Horizon 2020
European Union Funding
for Research & Innovation

Data Augmentation for Reliability & Fairness in Counselling Quality Classification

Vivek Kumar, Simone Balloccu, Wu Zixiu, Ehud Reiter,
Rim Helaoui, Diego Reforgiato Recupero & Daniele Riboni

Philips Research, University of Cagliari & Aberdeen

Motivation

- Scarcity of data in the public domain.
- Unbalancedness in data leading to unreliable classification.
- fficacy of Data Augmentation Techniques applied to complex domains, for instance, in mental health.
- Fairness and Bias assessment of classification models for real-world application of clinical NLP

Our Contribution

- We inspect the effects of data augmentation on classical machine (CML) and deep learning (DL) approaches for counselling quality classification.
- Our work is the first step towards increasing reliability and reducing the bias of classification models, as well as dealing with data scarcity and imbalance in mental health.
- We conduct the bias and fairness analysis considering the therapy topic as the sensitive variable.
- Finally, we implement a fairness-aware augmentation technique, showing how topic-wise bias can be mitigated by augmenting the target label with respect to the sensitive variable

Data Description

Our Experiments Use Three Datasets

Anno-MI:	It contains 110 high-quality and 23 low-quality Motivational Interviewing (MI) conversational dialogues from a total of 44 topics.
Anno-AugMI:	It is created in a topic-agnostic fashion, with the goal of obtaining a roughly balanced amount of HQ-MI and LQ-MI utterances across the entire dataset after applying augmentation pipeline **Anno-MI**.
Anno-FairMI:	To assess classification fairness, this dataset is created considering therapy topic (MI-topic) as sensitive variable consisting of all the therapist utterances from **Anno-MI**. Anno-FairMI aims to contain same amount of HQ-MI and LQ-MI utterances for each MI-topic

Distribution of Datasets

Dataset	Total utterances (no.)	High quality(%)	Low quality(%)
Anno-Mi	2601	91%	9%
Anno-AugMI	5302	45%	55%
Anno-FairMI	9154	50%	50%

