**Agenda:**

* Sensitivity analysis results
* Variation among base estimators
* Most important feature subsets
* Comparison to literature
* Figures and discussion of what to include in the poster

**Stacking Ensemble Results:**

* The random forest ensemble and linear regression ensemble models both have the best performance in different metrics. I chose the random forest ensemble because it had the highest performance in the metrics least influenced by outliers (relative error and MAE)

A table with numbers and letters

AI-generated content may be incorrect.



**Sensitivity Analysis:**

* Only trained the best-performing models on data from:
  + The lowest income level
  + The lower middle-income level
  + The upper middle-income level
  + The highest income level
* These re-training procedures were done as separate experiments.
* The retrained models’ performance on their specific dataset was compared to the random forest ensemble model trained on all data and tested on the sensitivity analysis dataset (i.e. full dataset filtered for a specific income level)

Split by country:

* Generally, the original model had higher error
  + As expected, as the model trained solely on countries from the same income level would be able to identify patterns without ‘noise’ from other income levels distorting the pattern.
    - It could be easier to isolate one pattern versus isolate and distinguish many.
* The difference in performance between the sensitivity analysis and original low-income models was smaller than the difference in performance between the models for the high and upper middle-income levels.
  + This may be due to overfitting of the low-income model, as there was less than 100 low-income samples in the dataset. The overfitting may have created a ‘performance cap’.
* The original model had smaller error compared to the sensitivity analysis model when tested on the lower middle-income dataset.
  + Potentially, countries in this data could reflect attributes more common among the low-income countries and/or the higher income countries, making it difficult to distinguish a single pattern that fit all the lower-middle countries. This may explain why having data from countries in a broader range of income levels may produce better performance.

A graph of a bar graph

AI-generated content may be incorrect.

Split by year:

* Unfortunately, many of the folds in the high-income dataset for correlation >0.8 did not have any remaining feature columns (none of the 11 features had sufficient non-missing data), meaning that there were too few predictions to use the original ensemble model.
* The difference between the two models was less prominent when the data was split by year, potentially due to between country differences being more relevant than across year differences
  + Alternatively, the countries may have all shown differences across years, making income level differences less relevant for prediction tasks
    - This may also explain why the sensitivity analysis now does better for the lower middle-income countries than the model trained on all data.

A graph of a bar graph

AI-generated content may be incorrect.

**Variance among base estimators**

* Variance increases as MMR increases, reflecting the small amount of data available for lower income, higher MMR countries

**A graph with blue dots

AI-generated content may be incorrect.**

**Feature Importance in Models Used by the Random Forest Ensemble**

Most important features for base models valued most highly by random forest ensemble.

Split by country

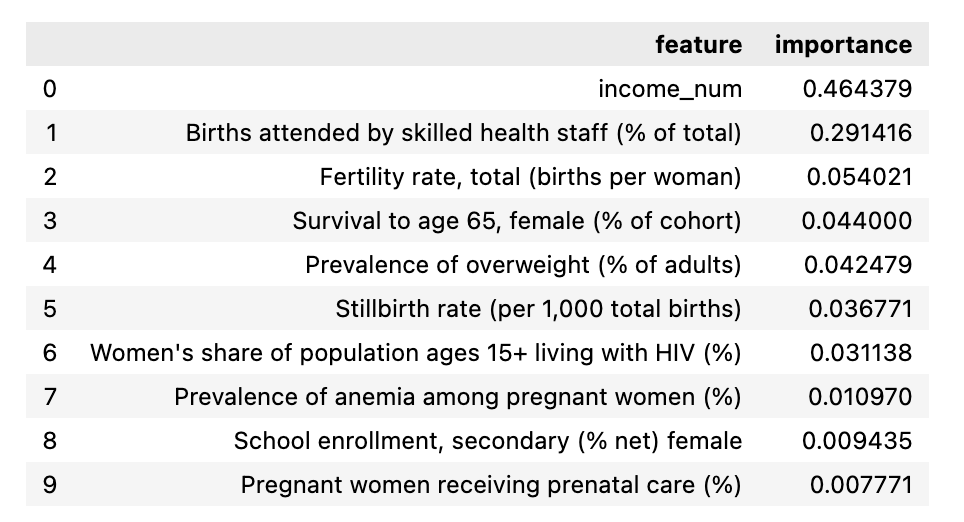
Most valued model by random forest ensemble:

A close-up of a list of women

AI-generated content may be incorrect.

