Artificial Intelligence In Medicine 145 (2023) 102676

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Medical informed machine learning: A scoping review and future research directions

Florian Leiser, Sascha Rank, Manuel Schmidt-Kraepelin, Scott Thiebes, Ali Sunyaev *

Department of Economics and Management, Karlsruhe Institute of Technology, Karlsruhe, Germany

ARTICLE INFO

Keywords:
Informed machine learning
Scoping literature review
Medical informatics
Domain knowledge
Machine learning

ABSTRACT

Combining domain knowledge (DK) and machine learning is a recent research stream to overcome multiple issues like limited explainability, lack of data, and insufficient robustness. Most approaches applying informed machine learning (IML), however, are customized to solve one specific problem. This study analyzes the status of IML in medicine by conducting a scoping literature review based on an existing taxonomy. We identified 177 papers and analyzed them regarding the used DK, the implemented machine learning model, and the motives for performing IML. We find an immense role of expert knowledge and image data in medical IML. We then provide an overview and analysis of recent approaches and supply five directions for future research. This review can help develop future medical IML approaches by easily referencing existing solutions and shaping future research directions.

1. Introduction

Despite the broad use of machine learning (ML) across applications, there are still limitations and challenges in the adoption of ML in medical practice [1]. Sufficient data availability and quality are essential for accurate predictions of ML models. In the medical domain, data gathering is difficult and expensive since diseases are constantly evolving [2], digitalization in medical institutions lags behind [3], and legal requirements need to be met for sensitive medical data [4].

Data availability, however, is not the only factor that hinders the adoption of ML in medicine. Whenever the decisions of ML models are in doubt, we refer to the decisions made by domain experts. Expert-based decisions can be more elaborate and take care of the specifics of each patient. Therefore, emerging research on explainable and interpretable ML provides explanations of the decisions made by ML approaches [5]. These improvements in explainability and specificity are often achieved by integrating experts' domain knowledge (DK) into ML models. This is called *informed ML (IML)* which "describes learning from a hybrid information source that consists of data and prior knowledge" [6]. Such hybrid learning can provide the benefits of both, expert-based and data-driven decisions [6].

Combining expert-based and data-driven decisions, however, results in customized solutions regarding the ML models used, the DK included in the models, or the step of applying the DK to achieve the benefits of overcoming a specific problem. Every customization solves a specific problem, but we currently lack a general understanding of frequent combinations, which could benefit practitioners to apply existing solutions to their related problems.

Previous works already investigated the current status of ML in medicine and argued that an improved inclusion of computers in clinical practice "may allow radiologists to further integrate their knowledge" [7,8]. Based on these results, first approaches were conducted summarizing the inclusion of DK into medical imaging algorithms [9]. However, in the study at hand, we do not limit ourselves to imaging applications but investigate the inclusion of DK in a broader sense. We refer to an existing taxonomy to assess and categorize the inclusion of DK by analyzing the three dimensions source of DK, structure of DK, and application step [6].

In addition to the inclusion-related dimensions, we also gather an overview of the motives for performing IML. All approaches conduct IML for a multitude of reasons, which range from improving the explainability of the approaches by including DK [10], over reducing the effort required by experts [11], to limited data availability [12]. These individual motives are all well on their own, but we require an overview of different motives to perform medical IML (MedIML). Such an overview could highlight previously undetected similarities between approaches and provide general conclusions on the motives for performing MedIML. We, therefore, pose the following research questions (RQs):

* Corresponding author at: Kaiserstr. 89, 76133 Karlsruhe, Germany.
E-mail address: sunyaev@kit.edu (A. Sunyaev).

https://doi.org/10.1016/j.artmed.2023.102676
Received 16 December 2022; Received in revised form 15 June 2023; Accepted 2 October 2023
Available online 19 October 2023

RQ1: What domain knowledge is integrated into medical machine learning approaches?
RQ2: How is domain knowledge integrated into medical machine learning approaches?
RQ3: What are the motives and potential improvements for MedIML*?*

To answer these research questions, we conduct a literature-based scoping study by following the method of Arksey and O'Malley [13]. Our search yielded 177 articles that were analyzed regarding the used ML models, the included DK, and the motives for performing IML. Since every paper also identifies potential drawbacks of their approach, we also analyzed the potential improvements specified by the authors. In the analysis, we identified clusters in the combination of the structure of the included DK and the ML models used in the approaches. We also found a broad range of motives and potential improvements in all directions, from which we derived five future research directions.

Building on previous works that have investigated IML approaches across domains [6], we synthesize the current state of knowledge on how DK is included in medical ML settings and thereby contribute to the emerging stream of literature on artificial intelligence in medicine [14]. The results of our study reveal a heterogeneous landscape of different ML approaches incorporating medical DK. We provide five directions for future research that researchers may use as starting points guiding their inquiries on MedIML. Our study reveals several specificities of MedIML in comparison to other domains of IML (e.g., the dominant role of expert knowledge over other types of DK). This strengthens our notion that collaboration and exchange between different context-specific IML research approaches will be key to unfold the full potential of IML to increase explainability and predictive performance in the future [5].

In the following section, we present the methodology conducted in this article. We follow with the results of our scoping study before we provide potential future research directions and limitations of our study in the discussion. We end the article with a brief conclusion.

2. Methods

For this scoping review, we followed a five-step approach based on the framework by Arksey and O'Malley [13] shown in Fig. 1. In the first step, we defined the research questions described above.

2.1. Identification of relevant studies

In the second step, we continued with the identification of relevant studies. To cover a wide range of publications, we searched the databases ACM Digital Library, AIS eLibrary, IEEE Xplore, ProQuest, PubMed, and Scopus with the search string: *TI-KE-AB*[(*"machine learning" OR "artificial intelligence" OR "deep learning"*) *AND* (*health* OR medic**) *AND* (*"domain knowledge" OR expert-based OR theory-guided OR theory-driven OR physics-informed OR physics-guided*)].

The search was performed on May 5, 2022 and yielded 2,243 hits (see Fig. 2). After removing pre-print articles published earlier than 2019 and duplicates, 1,521 articles remained. We decided to include pre-print articles from 2019 or later, since these articles can provide additional insights as the research stream of IML is fast-moving and relatively novel.

2.2. Study selection

In the third step, we selected a set of relevant studies by screening

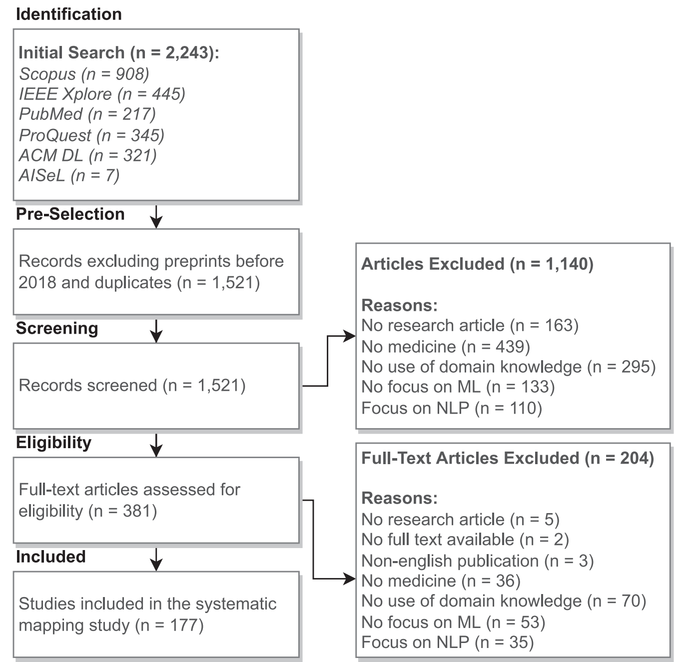

Fig. 2. PRISMA-Chart with exclusion criteria of the literature search.

their abstracts and, for articles deemed relevant based on the abstract, their full texts. We excluded 168 studies that turned out not to be peer-reviewed research articles (e.g., editorials or doctoral theses). In addition, we excluded articles that did not focus on medicine (n = 475), did not focus on ML (n = 186), or did not use DK (n = 365). For two articles, there were no full texts available to us, whereas three articles were not available in English. We also decided to exclude articles related to natural language processing (NLP, n = 145), since many approaches focused on medical reports and we want to focus on diagnostic approaches in this review. The remaining 177 articles were included in our review.

2.3. Charting the data

In the fourth step, we classified the set of relevant articles. The coding process was conducted by researchers with a predominant background in information systems and computer science. Researchers in other fields like medicine may have different assessments about parts of our classification scheme. For every article, we recorded the type of study based on a classification by Wieringa et al. [15] (i.e., evaluation research, solution proposal, validation research, philosophical paper, opinion paper, and personal experience papers). This classification scheme provides insights about the novelty of the publications and especially whether they validated existing approaches or constructed novel ones. We additionally assessed the structure of the training data as a data modality category. Regarding the ML approach, we recorded the ML task and the used ML model. For the facets related to the inclusion of DK, we followed the taxonomy by von Rueden et al. [6]. Thus, we analyzed the source, structure, and step of application of DK in each study. Regarding the source of DK, we differentiated between scientific knowledge, world knowledge, and expert knowledge.

