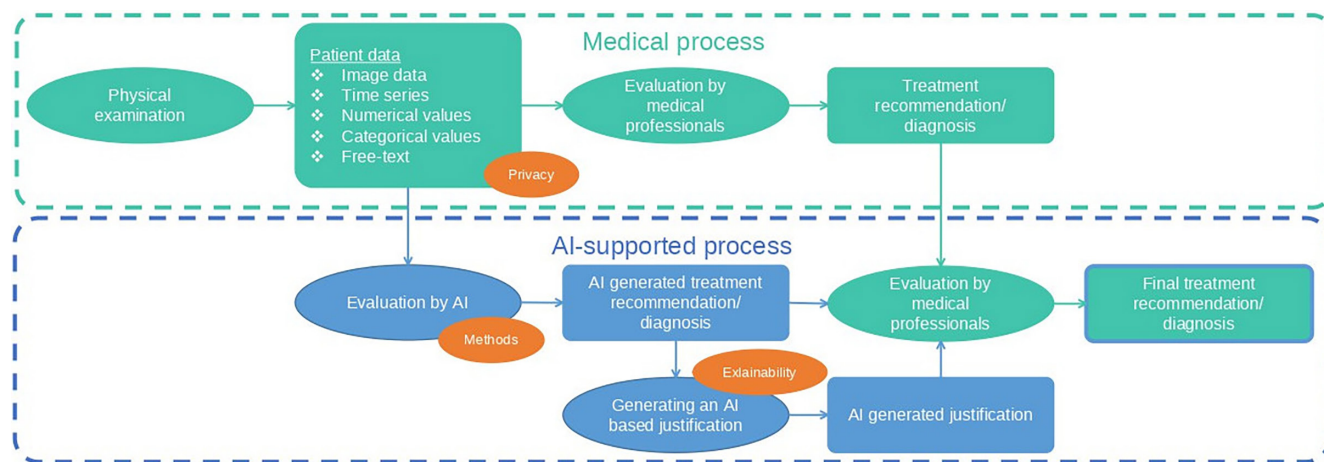


FIGURE 1 Extension of the medical process by an AI-supported process, which enables medical staff to support diagnosis and treatment recommendation. This shows an example of a process in which the medical process of generating a treatment recommendation is supported by an AI process. The basic idea is to perform validation after the process, which minimizes the risk of error in the recommendation. First, investigations are performed to collect patient data. This can be any type of relevant data. Next, evaluation is performed by medical professionals and a recommendation is made. The patient data would be evaluated using an AI while maintaining privacy. The AI makes a recommendation, which is presented to the healthcare professional- Ideally, an explanation should be generated by the AI that allows the recommendation proposal to be understood. After the evaluation by the specialist, a final recommendation can be made by the medical professionals, which might include suggestions or be supported by the AI-generated recommendation.

For both goals, radiological and pathological images as well as clinical parameters can be used as data sources. For example, there are two approaches of AI development based on analyzing histological images. In the first approach, the AI is trained to recognize already known indicators in histological images.^{14,15} This approach is considered less complex because it requires only a small cohort of training images and allows faster prediction based on known indicators. In the second approach, the AI is trained with a much larger cohort of patients, their histological images, and additionally, clinical information of their treatment courses. After the deployment phase, both approaches use the input provided by the doctors directly and present them the results. From user perspective, these systems can be used without any adjustments which provides an “end-to-end workflow” as the methodology is a black box for the end-user. It offers the chance of predicting treatment success in patients for whom this would not have been possible using already known indicators.^{14,15} Currently, only few studies on this approach exist as this requires large datasets to train the AI,¹⁴ for example, by predicting the success of immunotherapy in non-small cell lung cancer.^{14,16} In contrast, there are various clinical studies on the first approach, most commonly of histological biomarker for immunohistochemical and genetic analysis.^{17–20}

Until now, radiology offers the broadest range of applications for AI in medicine and oncology, in particular through radiomics.²¹ Radiomics represents a method for the quantitative description of medical images. For example, it is already possible that an AI trained on diffusion-weighted magnetic resonance images before and after (neoadjuvant) chemotherapy can predict response to therapy in terms of tumor shrinkage after chemotherapy.^{22–24} By such approaches, it would be possible to modify or stop chemotherapy in a more individual manner based on treatment success and adverse events.^{25–31} Recently, scientific approaches of AI were reported in tumor genomics. Here, AI approaches were able to arrange tumor samples based on their RNA and tumor characteristics in a multi-dimensional space³² as well as to predict response to chemotherapy based on tumor RNA.³³

Moreover, AI applications can be used to predict the survival of cancer patients, representing key information for any treatment decisions in oncology in order to adapt the treatment regime to achieve the highest possible survival outcome as well as quality of life.^{34–36} By using data that can be extracted from most medical records, it has already been possible to train AIs that have improved predictions of overall survival for patients.^{37,38} Moreover, a frequently used factor for AI development are patient-reported outcome measurements which may lead to decrease of mortality, while increasing sensitivity and specificity of the AI.³⁴

1.3 | Use of artificial intelligence applications in medical treatment recommendation

In the past two decades, only few approaches of computer-based decision support systems have been implemented.⁹ An early example is “OncoDoc,” which was developed for breast cancer patients.³⁹ By using a decision tree model based on medical guidelines, “OncoDoc” recommends a treatment option and explains the decision by showing the individual decision tree. However, a case study of OncoDoc’s successor “OncoDoc2” showed that 21.3% of the decisions were incorrect and did not follow the medical guidelines.⁴⁰ Another example is the “DESIRRE” project which was running from 2016 until 2020.⁴¹ This computer-based decision support system implements different medical guidelines, patient similarity, and the information’s and decisions from previous cases in a rule-based engine, which then gives different treatment recommendations for breast cancer patients.⁴² Until now, no systematic, reliable results have been published to discuss its potential in clinical practice.⁴³

The model “Watson for Oncology” (WFO), developed by IBM (International Business Machines Corporation, USA), represents one of the most relevant models in AI Oncology.^{44,45} The goal of WFO was to extract data from any type of medical record in order to make a treatment recommendation based on the most current evidence.⁴⁶ However, the use of WFO in routine clinical practice revealed various problems. One of the major problems was to extract the correct data from medical records. Therefore, despite the market implementation, WFO was still inferior to a physician in interpreting medical texts.⁴⁴ A major hurdle of WFO was the limiting use of the software outside the United States.⁴⁷ The reason for this was that WFO was primarily developed using US guidelines and patient data from a single hospital (*Memorial Sloan Kettering Cancer Center, New York, USA*).⁴⁵ This way, WFO was not accepted in real-world clinical applications across US states or other countries.^{47–49} Another major problem with WFO was the “black box,” meaning that WFO could not justify the decisions it made.^{46,50} Because of this and other problems, IBM discontinued the WFO program on January 7, 2022.

