Natural Language Processing, Wearables, and Their Combination in Healthcare: Opportunities, Challenges, and Considerations

Frank Rudzicz[1,2,3,4], Raymond Ng[5], Robert Wu[6], Mohamed Abdalla[2,4], Giuseppe Carenini[5], Kendall Ho[7]

[1] Li Ka Shing Knowledge Institute, St Michael's Hospital, Toronto, Canada,
[2] Department of Computer Science, University of Toronto, Toronto, Canada
[3] Surgical Safety Technologies Inc, Toronto, Canada
[4] Vector Institute for Artificial Intelligence, Toronto, Canada
[5] Department of Computer Science, University of British Columbia, Vancouver, Canada
[6] Toronto General Hospital Research Institute, University Health Network, Toronto, Canada
[7] Department of Emergency Medicine, University of British Columbia, Vancouver, Canada

Email: frank@spoclab.com, rng@cs.ubc.ca, robert.wu@uhn.ca, msa@cs.toronto.edu, carenini@cs.ubc.ca, kendall.ho@ubc.ca

***Abstract*—Natural Language Processing (NLP) is a growing field in data analytics that is increasingly applied in healthcare for assessment of patients' electronic records or patients' dialogues to unearth medical issues, emotions, or cognition. Meanwhile, active research is taking place in the use of sensors and wearables in tracking patient phenotypes and physiologic status in diagnosis and management of diseases. This paper's central purpose is to explore how natural language processing can be combined with wearables in healthcare, and how this approach can advance health service delivery and health system transformation. Our recommendations are to stimulate interest and lower barriers in order to implement this novel combination of solutions in healthcare industries, and to do so in consideration of various technical and ethical concerns.**

Keywords - natural language processing; digital health; wearables.

I. INTRODUCTION

Communication between health professionals and patients is a highly important area that affects diagnosis, treatment, and mitigation of acute and chronic diseases. In this era of interprofessional team based care, health professionals need to work together and often asynchronously to ensure smooth continuity of care of patients. Patients themselves can also benefit greatly from optimizing self-management in partnership with health professionals. Underpinning good communication amongst health professionals and patients is the comprehensive capturing and sharing of essential data of the patients in their expressions of symptoms, their physiologic changes, their responses to different types of therapies, and their physical, emotional, and social wellness. The current state of the art in capturing patient data to facilitate accurate tracking of disease progress is based primarily on observational narratives of health professionals, documented episodically in patients' electronic health records, incompletely shared between health professionals and patients, and without any significant patient input. Improvement in these areas conceivably can lead to improvement of disease management through interprofessional collaboration and partnership with patients through patient-centred care.

Modern information and communication technologies can contribute greatly to data capturing and documentation of patient-centric data and its analysis to support diagnosis and management, thereby enriching the therapeutic experiences for health professionals and patients. Data streams that can be tracked include numeric quantities, text, and physiologic signals, which can be valuable for disease management through big data analysis.

This paper will explore opportunities of Natural Language Processing (NLP) and wearables in health as follows. Section 2 gives an overview of NLP opportunities. Section 3 explores NLP in data mining and disease surveillance. Section 4 explores wearables in health. Section 5 posits how NLP and wearables can combine to contribute to health. Section 6 highlights issues and challenges of NLP and wearables in health. Section 7 makes recommendations and conclusions.

II. OPPORTUNITIES OF NLP IN HEALTH

NLP, or computational linguistics, is the study of using computer algorithms and tools to process or analyze natural language data. Key sub-areas of NLP include speech recognition, natural language understanding and natural language generation. In the past two decades, machine learning approaches become more and more popular for NLP, essentially relying more and more on the availability of training data.

 ISBN: 978-1-61208-763-4

NLP is an applied subarea of artificial intelligence that uses Machine Learning (ML) in a variety of tasks, including translation between languages, automatic summarization, speech recognition, information retrieval and search, and general areas of classification. To varying extents, each of these tasks can be expressed within the medical domain, including service delivery, and with each application there are particular considerations that must be taken into account. Proper application of these tools has the potential for significant transformation across the health system; improper application has the risk of significant loss of trust and unintended consequences. This provides the context for the potential impact of natural language processing for big data mining in the electronic health record.

Clinical decision support can assist through enabling better scoring systems, informed through NLP and structured data, such as lab values and vital signs. This can include scores that predict increased risk of an adverse event, including readmission, deterioration in hospital, and 1-year mortality. This, of course, assumes that clinicians will be able to use this information effectively, and that information with regards to deterioration or adverse events can be delivered in an actionable way. This will require better fine-tuning to reduce false alerts, and better design for human-computer interaction.

