

A vision for sustainable, equitable healthcare applications with XAI and transfer learning to augment neural models*

Aparna S. Varde^{1,*}, Benjamin S. B. Rasmussen^{2,‡}

¹SDU Center for Software Technology, Maersk McKinney Moller Institute (MMMI), Faculty of Engineering (TEK), University of Southern Denmark (SDU), Vejle, Denmark

²Department of Radiology and Faculty of Health Sciences, Radiology Research and Innovation Unit (UNIFY), Centre for Clinical Artificial Intelligence (CAI-X), Odense university hospital and University of Southern Denmark Odense, Denmark

Abstract

This paper presents a vision to propel approaches for sustainability and thus equity in healthcare applications with XAI (explainable AI) and transfer learning to augment deep learning neural models, e.g. multimodal ones (combining text, images and heterogeneous data). Transfer learning can promote scalability, i.e. hypotheses learned on one illness can be transferred to related settings. XAI solutions build trust and transparency in clinical decision-making (trace causes of errors, and map empirical observations to theories with factual reasoning etc.) As stated by real-life clinical practitioners: "Explainability is crucial to aid doctors to ensure that techniques work, and traceability helps improve solutions". It can thus benefit human-AI collaboration in medicine to enhance accessibility. The vision in this paper aims to make positive impacts on United Nations Sustainable Development Goals, e.g. Goal 3 on Good Health and Well-Being. It can play a key role in data analytics for real-life applications, especially by helping to advance global health and equity.

Keywords

AI in Healthcare, Commonsense Knowledge (CSK), Data Analytics for Social Good, Explainable AI (XAI), Knowledge Bases (KBs), Knowledge-Guided Machine Learning (KGML), Sustainable AI, Transfer Learning

1. Introduction

Background: Deep learning neural models in healthcare can be beneficial for a myriad of tasks such as assistance in disease diagnosis, interactive patient communication, and semantically-aware summarization of complex medical documents. Healthcare data can often be multimodal: plain text including doctor notes and research abstracts, medical images such as X-rays and MRI, audio recordings e.g. cough sounds, and video clips e.g. surgical procedures. Thus, to conduct data analytics in healthcare, neural models can possibly be used, especially multimodal ones, yet they pose issues. A few models and their issues are stated next.

Related Work: Multimodal neural models such as CLIP [1], Google [2], and GPT-4o [3] can process and generate text as well as images. CLIP (Contrastive Language-Image Pre-training) is a model that "learns how to connect images and text by learning simultaneous embeddings for both". Gemini can (among other things) fathom a photo of a nutritionist-recommended food item and provide its recipe or vice versa. A healthcare-specific multimodal model is Med-Gemini [4] that "builds upon Google's Gemini models by fine-tuning on de-identified medical data while inheriting Gemini's native reasoning, multimodal, and long-context abilities". While such multimodal models thriving on deep learning (DL) can be useful in healthcare, they pose issues e.g. lack of accuracy, errors without traceability, and conveying misinformation (which can be unintentional) [5], [6]. Moreover, DL models need colossal amounts of training data, and consume enormous amounts of energy. Training a neural model needs as much energy as 5 hybrid cars in their lifetime [7]. Thus, solely relying on DL can make

Published in the Workshop Proceedings of the EDBT/ICDT 2026 Joint Conference (March 24-27, 2026), Tampere, Finland

*Corresponding author.

✉ apva@mmpi.sdu.dk (A. S. Varde); benjamin.rasmussen@rsyd.dk (B. S. B. Rasmussen)

🌐 <https://portal.findresearcher.sdu.dk/en/persons/apva/> (A. S. Varde);

<https://cai-x.dk/om-cai-x/medarbejdere-og-organisering/benjamin-s-rasmussen> (B. S. B. Rasmussen)

🆔 0000-0002-3170-2510 (A. S. Varde); 0000-0002-7357-0755 (B. S. B. Rasmussen)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

AI-based solutions less affordable and accessible, posing equity issues. Furthermore, pure DL models lack explainability, hence adversely impacting interpretability, trust, evidence-based decisions as well as human-AI collaboration.



Figure 1: United Nations Sustainable Development Goals (UN SDGs); our vision aims to make positive impacts mainly on Goals 3, 10, and 11.

Motivation: Various issues with DL models motivate the need for greater sustainability and explainability. Sustainable healthcare is vital to make proposed solutions more accessible and hence equitable, aligning with the mission of UN SDGs (United Nations Sustainable Development Goals). The SDGs address healthcare, equity and sustainability (see Fig. 1). For instance, automated low-complexity disease diagnosis is imperative to promote global health initiatives. Likewise, adding explainability to AI-based healthcare solutions is highly desirable to aid meaningful communication (with reasoning) among medical staff and patients. It fosters more trustworthiness, consistency and evidence-based decision support. It also improves human-AI collaboration due to enhanced interpretation, trust and communication.

Proposed Vision: This vision paper advocates sustainable, equitable healthcare solutions. It aims to make positive impacts on UN SDGs (Goal 3: Good Health and Well-Being, Goal 10: Reduced Inequalities, Goal 11: Sustainable Cities and Communities). The crux of our vision is illustrated in Fig. 2 which synthesizes a few aspects such as trust in AI. In addition to exemplifying classical XAI methods, e.g. decision trees and counterfactuals, we propose the following main contributions.

1. *KGML (knowledge-guided machine learning) to add context and enhance accuracy*
2. *KB (knowledge base) design and commonsense knowledge (CSK) to supplement neural models*
3. *Exploration of transfer learning and low-complexity theories with efficiency*

An important aim is to ensure that healthcare is affordable and accessible to various regions across the globe. It promotes equity and contributes positively to the theme of data analytics in real-life applications.

