Biases and health inequalities pose a complex problem for infection AI models.

HARD OUTCOMES

MORTALITY

LENGTH OF STAY

There is a strong association between **sensitive attributes** (those not linkable or discriminatory) such as sex, socioeconomic status, ethnicity and and **significant health inequalities** including an individual's **infection risk**, **outcomes** and **antimicrobial resistance**

There is great concern artificial intelligence (AI) models developed to date suffer from **bias and lack of generalization**, due to the **excising inequalities** and biases engrained within training datasets. Thus, when developing AI solutions, it is important to ensure they are **un-biased** through **fairness metrics**

Equalized odds (EO) can be considered the most relevant **measure of fairness** in this scenario given we want to acknowledge and ideally minimize **false positives** (i.e., predicting survival for patients who die) as well as obtain **equal performance** across sensitive attributes groups

A RNN model was created for mortality and length of stay prediction using MIMIC-IV.

Dataset

HARVARD MEDICAL SCHOOL
Beth Israel Lahey Health
Beth Israel Deaconess Medical Center

- MIMIC-IV electronic health record database

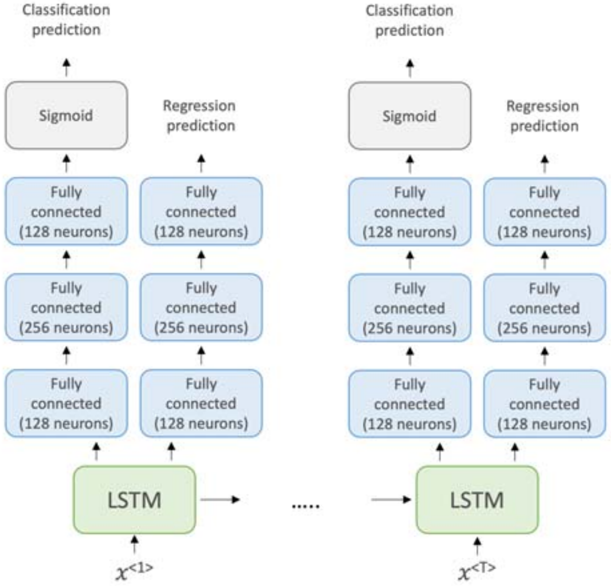
Population

- Patients who received **antibiotics** during an **ICU** stay

- Input features included **lab test** results and **clinical parameters**
- Features were normalised, **aggregated by day** for each unique stay
- Data split into training , validation and testing sets
- Many-to-many long short-term memory recurrent neural network (**LSTM-RNN**) was used as it considers the temporal nature of medical data
- Entire stay (**sequence** of days) used as an input

43 FEATURES → MORTALITY; LENGTH OF STAY (LOS)

Many-to-many RNN Model Architecture



Results were broken down by the **sensitive attribute classes** sex, socioeconomic status (i.e., insurance type), and ethnicity to evaluate the **equalized odds** (EO) **fairness** of the model

The model demonstrated some fairness across sex, but ethnicity biases were present.

- To attain **equalized odds** (EO) the true positive rate (TPR) across groups within a sensitive attribute class and the false positive rate (FPR) **across groups** must be equal or at least similar, meaning the model has balanced performance across the sensitive attribute
- Performance across ethnicities was **not very consistent**, with model outputs particularly differing between those groups **frequently and infrequently present** in the dataset such as white and native American or Asian populations (Native FPR undefined due to no individual dying in the test set)

Sensitive attribute	**Sex**		**Socioeconomic status**			**Ethnicity**						
Groups	Male	Female	Medic-aid	Medi-care	Other	White	Black	Hispanic	Asian	Native	Other	Un-known
True positive rate	0.82	0.79	0.82	0.76	0.85	0.85	0.78	0.78	0.79	0.57	0.71	0.63
False positive rate	0.42	0.44	0.34	0.43	0.45	0.48	0.41	0.45	0.30	NaN	0.30	0.28

Next steps include investigating appropriate action to reduce model biases.

Conclusion

- The model demonstrated **some equalized odds** (EO) fairness across sex, but **ethnicity biases** were present

- Biases within AI models are common particularly against **minority groups**

Future Work

- Recognise the inherent diversity and **discover bias** within infection related datasets and algorithms and take appropriate action to ensure they are **representative**

- Investigate methods such as **fairness constraints** and **fair adversarial representation learning** to best mitigate model biases and health inequalities, particularly against minority groups, to obtain **consistent performance** across the **intended patient population**

Contact email: william.bolton@imperial.ac.uk