A new approach that includes explainability is the KITTU project.⁵¹ The major goal of this multidisciplinary research project of physicians and experts in AI development is to support multidisciplinary treatment recommendations for patients with genitourinary cancer by using AI applications (Figure 4). One of the most important aspects in this context is to explain the recommendation of the system. With the help of the KITTU project, evidence-based treatment recommendation in oncology could be

increased to improve the quality of treatment and long-term survival of patients.

2 | CURRENT STATUS OF TECHNICAL APPLICATIONS OF ARTIFICIAL INTELLIGENCE DEVELOPMENT IN DECISION-MAKING TOOLS IN CLINICAL ONCOLOGY

The following chapter provides a dedicated overview of existing machine learning approaches in the medical domain with focus in clinical oncology, including relevant examples of use. Furthermore, it provides detailed information about the key challenges and important technical aspects when establishing machine learning in the specific domain of clinical oncology. The following subsections included on possible AI-architectures, explainability, and privacy as those are the three most important technical, medical, and ethical fields to consider for future developments of a decision support systems in clinical oncology.

2.1 | Current state of machine learning in medical applications

Since clinical decisions have become more and more complex, machine learning (ML) offers great potential to assist decision-making. Especially deep structures/deep

learning (DL) (Figures 2 and 3) which do not require feature engineering and depend on representation learning to infer data have shown to be successful.⁵² Importantly, these networks do not use pre-defined rules, but rather work with exemplary data and results. Datasets are used to train computational models, which involve mathematical optimization of neurons. AI models, such as deep neural networks (consisting of many nodes), receive information, process the data accordingly, and make recommendations. Exemplary applications include images of skin lesions to detect skin cancer: during a one-time training phase, the network is shown images and the corresponding information for each image showing skin cancer or not.⁵³ Systems like these process information patterns and is learning to make conclusions on the existence of skin cancer.

Besides the given examples, DL methods are not limited to image data and also work with tabular values, videos, audio, texts, or other sequences.⁵⁴ Potential applications include tasks such as classifications, prediction of values, detection of anomalies, object recognition/labeling, and transformations in the broadest sense. Neural networks require large amounts of data to be trained. Thus, the increasing availability of data and computing power of computer systems led to an increasing interest in AI methods.⁵⁵ DL in medical research is a challenging task for AI experts, because framing the medical problem involves separating two different technical terms for DL. The challenge lies on breaking down complex (human) tasks to a level that is understandable to machine language, but still includes the problem

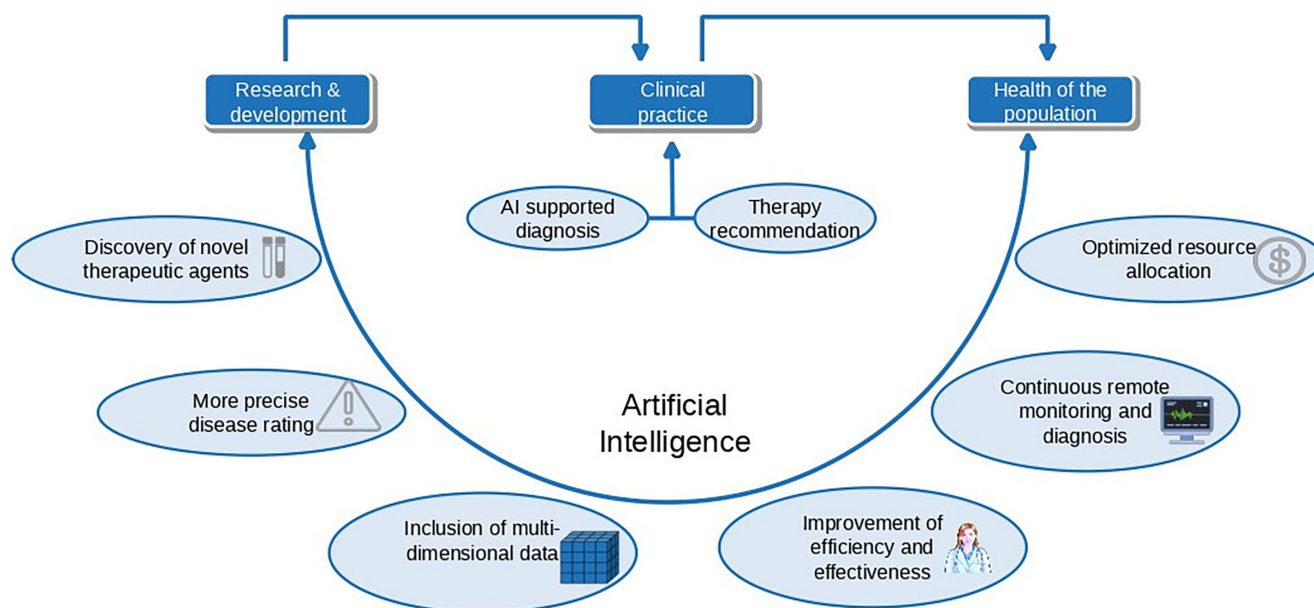


FIGURE 2 Artificial intelligence represents of the major current research fields in clinical medicine, in particular due to its opportunities in supporting diagnoses and treatment recommendations. This technical support might help to achieve a healthier global population and might have a positive impact on various application fields, such as resource optimization, efficiency, remote monitoring, and others.