2. Exploring the KGML Paradigm

A useful paradigm in recent years is knowledge-guided machine learning (KGML) which infuses domain knowledge in machine learning methods to add specific context, enhancing performance [8], [9]. Fig. 3 is an overview of the successful deployment of this paradigm in environmental science [10]. As seen here,

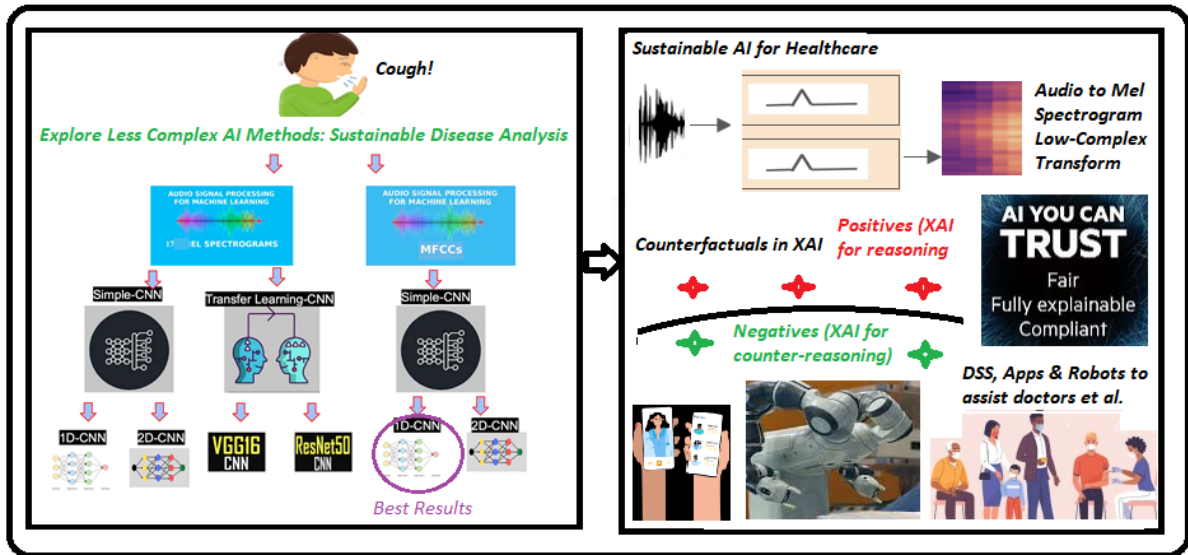


Figure 2: Investigating healthcare solutions: Low-complexity minimally invasive diagnosis (left); Multimodal analysis with images, audio-visual mapping, text, augmented by counterfactual reasoning & trustworthy solutions (right upper); Software development for human-AI collaboration to aid medical professionals (right lower).

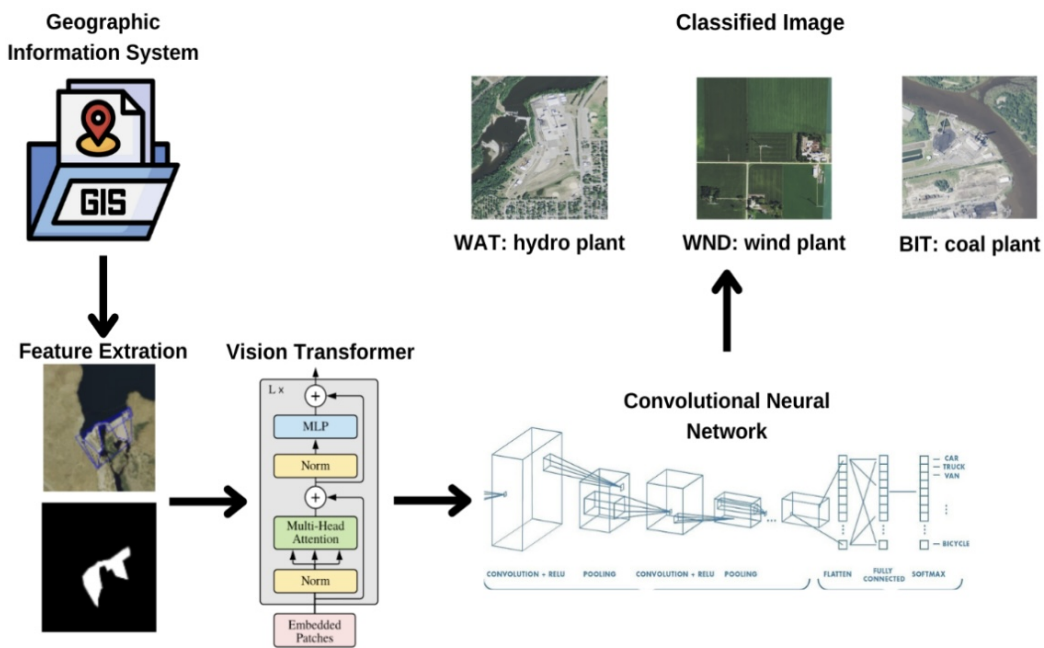


Figure 3: KGML example successfully deployed in environmental science [10]. Architecture analogous to this can be designed for healthcare as follows: ArcGIS can be replaced with documents in medical literature; ViT, CNN can be replaced with a text-cum-vision model e.g. CLIP; outputs can be: presence/absence of symptoms.

contextual information is extracted from ArcGIS and fed into an integrated framework of ViT (vision transformers) and CNN (convolutional neural networks) to enhance power plant classification. More specifically, this framework adds domain knowledge from GIS (geographic information system) data into ViT and CNN architectures to guide the learning with domain knowledge. This knowledge is integrated through spatial masks (SM) for adding context to power plant classification. It raises classification accuracy by approximately 10% across all power plant types, while also adding explainability, and saving training resources.

Likewise, our vision is to propose KGML in the healthcare domain as follows.

1. Work alongside medical experts to extract relevant domain knowledge.
2. Program using XAI methods to infuse the knowledge with multimodal DL models, e.g. CLIP.
3. Adapt the notion of RAG (Retrieval-Augmented Generation) in LLMs (large language models): medical documents to infuse RAG-context.
4. Test the performance and enhance the framework using human-in-the-loop for further optimization from a practical standpoint.