Scientific knowledge includes knowledge that can be formalized and

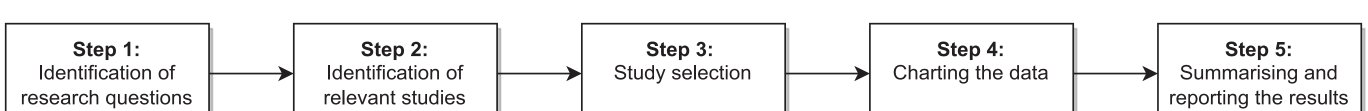

Fig. 1. Overview of the conducted study.

validated with scientific experiments like equations to represent the blood pressure in vessels [16]. World knowledge, or general knowledge, can be more or less formal and refers to facts known to almost everyone [17]. Expert knowledge is only held by a particular group of experts and is typically more informal [6]. As medical ML tasks include varying levels of expert knowledge held by different groups of experts, we extended the taxonomy by von Rueden et al. and further differentiated between common medical knowledge and specialized medical knowledge. Common medical knowledge refers to more general knowledge every doctor is likely to have (e.g., ICD-11 codes [18]). In contrast, specialized medical knowledge refers to knowledge from a medical specialty (e.g., difficulty of frames in detecting elbow fractures [19]).

Regarding the structure of DK, we also followed the categorization of von Rueden et al. and distinguished between the groups of algebraic equations, differential equations, simulation results, spatial invariances, logic rules, knowledge graphs, probabilistic relations, and human feedback. For the integration of DK into the machine learning model, we differentiated between training data, hypothesis set, learning algorithm, and final hypothesis. For a more detailed description of these categories, refer to [6].

To ensure a common understanding of the concepts, the first 20 articles were coded independently by two researchers and discussed subsequently. The remaining articles were split between the researchers of this article who then coded their split of articles, with regular discussions among the researchers throughout the coding process to ensure consistency.

To gain an understanding of the motivation for applying MedIML, we analyzed the motives stated by the authors in their manuscripts for including DK in the ML task. In addition, to discover the further potential of IML approaches, we coded the reported potential improvements. Afterward, two researchers grouped the identified motives and potential improvements individually. Cases of conflicting assignments were discussed and resolved among the researchers. The classification scheme with some exemplary manifestations is presented in Table 1. A full list of all identified papers and their manifestations can be found in the supplement.

Table 1
Coding scheme facets with explanations and exemplary manifestations.

Facet	Explanation	Manifestations
Type of study	The type of research conducted in the study	e.g., conceptual [20], quantitative [21], mixed methods [22]
ML model	The used ML algorithm	e.g., CNN [23], decision tree [24], SVM [25]
Data modality	The structure of the training data used in the ML model	e.g., imaging [26], tabular [27], time series [28]
ML task	The type of problem solved by the ML model	e.g., classification [29], regression [30], image segmentation [31]
Source of DK	The origin of the included DK	e.g., scientific knowledge [32], common medical knowledge [33]
Structure of DK	The representation of the included DK	e.g., algebraic equations [34], human feedback [35], spatial invariances [36]
Step of application	The place where DK is included in the ML pipeline	e.g., hypothesis set [37], learning algorithm [38], training data [39]
Motives for including DK	Reasons to include DK in a medical ML task stated by the authors	e.g., limited data [40], medical uncertainty [41], unexploited knowledge [42]
Potential improvements for including DK	Further ways of improving the inclusion of DK in medical ML tasks discovered by the authors	e.g., improve predictive performance [43], reduce manual effort [44], validate with practitioners [45]

3. Results

3.1. Characteristics of the included studies

The identified 177 studies were published between 1993 and 2022. Prior to 2006, 4 studies were published as shown in Fig. 3. For the following ten years, the inclusion of DK into medical ML algorithms remained inconspicuous, which is why until 2015, only 29 studies within our sample were published. Our findings reveal that MedIML is a nascent but quickly growing stream of research, as evidenced by the fact that after 2015, the number of papers published per year approximately doubled every year starting from 10 studies in 2018 to 58 manuscripts in 2021. In the first third of 2022 alone, we identified 12 studies combining DK and ML algorithms.

Fig. 4a shows that within the identified 177 papers, we find the majority of the studies developing and presenting novel approaches (n = 97). Two manuscripts conduct mixed methods meaning a combination of quantitative or qualitative data analysis, conceptual works, or literature reviews. One example conducts extensive interviews to gather expert data and conceptually develop a new algorithm [46]. The remaining manuscripts (n = 78) conduct a quantitative evaluation of their developed approaches, for example by comparing the developed extensions to baselines.

The most frequent outlets of the manuscripts are the pre-print platform arXiv (n = 15) and medical imaging journals and conferences, like IEEE Transactions on Medical Imaging (n = 8) or the International Symposium on Biomedical Imaging (n = 7) as seen in Fig. 4b. Other outlets, including Artificial Intelligence in Medicine, the Conference on Artificial Intelligence in Medicine in Europe, and the IEEE Journal of Biomedical and Health Informatics, each published 4 studies.

We see from Fig. 5a that a majority of publications address classification tasks (n = 114). Regression tasks (n = 24) and image segmentation tasks (n = 21) each make up about a fourth of the selected publications. Other ML problems include reinforcement learning (n = 2), association rule mining (n = 3), or synthetic data generation (n = 3). Regarding different data modalities, we find that about half of the publications work with image data (n = 85) as can be seen in Fig. 5b. Most of the remaining papers focus on tabular data (n = 50) and time series data (n = 39). Other data modalities include multimodal approaches and one approach inferring on knowledge graphs [47].

With publications mostly working with image data and classification tasks, it is no surprise that convolutional neural networks (CNNs) are the most frequently used ML models (n = 70) as seen in Fig. 5c. With the prevalence of manuscripts handling image data, our findings coincide with existing research indicating CNNs work best on image classification tasks [48,49]. Deep neural networks (DNNs, n = 16) and recurrent neural networks (RNNs, n = 11) contribute to the immense use of neural networks in general (n = 102). Bayesian networks are used (n = 17) as well as general adversarial networks (n = 10) and simple statistical approaches (n = 9).

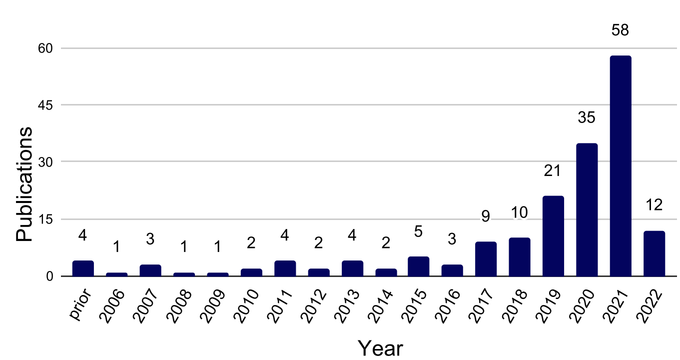

Fig. 3. Identified medical informed machine learning publications per year.

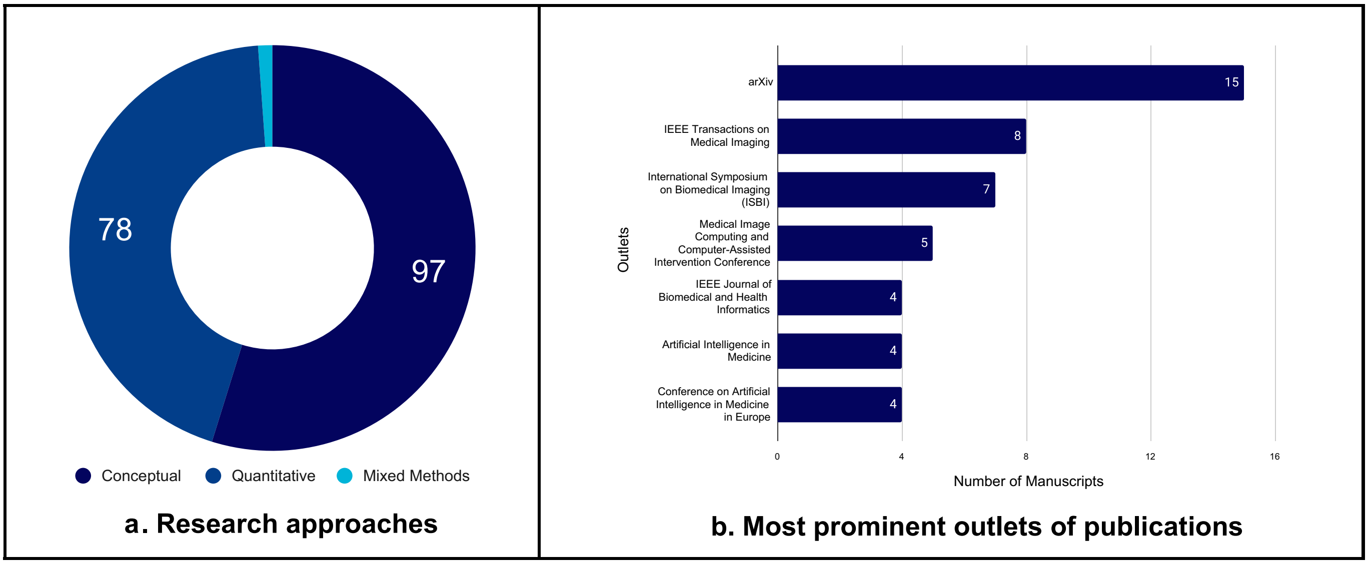

Fig. 4. Descriptive results on research approaches and outlets.

