Technology at the service of life

Loris Giulivi is a Ph.D. student at Politecnico di Milano. He has been working on deep learning research, with a particular focus on Explainable Artificial Intelligence.

Executive summary

This project aims at developing a system to help medical staff in school programs related to mental health. The core of the project consists in designing machine learning models able to identify mental illness through survey and medical record data. These models will be part of a more comprehensive system, to be used by medical professionals, that enables to monitor the students' mental health.

Objectives

- Design and train interpretable machine learning models to analyze survey and medical record data.
- Develop, in collaboration with medical staff, a comprehensive monitoring system, to be used in the context of school programs related to mental health.

Introduction

Mental health is paramount to having a good quality of life, especially for young people, whose mental well-being is likely to be carried into adulthood [1]. We often look at childhood and adolescence as times of freedom, exploration, maybe a few heartbreaks, but seldom we give weight to the turmoil that may be developing in a teenager's mind. Indeed, results from [2] demonstrated that over **half of all adults with a mental health problem had been diagnosed in childhood, with less than half of them having received the appropriate treatment** at the time of diagnosis.

Without treatment or intervention, many teenagers yield to the pressure of their mental health problems, resulting in self-harm and ultimately suicide. According to the World Health Organization [3], more than 700.000 people die by suicide each year, and for each suicide, there are twenty unsuccessful suicide attempts. Even though most victims are adults, **suicide is second in the leading causes of death in teenagers**. As much as we want to keep it distant from ourselves, the effects of suicide ripple through families and communities, so much so that it increases risks also for those close to the deceased [4]. This phenomenon, often hidden behind social stigma, is avoidable.

Proposed system

Treating mental health problems is possible, at individual, community, and national levels, and at different points during the development of the illness. With this project, we **aim at making use of Machine Learning (ML) to aid in the monitoring of mental health conditions in adolescents and young adults**. The final expected product is a system composed of models able to identify people that may be developing or suffering from a mental illness, and of a series of interfaces through which data can be collected and results can be analyzed. These models will be designed to be explainable, as such, results will be accompanied by descriptions of the model's reasoning. **Ultimately, this system may be used by medical professionals in school programs that focus on the students' mental health**.

Background

Mental illness is a type of health condition that affects a person's mind, emotions, and/or behavior. These include depression, schizophrenia, attention deficit disorder (ADHD), autisms spectrum disorders (ASD), and many more. Such disorders are highly prevalent, with an estimated 450 million people being affected worldwide [5].

Differently than other chronic conditions, which can typically be diagnosed through laboratory tests, **the diagnosis of mental illness relies on individual self-reports** [6], which are then analyzed by professionals. This process is straining for the health provider, who must provision expertise for lengthy periods. At the same time, however, the **data collected through self-report has the benefit of being ingestible by modern artificial intelligence algorithms** such as deep neural networks.

Machine Learning in mental health

Previous literature has attempted to utilize several data sources to extract patient information regarding their mental health. These include works that exploit *a)* clinical data, *b)* genomic data, *c)* vocal/expressive data, and *d)* social media data. In our monitoring scenario, we focus on non-invasive clinical evaluation of each patient, therefore, we are restricted to using clinical data that is either already available or that can be collected through self-report. Works such as [7] show that it is possible to predict future outcomes of depressive episodes from data in electronic health records (EHR). Moreover, EHRs have also been used in the prediction of suicide deaths [8]. These works make use of deep neural networks, highlighting the effectiveness of this kind of models in extracting patterns that relate to mental health. However, these results have been achieved by making use of basic models and without providing explanations for the model's results. **With this project, we aim at expanding on these concepts, and in turn exploit the full potential of deep models**. This can be achieved by implementing state-of-the-art ML models that make use of:

- Structured data (such as EHR or Likert-scale-based surveys).
- Unstructured data (such as text from open-ended answers).

