Collaborative Intelligent Decision Systems for Safe and Reliable AI-assisted Medical Image Diagnostics

Serge Dolgikh, PhD
National Aviation University,
Kyiv Ukraine

Abstract

The cost of diagnostic errors has been high in the developed world economics according to a number of recent studies and continues to rise. Up till now, a common process of performing image diagnostics for a growing number of conditions has been examination by a single human specialist (i.e., single-channel recognition and classification decision system). Such a system has natural limitations of unmitigated error that can be detected only much later in the treatment cycle, as well as resource intensity and poor ability to scale to the rising demand. At the same time Machine Intelligence (ML, AI) systems, specifically those including deep neural network and large visual domain models have made significant progress in the field of general image recognition, in many instances achieving the level of an average human and in a growing number of cases, a human specialist in the effectiveness of image recognition tasks. The objectives of the AI in Medicine (AIM) program were set to leverage the opportunities and advantages of the rapidly evolving Artificial Intelligence technology to achieve real and measurable gains in public healthcare, in quality, access, public confidence and cost efficiency. The proposal for a collaborative AI-human image diagnostics system falls directly into the scope of this program.

Keywords: Image diagnostics; Machine learning; transfer learning; collaborative Human-AI systems; intelligent decision systems; AIM.

Introduction

The cost of diagnostic errors has been high in the developed world economics according to a number of recent studies and continues to rise [1]. Up till now, a common process of performing image diagnostics for a growing number of conditions has been examination by a single human specialist (i.e., single-channel recognition and classification decision system). Such a system has natural limitations of unmitigated error that can be detected only much later in the treatment cycle, as well as resource intensity and poor ability to scale to the rising demand [2]. At the same time Machine Intelligence (ML, AI) systems, specifically those including deep neural network and large visual domain models have made significant progress in the field of classification and recognition of complex data, in many instances achieving the level of an average human and in a growing number of cases, a human specialist in the effectiveness of image recognition tasks [3,4].

Machine Intelligence models and systems (ML, AI) have a number of essential strengths, such as: stability and resilience with respect to the environmental factors and influences; superior operating performance in both time and volume; shorter training time and time to operation; accuracy in execution of intelligent tasks approaching and in a growing number of applications surpassing that of a human specialist; cost efficiency in operation. In a number of applications, ML/AI systems demonstrated the ability to identify characteristic types or “concepts” in complex realistic data [5,6] that can be instrumental in the analysis and description of its information structure.

On the other hand, integration of machine intelligence models in essential and critical public interest applications which include public health care is less than straightforward and can be impeded by insufficient understanding of the processes that lead to their decisions (explainability and trust challenges,

https://doi.org/10.33774/coe-2024-s7tc7 **ORCID:** https://orcid.org/0000-0001-5929-8954

[7,8]), preparedness of the public and general public trust, dependence on large bodies of trusted data describing the domain in sufficient detail and others.

A promising avenue in harnessing the power and the potential of machine intelligence in public interest applications has been developed over the years in the area of collaborative intelligent systems [9,10]. An inherent promise of this approach is the potential to use the machine and human intelligences in a collaborative process that used the respective strengths of each type while mitigating their downsides. It was shown that by designing decision systems in such a way that both human and machine components could contribute to the success of the resulting decision, a synergetic effect can be achieved with significant improvement in the accuracy, and as a consequence, a noticeable reduction in the diagnostic error and the associated with it cost [11].

The objectives of the AI in Medicine (AIM) program [12] were set to leverage the opportunities and advantages of the rapidly evolving Artificial Intelligence technology to achieve real and measurable gains in public healthcare, in quality, access, public confidence and cost efficiency. The proposal falls directly into the scope of this program.

Drivers of AI Integration in Public Health Care

The justification for the research in collaborative AI-human decision systems, as was briefly mentioned previously, is based on *four drivers* creating an opportunity for a successful integration of Machine Intelligence technology in the tasks of image diagnostics in the public healthcare system, with the potential to improve, measurably and significantly, the accuracy and productivity of the diagnostics tasks and processes.

Cost and Resources: rising cost of diagnostic errors and constraints on available resources in public healthcare systems.

Technology: recent advances in Machine Intelligence technology, bringing the level of image recognition success to that of a human.