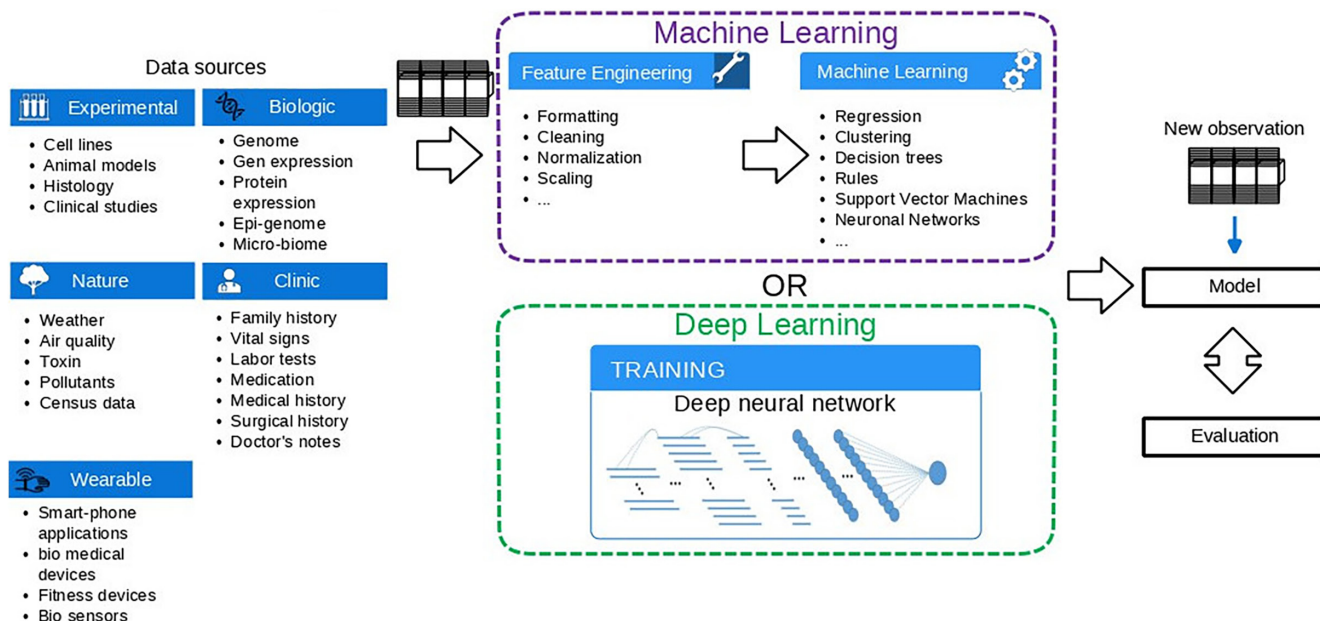


FIGURE 3 Machine learning can be deployed in various application fields. However, it requires feature engineering. Feature engineering is a complex approach that requires understanding of the data, which can be problematic in some cases. Deep learning can work with almost all kind of data and in addition do not require a costly feature engineering, making it possible to quickly explore the data without prior knowledge.

with its complexity. Moreover, data privacy and the largely individual representation of data (e.g., the small amount of training data and its lack of quality) are two key challenges. Therefore, optimal exploitation of datasets represents a major area of research. Furthermore, attempts are made to automatically generate missing data or to synthetically extend the datasets. Another goal of current AI research involves the integration of machine-readable expert knowledge, for example medical guidelines or clinical studies to push the performance. Lastly, the issue of explainability is one of the major challenges,^{56,57} as it is mandatory for any clinical application to ensure trust and responsibility.

2.2 | Methods of artificial intelligence development

In recent years, great focus has been placed on “convolutional neural networks” (CNNs), as they can be used in both, image and time series analyses (Table 1). Common issues that can be analyzed with CNNs are medical images including pathological classification of malignant tumors, skin tissues, and other diseases that are visually visible. These networks are able to score predominantly for high accuracy and fast analysis. However, especially in the case of time series, it makes sense to try out other methods in addition to CNNs. For example, “long short-term memories” (LSTMs)⁵⁸ which were designed to be applied

to time series. Another method, originally designed for text processing, is called “Transformer.”⁵⁹ They can handle long time series without losing accuracy. In addition, transformers also provide limited explainability without additional effort by using the attention that can be used to create a kind of heat map for the time series to emphasize on relevant parts.

Moreover, it is possible to process texts or tables using AI methods. For this purpose, an important feature is the process of “encoding.” Encoding or embedding of the texts is necessary because an AI has no understanding of words. To enable an AI to process words, they are mapped in a high-dimensional space. In this process, words that have context are positioned closer together than words without context. Based on encoding, the network can classify texts. Embeddings are not adapted to health terms, therefore it is necessary to train models in such a way that they produce outputs adapted to technical terms.⁶⁰ It is also possible to evaluate tables, by coding values of the table similarly to texts. Regarding medical applications, table analysis enables predictions to be made based on numerous properties which a human being can only take into account all at the same time with great effort. Models, which are used particularly with the analysis of flow texts, are Transformer and LSTMs. Both allow to capture the relationship between the individual words within a text. To give an outlook on the integration of machine learning into the clinical medical context, Table 1 illustrates how existing problems have been investigated

TABLE 1 Illustration how existing problems have been investigated using machine learning methods to give an outlook on the integration of machine learning into the clinical medical context.

References	Target/area of application	Data type	Artificial intelligence	Hybrid	Traditional approaches
(2019) ²⁶	Predicting the response to immunotherapy in patients with NSCLC	Image (2D)	CNN		
(2021) ¹⁸	Predicting the response to immunotherapy in patients with gastric cancer	Categorical, image (2D)	CNN		
(2019) ¹⁷	Identifying biomarkers to immunotherapy response from histological images of GI cancer	Categorical, image (2D)	CNN		
(2020) ²⁰	Predicting the response to immunotherapy in patients with colorectal cancer	Categorical, Image (2D)	CNN		
(2021) ³³	Predict the response to chemotherapy in patients with different tumors (5)	Categorical, numeric	CNN, VAE		SVM, XGBoost
(2019) ²²	Predicting the response to chemotherapy in patients with rectal cancer	Image (2D)		Radiomics	SVM
(2020) [43]	Predicting the response to chemotherapy in patients with rectal cancer	Image (2D)	CNN	Radiomics	
(2022) ²⁴	Predicting the response to chemotherapy in patients with breast cancer before and during therapy	Image (2D)	CNN	Radiomics	
(2022) ²⁵	Predicting the response of chemotherapy in patients with breast cancer	Image (2D)	CNN, Transformer	Radiomics	
(2021) (⁸⁵)	Predicting textbook outcome of pancreatectomy with sleep and plus data from a wearable	Categorical, numeric			SVM, KNN, Regression
(2021) ³⁶	Predicting overall survival of patients with gastric cancer	Categorical, image (2D)	CNN		
(2020) ³⁵	Predicting survival of patients with different (10) tumors	Categorical, image (2D)	CNN		
(2023) ³⁷	Predicting overall 3- and 5-year survival of patients with metastatic renal cell carcinoma	Categorical, numeric			SVM, KNN, tree-based, XGBoost, regression

using the above-mentioned methods. In addition to the aforementioned classification and prediction, these methods can also be used for other tasks, for example to mark and track certain features in video or image recordings. Finally, features like completion and generation of data can be used with tabular values and time series⁶¹ which enables to generate synthetic data that can improve the accuracy of other AI methods.

2.3 | Explainability of artificial intelligence

To justify the use of AI, it is essential to explain the predictions of these methods. This research field is covered by the so-called explainable artificial intelligence domain.⁵⁶

It attempts to find an explanation for predictions that is tangible to humans.