Explainable Artificial Intelligence

An open challenge that hinders the application of deep neural models, especially in critical fields such as the medical one, is their interpretability. Indeed, deep neural networks are complex models that operate as black boxes, whose decisions are hard to understand. **Black-box models cannot and shouldn't be blindly trusted**, as this would expose several safety risks. In the context of this project, application of black-box models would allow erroneous outputs to go unnoticed, with potentially dangerous effects. Moreover, correct predictions might also be refused by a professional if not substantiated by an explanation. Consequently, users might be deterred from trusting such a system.

In this project, we plan on designing models whose decisions are understandable. **In recent years, the field of explainable artificial intelligence (XAI) has proposed several methodologies to solve this problem**. The two main directions to obtain understandable model decisions are:

- Generating truthful explanations to existing models.
- Developing models that are interpretable by design.

For what regards the data domain of interest for this project, current XAI literature presents works spanning from the simplest partial dependence plots [9], that explain how different

features affect the model's output, to explanations of natural language processing (NPL) tools, such as those described in [10]. XAI techniques, when applied to our models, will allow medical staff to validate correct results and identify potential model mistakes, ultimately allowing to perform the correct diagnosis.

Mental health in schools

As stated in the introduction, **the focus of this project is monitoring the development of mental illness in children and teenagers. For this age group, the most effective environment in which to deploy these campaigns is in schools**. Indeed, the school community is suggested as the ideal forum for programs related to mental health [11]. A few initiatives around the globe have embraced this spirit, such as [12], showing how pervasive mental illness can be, and highlighting the need for more thorough prevention programs, such as the one proposed in this project.

Methods

The core of this project is realized in a series of interpretable models to be used by a medical professional to evaluate the mental well-being of patients subject to survey-based monitoring. The key ingredients that need to be defined are:

- The surveys and their corresponding administration strategies.
- The interpretable models that analyze the data.
- The interfaces through which data is acquired and the results are analyzed.

Survey

For what regards the survey details, we plan to rely on external expertise to craft questions that expose the most relevant patient information, while still preserving privacy. Indeed, the design of a mental health evaluation survey must be directed and supervised by experienced medical staff to ensure appropriateness of the contents. We expect survey responses to contain data in heterogeneous form, including categorical data (for multiple choice and Likert scale questions), text data (for open-ended questions), and possibly numerical data. Additionally, data from EHRs could be included to improve model accuracy. Ideally, these surveys will be administered as part of long-term or periodic mental health programs in schools, such that the development of mental illness can be analyzed through longitudinal studies.

Models

We propose two approaches to model design, which may also be used in conjunction. The first regards anomaly detection, the second regards classification. With the first approach, we construct models that can **identify patients that present anomalous characteristics**. The deep learning techniques most suitable for this scenario are those related to heterogeneous ensemble learning, such as [13]. In this setting, anomaly detection techniques could be used together with XAI methodologies to explain why a particular subject was deemed to be anomalous. Importantly, anomaly detection can be performed in an unsupervised manner, which means that the data doesn't need to be labeled as normal or anomalous. This significantly reduces the time needed to craft a dataset to be used to train the model.

For the classification approach, models would be able to recognize the specific illness a patient is suffering from. A large set of labeled data is needed for supervised training. EHR datasets, such as those used in [8], might be a starting point for the development of the

classification models. Nonetheless, a data collection and labelling effort would still be required to obtain the best performance. Moreover, since mental health is a many-faceted topic, categorization and classification of every illness might not be possible. In this setting, the techniques of choice are open set recognition [14] and zero-shot learning [15]. Like traditional classification networks, such models can classify data into pre-defined categories, additionally, these methodologies allow to also identify new unseen classes. Furthermore, by introducing prototype-based learning, such as in [16], these models could be rendered interpretable, allowing to obtain results which are naturally human-understandable.

For both approaches, time series information could be used to further improve the performance of the models, for example, by making use of change detection techniques [17]. This highlights the importance of long-term monitoring programs. Indeed, given a sequence of data points (survey responses) from a single person, these models would be able to identify trends in the responses, and better characterize the patient's illness, if any.

