CS-483: Big Data Mining - Progress Report
Interpreting Algorithmic Fairness via Aleatoric and Epistemic Uncertainties

Zohair Hashmi, Sajal Chandra, Rayaan Siddiqi and Sean Kudrna, University of Illinois at Chicago
Github Code: https://github.com/zohairhashmi/CS483-uncertainty-quantification.git

1. Introduction

1.1 Problem Description

The problem of fairness and uncertainty optimization via uncertainty characterization relates to the challenge of ensuring fairness and minimizing uncertainty in decision-making processes that rely on machine learning models. Machine learning models often rely on large amounts of data to make predictions or decisions, but this data can be biased or incomplete, leading to unfair or inaccurate outcomes. Additionally, there is often uncertainty in the data or the model itself, which can further complicate decision-making processes, this is prevalent in the problem of distinguishing between epistemic and aleatoric uncertainty. The goal of solving this problem is to develop techniques and algorithms that can characterize and quantify uncertainty, and use this information to optimize decision-making processes to ensure fairness and accuracy. By addressing this problem, we can improve the reliability and effectiveness of machine learning models in a wide range of applications, from healthcare to finance to criminal justice.

1.2 Goal

The goal of minimizing uncertainty via uncertainty characterization is to improve the reliability and accuracy of decision-making processes found commonly in the Machine Learning field. ML models often produce predictions with differing degrees of certainty which can lead to speculative outcomes. Through us performing characterization and quantifying the uncertainty in the model, we can develop strategies to minimize its negative impact on the model. At the end of the day, this allows for the overall performance of the model to improve and reduces the risk of incorrect decisions being matched and unintended consequences occurring. In this research we intend to address two important questions. 1) What is the relationship between aleatoric and epistemic uncertainty and the varying nature of the dataset? 2) How does uncertainty relate to different fairness metrics?

1.3 Motivation

The importance of uncertainty quantification of machine learning models in real-world applications cannot be overstated. For instance, in the field of autonomous driving, an ML model is used to make decisions and predictions about the car's surroundings, including detecting objects and pedestrians, predicting their behavior, and making decisions such as whether to stop or continue driving. Uncertainty in these predictions can lead to potentially dangerous situations, such as misjudging the speed or direction of a moving object. By quantifying the uncertainty of the model's predictions, developers can better understand the model's reliability, make more informed decisions, and ultimately improve the safety of autonomous vehicles. Furthermore, uncertainty quantification can also help with interpretability and explainability, enabling developers to better understand how the model makes decisions and detect potential biases or inaccuracies in the data.

The significant impact that uncertainty can have on machine learning models in real-world applications motivated our research paper. The consequences of inaccurate predictions can range from minor inconveniences to serious safety hazards. Therefore, it is essential to develop methods for quantifying uncertainty and assessing the reliability of models to ensure accurate and fair decision-making. Our research paper aimed to address this issue by exploring various techniques for uncertainty quantification in machine learning models and their applications in different domains. By investigating the strengths and limitations of different approaches, we aimed to provide insights and recommendations for researchers and practitioners who work with machine learning models in various applications.

While uncertainty quantification is critical for developing machine learning models, ensuring fairness in decision making is also crucial, particularly in applications that impact human lives like we exemplified in our example. Bias and discrimination in machine learning models can have serious consequences, perpetuating societal inequities and causing harm to disadvantaged groups. Therefore, our motivation for looking into the relationship between uncertainties and fairness was to address this issue and promote the development of more equitable and fair machine learning models. Through us being able to investigate how uncertainty affects fairness and exploring techniques to quantify and mitigate uncertainty-related biases, we aimed to provide insights and recommendations for researchers in our respective field. Ultimately, our goal was to contribute to the development of more reliable and accurate machine learning models in the future.

2. Tools used

2.1.1 AIF360 Fairness Metric

For quantification of fairness metrics we are using AIF 360 library. It is an open-source toolkit developed by IBM to help detect and mitigate bias in machine learning models. The library provides a set of metrics, algorithms, and explanations to help identify and reduce bias across the machine learning pipeline, from data preprocessing to model training and post-processing.

