A Robust Model for Integration of Artificial Intelligence Methods in Primary Care

Serge Dolgikh[1][0000-0001-5929-8954]

[1] National Aviation University, Komarova Ave., 1, 03058, Kiev, Ukraine
Solana Networks, 301 Moodie Dr., Ottawa, K2H9C4, Canada
serged.7@gmail.com

Abstract. The cost of patient care is rapidly increasing in the developed world and improving accuracy of screening and diagnostic testing as well as other areas of primary care can provide noticeable improvements in the recovery and cost efficiency of the health care systems. In this study the authors propose a simple yet robust model of parallel decision making incorporating machine and human expert competences whereby the strengths and advantages of Artificial Intelligence methods can be harnessed to improve the overall accuracy of essential testing, diagnostics and other critical areas of patient care while ensuring safety and complete human control over the course of diagnostics and treatment.

Keywords: Artificial Intelligence, Diagnostics, Primary care, Decision making models.

1 Introduction

The cost of primary care is rapidly increasing in the developed world and the accuracy of screening and diagnostic testing is one of the essential factors in the overall cost of health care systems. The cost of misdiagnosing can be significant both in the case of undetected serious condition resulting in prolonged recovery and higher cost of treatment, as in the case of a false positive diagnosis leading to higher cost of subsequent testing and possible emotional impact on the patient and their family. Directly on the cost of direct consequences of misdiagnosis, "1 million added days in hospital and $750 million in extra health-care spending" may be attributable to medical errors by doctors, hospitals, and pharmacists", according to the Canadian Institute for Health Information's (CIHI) examination of patient safety in Canada [1], while "improving patient safety in US Medicare hospitals estimated to have saved US $28 billion" [2]. High cost of diagnostics errors to the patients as well as to the primary care system was highlighted in the World Health Organisation's Technical Series on Safer Primary Care report on diagnostics errors [3].

The causes of misdiagnosis are complex and while no perfect or simple solution has been found for this serious problem, it is clear that personal and environmental influences on the human operators in the field is one of the contributing factors. It is

2019 IDDM Workshops.

well known that the performance of even professional and highly trained personnel may vary in time and be affected by multiple factors such as physical condition, mood, fatigue, stress and others. In particular, the burnout syndrome is well known among professionals whose work involves conditions of high and constant stress, responsibility for life and well-being of other people such as military personnel, pilots, medical professionals, teachers, social workers [4].

On the other hand, the advances in the field of Artificial Intelligence technologies over the past decades have brought the performance of machine systems in some areas to the level of human experts, including in health care applications [5,6]. Unlike human practitioners, machine systems offer performance on a stable level not affected by personal and transient factors. These developments offer opportunities to significantly improve the performance of essential diagnostics practices and procedures via incorporation of pre-trained in the diagnostic area high performance machine intelligence systems.

However, the introduction of such complex systems in direct human care can bring serious challenges of their own, particularly in the areas of trust and confidence in the system that employs such components: the internal operation of complex machine systems such as deep learning neural networks used in high accuracy image analysis is not very well understood at the time of writing and trusting them with an essential treatment decision can be seen as premature at this point, and less than clear if achievable in the longer perspective. Quoting Dr Raj Jena at Addenbrooke's hospital in Cambridge "if you are a deep learning algorithm, when you fail you can often fail in a very unpredictable and spectacular way", stressing that applications of machine intelligence systems will need robust real-world testing [6,7].

Taking into account these challenges and opportunities, the authors undertook the study to investigate possibilities of safe and efficient introduction of Artificial Intelligence methods in the operational practices of primary care and proposed a simple yet robust model whereby high accuracy machine methods can be harnessed to improve the accuracy of essential testing, diagnostics and other critical areas of health care without any compromised of safety, trust and confidence in the system.

The motivation for this study is:

- to investigate opportunities and models of incorporating high performance Artificial Intelligence methods into the diagnostics practices to improve the accuracy and cost efficiency of essential diagnostics without compromising safety, trust and confidence in the system, and
- to propose a general approach to incorporating machine intelligence systems with the potential to measurably improve the diagnostics outcomes while complying with the requirements of safety and full human control over the processes of diagnostics and treatment.

2 Background: Challenges and Shortcomings of the Current Practice

In many health care systems and institutions, both private and public, the diagnostics following an essential test is performed by a single human practitioner and passed on to the next stage in the patient care chain that often takes it as a given with no further feedback or analysis [1,3]. This practice may create a single link chain model in which the accuracy of the entire chain is dependent and determined by that of a single link, with correct diagnostics playing primary and sometimes critical role in the outcome of the treatment.