Complementarity: complementary, mutually contributing and complementing nature of human and machine intelligences creates an opportunity for cooperative and collaborative work process with improved outcomes in the targeted tasks.

Trust and safety: the solution has to have uncompromised safety and full human control over the resulting decision.

The drivers and opportunities associated with them set the foundation for a collaborative, joint decision framework of human and machine intelligences that can use the strengths of either while mutually compensating for the limitations and shortcomings, to improve the outcome of the collective decision in image diagnostic tasks. The potential and the window of opportunity for the development and integration of collaborative AI-human intelligent decision systems is illustrated in **Figure 1**.

Figure 1. Drivers for the advent of synergetic collaborative AI-human decision systems.

The drivers described above lay the ground and provide the incentive for research into intelligent systems that could harness the strong sides of either type of intelligence to produce superior outcome, within the framework of expectations and requirements, to the use of either system on its own.

Complementarity and Synergetic Potential of Collaborative Intelligent Systems

A collaborative intelligent decision system outlined in this article is based on the observation that due to the complementary nature of the human and machine intelligences they may not be expected to make “many” mistakes in the same situations and cases; and as a consequence, an opportunity emerges for the creation of synergetic intelligent systems where human and machine channels would be able to complement and correct each other.

For example, a human practitioner can be tired, stressed or temporarily distracted [13] whereas the Machine Intelligence component of the system would not be affected by these factors. On the other hand, a machine system can make spurious classification mistakes, “hallucinations” [14] that are easily detected by a human specialist.

These and other constituent parts of the complementarity of the human and machine intelligences are based on the fact that, while being capable of achieving high degree of accuracy in recognition tasks, humans and machine systems learn differently, with different data, in quite different ways and processes, and so on, as described in the **Figure 2** below.

https://doi.org/10.33774/coe-2024-s7tc7 **ORCID:** https://orcid.org/0000-0001-5929-8954

Intelligence	Training	Operation
Human	▪ Focus on generalization ▪ Smaller number of cases: patterns and associations, generalization ▪ Longer time to competence (learning curve) ▪ Consultation and collaboration	▪ Influenced by environment and personal factors ▪ Influenced by learning quality, competence ▪ Training through experience ▪ <u>Incremental approach to solution</u> ▪ Association and creativity
Machine Intelligence (conventional models)	• Needs massive sets of annotated cases • Depends on high confidence annotations (trusted prior knowledge) • Learning via informative (differentiating) features • Shorter training time	• Stable performance with respect to environment and individual factors • Cannot easily correct wrong decision • Explainability challenge • Limited creativity

Figure 2. Essential differences between human and machine intelligences.

Based on the observations discussed above, obtaining the decision inputs of both human practitioner and a machine intelligence system (Automated Recognition System) and combining them in producing the diagnostic decision can help in detecting possible errors of the either component, and improve, to a significant extent according to the published research [11], the outcome of the collaborative decision.

To this end, in the proposed approach, a collaborative AI-human intelligent system with parallel decision channels is envisioned, that is expected to take full advantage of the strengths of the human and machine intelligences while mitigating the chance of error in a combined effort to achieve the best diagnostic outcome. This is achieved by a multi-channel concurrent intelligent decision process that mitigates the probability of a decision error to the second power of the probability of error in a conventional diagnostic workflow.

This solution is illustrated in a possible architecture of a collaborative decision system with parallel human and machine channels that make independent decisions on the provided inputs, for an example in our case, diagnostic images. Considering the diagram of the high-level architecture of such a system illustrated in **Figure 3**, one can observe that the following scenarios are possible in the first, concurrent phase of the collaborative system:

A. Both human specialist and the machine system produce an erroneous decision. It could be a complex or rare case, where could struggle and err simultaneously or a random coincidence of errors of the components (channels). In that case, one can expect that due to complementary nature of human and machine intelligences discussed above, this possibility would be suppressed as roughly the multiple of the failure (i.e., error) rates of the components. This case would produce a wrong diagnostic decision, and a diagnostic error.
B. The channels agree on a correct decision. The decision is accepted in as the final one with no need for second opinion or verification.
C. This scenario can transpire if the initial decisions of the channels "disagree"; then the final decision is referred to a human expert in the diagnostic area. This case can produce an erroneous result only if the expert and one of the channels err simultaneously (i.e., on the same case, input); then, the error of the final decision will be strongly suppressed in this case as well.