One of the most versatile and effective approaches is to look at the relevant data, for example, a heat map that highlights the relevant areas (Figure 5).⁶² To achieve this highlighting, countless different methods have been developed, which can be divided into three categories: gradient-based, permutation-based, or replacement-based. In gradient-based methods, the backpropagation algorithm utilized during training is used to calculate which information features of the input have led to the output. During the calculation, the network returns a value that represents the influence of each input value. However, these heat maps are not very accurate due to the gradient, resulting into precise details that are not useful for understanding. Permutation-based approaches modify the input and compute a change

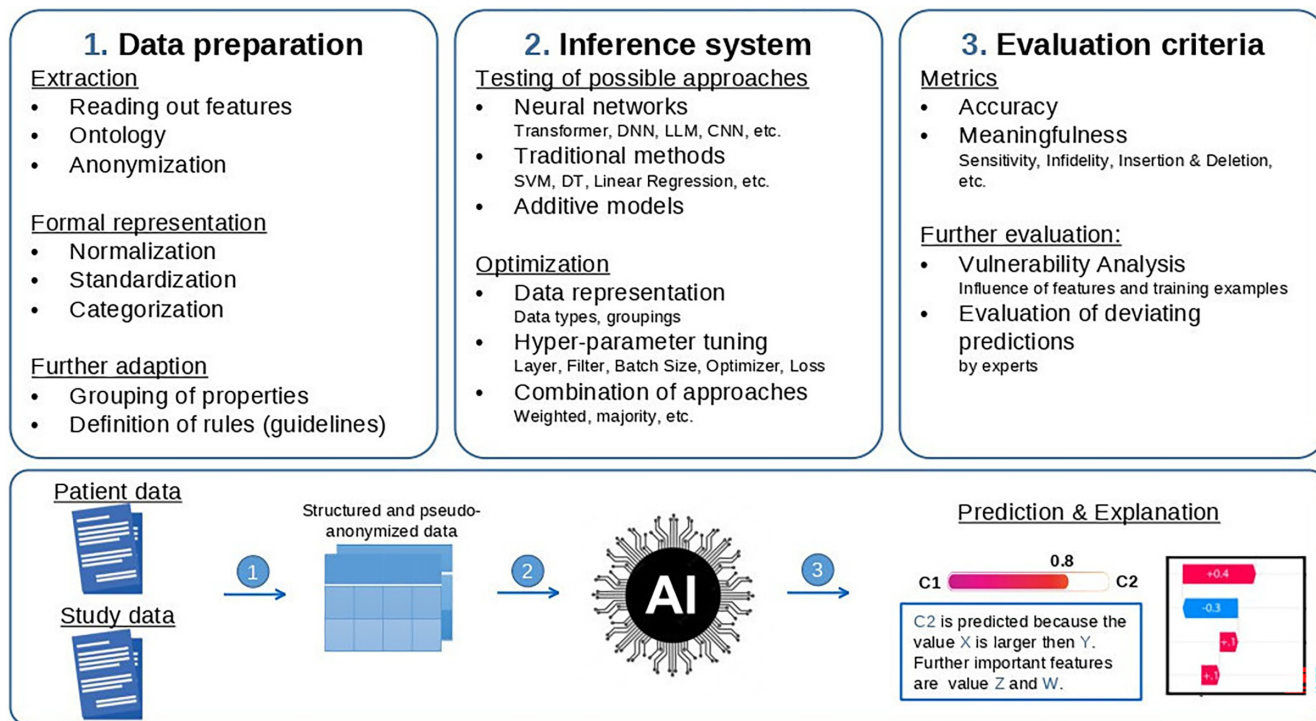


FIGURE 4 Illustration of different working steps required to develop a trustful system for medical use cases. First (1), a reliable data extraction is required to gather information from different sources and standardize it in a way that is readable for the machine learning. Second (2), an inference system needs to be developed and optimized for the use case. To achieve the best result, multiple approaches need to be tested, optimized and possible combinations need to be evaluated. Finally (3), the accuracy and explainability need to be evaluated including further analysis by experts and a vulnerability analysis to ensure a reliable and trustful system.

in the output. An assumption made here is that modifying relevant input values also results in a change of the output. This approach contains less noise and is easier to interpret. The third category attempts to have the complex AI model mimicked by a simpler explainable model.

There are other ways to explain predictions of AI models, so-called ante hoc methods. The goal is to modify the AI models in a way that their prediction can be interpreted directly. This usually leads to a decrease in accuracy. A well-known method is the so-called “attention,”⁶³ which is used by transformers. Here, a heat map of the input is already created during the prediction. In contrast to most other ante hoc explainability methods, transformers show very high performance and are less bound to an architecture. Prototypes are a widely used way to explain the prediction of an AI network. Prototypes are specific salient features/patterns. For example, a specific pattern within an image that is crucial in certain diseases. For a prediction, the AI compares the input data with the prototypes it has learned on its own and explains its decision by highlighting the recognized prototype (saliency/pattern). Another possibility for achieving explainability is to use a second network to compress the input in such a way that it shows only the features that are used by the actual network.⁶⁴

2.4 | Privacy and data security

Neural networks are learning based on a training datasets and store information about the data in this context. It is possible to reconstruct training data to obtain sensitive information. Additionally, it requires prevention to protect the models against external attacks for data extraction. It is important to understand attack methods and their impact on the models. One common attack method is the membership attack,⁶⁵ in which an attempt is made to obtain information about training data from the model by means of targeted queries. In conjunction with this, it is possible to reconstruct the dataset and then train a custom model. With such a model reconstruction,⁶⁶ an attempt is made to train a model of one's own, which copies the original, by making special queries to the private model. Many more approaches exist to steal sensitive information from AI methods. Classical anonymization can only prevent this to a limited extent, since there are also methods to reverse this.

There are different methods that can be used to protect against such attacks. One of the best-known is “homomorphic encryption,” in which the model and data are encrypted in the following way⁶⁷ that only the parties involved can use them. This makes it possible to protect a model without suffering a loss of performance.