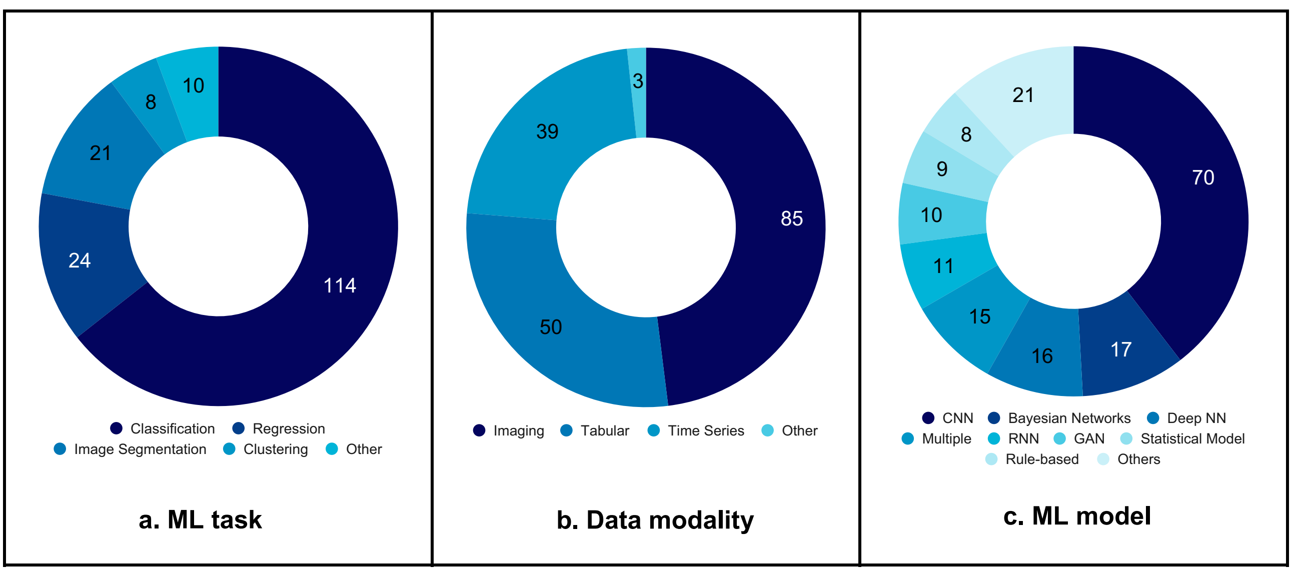

Fig. 5. Descriptive results on ML tasks, data modalities, and ML models.

3.2. Inclusion of domain knowledge in medical machine learning

In the selected publications, most papers include DK in ML models in one way. However, we identify three papers using multiple sources or structures of DK [11,50,51]. With the inclusion of multiple sources of DK, we analyze 181 combinations of DK and ML models out of 177 manuscripts. When investigating the relation of source, structure, and application step of DK, we see interesting results in Fig. 6.

Scientific knowledge (n = 17) and world knowledge (n = 13) play an inferior role and most publications rely on expert knowledge (n = 151). As described above, we further divide expert knowledge into common medical knowledge (n = 68), indicating available DK to all or most medical experts, and specialized medical knowledge (n = 83), which requires application-specific knowledge. This heavy use of expert knowledge demonstrates the relevance of medical experts, which cannot be replaced by factual knowledge in the medical domain where many decisions need to be made on single case assessment.

Overall, all eight structures of DK are used in the publications led by spatial invariances (n = 41) and probabilistic relations (n = 37). Simulation results are the least frequent (n = 2). Knowledge graph representations of DK, like ICD-11 codes, rank third in the structure of DK across all identified publications (n = 30). The manuscripts apply DK in all steps of the ML pipeline, most frequently in the hypothesis set (n = 73) and learning algorithm (n = 50). Since both steps are directly related to the ML models themselves and not the related training data or final hypothesis, this indicates that adapting the design of ML models provides the best possibilities to include DK. The limited scientific knowledge present in the identified manuscripts is mostly represented as algebraic (n = 6) or differential equations (n = 8). All differential equations come from scientific knowledge in the medical domain, like [52] who use Navier-Stokes equations to predict the drop in blood pressure. The remaining three manuscripts use scientific knowledge in the form of spatial invariances and probabilistic relations.

Spatial invariances are also frequently used to represent world knowledge (n = 5) along with knowledge graphs (n = 3). In other manuscripts, world knowledge is represented as algebraic equations, probabilistic relations, or logic rules. As an example, Molina et al. transformed brainstem auditory evoked potentials into symbolic

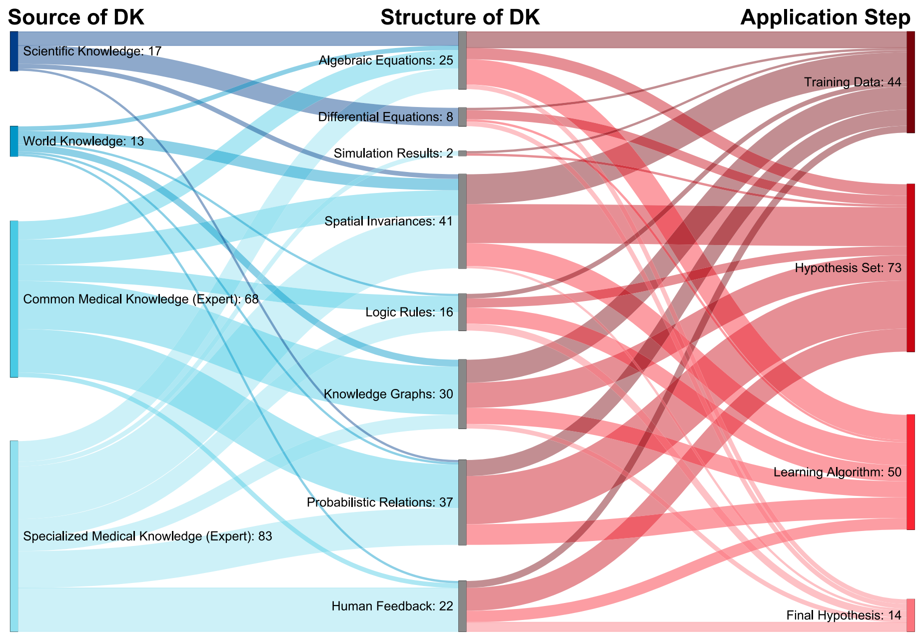

Fig. 6. Alluvial chart on the inclusion of domain knowledge.

patterns like "valleys" or "peaks" which can be transformed and understood by anyone. This transformation is then used to discover patterns for new predictions [53].

Common medical knowledge is frequently represented in the form of knowledge graphs (n = 21), such as ICD-11 codes, or probabilistic relations (n = 19), like following the estimated distribution of ECG domains [54]. Other representations of common medical knowledge include spatial invariances (n = 11), algebraic equations (n = 8), and logic rules (n = 7). Specialized medical knowledge is often represented by spatial invariances (n = 23) or human feedback (n = 19), for example, by assigning different symptoms to diagnoses [55].

Regarding the different structures of DK, algebraic equations are included along the entire ML model pipeline but most frequently in the learning algorithm (n = 11). The most frequent representations of knowledge, namely spatial invariances, probabilistic relations, and knowledge graphs are used across different steps of the ML pipeline except for the final output indicating broad application possibilities for those representations. Human feedback is included in all steps of the ML pipeline making it a versatile structure of DK.

