

Capturing Human Intelligence for Modelling Cognitive-based Clinical Decision Support Agents

Ali Rezaei-Yazdi*, Christopher Buckingham

Abstract

The success of intelligent agents in clinical care depends on the degree to which they represent and work with human decision makers. This is particularly important in the domain of clinical risk assessment where such agents either conduct the task of risk evaluation or support human clinicians with the task. This paper provides insights into how to understand and capture the cognitive processes used by clinicians when collecting the most important data about a person's risks. It attempts to create some theoretical foundations for developing clinically justifiable and reliable decision support systems for initial risk screening. The idea is to direct an assessor to the most informative next question depending on what has already been asked using a mixture of probabilities and heuristics. The method was tested on anonymous mental health data collected by the GRiST risk and safety tool (www.egrist.org).

Keywords: Intelligent agents, Clinical Intelligence, Clinical Decision Support Systems, Dynamic data collection, Risk assessment, Healthcare, eHealth

This work was part funded by the EIT Health GRaCE-AGE project.

1 Introduction

Successful intelligent agents are those which can learn and utilise human expertise. This is of particular importance in the health-care sector where computerised systems aim to facilitate important clinical tasks including risk assessment [1, 2]. For mental health, evaluating risks such as suicide and self-harm is particularly challenging, because the symptoms, motives and deterrents are often abstract and difficult to measure. Here, intelligent agents such as certain types of Clinical Decision Support Systems (CDSSs) can play an important role by guiding users towards the most informative data features and advising on the risks associated with the collected data. If they are to be used in real-world settings, it is crucial that these computer agents are completely reliable.

Although recent reviews on the impact of CDSSs on the quality of healthcare have suggested an overall improvement [3, 4], difficulties remain with modelling and prediction of risk [5, 6]. The underlying problem is that many CDSSs cannot

*Corresponding Author: Ali Rezaei-Yazdi, Aston University, Aston Triangle, Birmingham, B4 7ET; E-mail: rezaeiya@aston.ac.uk.

reliably represent the clinicians' mental model of how the system should work [7, 8] i.e. accommodating clinicians' cognitive work flow [9]. Many intelligent agents lack a functional model allowing meaningful interpretation of results in terms of the features used for their risk classifications [10]. All these factors result in a reluctance to use such systems for major decision making tasks such as risk judgements because clinicians trust their own expertise and experience more than the complex models that underpin the intelligent systems [6, 11].

To gain the trust of clinicians, CDSSs must support their cognitive work flow and mental models of decision-making [8, 12]. This requires knowledge of how clinicians think and collect data, and how the data they collect work together to determine judgements of patients' health.

The goal of this research is to understand how clinicians are alerted to the most pertinent information for evaluating risks. Their data-collection processes can then be simulated by intelligent agents so that they point assessors to the right data at the right time for making quick and effective clinical decisions.

The research analyses a database of patient profiles associated with clinical risk judgements in order to understand how the relationships within the data influence those judgements. It does not aim to introduce a new model or algorithm, but rather focuses on the theoretical investigations necessary for building a clinically justifiable model to be used for clinical intelligent agents. The paper will provide the study context and then explain the methodology in detail. The results will be described and discussed followed by conclusions and future work.

2 Context

The research was conducted on data collected by the Galatean Risk & Safety Tool, GRiST,[13] which is a clinical decision support system that helps assessors evaluate risks of suicide, self-harm, harm to others, self-neglect, and vulnerability. This study focused on suicide and used 30,000 cases of completed risk assessments provided by mental-health practitioners through their normal practice. The data are automatically stored without personal identification information before any researcher has access to them and so are completely anonymous from the start.¹

GRiST is based on a psychological model of classification for representing clinical expertise [14]. The 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 leaves of the tree are reached, such as anger and anxiety. Figure 1 shows a simplified part of the suicide risk tree.

Although GRiST has several hundred leaf nodes, and therefore a very large potential data set for each patient, the actual number of questions is much lower because the tree provides a top-down gateway to the relevant concepts.