After running our experiments on demo adult data provided by aif360 library, we have concluded to use the following final fairness metric for evaluating our results:

- **Equal Opportunity Difference (EOD):** Returns the difference in recall scores (TPR) between the unprivileged and privileged groups.

$$EOD = P(\hat{y}=1 \mid y=1, p=1) - P(\hat{y}=1 \mid y=1, p=0)$$

2.1.2 Keras_Uncertainty

Keras_uncertainty is a python library that provides a set of tools for quantifying uncertainty in deep learning models, specifically in the form of epistemic and aleatoric uncertainty. It provides functions and layers for computing and incorporating uncertainty estimates into the loss function during model training, as well as tools for evaluating and visualizing uncertainty in model predictions. By incorporating uncertainty estimates into deep learning models, Keras Uncertainty aims to improve the reliability and interpretability of these models for real-world applications. More on this in the following section.

2.2.1 Models

For the purpose of this project, we intend to utilize the following Deep Learning model to train our dataset and estimate uncertainties:

- **Deep Ensemble Learning**: Ensemble Learning is a technique where multiple models, called "base learners," are trained and combined to make predictions. By capitalizing on the strengths of individual models, this approach yields more accurate and stable results. In our research project, we applied Ensemble Learning to enhance the overall performance of our deep neural networks.

- **Drop Out Model**: Overfitting is a common challenge in deep neural networks, where the model becomes too specialized in the training data and struggles to generalize to new data. The Drop Out model is a regularization technique designed to address this issue. During training, a random subset of neurons is deactivated in each forward pass, which encourages the model's ability to generalize and make accurate predictions on new data. We incorporated the Drop Out model into our deep neural networks to improve their robustness.

- **Deep Ensemble Learning with Dropout Technique:** The project employs a combination of dropout and ensemble learning techniques, where the base learner is replaced by a dropout model. The ensemble model is iterated using the dropout technique. This approach has proven to be effective in generating results that outperform those obtained from each individual model.

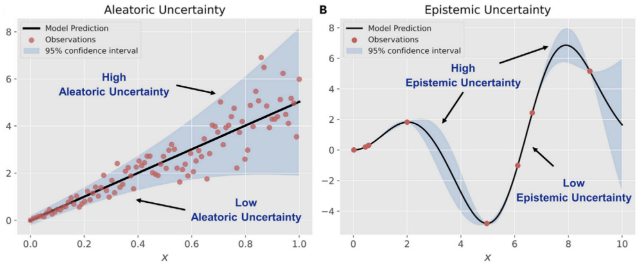
2.2.2 Model Integration with Keras

- **Keras Library:** The project used the Keras library to implement and train deep neural networks with the Ensemble Learning and Drop Out techniques. Keras simplified the implementation process and supported various neural network architectures. The integration with TensorFlow also facilitated the development and training of complex models, leading to improved accuracy and robustness.

3. Sources of Uncertainty

Uncertainty can come into the machine learning pipeline at different stages- and in different forms. In this project, we explored two types of uncertainties and their relationship with fairness of a machine learning model's prediction.

A comparison of the two is depicted in this diagram.



3.1 Aleatoric Uncertainty

This is the uncertainty caused by variance in the observed data. When a machine learning model is trained on data that is noisy, it leads to a higher error in the output. This is called Aleatoric uncertainty. For our experimentation, we intentionally introduced variance in the dataset by adding random noise in a certain subset of the training data- ranging from 10% to 50%. Next, we used the modified data to train the model and observed its predictions.

3.2 Epistemic Uncertainty

This uncertainty is caused by a lack of knowledge or lack of enough data. When the machine learning model does not get enough data to train on, it tends to overfit on the training data, and shows high variance in its prediction for other parts of the data space. To study epistemic uncertainty, we trained the model on controlled amounts of training data and compared its prediction's uncertainty levels.

4. Methodology

4.1 Adult Dataset

We utilized the adult dataset provided by the aif360 library to train Neural Network models, allowing us to evaluate aleatoric and epistemic uncertainties, as well as EOD scores. This dataset is widely used for evaluating and benchmarking fairness metrics in machine learning models. It includes demographic information and features such as age, education level, work class, and income of individuals, with a binary label indicating whether their income is above or below $50,000 per year.