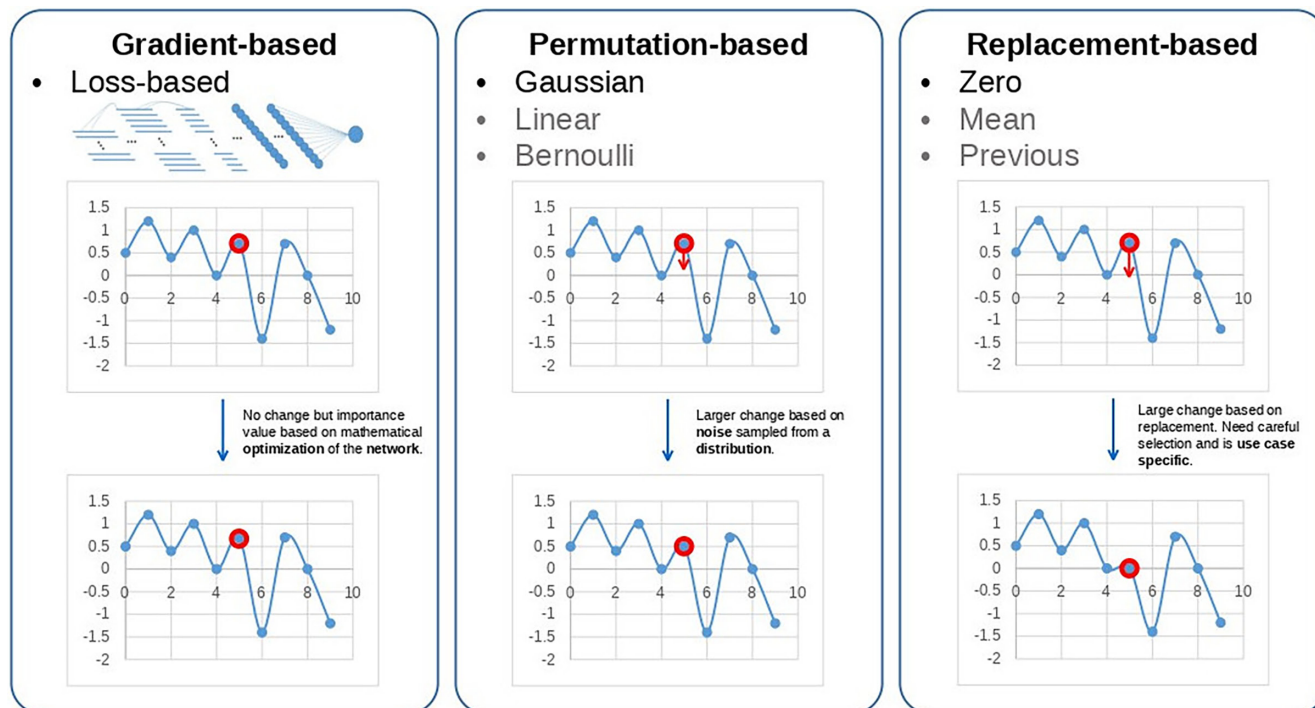


FIGURE 5 Illustration of different concepts used in attribution methods to produce a “heat map.” Gradient-based approaches depend on the loss and require a backward pass of the data to mathematically compute the importance value of each point. Permutation-based approaches only require forward passes and change the input to understand the impact of input changes. Replacement-based approaches are related to the previous mentioned approaches but are different in the way that they do not perform a small change of the point but rather completely remove the point or set it to a fixed value.

Unfortunately, it is not yet possible to apply this approach to larger networks because the additional computational cost is enormous. One option to avoid this is to encrypt the model only after training, thus protecting the model after training but not during the training phase. A widely used alternative is “Differential Privacy.”⁶⁸ Here, data and optimizers of the network are already modified during training so that the resulting model protects the sensitive data. “Noise” is added during training, and security depends on the strength of the noise. This method can be applied to almost all network architectures, but it weakens the accuracy of the models. There is also the option to have a model which learns artificially distributed.⁶⁹ Especially, the distributed processing can be used to aggregate data across multiple institutions to create datasets large enough to train neural networks, however, the sharing involves further security aspects. During the training process, the gradients used for optimization are rounded. In the second step, the average of the computed gradients of these models is used to update the model. The generated average of data information makes the model less sensitive to attacks. The method can be combined with differential privacy models. In generative approaches, synthetic data is generated. A protected generative model uses the sensitive data to generate synthetic data that cannot be distinguished from the original data by an attacker. To do

this, the generative models, usually generative adversarial networks (GANs)⁷⁰ or variational autoencoders (VAEs) are trained using differential privacy. The generated data can be used by a model without any protection mechanism.

2.5 | Usage of medical data: Representation, types, quality, and quantity

In addition to the distinction between imaging, text, video, and tabular data, a more precise division is necessary, especially for tabular data. Data are distinguished in numeric and categorical data. Numerical data can be used directly by AI methods for the most part, while categorical data must be converted into a machine-readable format. Ordinal data include a relationship between the values, while nominal data has no relation. However, both must be transformed into categories. One approach that transforms categories into numerical values without a cardinal relationship is “one-hot encoding.” For textual data, it is further required to design a context-specific vocabulary (Embedding) that maps the words to a numerical representation.

The amount of data is one of the biggest problems regarding any medical projects of AI development.

Specifically, AI models are known to handle a larger amount of data than traditional models such as support vector machines,⁷¹ linear regression,⁷² or decision trees. Furthermore, there is no strict rule of how many samples are required to successfully learn a task, leading to the problem that collecting data results in the risk of having not enough samples or samples with too much variety to establish a working algorithm.

Finally, one of the largest challenges is the preparation of the data, as the representation used by physicians and AI differs largely. This means that in most cases, many efforts are required to first extract and then transform the data into a representation that is suitable for an AI system. Depending on the chosen model, this extraction involves a feature as shown in Figure 3. The data preparation further includes the adjustment of data quality. As the medical process is continuous and improves over time, the data representation changes leading to different representation of similar features, empty data fields, different terms, and some mistakes in the data. All those aspects need to be addressed during the extraction and preparation. Therefore, one need to implement rules, incorporate logical rules, and ontologies to merge terms.

2.6 | Continuous lifelong learning: Adaption of approaches

One of the biggest challenges is to keep the approaches updated. In the medical domain, there are always new medications or other treatment options that need to be considered by an automated system. If this is not the case, the system works with a snapshot of patient data and does neither improve nor keep the quality at a desired level. One naive approach is to retrain the system from scratch, but this can result in huge resource consumption and requires time. Another approach is to fine-tune the model using some of the old data and new data. Fine-tuning is much faster and requires less resources, furthermore, it benefits from the old system that already learned the concepts of the task. However, the term fine-tuning is mainly used when a network is adapted to a new dataset. Modifying the task without forgetting old relevant information is called continuous lifelong learning. There are several works that review this kind of approaches.^{73,74} Furthermore, scientific attempts showed that it is possible to adapt a network to new data⁷⁵ making it possible to create AI approaches that learn continuously to always produce state-of-the-art results. However, it has to be mentioned that even in such a scenario, it is important to update the AI approaches using novel technology such as better architecture and apply the life-long learning on these approaches.

