Selective Regression Under Fairness Criteria

Abhin Shah, Yuheng Bu, Joshua Ka-Wing Lee, Subhro Das,
Rameswar Panda, Prasanna Sattigeri, Gregory W. Wornell

MIT-IBM Watson AI Lab

Selective Prediction

Prediction with a reject-option

Selective Prediction

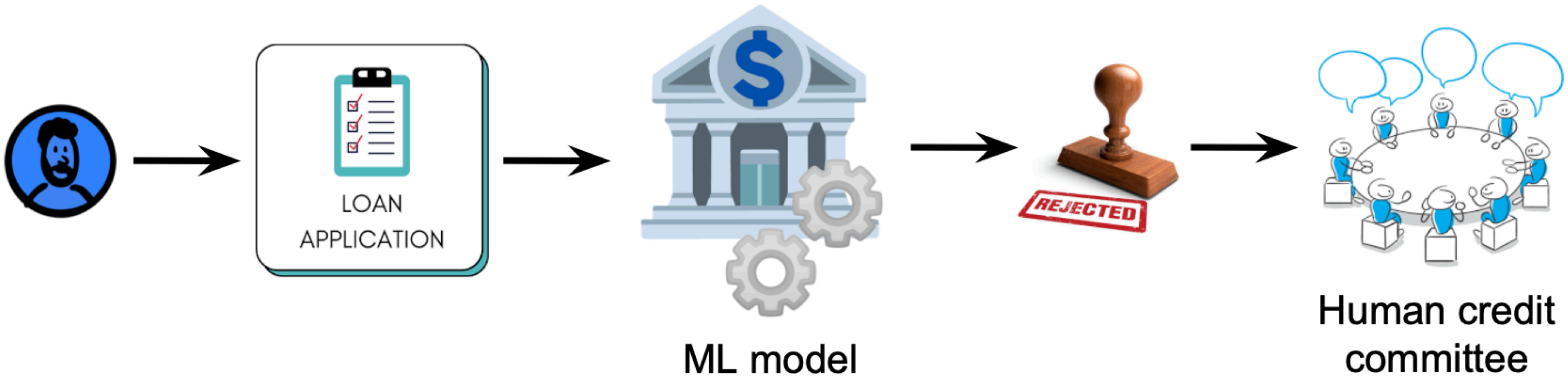
Prediction with a reject-option

- A trustworthy machine learning system → reliably communicate the uncertainty in its predictions.

Selective Prediction

Prediction with a reject-option

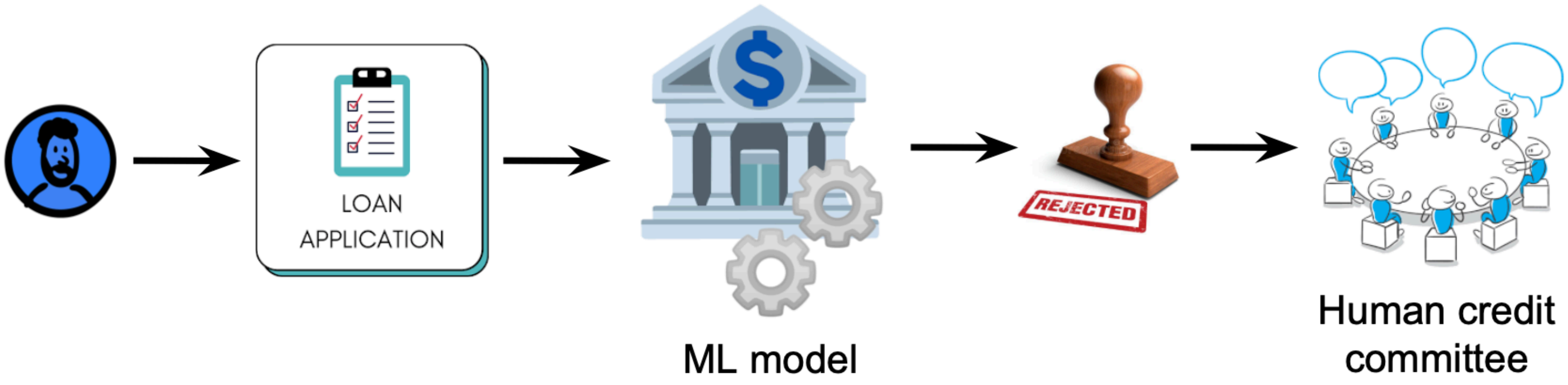
- A trustworthy machine learning system → reliably communicate the uncertainty in its predictions.



Selective Prediction

Prediction with a reject-option

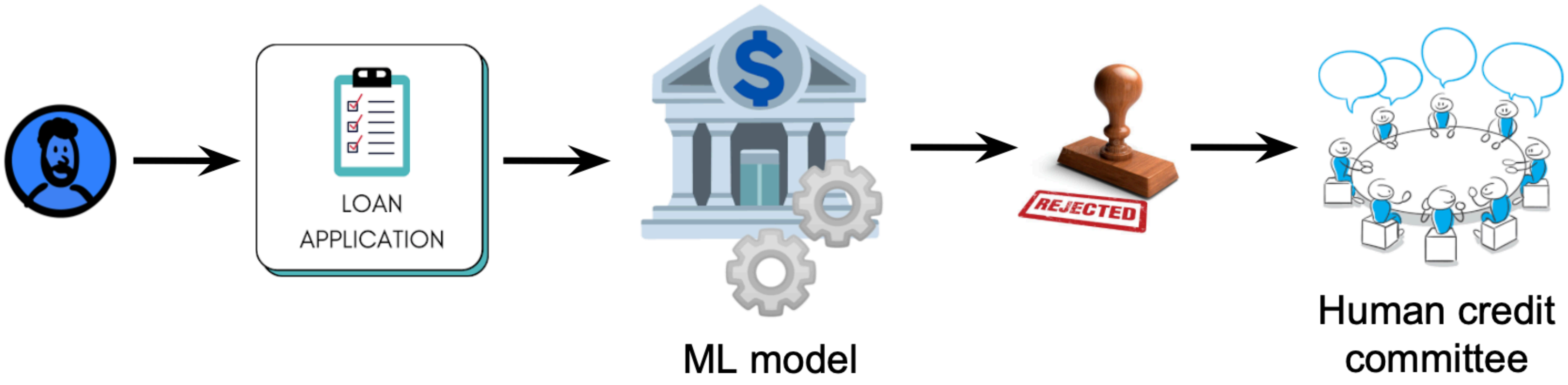
- A trustworthy machine learning system → reliably communicate the uncertainty in its predictions.
- If the uncertainty in a prediction is high → the prediction can be rejected to avoid potentially costly errors.



