Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 96 (2016) 577 – 586

Procedia Computer Science

20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2016, 5-7 September 2016, York, United Kingdom

Introducing a pilot data collection model for real-time evaluation of data redundancy

Ali Rezaei-Yazdi, Christopher Buckingham

Aston University, Aston Triangle, Birmingham, United Kingdom, B4 7ET

Abstract

In order to reduce serious health incidents, individuals with high risks need to be identified as early as possible so that effective intervention and preventive care can be provided. This requires regular and efficient assessments of risk within communities that are the first point of contacts for individuals.

Clinical Decision Support Systems CDSSs have been developed to help with the task of risk assessment, however such systems and their underpinning classification models are tailored towards those with clinical expertise. Communities where regular risk assessments are required lack such expertise. This paper presents the continuation of GRiST research team efforts to disseminate clinical expertise to communities. Based on our earlier published findings, this paper introduces the framework and skeleton for a data collection and risk classification model that evaluates data redundancy in real-time, detects the risk-informative data and guides the risk assessors towards collecting those data. By doing so, it enables non-experts within the communities to conduct reliable Mental Health risk triage.

Peer-review under responsibility of KES International

Keywords: Dynamic data collection, Clinical Decision Support Systems, Risk assessment, Risk classification, Healthcare, Mental Health

1. Introduction

Effective risk assessment is vital for enhancing health care delivery and ultimately patient outcomes particularly in the mental health sector[1]. Various studies show that failure to detect health problems in the early stages of development results in higher mental-health risks including suicide[2]. On the other hand, early detection and effective preventive measures in psychosis services have high impacts in reducing the probability of mental health risks and the risk of suicide in particular[3 4] .

The problem is that risk assessment and the prediction of the level of risk while clearly extremely important, is difficult[4] and a major clinical challenge[5] and requires clinical expertise. This is for various reasons including the

* Corresponding author: Ali Rezaei-Yazdi, Aston University, Aston Triangle, Birmingham, UK, B4 7ET. Email: rezaeiya@aston.ac.uk.
** This work was supported through the Aston Research Centre for Healthy Ageing (ARCHA).

Peer-review under responsibility of KES International
doi:10.1016/j.procs.2016.08.237

complexity of factors influencing the level of risk, rarity of the serious incidents[6,7] fuzzy nature of many health symptoms and risks, missing data,[8,9] and the time that is needed for accurate risk assessments[10,11].

In addition to the general difficulties of risk prediction, effective risk assessments are those with which individuals at high risk are identified and monitored as early as possible so that preventive measures are provided in time. In other words, the risk assessments' objective should be to predict, not just for example who successfully will commit suicide, but who is likely to try and attempt suicide. The focus should be to stop any attempt and this is only possible if potentially high risk individuals are identified early enough and hence referred to specialist for detailed monitoring and assessment. This means that risk assessment needs to be conducted on a community level that is the first point of contact for most people[12]. Community level include social workers, 111 paramedics and other front line services which do not have specialist clinical expertise but are faced with patients on daily basis.

To enable risk assessment to be carried out on a community level, the clinical expertise necessary for risk assessment and prediction need to be made available to non-experts. This is the aim of this research paper: Building on top of our initial research started in 2014 and reflected in a previous paper[13] this paper aims to shed further lights towards enabling non-experts to use expertise extracted from clinicians to conduct fast but reliable risk triage. It does that by introducing and testing the initial skeleton of a risk triage data collection and risk classification model. Thus the objective is to probe the appropriateness of our previous research as indicated in our previous paper[13] and taking the first step towards an ultimate data collection and risk classification model. The final model is aimed to guide non-experts risk assessors through huge pools of possible data to collect only those data that are relevant and informative for the purpose of quick safe triage that is specific to the particular patient under assessment.

We first describe the context including the Clinical Decision Support System based on which our research was conducted as well as the former findings which form the conceptual framework for this paper. After that comes the methodology followed by introducing our initial data collection and risk classification model and evaluating its performance. The paper ends with describing the future work.