We have the following example inputs garnered by interacting with real-life physicians [11], [12]: *"For the purpose of integrating clinical context in deep learning based pneumonia classification, conventional deep learning models for pneumonia classification often rely solely on raw radiographic data. However, diagnostic accuracy can be significantly enhanced by augmenting these models with structured or textual clinical context, mirroring the integrative reasoning employed by us physicians"*. Thus, in clinical context integration, consider the scenario next.

Scenario 1: *KGML for clinical context in pyrexia analysis*

Inputs: Case histories of patients e.g. "72-year-old patient presents with pyrexia (39° C), an elevated C-reactive protein (CRP) level, and a productive cough persisting for three days. Physical examination reveals bilateral basal crackles."

KGML model as (DL + RAG): (1) Infuse variables with weights & markers given by doctors (as RAG data) into DL (2) This KGML model moves beyond simple pattern recognition (with pure DL) to interpret findings in a broader diagnostic framework (with RAG). (3) It can differentiate visually similar pathologies, e.g. bacterial pneumonia vs. cardiogenic pulmonary edema, by weighing patient age, inflammatory markers, and symptomatic presentation.

Benefits: (1) Synergy of visual data & clinical knowledge, best demonstrated by annotated radiographic evidence (e.g. model identifies focal consolidation in lower lung lobe if paired with metadata such as "consolidation + fever + elevated inflammatory markers"). (2) More accuracy in assigning high probability to an infectious etiology. (3) Holistic approach, lowers diagnostic ambiguity. (4) AI-driven analysis by knowledge synthesis, i.e. evidence-based medicine.

Recent successes with RAG-based LLMs [13], [14], [15] inspire its deployment in KGML for healthcare. A vital aspect is amalgamation of contextual domain knowledge in DL. It can be achieved by knowledge base (KB) design to curate relevant domain knowledge with adequate use of commonsense knowledge (CSK).

3. KB Design and Commonsense

The design of knowledge bases is a one-time process; each designed KB can be used recurrently (to add context in a DL model) without training it on millions of data samples in each scenario [16], [17], [18] It can be highly useful in medicine. Consider the following scenario.

Scenario 2: *Automated diabetes diagnosis (XAI through KBs)*

Inputs: Documents of patients' conditions, family history, medical symptoms, e.g. "41-year old female patient has sugar level slightly above normal, she conceived at age 37, has a diabetic mother, no one else in the family is diabetic".

KB Usage to Augment DL: Extract domain-knowledge e.g. (1) Diabetes is usually inherited from paternal sides; (2) Hereditary diabetes often skips a generation; (3) Pregnancy with age>35 can cause gestation diabetes.

Benefits: (1) Adds context & explainability to diagnosis; (2) Saves resources by infusing KB info one-time (vs. making a DL model learn repeatedly from a vast number of cases), (3) Fosters automated sustainable diagnosis for widespread use, promoting equity.

Such benefits are highly desirable as cited in the literature [19]. Similar claims with XAI can be made in various fields (radiology, cardiology, psychiatry) for decision support in diagnosis, as elaborated by doctors [20], [21].

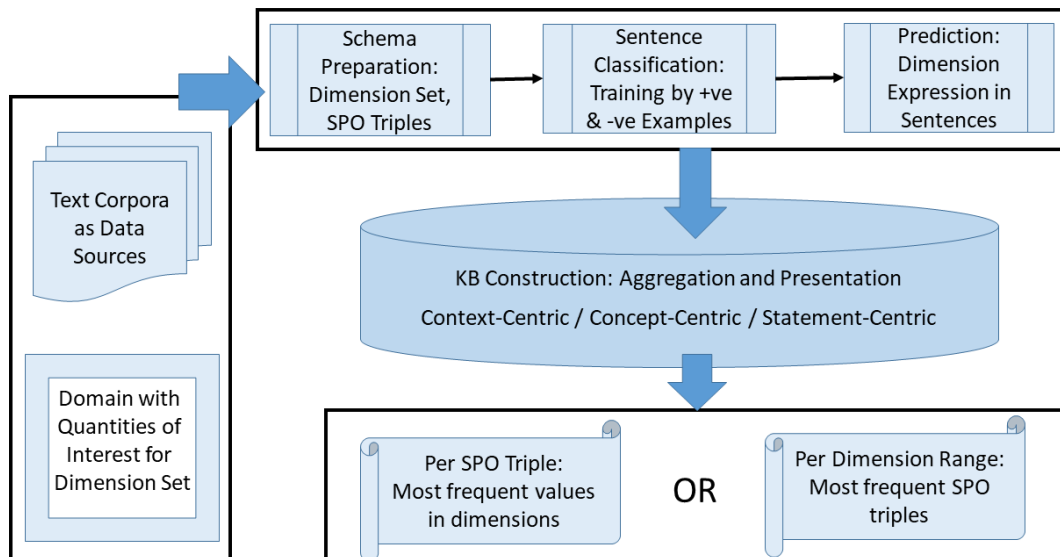


Figure 4: Proposed approach for the development of knowledge bases, incorporating natural language, quantitative data and other aspects

Another important aspect is quantitative information. Numeric values can be mapped to domain-specific terms to facilitate reasoning, e.g. *"normal body temperature"* instead of *"values around 36.9 degrees Celsius, or 98.4 degrees Fahrenheit"*. It can be helpful to reason from inputs provided (in patient data) along with corresponding medical knowledge. The main claim is that we can learn substantially by traceability from KBs, and use that knowledge for domain-specific reasoning, augmenting classification offered by DL models in decision support.

Our vision for KB design is synopsised in Fig. 4. Through medical literature study and domain expert consultation, we propose to annotate domain-specific assertions with contextual information. The resulting KB can be aggregated by multiple methods as listed next.