3 | FUTURE PERSPECTIVES IN TECHNICAL APPLICATIONS OF ARTIFICIAL INTELLIGENCE IN CLINICAL ONCOLOGY

3.1 | Perspectives in usage of artificial intelligence applications in medical treatment recommendation

Data should ideally always be up-to-date for the purpose of “AI training.”⁷⁶ Qualitative and quantitative, high-quality data are an important factor in increasing the evidence of AI. To further increase evidence, it is necessary for an AI to be validated by external validation. The example of WFO showed that this is not granted at all. One of the central problems of WFO was the fact that it had not been trained with representative data.⁴⁵ Therefore, it is of great importance that training data of AI are representative for all patient groups.⁷⁶ The highest possible level of evidence provides patients a high level of safety in treatment. In this context, it is important to know who is legally responsible if an AI in case of potential user mistakes with consequences for the health of patients. Thus, it is mandatory that physicians are well-trained in the use of AI applications.⁷⁷

3.2 | Opportunities and limitations

A major goal towards the best possible treatment in oncology is personalized medicine.⁷⁶ For this purpose, AI applications allow a maximum high number of data and images of a single patient to be analyzed automatically in a very short time in order to determine the best treatment available^{11,78} (Figure 4). This capability is particularly helpful with complex oncological diseases, which to this date are evaluated and treated by MCCs (Figure 1). Here, AI could help physicians by focusing on the key information from large databases of patients' characteristics and from trial information of available treatments, which would be analyzed by AI applications.⁷⁹ AI could incorporate all existing, relevant medical publications of available treatment regimens into their decision without their own interpretation, which are not (yet) known to a treating physician.^{44,47,80} Furthermore, human errors or physician-dependent individual variation could be decreased.⁸¹ By processing data quickly and automatically, an AI would be able to reduce time in various settings during the treatment process in oncology. AI applications could offer a cost-reducing effect by reducing working time and replace expensive tests and examinations.^{17,19,20}

However, we need to be aware of AI's limitations and current problems. A major problem is the “black box”

which describes the problem of the missing explainability of decisions by AI.⁴⁶ The ability of an AI to justify its decision is almost mandatory for a clinical implementation. There are already a few, promising AI applications in predicting treatment response that can justify their decisions and are referred to as “white boxes.”^{82–84} Finally, we need to consider the immense financial investment and associated financial risk that would be necessary to develop and implement AI in global healthcare systems.⁷⁷

4 | CONCLUSION

Applications of AI already offer many potential advantages and will be able to offer even more benefits in the future. These applications are only as good as the data is on which they are trained. Thus, it is essential to use data of highest quality and quantity. Both conditions can be significantly achieved through maximum digitization in our healthcare system. Data which will be used to train any AI should be generated using standards that are as uniform as possible.⁷⁸ In future, such a standardization should take place on a global level across different healthcare system in order to improve AI developments and applications globally.⁷⁶ Finally, it will be mandatory to explain any AI-generated recommendation in order to achieve a successful implementation of any AI applications in clinical practice. Nevertheless, the final decision should always be made by the physicians in a joint decision-making process with the patients.

AUTHOR CONTRIBUTIONS

Gregor Duwe: Conceptualization (equal); funding acquisition (equal); supervision (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Dominique Mercier:** Conceptualization (equal); resources (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Crispin Wiesmann:** Resources (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Verena Kauth:** Writing – original draft (equal); writing – review and editing (equal). **Kerstin Moench:** Writing – original draft (equal); writing – review and editing (equal). **Markus Junker:** Conceptualization (equal); funding acquisition (equal); validation (equal); writing – review and editing (equal). **Christopher C. M. Neumann:** Supervision (equal); validation (equal); writing – review and editing (equal). **Axel Haferkamp:** Conceptualization (equal); supervision (equal); validation (equal); writing – review and editing (equal). **Andreas Dengel:** Conceptualization (equal); funding acquisition (equal); supervision (equal); validation (equal); writing – review and editing (equal). **Thomas Höfner:**

Conceptualization (equal); funding acquisition (equal); supervision (equal); validation (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This work was supported by the German Federal Ministry of Education and Research, grant number: 16SV9053.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

ETHICS STATEMENT

No ethics approval is required as data is freely available online in the public domain.

ORCID

Gregor Duwe  <https://orcid.org/0000-0001-7008-1238>

REFERENCES

1. Darling G. The impact of clinical practice guidelines and clinical trials on treatment decisions. *Surg Oncol.* 2002;11(4):255-262.
2. Williams CJ. Evidence-based cancer care. *Clin Oncol.* 1998;10(3):144-149.
3. Pluchino LA, D'Amico TA. National Comprehensive Cancer Network Guidelines: who makes them? What are they? Why are they important? *Ann Thorac Surg.* 2020;110(6):1789-1795.
4. Vinod SK. Decision making in lung cancer—how applicable are the guidelines? *Clin Oncol (R Coll Radiol).* 2015;27(2):125-131.
5. Wright FC, de Vito C, Langer B, Hunter A, Expert Panel on Multidisciplinary Cancer Conference Standards. Multidisciplinary cancer conferences: a systematic review and development of practice standards. *Eur J Cancer.* 2007;43(6):1002-1010.
6. Schreyer AG, Dendl LM, Antoch G, Layer G, Beyer L, Schleder S. Interdisciplinary tumor boards in the radiological routine: current situation based on an online survey in Germany. *Radiologe.* 2020;60(8):737-746.
7. Bundesausschuss G. Tragende Gründe zum Beschlussentwurf des Gemeinsamen Bundesausschusses über Regelungen zur Konkretisierung der besonderen Aufgaben von Zentren und Schwerpunkten gemäß § 136c Absatz 5 SGB V (Zentrumsregelungen)—Erstfassung. 2020.
8. El Saghier NS, Keating NL, Carlson RW, Khoury KE, Fallowfield L. Tumor boards: optimizing the structure and improving efficiency of multidisciplinary management of patients with cancer worldwide. *Am Soc Clin Oncol Educ Book.* 2014;34:e461-6.
9. Klarenbeek SE, Weekenstroom HHA, Sedelaar JPM, Fütterer JJ, Prokop M, Tummers M. The effect of higher level computerized