In Fig. 7 we compare the structure of the included DK to the applied

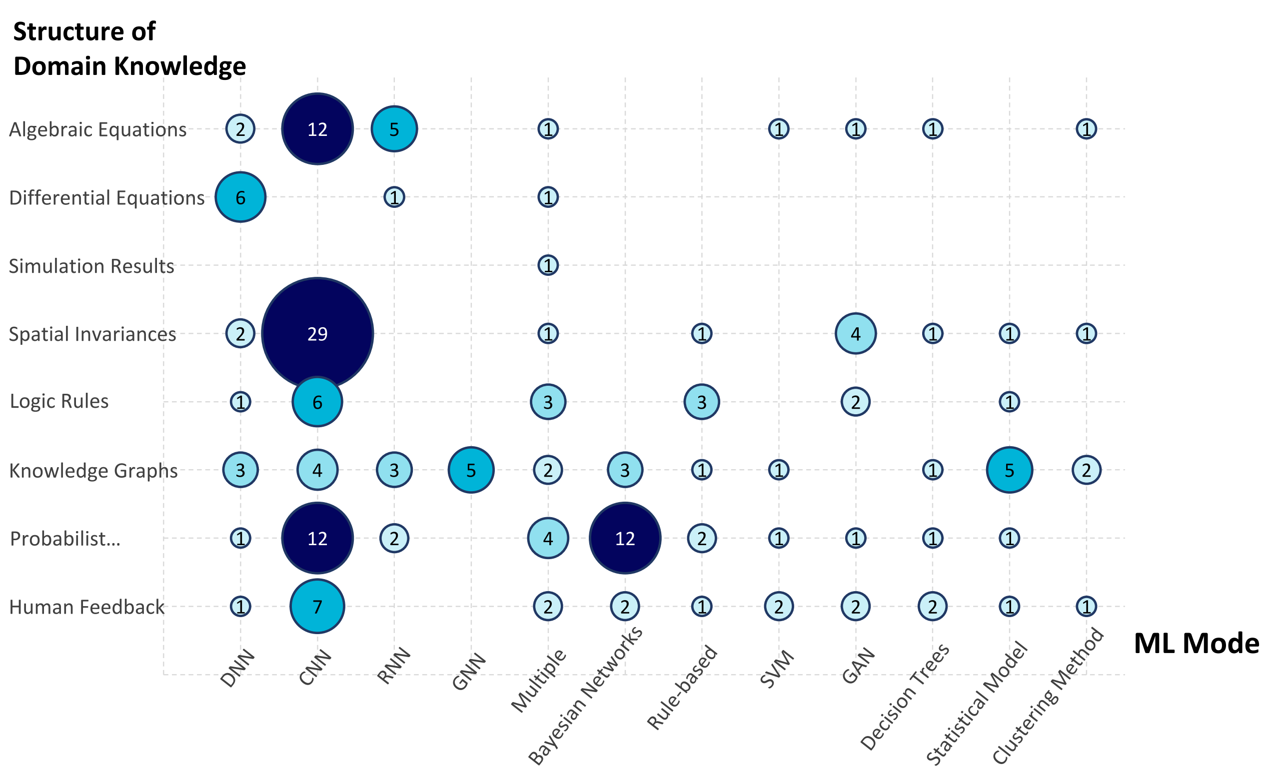

Fig. 7. Bubble chart showing the structure of domain knowledge per ML model.

ML models. We see that the most frequent ML models (i.e., CNNs) use almost every structure of DK. Spatial invariances are the most frequent structure of DK included in CNNs (n = 29), for example [56] who transform spatial constraints of dental images into a loss function of a CNN.

Following in frequency are algebraic equations (n = 12) and probabilistic relations (n = 12). The only structures of DK we have not seen included in CNNs are differential equations and simulation results. This is interesting because differential equations themselves are used by deep and recurrent neural networks, for example, to estimate patient-specific blood flows [16].

Probabilistic relations are most frequently used in Bayesian networks. The relations mimic the DK in the hypothesis set so that, for example, the connection of patient attributes can be used to predict the disease pathogenesis in obstructive sleep apnea [30]. Knowledge graphs are used to inform statistical models and neural networks, in particular graph neural networks.

3.3. Motives for medical informed machine learning

Next, we investigate the manifold motives for including DK in medical ML models. Since some papers provide multiple motives, the total number exceeds 177 as is shown in Fig. 8. Frequent motives are regarding the medical domain, either using previously unexploited knowledge (n = 48) or overcoming the limited availability of experts (n = 23). Other motives focus on data issues like poor data quality (n = 17), complex data sets (n = 20), and limited data availability (n = 47) or practical limitations due to insufficient predictive performance (n = 43), difficult practical integration (n = 9), and a lack of trust in ML models (n = 34).

Fig. 9 shows the relations between the motives to use DK and the structure of the DK used in each approach. The most frequent structure of DK, namely spatial invariances, is included in 13 manuscripts to use previously unexploited knowledge. For example, Xie et al. emphasize the limited use of "specific properties of [the] medical domain" [58]. In their paper, they improve the semi-supervised detection of breast cancer by including properties of segmentation masks and focusing on a region of interest. In other cases, spatial invariances are used to mitigate limited data availability (n = 12) or to overcome insufficient predictive performance (n = 10), like an improved segmentation of deformation fields in the brain achieved by refining segmentation masks with specialized medical expert knowledge [36].

Improving predictive performance is also solved by including algebraic equations (n = 12) and probabilistic relations (n = 7). Probabilistic relations like population-level trends [59] are in turn used to target all identified motives, especially to capture medical uncertainty (n = 8) and to account for limited expert availability (n = 6).

Besides spatial invariances and probabilistic relations, knowledge graphs and human feedback are the two structures of DK that target all identified motives. Knowledge graphs are included in ML algorithms to use previously unexploited knowledge (n = 10), to increase trust in the algorithms (n = 10), and to mitigate limited data availability (n = 9). This is shown for example when a Bayesian network follows a combination of expert knowledge and data, providing easier interpretation of predictions compared to purely expert-driven networks [60].

Human feedback is distributed evenly across all eight motives and is included to overcome limited data availability in five cases such as combining many domain expert-based models and validating them on data to achieve an ensemble of domain expert models [61]. Human feedback is included in four manuscripts to overcome limited expert availability. For example, the inclusion of previous experiences with dose volumes into the loss function improves the efficiency and predictive performance of models predicting radiation treatment [62].

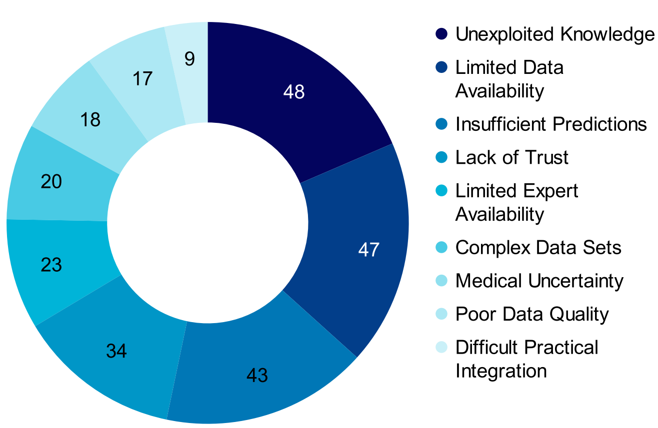

Fig. 8. Perceived motives to include domain knowledge in medical ML models.

3.4. Potential improvements in informing medical machine learning

During the inclusion of DK, most manuscripts identified and stated further potential improvements of their MedIML approaches. These improvements frequently revolve around future investigation, which can be seen from Fig. 10. The improvements range wide from the inclusion of additional knowledge (n = 34) like special constraints in individual cases [52] to the use of additional data (n = 41) so that the next step is to "apply the method to different prediction tasks" [63].

59 publications state they want to extend their experiments to validate the approach, which makes it the most frequent perceived improvement. This extension of experiments includes for example a "more comprehensive evaluation" of the approach [64] or the adaptation of the approach from labeled to unlabeled data sets [65].

The remaining publications discover challenges in the validation of the algorithm with practitioners (n = 10), improving the explainability of their approaches (n = 9), the deployment in practice (n = 13), and a reduction of the manual effort (n = 8) for example by automating the detection of brain tumors [66]. The authors of 42 publications do not specify the challenges they would like to target in the future.

Comparing the structure of DK to the potential improvements identified in the manuscripts, the inclusion of algebraic (n = 14) and differential equations (n = 9) needs to be validated in further experiments in the medical domain. For example, including expert-driven ordinary differential equations to predict the progression of COVID-19 improves the predictive performance but needs further validation for the prediction in other diseases [67]. Overall, validating the developed approaches with further experiments is the most frequently observed challenge in MedIML.

We see from Fig. 11 that using more data is the second most frequently stated potential improvement, discovered across most structures of DK, especially in probabilistic relations (n = 14) like Hong et al. who want to extend their knowledge-guided attention network with "other health data such as electronic health records" [68]. Using more data is also a potential improvement for DK represented as knowledge graphs (n = 7) and algebraic equations (n = 6).

In manuscripts using probabilistic relations, potential improvements include additional knowledge (n = 11) and extending their experiments (n = 9). For example, the model developed by Kong et al. could benefit by including not only static decision guidelines but also by including dynamic risk assessments [69]. Potential improvements in predictive performance are frequently stated by manuscripts including spatial invariances (n = 11) like using the available data to "test the performance of the proposed framework" [56].