- Context-centric: Fix ranges on a (sub-)set of dimensions, then obtain contextualized values, e.g. consider "Oral (Mouth): 35.8–37.3° C (or 96.4–99.1° F)" as "normal" for the dimension temperature, then obtain the subset of all assertions that appeared with matching context.
- Concept-centric: Center on a concept as subject, and look at most frequent contexts, as well as statements for these contexts.
- Statement-centric: For an individual statement, focus on distributions over its contexts.

Hence, we envision KB construction as follows.

1. Define a set of dimensions e.g. temperature, blood pressure. Schema can be (S/SP/SPO): S, P, O are subject, predicate, object respectively.
2. Collect positive examples of sentences on these dimensions e.g. "Stage1 Hypertension: Systolic (130–139 mmHg), Diastolic (80–89 mmHg)".
3. Use these examples, along with random negatives, to train one sentence classifier per dimension, which predicts the dimension-expression of that sentence: 1 if the dimension is expressed, 0 if not (e.g. 1 if hypertension occurs, 0 if not).
4. At KB construction time, run each classifier on each sentence in the corpus, and whenever the dimension-expression-prediction is above some threshold (e.g. > 0.75) assume that the dimension is expressed. Then use a prompting approach to obtain the qualitative expression, respectively, grab the number in the sentence and consider it as quantitative expression.

The analysis of the intended output can then be assisted by two different approaches, stated below.

- (A) Start from an S or SP or SPO, observing the most frequent values along each dimension.

- (B) Starting with a dimension range e.g. 35.8–37.3° C, observe the most frequent S/SP/SPO. Both these approaches can be used for reasoning, and this contextual information in KBs can help to augment DL models to enhance learning.

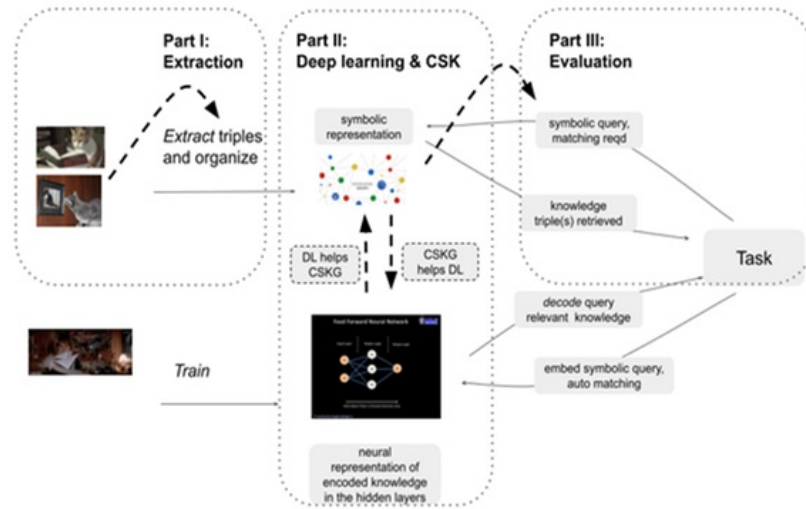


Figure 5: Commonsense knowledge (CSK) extracted, compiled, evaluated; it can be used to complement DL [22]

In addition to KBs with domain knowledge, it is important to harness everyday facts comprising commonsense knowledge (CSK), too obvious for humans but highly difficult for machines. CSK can be imperative in machine intelligence [23]. Hence, we propose to utilize CSK which can be discovered from online sources and compiled into CSKBs (commonsense knowledge bases). Methods for CSK storage and its usage to complement DL have been presented in an interesting tutorial [22]. Its crux is synopsised in Fig. 5 which is self-explanatory.

A recent study [24] leveraging the CSKB ConceptNet [25] proved highly effective in improving multipurpose robot perception, providing better results than LLM-based guidance with ChatGPT [26]. Vital aspects here included consistency and traceability, both guaranteed by CSK (not by the LLM). In line with such research, we propose that the training of medical AI systems for decision support can occur better when systems are CSK-infused.

Consider an example of a tool called CSK-Detector (commonsense knowledge based object detector) [27] (see Fig. 6) developed using XAI with CSK extracted from a CSKB (commonsense knowledge base) called DICE [28] entailing commonsense clauses of plausibility, typicality, remarkability and salience. CSK-Detector has been successfully applied to domestic robotics and can be extended to medical robotics as well. In a medical context, such a tool can be useful as indicated in the scenario next.

Scenario 3: *Object detection in medical robots with CSK*

Inputs: Images from medical domains that need task-relevant classification, e.g. photo P1 from a hospital where jewelry is on object in the room, photo P2 with a surgical table in a room.

CSK-based analysis: Clauses in CSKB e.g. (1) *Plausibility:* Object=Jewelry \Rightarrow Image \neq OperationTheater; (2) *Salience:* Object=SurgicalTable \Rightarrow Image=OperationTheater; (3) *Classification:* P1 is surely not an operation theater; P2 is highly likely to be an operation theater; (4) *Task-Relevance:* Medical robot can be fed with task-specific context to recognize relevant images.

Benefits: (1) Adds interpretability good in robot training, esp. human-robot collaboration; (2) Reduces complexity by at least 10^3 orders of magnitude (#objects: 100s vs. #images: millions) good in sustainable healthcare; (3) Reduces ambiguity, enhances trust.