4.2 Epistemic Uncertainty Quantification

To measure the epistemic uncertainty in the dataset, we resampled it into datasets of varying sizes, ranging from 10% to 100% of the original dataset. By doing this, we were able to observe how the nature of our model's predictions changed with different levels of data knowledge. We then trained the dropout model, deep ensemble learning model, and deep ensemble learning with dropout model using each of the resampled datasets to quantify the epistemic uncertainty. The predictions generated by these models were then used to estimate the fairness of the dataset using the Equalized Opportunity Difference (EOD) metric.

4.2 Aleatoric Uncertainty Quantification

To measure the aleatoric uncertainty in the dataset, we employed a comparable methodology. Rather than modifying the dataset by removing or adding data, we replaced a fraction of it with random noise ranging from 10% to 50%. For each substitution, the models were trained, and their uncertainties were measured. The predictions obtained from the noisy dataset were then used to determine the fairness values, similar to the truncated dataset approach.

5. Results

5.1 Epistemic Uncertainty vs EOD

In this project report, we present the findings for the Ensemble and Dropout models' combination only. The individual results are documented in the project notebook. As shown in Figure 2, our results support the hypothesis that the Epistemic uncertainty decreases as the dataset knowledge or information increases. However, we observed no correlation between the Epistemic uncertainty and the predicted fairness (Fig 4). This could be attributed to the random sampling of datasets that led to an inconsistent alteration of the dataset's inherent nature, having minimal impact on the model's predictability against a particular protected attribute.

Therefore, the relationship between mean epistemic uncertainty and mean equal opportunity shows no correlation as well, as observed in Fig 4.