* This and the 3rd most valued model place most of their importance on 1 feature
* Many of these features are valued in the literature (contraception prevalence, infant mortality, literacy, wasting, HIV)

Second most valued model by random forest ensemble:



Third most valued model by random forest ensemble:

A close-up of a table

AI-generated content may be incorrect.

Negligible contribution to random forest ensemble model:

A screenshot of a table

AI-generated content may be incorrect.

* Put most of its importance on a single feature
* Included mortality due to other factors (e.g. traffic injuries)
* Places more importance on factors involving males than the more valued models

Split by year:

Most valued model by random forest ensemble:

A close-up of a white background

AI-generated content may be incorrect.

Second most valued model by random forest ensemble:

A screenshot of a white and black text

AI-generated content may be incorrect.

* Very similar to most important factors for the split by country data
  + Potentially, because the split by country data already relies on factors that are important for long term health (e.g. labour and education), not just short-term country-specific factors

**Comparison to literature values:**

During my literature review, I identified 3 methods for estimating the maternal mortality ratio.

* Micro-simulation based on reproductive lifecycles
* Global Burdan of Disease Study
* UN MMEIG Estimates

These studies all gave upper and lower bounds for their estimates. I calculated the percentage of my model’s estimates that fell within the literature estimates’ 95% confidence intervals. I also calculated the percentage of national estimates (the values my model has been trained to predict) fall within these confidence intervals.

|  |  |  |
| --- | --- | --- |
| **Literature Estimate** | **Percentage Coverage of My Estimates (%)** | **Percentage Coverage of National Estimates (%)** |
| Micro-simulation | 64.18 | 64.18 |
| Global Burden of Disease | 27.97 | 32.20 |
| UN MMEIG | 22.88 | 20.76 |

As a further point of comparison, a study estimating the sub-national maternal mortality rates in Kenya using a multiple regression linear model achieved a MSE of 30 per 100,000 live births and an R-squared value of 75.5%. In contrast, my model had an MSE of 2,161 deaths per 100,000 live births and an R-squared value of 94.4%.

The following plots show the percentage difference between my predictions and the literature’s estimates. It was calculated as:

The plots demonstrate that the percentage difference was generally negative, indicating that my estimates were smaller than those in the literature. This translates to my model predicting lower maternal mortality ratios than given in the literature. Likely, this is due to the lack of low-income, high MMR data in my dataset (only 78 samples from the lowest income countries, and as seen in the table below, these samples had to cover a very large range of values).