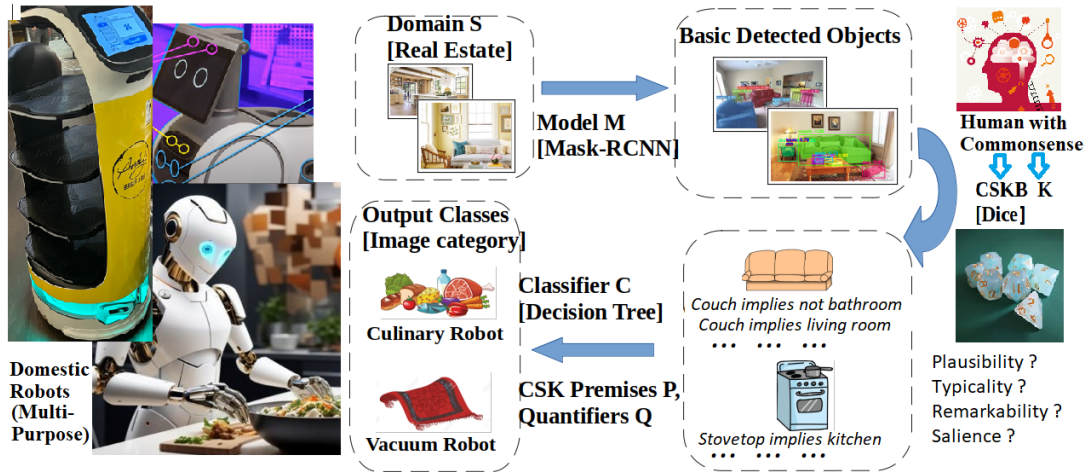


Figure 6: CSK-Detector for object-based image detection with simple commonsense facts [27] so it adds explainability to classification, and hugely saves resources as it trains on distinct individual objects (order of 100s) vs. distinct full images (order of millions); it is applied to domestic robotics here, and can be useful in medical robotics to add interpretability and enhance sustainability in healthcare.

Scenarios such as these exemplify the importance of commonsense-based classification, contributing to XAI for enhanced comprehension, as well as sustainable AI for enhanced affordability, energy savings and accessibility. Thus, we advocate the use of CSK in various aspects of sustainable healthcare including medical robot training and human-robot collaboration.

4. Transfer Learning and Simplicity

The paradigm of transfer learning entails discovering hypotheses on smaller datasets and transferring the learned hypotheses to much larger data. In adapting this paradigm to healthcare, there are many challenges, some of which are listed as follows.

- Adequate selection of reasonably-sized data spanning the gamut of the enormous sample space
- Suitable choice of models for accurate learning with low complexity
- Testing the learned hypotheses in real-life scenarios with active involvement of domain experts

In recent work [29], chest X-rays are subjected to transfer learning (with computer vision models) to classify them as COVID, pneumonia or healthy (See Fig. 7). The approach yields promising results on unseen test data from real-life datasets. Motivated by these and other success stories [30], [31], we propose to harness transfer learning by deploying multimodal models such as Gemini, CLIP, GPT-4o on heterogeneous data comprising medical literature, doctor notes, images, audio and video. This process is harder than learning from imagery alone, hence in addition to the challenges above, we face the following additional issues.

- Sufficient and holistic heterogeneous data procurement for multimodal learning
- Data transformation and data augmentation, preferring simplistic approaches
- Procurement of minimal well-grounded data samples to aid learning with maximal accuracy
- Extrinsic evaluation on multiple real-world datasets comprising various data types (with and without human-in-the-loop)

These and other issues provide the scope for further research that we propose to address as a part of the proposed work in this vision paper.

A related paradigm is exploring low-complexity solutions. In recent work [32], a 1D-CNN (convolutional neural network) interestingly yielded higher accuracy than more complex models in automated

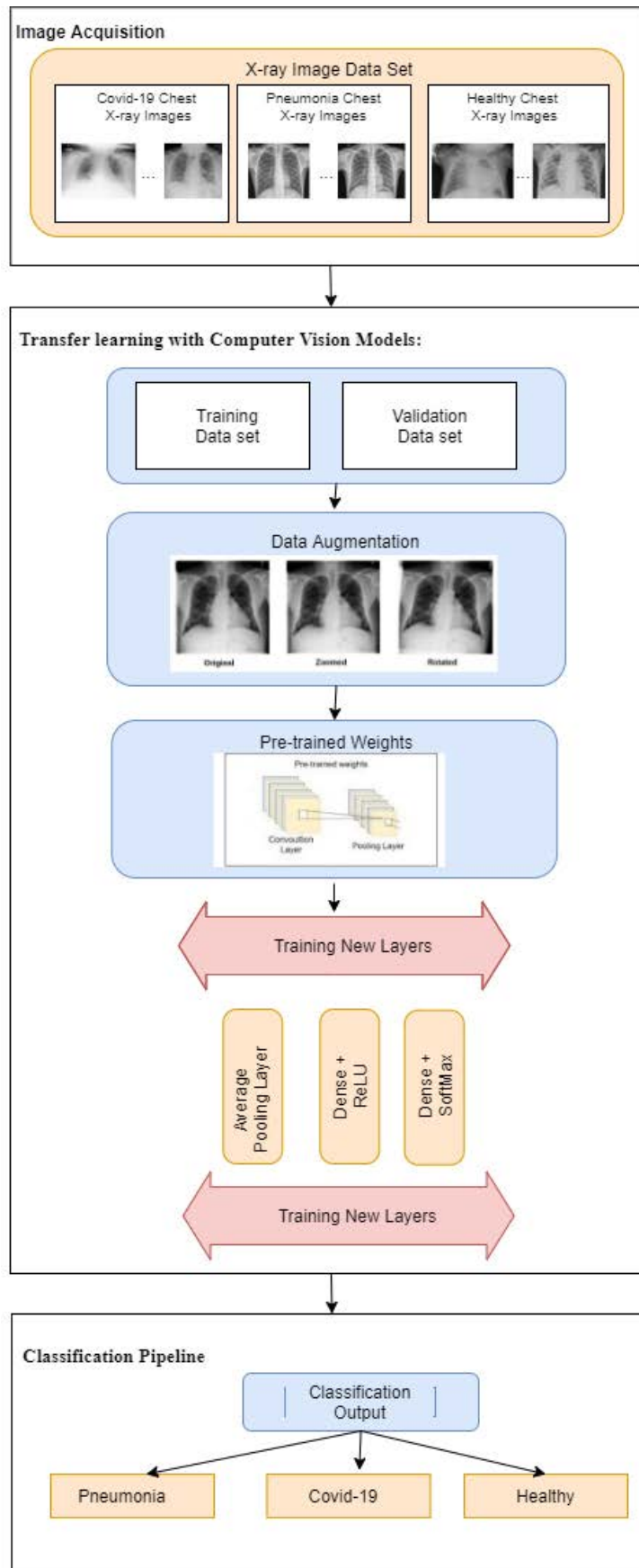


Figure 7: Transfer learning pipeline for Covid, pneumonia detection with computer vision models [29]; can be adapted to other disease diagnostics with scalability & fine-tuning, deploying state-of-the-art multimodal models on heterogeneous data