2. Context and the Conceptual Framework

Clinicians need to be able to carry out varied level of risk assessments based on patients circumstances and context. In each case, they require a different set of data to be collected in a different way with perhaps different order of collection or importance. This implies that clinicians are faced with so much data from different sources, different patients and different contexts with ever evolving information[14] and their judgement may even then be affected by factors beyond the patients presentation (i.e. the data collected)[15].

To aid clinicians with the task of risk assessment, Clinical Decision Support Systems CDSSs have been developed to incorporate research and evidence-based approaches and techniques into daily practice and hence aid clinicians in their clinical work and in particularly clinical decision making[16,17,18].

Galetean Risk & Safety Tool GRiST is a mental health Clinical Decision Support System that aims to assist the early detection of multiple risks, including suicide, self-harm, harm to others, self-neglect, and vulnerability amongst people with mental health problems[19]. It is built using a psychological model of human expertise that attempts to approach risk assessment in the way clinicians do. It aims to combine human expertise together with actuarial tools to help mental health clinicians in their daily risk assessment tasks. GRiST is used by various sectors including NHS secondary Mental Health Trusts, charities as well as IAPT services. In the future it will also be available for use in other front line agencies such as the criminal justice system, housing associations, accident and emergency departments.

This paper presents part of an effort by the GRiST research team to disseminate clinical expertise to the wider community by working towards a critical risk screening version of GRiST to be used by non-experts risk assessors on daily basis for fast risk triage.

GRiST has been developed based on a specific psychological model of classification in which clinical expertise is represented in a way whereby risk nodes such as suicide are hierarchical trees that are deconstructed into progressively more granular concepts (branches) such as current intention and feelings/emotions until the input data nodes or leaves of the tree (e.g. anger) are reached. Figure 1 shows a simplified portion of the suicide knowledge hierarchy.

While GRiST has several hundred leaf nodes, and therefore a very large potential data set for each patient (137 data features only for risk of suicide), the actual number of questions that is applicable to individual patients is much

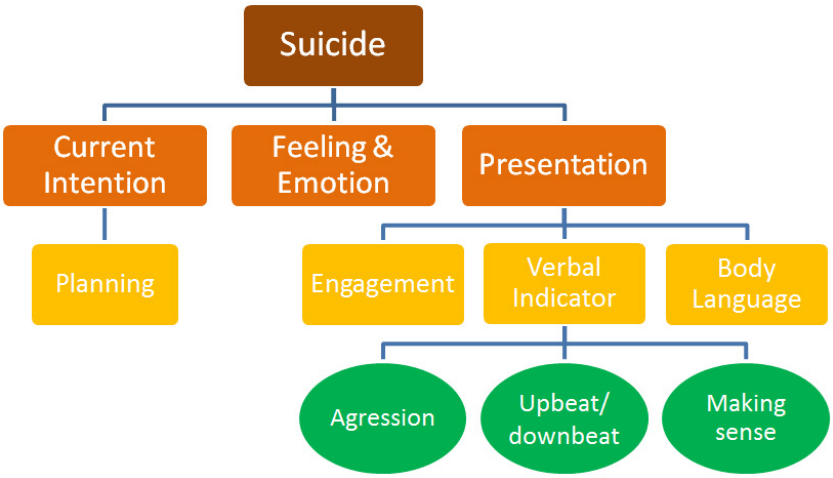

Fig. 1. Part of the suicide knowledge tree.

lower because the tree provides a top-down gateway to the the patient data by a series of layer questions that determine whether or not a particular branch of the tree, and thus its associated leaf node questions, are relevant for that patient.

For example, there is a question asking whether the risk assessor is concerned with patient's *Presentation during the assessment* (as shown in Figure 1). If not, this part of the tree is immediately closed off with no further exploration. Hence the knowledge hierarchy imposes constraints on the order of asking questions because the branch questions have to be asked before the leaf questions are reached. When assessors have finished asking questions, they are asked to provide their overall risk judgement for the patient in the form of a score ranging from 0 (no risk) to 10 (maximum risk).

Therefore, the idea is that not all the branches and parts of the tree are relevant for the patient under assessment particularly for the initial triage of risk and this is easily witnessed by the fact that some branches of the tree are left unanswered for patients. Therefore, the task is to find out which questions are risk informative for which patients and which questions can be made redundant.