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

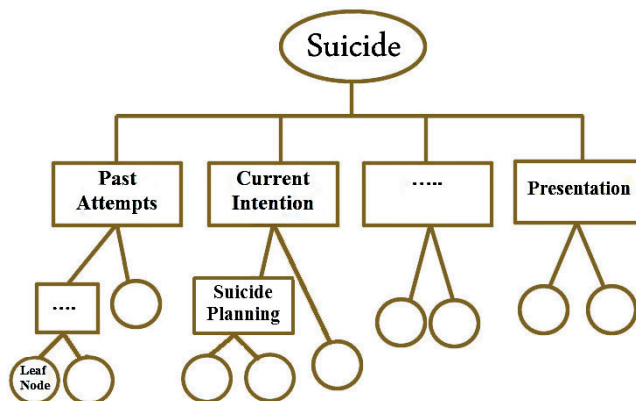


Figure 1: Part of the suicide risk knowledge tree.

For example, a top-level question is whether the patient has any history of suicide attempts and if the answer is ‘NO’ then all the history questions remain hidden and unasked. 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).

The idea is that not all the branches of the tree or all the questions of a branch are necessarily relevant for the patient under assessment. The task is to understand which branch variables (i.e. answers to branch questions) influence risk judgement, to what degree, and how these influences impact on the process of data collection. Data collection can then be guided only to those parts of the knowledge tree that are required for the particular risk assessment and particular patient under assessment, thereby minimising the number of questions needing to be asked. The specific objectives are to understand what drives the data collection behaviour of clinicians and use this to find a minimum set of variables required to provide accurate risk judgements for any particular patient.

3 Methodology

Bayes Theorem describes how the degree of belief in an event should change to account for evidence and is widely utilised for analysing clinical decision making [16, 17]. In this study, it was applied to understand how clinical judgements are affected by the sequence of data collection. The aim was to determine which questions would *change* the judgement the most, in the light of data already collected. Guiding assessors to these questions would provide the least number of questions required to provide the most informative risk evaluation.

The method is not as simple as described so far because each branch can have two answers, YES or NO, and they 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. Hence a question may only be useful for

one answer but not the other and the likelihood of each answer affects the overall informativeness of the question. Furthermore, the informativeness of an answer changes depending on what has already been asked and the conditional probabilities have to be updated for each new answer.

The conditional probability of an event ‘A’ given another event ‘B’ is denoted in this paper using the following format:

$P(A|B)$ reads as probability of A given B.

The condition (in the above example, B) is taken as the **I**ndependent **V**ariable, the *IV*, and the event whose probability we are trying to calculate (in the above example, A) is called the **D**ependent **V**ariable, the *DV*. Using this approach, the investigation was carried out in 2 main stages.

1. **Stage 1:** To investigate how answers to the questions of the knowledge tree influence clinicians’ final risk judgement. First, we looked at how each variable changed the suicide risk judgement on its own, to give a measure of its predictive power. Second, we explored how variables change the probability of risk in combination with other variables, to find combinations of variables that are particularly predictive compared to each one separately.
2. **Stage 2:** To investigate whether clinicians’ decide that it is not worth pursuing a line of reasoning. If clinicians stop asking questions in one part of the tree because of the answers they have obtained elsewhere, it should be possible to detect redundant parts of the tree. Assessors can then be guided away from these and towards more useful lines of questioning with respect to risk evaluation.

4 Findings

The risk judgements were categorised into three groups: High Risk (risk scores of 7 and above), Medium Risk (risk scores 4,5,6) and Low Risk (risk scores of 3 and below). These groupings made it easier to measure the degree of dependency between variables and the risk by treating each risk group as a separate variable. Since determining the high risk patients is most important, the investigation concentrated on measuring the variables influence on High Risk. What follows is the result of analyses and the semantic conclusions drawn from them.

4.1 Stage 1: Influence of variables and their combinations on risk judgements)

With a total of 28 branch and sub-branch variables, 28 pairs of conditional probabilities of the risk given a question’s answers were produced because only the “yes” answers were analysed (the “no” answers are, by definition, not indicative of high risk). Then the answers for each question were paired with answers for the other variables in combinations of pairs and triples. Any more than three

variables reduce the sample size significantly and also became unmanageable in scale: 28 variables combined in groups of two produces 378 combinations and in groups of three produces 3,276 combinations based on combination factorial calculations. Each combination was then compared against the high risk variable to give a total number of 3,654 conditional probability calculations.