A computational model that captures patient speech patterns by NLP can be deconstructed into three components: 1) topic modeling to highlight and prioritize topics in patients' communication; 2) sentiment analysis to detect patient changes in affect, such as anxiety or depression, that interconnect with patient state of wellness and need for hospital visits; and 3) discourse coherence analysis to gauge patient coherence in their thinking and detect onset of mental dysfunction or delirium. These three components act synergistically to analyze the totality of the patient's speech pattern and content for optimal. For example, in patients with depression, they ruminate on certain morbid thoughts and topics, use words associated with depression and anxiety, and may have incoherent speech that are difficult for others to understand when the depression is profound or associated with delusions. Topic modeling can pick up the rumination, sentiment analysis can identify words with negative emotional connotations, and discourse coherence can highlight disturbed flow or logic of thinking. Liu *et al.* [1] demonstrated that automated NLP screening could identify 10.3% of 27,002 comments in a microblog were suggestive of suicidality with high precision (0.86), recall (0.78), F-measure (0.86), and accuracy (0.88).

III. NLP IN DATA MINING AND DISEASE SURVEILLANCE

Word embeddings (i.e., 'distributed representations') are dense numeric representations of words which serve as input to a wide array of ML methods. Typically, by learning to do automatic word prediction accurately (through measures of statistical perplexity on training data), these embeddings induce latent dimensions that encode aspects of morphology, syntax, and even semantics. The results therefore can capture meaningful relationships among concepts in the data not afforded by traditional methods.

Publicly-available, pre-trained word embedding models may suffice for certain tasks, but typically training on text within a single domain leads to improved performance [2]. Publicly-available pre-trained clinical word embedding models have been trained on PubMed abstracts and full-text PubMed documents; however, those corpora constitute very different language than clinical notes, limiting transferability.

NLP can uncover information in unstructured data that is not always present (or is inconsistent) in the structured data, including predicted cancer recurrence, whether a patient smokes, and drug treatment patterns [3]. However, traditional text search is typically unsuitable because of spelling differences, synonyms, and general ambiguity. Dubois and Romano [4] trained word embeddings using anonymized notes, treating negative counterparts of words as a single word, since negations are important in evaluation. They also trained embeddings on journal abstracts (MCEMJ) from OHSUMED [5] and evaluated their embeddings on several tasks (e.g., disease and mortality prediction) at the word-, note-, and patient-levels. Embeddings trained on medical notes resulted in greater downstream accuracy than those trained on abstracts.

Unlike other domains, larger data sets (i.e., 'corpora') do not necessarily produce better biomedical models [6]. Wang *et al.* [3] found that word models trained from electronic medical records had the best F1 score (0.90, a geometric mean of recall and precision) on a clinical information extraction task, compared with models trained on larger (but more general) data sets.

Beyond *learning* word embeddings using contextual information during training, it is also important to use context during *inference*, which means modifying outputs in different scenarios. For example, if one knows the author typing a clinical note, the word embedding space may shift. This generalizes to other covariates which may have an effect on the language. Nguyen *et al.* [7], for example, used convolutional neural networks to train an end-to-end deep learning system that learned to extract features from medical records and predicted future risk automatically. Some simple extensions were not taken in that work (e.g., long-term dependencies were simply captured through a max-pooling operation). A more careful consideration of effects that have a temporal component, such as the progression of symptoms or other disease trajectories, ought to enable more accurate predictions.

IV. OPPORTUNITIES FOR WEARABLES IN HEALTH

Commercial devices, such as Apple Watch and Fitbit, have fueled the popularity of qualified self [8] – the measuring of individuals' own physiologic signals so as to

 ISBN: 978-1-61208-763-4

gain insights into their own health and wellness. A 2017 survey by the Dublin Quantified self-group [9] found that some of the most common variables being tracked were: steps, sleep, weight, heart rate, diet, and exercise. The reasoning behind the tracking and data usage included "to make change to my life", "I learn something new about myself", "Monitor the effect of a new regime", to motivate myself", and "to make myself accountable". These are powerful reflections of individuals committed and ready to make changes through optimizing self-management.