Statistics of Datasets

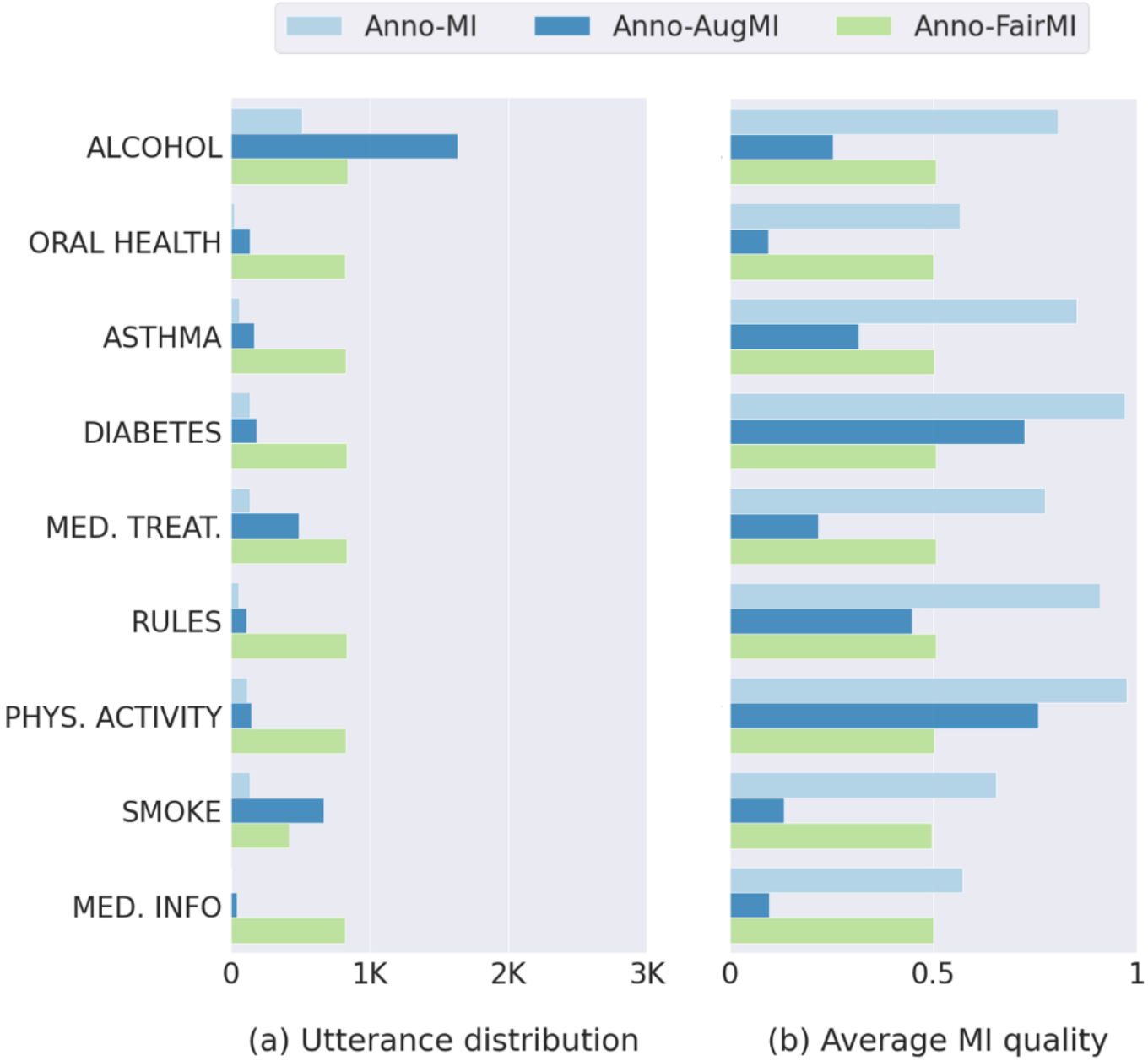


Figure : Sensitive variable statistics for each dataset. We show topic-wise

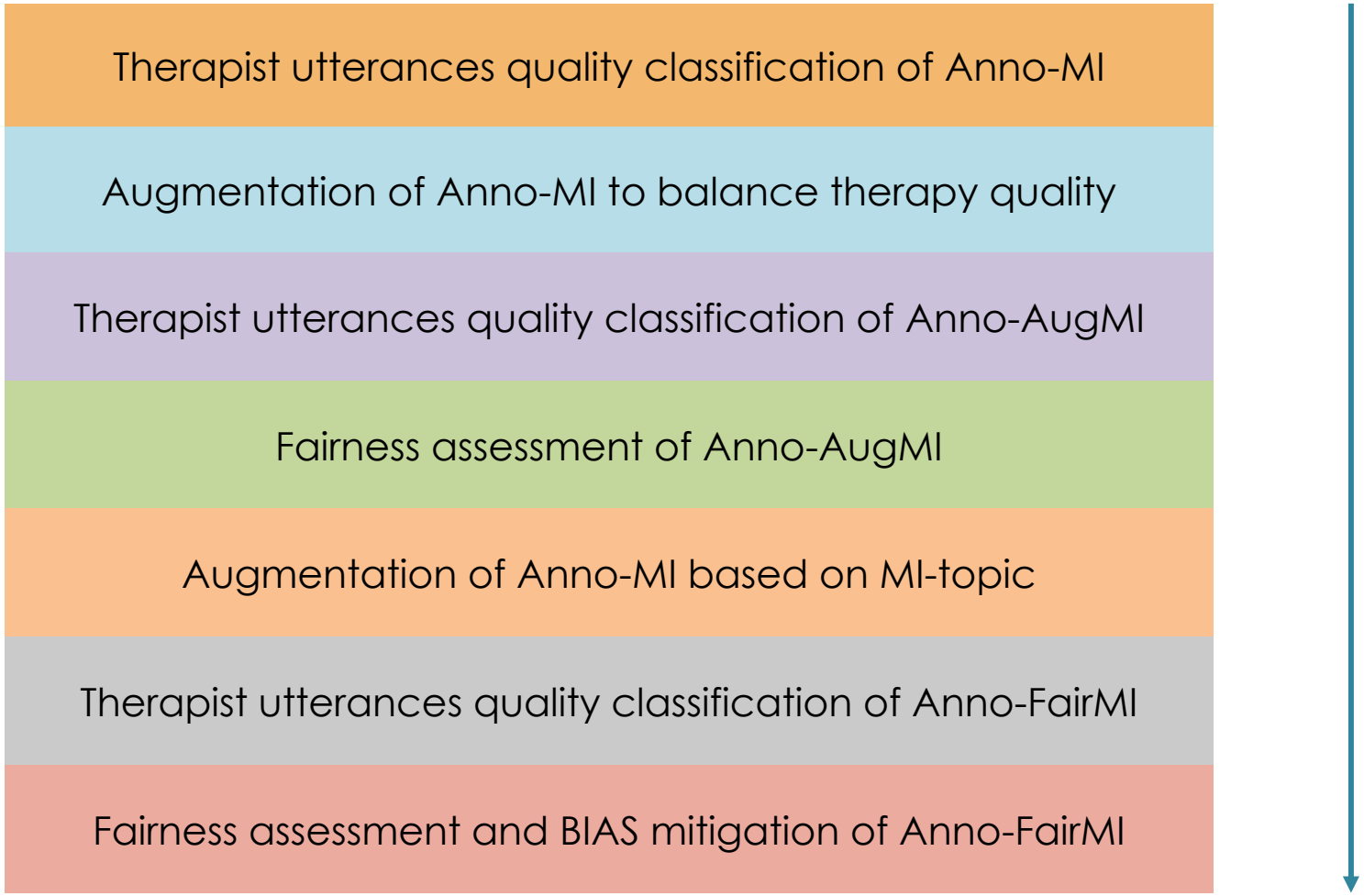
(a) Utterances distribution

&

(b) Average therapy quality.

Experimental Setup

We design a series of experiments, where each experiment's input is based on