The combinations producing the most influence on high risk were selected, with a sample shown in Table 1. The table shows variables' influence on the probability of High Risk on an individual basis as well as jointly. The CP column gives the conditional probability of the DV given the IV and the change in the probability of DV caused by the IV is the fourth ΔP column. The change in probability, ΔP , shows how influential the variable is on the clinicians' risk judgements.

Independent Variable	Dependant Variable	CP	ΔP
Current Intention	High Risk	0.37	0.32
Presentation	High Risk	0.22	0.17
Suicide Triggers	High Risk	0.14	0.09
Ideation	High Risk	0.13	0.08
Self-Worth	High Risk	0.12	0.07
Motive&Engagement	High Risk	0.11	0.06
Feeling&Emotion	High Risk	0.09	0.04
Personality	High Risk	0.09	0.04
Presentation & Current Intention	High Risk	0.51	0.48
Current Intention & Suicide Planning	High Risk	0.46	0.41
Presentation & Self-Worth	High Risk	0.32	0.28
Presentation & Verbal Indicators	High Risk	0.30	0.25
Presentation & Personality	High Risk	0.22	0.17
Suicide ideation & Personality	High Risk	0.20	0.14

Table 1: A selection of most influential variables which individually or in combination influence clinicians' judgement of risk.

Figure 2 and 3 visualise the result of Table 1. Figure 2 shows the variables with strongest predictive power (influence) on the probability of High Risk on individual basis. Figure 3, on the other hand, shows the variable combinations with strongest predictive power on High Risk.

Table 1 and the associated figures reveal important insights into the clinicians' pattern of response. Some variables often appear in combinations with high predictive power as well as being strong independent predictors. It would make sense if these were primary drivers of clinical decisions and they include Current Intention to commit suicide (CI), Suicide Triggers (ST), Suicide Ideation (SI), Presentation of patients during assessment (PR), Self-Worth (SW) and Motivation & Engagement (M&E).

Furthermore, some variables' show increased predictive power in combination

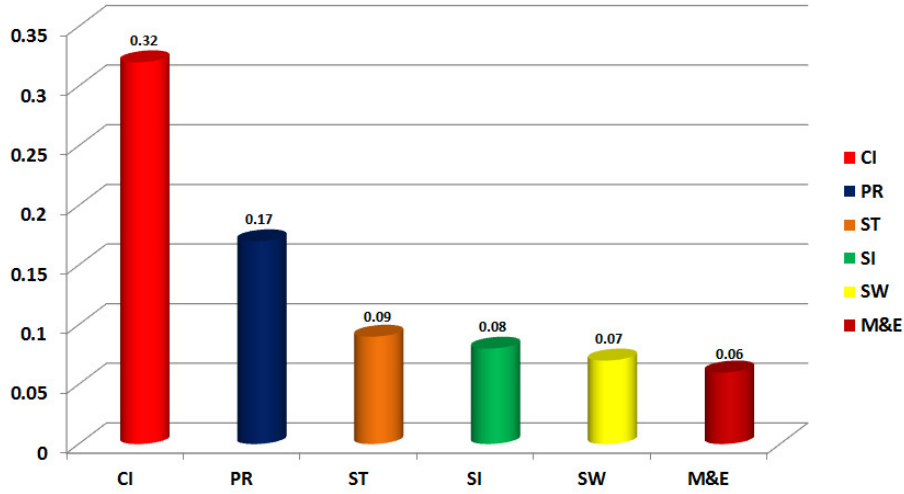


Figure 2: Variables with strongest probability impact on High Risk on individual basis.

with other variables. For example, Self-Worth (SW), on its own, changes the probability of High Risk (HR) by 0.07 but when combined with the variable Presentation (PR) its influence on HR increases to 0.10:

$$\Delta P(HR|SW) = 0.07 \neq \Delta P((HR|PR)|SW) = 0.10$$

\implies SW demonstrates interactive behaviour with PR.