2.1. ***Previous conceptual findings forming the basis for this research***

Our initial studies as presented in our paper in 2014 [13] and the subsequent studies provided evidence that the patients emerging profiles have a crucial influence on what data become relevant and risk informative. In other words, the risk informativeness of data (i.e. assessment questions) evolves as more data are disclosed (i.e. questions are answered) about the patients. For example, our paper [13] showed that the branch question 'Past Suicide Attempt' is redundant for the purpose of risk triage when clinicians answered YES to another branch question 'Current Intention'. However if clinicians give a NO answer to Current Intention, then the question on Past Suicide Attempts become informative and hence required for the triage of risk.

In particular, we found that while a group of questions are important for the triage of risk in all context, clinicians do not base their risk judgement on only a this small set of questions. In contrast, depending on the context which is dictated by the collected data (i.e. answers to assessment questions), different set of questions can become informative or redundant and hence required for the triage of risk. Secondly, the perception of clinicians on the knowledge attained from data depend on all other data they collect during the assessment. In other words, they do not evaluate the informativeness of a question in isolation from other questions. Hence a previously informative piece of information can potentially become completely redundant for the evaluation of risk because clinicians feel that the new information adds little or nothing to their perception of the level of risk. Similarly, a piece of data that is made redundant by the clinicians at a certain point during the assessment can potentially become very informative later on during the same assessment. Figure 2 demonstrates such dynamism in the informativeness of data. As this figure symbolically shows, the informativeness variability in data is similar to electrocardiogram ECG of the heart rate: It can become flat

which means death of informativeness (i.e. complete redundancy) and surprisingly it can comeback to life (i.e. high informativeness) at a later point during the assessment.

Fig. 2. Data informativeness life line.

The question therefore is to find a mechanism that guarantees the best set of questions and the best order for the questions to be asked. However, due to this dynamic and unfixed nature of data redundancy, the optimal order in which the questions need to be asked to maximise the accuracy of prediction will have to be decided dynamically.

Using these conceptual framework as indicated in [13], in this paper and a subsequent paper (that is awaiting publication), we investigate the usefulness of our findings by proposing and testing an initial skeleton of a data collection and risk prediction model that attempt to dynamically guide non-experts towards informative data by detecting redundant data in real time and hence producing a safe triage of risk.

This study focused on the risk of suicide and used 30,000 cases of completed risk assessments provided by mental-health practitioners through their normal practice. The data are already anonymous by the collection process, which means they do not need any additional anonymisation to remove personal identification information, and ethics approval was obtained both from the health services and the university[1].

It should also be noted that this study focused only on branch and sub-branch questions of which there are 28. The reason for focusing on only branch and sub-branch questions is because they are the most abstract intuitive questions which can be comprehended and dealt with by non-experts. Detailed low level questions might need clinical expertise and hence should not be used for triage of risk by non-experts.

3. Methodology

The task is to work out the amount of informativeness or similarly the redundancy of the questions in terms of how much knowledge they can give about the risk associated with the patients. The aim is to find the questions and the most optimal order that would result in most knowledge being given about the patients' level of risk.

To quantify the magnitude of the level of informativeness, a Bayesian approach was taken to analyse how accumulating evidence influenced the evolving suicide risk evaluations. The centre of focus was on the concept of conditional probability. The conditional probability of an event 'A' given another event 'B' is denoted in this paper using the following format:

$$P(A|B) \text{ reads as probability of A given B.}$$

The reason for using the concept of Bayesian probabilities is because this is what human cognitively use and are influenced with in their daily life when making decisions[20]. This is why probability approaches and particularly the Bayes theorem are among the most widely employed approaches towards clinical decision making[21,22]. Bayes theorem describes how the degree of belief in an event should change to account for evidence. In simple words, how

[1] Ethics approval was obtained from NRES Committee East Midlands, 13/EM/0007, and Aston University.

the evidence changes the degree of certainty (in this case probability) in the event. In the context of this research it helps us to understand how one piece of data that is in the form of an answer to a question (i.e. the evidence) changes the degree of our certainty about the probability of risk. Understanding this change in the amount of certainty or probability is crucial in determining the informativeness of data and hence usefulness of assessment questions.