Selective Prediction

Prediction with a reject-option

- A trustworthy machine learning system → reliably communicate the uncertainty in its predictions.
- If the uncertainty in a prediction is high → the prediction can be rejected to avoid potentially costly errors.
- Selective prediction → can abstain from making a decision



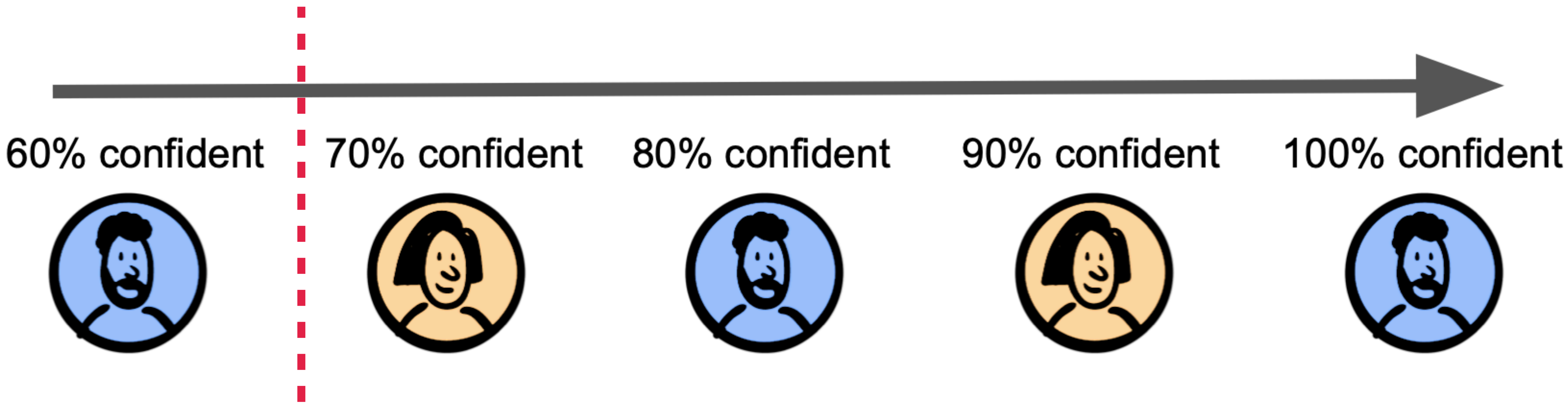
Selective Prediction

Prediction with reject-option

Selective Prediction

Prediction with reject-option

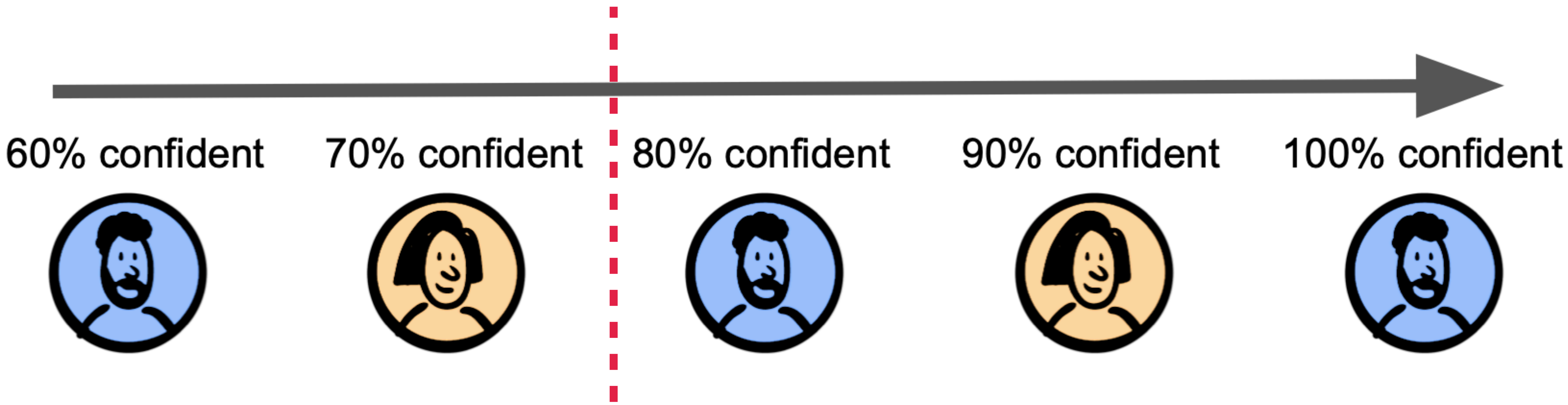
- If we have confidence measure for each prediction → abstain from decision making if our confidence is below a certain threshold.



Selective Prediction

Prediction with reject-option

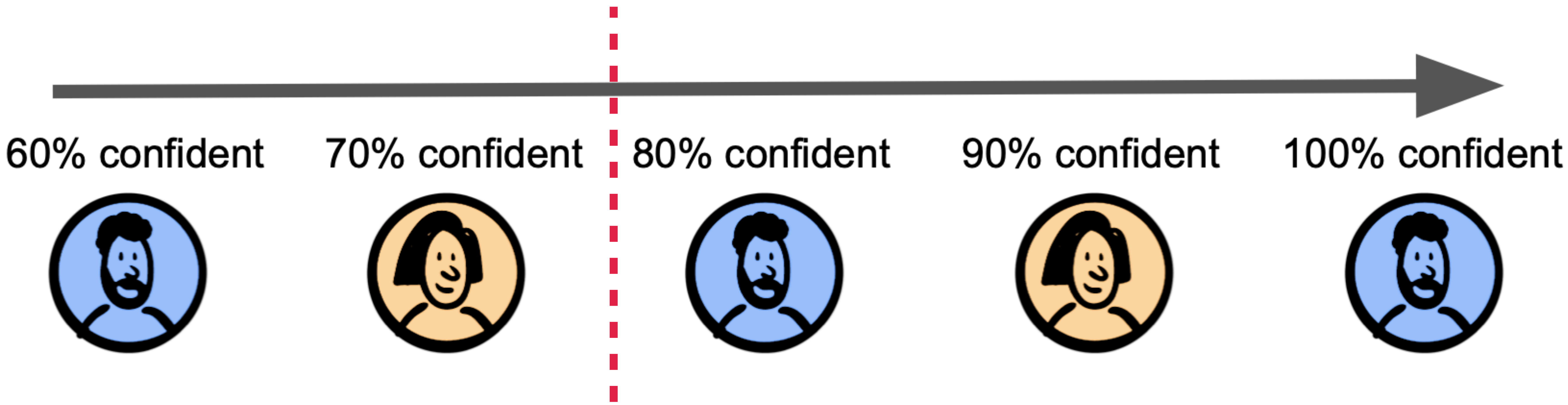
- If we have confidence measure for each prediction → abstain from decision making if our confidence is below a certain threshold.
- With a good confidence measure → increasing the threshold results in a better performance.



