



Using artificial intelligence to analyse and teach communication in healthcare

Phyllis Butow ^{a,*}, Ehsan Hoque ^b

^a University of Sydney, School of Psychology, Centre for Medical Psychology and Evidence-Based Medicine (CeMPED), Sydney, Australia

^b University of Rochester, Rochester Human-Computer Interaction Group, Rochester, New York, USA

ARTICLE INFO

Article history:

Received 26 August 2019

Received in revised form

15 January 2020

Accepted 16 January 2020

Available online 17 January 2020

Keywords:

Artificial intelligence

Machine learning

Communication

Healthcare

ABSTRACT

Communication is a core component of effective healthcare that impacts many patient and doctor outcomes, yet is complex and challenging to both analyse and teach. Human-based coding and audit systems are time-intensive and costly; thus, there is considerable interest in the application of artificial intelligence to this topic, through machine learning using both supervised and unsupervised learning algorithms. In this article we introduce health communication, its importance for patient and health professional outcomes, and the need for rigorous empirical data to support this field. We then discuss historical interaction coding systems and recent developments in applying artificial intelligence (AI) to automate such coding in the health setting. Finally, we discuss available evidence for the reliability and validity of AI coding, application of AI in training and audit of communication, as well as limitations and future directions in this field. In summary, recent advances in machine learning have allowed accurate textual transcription, and analysis of prosody, pauses, energy, intonation, emotion and communication style. Studies have established moderate to good reliability of machine learning algorithms, comparable with human coding (or better), and have identified some expected and unexpected associations between communication variables and patient satisfaction. Finally, application of artificial intelligence to communication skills training has been attempted, to provide audit and feedback, and through the use of avatars. This looks promising to provide confidential and easily accessible training, but may be best used as an adjunct to human-based training.

© 2020 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Health communication

Communication in healthcare is a core clinical skill essential to effective clinical diagnosis, treatment decision-making and achievement of optimal patient outcomes. Research has demonstrated strong positive relationships between health professionals' communication skills and patients' capacity to understand, recall and follow medical recommendations, self-manage chronic illnesses, and adopt preventive health behaviors [1–6]. In addition, studies have shown that the clinician's ability to listen and empathize with patients' emotions can have a profound effect on patients' psychological and functional outcomes, as well as their experience of, and satisfaction with, care [7,8].

Furthermore, clinicians' experience of and confidence in their

communication with patients can impact their own levels of occupational satisfaction, as well as of stress and burnout [9,10]. Poor physician communicators also face a higher risk of being sued by dissatisfied patients. One study that examined plaintiff depositions [11] found that 71% of the malpractice claims were initiated as a result of a physician-patient relationship problem, with most litigious patients perceiving their physicians as uncaring, poor deliverers of medical information, and poor listeners.

Recognition of the fundamental role of communication in health care has informed the development of doctrines such as informed consent, shared decision-making and patient-centred care, now enshrined in legislation in many countries, thus legally requiring health practitioners to practice in ways that conform to these approaches. There is, however, good evidence that health practitioners vary widely in their communication skills, and that these skills do not necessarily improve over time without intervention [12]. In response to these documented deficiencies, communication skills training is now provided in most medical and nursing programs, as well as for specialist trainees in disciplines such as

* Corresponding author. School of Psychology, Lifehouse Level 6-North (C39Z), University of Sydney, NSW 2006, Australia.

E-mail address: phyllis.butow@sydney.edu.au (P. Butow).

medical oncology and palliative care. However, some studies have suggested that such training teaches only basic skills, and that these are not sustained into clinical practice [13,14].

It is important that communication skills training is theoretically and empirically based, and that current skills can be reliably and validly assessed. Furthermore, trainees need to be provided with concrete and accurate feedback, while skills should be reliably audited over the long term to ensure standards are maintained. However, this is not easy to achieve. Communication is complex, with messages sent via verbal, para-verbal (e.g. voice tone) and non-verbal (e.g. eye-gaze, expressions) channels. Multiple participants (doctor, nurse, patient, caregiver) may be involved in the interaction, each with differing expectations, goals, roles, capabilities and vulnerabilities [15]. Tasks within consultations can be diverse, including information gathering, education, decision-making, relationship-building and managing emotions. Some topics discussed in medical consultations, such as prognosis, end of life care, pain and lost fertility, can be highly emotive. Thus, building comprehensive theory, and gathering empirical data to support training, can be difficult.