Algorithm 1 Mel-Spectrogram to Reduce Complexity

Audio signal x , Number of Mel filters M , Window size W , Hop length H , Sampling rate F_s

Mel-Spectrogram MS

1: Load audio signal x from the dataset

2: Apply pre-emphasis to the audio signal (optional): $y[n] = x[n] - \alpha \cdot x[n - 1]$

3: Divide the pre-emphasized signal into frames of window size W and hop length H

4: Foreach frame

a: Apply a window function (e.g., Hann) to the frame: $windowed_frame = window \cdot frame$

b: Compute Mel-Spectrogram: $MS = librosa.feature.melSpectrogram(y = x, sr = F_s, n_mels = M, window = window, hop_length = H)$

5: Return the Mel-Spectrogram MS

non-invasive tuberculosis detection. The study comprised audio-visual mapping of cough samples to Mel Spectrograms and numerical mapping to MFCC (Mel Frequency Cepstral Coefficients) followed by analysis with multiple DL models. A vital aspect of the methodology for low-complexity TB detection is summarized in Algorithm 1. Using MFCC (which captured domain knowledge), the complexity of the analysis was significantly reduced (as opposed to training directly on the audio-visual data). Furthermore, a simple 1D CNN surprisingly gave the best results. It can be reasoned in layperson's terms by drawing analogy with intuitive human reasoning to get the right answer (vs. making things more complex by overthinking). In computational terms, this logic matches the Occam's Razor principle of preferring simpler theories to complex ones if both offer good results, as noticed in the literature [33], [34], [35], [36], [37] in research spanning computational estimation and other works.

Likewise, Mel Spectrograms have been useful in other studies [38], [39] involving natural language data. Thriving on such research, we propose to investigate simple theories with several multimodal models to analyze heterogeneous data in healthcare. We further propose to deploy the least complex models that yield the greatest accuracy for the purpose of transfer learning.

Our proposed vision for low-complexity transfer learning entails the following steps.

1. Explore several models to discover the low complex ones yielding the high accuracy
2. Deploy transfer learning on suitable heterogeneous low-sized datasets (small data), starting with the least complex model yielding accuracy above thresholds, proceed in ascending order
3. Harness the best transfer learning model, i.e. the best combination of low complexity and high-accuracy, for transferring the learned hypotheses to larger datasets (big data)
4. Use hypotheses learned on big data for healthcare decision support e.g. disease diagnostics

This process can help to draw good inferences while also saving resources and offering low-cost solutions. It promotes widespread automated medical diagnoses, aiding UN SDGs missions in Goals 3, 10 and 11 with positive impacts on healthcare, equity and sustainability.

5. Experiments and Discussion

Initial experimentation has been conducted with datasets from healthcare and environmental domains spanning paradigms such as KGML and transfer learning in our own relevant prior work [29], [10], [32] that motivated this vision paper. We tabulate a synopsis of the experimental results in Tables 1, 2, and 3 here.

Note that KGML adds explainability with RAG-based context capturing domain knowledge (in this case, spatial masks to add context to the power plants analogous to the manner in which a domain expert would guide the classification). Similarly, medical domain experts can add context to guide classification by providing use cases that help to pinpoint a solution. Additionally, KB design along with CSK can further augment DL models by supplementing them with domain knowledge with semantics as well as more intuitive everyday facts with worldly knowledge and pragmatics. This can help make better decisions with explainability.

Category	F1 Score (Baseline)	F1 Score (KGML)
WND (Wind)	0.85	0.89
SUN (Solar)	0.81	0.88
BIT (Biomass / Coal)	0.25	0.48
NG (Natural Gas)	0.77	0.81
WAT (Hydroelectric)	0.78	0.87

Table 1

KGML to classify power plants; baseline is CNN+ViT; KGML model adds context by Spatial Masks with domain knowledge from GIS; experiments analogous to these can be run with KGML in medicine (envisioned in Section 2).

Model for TL	No. of Samples	Accuracy
VGG16	350	98%
VGG16	450	95%
VGG19	50	97%
VGG19	250	99%
VGG19	500	97%
ResNet101	250	75%
ResNet101	500	83%

Table 2

Transfer learning to classify Covid / pneumonia patients; such experiments can be run with multimodal models, esp. on least complex model with highest accuracy (envisioned in Section 4).

Approach	F1 Score	Accuracy
2D CNN on Mel Spectrogram	0.33	72%
1D CNN on Mel Spectrogram	0.66	81%
VGG16 on Mel Spectrogram	0.57	79%
ResNet50 on Mel Spectrogram	0.58	76%
2D CNN on MFCC	0.85	77%
1D CNN on MFCC	0.93	91%

Table 3

Experiments on real-life tuberculosis data, simple 1D CNN on MFCC gives best results (domain knowledge in MFCC), aligns with Occam's Razor principle preferring simpler theories; more experiments can be run on other medical data to explore if simpler models yield excellent results, they can be used for transfer learning.

Regarding transfer learning and low complexity diagnosis, further studies can be motivated thriving upon the success of our prior work. These include investigating multimodal models with transfer learning, as well as conducting studies using simplistic models to explore their effectiveness in learning hypotheses that can be used to analyze medical data. Our proposed approach of investigating the least complex models and deploying them in transfer learning (as envisioned in Section 4) is worthy of further experimentation, and constitutes part of our ongoing work. Such work can make good impacts on sustainable healthcare.