Selective Prediction

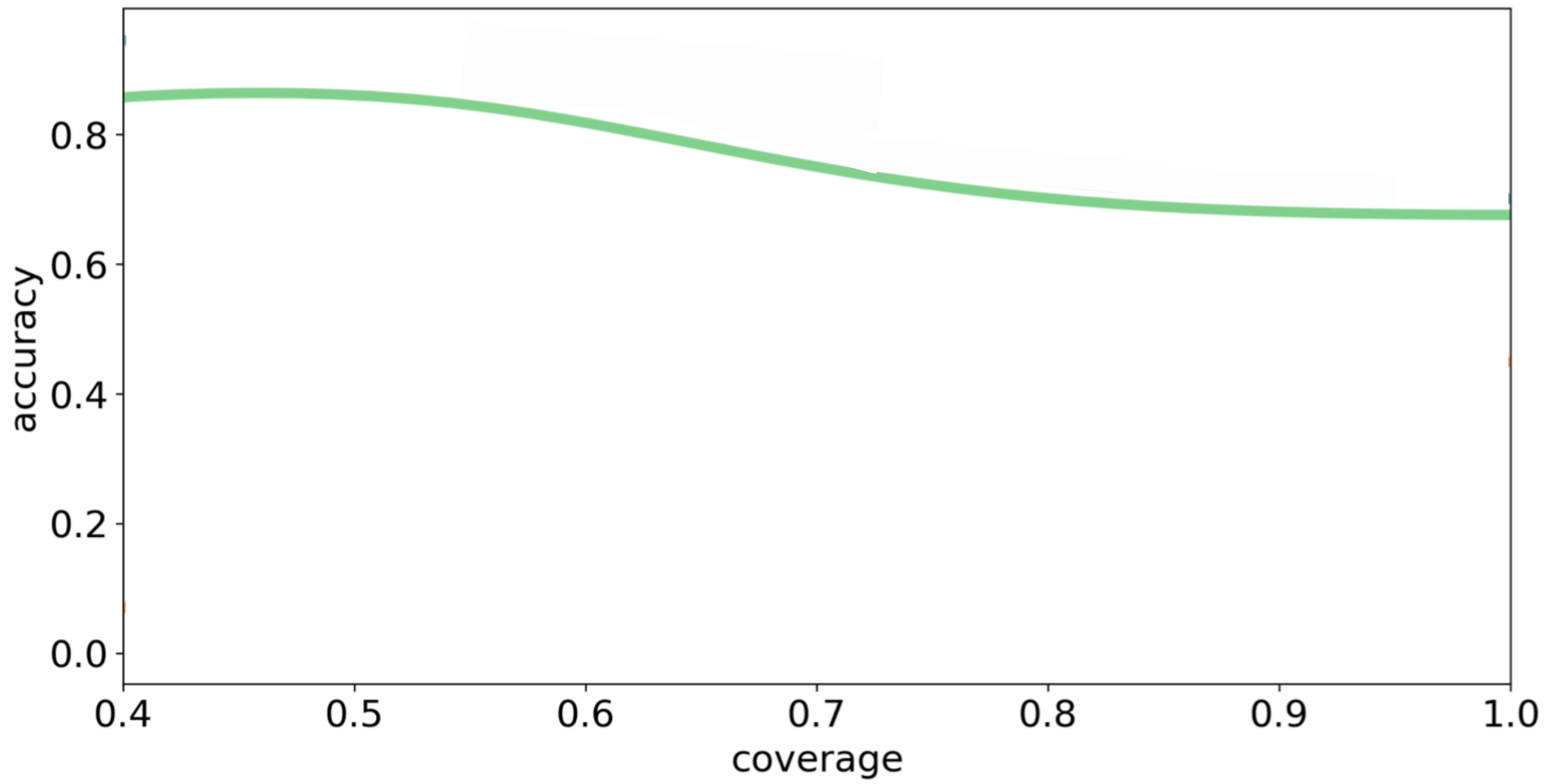
Prediction with reject-option

- If we have confidence measure for each prediction → abstain from decision making if our confidence is below a certain threshold.
- With a good confidence measure → increasing the threshold results in a better performance.
- Tradeoff → we have predictions for a fewer samples (i.e., low coverage).