While there is evidence supporting models such as shared-decision-making and patient-centred care [16,17], there remains significant controversy in the field. For example, some studies have reported data suggesting that many patients do not want to share decisions when they feel they have inadequate expertise and are vulnerable and in need of reassurance rather than autonomy [18]. Some patients report wanting a paternalistic style of communication from their doctors for some medical tasks (such as when explaining the diagnosis and process of treatment), while preferring a more patient-centred style when discussing emotive issues such as prognosis [19]. Fundamental to resolving these controversies, and emerging with a sufficiently nuanced understanding of doctor-patient communication, is the accurate documentation of existing communication. Only then, can the relationships between communication and patient outcomes be reliably explored.

2. Coding health-professional-patient encounters

Attempts to document medical communication have largely utilized interaction analysis systems. Systems analyses “decompose a system into its component pieces for the purpose of studying how well those component parts work and interact to accomplish their purpose” [20]. Interaction analysis systems analyse communication between the doctor, patient, family, and other health professionals (HPs) in a qualitative and/or quantitative fashion. Interaction analysis systems typically describe task oriented and/or socio-emotional behaviours, but differ in their clinical focus (e.g., general practice or specialty), extent of coverage (whole consultation or specific behaviours only), and communication modes encoded (verbal, paraverbal, non verbal, or all) [21].

Ong et al. [22] conducted a systematic review of whole consultation interaction analysis systems, identifying twelve. Other systems have since emerged, such as CN-LOGIT (later re-named CANCODE) [23] and the Siminoff Communication Content and Affect Program (SCCAP) [24]. However, the most commonly applied interaction analysis system is the Roter Interaction Analysis System (RIAS) [25]. The RIAS codes every doctor and patient utterance into one of 37 mutually-exclusive and exhaustive categories. The RIAS captures *socio-emotional behaviours* e.g. agreement, showing concern, reassurance; and *task-oriented behaviours* e.g. giving directions, asking medical/therapeutic questions, giving lifestyle-related information. These categories can be combined to reflect the total amount of talk in broader categories such as patient-centred exchanges. Additionally, global ratings of anger, anxiety, dominance, interest, responsiveness, and warmth are allocated.

Some interaction analysis systems look for specific behaviours within a consultation, rating them as present or absent. Sometimes an overall qualitative rating is also applied (such as basic/extended, or poor/good). Such interaction analysis systems record aspects such as: response to emotion [26], information giving [27], shared decision making [28] and patient centred care [29,30]. Both whole consultation and more specific coding systems have been shown to be reliable and valid, and have generated much useful data about health-professional-patient communication [21].

Most of these interaction analysis systems, however, are operationalised by hand coding, with or without some computerised support (for example, to time exchanges). They are laborious, time-consuming and expensive to use. Funding for such exercises may be available for research, but is rarely available within educational facilities to enable accurate feedback to trainees on their communication skills, or at population levels to enable auditing of communication. Artificial intelligence (the ability of a computer algorithm or computer-controlled robot to perform tasks commonly associated with intelligent beings) [31] holds great promise in making this process much more cost-effective, as well as allowing exploration of communication outside of established theories. Ultimately, inexpensive, easy and always available communication audit through artificial intelligence could transform the field by providing evidence-based communication targets that may become required practice.

3. Artificial intelligence

While the exact definition of artificial intelligence (AI) remains controversial, there is significant enthusiasm regarding its potential in this field, based on its latest successes in machine learning and deep learning. Machine learning allows the development of novel algorithms that can automatically make decisions by relying on patterns and inferences, without the need for any explicit instructions. There are two types of learning algorithms: 1) supervised and 2) unsupervised. Supervised algorithms require availability of labelled data from which to learn. The performance of the algorithms is determined by testing them on “unseen” data with known labels. For example, if the algorithm performs with 95% accuracy on unseen data, it can then be deployed to make decisions in real-life scenarios, with the caveat that it will make mistakes 5% of the time. Nevertheless, a well-recognised problem with AI is transferability between settings, which is similar to the contrast between efficacy and effectiveness in traditional intervention research.

In many practical applications, the non-availability of labelled data, or inaccurate or biased labelling, are the major bottlenecks delaying or preventing effective supervised learning. For those scenarios, a different set of algorithms exist that can group unlabelled data based on similarities, patterns and differences without any prior information. This group of algorithms is called unsupervised learning.

While the application of machine learning on carefully crafted data has shown great promise, its utility in the real world is still limited by its ability to encompass the full complexity of human communication. For example, training an AI system to recognize human nonverbal behaviour, essential to understanding communication, is extremely complex. Consider the extensively studied computational challenge of playing chess. In chess, the first player can open with any of 20 moves, and the second player can do the same; thus, after the first two moves, there are $20 \times 20 = 400$ outcomes to specify. By extension, according to Avinash K. Dixit and his coauthors [32], the number of possible moves in chess is in the order of 10^{120} . A computer making a billion calculations a second would take approximately 3×10^{103} years to consider all these

possible moves. Because of the complexity involved, chess has become a common platform for computer theorists designing new optimization algorithms and performing complexity analysis.