Specific inputs including use cases, proofs of concept, and example-based reasoning constitute areas of further study in this proposed work. The design of such inputs is a non-trivial task which motivates further research. It entails solid work involving active interactions with clinical practitioners, medical researchers and other healthcare professionals. Once such inputs are formulated, they can provide RAG-based context to enhance KGML. Furthermore, automated discovery of domain knowledge can be conducted to curate KBs which can then be used for further analysis and reasoning. All of these constitute contributions to the XAI work envisioned in this paper.

As relevant discussion, we include another vital comment from our interaction with real-life clinical practitioners [11], [12]: *"Recent advancements in machine learning allow for the prediction of five-year breast cancer risk through the integration of mammographic imaging and clinical variables. Given that these estimates directly influence subsequent clinical interventions, providing clear rationales for such*

predictions is vital. Transparent models enable doctors to evaluate the clinical validity of high-risk outputs, ensuring that AI serves as an interpretable aid in the decision-making process".

Both explainability and sustainability are highly desirable in next-generation AI applications [40], [41], [42], [43], [44], [45]. Our vision in this paper is to promote XAI and sustainable AI in the healthcare domain. This will assist broader missions of accessibility, affordability, and human-AI collaboration in healthcare, striving more towards achieving the much-needed global equity.

6. Conclusions and Impacts

Our proposed research in this vision paper advocates sustainable AI in healthcare, mainly by proposing solutions with XAI and transfer learning. Scientific contributions in this vision entail designing methods to balance accuracy (crucial to healthcare) versus complexity (for affordability and accessibility). Further implementation of the proposed work in our vision can pave the way to build AI tools (e.g. decision support systems and mobile apps) for highly accurate, minimally-invasive, evidence-based, energy-efficient, low-cost diagnostics.

The proposed work in our envisioned research can be particularly useful in areas where doctor-patient ratios are low, as it can help in greater affordability and accessibility. Hence, it can make significantly positive impacts on equity in healthcare. Consequently, it can contribute positively towards the United Nations missions in their Sustainability Development Goals, especially Goals 3, 10 and 11 related to Good Health and Well-Being, Reduced Inequalities, and Sustainable Cities and Communities, respectively. This work can be interesting to data scientists, AI professionals, domain experts in healthcare, as well as stakeholders who can be the potential users of the proposed methods and the software tools resulting from them. The multidisciplinary appeal of this vision paper is well-suited to fit the theme of data analytics in real-life applications.

Acknowledgments

The work presented in this article is supported by *Novo Nordisk Foundation grant NNF25OC0110035*. The authors acknowledge the help of *Sune Lundo Sorensen from SDU* during the compilation of the camera-ready paper in the required format on Overleaf. Some of the commonsense knowledge and KB design work in this paper is related to earlier research visits made by Aparna Varde at *Max Planck Institute for Informatics* at Saarbrücken, Germany.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] OpenAI, Clip: Connecting text and images, <https://openai.com/index/clip/>, 2021.
- [2] G. Team, R. Anil, S. Borgeaud, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, K. Millican, et al., Gemini: a family of highly capable multimodal models, arXiv preprint arXiv:2312.11805 (2023).
- [3] A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, A. Radford, et al., Gpt-4o system card, arXiv preprint arXiv:2410.21276 (2024).
- [4] G. Corrado, J. Barral, Advancing medical ai with med-gemini, <https://openai.com/index/clip/>, 2024.
- [5] F. Demrozi, M. Farmanbar, K. Engan, Multimodal ai (mmai) for next-generation healthcare: Data domains, algorithms, challenges, and future perspectives, *Current Opinion in Biomedical Engineering* (2025) 100632.
- [6] A. Kline, H. Wang, Y. Li, S. Dennis, M. Hutch, Z. Xu, F. Wang, F. Cheng, Y. Luo, Multimodal machine learning in precision health: A scoping review, *NPJ digital medicine* 5 (2022) 171.
- [7] K. Hao, Training a single ai model can emit as much carbon as five cars in their lifetimes, *MIT Technology Review* June (2019).
- [8] A. Karpatne, R. Kannan, V. Kumar, Knowledge guided machine learning: Accelerating discovery using scientific knowledge and data, CRC Press, 2022.
- [9] L. Liu, W. Zhou, K. Guan, B. Peng, S. Xu, J. Tang, Q. Zhu, J. Till, X. Jia, C. Jiang, et al., Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems, *Nature communications* 15 (2024) 357.

