Nurses and midwives play important roles in the healthcare system as they provide highly skilled and often complex care in both hospitals and communities. To protect and prioritize the safety of the public from harmful practices, most countries have specific health professional regulators to set rules, monitor and shape the practice of nurses and midwives. When concerns over a nurse or midwife’s practice are raised, a formal complaint can be submitted to the regulator, and investigations will be performed to decide further actions. Processing complaints is highly time-consuming and costly hence, the need for effective tools to support investigations is crucial. In this paper, we present a decision support system to improve the efficiency of complaints investigation for nursing and midwifery regulators, by employing state-of-the-art machine learning and natural language processing (NLP) techniques with a human-in-the-loop.

Background and Objectives

The most essential functionality regulatory agencies need is to be able to predict the Risk Level of the case, as it allows them to prioritise the high-risk cases and better manage the workload. To make this prediction, we formulated the problem as a binary classification task and developed an ensemble model. Our system also provided some additional information to further support the decision-making process of the regulator and help them interpret the prediction results:

- **Confidence scores** to assess the certainty of each prediction
- **LIME** to provide explanations for each prediction
- **Reference to similar past cases**
- **Natural Language Inference (NLI) models to detect non-compliance**

System Design

The user interface of our system is presented in Figures 2 and 3. The interface includes the following components:

- **Risk level prediction (ensemble)**
- **Confidence prediction**
- **Linking to regulations**
- **Linking to similar cases**

Evaluation

Results

We presented the first system to support complaints investigation for nursing and midwifery regulators. The system exploits state-of-the-art text classification, summarisation, semantic similarity measurement and NLI techniques, and provides different types of information to assist the regulators, including risk level assessment (with highlighted words as explanations), similar past cases, and non-compliance to regulations. Also, gender de-biasing operations are performed to reduce systemic gender biases. Feedback received from domain experts confirmed the system’s usefulness and potential. We hope this work will inspire more AI/NLP-based decision support systems across different jurisdictions, and encourage further collaborations between NLP researchers and regulatory bodies.

Fig. 2: A screenshot of the result page for a fictitious complaint. The page consists of (1) the complaint text (2) the predicted risk level, probability, and confidence (3) word importance scores provided as the explanation by LIME (4) similar past cases (5) non-compliance to regulations (6) the final decision to be given by a case manager.

Fig. 3: Performance (mean ± standard deviation) of the risk classifiers, averaged over 10 random splits.

Fig. 4: Performance of different gender-bias methods. O″ to G″ in the bottom column stand for original and gender-debiased embeddings, respectively.

Reference to similar past cases

• The identified relevant regulations, participants provided moderate ratings at 2.8 for each of them. However, lower ratings (1.8) were given on the similar cases found by the system.

Our system is primarily implemented with Flask 1.0.2, Bootstrap 4.1.3 and Charts.js 2.5.4. We invited five regulatory staff from NMC to use and evaluate our system. All participants found the usability and responsiveness of the system highly satisfactory, with average scores at 4.4 and 4.2 respectively. With respect to the quality of the risk predictions, explanations (i.e., the highlighted words), and the identified relevant regulations, participants provided moderate ratings at 2.8 for each of them. However, lower ratings (1.8) were given on the similar cases found by the system.