- clinical decision support systems on oncology care: a systematic review. *Cancer*. 2020;12(4):1032.
10. Walsh S, de Jong EEC, van Timmeren J, et al. Decision support systems in oncology. *JCO Clin Cancer Inform*. 2019;3:1-9.
 11. Farina E, Nabhen JJ, Dacoregio MI, Batalini F, Moraes FY. An overview of artificial intelligence in oncology. *Future Sci OA*. 2022;8(4):Fso787.
 12. Benary M, Wang XD, Schmidt M, et al. Leveraging large language models for decision support in personalized oncology. *JAMA Netw Open*. 2023;6(11):e2343689.
 13. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol*. 2022;19(2):132-146.
 14. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br J Cancer*. 2021;124(4):686-696.
 15. Yousif M, van Diest PJ, Laurinavicius A, et al. Artificial intelligence applied to breast pathology. *Virchows Arch*. 2022;480(1):191-209.
 16. Madabhushi ASH, Wang X, Barrera C, Velcheti V. *Predicting Response to Immunotherapy Using Computer Extracted Features of Cancer Nuclei from Hematoxylin and Eosin (H&E) Stained Images of Non-small Cell Lung Cancer (NSCLC)*. O. Case Western Reserve University; 2019.
 17. Kather JN, Pearson AT, Halama N, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med*. 2019;25(7):1054-1056.
 18. Muti HS, Heij LR, Keller G, et al. Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: a retrospective multicentre cohort study. *Lancet Digi Health*. 2021;3(10):e654-e664.
 19. Echle A, Laleh NG, Schrammen PL, et al. Deep learning for the detection of microsatellite instability from histology images in colorectal cancer: a systematic literature review. *ImmunoInformatics*. 2021;3-4:100008.
 20. Echle A, Grabsch HI, Quirke P, et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. *Gastroenterology*. 2020;159(4):1406-1416.e11.
 21. Avanzo M, Stancanella J, El Naqa I. Beyond imaging: the promise of radiomics. *Phys Med*. 2017;38:122-139.
 22. Yi X, Pei Q, Zhang Y, et al. MRI-based Radiomics predicts tumor response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Front Oncol*. 2019;9:552.
 23. Fu J, Zhong X, Li N, et al. Deep learning-based radiomic features for improving neoadjuvant chemoradiation response prediction in locally advanced rectal cancer. *Phys Med Biol*. 2020;65(7):075001.
 24. Gu J, Tong T, He C, et al. Deep learning radiomics of ultrasonography can predict response to neoadjuvant chemotherapy in breast cancer at an early stage of treatment: a prospective study. *Eur Radiol*. 2022;32(3):2099-2109.
 25. Skarping I, Larsson M, Förnvik D. Analysis of mammograms using artificial intelligence to predict response to neoadjuvant chemotherapy in breast cancer patients: proof of concept. *Eur Radiol*. 2022;32(5):3131-3141.
 26. Wang PP, Deng CL, Wu B. Magnetic resonance imaging-based artificial intelligence model in rectal cancer. *World J Gastroenterol*. 2021;27(18):2122-2130.
 27. Wesdorp NJ, Hellingman T, Jansma EP, et al. Advanced analytics and artificial intelligence in gastrointestinal cancer: a systematic review of radiomics predicting response to treatment. *Eur J Nucl Med Mol Imaging*. 2021;48(6):1785-1794.
 28. Trebeschi S, Drago SG, Birkbak NJ, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol*. 2019;30(6):998-1004.
 29. Bibault JE, Giraud P, Housset M, et al. Deep learning and radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. *Sci Rep*. 2018;8(1):12611.
 30. Vaidya P, Bera K, Gupta A, et al. CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in stage I, II resectable non-small cell lung cancer: a retrospective multicohort study for outcome prediction. *Lancet Digit Health*. 2020;2(3):e116-e128.
 31. Granzier RWY, van Nijnatten TJA, Woodruff HC, Smidt ML, Lobbes MBI. Exploring breast cancer response prediction to neoadjuvant systemic therapy using MRI-based radiomics: a systematic review. *Eur J Radiol*. 2019;121:108736.
 32. Way GP, Greene CS. Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders. *Pac Symp Biocomput*. 2018;23:80-91.
 33. Wei Q, Ramsey SA. Predicting chemotherapy response using a variational autoencoder approach. *BMC Bioinformatics*. 2021;22(1):453.
 34. Parikh RB, Hasler JS, Zhang Y, et al. Development of machine learning algorithms incorporating electronic health record data, patient-reported outcomes, or both to predict mortality for outpatients with cancer. *JCO Clin Cancer Inform*. 2022;6:e2200073.
 35. Wulczyn E, Steiner DF, Xu Z, et al. Deep learning-based survival prediction for multiple cancer types using histopathology images. *PLoS One*. 2020;15(6):e0233678.
 36. Huang B, Tian S, Zhan N, et al. Accurate diagnosis and prognosis prediction of gastric cancer using deep learning on digital pathological images: a retrospective multicentre study. *EBioMedicine*. 2021;73:103631.
 37. Barkan E, Porta C, Rabinovici-Cohen S, Tibollo V, Quaglini S, Rizzo M. Artificial intelligence-based prediction of overall survival in metastatic renal cell carcinoma. *Front Oncol*. 2023;13:1021684.
 38. Jang W, Jeong C, Kwon KA, et al. Artificial intelligence for predicting five-year survival in stage IV metastatic breast cancer patients: a focus on sarcopenia and other host factors. *Front Physiol*. 2022;13:977189.
 39. Séroussi B, Bouaud J, Antoine É-C. OncoDoc: a successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer. *Artif Intell Med*. 2001;22(1):43-64.
 40. Bouaud J, Spano JP, Lefranc JP, et al. Physicians' attitudes towards the advice of a guideline-based decision support system: a case study with OncoDoc2 in the management of breast cancer patients. *Stud Health Technol Inform*. 2015;216:264-269.
 41. Larburu N, Muro N, Arrue M, Álvarez R, et al. DESIREE—a web-based software ecosystem for the personalized, collaborative and multidisciplinary management of primary breast cancer. in 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom) 2018.
 42. Larburu N, Muro N, Macía I, Sánchez E, Wang H, et al. Augmenting guideline-based CDSS with experts. *Knowledge*. 2017;5:370-376.