Therefore, the driver for computing the level of informativeness is how much the answer to a question changed the prior probability of risk, with questions being favoured the bigger the difference in posterior probability. For example, our analysis shows that Current Intention (CI) causes a much larger change (denoted by Δ) in the probability of High Risk (HR)[2] in comparison to Presentation (PR) and hence CI has a higher level of informativeness:

$$P(HR) = 0.05, P(HR|CI) = 0.37 \implies \Delta P = 0.32$$

$$P(HR) = 0.05, P(HR|PR) = 0.22 \implies \Delta P = 0.17$$

The method is not quite as simple as described so far because each branch can have two answers: YES or NO. The answers will not have the same influence on the posterior probability; the 'YES' answer may cause a large change with the 'no' answer having no effect at all or vice versa. Hence a question may only be worth asking if one of the answers is given and the order of questions need to weigh up the probability that the influential answer is actually given when determining which question to ask. This means that in addition to the amount change ΔP the questions cause on the probability of risk, the actual probability of the answers to the questions needs to be taken into account.

Furthermore, as mentioned before, our analysis reflected in our 2014 paper[13] and the subsequent studies showed that a question level of informativeness (i.e. the change it causes on risk) is dependent on all the questions that have been asked before it. This is due to the fact that the knowledge that is disclosed by answering a question might be embedded (to some degree) by answer to another question. For example, our analysis showed that the question on Personality (PER) increases the probability of High Risk by 0.04, however if the question on Presentation (PR) has been asked prior to Personality, then the change Personality cause on the probability of High Risk is 0.0 meaning that Presentation contain the knowledge that Personality has on risk:

$$\left.\begin{array}{l} \text{P(HR | PR)} = 0.22 \\ \text{P(HR | PER)} = 0.09 \\ \text{P(HR | PR} \cap \text{PER)} = 0.22 \end{array}\right\} \implies \text{PER is redundant if PR has been asked}$$

Putting these findings together, the function that evaluates the informativeness of questions needs to take into account three factors:

1. All questions which have been asked on the path so far
2. The change ΔP that the answers to the questions cause on the probability of risk (which will be dependent on 1)
3. The actual probability of the given answers (i.e. How probable are the answers to a question)

These three factors form the basis for our initial data collection and risk prediction model. In the next section we operationalise this model and its forming elements, we then implement and validate its accuracy.

4. Our Bayesian-based initial data collection and risk classification model

Operationalising informativeness of questions is achieved by setting up a function named as Evaluation Function (EF) taking into consideration the tree main factors discussed in the previous section i.e. answers to all previously collected questions, the question, Q, we might ask given those answers and the likelihood of the answers that are given to Q. The question evaluation function decides which path the assessment would take to gain the best possible knowledge before it stops.

[2] High Risk are risk scores of 7 and over

Each question in the context of our analysis has two answers YES or NO and the preferred question is the one that gives information about the risk irrespective of whether the answer is YES or NO. This means that the EF of a question, EF(Q), is determined by evaluating each answer, A, and summing across them. It needs to be reminded that the evaluation for an answer A is conditional on all the other data preceding A on the path. So A is really A& PathA, where PathA represents all the preceding questions along the path. What follows is the construction of our Evaluation Function by describing its elements.

1. **The change in probability of risk, ΔP**

   The predictive power of a question primarily comes from how much the answer to it can change the probability of risk; the more it can change the risk (whether increase or decrease it) the more powerful it is believed to be. As each question has two potential answers YES and NO with each causing a certain amount of change in the probability of risk, the EF for an answer A can hence be initiated to be the following:

   $$EF(A) = |P(R|A \ \& \ PathA) - P(R)|$$

   Where A is either YES or NO. This means that the EF for one question with two possible answers, YES and NO, will be:

   $$EF(Q) = EF(YES) + EF(NO) = |P(R|YES) - P(R)| + |P(R|NO) - P(R)|$$

   The EF of the questions can hence be generalised and formalised in the following format:

   $$EF(Q) = \sum_{i=1}^{A} |P(R|a_i) - P(R)| \tag{1}$$

   where A is the total number of different answers, a_i, for a question.