Selective Classification

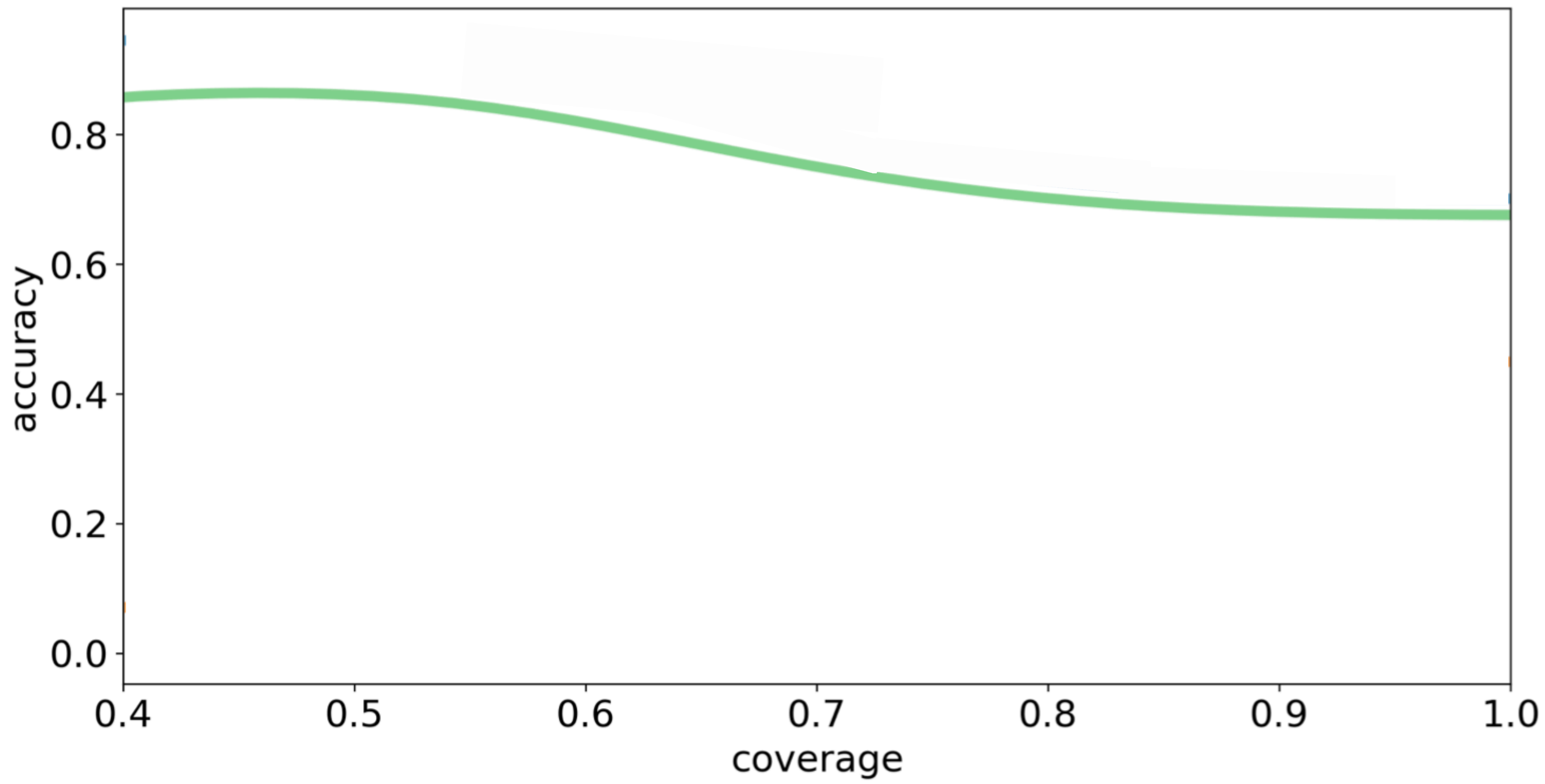
Prior Work



Selective Classification

Prior Work

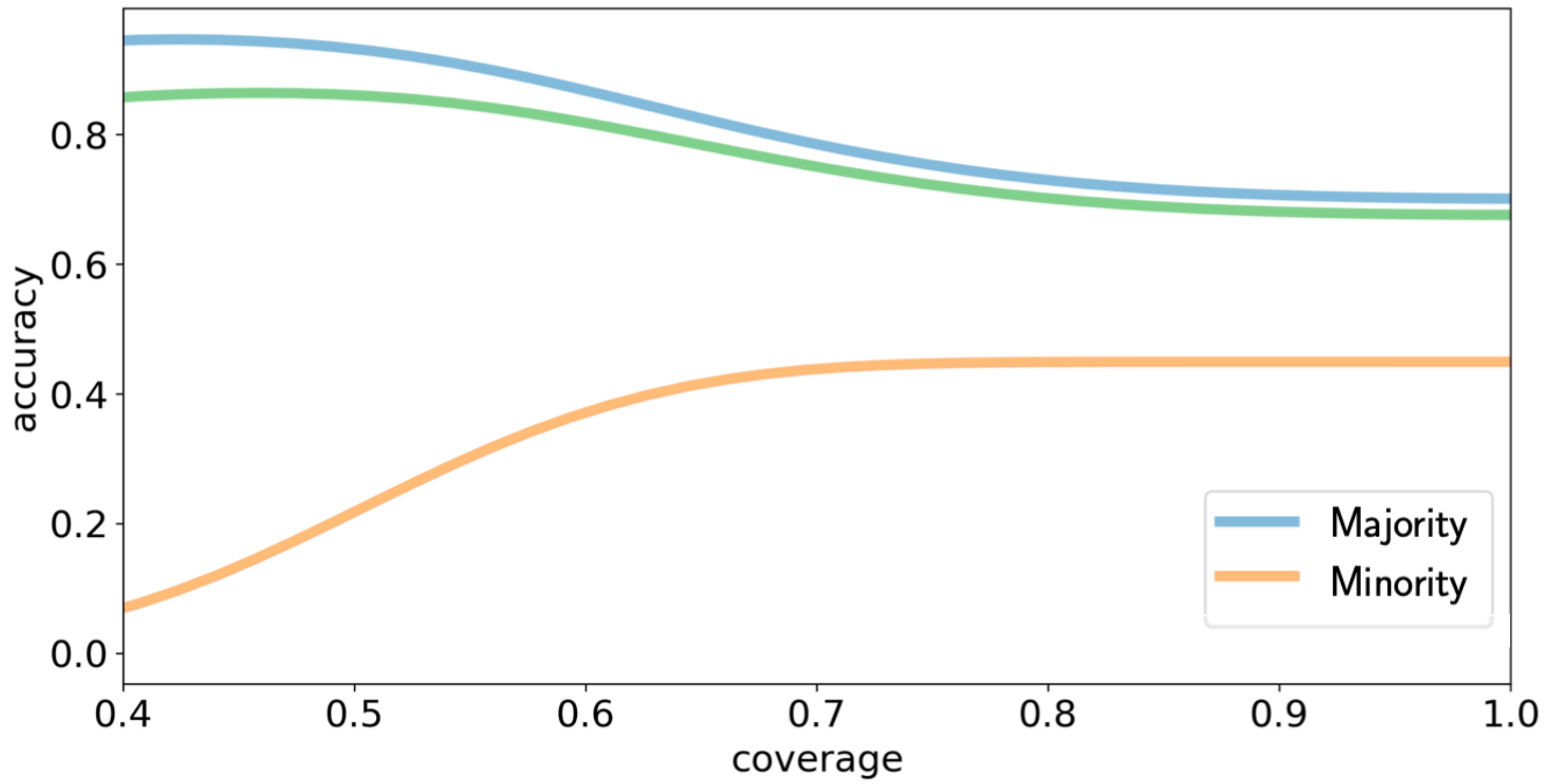
- Classifiers can have good average performance but may perform poorly on certain protected / sensitive groups [Jones et al. 2020].