Human nonverbal behaviors are by comparison significantly more complex. Using the 43 muscles of our face, it is possible for humans to produce approximately 10,000 unique facial expressions at any given time (which, for humans, can be counted in milliseconds) [33]. In other words, for two people, the number of possible facial paths after the opening “move” is $10,000 \times 10,000 = 100,000,000$. Adding to the challenge, while a chess player might have up to a minute between moves, those engaged in social interactions have barely a moment to register and interpret each facial movement. People who have trouble deciphering the social “games” others play confront an important challenge: the instantaneity of changing facial expressions. Factor in the complexity imposed by other nonverbal modalities—such as prosody (the pattern of stress and intonation in a language), gesture, and body movement—and the range of possible social–emotional moves in human interaction is truly astronomic, dwarfing that of chess. Rich real-time recognition and understanding of natural nonverbal communication is thus a great challenge for AI, requiring breakthroughs in multiple areas. Despite the challenges, however, the AI community has progressed significantly in developing computational frameworks that can model the interplay, redundancy, and dependency among behavioural modalities. We provide a brief summary of some of the key developments that have allowed new opportunities for practitioners and non-experts to take advantage of them.

Over the last few years, the field of AI has seen tremendous growth in terms of enabling new practical and reliable systems. After decades of effort in speech recognition, in 2017, Microsoft researchers reported a historical milestone by creating a technology that recognizes words in a conversation as well as a person does. The speech recognition accuracy was optimized by using several deep learning models. The automated speech recognition system yielded a word error rate which was as low as that of 4 professional human transcribers working together on the same benchmark [34]. Being able to reliably transcribe speech from an audio file has enabled the automated sensing of many other psycholinguistic cues, including word level prosody analysis, pauses, energy, and intonation.

In addition to machine recognition of what has been said and how this has been said, there is also a significant need to automate the understanding of the semantics and sentiment of the utterances. With many years of effort, IBM Watson recently released a prototype that understands emotion and communication style in text [35]. It works on a document as well as on sentences. On a document level, the system can provide an overall tone of the document, and on the sentence level, it is possible to identify specific content areas with strong emotional undertone. The types of sentiment captured include anger, fear, joy, sadness, analytical, confident and tentative. IBM Watson provides a cloud Application Programming Interface (API) allowing anyone with limited knowledge of AI to access it and use it to develop interactive applications.

The face is perhaps the most important channel for non-verbal communication. Facial expressions encode many different affective states, including confusion, stress, pain and empathy, which are very relevant and important states in healthcare communication. Paul Ekman first introduced the notion of a Facial Action Coding System which is a taxonomy of facial movements, correlated with certain expressions [33]. For example, lip corner pull (AU 12) and cheek raiser (AU6) are often correlated with a genuine smile. After many years of research, there is now a commercially available tool called Affectiva and an open source toolbox called openFace [36]. They both can automatically and in realtime recognize the facial

landmark, estimate head pose and eye gaze, and using those data calculate action units as defined by Paul Ekman. This new development now allows for an automated coding of facial movement and mental states which otherwise is very laborious and expensive to do with human experts. In the following sections, we examine how these systems have been applied to, and evaluated in, the health setting.

4. Artificial intelligence applications in communication research

Examples of metrics that could be identified through AI, guided by communication theory, are: a) the proportional time spent by HP and patient talking during the consultation, as an indicator of the power dynamic and the HP’s willingness to listen; b) overlapping talk or interruptions, as an indicator of respect and acceptance of the other’s agenda; c) the number of pauses longer than two seconds, as an indicator of comfort with silence and encouragement to continue; laughter and social talk as indicators of relationship-building and trust; d) word and sentence length and structure, and use of clinical jargon, as indicators of complexity and ease of understanding; e) turn-taking as an indicator of the interactivity of the consultation; and intonation, pitch, energy and pace, as an indicator of stress, anger, or interest. Once identified, associations between these attributes with important patient outcomes can be explored.

5. Is AI reliable, timely and valid in the context of communication analysis?

Many studies are in the formative stage of establishing the reliability and validity of AI communication coding, compared with human coders. An example is Durieux et al.’s analysis of audiotapes of at least one of the first three consultations between palliative care practitioners and 225 patients with advanced cancer (heterogeneous cancers, including breast cancer) [37]. The analysis used a machine learning approach to identify a relatively simple characteristic, pauses, which were further categorised as connective silences (which feel comforting, affirming, and safe) or not, by human coders. The study showed that the machine learning algorithm demonstrated moderate to substantial reliability (kappa 0.62; 95% confidence interval: 0.47–0.76). Human Coders alone required 61% more time than the machine learning method. No connective silences were missed by the machine learning screening algorithm.