Interfaces

The last element of the system to be defined are the interfaces through which the models can be operated. This matter is twofold, as **interfaces are to be provided for data collection and results analysis**. We expect the former to be relatively straightforward, as survey answering applications are widely available. Nonetheless, collaboration with UX engineers is needed to obtain optimal results. A more significant effort is instead required for the development of the interface through which results and explanations are given. Indeed, this human-machine point of contact is critical in the delivery of the information processed by the model. A suboptimal design of this interface may substantially alter the user's behavior in response to the model's output. For these purposes, **XAI techniques could be compared and evaluated in user studies** where target users (medical professionals) would be asked to utilize the systems. The configuration that is easiest to interpret would then be used in the final product.

Development and deployment

In the previous section, we have detailed the three main elements constituting this project: the models, the survey, and the interfaces. The development of these components is expected to be composed of four main phases:

1) Survey design, including preliminary studies regarding the administration strategy.
2) Model design and data collection.
3) Interface and application design, comprising user studies on the effectiveness of the XAI techniques.
4) System integration and deployment in pilot school programs.

Clearly, these stages are not independent, as such, **a *waterfall* development process** [18] **is not applicable**. Interface design may significantly affect how the models are constructed, especially based on feedback on the XAI techniques employed. Furthermore, first deployments of the system to pilot school programs may uncover flaws in the survey design and trigger further developments. Nonetheless, we propose below a possible timeline for the development of the project and subsequent deployment in a pilot study. This would be useful in the evaluation of the proposed methods, even though it would be far from an effective monitoring strategy, which would instead require long-term efforts.

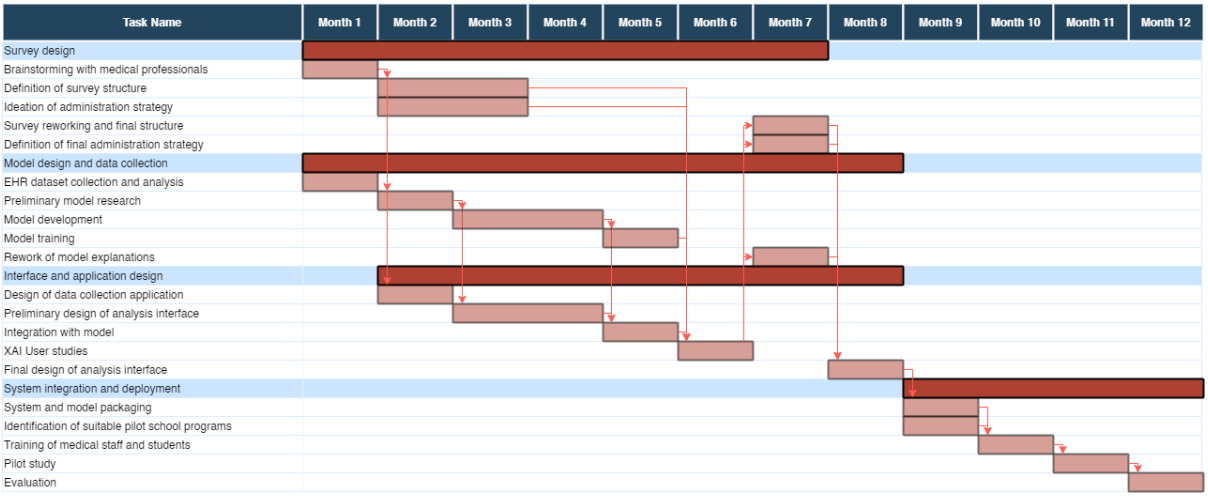

Figure 1 - Timeline for development and deployment of the project on a pilot study

Importantly, constant collaboration and oversight from medical staff is required throughout the development. Indeed, to guarantee success of the project, we must ensure that all parts are compliant with specific protocols that apply in the medical domain, including, but not limited to, legal and safety requirements.