Selective Classification

Prior Work

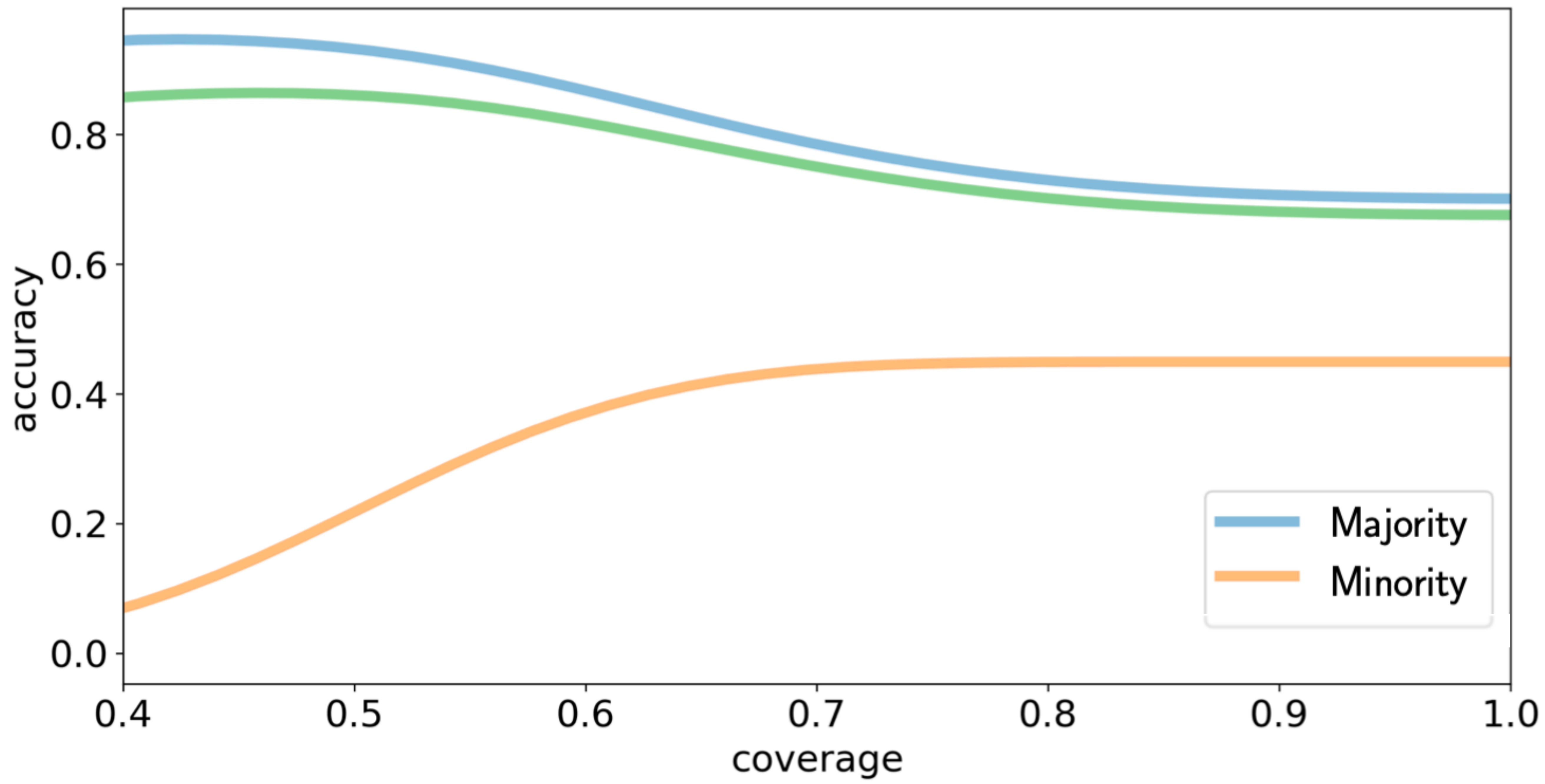
- Classifiers can have good average performance but may perform poorly on certain protected / sensitive groups [Jones et al. 2020].



Selective Classification

Prior Work

- Classifiers can have good average performance but may perform poorly on certain protected / sensitive groups [Jones et al. 2020].



- To mitigate such disparities, recent works [Lee et al., 2021; Schreuder & Chzhen, 2021] proposed methods for performing fair selective classification.

Selective Regression

Designing an Uncertainty measure

Selective Regression

Designing an Uncertainty measure

Classification → learned using the softmax output (of an existing classifier)

Selective Regression

Designing an Uncertainty measure

Classification → learned using the softmax output (of an existing classifier)

Regression → no direct method to learn from an existing regressor designed only to predict the conditional mean