This means that when clinicians are concerned with the ‘Presentation’ of patients, asking about Self-Worth would increase clinicians’ chance of judging the patients as high risk, if self worth was, indeed, an issue. To guide assessors to the most important data, the system needs to know both the independent influence of a variable and also how much it is likely to change the impact of risk in conjunction with others already asked.. Interaction of variable influences is the key because the addition of a second or third IV sometimes influences the probability of high risk sometimes does not. For example, according to Table 1, Presentation, on its own changes the probability of High Risk by 0.17 but if the variable Personality (PER) is added, the risk is hardly affected (this is demonstrated in Figure 3 by having the column representing Presentation and Personality as completely blue). However, when Personality is added to Suicide Ideation, the probability of High Risk increases by 0.07. It shows how the strength of interactions dynamically changes as new answers are added.

The results so far suggest that there are a limited number of features which strongly influence the clinical risk judgement. Furthermore, although the influences on risk change as new answers are added to previous ones, if the previous answers are already strong influencers, new answers make little difference: the strong influencers are making others redundant, which is an important factor in determining the order of data collection if minimising the questions required is important (which is the case for screening tools). This is investigated in more

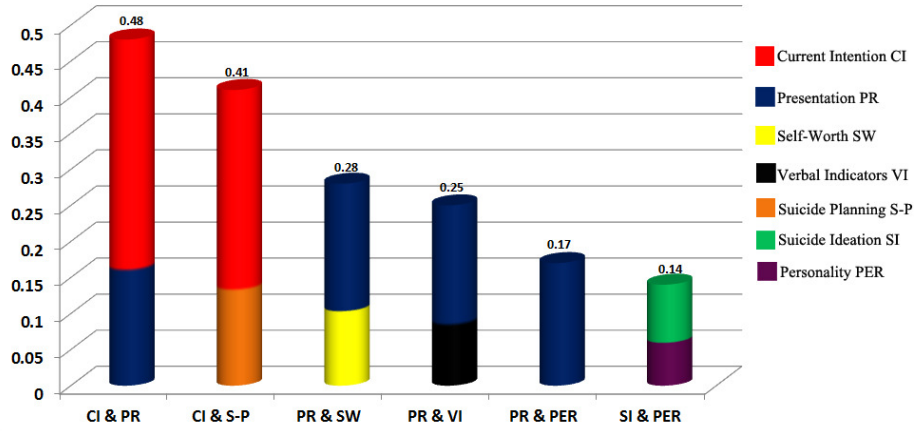


Figure 3: Combination of variables with strongest probability impact on High Risk. Different proportion of colouring demonstrates the power of each variable in the combinations. For example the first column shows the combination of Current Intention and Presentation with Current Intention having a more impact on risk, hence the column is dominated in red.

detail in Stage 2.

4.2 Stage 2: identifying redundant data

Having identified the most influential variables, it is important to clarify whether they mask all the other variables or whether there are certain combinations that will still be informative. It may also be the case that a particular variable is redundant for a certain sequence of answers but later becomes influential if a new one is added.

Redundancy was investigated using *Conditional Independence* between IVs and the DV: is the DV (high risk) independent of an IV when that IV occurs in conjunction with another IV? If so, then only the second IV is required. This can be written in terms of conditional probabilities as:

$$P(DV|IV2 \cap IV1) = P(DV|IV2)$$

where IV2 is the independent variable that is fully incorporating the influence of IV1. A variable can only be considered redundant if none of its answers (YES or NO) can change the probability of the risk, given any answer (YES or NO) to the other variables. Therefore, the first redundancy test checks to see what effects the different answers for an IV would have on the redundancy of the other, test IV (T-IV).

The second redundancy test examines the effects of different answers to the T-IV on redundancy of itself in relation to the risk. In other words, it checks to see how different answers to T-IV change its level of influence on the risk given that the first IV is present.

Redundancy Test 1:

$$\frac{\text{IF } P(\text{HR} \mid \text{IV}=\text{YES} \ \& \ \text{T-IV}) \approx P(\text{HR} \mid \text{IV}=\text{YES})}{\text{AND}} \left. \vphantom{\frac{\text{IF } P(\text{HR} \mid \text{IV}=\text{YES} \ \& \ \text{T-IV}) \approx P(\text{HR} \mid \text{IV}=\text{YES})}{\text{AND}}} \right\} \Rightarrow \text{T-IV REDUNDANT}$$

$$\text{IF } P(\text{HR} \mid \text{IV}=\text{NO} \ \& \ \text{T-IV}) \approx P(\text{HR} \mid \text{IV}=\text{NO})$$