Conclusion

In this work, we propose a new system aimed at monitoring mental status in children and teenagers, with the ultimate purpose of aiding medical staff in the context of mental health programs in schools. The project consists in the realization of a system comprised of deep learning models able to identify and classify people that suffer from or are developing a mental illness. Moreover, we propose integration of these models in an explainable artificial intelligence scenario. In this context, medical professionals making use of the system will be able to understand and validate the model's results and act accordingly. Furthermore, this system will be developed in collaboration with medical staff in order to achieve compliance with legal and ethical requirements. Ultimately, we expect this system to allow broader mental health programs than what is currently available, enabling improvements to the prevention of mental illness amongst the most vulnerable groups.

Bibliography

[1] WHO, "Adolescent Mental Health - Mapping actions of nongovermental organizations and other international development organizations," 2012. [Online]. Available: http://tinyurl.com/j2yzanq.

[2] T. Kim-Cohen, A. Caspi, T. Moffit, H. Harrington, B. Milne and R. Pulton, "Prior juvenile diagnoses in adults with mental disorder: developmental," *Arch Gen Psychiatry,* 2003.

[3] WHO, "Suicide Prevention.," 2 February 2022. [Online]. Available: https://www.who.int/health-topics/suicide#tab=tab_1.

[4] M. Pompili, P. Girardi, A. Ruberto, G. Angeletti and R. Tatarelli, "Stigma E Rischio Di Suicidio," *Psichiatria e Psicoterapia,* pp. 36-47, 2006.

[5] WHO, "The World Health Report 2001: Mental Health: New," The World Health Report 2001: Mental Health: New, Geneva, 2001.

[6] M. Hamilton, "Development of a rating scale for primary depressive illness," *J. Soc. Clin. Psychol.,* p. 278–296, 1967.

[7] T. Pham, T. Tran, D. Phung and S. Venkatesh, "Predicting healthcare trajectories from medical records: a deep learning approach," *J. Biomed. Inform.,* p. 218–229, 2017.

[8] S. Choi, W. Lee, J. Yoon, J. Won and D. Kim, "Ten-year prediction of suicide death using Cox regression and machine learning in a nationwide retrospective cohort study in South Korea.," *J. Affect. Disord.,* p. 8–14, 2018.

[9] T. Hastie, R. Tibshirani and J. Friedman, Elements of Statistical Learning Ed. 2, Springer, 2009.

[10] A. Joshi, P. Bhattacharyya and M. J. Carman, "Automatic Sarcasm Detection: A Survey," *ACM Computing Surveys,* p. 1–22, 2017.

[11] J. Wells, J. Barlow and S. Stewart-Brown, "A systematic review of universal approaches to mental health promotion in schools.," *Health Education,* p. 197–220, 2003.

[12] CDC, "Mental Health Surveillance Among Children," U. S. Centers for Disease Control and Prevention, 2013.

[13] Y. Zhong, W. Chen, Z. Wang, Y. Chen, K. Wang, Y. Li, X. Yin, X. Shi, J. Yang and K. Li, "HELAD: A novel network anomaly detection model based on," *Computer Networks,* p. 169, 2020.

[14] C. Geng, S. J. Huang and S. Chen, "Recent advances in open set recognition: A survey," *IEEE transactions on pattern analysis and machine intelligence,* pp. 3614-3631, 2020.

[15] W. Wang, V. Zheng, H. Yu and C. Miao, "A survey of zero-shot learning: Settings, methods, and applications," *ACM Transactions on Intelligent Systems and Technology,* pp. 1-37, 2019.

[16] C. Chen, O. Li, D. Tao, A. Barnett, C. Rudin and J. Su, "This looks like that: deep learning for interpretable image recognition," *Advances in neural information processing systems,* p. 32, 2019.

[17] S. Aminikhanghahi and D. Cook, "A survey of methods for time series change point detection," *Knowledge and information systems,* 2017.

[18] K. Petersen, W. Claes and B. Dejan, "The waterfall model in large-scale development," *International Conference on Product-Focused Software Process Improvement,* 2009.