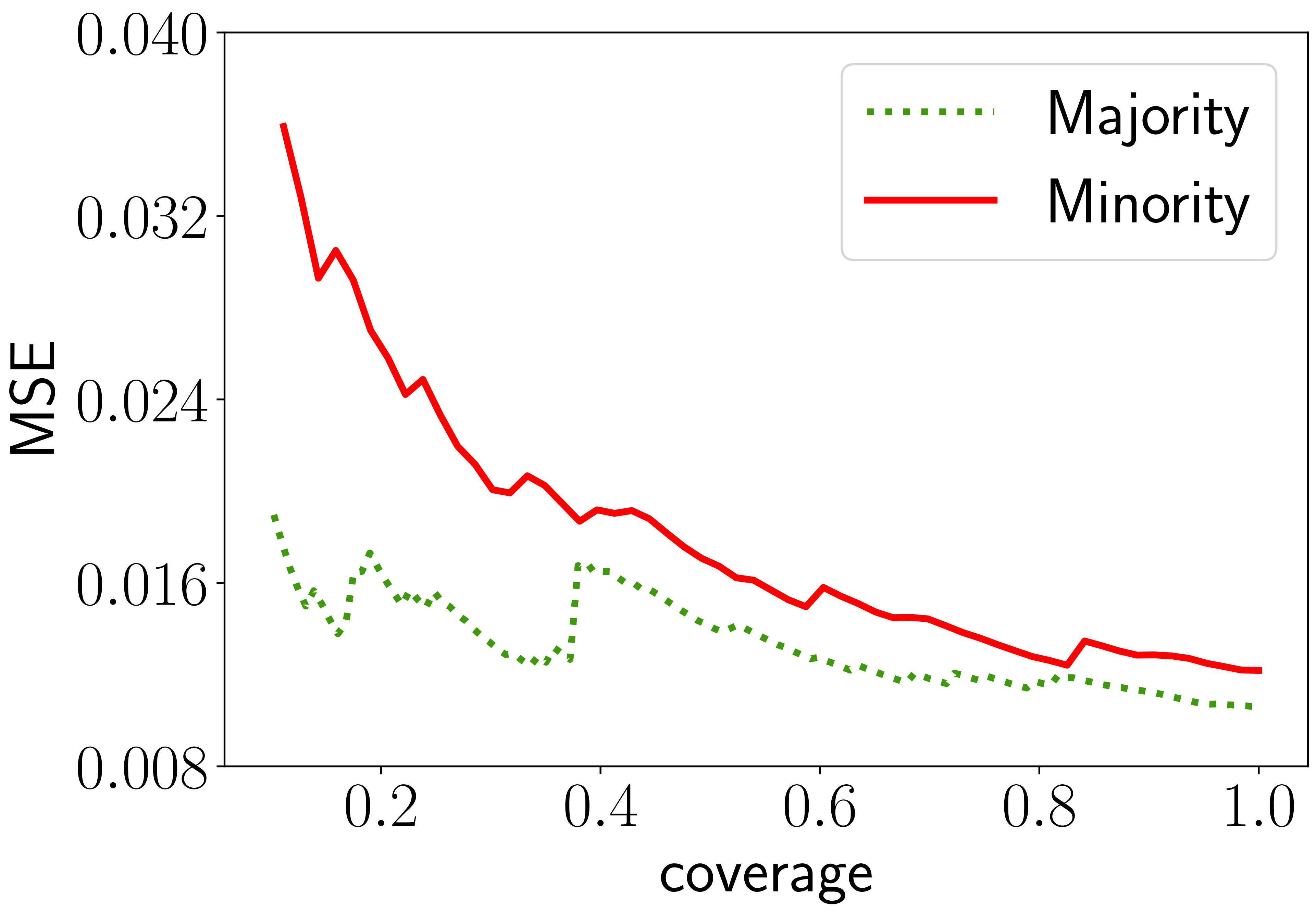
Biases in Selective Regression

Contributions

Biases in Selective Regression

Contributions

- We show that selective regression, like selective classification, can decrease the performance of some protected groups when coverage is reduced.



Insurance dataset

Fair Selective Regression

Contributions

Fair Selective Regression

Contributions

- Monotonic selective risk → requires the risk (i.e., MSE) of each subgroup to monotonically decrease with a decrease in coverage.

Fair Selective Regression

Contributions

- Monotonic selective risk → requires the risk (i.e., MSE) of each subgroup to monotonically decrease with a decrease in coverage.
- Monotonic selective risk is met if a feature representation:

Fair Selective Regression

Contributions

- Monotonic selective risk → requires the risk (i.e., MSE) of each subgroup to monotonically decrease with a decrease in coverage.
- Monotonic selective risk is met if a feature representation:
    1. satisfies the standard sufficiency criterion or

Fair Selective Regression

Contributions

- Monotonic selective risk → requires the risk (i.e., MSE) of each subgroup to monotonically decrease with a decrease in coverage.
- Monotonic selective risk is met if a feature representation:
    1. satisfies the standard sufficiency criterion or
    2. is calibrated for mean and variance.

Fair Selective Regression

Contributions

- Monotonic selective risk → requires the risk (i.e., MSE) of each subgroup to monotonically decrease with a decrease in coverage.
- Monotonic selective risk is met if a feature representation:
  1. satisfies the standard sufficiency criterion or
  2. is calibrated for mean and variance.
- Two algorithms:

Fair Selective Regression

Contributions

- Monotonic selective risk → requires the risk (i.e., MSE) of each subgroup to monotonically decrease with a decrease in coverage.
- Monotonic selective risk is met if a feature representation:
  1. satisfies the standard sufficiency criterion or
  2. is calibrated for mean and variance.
- Two algorithms:
  1. impose the sufficiency criterion by regularizing an upper bound of conditional mutual information.

Fair Selective Regression

Contributions

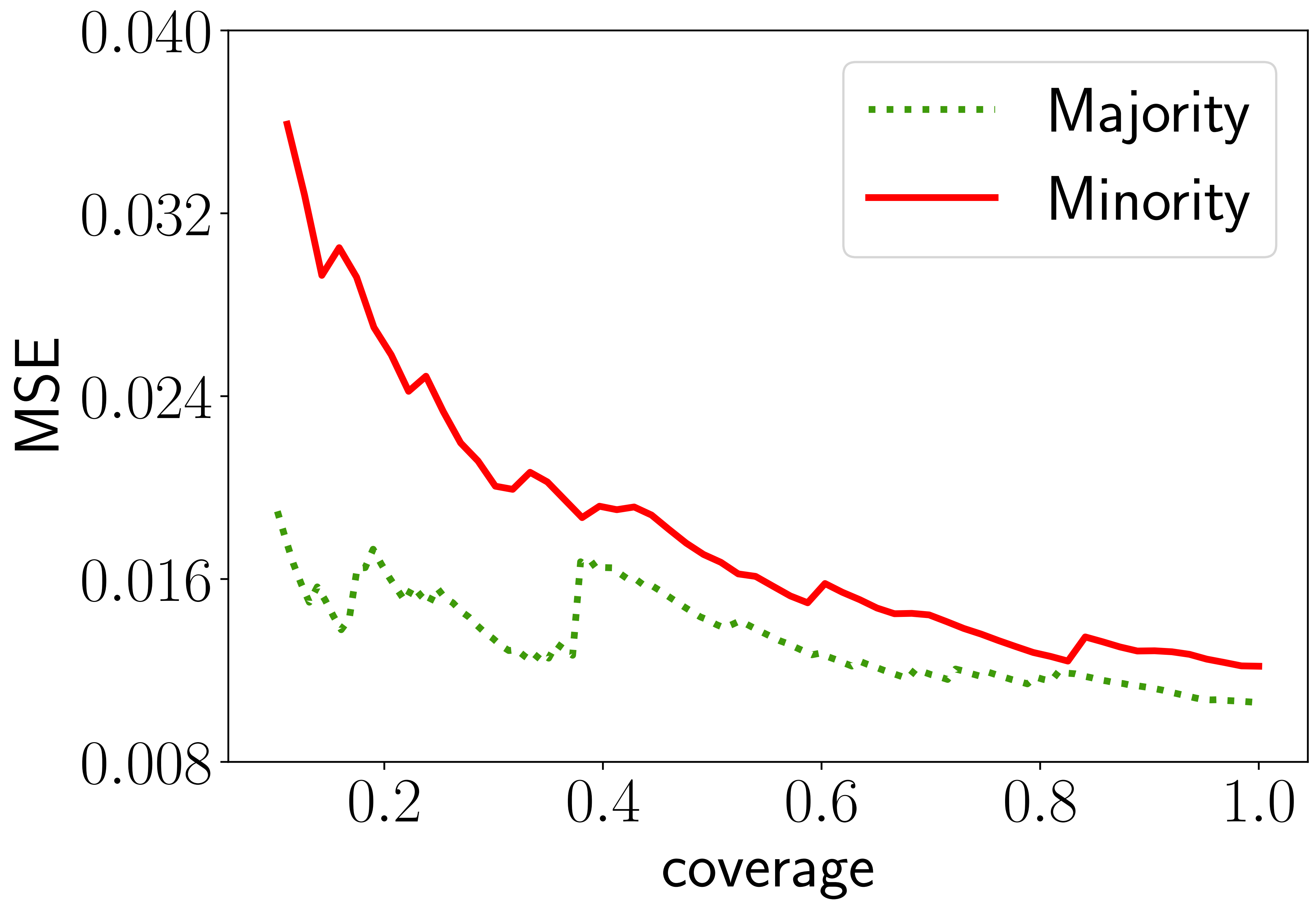
- Monotonic selective risk → requires the risk (i.e., MSE) of each subgroup to monotonically decrease with a decrease in coverage.
- Monotonic selective risk is met if a feature representation:
  1. satisfies the standard sufficiency criterion or
  2. is calibrated for mean and variance.
- Two algorithms:
  1. impose the sufficiency criterion by regularizing an upper bound of conditional mutual information.
  2. impose the calibration for mean and variance by regularizing a contrastive loss.