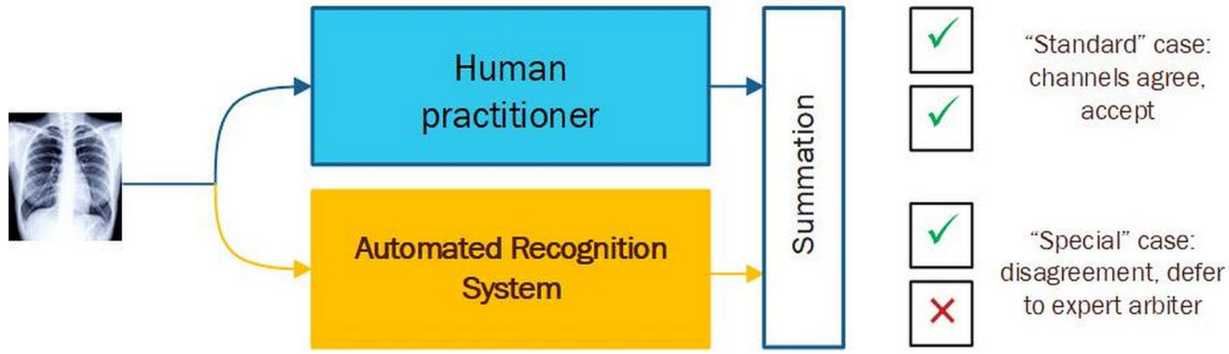

Figure 3. Collaborative AI-human intelligent decision system with concurrent channels.

As can be concluded, in this case an error of the final diagnostic decision <u>will be suppressed at least to the second power of the characteristic error of an individual channel</u>. The benefit of the collaborative process comes afore most in the case of mixed decisions, whereby erroneous decisions by individual channels are "caught" by the concurrent channel and referred to an expert arbiter in the area/condition (**Figure 4**).

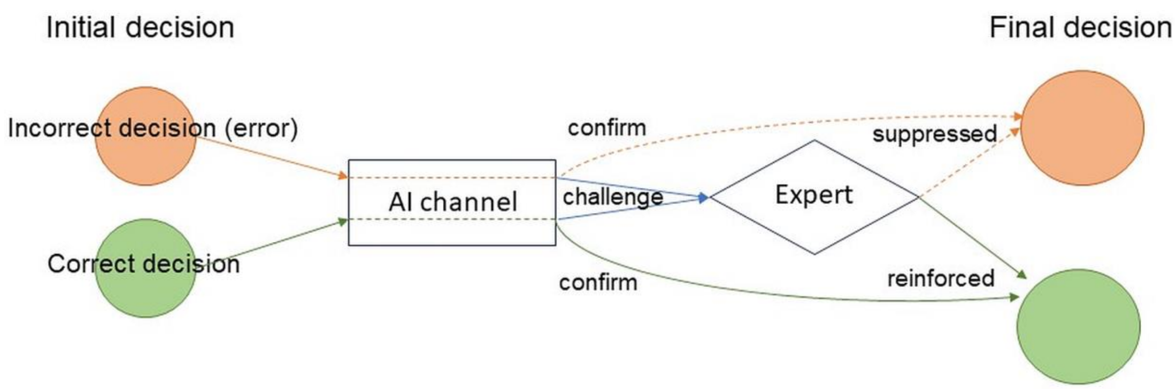

Figure 4. Reducing diagnostic error via intelligent collaboration of concurrent decision channels.

The analysis above demonstrates that integration of a concurrent machine intelligence system can indeed improve the outcome of essential and critical decision via a mechanism of suppression of routine errors and reinforcement of correct decisions.

In contrast, in the conventional "single link" diagnostic workflow, the possibility of an erroneous diagnostic decision is not suppressed by any mechanism and it may not be detected until much later in the treatment cycle [15].

The proposed system would have a potential to eliminate many or most of routine errors that can account for majority of diagnostic errors and can have the following advantages, compared to conventional "single chain" diagnostics workflows.