4. Discussion

4.1. Research directions for medical informed machine learning

Since research on IML and in particular on MedIML is still nascent,

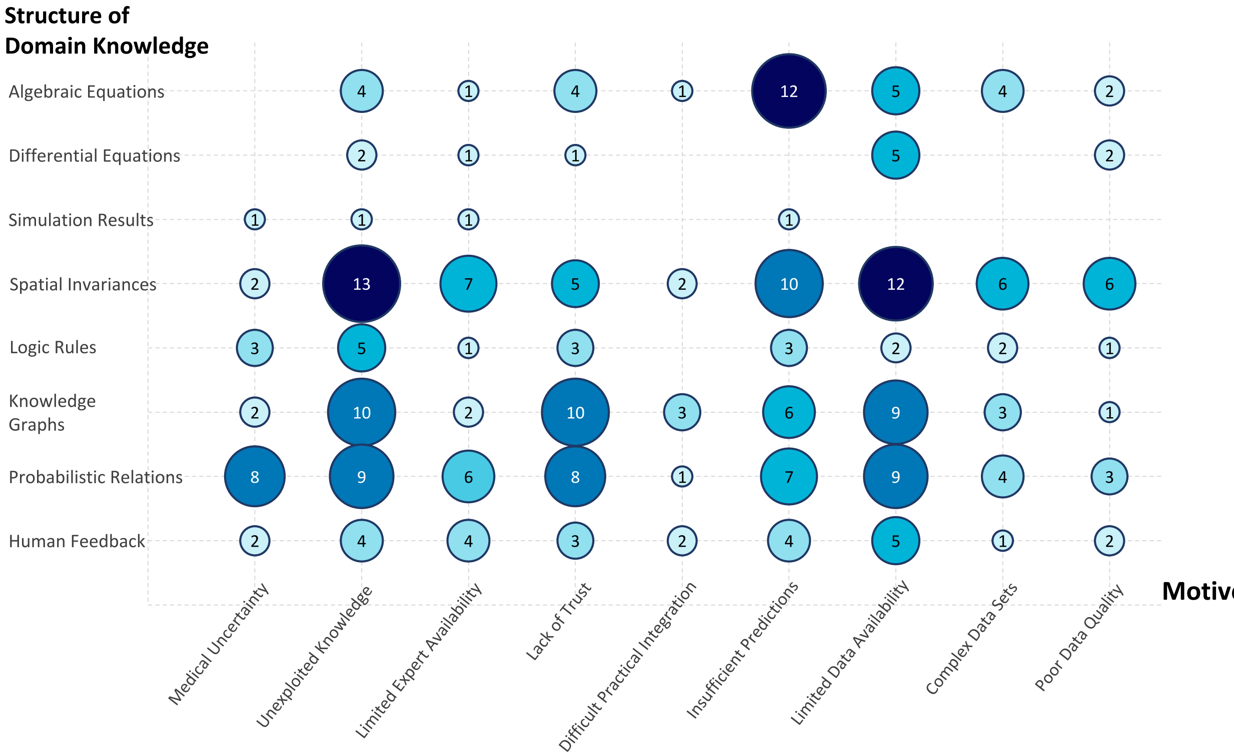

Fig. 9. Bubble chart on the structure of domain knowledge per motive.

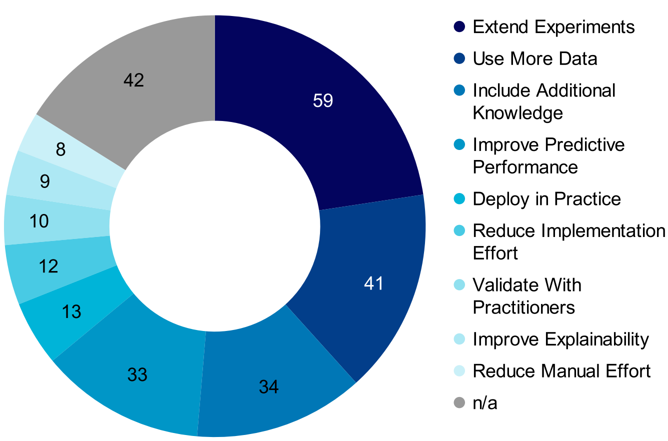

Fig. 10. Potential improvements of including domain knowledge.

there remain many open questions and knowledge gaps. Toward that end, we identified five directions for future research that build on our outlined findings. For every research direction, we first present the current status identified in this study before we propose exemplary starting points for future analysis that are meant to be adapted for individual research efforts.

4.1.1. Direction 1: Examining experts' decision-making processes

In our data, we observed heavy use of expert knowledge in MedIML approaches (85 %). To provide a more fine-grained analysis of the use of expert knowledge in MedIML, we further differentiated between common and specialized medical expert knowledge. Following this distinction, approaches utilizing specialized medical expert knowledge account for 47 %, whereas common medical expert knowledge is used in 38 % of the approaches. With this heavy use of expert knowledge, we see that experts provide essential information to guide ML models. However, the more specialized the required knowledge gets, the fewer experts are available for cooperation. To reduce the effort for these specialized experts, research should investigate their decision-making processes. With that research, the decision-process can be incorporated into ML models and therefore reduce the required time of experts in the long term. Research has already started to investigate the decisions made by experts and identified an essential role of intuition therein [70]. While these approaches provide an initial step, a better understanding of experts' decision-making processes might guide ML models without spending immense effort on individual adaptations.

4.1.2. Direction 2: Improving working environments of experts for knowledge extraction

In our sample, human feedback was incorporated across all application steps. The most prominent application step was the hypothesis set, which accounts for 45 % of the papers utilizing human feedback, while 14 % used human feedback in the training data step. This indicates that experts provide essential information and enhance ML models at every stage.

This versatile use of human feedback shows the relevance of human decisions for ML models. By leveraging expert-based and data-driven insights, patient health could be improved. Future projects should enhance the working environments of experts to account for this coexistence and the benefits for both approaches. This includes experts assisted by ML models in their decisions and experts enhancing ML models by explaining and incorporating their knowledge. The first steps in this direction include human-in-the-loop approaches, where experts iteratively label imaging data [71], and the development of persuasive annotation tools to enhance the annotation quality made by experts [72].

4.1.3. Direction 3: Developing guidelines for domain knowledge inclusion

When we investigated the structure of the knowledge included in each ML approach, several frequently used combinations emerged. We found, for example, spatial invariances frequently included in CNNs, like [58] who included regions of interest in the learning process of the CNN for detecting breast cancer. We also frequently saw probabilistic

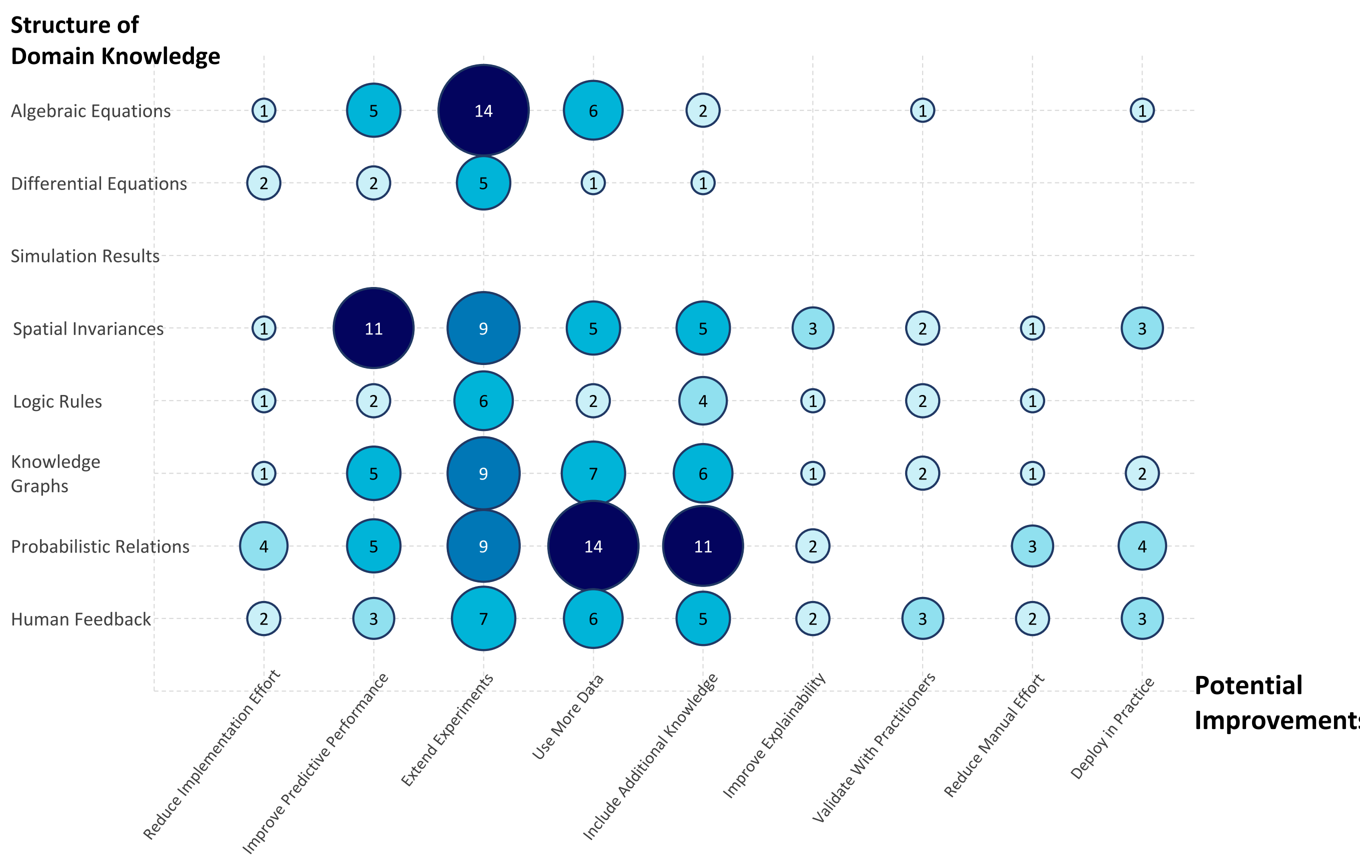

Fig. 11. Mapping of potential improvements on knowledge representations.

relations represented in the structure of Bayesian networks (e.g., [73]). While such clusters indicate that those combinations are frequently used together, they do not provide insights that some ML model types are better suited for certain knowledge representations than others. We, therefore, suggest that future research develops guidelines on the inclusion of DK into ML approaches. Such guidelines could take the available DK into account and then provide the next steps on what types of ML models could represent the DK best.