Empirical Results

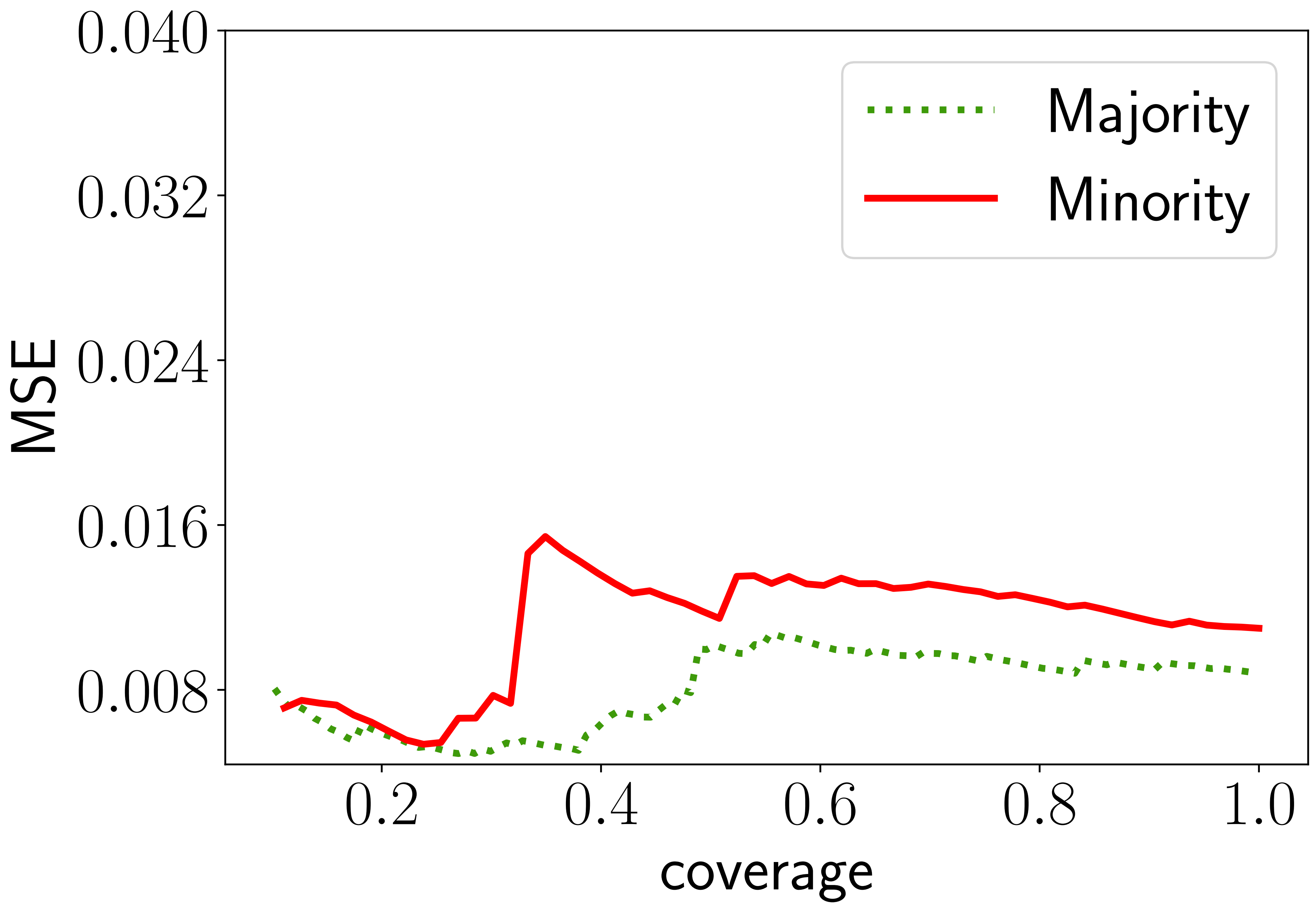
Insurance dataset

Empirical Results

Insurance dataset



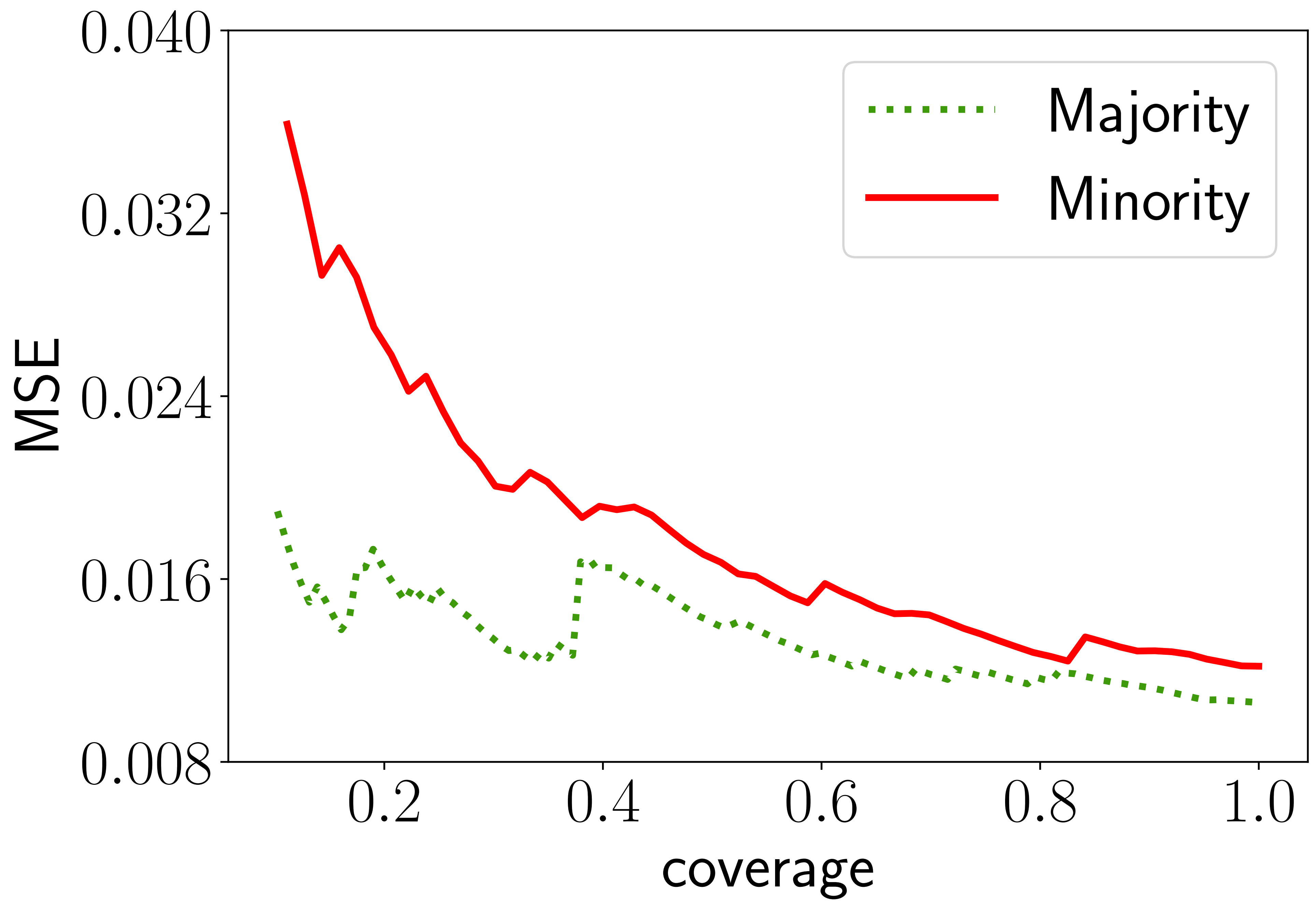