Similarly, Mayfield et al. compared machine learning with human coding of 415 audiotaped routine outpatient visits of HIV patients [38]. They trained the machine to identify information giving (doctor) and requesting (patient) speech acts and calculated the ratio of information giving to requests. Automated coding produced moderate reliability with human coding (accuracy 71.2%, $\kappa = 0.57$), with high correlation between machine and human prediction of the information-giving ratio ($r = 0.96$).

In a more ambitious project, Gaut et al. [39] explored the ability of machine learning to identify subjects (161 possible codes) and patient symptoms (48 possible codes) discussed in audiotaped psychotherapy encounters. Subjects and symptoms were coded for each session in binary form (discussed or not). In addition, talk-turns in which a symptom was discussed were rated for the representativeness of that symptom on a scale of 1 (atypical) to 7 (typical), to test the system’s ability to code more fine-grained detail. Six graduate students also rated the representativeness of symptoms.

The authors compared the ability of the semi-supervised machine learning labelled latent Dirichlet allocation (L-LDA) model to

learn associations between text and codes, to predict codes in psychotherapy sessions, and to identify specific passages of within-session text representative of a session code, and compared it with a baseline lasso regression model. The L-LDA model out-performed the logistic regression model at predicting the occurrence of codes at the session level (average area under the ROC curve (AUC) score over all codes for the L-LDA model was 0.789 (SD = 0.137) compared to 0.702 (SD = 0.145) for the regression model, and was able to identify specific talk-turns representative of symptom codes, although not quite as reliably as human coders (e.g. human coders had an average AUC score of 0.94 compared to 0.89 for the L-LDA model, for the symptom anger). The authors concluded that the L-LDA model has the potential to be an objective, scalable method for accurate automated coding of psychotherapy sessions. It certainly shows promise, although translating these findings into the real world to demonstrate utility has yet to be done.

6. Can AI studies identify specified communication elements and demonstrate their association with patient satisfaction?

Other studies have used AI algorithms to identify pre-specified characteristics in health professionals' communication (thought to be important on theoretical grounds) and examine their association with patient outcomes, usually collected via patient reported outcome measures (PROMs). An example is work by Wallace et al., [40] who divided doctors treating HIV patients into those who received positive patient ratings on a post-consultation survey, versus those who received negative ratings. A number of doctor variables were audited by AI, and associated with these doctor clusters. They found that doctors who did more advising without permission (doctor-centred) were significantly more likely to be in the poorly rated cluster ($p < 0.001$).

In the Mayfield study quoted above [38], the researchers explored associations between the ratio of information giving to patient requests, and patient reports of communication quality from post-visit surveys. The regression significantly predicted four of five patient-reported measures of communication quality ($r = 0.263-0.344$).

AI can also produce novel and helpful ways to display data. An example is the Discursis program, designed to display a conversation diagonally turn-by-turn, and show the extent to which speakers repeat their own conceptual content in the short, medium or long term, or how much they engage with speakers' content (see Fig. 1). This is based on Platt et al.'s assertion [41] of the importance of engagement (showing interest in and rapport with the patient's perspective) and enlistment (inviting the patient to collaborate on decision-making and treatment planning). Angus et al. [42,43] used this method to display data from medical consultations, and were able in a small demonstration set, to identify communication patterns that distinguished between doctors who did and did not balance task orientation and rapport.

While theory can guide algorithm development for AI analysis, unsupervised learning freely looks for patterns that optimally distinguish participants known to vary on important outcomes, potentially revealing new behaviours not previously thought to be important communication determinants. Epstein and colleagues recently used this approach in a fascinating study of communication in the context of stage 3 and 4 cancer [44]. This team analysed transcripts of 122 visits between cancer patients and their oncologists, and post-visit patient ratings of satisfaction with physician's communication. They employed AI unsupervised clustering of conversation features into "styles" and used machine learning models to automatically predict whether a doctor-patient interaction was rated high or not on patient satisfaction, with a best-performing 71% test set accuracy. The results showed that doctors

who spoke more words ($p < 0.05$), and repeated themselves less, were more likely to be rated high on patient satisfaction (see Fig. 2), although the latter finding did not quite reach significance ($p=0.06$). This finding flies in the face of theories of communication regarding behaviours that facilitate patient understanding and shared decision-making, which would predict the opposite. The authors conclude that further research is required to understand more about the context, and patient and doctor factors, that might explain these results. Nonetheless, this research demonstrates the value of unsupervised learning in raising new hypotheses and challenging accepted assumptions. It could also be argued that these results reflect limitations in this approach which may not take into account the complexity of the situation, and the factors modifying or mediating these relationships, which human qualitative coding can better achieve.