Redundancy Test 2:

$$\frac{\text{IF } P(\text{HR} \mid \text{IV} \ \& \ \text{T-IV} = \text{YES}) \approx P(\text{HR} \mid \text{IV})}{\text{AND}} \left. \vphantom{\frac{\text{IF } P(\text{HR} \mid \text{IV} \ \& \ \text{T-IV} = \text{YES}) \approx P(\text{HR} \mid \text{IV})}{\text{AND}}} \right\} \Rightarrow \text{T-IV REDUNDANT}$$

$$\text{IF } P(\text{HR} \mid \text{IV} \ \& \ \text{T-IV}=\text{NO}) \approx P(\text{HR} \mid \text{IV})$$

The third test complements the previous two by checking to see if a variable which has been made (by the first two tests) redundant remains redundant permanently or whether it can become informative at a later point during the assessment when further variables have been collected.

Redundancy Test 3:

$$\frac{\text{Given T-IV is redundant against IV}}{\text{IF } P(\text{HR} \mid \text{IV} \ \& \ \text{additional IVs} \ \& \ \text{T-IV})} \left. \vphantom{\frac{\text{Given T-IV is redundant against IV}}{\text{IF } P(\text{HR} \mid \text{IV} \ \& \ \text{additional IVs} \ \& \ \text{T-IV})}} \right\} \Rightarrow \text{T-IV REDUNDANT}$$

$$\approx$$

$$P(\text{HR} \mid \text{IV1} \ \& \ \text{additional IVs})$$

The idea is that if the redundant features stay redundant throughout the assessment, then they can be removed from the sequence of questions. The method was tested using a selection of the strongest variables as found in Table 1 along with some of the weaker variables, with a variable deemed redundant if and only if it passes all the three redundancy tests.

The results were interesting: Some variables pass none of the tests, some pass the first test but not the second and some pass the first and second but not the third. For example, Tables 2, 3 and 4 show the result of the 3 redundancy tests on Current Intention, Presentation, Feeling&Emotion and Mental Faculty. In these tests, Current Intention is taken as the IV and the other three are taken as the T-IVs. The tables show that Presentation is redundant if and only if Current Intention is given a NO value. Otherwise Presentation can independently change the level of risk. Presentation thus fails the first test of redundancy as demonstrated in Table 2 and is not investigated further. On the other hand, Feeling&Emotion passes the first test of redundancy, which gives the impression that as long as Current Intention is collected (whether its value is YES or NO) Feeling&Emotion is redundant. However it fails the second test (as demonstrated in Table 3) meaning Feeling&Emotion does not necessarily stay redundant if it has a NO value. In other words, while a YES value to Feeling&Emotion makes it redundant in the presence of either Current Intention answer, a NO value will actually make Feeling&Emotion informative (i.e. independently causes a substantial change on the probability of risk).

In contrast to Feeling&Emotion, Mental Faculty passes the first two tests. This means if Current Intention is collected (regardless of the value it takes), then Mental Faculty will always be redundant regardless of the value it is given (demonstrated by Tables 2 and 3). This might suggest that Mental Faculty is completely redundant but surprisingly Mental Faculty does not pass the third

Redundancy Test 1				ΔP	Threshold
IV = Current Intention CI , T-IV = Presentation PR					>0.05
P(High Risk CI = Yes & PR = YES) = 0.51	\neq	P(High Risk CI = Yes) = 0.37		0.14	Over
P(High Risk CI = NO & PR = YES) = 0.04	\approx	P(High Risk CI = NO) = 0.01		0.03	Within
IV = Current Intention CI , T-IV = Feeling&Emotion FE				ΔP	Threshold
IV = Current Intention CI , T-IV = Feeling&Emotion FE					>0.05
P(High Risk CI = Yes & FE = YES) = 0.39	\approx	P(High Risk CI = Yes) = 0.37		0.02	Within
P(High Risk CI = NO & FE = YES) = 0.02	\approx	P(High Risk CI = NO) = 0.01		0.01	Within
IV = Current Intention CI , T-IV = Mental Faculty MF				ΔP	Threshold
IV = Current Intention CI , T-IV = Mental Faculty MF					>0.05
P(High Risk CI = Yes & MF = YES) = 0.40	\approx	P(High Risk CI = Yes) = 0.37		0.03	Within
P(High Risk CI = NO & MF = YES) = 0.01	\approx	P(High Risk CI = NO) = 0.01		0.0	Within