Baseline



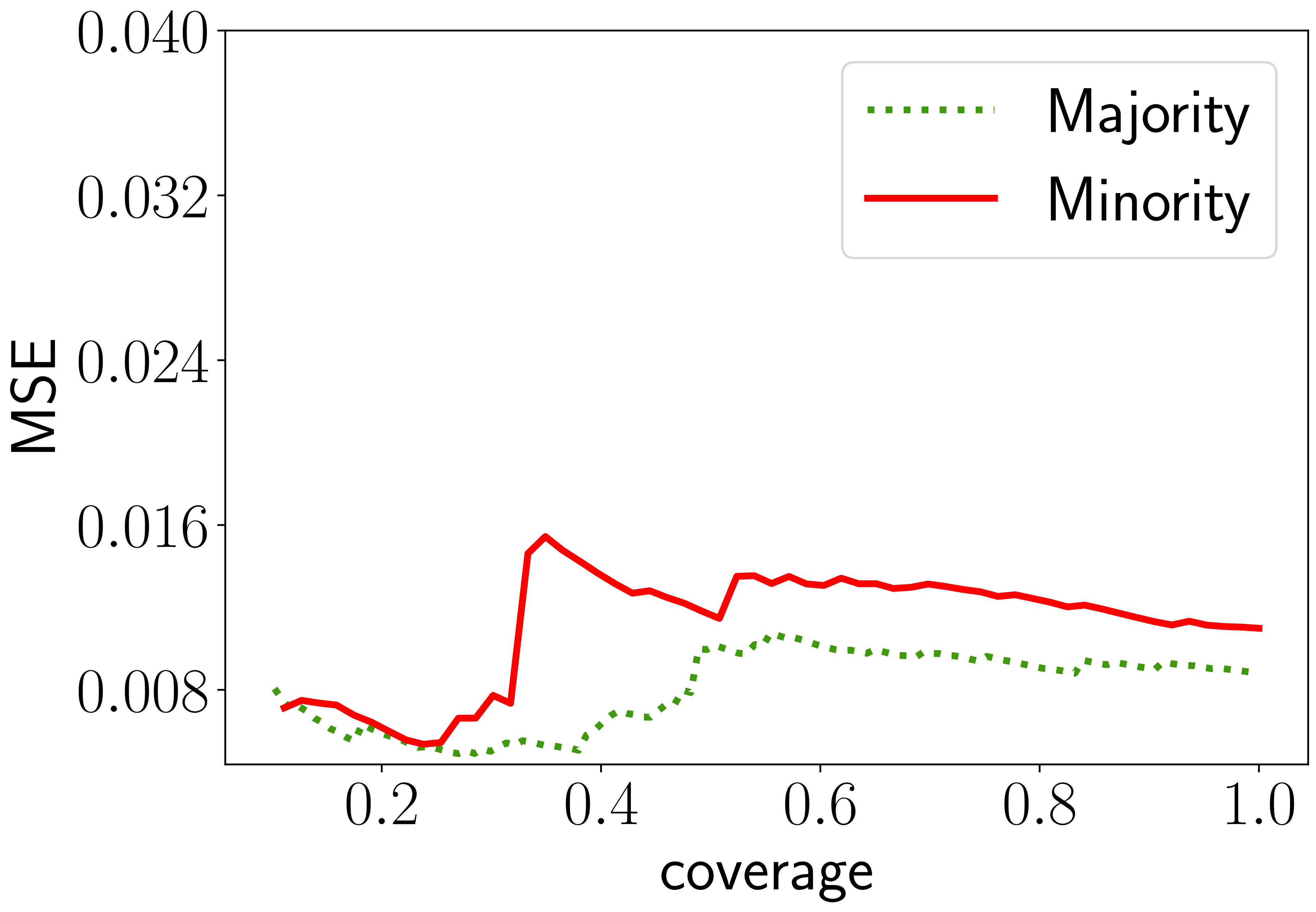
Our method

Empirical Results

Insurance dataset



Baseline



Our method

Poster #1108