The logical consequence of this observation is that the efficiency of the chain cannot exceed that of any single link, and the error rate in the diagnostics phase may drive down the overall efficiency, both in terms of the patient outcome and the cost to the system, of the entire chain.

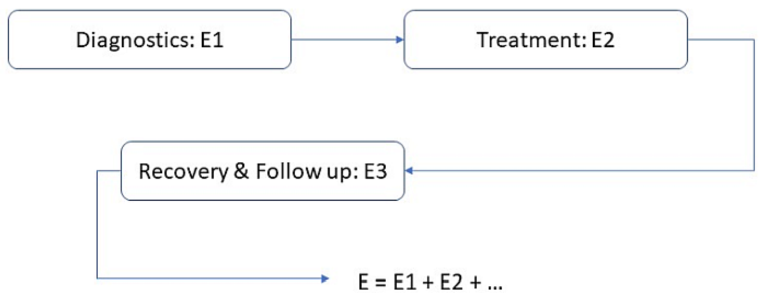

Fig. 1. A single-link chain diagnostics and treatment model

On the other hand, the ability to reduce the incidence of essential errors is limited by the factors of human nature that is essentially impacted by the condition and the environment; and the cost and resource limitations in the system that do not allow significant duplication of processes to reduce the overall error. For example, to reduce the single link error, the system would need a second opinion on every diagnostics test or decision, resulting in the doubling of the cost of the diagnostics system, the direction that is rarely acceptable.

The advances in machine intelligence methods and systems over the last decade can offer an avenue toward a solution of this complex and costly problem as the cost of deploying a pre-trained in a specific diagnostics area high accuracy and high performance machine intelligence component can be negligible compared to educating and hiring hundreds of human practitioners, and its performance is more stable and not affected as much by internal or environment factors.

However, as mentioned earlier, any such development must be cautious and deal with the issues of trust and confidence in machine based decision-making systems that at this time cannot be taken for granted [6]. The challenge therefore lies in creating combined, hybrid human-machine expertise decision-making models that would be able to combine the benefits of accuracy, high performance and stability offered by machine systems with trust and confidence of complete and uncompromised human control over the outcome of the diagnostics and treatment. Such an approach is investigated and proposed in this study.

3 Multi-Channel Parallel Decision-Making Model

3.1 Decision Functions: Cumulative and Conflict

Let's suppose a decision-making system has multiple decision making channels C_1, .. C_n and the final decision will be obtained from the partial decisions of the channels by a certain summation process that can be described by a "cumulative function" taking as input the partial decisions of the channels and producing the final decision:

$$D = S(C_1, \dots C_n)$$

In the simplest case, the channel decisions can have Boolean value of True (condition detected) or False (normal, no condition) and one of the simplest forms of the cumulative function could be the logical OR of the channel decisions:

$$S(C_1,..C_n) = OR\ (C_1, ..\ C_n)$$

Similarly to the cumulative function, the "conflict function" is defined as another perspective on the cumulative set of the decisions of the channels, that in the simplest form can be defined as the logical sum of pair-wise comparisons of the channel decisions:

$$X(C_1, ..C_n) = OR((C_1 == C_2), (C_1 == C_3)\ ..)$$

Thus, the meaning of the cumulative function is: "at least one channel has detected the condition" and that of the conflict function, "there's at least one conflict between the decisions of the channels". These definitions are summarized in Table 1.

Table 1. Cumulative and Conflict decision functions.

Channel 1	Channel 2	Cumulative, S	Conflict, X
True	True	True	False
True	False	True	True
False	True	True	True
False	False	False	False

3.2 Accuracy

In the next step the accuracy of the channels and how it affects the accuracy of the system as a whole will be analyzed. Suppose the mean accuracies of the two channels are A_1 and A_2, respectively. It easily follows from the definitions of cumulative and conflict functions above that the probabilities of an agreement (no conflict) and a conflict of the channels under that assumption will be as follows:

$$P_{agr} = A_1 \times A_2 + (1-A_1) \times (1-A_2) = 1 + 2\, A_1 \times A_2 - (A_1+A_2)$$

$$P_{conf} = A_1 \times (1-A_2) + A_2 \times (1-A_1) = A_1 + A_2 - 2\, A_1 \times A_2, \qquad (1)$$

and obviously,

$$P_{conf} = 1 - P_{agr}$$

We shall now introduce into the model the third channel, sequential to the parallel channels C_1 and C_2 that takes the input of the channels as well as values of S and X and produces the final decision:

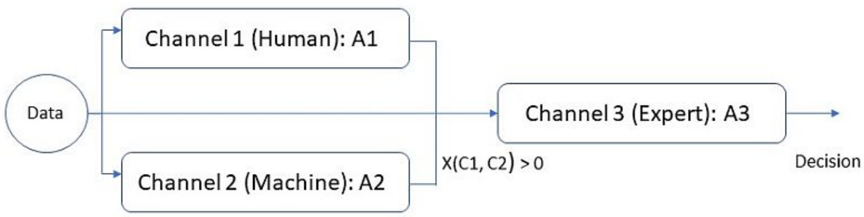

Fig.2. A multi-channel parallel system with expert channel

The further constraint that will be imposed in this model is that the final "expert" channel will be involved only in the case of a conflict between the parallel channels, that is, if $X(C_1, C_2)$ = True. It will also be assumed that the accuracy of the expert channel $A_3 > A_1, A_2$.

From (1) the probability of the correct decision of the expert channel can then be calculated as:

$$P_{exp} = P_{conf} \times A_3 \qquad (2)$$

and the error of the expert channel, as:

$$E_{exp} = P_{conf} \times (1-A_3)$$

Finally, from (1) and (2) one can estimate the overall error in the three-channel system as:

$$E_{tot} = (1-A_1) \times (1-A_2) + P_{conf} \times (1-A_3) \qquad (3)$$

4 ML Applications in a Multi-Channel Hybrid System

In this section the authors shall apply the analysis of the multi-channel decision making system defined in the previous section in a real-world diagnostics system composed of the following elements:

1. A human diagnostic practitioner trained in the diagnostics domain, representing the first channel of the parallel channel decision-making system, characterized by a certain mean accuracy of decision A_1
2. A machine intelligence system pre-trained in the diagnostic domain representing the second parallel channel of the decision-making system with mean accuracy of A_2
3. A data collection and processing unit that combines the results of the channels producing the cumulative and conflict outputs as described in the previous section.
4. An expert human practitioner called to make the final decision in the case of a conflict between the channels as described in the previous section.

Also, the additional assumptions are:

a) On average, the accuracies of the human and machine channels are in the same range [5,6], and
b) The accuracy of the expert channel in the final stage of the model is higher than that of either of the human or the machine channels in the parallel stage.

A system designed in this way may have a number of essential advantages over the traditional single-chain model described in Section 2. First, it wouldn't introduce significant overhead in time or effort, other than in the cases where such an exercise would be justified by the complexity of the case. If both human and machine channels agree on the initial assessment, the expert channel will not be involved. And due to high operational efficiency of the machine system and the fact that it can be used in the 24 × 365 regime, in most cases its result would be ready for evaluation well before those of the human practitioner, whereas the time and the additional cost of combining the results of the channels in a modern computer system can be evaluated as negligible.

Secondly, such a system allows to free the highly knowledgeable and high demand expert resources only for the most challenging cases where higher level of expertise is warranted. Such limited resources can be involved in a highly efficient distributed organization on a regional or even national level with remote access to all necessary data, tests and case history.

Thirdly, as will be reported in the results section it allows to substantially increase the overall accuracy of the diagnostics process through combining human and machine expertise in parallel decision-making channels resulting in measurable reduction of the overall incidence of errors in the diagnosis phase and as a direct consequence noted in the aforementioned studies, improving the outcome as well as cost efficiency of the entire treatment chain.

Finally, it is worth mentioning that the marginal cost of deployment of a pre-trained and pre-tested in the given diagnostics area machine intelligence system can be minimal, comparable to that of a routine operation of installing or upgrading software packages thus offering a measurable addition of value and quality of care with minimal extra cost.

5 Results

In this section the authors report the estimations of the gain in accuracy of the final diagnostics decision based on realistic values of the current diagnostics accuracy reported in the literature.

In this analysis, following [5] and other reports it will be assumed that in the diagnostics domain of interest the accuracy of machine intelligence system has reached or approached the average accuracy of a qualified, but not necessarily expert human practitioner. Thus, the machine system is considered in the analysis to be a peer to an average human practitioner in the given diagnostics area, but not necessarily to a distinguished expert.