- [10] B. Austin-Gabriel, A. S. Varde, H. Liu, Geoinformatics-guided machine learning for power plant classification, AAAI Conference Bridge Program arXiv preprint arXiv:2502.01039 (2025).
- [11] OUH, Odense university hospital, <https://en.ouh.dk/>, 2026.
- [12] CAI-X, Odense university hospital, <https://cai-x.com/>, 2026.
- [13] P. Ardimento, M. L. Bernardi, M. Cimitile, Teaching uml using a rag-based llm, in: 2024 International Joint Conference on Neural Networks (IJCNN), IEEE, 2024, pp. 1–8.
- [14] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, H. Wang, H. Wang, Retrieval-augmented generation for large language models: A survey, arXiv preprint arXiv:2312.10997 2 (2023).
- [15] V. Zolfaghari, N. Petrovic, F. Pan, K. Lebioda, A. Knoll, Adopting rag for llm-aided future vehicle design, in: 2024 2nd International Conference on Foundation and Large Language Models (FLLM), IEEE, 2024, pp. 437–442.
- [16] A. Varde, N. Tandon, S. Nag-Chowdury, G. Weikum, Common sense knowledge in domain-specific knowledge bases, Technical Report, Max Planck Institute for Informatics, 2015.
- [17] A. Garg, N. Tandon, A. S. Varde, I am guessing you can't recognize this: Generating adversarial images for object detection using spatial commonsense, in: AAAI conference, 2020, pp. 13789–13790.
- [18] A. Varde, S. Razniewski, T.-P. Nguyen, G. Weikum, Data Mining on Environmental Quantities with Commonsense Reasoning, Technical Report, Max Planck Institute for Informatics, 2021.
- [19] I. FASTERHOLDT, M. GERSTRØM, B. S. B. RASMUSSEN, K. B. YDERSTRÆDE, K. KIDHOLM, K. M. PEDERSEN, Cost-effectiveness of telemonitoring of diabetic foot ulcer patients, *Health Informatics Journal* 24 (2018) 245–258.
- [20] M. Mottaqi, P. Zhang, L. Xie, Integrating interpretable machine learning and multi-omics systems biology for personalized biomarker discovery and drug repurposing in alzheimer's disease, *bioRxiv* (2025) 2025–03.
- [21] J. A. Bojsen, M. T. Elhakim, O. Graumann, D. Gaist, M. Nielsen, F. S. G. Harbo, C. H. Krag, M. V. Sagar, C. Kruuse, M. P. Boesen, et al., Artificial intelligence for mri stroke detection: a systematic review and meta-analysis, *Insights into Imaging* 15 (2024) 160.
- [22] S. Razniewski, N. Tandon, A. S. Varde, Information to wisdom: Commonsense knowledge extraction and compilation, in: Proceedings of the 14th ACM international conference on web search and data mining, 2021, pp. 1143–1146.
- [23] N. Tandon, A. S. Varde, G. de Melo, Commonsense knowledge in machine intelligence, *ACM SIGMOD Record* 46 (2018) 49–52.
- [24] R. Hidalgo, A. S. Varde, J. Parron, W. Wang, Incorporating commonsense knowledge to enhance robot perception, *IEEE Transactions on Automation Science and Engineering* 22 (2025) 15488–15501. URL: <https://doi.org/10.1109/TASE.2025.3565191>.
- [25] R. Speer, J. Chin, C. Havasi, Conceptnet 5.5: An open multilingual graph of general knowledge, in: Proceedings of the AAAI conference on artificial intelligence, volume 31, 2017.
- [26] OpenAI, Chatgpt, <https://chatgpt.com/>, 2022.
- [27] I. Chernyavsky, A. S. Varde, S. Razniewski, Csk-detector: Commonsense in object detection, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE, 2022, pp. 6609–6612.
- [28] Y. Chalier, S. Razniewski, G. Weikum, Dice: A joint reasoning framework for multi-faceted commonsense knowledge., in: ISWC (Demos/Industry), 2020, pp. 16–20.
- [29] A. S. Varde, D. Karthikeyan, W. Wang, Facilitating covid recognition from x-rays with computer vision models and transfer learning, *Multimedia Tools and Applications* 83 (2024) 807–838.
- [30] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, A comprehensive survey on transfer learning, *Proceedings of the IEEE* 109 (2020) 43–76.
- [31] K. Weiss, T. M. Khoshgoftaar, D. Wang, A survey of transfer learning, *Journal of Big data* 3 (2016) 9.
- [32] J. Yadav, A. S. Varde, H. Liu, G. Antoniou, L. Xie, Audiovisual multimodal cough data analysis for tuberculosis detection, in: 2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA), IEEE, 2024, pp. 1–8.
- [33] A. S. Varde, Computational estimation by scientific data mining with classical methods to automate learning strategies of scientists, *ACM Transactions on Knowledge Discovery from Data (TKDD)* 16 (2022) 1–52.
- [34] J. McFadden, Razor sharp: The role of occam's razor in science, *Annals of the New York Academy of Sciences* 1530 (2023) 8–17.
- [35] A. S. Varde, E. A. Rundensteiner, C. Ruiz, D. C. Brown, M. Maniruzzaman, R. D. Sisson, Designing semantics-preserving cluster representatives for scientific input conditions, in: Proceedings of the 15th ACM international conference on Information and knowledge management, 2006, pp. 708–717.
- [36] A. S. Varde, E. A. Rundensteiner, C. Ruiz, D. C. Brown, M. Maniruzzaman, Effectiveness of domain-specific cluster representatives for graphical plots, in: ACM SIGMOD (IQIS workshop), 2006.
- [37] T. Blanchard, T. Lombrozo, S. Nichols, Bayesian occam's razor is a razor of the people, *Cognitive science* 42 (2018) 1345–1359.
- [38] N. Shao, R. Zhou, P. Wang, X. Li, Y. Fang, Y. Yang, X. Li, Cleanmel: Mel-spectrogram enhancement for improving both speech quality and asr, arXiv preprint arXiv:2502.20040 (2025).
- [39] C. Fan, S. Zhang, J. Zhang, E. Liu, X. Li, G. Zhao, Z. Lv, Dmf2mel: A dynamic multiscale fusion network for eeg-driven mel spectrogram reconstruction, in: Proceedings of the 33rd ACM International Conference on Multimedia, 2025, pp. 6977–6985.
- [40] P. Basavaraju, A. S. Varde, Supervised learning techniques in mobile device apps for androids, *ACM SIGKDD Explorations* 18 (2017) 18–29.
- [41] L. Paulino, C. Hannum, A. S. Varde, C. J. Conti, Search methods in motion planning for mobile robots, in: Proceedings of SAI Intelligent Systems Conference, Springer, 2021, pp. 802–822.

- [42] K. Gandhe, A. S. Varde, X. Du, Sentiment analysis of twitter data with hybrid learning for recommender applications, in: 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE, 2018, pp. 57–63.
- [43] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, et al., Explainable ai (xai): Core ideas, techniques, and solutions, *ACM computing surveys* 55 (2023) 1–33.
- [44] D. Lakshmi, R. S. Tiwari, R. K. Dhanaraj, S. Kadry, *Explainable AI (XAI) for sustainable development: Trends and applications*, CRC Press, 2024.
- [45] A. Van Wynsberghe, Sustainable ai: Ai for sustainability and the sustainability of ai, *AI and Ethics* 1 (2021) 213–218.