2. **The probability of each answer, P(A)**

   As explained before, the probability of the answers need to be taken into account when deciding which question to be asked next. An answer might be very informative but the probability of it actually being given by the assessor might be extremely low. This is why, the objective should be to ask the question that is most *likely* to provide a large change in posterior probability of the risk, which means taking into account the answer that is likely to be given. For example, if question QA has a change in high risk probability of 0.4 for YES but no change for NO and QB has a change in high risk of 0.2 for YES and 0.1 for NO, which one do we choose? If we knew the answer would be YES for QA, then this is the one to ask but if the probability of a YES response is 0.1 compared to 0.5 for QB, then QB looks like being a better bet, especially as both QB answers gives some change compared to only one for QA. In this way, the EF for an answer A can be improved to be the following:

   $$EF(A) = |P(R|A \ \& \ PathA) - P(R)| + P(A)$$

   The EF of the question, can then be improved to:

   $$EF(Q) = \sum_{i=1}^{A} \Big(|P(R|a_i) - P(R)| + P(a_i) \Big) \tag{2}$$

   where A is the total number of different answers, a_i.

3. **Statistical significant constraint**

   The question that is selected by the EF needs to be statistically significant in its dependency with the risk. In theory, it is possible that the risk probability is shown to be changed by a questions answer, however there would

be no evidence of such impact being statistically significant. We need to be sure that the change that a question causes on the probability of risk has not happened by chance. Therefore, statistical significance test is applied to each question when its EF is being calculated. If the question is not passing the significant threshold, then there is no point in considering this question as a candidate. For significant testing, we used Pearson chi-square test of independence with significant level of 0.05. The model evaluates the statistical significant of each question and will stop data collection when no question cause statistically significant change on the risk.

Using this Evaluation Function, the data collection model starts the assessment by ranking the questions based on their level of informativeness and choosing the most informative question to ask. Based on the answer to the first questions, it then re-evaluates the informativeness of the remaining questions and chooses the next informative question to ask. By asking each question and providing answers, the population is split towards more homogeneous risk groups (i.e. samples in which patients have as close risk scores as possible). The model carries on by re-evaluating and asking questions until no more pieces of data can prove to be informative. At the end of data collection, the patient will be in a sample from which a classification needs to be made.

4.1. ***Classification of risk and validating its performance***

Once the data collection has stopped, the risk classification of patient starts. At the end of data collection there are two scenarios. Scenario one is when the patient is in a sample where all cases belong to the same risk score (e.g. all are 7) in which case we have a perfect classification. Scenario two is when the patient is in a sample with different risk scores. In this case, we take the mean of risk scores to be the class label i.e. the classification.

We then evaluated the accuracy of the classification result. The results were measured as the accuracy percentage on two levels of 'Correct' classification and '±1 tolerance'. *Correct* classification means that the patient's risk is classified as having the exact same risk score which was given by the clinician (100% correct). The ±1 *tolerance* allows the classification to be assumed correct as long as it is within 1 class (i.e. risk score) from the actual risk score. This means a patient with risk score of 2 can be classified as 1, 2 or 3. This level of tolerance is firstly clinically acceptable since a granularity of 11 classes (risk scores 0 to 10 in GRiST) is higher than the semantic groups used by many classification tools to distinguish patient risk. Secondly, for our model which is intended for risk triage, the ±1 tolerance results are more significant than the exact ones because as mentioned, our ultimate classification model is looking to gain only a good approximation to the risk (risk triage), not an exact prediction, but it wants a good approximation for all patients. Therefore, for the objective of the risk triage, a less accurate 100% correct prediction with a higher accuracy for the ±1 tolerance is much more preferred to a more accurate 100% correct classification but a lower accuracy in the tolerance which means more error (i.e. those which are predicted outside the tolerance). Notice that the ±1 tolerance includes the 100% correct classifications and hence a less accurate result for the tolerance means a higher error rate.

Table 1 shows the result of our performance evaluation. As it shows, our initial model produces a rather decent classification of 0.78 for the ±1 category. This means with the use of only very high level questions (branch questions), this initial model classifies 78% of the cases correctly within the tolerance limit.