For application of the proposed diagnostics model and illustration of its potential several different diagnostics areas were taken with the data on accuracy of diagnostics procedures and incidence of errors from comprehensive studies of diagnostics errors in primary care [8,9]:

(1) Internal conditions (e.g. COPD, Rheumatoid arthritis), 2004, [9]: diagnostic error incidence 13%, not including false positive cases. Adjusted to 20% to account for false positives.
(2) Asthma, [8]: diagnostic error of up to 30% within reasonable timeframe (wrong diagnosis or no diagnosis)
(3) Mammography, [9]: 10% and above
(4) Common across multiple diagnostic areas [9]: 13-15% excluding false positives.

In Table 2, the multi-channel decision-making model has been applied to the above conditions based on the analysis of the model accuracy in Section 3. As can be observed from these results based on reported incidence of diagnostic errors, the improvement in the accuracy of diagnostics resulting from introduction of a multi-channel decision-making system with an incorporated AI channel ranged from 8% to 13%.

Table 2. Comparative accuracy, single chain vs multi-channel hybrid models.

Condition	Human	Machine	Expert	Multi-channel hybrid model	Single chain model
(1)	80	80	90	92.8%	80%
(2)	75	75	85	88.1%	75%
(3)	90	90	95	98.1%	90%
(4)	85	85	90	95.2%	~85%

These results clearly demonstrate that incorporation of machine intelligence systems as a parallel source of opinion in the decision-making process with a human expert follow-up can significantly improve the accuracy of diagnostics in most reviewed areas with measurable potential benefits for the patients and for the primary care system.

6 Discussion

The results reported in the previous section demonstrate that the accuracy of routine diagnostics and the consequent outcome as well as the cost efficiency of the diagnostics phase can be significantly improved by harnessing the capabilities and advantages of machine intelligence systems as a parallel decision-making channel to that of a human practitioner, as in the standard practice of the day.

This conclusion, and the ensuing results are based on the assumption that the probability distributions of channel errors are primarily independent, as illustrated in Fig.3. In this case, the probability of a conflict between the channel can indeed be estimated as in (3).

The authors will attempt to justify this assumption as reasonable. Indeed, as has been pointed by multiple studies, e.g. [4], human performance in critical tasks is often affected by the factors of their condition and environment which machine systems are less dependent upon and affected by. Consequently, it can be expected that errors caused by these factors would not be correlated between the channels.

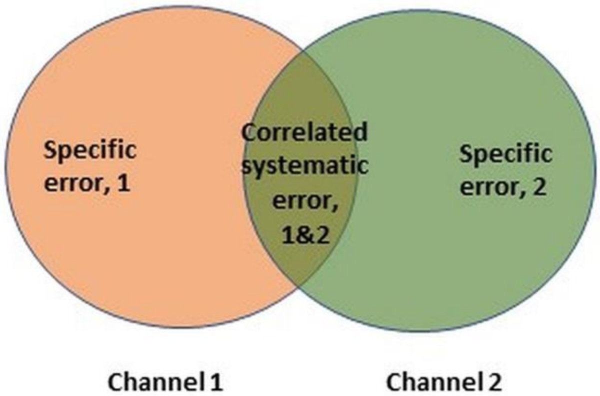

Fig.3. Specific and correlated systematic error in a parallel multi-channel system

Another cause of correlation of erroneous decisions can lie in the specifics of education and experience of the human practitioner vs. the machine system. Again, it can be observed, that the machine system would likely be trained with a much broader and larger sets of data, across geographic as well as individual practice spectrum, reducing the likelihood of correlated systematic errors with any individual human expert. And vice versa, any systematic or system errors in development and / or training of the machine systems are less likely to be reflected in the education and practice of a human practitioner reducing the likelihood of correlated errors. For these reasons the authors believe that the assumption of independence of human and machine decision-making can be made at least as a first approximation in evaluating the accuracy of hybrid decision-making systems with multiple parallel channels.

Importantly, the model equally addresses both channels of potential error in the single chain scenario: false negative cases that may cause deterioration of the condition and the prognosis due to undetected condition, resulting in prolonged treatment, less positive prognosis and an increase in the overall cost of treatment; and false positive ones, that may lead to unnecessary further testing and treatment and cause emotional discomfort to the patient and their families. In either case, if a disagreement in the decisions between the channels is detected, the case is brought to the attention of a leading expert in the field with improved chance of the correct decision.