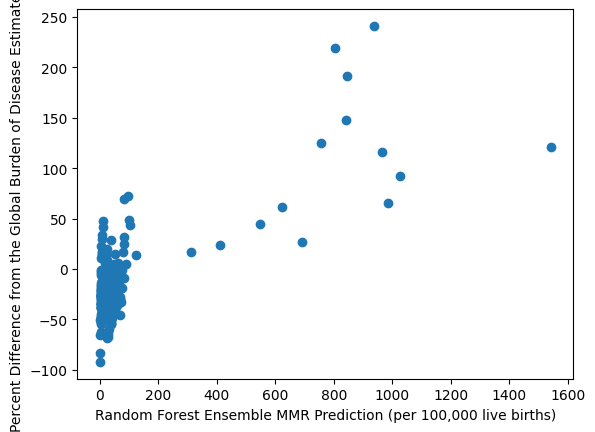
Micro-simulation

A graph of blue dots

AI-generated content may be incorrect.A graph of a number of blue bars

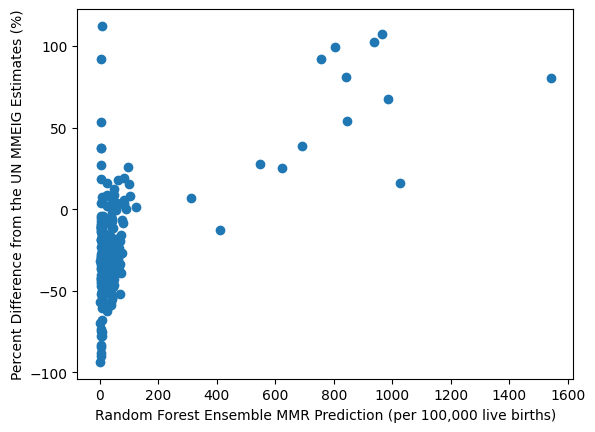
AI-generated content may be incorrect.

Global burden of disease

 A blue graph with white text

AI-generated content may be incorrect.

UN MMEIG

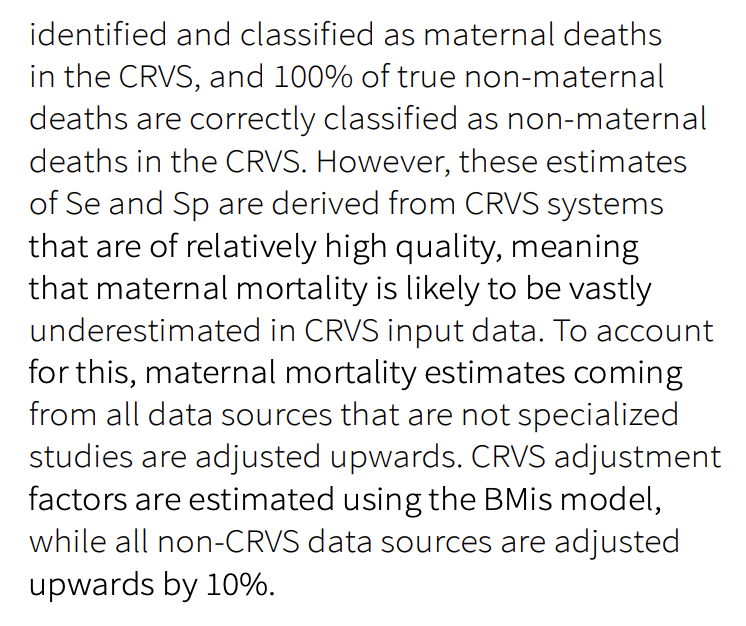
 A blue graph with white text

AI-generated content may be incorrect.

**Interpretations:**

* This difference from the literature was relatively expected, as every study involving the estimate of maternal mortality data reports the limitations imposed by under-reporting and misclassification of maternal deaths in countries’ national civil registration systems. Therefore, the three models shown above all include a component for adjusting the national estimates of maternal deaths to increase the number of maternal deaths to a more realistic level. In contrast, my model solely predicts the nationally reported maternal mortality statistics, meaning that the predictions have not been adjusted for underreporting. For example, in the report detailing the UN MMEIG model’s methods:

A close-up of a white text

AI-generated content may be incorrect. 

* + Possible ways of dealing with this in my thesis:
    - As a future extension of my model, state that it would be interesting to combine my model to estimate national statistics with a secondary model that estimates the adjustment factor needed to account of underreporting (reminiscent of the UN MMEIG BMis model). My model will save effort of collating national statistics, and when no maternal mortality data is found for a country, the estimates will be driven by a higher number of socio-economic and health-related covariates than the literature models.
    - Try to do something like this now.
* A previous study found that the micro-simulation and UN MMEIG results were very similar, and higher than the estimates produced by the GBD study. This bears out in the results presented above, as the GBD estimates had a higher range of positive percentage differences with my results than the other models.
  + This is shown in the following figure.
  + Additionally, this figure shows that the GMatH model (micro-simulation) had much larger error bounds than the UN MMEIG or GBD models, potentially explaining the much higher percentage of my results that are within the bounds of the micro-simulation model than the other models.

