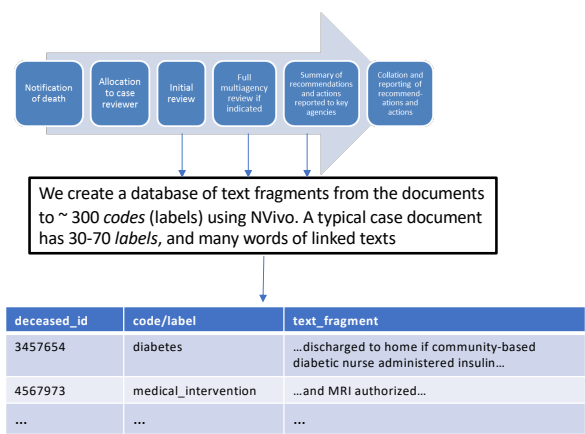




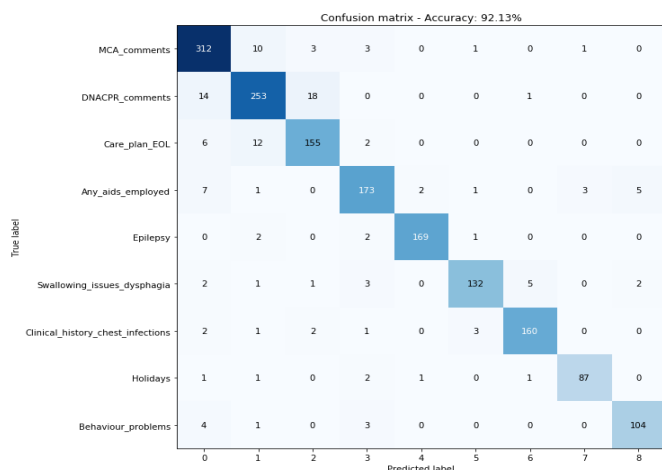
A unique labelled database of health-care and social-care texts for machine learning



Manual coding of labels

To aid analysis of these reviews, we have a team of *coders*, who label fragments of the review texts to about 300 codes (labels) using NVivo. This has generated a unique database of labelled texts in the health and social care domain,

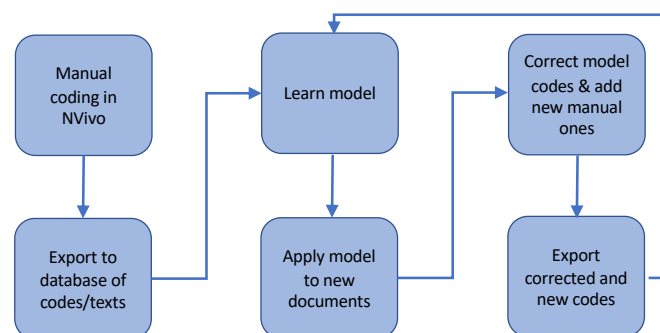
This *coding* is an arduous and expensive task. So we are looking to support it through semi-automated classification of review sentences/paragraphs to our 300 codes using machine learning.



Introduction

People with a learning disability die some 20 years earlier than the general population. The *LeDeR* programme studies the factors that may contribute to such early deaths and can thus suggest policy changes to reduce them.

LeDeR collects notification of deaths and then oversees the writing of case reviews by health professionals. We currently have ~7000 notified cases, of which ~3000 have completed written reviews.



Machine Learning

Initial experiments are promising. Using a support vector classifier on a small sample of important codes (labels) we reach an accuracy of over 90% (see heatmap on left), although this may decline as we add more challenging codes. The table below shows results on unseen texts. The system does well on most texts, failing only on the last example, where the word *enjoyed* has led to a false class label of *Holidays*.

But as human coders will check any suggested labels, an accuracy of even 70-80% will still improve throughout and consistency of the final labelled database. We will use a semi-automated approach, in which after the predicted labels are checked by human coders the new data are used to refine the model (see flowchart above).

| New (previously unseen) text | Predicted class | Probability |
|---|------------------------------------|-------------|
| She required a shower chair to be able to have a safe daily shower | <i>Any_aids_employed</i> | 0.97 |
| She was on a textured diet and fortified drinks | <i>Swallowing_issues_dysphagia</i> | 0.96 |
| Her mobility had decreased and she needed to use a wheeled walker frame in the home, however she forgot | <i>Any_aids_employed</i> | 0.99 |
| She enjoyed living in her Somerset bungalow | <i>Holidays</i> | 0.90 |