the output of the preceding ones. The experimental setup is as follows:

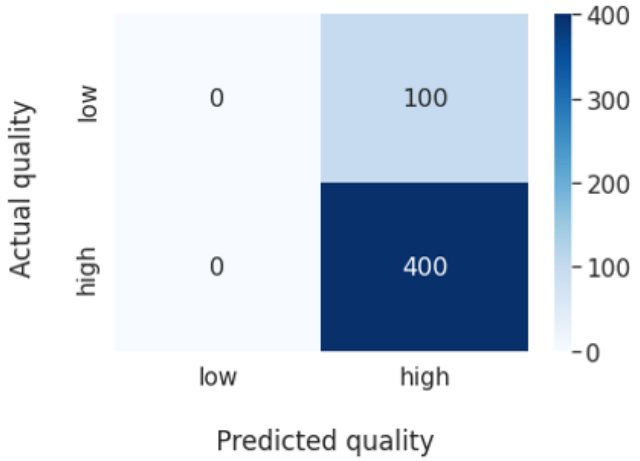


RESULTS

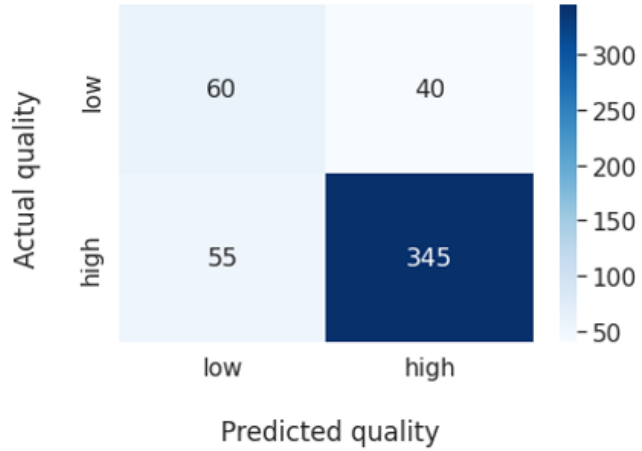
Result: Performance of CML and DL approaches with Anno-MI, Anno-AugMI, Anno-FairMI. For each dataset we report Balanced Accuracy and F1 score calculated with regards to MI quality.