As a variety of newer and innovative wearables come to the market, the opportunity to apply these devices for continuous tracking of patients for disease diagnosis and management is not only imaginable, but becoming reality. For example, tracking physiologic measures with a wearable necklace was found very promising for predicting hospitalization of patients [10]. Also, continuous monitoring in a hospital ward setting has found a dramatic reduction of unexpected cardiac arrest and length of stay of patients [11]. The Apple watch has recently been certified by Food and Drug Administration (FDA) to be a medical device for diagnosing and tracking atrial fibrillation [12]. The opportunity is ripe to introduce wearables into the health system as an integral part of patient surveillance and monitoring, and incorporate these data into the electronic health record to document patient digital phenotypes [13] and characteristic signatures of their diseases in their well management or exacerbation stages so as to better support and monitor patients' progress. This field of physiological informatics is rapidly rising in recognition and prominence in disease diagnosis, management, and even early prediction [14].

V. COMBINING NLP AND WEARABLES IN HEALTH

The discipline of medicine is built on the foundation of history taking and physical examination of patients in order to carefully deliberate on the constellation of symptoms that the patients experience, and the patterns of these symptoms to suggest the diagnoses and management approaches. With the promise of NLP and wearables, we have now the opportunity to digitize and record these interactions not only narratively in traditional ways, but also digitally in a brand new and innovative approach. This digitization can support diagnosis, management, and data analytics to learn about these individual patients and their personalized patterns of illnesses and their unique responses to therapies. The capturing of this "digital mirror" of the patients is unprecedented, and can truly revolutionize how medicine will be practiced in the future.

The combination of NLP and wearables will tremendously increase the amount of information that we will have about patients in our Emergency departments and admitted to hospital. Wearables will provide continuous real-time data on patient status, including vitals, activity, and sleep. Applying NLP to unstructured data, including clinical notes, and communication between providers, will complement that information. The opportunities could be vast and include significant improvement in the domains of clinical decision support, quality improvement, and research.

VI. CHALLENGES AND CONSIDERATIONS

To develop, implement, and deploy NLP approaches in biomedical contexts, it is necessary to define a principled framework for systematic identification and response to ethical dilemmas that may arise. As biomedical NLP lies at the intersection of NLP and biomedicine, it inherits ethical principles and conceptual frameworks developed for both parent-fields, such as those laid out in the International Code of Medical Ethics (adopted by the World Medical Organization) and ACM's Code of Ethics. These, and others, serve as a good starting point to investigate and further develop ethical questions that may arise in biomedical NLP applications. However, critically, these existing frameworks do not cover emerging ethical concerns, which fall beyond the scope of many established frameworks. We take the approach of Hirst [15] in discussing possible areas of concern when it comes biomedical NLP by exploring three areas: 1) concerns regarding how research is carried out, 2) concerns regarding motivations for the research, and 3) concerns about the unintended consequences of our research.

First, regarding how research is performed, we inherit the general ethical guidelines of research ethics. While formulations may differ in detail, the foundations are largely agreed upon. Here, we highlight possible issues facing biomedical NLP researchers which affect how research should be performed. One such issue is the problem of demographic skew. Because of the data available for training, ML algorithms often underperform for minority demographics [16]-[18] and this can have life-altering consequences for our patients. Additionally, upholding the privacy of patients is of vital importance and greatly affects how we do research. However, in addition to basic precautionary steps, such as encrypting hard-drives and password protection, we must take additional steps to protect the data we work with. The literature presents many ways that the privacy of patients can be compromised, including adversarial attacks [19], re-identification from released models [20], and re-identification of hidden variables from de-identified datasets [21]. We must be aware of the vulnerabilities of the techniques used and take the appropriate measures to protect our patients.

Second, regarding the motivation for research, we must, as our parent fields do, proceed with the intention of improving patients' lives and alleviating suffering. This was at the core of Hippocratic oath, which was modernized at the Declaration of Geneva, and also adapted for the general researcher by many including Joseph Rotblat who proposed a "Hippocratic Oath for Scientists" [22]. As scientists, we must place the well-being of patients first and take actions

 ISBN: 978-1-61208-763-4

in line with the aforementioned principles. Concretely, in a biomedical application of NLP, this could translate to being aware of the issue of over- or under-exposure [23]. Overexposure, the increased mainstream attention to topics in health, helps to create biases that lead to discrimination regarding project choices, and approaches. To those of us applying biomedical research to clinical settings, we must ensure not to over-complicate our work. The most novel approaches in the literature often lack the rigorous testing required by healthcare systems. When the results of various systems are comparable, we should opt to use the simpler and more studied of those systems. To those pushing the state of the research, we must not ‘oversell’ our results and the capabilities of our systems. Doing so not only hurts trust in our work when they fail to match the oversold promises, but also distract from the more useful discussions possible around our work regarding limitations and possible improvements.