7. Artificial intelligence applications in training and audit

Glyn Elwyn and colleagues have recently provided a useful overview of potential applications of AI in medical communication training [45]. These authors suggest that AI analysis of communication elements such as words and phrases, turn-taking, tone and style could be used to provide detailed and confidential feedback and comparison with peers to trainees, in a safe, confidential setting, at regular intervals. Such methods could additionally be used as part of formal assessments of communication skills using metrics that are standardized, repeatable and objective. These and other authors have emphasised the need for automated approaches to be able to comprehensively provide all health professionals with the opportunity to receive feedback on their communication, at reasonable cost. These communication training modules can be rapidly scalable ensuring equality in medical training.

Whether this approach would be acceptable and useful to trainees remains to be seen. Texts on communication skills training suggest the importance of maintaining safety and dignity during this training, using a learner-centred approach which focuses on what the learner wants to explore, as well as receiving feedback perceived as authentic from simulated patients, and having the opportunity to try different strategies in repeated encounters [46]. Skilled facilitators can ensure these elements are present during communication skills training and can learn to adapt; it is not clear that AI feedback will achieve the same. Perhaps, the ideal scenario would be to allow AI driven training as supplemental materials in addition to expert human facilitators.

Others have suggested the use of avatars (a computerised image to represent a person within a virtual reality environment), in this case a patient, in order to train communication skills, arguing that they have advantages over simulated patient (SP) encounters in roleplay [47]. Avatars can be created to depict physical symptoms and deformities more easily than SPs, can be used repetitively to improve behaviour, and provide a secure, confidential, low-risk environment that is less threatening to students [47].

For example, Pan et al. [48] have investigated the use of avatars to provide health professionals with the opportunity to respond to ethical dilemmas and challenging situations, in this case, a patient and daughter strongly requesting antibiotics for a probably viral infection. Participants in their study (12 experienced GPs and 9 trainees) reported high immersion in the interaction, and that the avatar produced believable and appropriate responses. Furthermore, experienced doctors were more resistant to prescribing antibiotics than junior trainees, suggesting that the methodology could distinguish differential levels of skill.

Kleinsmeith et al. [47] investigated the use of avatars to teach empathy, a somewhat more challenging goal than teaching argument or information skills. 110 3rd year medical students interacted