Table 2: First redundancy test applied to Presentation, Feeling&Emotion and Mental Faculty against Current Intention. Feeling&Emotion and Mental Faculty pass the test. Presentation fails the test as the change it causes on High Risk is **over** the threshold

Redundancy Test 2				ΔP	Threshold
IV = Current Intention CI , T-IV = Feeling&Emotion FE					>0.05
P(High Risk CI = Yes & FE = Yes) = 0.39	\approx	P(High Risk CI = Yes) = 0.37		0.02	Within
P(High Risk CI = Yes & FE = No) = 0.26	\neq	P(High Risk CI = Yes) = 0.37		0.11	Over
IV = Current Intention CI , T-IV = Mental Faculty MF				ΔP	Threshold
IV = Current Intention CI , T-IV = Mental Faculty MF					>0.05
P(High Risk CI = Yes & MF = Yes) = 0.40	\approx	P(High Risk CI = Yes) = 0.37		0.03	Within
P(High Risk CI = Yes & MF = No) = 0.37	=	P(High Risk CI = Yes) = 0.37		0.0	Within

Table 3: Second redundancy test applied to Feeling&Emotion and Mental Faculty against Current Intention. Mental Faculty passes the test. Feeling&Emotion fails as the change it cases on High Risk is **over** the threshold when it is given a ‘NO’ answer

Redundancy Test 3				ΔP	Threshold
IV = Current Intention CI , T-IV = Mental Faculty MF Additional Variable = Feeling&Emotion FE					>0.05
P(High Risk CI = Yes & MF = Yes) = 0.40	\approx	P(High Risk CI = Yes) = 0.37		0.03	Within
P(High Risk CI = Yes & FE= Yes & MF = Yes) = 0.46	\neq	P(High Risk CI = Yes & FE= Yes) = 0.39		0.07	Over

Table 4: Third redundancy test applied to Mental Faculty against Current Intention. Mental Faculty fails the test as it regains its informativeness after being made redundant once.

test. Table 4 shows that Mental Faculty which was previously made redundant with Current Intention, becomes informative ($\Delta P > \text{Threshold}$) when a second IV (i.e. Feeling&Emotion) is added to Current Intention. In other words, Mental Faculty comes back to be informative when Feeling&Emotion has been given a YES value. In clinical terms this means that when clinicians collect Current Intention and are also concerned with Feeling&Emotion, then the questions on Mental Faculty would influence their judgement of risk but not if they are not concerned with Feeling&Emotion.

These findings point out an important concept: the pattern of variable redundancy is not fixed but evolves throughout the assessment. In other words, a variables redundancy (and informativeness) is a dynamic phenomenon and changes as the value to other variables become available.

4.3 Clinical conclusions from the result

This investigation provides preliminary suggestions about how clinical assessments might be optimised:

1. There are a group of features that are always influence the clinical judgements and these are absolutely required in all contexts.
2. Depending on the context, which is dictated by the data collected during the course of the assessment, less influential features can actually be informative and hence required for evaluating the level of risk.
3. Clinicians do not evaluate the predictiveness of a variable in isolation from other variables. Rather, they base their evaluation on the combination of answers already obtained. Hence a new answer that is informative on its own can be completely redundant in the presence of other data because clinicians feel that the new information adds little or nothing to their existing understanding of the level of risk.
4. The informativeness of data fluctuates throughout an assessment. Not only does a variable's predictive power depend on the preceding variables but it can also change depending on the future data too. This dynamic life line is illustrated by Figure 4.



Figure 4: Data informativeness life line

5. The dynamic nature of data redundancy means the optimal order of asking questions is also dynamic: it has to be decided in real time.

5 Conclusion & Future work

In order for clinical decision support systems CDSSs to be effective, they have to be intelligent agents that simulate the clinical judgements of a human clinician. To achieve this, such intelligent agents have to be able to replicate clinicians' expertise and way of reasoning. This research has explored this aim by analysing a database of mental health risk assessments. The research aimed to understand the factors and circumstances influencing humans (i.e. clinicians) decisions on the data they need to collect. Specifically, it tried to find out whether there is an optimal order of answers that produce the most accurate risk judgements with the least number of questions. This was pursued by analysing the probabilistic relationships between variables and clinical risk judgements at multiple levels.