It can be noted further that in the longer term the performance, i.e. in the case under consideration, diagnostics accuracy of machine systems can be expected to improve further and eventually surpass not only the average but even the expert ability of humans as has been the case with Chess and Go games [10,11], potentially resulting in further potential gain in the overall accuracy and the resulting efficiency of the multi-channel diagnostics models. In such an event, the efficiency and justification for the expert channel in the proposed model can be called into question, as due to a higher error rate it could actually reverse some of the corrections made by the more accurate machine channel. However, at the time this possibility appears to be remote, both in time and the state of technology.

In conclusion one needs to comment on the monitoring of the operational performance of the diagnostics system that is a necessary and very important phase in applications of any automated systems especially in the areas where it can affect the well-being and health of human population. As has been noted earlier in the section, one possible source of unaccounted error in the proposed type of systems that cannot be completely eliminated can be a systematic correlated error that may cause simultaneous failure of the channels.

An example of cases causing such systematic failures can be a subset of rare, non-standard, novel or substantially deviating from the norm in the diagnostics area cases where neither the human practitioner nor the machine system have received sufficient training or experience. While for aforementioned reasons the authors consider possibility of such errors as reasonably low on the average across the domains, it can certainly be an issue in specific diagnostics areas.

One approach to address such systematic issues could be to trace the diagnostic decision to the eventual outcome of the treatment. Availability and the analysis of such data would allow to identify, track and resolve this type of systematic errors by adding them in the curriculum and practice of both human and machine diagnostics practitioners.

7 Conclusion

The realities of aging population are driving the cost of health care system in the developed countries ever upwards calling for innovative approaches to increase the efficiency of the system while retaining and enhancing its reliability, quality of care and safety. Such opportunities can be found in harnessing the benefits of machine intelli-

gence methods in applications in essential patient care that can substantially improve the accuracy of the diagnostics systems while retaining full control over its operation.
The proposed model of combining human and machine expertise into a single synergetic operational system offers a number of significant advantages over the traditional “single-chain” models:
- demonstrated significant improvement in overall accuracy of diagnostics resulting in reduction in unnecessary spending and improved patient care;
- with minimal incremental cost of development and deployment;
- flexible: the model can be easily adaptable and transferrable to different areas of patient care;
- does not introduce any additional delay due to high performance of the parallel machine channel;
- allows the optimal use of the expert resources only in the cases that require their attention and involvement;
- fully compatible with distributed, high performance and outstanding quality service delivery operational models;
- combines strengths and advantages of the human and machine expertise for a significant improvement to the current practice;
- while retaining complete and uncompromised human control over the diagnostics and treatment.
The authors believe and fully expect that development and introduction into operational practice of primary care of hybrid and synergetic human-machine service delivery models of the proposed type and ones similar to it in the near future will have the potential to significant improve the quality, reliability, safety and efficiency of the patient care systems and may facilitate new ideas and approaches in further research, development and improvements in operational practice in this essential for the continuous well-being of the society field.

References

1. Kondro, W.: Canadian report quantifies cost of medical errors. Lancet 363 (9426), 2059 (2004).
2. Auraaen, A.: The economics of patient safety. OECD study report, https://www.paqs.be/fr-BE/Ressources/Livre-Mauve/PAQS-Livre-Mauve-20180614-02-OECD-PDF, last accessed 2019/09/27.
3. Sheikh A., Donaldson L., Westfall-Bates D. et.al.: Diagnostics errors. WHO Technical series on safer primary care, https://apps.who.int/iris/bitstream/handle/10665/252410/9789241511636-eng.pdf, last accessed 2019/09/27.
4. Sidorov, P.: The burnout syndrome in communicative professional workers. Gigiena i Sanitariia 3, 29-33 (2008).
5. Liu, X., Faes L., Kale A, et al.: 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 (2019).

6. AI equal with human experts in medical diagnosis, study finds. Guardian, https://www.theguardian.com/technology/2019/sep/24/ai-equal-with-human-experts-in-medical-diagnosis-study-finds, last accessed 2019/09/24.
7. Roberts, J.: Thinking machines: the search for Artificial Intelligence. Distillations 2 (2), 14–23 (2016).
8. Costopoulou O., Delaney B., Munro C.: Diagnostic difficulty and error in primary care – a systematic review. Oxford University Press (2008).
9. Graber M.: The incidence of diagnostic error in medicine. BMJ Quality and Care 22 (2), 21 (2012).
10. Hsu, Feng-hsiung: Behind Deep Blue: Building the Computer that Defeated the World Chess Champion. Princeton University Press (2002).
11. Silver D. Shrittwieser J. Simonyan K. et al.: Mastering the game of Go without human knowledge. Nature 550, 354–359 (2017).