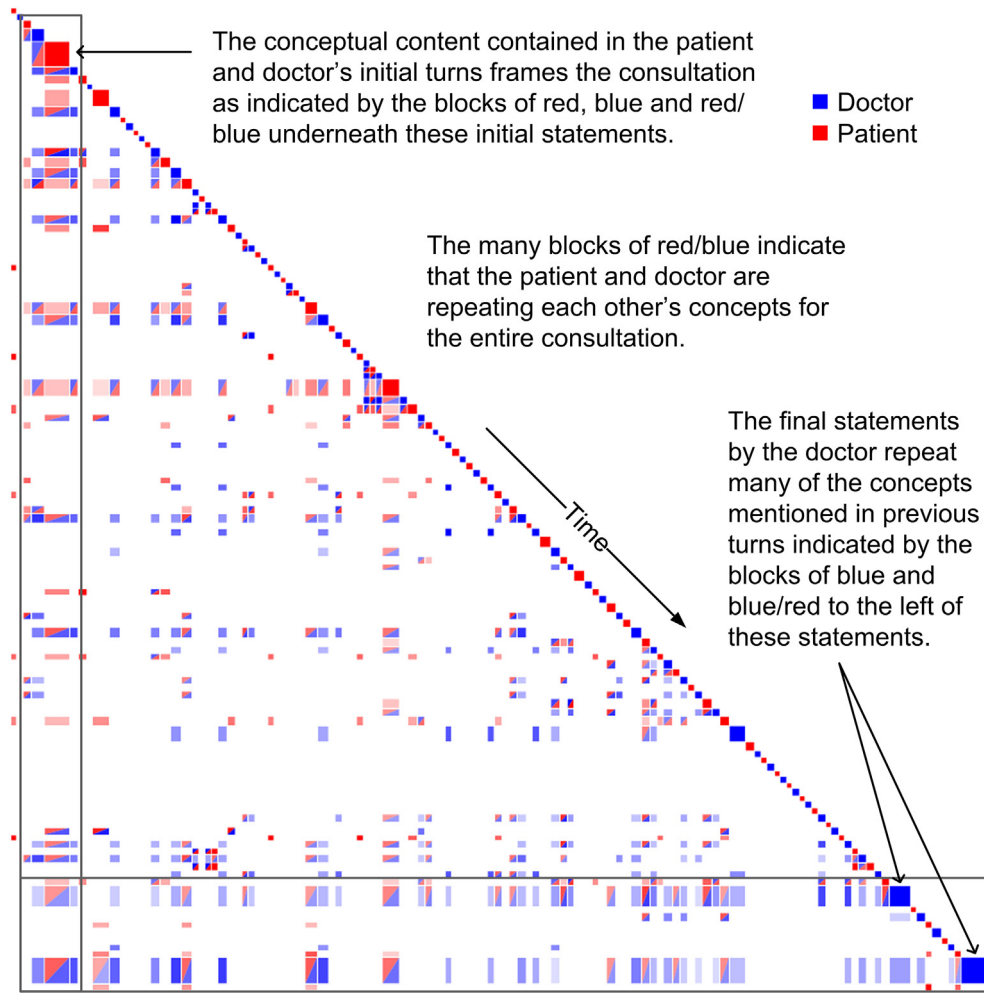


Fig. 1. Example of computerised visualisation of a consultation used by Angus et al., 2012.

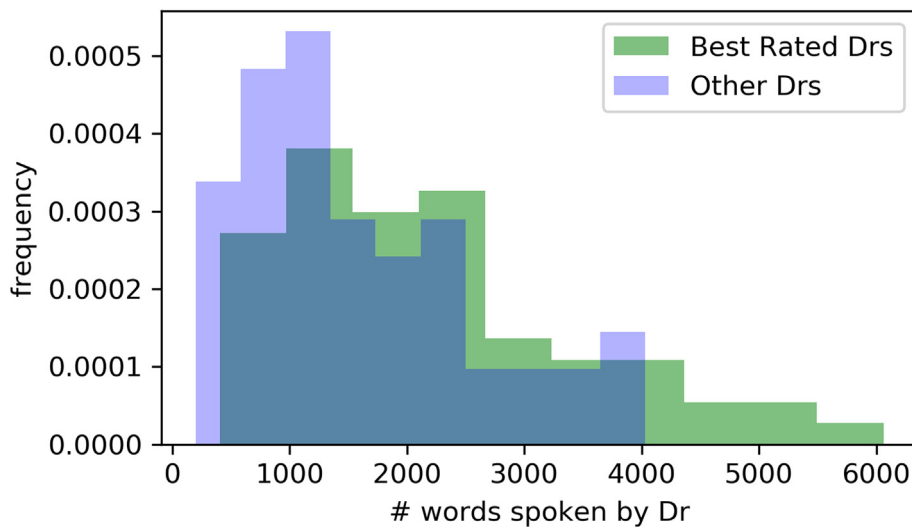


Fig. 2. Comparative histograms of the number of words spoken by doctor for best rated doctors and other doctors. From Sen et al., 2017.

with an avatar and an SP (in random order) that amongst other responses, provided four cues for empathy, by expressing in varying levels of intensity, concerns or emotions. The study found that the

students responded more empathetically and with longer replies, to the avatar than to the SP, perhaps because they felt less pressure to respond quickly. The authors suggested that this less pressured

environment may have facilitated cue recognition, which could be generalised to real interactions, although that remains to be tested.

8. Limitations and future directions

Clearly this application of AI and machine learning to the field of health professional communication is an emerging area, with much potential for further development. The possibility of analysing large amounts of data in a very cost-effective manner is enticing, while the potential to offer individualised, objective and repeated feedback to large numbers of learners is exciting. Often, an individual may feel hesitant to seek help from a human trainer due to scheduling constraints, or the possible stigma involved in seeking more practice time. An AI driven training system which is available whenever and wherever, addresses such constraints and takes away the stigma associated with seeking practice/help to enable more effective communication. However, to date, most applications have been limited, targeting only a few concepts, and not yet investigating complex inter-relationships between variables. It remains important that we generate evidence about the feasibility, reliability, acceptability and effectiveness of this approach before broad implementation. Further research will need to explore how authentic the experience of obtaining computerised feedback is to learners, and whether machine learning can produce insights that reflect the true complexity of health-professional-patient communication.

Funding source

There was no funding for this paper.

Ethical approval

Ethical approval was not required for this overview paper.

Declaration of competing interest

The authors have no conflicts of interest to declare.