Table 1. Model Overall Classification Performance

100% Correct	**±1 Correct**	**Error**
0.34	0.78	0.22

In the next stage, we wanted to know for which group of patients (i.e. high or medium or low risk individuals) our model performs better. In other words, we wanted to know the error rate of 0.22 as shown in table 1 refers to which group of patients. Our ultimate final model is (and will be) primarily aimed at detecting high risk individuals and therefore it is important for our model *not* to classify high risk individuals incorrectly. Similarly it is important for our model to classify low risk individuals correctly too because correctly classifying low risk means confidently ruling out high risk.

To achieve this, we conducted a risk by risk classification evaluation to see how well our model predicts individual risk scores. The result can be found in Table 2, and Figure 3 displays the result in the form of a graph.

Table 2. Model classification performance on risk level by level basis

Risk	Frequency	100% Correct	±1 Correct
1	3567	0.35	0.83
2	16911	0.33	0.84
3	3125	0.38	0.78
4	3142	0.28	0.74
5	1337	0.28	0.68
6	1259	0.19	0.62
7	483	0.36	0.77
8	150	0.48	0.88
9	26	0.43	0.85

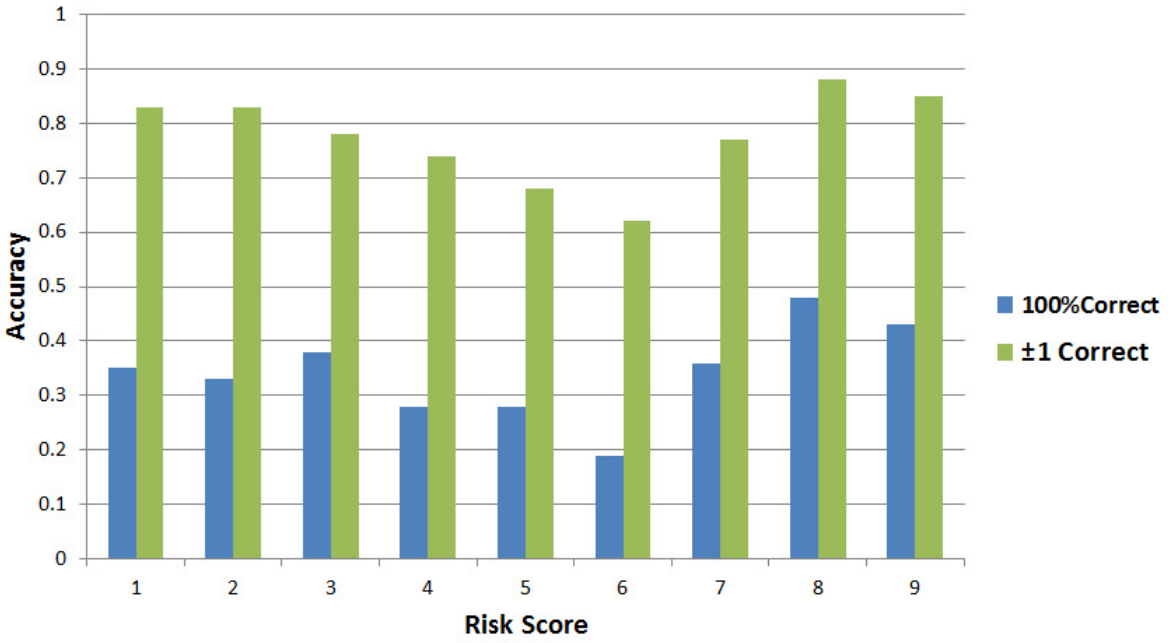

Fig. 3. Prediction Accuracy of our pilot model at risk-by-risk level.

As Table 2 and Figure 3 show, our classification model performed pretty well on detecting the high risk and low risk cases (i.e. risk scores of 7 and over for high risk and risk scores of 3 and below for low risk). As the results demonstrate while this initial model performs not particularly well in detecting medium risk cases, it performs significantly better in detecting low risk and high risk cases. This is a very promising result for our ultimate risk triage model that is currently under development because as mentioned before the aim of our final model is to detect high risk individuals so that effective referral can be made to clinical specialists. Thus it is important to be able to comfortably rule out or rule in high risk patients. By classifying the *low* risk patients correctly most the time, we are comfortably ruling out the possibility of missing high risk individuals. Similarly, by classifying *high* risk individuals correctly we make sure that high risk patients are actually being recognised and referred.