43. Pelayo S, Bouaud J, Blancafort C, Lamy J-B, et al. Preliminary qualitative and quantitative evaluation of DESIREE, a decision support platform for the management of primary breast cancer patients. *AMIA Annu Symp Proc.* 2020;1012-1021.
44. Schmidt C. M. D. Anderson breaks with IBM Watson, raising questions about artificial intelligence in oncology. *J Natl Cancer Inst.* 2017;109(5).
45. Harish V, Morgado F, Stern AD, das S. Artificial intelligence and clinical decision making: the new nature of medical uncertainty. *Acad Med.* 2021;96(1):31-36.
46. Keikes L, Medlock S, van de Berg D, et al. The first steps in the evaluation of a “black-box” decision support tool: a protocol and feasibility study for the evaluation of Watson for oncology. *J Clin Transl Res.* 2018;3:411-423.
47. Jie Z, Zhiying Z, Li L. A meta-analysis of Watson for oncology in clinical application. *Sci Rep.* 2021;11(1):5792.
48. Suwanvecho S, Suwanrusme H, Jirakulaporn T, et al. Comparison of an oncology clinical decision-support system's recommendations with actual treatment decisions. *J Am Med Inform Assoc.* 2021;28(4):832-838.
49. Li T, Chen C, Zhang SS, et al. Deployment and integration of a cognitive technology in China: experiences and lessons learned. *J Clin Oncol.* 2019;37(15_suppl):6538.
50. Sarre-Lazcano C, Armengol Alonso A, Huitzil Melendez FD, et al. Cognitive computing in oncology: a qualitative assessment of IBM Watson for oncology in Mexico. *J Clin Oncol.* 2017;35(15_suppl):e18166.
51. Duwe G, Junker M, Mercier D, et al. A0518—KITTU: artificial intelligence supported clinical decisions in uro-oncology—an interdisciplinary study. *Eur Urol.* 2023;83:S741.
52. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell.* 2013;35:1798-1828.
53. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115-118.
54. Kollias D, Tagaris A, Stafylopatis A, Kollias S, Tagaris G. Deep neural architectures for prediction in healthcare. *Complex Intell Syst.* 2018;4(2):119-131.
55. Pandey SK, Janghel RR. Recent deep learning techniques, challenges and its applications for medical healthcare system: a review. *Neural Process Lett.* 2019;50(2):1907-1935.
56. Saraswat D, Bhattacharya P, Verma A, et al. Explainable AI for healthcare 5.0: opportunities and challenges. *IEEE Access.* 2022;10:84486-84517.
57. Adadi A, Berrada M. Explainable AI for healthcare: from black box to interpretable models. *Embedded Systems and Artificial Intelligence: Proceedings of ESAI.* Springer; 2020.
58. Graves A. Long short-term memory. *Supervised Sequence Labelling with Recurrent Neural Networks.* Berlin; 2012:37-45.
59. Zerveas G, Jayaraman S, Patel D, Bhamidipaty A, et al. A transformer-based framework for multivariate time series representation learning. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining 2021.
60. Johnson JM, Khoshgoftaar TM. Medical provider embeddings for healthcare fraud detection. *SN Comput Sci.* 2021;2(4):276.
61. Hernandez M, Epelde G, Alberdi A, Cilla R, Rankin D. Synthetic data generation for tabular health records: a systematic review. *Neurocomputing.* 2022;493:28-45.
62. Jin D, Sergeeva E, Weng WH, Chauhan G, Szolovits P. Explainable deep learning in healthcare: a methodological survey from an attribution view. *WIREs Mech.* 2022;14(3):e1548.
63. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. *Neurocomputing.* 2021;452:48-62.
64. Palacio S, Folz J, Hees J, et al. What do deep networks like to see? In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3108–3117.
65. Hu H, Salci Z, Sun L, Dobbie G, Yu PS, Zhang X. Membership inference attacks on machine learning: a survey. *ACM Comput Surv.* 2022;54:1-37.
66. Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security.* Association for Computing Machinery; 2015:1322-1333.
67. Aono Y, Hayashi T, Wang L, Moriai S. Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Trans Inf Forensics Secur.* 2017;13(5):1333-1345.
68. Abadi M et al. Deep learning with differential privacy. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security.* Association for Computing Machinery; 2016:308-318.
69. Konečný J, McMahan HB, Yu FX, et al. Federated learning: strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492 2016.
70. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: an overview. *IEEE Signal Process Mag.* 2018;35(1):53-65.
71. Noble WS. What is a support vector machine? *Nat Biotechnol.* 2006;24(12):1565-1567.
72. Su X, Yan X, Tsai CL. Linear regression. *Wiley Interdiscip Rev Comput Stat.* 2012;4(3):275-294.
73. De Lange M, Aljundi R, Masana M, et al. A continual learning survey: defying forgetting in classification tasks. *IEEE Trans Pattern Anal Mach Intell.* 2022;44(7):3366-3385.
74. Parisi GI, Kemker R, Part JL, Kanan C, Wermter S. Continual lifelong learning with neural networks: a review. *Neural Netw.* 2019;113:54-71.
75. Ling CX, Bohn TA. A deep learning framework for lifelong machine learning. 2021. ArXiv,abs/2105.00157.
76. Shreve JT, Khanani SA, Haddad TC. Artificial intelligence in oncology: current capabilities, future opportunities, and ethical considerations. *Am Soc Clin Oncol Educ Book.* 2022;42:1-10.
77. Tătaru OS, Vartolomei MD, Rassweiler JJ, et al. Artificial intelligence and machine learning in prostate cancer patient management-current trends and future perspectives. *Diagnostics.* 2021;11(2):354.
78. Sebastian AM, Peter D. Artificial intelligence in cancer research: trends, challenges and future directions. *Lifestyles.* 2022;12(12):1991.
79. Stenzl A, Sternberg CN, Ghith J, Serfass L, Schijvenaars BJA, Sboner A. Application of artificial intelligence to overcome clinical information overload in urological cancer. *BJU Int.* 2022;130(3):291-300.
80. Price W, Nicholson I. Medical AI and contextual bias. *Harv JL & Tech.* 2019;33:65.
81. Huang EP, O'Connor JPB, McShane LM, et al. Criteria for the translation of radiomics into clinically useful tests. *Nat Rev Clin Oncol.* 2023;20(2):69-82.

82. Prelaj A, Galli EG, Miskovic V, et al. Real-world data to build explainable trustworthy artificial intelligence models for prediction of immunotherapy efficacy in NSCLC patients. *Front Oncol.* 2022;12:1078822.
83. Shimizu H, Nakayama KI. Artificial intelligence in oncology. *Cancer Sci.* 2020;111(5):1452-1460.
84. Foersch S, Glasner C, Woerl AC, et al. Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. *Nat Med.* 2023;29(2):430-439.
85. Cos H, Li D, Williams G, et al. Predicting Outcomes in Patients Undergoing Pancreatectomy Using Wearable Technology and Machine Learning: Prospective Cohort Study. *J Med Internet Res.* 2021;23:e23595.

How to cite this article: Duwe G, Mercier D, Wiesmann C, et al. Challenges and perspectives in use of artificial intelligence to support treatment recommendations in clinical oncology. *Cancer Med.* 2024;13:e7398. doi:[10.1002/cam4.7398](https://doi.org/10.1002/cam4.7398)