References

- [1] Duffy FD, Gordon GH, Whelan G, Cole-Kelly K, Frankel R. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. *Acad Med* 2004;79:495–507.
- [2] Heisler M, Bouknight RR, Hayward RA, Smith DM, Kerr EA. The relative importance of physician communication, participatory decision-making, and patient understanding in diabetes self-management. *J Gen Intern Med* 2002;17:243–52.
- [3] Renzi C, Abeni D, Picardi A, Agostini E, Melchi CF, Pasquini P, Prudu P, Braga M. Factors associated with patient satisfaction with care among dermatological outpatients. *Br J Dermatol* 2001;145:617–23.
- [4] Safran DG, Taira D, Rogers WH, Kosinski M, Ware JE, Tarlov AR. Linking primary care performance to outcomes of care. *J Fam Pract* 1998;47(3):213–20.
- [5] Sullivan LM, Stein MD, Savetsky JB, Samet JH. The doctor-patient relationship and HIV-infected patients' satisfaction with primary care physicians. *J Gen Intern Med* 2000;15:462–9.
- [6] Zachariae R, Pederson CG, Jensen AB, Ehmrooth E, Rossen PB, Von der Maase H. Association of perceived physician communication style with patient satisfaction, distress, cancer-related self-efficacy, and perceived control over the disease. *Br J Canc* 2003;88:658–65.
- [7] Kim SS, Kapliwicz SA, Johnston MV. The effects of physician empathy on patient satisfaction and compliance. *Eval Health Prof* 2004;27(3):237–51.
- [8] Lelorain S, Brédart A, Dolbeault S, Sultan S. A systematic review of the associations between empathy measures and patient outcomes in cancer care. *Psycho Oncol* 2012;21:1255–64.
- [9] Vermeir P, Degroote S, Vandickx D, Mariman A, Deveugele M, Peleman R, et al. Job satisfaction in relation to communication in health care among nurses: a narrative review and practical recommendations. *Sage Open*; 2017. p. 1–11. <https://doi.org/10.1177/258244017788486>.
- [10] Ramirez AJ, Graham J, Richards MA, Cull A, Gregory W, Leaning MS, et al. Burnout and psychiatric disorder among cancer clinicians. *Br J Canc* 1995;71(6):1263–9.
- [11] Beckman HB, Markakis KM, Suchman AL, Frankel RM. The doctor-patient relationship and malpractice. Lessons from plaintiff depositions. *Arch Intern Med* 1994;154(12):1365–70.
- [12] Maguire P, Fairbairn S, Fletcher C. Consultation skills of young doctor: II – most young doctors are bad at giving information. *BMJ* 1986;292:1576–8.
- [13] Bachmann C, Roschlaub S, Harendza S, Keim R, Scherer M. Medical students' communication skills in clinical education: results from a cohort study. *Patient Educ Couns* 2017;100:1874–81.
- [14] Wouda JC, van de Wiel HBM. The communication competency of medical students, residents and consultants. *Patient Educ Couns* 2012;86:57–62.
- [15] Rocque R, Leanza Y. A systematic review of patients' experiences in communication with primary care physicians: intercultural encounters and a balance between vulnerability and intergrity. *PLoS One* 2015. <https://doi.org/10.1371/journal.pone.0139577>.
- [16] Scalia P, Durand MA, Berkowitz JL, Ramesh NP, Faber MJ, Kremer JAM, Elwyn G. The impact and utility of encounter patient decision aids: systematic review, meta-analysis and narrative synthesis. *Patient Educ Couns* 2019;102(5):817–41.
- [17] Epstein RM, Street RL. The values and value of patient centred care. *Ann Fam Med* 2011;9(2):100–3.
- [18] Robinson A, Thomson R. Variability in patient preferences for participating in medical decision making: implication for the use of decision support tools. *Quality in Health Care* 2001;10(Suppl 1):i34–8.
- [19] Dowsett SM, Saul JL, Butow PN, Dunn SM, Boyer MJ, Findlow R, Dunsmore J. Communication styles in the cancer consultation: preferences for a patient-centred approach. *Psycho Oncol* 2000;9(2):147–56.
- [20] Bentley LD, Whitten JL. *Systems analysis and design for the global enterprise*. seventh ed. McGraw-Hill Education; 2007. p. 160.
- [21] Butow P, Ford S. Issues in coding cancer consultations: interaction analysis systems. In: Kissane D, Bultz B, Butow P, Finlay I, editors. *Handbook of communication in oncology and palliative care*. Oxford UK: Oxford Press; 2010.
- [22] Ong LML, De Haes JCJM, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. *Soc Sci Med* 1995;40:903–18.
- [23] Butow PN, Dunn SM, Tattersall MHN, Jones QJ. Computer-based interaction analysis of the cancer consultation. *Br J Canc* 1995;71:1115–21.
- [24] Siminoff LA, Step MM. A comprehensive observational coding scheme for analysing instrumental, affective, and relational communication in health care contexts. *J Health Commun* 2011;16(2):178–97.
- [25] Ong LML, Visser MRM, Kruyver IPM, Bensing JM, Van Den Bruink Muinen A, Stouthard JML, et al. The Roter interaction analysis system (RIAS) in oncological consultations: psychometric properties. *Psycho Oncol* 1998;7:387–401.
- [26] Del Piccolo L, de Haas H, Heaven C, Zimmerman C, Finset A. Development of the Verona coding definitions of emotional sequences to code health providers' responses (VR-CoDES-P) to patient cues and concerns. *Patient Educ Couns* 2011;82(2):149–55.
- [27] Koedoot CG, Oort FJ, de Haan RJ, Bakker PJ, de Graeff A, de Haes JC. The content and amount of information given by medical oncologists when telling patients with advanced cancer what their treatment options are. *palliative chemotherapy and watchful-waiting*. *Eur J Cancer* 2004;40(2):225–35.
- [28] Elwyn G, Hutchings H, Edwards A, Rapport F, Wensing M, Cheung WY, Grol R. The OPTION scale: measuring the extent that clinicians involve patients in decision-making tasks. *Health Expect* 2005;8(1):34–42.
- [29] Street RL, Gordon HS, Ward MM, Krupat E, Kravitz RL. Patient participation in medical consultations: Why some patients are more involved than others. *Med Care* 2005;43(10):960–9.
- [30] Brown JB, Stewart M, Ryan B. Assessing communication between patients and physicians: the measure of patient-centred communication (MPPC). In: Working paper series, paper #95-2. second ed. Longon Ontraio, Canada. Thames Valley Family Practice Research Unit and Centre for Studies in Family Medicine; 2001.
- [31] Russell SJ, Norvig P, Davis E. *Artificial intelligence: a modern approach*. third ed. Prentice Hall; 2010.
- [32] Dixit AK, Skeath S, Reiley DH. *Games of strategy*. New York: Norton; 2009.
- [33] Ekman P, Friesen W. *Facial action coding system: a technique for the measurement of facial movement*. Consulting Psychologists Press; 1978.
- [34] <https://www.microsoft.com/en-us/research/blog/microsoft-researchers-achieve-new-conversational-speech-recognition-milestone/>.
- [35] <https://www.ibm.com/watson/services/tone-analyzer/>.
- [36] Baltrušaitis T, Robinson P, Morency L-P. Openface: an "open source facial behavior analysis toolkit," in applications of computer vision (WACV), 2016 IEEE winter conference on. *IEEE*; 2016. p. 1–10.
- [37] Durieux BN, Gramling CJ, Manukyan V, Eppstein MJ, Rizzo DM, Ross LM, Ryan AG, Niland MA, Clarfeld LA, Alexander SC, Gramling R. Identifying connectional silence in palliative care consultations: a tandem machine-learning and human coding method. *J Palliat Med* 2018;21(12):1755–69.
- [38] Mayfield E, Laws MB, Wilson IB, et al. Automating annotation of information giving for analysis of clinical conversation. *Am Med Inform Assoc* 2014;21:e122–8.
- [39] Gaut G, Steyvers M, Imel ZE, Atkins DC, Smyth P. Content coding of psychotherapy transcripts using labeled topic models. *IEEE J of Biomedical and health informatics* 2017;21(2):476–87.
- [40] Wallace BC, Dahabreh IJ, Trikalinos TA, Laws MB, Wilson I, Charniak E. Identifying differences in physician communication styles with a log-linear