Although the classification performance is not high all the time, however the result indicates a potential for a successful data collection and risk classification model once our final evaluation function is designed. This is because we have already seen a pretty decent performance taking into consideration the constraints we have on our model in comparison to ordinary classification models. Most classification models are post-hoc analyses whereby the whole data set have already been collected and available and hence those models have access to a large pool of data and have the freedom of choosing whichever features they like and as many as they need. This is different to our risk

triage model: Our model is not post-hoc which means it does not know in advance how questions will be answered. Also, this model does not have access to every single possible data because it simply does not (and does not want to) collect all data (otherwise it contradicts its purpose). Therefore, for making risk prediction, this model is constrained by those data which it collects which means it needs to make its prediction using only the data it collects along the path. Furthermore, our initial risk triage model only collects high level branch and sub-branch questions which are easy and intuitive to manage by the non-experts. This is different to ordinary data-driven models where the model has access to detailed low level data features which potentially can provide more information on risk.

Overall, our initial data collection and risk classification model confirms the suitability of our previous analysis as described in our 2014 paper and the probabilistic approach adopted by this research. Our model reduces the data dimension from 175 questions to a maximum of 10 questions while keeping the risk prediction accurate enough for a safe triage to be made. Furthermore, This model produces a risk triage using only the high level intuitive data, collection of which does not need clinical expertise.

5. Conclusion & future work

This paper aimed to produce the skeleton of a data collection and risk classification model for mental health risk triage to be used by the non-experts based on the former analyses which had been done on risk assessments collected by Galatean Risk & Safety Tool GRiST. The analyses had shown that the redundancy of data features (i.e. usefulness of assessment questions) has to be decided dynamically as the level of redundancy and informativeness of data features evolves throughout the course of assessment. Using the factors which had been identified in our analysis we modelled an initial pilot study data collection mechanism based on Bayesian probability analyses.

The data collection mechanism evaluates the informativeness of question based on the change that each question causes on the probability of risk in the light of all previously collected data along side the probability of the the questions' answers. The model iteratively evaluates questions, select most informative ones and stop when no more question is proved to be able to reliably change the level of risk. Once the data collection is complete, the model predicts the risk level. We evaluated the classification accuracy of our model and we found a decent performance considering the constraints.

Major work is currently being undertaken to bring this model into full realisation. Therefore, a number of major future works for this research paper is identified. For example, we had found that the probability of the answers to a question has to be taken into account when deciding on the informativeness. However, further investigation needs to be made on how much emphasis should be given to the probability of the answers in comparison to the potential change they can cause on the risk. In other words, it is unclear whether clinicians prefer questions which can be very informative even though the informative answers are very unlikely to be given or rather they prefer questions which are less informative but the answers are more probable. Therefore, The model needs to include the probability of each answer but with a parameter indicating the correct level of emphasis that should be give to the probability. Similarly, our analysis had shown that the redundancy or the informativeness of questions are unfixed and hence a questions informativeness dynamically changes as more and more data are revealed by other questions. This means that a question can be little informative in itself but it puts on the agenda another question that is much more likely to have an answer that gives a high amount of information, when that answer would be very unlikely if you had not collected the little informative question. Therefore, the evaluation of a question should also take into account the likelihood of increasing the informativeness of future questions. Thus how much a question will increase the informativeness of future questions should be taken into account in the evaluation function.

Once the identified future works are completed then we will have a fully reliable system that can dynamically decide which questions are informative in a stage by stage manner, based on particular patient profile, and lead the assessors through the huge pool of data towards only information that help them improve their ability for evaluating the level of risk. The result of such dynamic system is the collection of a different set of data features for each patient to facilitate their specific evaluation of risk. Consequently, we will have an accurate triage of risk that is doable by non-experts and hence can be conducted on daily basis on a community level.