A graph of the birth rate of the infant

AI-generated content may be incorrect.

<https://www.nature.com/articles/s41591-023-02310-x/figures/3>

**Figures:**

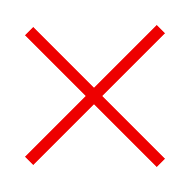
* What to include in the poster:
  + Which results and which segments of the discussion?
* Model results (show both split by country and split by year)
* PCA
* Comparison to literature plots
* Sensitivity analysis



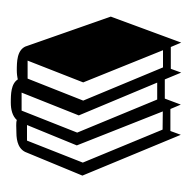
No feature selection

Merged data

Feature selection



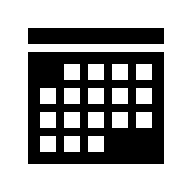
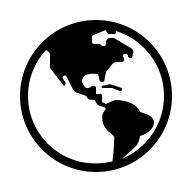
From literature



Correlation Strength

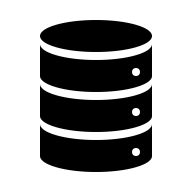


Incoming data for train/test split



Split by country

Split by year



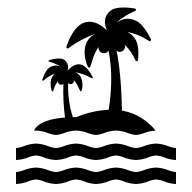
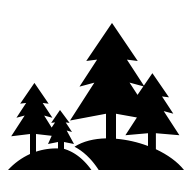
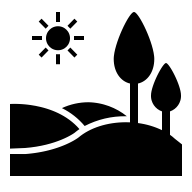
Splitting into 5 cross-validation folds

No missing data removal

90% missing data threshold

95% missing data threshold

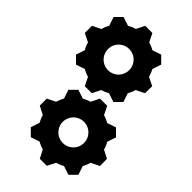
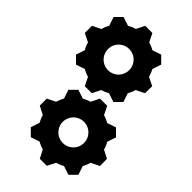
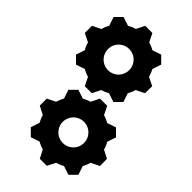
85% missing data threshold



Random Forest

XGBoost

LightGBM



SVM Stacking Ensemble

Elastic Net Stacking Ensemble

Random Forest Stacking Ensemble

Voting Ensemble

Ensemble Model

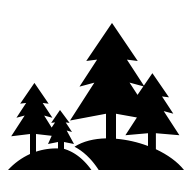
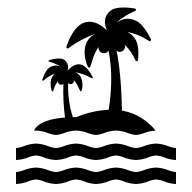
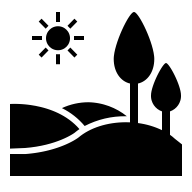
Ensemble Model:

Combines 300 predictions from base estimators

100 Predictions

100 Predictions

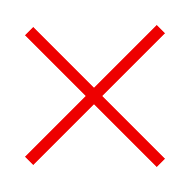
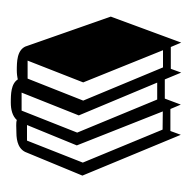
100 Predictions



Random Forest

XGBoost

LightGBM



Final Ensemble Model

300 MMR Predictions

Final MMR Prediction

1

1

1

1

1

Fold 0

Fold 4

Fold 3

Fold 2

Fold 1

5

5

5

5

No Missing Data Threshold

Missing Data Threshold = 95%

Missing Data Threshold = 90%

Missing Data Threshold = 85%

20

20

20

20

20

Correlation >0.8

Correlation >0.7

Correlation >0.6

Literature

No Feature Selection