Additionally, guidelines should include the motives for performing IML. In our review, we saw the motives to be manifold and ranging from issues in data availability and quality to the limited availability of experts. All nine identified motives are almost evenly distributed in their frequency. As such, we could not detect any clusters regarding the motives of applying IML and the structure of the DK included, which might be due to the novelty of the field and the previous lack of investigation on why to include certain structures of DK. Newly developed guidelines might identify pipelines that use the motives of performing IML to identify the best-suiting DK, which could then provide the best-suited inclusion in ML models. Overall, these guidelines must be easy to adapt for specific use cases and should only provide first steps in the development of MedIML solutions.

Based on frequent combinations within our study, we developed a potential guideline shown in Fig. 12. In this draft, we envision a requirement on the motive for performing IML for the guideline. The guideline in turn should provide insights on what structures of DK work best to achieve this motive. Our example in Fig. 12 shows *unexploited knowledge* as a motive for performing MedIML which was frequently solved in our sample by using *spatial invariances* (cf. Fig. 9). The structure of DK, in turn, motivates the best-suited ML model, in our case (convolutional) *neural networks* (cf. Fig. 7). This draft is purely based on our sample and the findings of this study and we encourage researchers and practitioners to develop guidelines by interviewing domain experts.

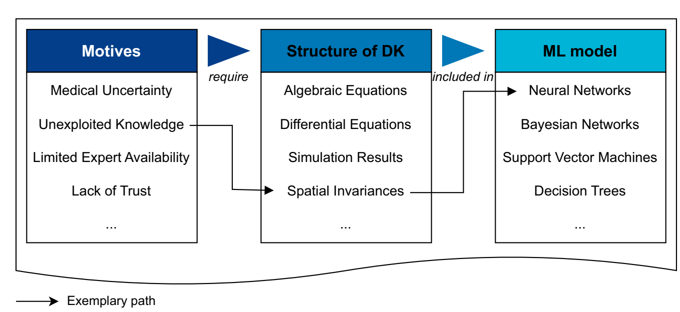

Fig. 12. Exemplary guideline to include domain knowledge.

4.1.4. Direction 4: Exploring combinations of multiple informed machine learning approaches

The results of our literature review further highlight multiple papers with customized MedIML approaches. Our sample indicates that despite manifold motives (e.g., improvements in explainability or robustness), potential improvements mainly revolve around experimental enhancements like using additional data or applying different data sets. We found only a few papers aiming at further improvements in the explainability or robustness of their approaches (e.g., [50,74]).

One potential mitigation strategy could be a cascade of IML approaches, where multiple approaches are executed after another. This, however, could result in very complex pipelines which require increased computational power while providing only minor benefits in predictive or explanatory performance. Combining multiple IML approaches might overcome multiple issues as well and could gain the benefits of all combined approaches. However, an aggregation of IML approaches might be complex to design and highly situational in their application, which raises the question of how approaches could be combined best.

One promising research direction could be to ensure the privacy-preserving exchange of IML approaches via a combination of federated learning and IML, where each client trains a local IML approach. The local models are then aggregated to a joint model by a central server, thus leveraging the DK present at every client [14].

4.1.5. Direction 5: Increasing domain-specific understanding of informed machine learning

Lastly, there are several differences in our domain-specific sample compared to the domain-agnostic taxonomy of von Rueden et al. [6]. Besides the immense use of expert knowledge explained in direction 1, our sample also showed heavy use of image data (48 %). While models trained on medical image data provide important and easy-to-acquire support in diagnoses, other data modalities like electrocardiogram (ECG) signals might be insightful as well. Some approaches conducted on imaging data might also be applicable to other data modalities like the approach of Sekuboyina et al. [75] whose multi-label approach should be extendable to time series or tabular data as well.

We did, however, not only identify differences but also similarities compared to the same taxonomy. Within our sample, DK was included across the inclusion steps similarly to the domain-agnostic taxonomy [6]. In our sample, as well as in a domain-agnostic sample, human feedback plays a major role indicating the relevance of experts in every domain. Researchers and practitioners might benefit from investigating IML approaches in other domains to gain an understanding of common transformations or representations of domain knowledge.

We, therefore, encourage other researchers to replicate this study for other domains to strengthen our findings and to identify other domain-specific characteristics. In other domain-specific reviews, more similarities and differences could be identified providing even more insights into the future directions of IML.

4.2. Implications

This study provides an overview of the status of the literature on IML in medicine. This overview can guide researchers in developing new approaches or adapting existing ones to their use case. The five identified research directions could enhance new research proposals that ultimately improve IML approaches not only in medicine but potentially in other domains as well.

Our survey, especially the flow diagram presented in Fig. 6, can help practitioners to identify suitable approaches to incorporate medical DK into ML models. Depending on the available DK, suitable preprocessing steps can be identified and executed. Additionally, if MedIML should be performed for a specific reason, our mappings in Fig. 9 provide insights into frequently used structures of DK which can serve as starting points.

4.3. Limitations

Our study is subject to some limitations. First, our classification scheme is subjective to some degree. For example, the distinction between the sources of DK (scientific knowledge, world knowledge, and expert knowledge) is not always trivial. We addressed this concern by letting two researchers code the first 20 papers independently and aligning the understanding of each coding dimension. Throughout the coding process, we conducted regular discussions between the researchers to resolve inconsistencies and ensured a common understanding of the concepts between all researchers.

The identified motives and potentials for improvement are limited to statements made by the authors in their respective papers. There may be additional motives for performing MedIML and additional potentials for improvement, which we could not identify. Future research could resolve the limitations by conducting interviews with MedIML researchers to reveal additional implicit and explicit motives for including DK in medical ML.

5. Conclusion

IML is increasingly used to combine the benefits of conventional ML models and human DK. In this scoping study, we investigated the current state of research on IML in medicine. Based on an existing taxonomy, we analyzed the required DK, the applied ML models, motives for performing IML, and potential improvements for all approaches. Compared to previous domain-agnostic studies, we identified various interesting differences in MedIML, most notably the predominant role of expert knowledge over more general types of DK and the prevalence of image data in the medical domain. Our review describes several auspicious directions to advance the state of MedIML by formalizing medical expert knowledge and identifying suitable approaches for different motives and types of DK. Research in other domains may benefit from complementary reviews to uncover similar domain-specific differences, so multidisciplinary approaches can be developed.

Declaration of competing interest

The authors do not have any conflicts of interest to disclose.

Acknowledgment

This research was partially funded by the Federal Ministry for Economic Affairs and Climate Action of Germany (BMKW). Grant number: 01MJ21002D.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.artmed.2023.102676.

References

[1] Wang F, Preininger A. AI in health: state of the art, challenges, and future directions. Yearb Med Inform Aug. 2019;28(01):016–26. https://doi.org/10.1055/s-0039-1677908.

[2] Frank SA. Immunology and evolution of infectious disease. Princeton University Press; 2002. https://doi.org/10.1515/9780691220161.

[3] Guse R, Thiebes S, Hennel P, Rosenkranz C, Sunyaev A. How do employees perceive digital transformation and its effects? A theory of the smart machine perspective. In: Proceedings of the 43rd International Conference on Information Systems (ICIS), Association for Information Systems (AIS), 9th–14th December 2022, Copenhagen, Denmark. Association for Information Systems (AIS); 2022.

[4] Dick RS, Steen EB, Detmer DE. The computer-based patient record: an essential technology for health care. In: Institute of Medicine (US) committee on improving the patient record. revised ed. Washington, D.C.: National Academies Press; 1997. p. 5306. https://doi.org/10.17226/5306.

[5] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell May 2019;1(5):206–15. https://doi.org/10.1038/s42256-019-0048-x.

[6] von Rueden L, et al. Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems. IEEE Trans Knowl Data Eng 2021;35(1):614–33. https://doi.org/10.1109/TKDE.2021.3079836.

[7] Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med Apr. 2019;380(14):1347–58. https://doi.org/10.1056/NEJMra1814259.

[8] Giger ML. Machine learning in medical imaging. J Am Coll Radiol Mar. 2018;15(3):512–20. https://doi.org/10.1016/j.jacr.2017.12.028.

[9] Xie X, Niu J, Liu X, Chen Z, Tang S, Yu S. A survey on incorporating domain knowledge into deep learning for medical image analysis. Med Image Anal Apr. 2021;69:101985. https://doi.org/10.1016/j.media.2021.101985.

[10] Dravid A, Schiffers F, Gong B, Katsaggelos AK. medXGAN: visual explanations for medical classifiers through a generative latent space. In: Preprint, arXiv; 2022. https://doi.org/10.48550/ARXIV.2204.05376.