References

1. Federico Cismondi, André S Fialho, Susana M Vieira, Shane R Reti, João Sousa, and Stan N Finkelstein. Missing data in medical databases: Impute, delete or classify? *Artificial intelligence in medicine*, 58(1):63–72, 2013.
2. Marcus Richards and Rosemary Abbott. Childhood mental health and adult life chances in post-war britain: insights from three national birth cohort studies. 2009.
3. Swaran P Singh. Early intervention in psychosis. *The British Journal of Psychiatry*, 196(5):343–345, 2010.
4. Keith Hawton, Carolina Casañas i Comabella, Camilla Haw, and Kate Saunders. Risk factors for suicide in individuals with depression: a systematic review. *Journal of affective disorders*, 147(1):17–28, 2013.
5. Chris Poulin, Brian Shiner, Paul Thompson, Linas Vepstas, Yinong Young-Xu, Benjamin Goertzel, Bradley Watts, Laura Flashman, and Thomas McAllister. Predicting the risk of suicide by analyzing the text of clinical notes. *PloS one*, 9(1):e85733, 2014.
6. Megan Chesin and Barbara Stanley. Risk assessment and psychosocial interventions for suicidal patients. *Bipolar disorders*, 15(5):584–593, 2013.
7. Matthew K Nock, Guilherme Borges, Evelyn J Bromet, Christine B Cha, Ronald C Kessler, and Sing Lee. Suicide and suicidal behavior. *Epidemiologic reviews*, 30(1):133–154, 2008.
8. Wolfgang Aigner, Silvia Miksch, Wolfgang Muller, Heidrun Schumann, and Christian Tominski. Visual methods for analyzing time-oriented data. *Visualization and Computer Graphics, IEEE Transactions on*, 14(1):47–60, 2008.
9. Daniel F Heitjan. Annotation: what can be done about missing data? approaches to imputation. *American Journal of Public Health*, 87(4):548–550, 1997.
10. Robert I Simon. Imminent suicide: The illusion of short-term prediction. *Suicide and Life-Threatening Behavior*, 36(3):296–301, 2006.
11. Alex D Pokorny. Prediction of suicide in psychiatric patients: report of a prospective study. *Archives of general psychiatry*, 40(3):249–257, 1983.
12. Health Service Executive: Ireland. Risk management in mental health services.
13. A Rezaei-Yazdi and CD Buckingham. Understanding data collection behaviour of mental health practitioners. *Studies in health technology and informatics*, 207:193, 2014.
14. Kathy Lemon Osterling and Michael J Austin. The dissemination and utilization of research for promoting evidence-based practice. *Journal of evidence-based social work*, 5(1-2):295–319, 2008.
15. Jon Glasby, Martin C Calder, Kerry Baker, Hazel Kemshall, and Bernadette Wilkinson. *Good practice in assessing risk: Current knowledge, issues and approaches (introduction By Kemshall. H*. Jessica Kingsley Publishers, 2011.
16. Rick Goud, Mariette van Engen-Verheul, Nicolette F de Keizer, Roland Bal, Arie Hasman, Irene M Hellemans, and Niels Peek. The effect of computerized decision support on barriers to guideline implementation: a qualitative study in outpatient cardiac rehabilitation. *International Journal of Medical Informatics*, 79(6):430–437, 2010.
17. K Nirantharakumar, YF Chen, T Marshall, J Webber, and JJ Coleman. Clinical decision support systems in the care of inpatients with diabetes in non-critical care setting: systematic review. *Diabetic Medicine*, 29(6):698–708, 2012.
18. F Sonnenberg and C Hagerty. Computer-interpretable clinical practice guidelines. *Where are we and where are we going*, pages 145–158, 2006.
19. Sherif Hegazy and Chris D Buckingham. A method for automatically eliciting node weights in a hierarchical knowledge-based structure for reasoning with uncertainty. *International Journal On Advances in Software*, 2(1):76–83, 2009.
20. Janet A. Sniezek. Cognitive processes of decision making. *Salem Press Encyclopedia of Health*, 2014.
21. Steven N Byers and Charlotte A Roberts. Bayes' theorem in paleopathological diagnosis. *American journal of physical anthropology*, 121(1):1–9, 2003.
22. David J Spiegelhalter, Keith R Abrams, and Jonathan P Myles. *Bayesian approaches to clinical trials and health-care evaluation*, volume 13. John Wiley & Sons, 2004.