- transition component model. Proceedings of the 28th AAAI conference on artificial intelligence 214.
- [41] Platt FW, Gordon GH. *Field guide to the difficult patient interview*. Lippincott Williams & Wilkins; 2004.
- [42] Angus D, Smith AE, Wiles J. Conceptual recurrence plots: revealing patterns in human discourse. *IEEE Transactions on visualisation and computer graphics* 2012;18:988–97.
- [43] Angus D, Smith A, Wiles J. Human communication as coupled time series: quantifying multi-participant recurrence. *IEEE Trans Audio Speech Lang Process* 2012;20:1795–807.
- [44] Sen T, Ali MR, Hoque ME, Epstein R, Duberstein P. Modelling doctor-patient communication with affective text analysis. 2017 seventh international conference on affective computing and intelligent interaction (ACII).
- [45] Ryan P, Luz S, Albert P, Vogel C, Normand C, Elwyn G. Using artificial intelligence to assess clinicians' communication skills. *BMJ* 2019;364:1161. <https://doi.org/10.1136/bmj.1161>.
- [46] Kurtz SM, Silverman J, Draper J. *Teaching and learning communication skills in medicine*. Radcliffe Medical Press; 1998.
- [47] Kleinsmith A, Rivera-Gutierrez D, Finney G, Cendan J, Lok B. Understanding empathy training with virtual patients. *Comput Human Behav* 2015;52:151–8.
- [48] Pan X, Slater M, Beacco A, Navarro X, Bellido Rivas AI, Swapp D, et al. The responses of medical general practitioners to unreasonable patient demand for antibiotics - a study of medical ethics using immersive virtual reality. *PLoS One* 2016;11(2):e0146837. <https://doi.org/10.1371/journal.pone.0146837>.