	SVM		Random Forest		Bi-LSTM(DNN)	
Dataset	**Bal. Acc.**	**F-1 Score**	**Bal. Acc.**	**F-1 Score**	**Bal. Acc.**	**F-1 Score**
Anno-Mi	50.00	44.44	50.75	46.34	50.00	44.44
Anno-AugMI	48.87	38.12	50.37	45.78	**73.12**	**71.85**
Anno-FairMI	53.87	48.15	51.00	50.99	64.13	59.50

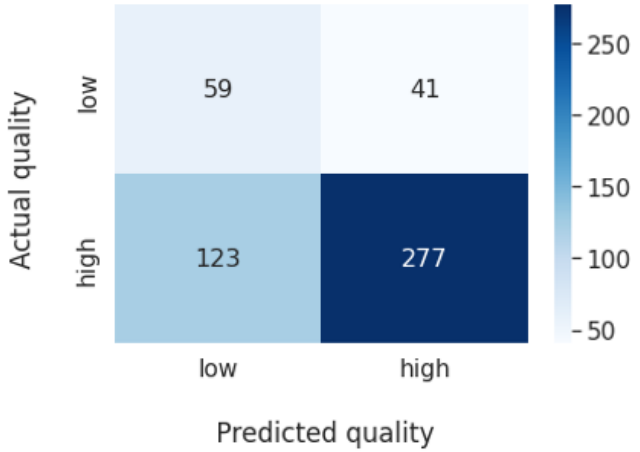
RESULTS



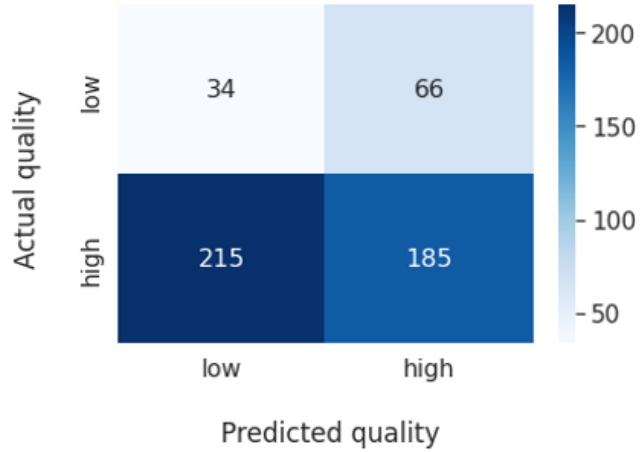
(a) Anno-MI



(b) Anno-AugMI



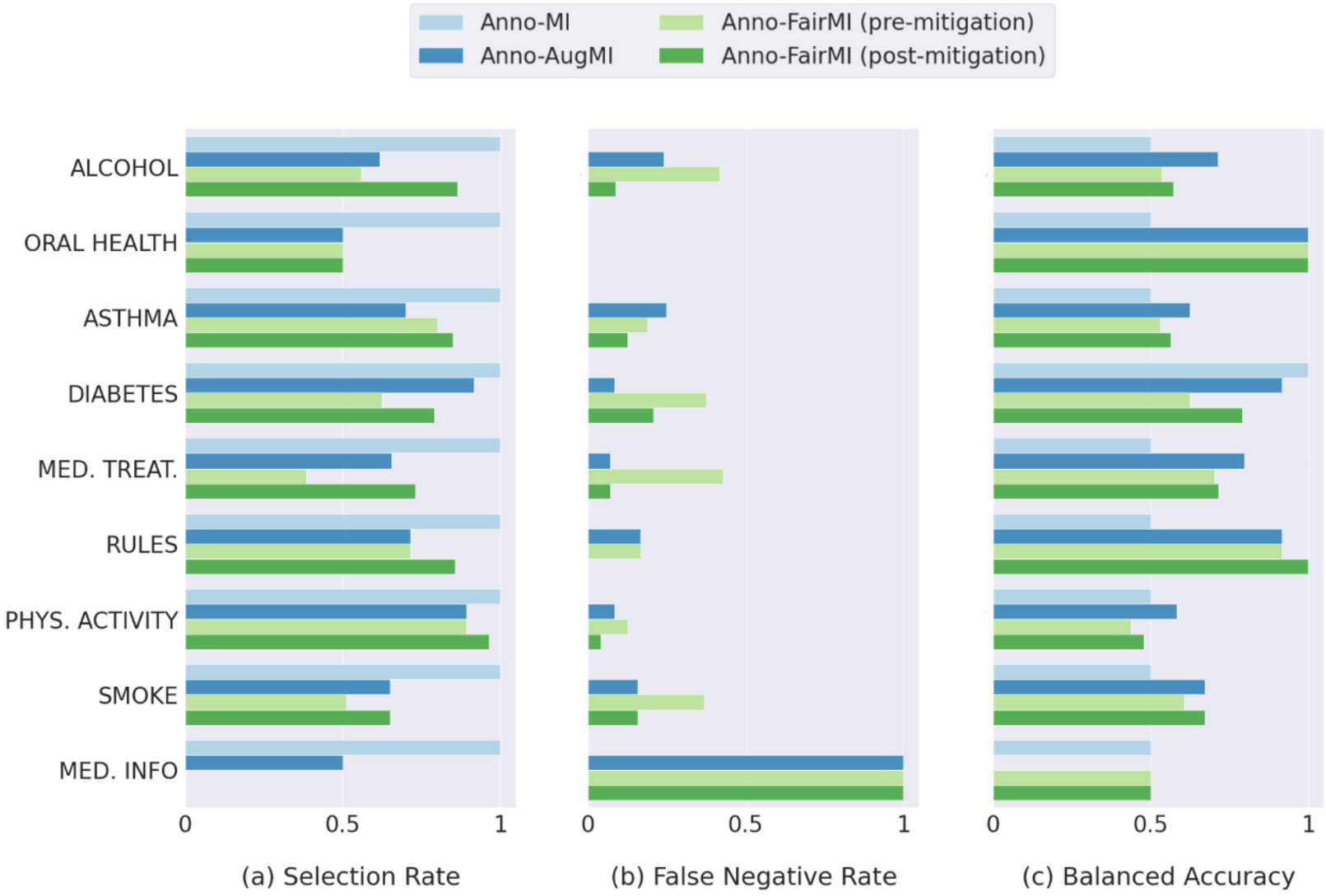
(b) Anno-FairMI (pre-mitigation)



(d) Anno-FairMI (post-mitigation)

Confusion matrix for the Bi-LSTM trained on each dataset. For Anno-FairMI we provide pre and post-mitigation matrix.

RESULTS



Fairness assessment and BIAS mitigation for Bi-LSTM on each dataset.

Conclusions

- We evaluated our approaches on a classification task, aimed at recognising therapy quality.
- Our results show a promising accuracy increase for DL classifiers by using augmented datasets, especially Anno-AugMI.
- The obtained results motivate us to consider other target attributes in future works, such as client talk type or therapist behaviour, also extending to other tasks like forecasting.
- The fairness assessment and BIAS mitigation show that Anno-FairMI is too sensitive to unseen topics, opening interesting future work on the adoption of more advanced augmentation techniques.
- Overall, we consider target-aware augmentation effective at addressing the challenges of unbalanced and scarce data in the mental health domain.

Thank You