[11] Lian J, et al. A structure-aware relation network for thoracic diseases detection and segmentation. IEEE Trans Med Imaging Aug. 2021;40(8):2042–52. https://doi.org/10.1109/TMI.2021.3070847.

[12] Shi G, et al. Knowledge-guided synthetic medical image adversarial augmentation for ultrasonography thyroid nodule classification. Comput Methods Programs Biomed Nov. 2020;196:105611. https://doi.org/10.1016/j.cmpb.2020.105611.

[13] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol Feb. 2005;8(1):19–32. https://doi.org/10.1080/1364557032000119616.

[14] Rieke N, et al. The future of digital health with federated learning. NPJ Digit Med Dec. 2020;3(1):119. https://doi.org/10.1038/s41746-020-00323-1.

[15] Wieringa R, Maiden N, Mead N, Rolland C. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. Requir Eng Mar. 2006;11(1):102–7. https://doi.org/10.1007/s00766-005-0021-6.

[16] Azmat M, Tu E, Branch KR, Alessio A. Machine learned versus analytical models for estimation of Fractional Flow Reserve (FFR) from CT-derived information. In: Gimi BS, Krol A, editors. Medical imaging 2021: biomedical applications in molecular, structural, and functional imaging. United States: SPIE; Feb. 2021. p. 29. https://doi.org/10.1117/12.2581701. Online Only.

[17] Yu W, Zhou H, Choi Y, Goldin JG, Teng P, Kim GHJ. An automatic diagnosis of idiopathic pulmonary fibrosis (IPF) using domain knowledge-guided attention models in HRCT images. In: Drukker K, Mazurowski MA, editors. Medical imaging 2021: computer-aided diagnosis. United States: SPIE; Feb. 2021. p. 67. https://doi.org/10.1117/12.2579735. Online Only.

[18] Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. BMC Med Inform Decis Mak Nov. 2021;21 (S6):206. https://doi.org/10.1186/s12911-021-01534-6.

[19] Luo J, Kitamura G, Doganay E, Arefan D, Wu S. Medical knowledge-guided deep curriculum learning for elbow fracture diagnosis from x-ray images. In: Drukker K, Mazurowski MA, editors. Medical imaging 2021: computer-aided diagnosis. United States: SPIE; Feb. 2021. p. 35. https://doi.org/10.1117/12.2582184. Online Only.

[20] Mei J, Xia E. Knowledge learning symbiosis for developing risk prediction models from regional EHR repositories. Stud Health Technol Inform Aug. 2019;264: 258–62. https://doi.org/10.3233/SHTI190223.

[21] Qoku A, Buettner F. Encoding domain knowledge in multi-view latent variable models: a Bayesian approach with structured sparsity. In: Preprint arXiv; 2022. https://doi.org/10.48550/ARXIV.2204.06242.

[22] Guerrero JM, Martínez-Tomás R, Rincón M, Peraita H. Diagnosis of cognitive impairment compatible with early diagnosis of Alzheimer's disease: a Bayesian network model based on the analysis of oral definitions of semantic categories. Methods Inf Med 2016;55(01):42–9. https://doi.org/10.3414/ME14-01-0071.

[23] Chen C, Wang Y, Niu J, Liu X, Li Q, Gong X. Domain knowledge powered deep learning for breast cancer diagnosis based on contrast-enhanced ultrasound videos. IEEE Trans Med Imaging Sep. 2021;40(9):2439–51. https://doi.org/10.1109/TMI.2021.3078370.

[24] Nelson CA, Bove R, Butte AJ, Baranzini SE. Embedding electronic health records onto a knowledge network recognizes prodromal features of multiple sclerosis and predicts diagnosis. J Am Med Inform Assoc Jan. 2022;29(3):424–34. https://doi.org/10.1093/jamia/ocab270.

[25] Andrushia A, Sagayam K, Dang H, Pomplun M, Quach L. Visual-saliency-based abnormality detection for MRI brain images—Alzheimer's disease analysis. Appl Sci Oct. 2021;11(19):9199. https://doi.org/10.3390/app11199199.

[26] Vakanski A, Xian M, Freer PE. Attention-enriched deep learning model for breast tumor segmentation in ultrasound images. Ultrasound Med Biol Oct. 2020;46(10): 2819–33. https://doi.org/10.1016/j.ultrasmedbio.2020.06.015.

[27] Nori N, Kashima H, Yamashita K, Ikai H, Imanaka Y. Simultaneous modeling of multiple diseases for mortality prediction in acute hospital care. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. Sydney NSW Australia: ACM; Aug. 2015. p. 855–64. https://doi.org/10.1145/2783258.2783308.

[28] Chen Z, Wu M, Cui W, Liu C, Li X. An attention based CNN-LSTM approach for sleep-wake detection with heterogeneous sensors. IEEE J Biomed Health Inform Sep. 2021;25(9):3270–7. https://doi.org/10.1109/JBHI.2020.3006145.

[29] Wen G, Ma J, Hu Y, Li H, Jiang L. Grouping attributes zero-shot learning for tongue constitution recognition. Artif Intell Med Sep. 2020;109:101951. https://doi.org/10.1016/j.artmed.2020.101951.

[30] Huynh PK, Setty A, Phan H, Le TQ. Probabilistic domain-knowledge modeling of disorder pathogenesis for dynamics forecasting of acute onset. Artif Intell Med May 2021;115:102056. https://doi.org/10.1016/j.artmed.2021.102056.

[31] Borges P, et al. Physics-informed brain MRI segmentation. In: Burgos N, Gooya A, Svoboda D, editors. Simulation and synthesis in medical imaging. Lecture Notes in Computer Science, vol. 11827. Cham: Springer International Publishing; 2019. p. 100–9. https://doi.org/10.1007/978-3-030-32778-1_11.

[32] Quinn JA, Williams CKI, McIntosh N. Factorial switching linear dynamical systems applied to physiological condition monitoring. IEEE Trans Pattern Anal Mach Intell Sep. 2009;31(9):1537–51. https://doi.org/10.1109/TPAMI.2008.191.

[33] Oh S, Lee M. A shallow domain knowledge injection (SDK-injection) method for improving CNN-based ECG pattern classification. Appl Sci Jan. 2022;12(3):1307. https://doi.org/10.3390/app12031307.

[34] Huang Y, et al. CHI: a contemporaneous health index for degenerative disease monitoring using longitudinal measurements. J Biomed Inform Sep. 2017;73: 115–24. https://doi.org/10.1016/j.jbi.2017.07.003.

[35] Banerjee R, Ghose A, Sinha A, Pal A, Mandana KM. A multi-modal approach for non-invasive detection of coronary artery disease. In: Adjunct proceedings of the 2019 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2019 ACM international symposium on wearable computers. London United Kingdom: ACM; Sep. 2019. p. 543–50. https://doi.org/10.1145/3341162.3349331.

[36] Zhang Y, Liu R, Li Z, Liu Z, Fan X, Luo Z. Coupling principled refinement with bi-directional deep estimation for robust deformable 3D medical image registration. In: 2020 IEEE 17th international symposium on biomedical imaging (ISBI). Iowa City, IA, USA: IEEE; Apr. 2020. p. 86–90. https://doi.org/10.1109/ISBI45749.2020.9098526.

[37] Sobecki P, Jóźwiak R, Sklinda K, Przelaskowski A. Effect of domain knowledge encoding in CNN model architecture—a prostate cancer study using mpMRI images. PeerJ Mar. 2021;9:e11006. https://doi.org/10.7717/peerj.11006.

[38] Ma F, Gao J, Suo Q, You Q, Zhou J, Zhang A. Risk prediction on electronic health records with prior medical knowledge. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. London, United Kingdom: ACM; Jul. 2018. p. 1910–9. https://doi.org/10.1145/3219819.3220020.

[39] Shknevsky A, Shahar Y, Moskovitch R. The semantic adjacency criterion in time intervals mining. In: Preprint arXiv; 2021. https://doi.org/10.48550/ARXIV.2101.03842.

[40] Yin C, Zhao R, Qian B, Lv X, Zhang P. Domain knowledge guided deep learning with electronic health records. In: 2019 IEEE international conference on data mining (ICDM). Beijing, China: IEEE; Nov. 2019. p. 738–47. https://doi.org/10.1109/ICDM.2019.00084.

[41] Yu W, Zhou H, Choi Y, Goldin JG, Kim GHJ. Mga-net: multi-scale guided attention models for an automated diagnosis of idiopathic pulmonary fibrosis (IPF). In: 2021 IEEE 18th international symposium on biomedical imaging (ISBI). Nice, France: IEEE; Apr. 2021. p. 1777–80. https://doi.org/10.1109/ISBI48211.2021.9433956.

[42] Zhang M, King CR, Avidan M, Chen Y. Hierarchical attention propagation for healthcare representation learning. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. Virtual Event, CA, USA: ACM; Aug. 2020. p. 249–56. https://doi.org/10.1145/3394486.3403067.

[43] Linial O, Ravid N, Eytan D, Shalit U. Generative ODE modeling with known unknowns. In: Proceedings of the conference on health, inference, and learning. Virtual Event, USA: ACM; Apr. 2021. p. 79–94. https://doi.org/10.1145/3450439.3451866.