The findings were used to extract key lessons about the way clinicians' risk judgements respond to the sequence of data collected. The most important is that the informativeness of variables have different patterns during the assessment. Some become redundant in the light of previous answers and will never then become informative; others are always informative; and some may be redundant with certain sequences of answers but then become informative in the light of a new answer, which means that their initial redundancy has not consigned them to the dustbin as far as asking the question. There is no fixed order of questions that will be the optimal one for all patients.

The challenge is to use this extracted clinical intelligence to produce an intelligent decision support agent that is able to communicate with the human risk assessors, take input from them and then find and collect the best set of data features for each different patient and context. The model will then use the collected data and the induced knowledge to provide intelligent advice for the human assessors. Exactly how to do this is currently under investigation. One of the most significant outcomes will be a system that can provide screening questions and accompanying risk advice for people without a mental-health background, based on the inherent expertise of the clinical judgements residing in the GRiST database. It will be particularly useful for primary care and emergency services who have to make fast and reliable mental-health triage decisions for people in the community.

References

- [1] T Allan Pryor. Development of decision support systems. *International journal of clinical monitoring and computing*, 7(3):137–146, 1990.
- [2] Harry Campbell, Rona Hotchkiss, Nicola Bradshaw, and Mary Porteous. Integrated care pathways. *BMJ: British Medical Journal*, 316(7125):133, 1998.
- [3] John Fox, Vivek Patkar, and Richard Thomson. Decision support for health care: the proforma evidence base. *Journal of Innovation in Health Informatics*, 14(1):49–54, 2006.

- [4] Kensaku Kawamoto, Caitlin A Houlihan, E Andrew Balas, and David F Lobach. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj*, 330(7494):765, 2005.
- [5] Federico Sanchez. *Suicide Explained: A Neuropsychological Approach*. Xlibris Corporation, 2007.
- [6] K Brunton. The evidence on how nurses approach risk assessment. *Nursing times*, 101(28):38, 2005.
- [7] Ellen Kilsdonk, Linda W Peute, Rinke J Riezebos, Leontien C Kremer, and Monique WM Jaspers. From an expert-driven paper guideline to a user-centred decision support system: A usability comparison study. *Artificial Intelligence in Medicine*, 59(1):5–13, 2013.
- [8] Linda WP Peute and Monique WM Jaspers. The significance of a usability evaluation of an emerging laboratory order entry system. *International Journal of Medical Informatics*, 76(2):157–168, 2007.
- [9] Dean F Sittig, Adam Wright, Jerome A Osheroff, Blackford Middleton, Jonathan M Teich, Joan S Ash, Emily Campbell, and David W Bates. Grand challenges in clinical decision support. *Journal of Biomedical Informatics*, 41(2):387–392, 2008.
- [10] Dayle L Sampson, Tony J Parker, Zee Upton, and Cameron P Hurst. A comparison of methods for classifying clinical samples based on proteomics data: a case study for statistical and machine learning approaches. *PloS one*, 6(9):e24973, 2011.
- [11] David JA Dozois. Psychological treatments: Putting evidence into practice and practice into evidence. *Canadian Psychology/Psychologie canadienne*, 54(1):1, 2013.
- [12] A Rezaei-Yazdi and CD Buckingham. Understanding data collection behaviour of mental health practitioners. *Studies in health technology and informatics*, 207:193, 2014.
- [13] GRiST. Galatean risk and safety tool. www.egrist.org, 2014. accessed May 15th.
- [14] Christopher Buckingham. Psychological cue use and implications for a clinical decision support system. *Informatics for Health and Social Care*, 27(4):237–251, 2002.
- [15] Janet A. Snizek. Cognitive processes of decision making. *Salem Press Encyclopedia of Health*, 2014.
- [16] Steven N Byers and Charlotte A Roberts. Bayes’ theorem in paleopathological diagnosis. *American journal of physical anthropology*, 121(1):1–9, 2003.
- [17] 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.