Third, we must be aware of the unintended consequences of our research and work to mitigate the negative effects of our work. Researchers should be aware that ML methods serve to further propagate the biases implicitly and explicitly present in our society [24] unless methods to remove and prevent such bias are taken [25]. Accurately determining predisposition and risk for a disease given publicly available text, such as social media, would enable medical professionals to improve quality of life but would also allow insurance companies the opportunity to deny coverage thereby *reducing* quality of life.

Using patient speech patterns to build models that inform disease management raises several important questions. AI systems for NLP currently generate imprecise analyses of patients’ explicit meaning and implicit affect and cognition. Who will be responsible for erroneous therapeutic decisions? How will these percolate into therapeutic decisions? Would NLP allow health professionals to be more empathic to these patients and sensitive to their physical and mental management needs? If information collected regarding management may reveal secondary or incidental findings (e.g., emotional symptoms, signs of domestic abuse) that may be shared with clinicians outside of the initial or immediate treatment team, how will patient consent, privacy and confidentiality be handled? In our multicultural communities, how will NLP from patients of different linguistic histories or first languages affect efficacy, and how will sources of bias in the data be mitigated against?

Meanwhile, the applications of wearables in healthcare also have their challenges. For health system to adopt these technologies, they need to be medical grade requiring institutional approval, such as FDA. Even though consumers’ grade devices can still lead to positive effects, such as step tracking to promote exercises, the variability of the different step trackers in the market place [26] make these devices difficult to fully be deployed in a health context. As more wearable devices come to the market that are medically certified, their application will likely be more pervasive.

Even though there are many observational studies and case series about the use of wearables in healthcare that are promising, convincing evidence, such as randomized controlled trials, have so far not demonstrated unequivocal health benefits for patient management [27]. While many of these studies focus on outcomes in health, some authors argue that the patient experience and health system benefits should also be taken into account [28].

VII. CONCLUSION

This paper has identified a rapidly evolving domain of research and areas of exploration to this field. Beyond research, we also need to identify approaches towards change management and knowledge translation at this early stage of NLP and wearables in healthcare, so that when evidence of their benefits are clearly demonstrated, adoption of these practices will not be prolonged to delay reaping of the patients and health system advantages in using these technologies. At the same breath, we should not enthusiastically adopt these technologies now without judicious research and analysis to truly tease out their benefits, so as to ensure evidence-informed policy translation in incorporating these technologies into health system provision of care.

A governance, education, and ethics body, aimed at researchers, should collate and disperse general information regarding issues and concerns of AI in health, including the application of NLP and wearables. This includes best practices and policies regarding analysis of data. Such an organization would serve as i) a central hub for knowledge, hosting common issues, their solutions, and ii) an active educational force, holding workshops, and developing curricula.

The potential impact of NLP in healthcare is massive and largely untapped. However, with this potential good comes a lot of potential harm, and researchers should devote energy to ensuring that minimal harm befalls their patients and society at large.

REFERENCES

[1] X. Liu et al. “Proactive Suicide Prevention Online (PSPO): Machine Identification and Crisis Management for Chinese Social Media Users With Suicidal Thoughts and Behaviors,” *J Med Internet Res*, 21(5):e11705, 2019. doi: 10.2196/11705

[2] S. Pyysalo, F. Ginter, H. Moen, T. Salakoski, and S. Ananiadou, “Distributional semantics resources for biomedical text processing,” In LBM 2013, pp. 39–44, 2013.

[3] Y. Wang et al., A comparison of word embeddings for the biomedical natural language processing. pp.1–21, 2018. URL http://arxiv.org/abs/1802.00400 [retrieved: January 2020].

[4] S. Dubois and N. Romano, Learning effective embeddings from medical notes. Technical report, 2017. URL https://web.stanford.edu/class/cs224n/reports/2744372.pdf [retrieved: January 2020].

[5] Xmdv Tool. Ohsumed DataSet, URL http://davis.wpi.edu/xmdv/datasets/ohsumed.html [retrieved: January 2020].

[6] B. Chiu, G. Crichton, A. Korhonen, and S. Pyysalo, “How to train good word embeddings for biomedical NLP,” Proceedings of the 15th Workshop on Biomedical Natural Language Processing, pp. 166–174, 2016.