[44] Yet B, Marsh W, Morrissey D. Towards an evidence-based decision support tool for management of musculoskeletal conditions. Stud Health Technol Inform 2018;255: 175–9.

[45] Leung KH, et al. A physics-guided modular deep-learning based automated framework for tumor segmentation in PET. Phys Med Biol Dec. 2020;65(24): 245032. https://doi.org/10.1088/1361-6560/ab8535.

[46] Guo X, et al. An expert-in-the-loop paradigm for learning medical image grouping. In: Bailey J, Khan L, Washio T, Dobbie G, Huang JZ, Wang R, editors. Advances in knowledge discovery and data mining. Lecture notes in computer science, vol. 9651. Cham: Springer International Publishing; 2016. p. 477–88. https://doi.org/10.1007/978-3-319-31753-3_38.

[47] Zhang MY, Du RZ. A real-time inference method of graph attention network based on knowledge graph for lung cancer. In: 2021 5th international conference on digital signal processing. Chengdu China: ACM; Feb. 2021. p. 326–31. https://doi.org/10.1145/3458380.3458436.

[48] Jarrett K, Kavukcuoglu K, Ranzato MA, LeCun Y. What is the best multi-stage architecture for object recognition?. In: 2009 IEEE 12th international conference on computer vision. Kyoto: IEEE; Sep. 2009. p. 2146–53. https://doi.org/10.1109/ICCV.2009.5459469.

[49] LeCun Y, et al. Handwritten digit recognition with a back-propagation network. In: Touretzky D, editor. Advances in neural information processing systems. Morgan-Kaufmann; 1989.

[50] Pezzotti N, et al. An adaptive intelligence algorithm for undersampled knee MRI reconstruction. IEEE Access 2020;8:204825–38. https://doi.org/10.1109/ACCESS.2020.3034287.

[51] Antal P, et al. Extended Bayesian regression models: a symbiotic application of belief networks and multilayer perceptrons for the classification of ovarian tumors. In: Lecture notes in computer science. vol. 2101. Springer; Jan. 01, 2001. p. 177–87.

[52] Fossan FE, et al. Machine learning augmented reduced-order models for FFR-prediction. Comput Methods Appl Mech Eng Oct. 2021;384. https://doi.org/10.1016/j.cma.2021.113892.

[53] Molina ME, Perez A, Valente JP. Classification of auditory brainstem responses through symbolic pattern discovery. Artif Intell Med Jun. 2016;70:12–30. https://doi.org/10.1016/j.artmed.2016.05.001.

[54] Jia Z, Shi Y, Saba S, Hu J. On-device prior knowledge incorporated learning for personalized atrial fibrillation detection. ACM Trans Embed Comput Syst Oct. 2021;20(5s):1–25. https://doi.org/10.1145/3476987.

[55] Quinteros J, et al. Diagnostic test suggestion via Bayesian network of non-expert assisted knowledge base. In: 2018 20th international conference on advanced communication technology (ICACT), Chuncheon-si Gangwon-do. Korea (South): IEEE; Feb. 2018. p. 340–6. https://doi.org/10.23919/ICACT.2018.8323748.

[56] Huang J, Yan H, Li J, Stewart HM, Setzer F. Combining anatomical constraints and deep learning for 3-D CBCT dental image multi-label segmentation. In: 2021 IEEE 37th international conference on data engineering (ICDE), Chania. Greece: IEEE; Apr. 2021. p. 2750–5. https://doi.org/10.1109/ICDE51399.2021.00319.

[58] Xie X, Niu J, Liu X, Li Q, Wang Y, Tang S. DK-consistency: a domain knowledge guided consistency regularization method for semi-supervised breast cancer diagnosis. In: 2021 IEEE international conference on bioinformatics and biomedicine (BIBM). Houston, TX, USA: IEEE; Dec. 2021. p. 3435–42. https://doi.org/10.1109/BIBM52615.2021.9669494.

[59] Yu W, Zhou H, Goldin JG, Wong WK, Kim GHJ. End-to-end domain knowledge-assisted automatic diagnosis of idiopathic pulmonary fibrosis (IPF) using computed tomography (CT). Med Phys May 2021;48(5):2458–67. https://doi.org/10.1002/mp.14754.

[60] Velikova M, Lucas PJF, Samulski M, Karssemeijer N. On the interplay of machine learning and background knowledge in image interpretation by Bayesian networks. Artif Intell Med Jan. 2013;57(1):73–86. https://doi.org/10.1016/j.artmed.2012.12.004.

[61] Gavrishchaka V, Yang Z, Miao R, Senyukova O. Advantages of hybrid deep learning frameworks in applications with limited data. Int J Mach Learn Comput 2018;8(6): 549–58.

[62] Nguyen D, et al. Incorporating human and learned domain knowledge into training deep neural networks: a differentiable dose-volume histogram and adversarial inspired framework for generating Pareto optimal dose distributions in radiation therapy. Med Phys Mar. 2020;47(3):837–49. https://doi.org/10.1002/mp.13955.

[63] Radovanovic S, Vukicevic M, Kovacevic A, Stiglic G, Obradovic Z. Domain knowledge based hierarchical feature selection for 30-day hospital readmission prediction. Artif Intell Med 2015;9105:96–100. https://doi.org/10.1007/978-3-319-19551-3_11.

[64] Yu C, Ren G, Liu J. Deep inverse reinforcement learning for sepsis treatment. In: 2019 IEEE international conference on healthcare informatics (ICHI). Xi'an, China: IEEE; Jun. 2019. p. 1–3. https://doi.org/10.1109/ICHI.2019.8904645.

[65] Li K, Wang S, Yu L, Heng PA. Dual-teacher++: exploiting intra-domain and inter-domain knowledge with reliable transfer for cardiac segmentation. IEEE Trans Med Imaging Oct. 2021;40(10):2771–82. https://doi.org/10.1109/TMI.2020.3038828.

[66] Huang M, Yang W, Yu M, Lu Z, Feng Q, Chen W. Retrieval of brain tumors with region-specific bag-of-visual-words representations in contrast-enhanced MRI images. Comput Math Methods Med 2012;2012:1–17. https://doi.org/10.1155/2012/280538.

[67] Qian Z, Zame WR, Fleuren LM, Elbers P, van der Schaar M. Integrating expert ODEs into neural ODEs: pharmacology and disease progression. In: 35th conference on neural information processing systems (NeurIPS 2021). Virtual only: arXiv; 2021. https://doi.org/10.48550/ARXIV.2106.02875.

[68] Hong S, Xiao C, Ma T, Li H, Sun J. MINA: multilevel knowledge-guided attention for modeling electrocardiography signals. In: Proceedings of the twenty-eighth international joint conference on artificial intelligence (IJCAI-19). Macao, China: arXiv; 2019. https://doi.org/10.48550/ARXIV.1905.11333.

[69] Kong G, Xu D-L, Body R, Yang J-B, Mackway-Jones K, Carley S. A belief rule-based decision support system for clinical risk assessment of cardiac chest pain. Eur J Oper Res Jun. 2012;219(3):564–73. https://doi.org/10.1016/j.ejor.2011.10.044.

[70] Baylor AL. A U-shaped model for the development of intuition by level of expertise. New Ideas in Psychol Dec. 2001;19(3):237–44. https://doi.org/10.1016/S0732-118X(01)00005-8.

[71] Zheng H, Chen Y, Yue X, Ma C. Deep interactive segmentation of uncertain regions with shadowed sets. In: Proceedings of the third international symposium on image computing and digital medicine - ISICDM 2019. Xi'an, China: ACM Press; 2019. p. 244–8. https://doi.org/10.1145/3364836.3364885.

[72] Warsinsky S, Schmidt-Kraepelin M, Thiebes S, Wagner M, Sunyaev A. Gamified expert annotation systems: meta-requirements and tentative design. In: Drechsler A, Gerber A, Hevner A, editors. The transdisciplinary reach of design science research. Lecture notes in computer science, vol. 13229. Cham: Springer International Publishing; 2022. p. 154–66. https://doi.org/10.1007/978-3-031-06516-3_12.

[73] Liu B, Li Y, Ghosh S, Sun Z, Ng K, Hu J. Complication risk profiling in diabetes care: a Bayesian multi-task and feature relationship learning approach. IEEE Trans Knowl Data Eng Jul. 2020;32(7):1276–89. https://doi.org/10.1109/TKDE.2019.2904060.

[74] Kissas G, Yang Y, Hwuang E, Witschey WR, Detre JA, Perdikaris P. Machine learning in cardiovascular flows modeling: predicting arterial blood pressure from non-invasive 4D flow MRI data using physics-informed neural networks. Comput Methods Appl Mech Eng Jan. 2020;358:112623. https://doi.org/10.1016/j.cma.2019.112623.

[75] Sekuboyina A, Onoro-Rubio D, Kleesiek J, Malone B. A relational-learning perspective to multi-label chest X-ray classification. In: 2021 IEEE 18th international symposium on biomedical imaging (ISBI). Nice, France: IEEE; Apr. 2021. p. 1618–22. https://doi.org/10.1109/ISBI48211.2021.9433786.