- Quality: significant improvement in overall accuracy of diagnostics decisions and associated reduction in the follow up spending in the system, improved patient care and overall quality.
- Performance: improves throughput of the diagnostic tasks through the system; balanced and scalable operational model, fully compatible with distributed, high performance and outstanding quality operational models of public service delivery.
- Stimulate optimal use of the expert resources only in those cases that require their attention and expertise.
- Safety and trust: retained full human control over the diagnostic decision.

https://doi.org/10.33774/coe-2024-s7tc7 **ORCID:** https://orcid.org/0000-0001-5929-8954

- Cost efficiency: small cost of development, deployment and operation compared to massive cost saving in the health system to due reduced error and improved outcomes.
- Flexibility: the solution is adaptable to different conditions and types of input data and can be transferred to different areas of integration of Machine Intelligence in medical applications, as well as collaborative decision systems in other areas of application.

The potential for improvement described above can be attributed to the emerging opportunity to combine the strengths and advantages of human and machine intelligences for a significant improvement in the quality of diagnostic decisions over the current practice, while retaining complete and uncompromised human control over the process of diagnostics and treatment.

Moreover, another important advantage of the proposed collaborative system is the potential for continuous improvement. Indeed, cases that resulted in eventual diagnostic errors of all discussed types can be reviewed by the experts in the diagnostic area/condition and integrated into the training processes for human specialists and machine systems. For example, expanding training sets with new samples, drawing attention to certain cases in instruction of practitioners) to attain further gain of quality in each new iteration of the system.

Thus, not only the proposed system can be expected to achieve a substantial improvement in quality at the time of release; but it can be integrated naturally into a lasting process of continuous iterative quality improvement over the lifecycle of the solution.

A combination of high-performance diagnostic AI models with intelligent multi-channel decision system incorporating human in the loop for maximum safety in the operational practice can produce, according to the published results [11] a significant improvement in the accuracy of image diagnostic, and correspondingly significantly reduce the negative impacts and cost of misdiagnosis in the evaluated conditions and areas of public healthcare.

Further Opportunities for Integration of AI in Image Diagnostic

A practical approach to planning of large-scale integration of collaborative AI-assisted decision systems and models in the practice of diagnostics can include a blueprint proposal for a regional or multi-national anonymous database of diagnostic images that can be used in the research, including development of descriptive core image generative and domain models that can be adapted to specific diagnostic conditions and requirements with minimal effort and lead time. One of many such opportunities can be based on the well-researched practice of transfer learning in the area of image recognition. Based on availability of dataset(s) of representative samples and the verified methods in transfer learning, effective diagnostic systems for a broad range of conditions can be developed in in a short time, based on established and verified general framework.

Another promising approach being widely developed these days is the development of specific large domain models in diagnostic imaging. Due to limitations of the format, these research directions and approaches will be discussed in more detail elsewhere.

Practical Implementation of Collaborative Decision Systems

Projects in practical development of collaborative intelligent decision systems can comprise the following key phases and activities.

- A comprehensive review of the current state of the art in image processing, classification and recognition including in the domain of medical image diagnostics and specific medical conditions.

https://doi.org/10.33774/coe-2024-s7tc7 **ORCID:** https://orcid.org/0000-0001-5929-8954 Content not peer-reviewed by Cambridge University Press. **License:** All Rights Reserved

- Research, review, collection, compilation and acquisition of sufficient, by size, representativity, etc., datasets of diagnostic images.
- Research and development of image recognition models, generative and large domain models, preprocessing and other methods for the AI-based component of collaborative models.
- Prototype implementation of the collaborative decision system with realistic operational characteristics.
- Verification, corrections, adjustment and optimization of the collaborative decision system.
- Reviews, information exchanges, presentations, demonstrations and discussions with the practitioner and expert specialists in the field.
- Test deployments, collection of feedback, further improvements, tuning and optimizations for mass-scale integration.
- Integration of the tracking information processes and continuous improvement processes, procedures and policies.

Conclusions

To tackle the actual and increasingly pressing challenge of the volume, cost and quality of diagnostic decisions in modern public health systems, integration of Machine Intelligence appears to be an obvious direction to the solution. As we discussed in this work, the perception of simplicity in this program can be misguided and lead to unexpected and unwanted consequences.