[7] P. Nguyen, T. Tran, N. Wickramasinghe, and S. Venkatesh, “Deepr: A convolutional net for medical records,” *IEEE Journal of Biomedical and Health Informatics*, 21(1), pp.22–30, 2017.

[8] Quantified Self. Quantified Self: Self knowledge through numbers. URL http://quantifiedself.com/ [retrieved: January 2020].

[9] Quantified Self. Dispatch from QS Dublin: Results from a QS Community Survey. URL http://quantifiedself.com/2017/02/quantified-self-health-wellness-now-future-dispatch-qs-dublin/ [retrieved: January 2020].

[10] R. M. Khandwalla et al., “Predicting heart failure events with home monitoring: use of a novel, wearable necklace to measure stroke volume, cardiac output and thoracic impedance,” *J Am Coll Cardiol* 2016;67:1296. doi:10.1016/S0735-1097(16)31297-9

[11] H. Brown, J. Terrence, P. Vasquez, D. W. Bates, and E. Zimlichma, “Continuous monitoring in an inpatient medical-surgical unit: a controlled clinical trial,” *Am J Med* 2014;127:226–32. doi:10.1016/j.amjmed.2013.12.004

[12] Healthcare IT News. Apple unveils Watch Series 4 with FDA-approved ECG. URL https://www.healthcareitnews.com/news/apple-unveils-watch-series-4-fda-approved-ecg [retrieved: January 2020].

[13] S. Gehrmann et al., Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. *PLoS ONE*, 13(2): pp.1–19, 2018.

[14] J. Bai, L. Shen, H. Sun, and B. Shen, “Physiological Informatics: Collection and Analyses of Data from Wearable Sensors and Smartphone for Healthcare,” *Adv Exp Med Biol*, 1028:17-37, 2017. doi: 10.1007/978-981-10-6041-0_2

[15] H. Hirst, “Keynote: NLP ethics in a dangerous time,” In Proceedings of the First ACL Workshop on Ethics in Natural Language Processing, 2017.

[16] J. Anna, D. Hovy, and A. Søgaard, "Challenges of studying and processing dialects in social media," In Proceedings of the Workshop on Noisy User-generated Text, pp.9-18, 2015.

[17] D. Hovy, “Demographic factors improve classification performance,” In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 1, pp. 752-762, 2015.

[18] R. Tatman, “Gender and dialect bias in YouTube's automatic captions,” In Proceedings of the First ACL Workshop on Ethics in Natural Language Processing, pp. 53-59, 2017.

[19] J. Ebrahimi, A. Rao, D. Lowd and D. Dou, “HotFlip: white-box adversarial examples for text classification,” In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistic. 2, pp. 31-36, 2018.

[20] M. Fredrikson, S. Jha, and T. Ristenpart, “Model inversion attacks that exploit confidence information and basic countermeasures,” In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pp. 1322-1333, 2015.

[21] A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” In Security and Privacy, IEEE Symposium, pp. 111-125, 2008.

[22] J. Rotblat, “A Hippocratic Oath for scientists,” *Science*, 286, pp.1475-1475, 1999.

[23] D. Hovy and S.L. Spruit, “The social impact of natural language processing,” In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2, pp.591-598, 2016.

[24] S. Kiritchenko and S. Mohammad, “Examining gender and race bias in two hundred sentiment analysis systems,” In Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, pp. 43-53, 2018.

[25] T. Bolukbasi, K.W. Chang, J. Y. Zou, V. Saligrama, and A.T. Kalai, “Man is to computer programmer as woman is to homemaker? debiasing word embeddings,” In 30th Conference on Neural Information Processing Systems, 2016.

[26] P. Alinia et al., “How accurate is your activity tracker? A Comparative Study of Step Counts in Low-Intensity Physical Activities,” *JMIR mHealth uHealth* 2017;5:e106. doi:10.2196/mhealth.6321

[27] A. Vegesna, M. Tran, M. Angelaccio, and S. Arcona, “Remote patient monitoring via non-invasive digital technologies: A Systematic Review,” *Telemed e-Health* 2017;23: pp.3–17. doi:10.1089/tmj.2016.0051.

[28] P. Zanaboni et al., “Methods to evaluate the effects of internet-based digital health interventions for citizens: Systematic review of reviews,” *J Med Internet Res* 2018;20. doi:10.2196/10202.