To avoid them, we first formulated essential expectations and requirements for intelligent systems with integrated AI components, in both technical and social domains, including, importantly, critical matters of explainability and trust.

An approach to development of collaborative human-AI intelligent decision systems discussed here proposes a framework for collaborative decision process that taps into the strong side of each type of intelligence while mitigating their respective downsides and weaknesses. As a result, a noticeable improvement in the outcome, measured by the overall diagnostic error can be expected in the domains and conditions where effective integration of machine intelligence is possible and warranted. An additional positive effect of the discussed framework and architecture is the complete traceability of the process that allows effective and straightforward initiation of continuous improvement feedback loops.

The authors expect that the findings and the discussion presented in this work will be of benefit to the research and general community and will contribute to the program of development of performant, accurate, transparent and responsible collaborative intelligent systems for the ultimate benefit of the society.

REFERENCES

1. Sheikh A., Donaldson L., Westfall-Bates D. et.al. (2016) Diagnostics errors. WHO Technical series on safer primary care. Available online: https://apps.who.int/iris/bitstream/handle/10665/252410/9789241511636-eng.pdf (accessed September, 2019).
2. Kondro, W. (2004). Canadian report quantifies cost of medical errors, Lancet 363 (9426) 2059.
3. Liu, X., Faes L., Kale A et al. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis, Lancet 1 (6) 271-297.

https://doi.org/10.33774/coe-2024-s7tc7 **ORCID:** https://orcid.org/0000-0001-5929-8954 Content not peer-reviewed by Cambridge University Press. **License:** All Rights Reserved

4. The Guardian (2019). AI equal with human experts in medical diagnosis, study finds | Artificial intelligence (AI). The Guardian News & Media, London, United Kingdom. Available online: https://www.theguardian.com/technology/2019/sep/24/ai-equal-with-human-experts-in-medical-diagnosis-study-finds (accessed on April 2021).
5. Rodriguez, R.C., Alaniz, S., Akata, Z. (2019). Modeling conceptual understanding in image reference games. In: Advances in Neural Information Processing Systems, Vancouver, Canada 13155–13165.
6. Dolgikh, S. (2021). Low-dimensional representations in unsupervised generative models. In: 20th International Conference Information Technologies - Applications and Theory (ITAT 2020), Slovakia CEUR-WS.org 2718, pp. 239–245.
7. Bartneck, C., Lütge, C., Wagner, A., Welsh, S. (2020). Trust and fairness in AI systems. In: An Introduction to Ethics in Robotics and AI, SpringerBriefs in Ethics. Springer, Cham.
8. Bogen, M. (2019). All the ways hiring algorithms can introduce bias, Harvard Business Review, 06.05.2019.
9. Wang, D., Churchill, E., Maes, P., Fan, X., Schneiderman, B. et al. (2020). From Human-Human collaboration to Human-AI collaboration: designing AI systems that can work together with people. In: CHI Conference on Human Factors in Computing Systems (CHI '20) 1-6.
10. Papadopoulos, G.T., Antona, M., Stephanidis, C. (2021). Towards open and expandable cognitive AI architectures for large-scale multi-agent human-robot collaborative learning. In: IEEE Access, 9, 73890-73909.
11. Dolgikh, S. (2021) *A Collaborative Model for Integration of Artificial Intelligence in Primary Care*. Journal of Human, Earth and Future, Elsevier, 2 (4).
12. Kulikowski, C.A. (2019) *Beginnings of Artificial Intelligence in Medicine (AIM): Computational Artifice Assisting Scientific Inquiry and Clinical Art - with Reflections on Present AIM Challenges*. Yearbook of Medical Informatics 28 (1) :249-256.
13. Sidorov, P. (2008). The burnout syndrome in communicative professional workers. Gigiena i Sanitariia 3 29-33.
14. Ji, Z., Lee, N., Frieske, R., Yu, T.S., et al. (2022). Survey of hallucination in Natural Language Generation. ACM Computing Surveys.
15. Costopoulou O., Delaney B., Munro C. Diagnostic difficulty and error in primary care – a systematic review. Oxford University Press, 2008.