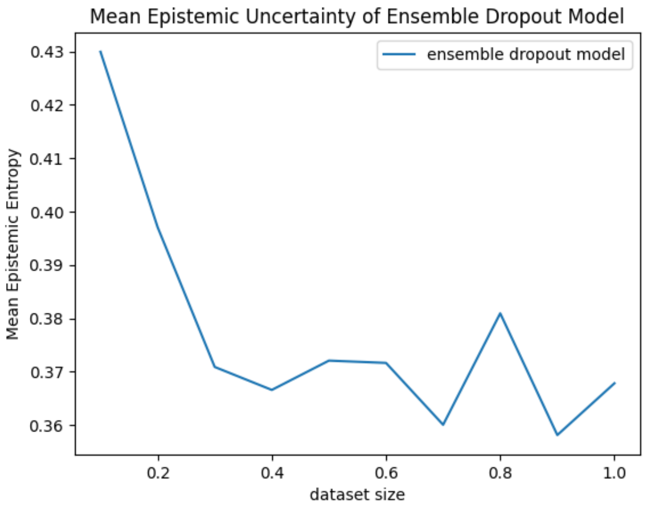


Fig 2. Mean Epistemic Uncertainty vs Dataset size

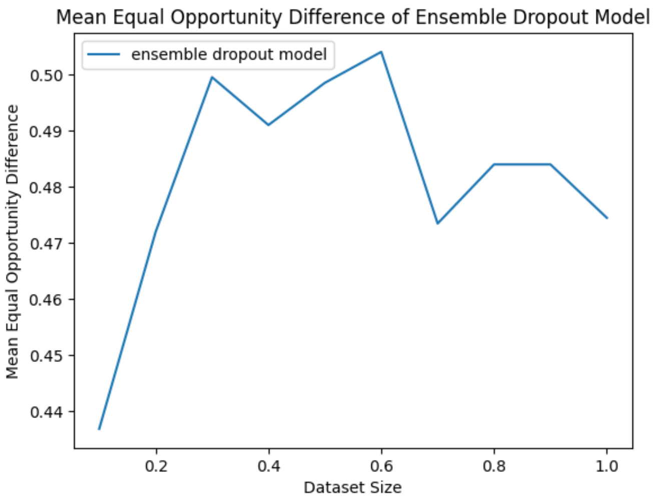


Fig 3. Mean Equal Opportunity Difference vs Dataset Size

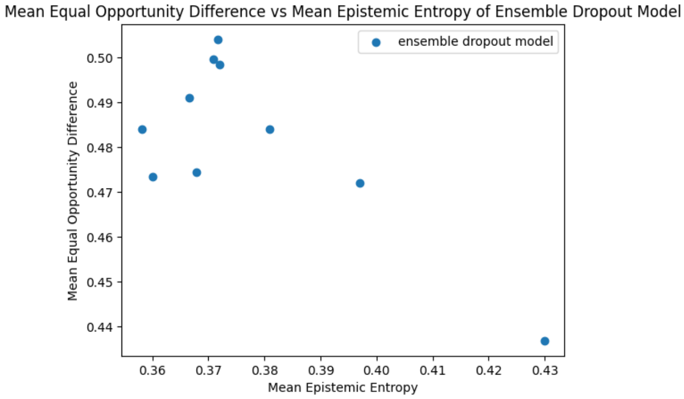


Fig 4. Mean Equal Opportunity Difference vs Mean Epistemic Uncertainty

5.2 Aleatoric Uncertainty vs EOD

Our project's results demonstrated the impact of varying levels of noise on Aleatoric uncertainty and the resulting fairness metric. Our hypothesis regarding aleatoric uncertainty was confirmed as the mean aleatoric uncertainty consistently increased with the addition of noisy samples in the dataset, indicating that noise disrupts the actual data pattern and leads to more uncertain results (Fig 5).

In contrast, an interesting finding from our study was that the fairness of the dataset showed a negative correlation with the level of noise added. This trend can be explained by the fact that the introduction of random noise reduces the inherent biases that the dataset may have towards a particular group of our protected attribute (Fig 6).

The plot of the mean equal opportunity difference against the mean aleatoric uncertainty revealed a negative correlation between increasing levels of aleatoric uncertainty and the equal opportunity difference of the dataset. Specifically, as the level of aleatoric uncertainty increased, the equal opportunity difference tended to decrease (Fig 7).

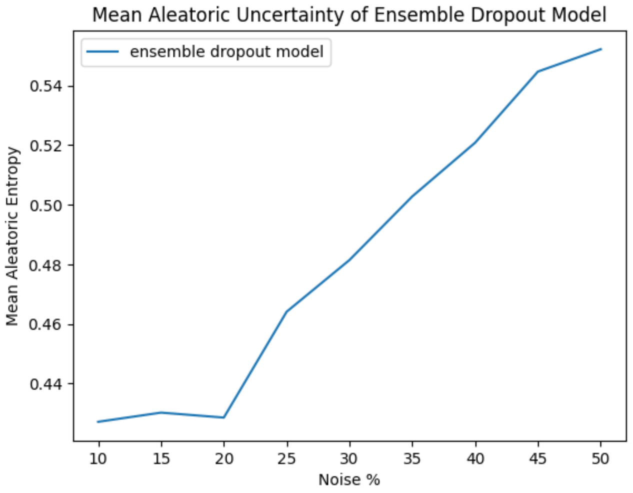


Fig 5. Mean Epistemic Uncertainty vs Dataset size

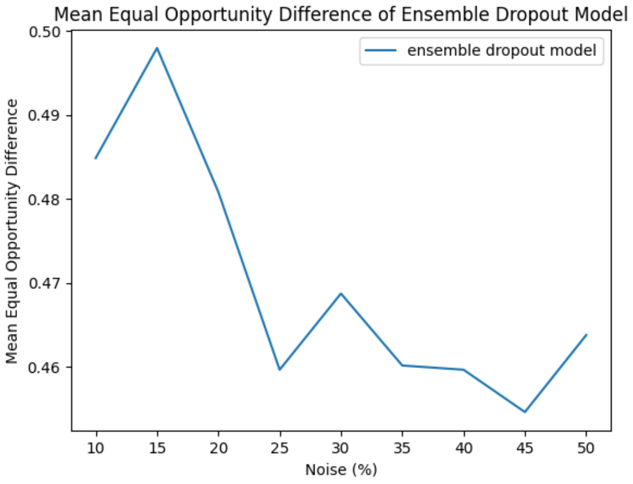


Fig 6. Mean Equal Opp. Difference vs Noise %

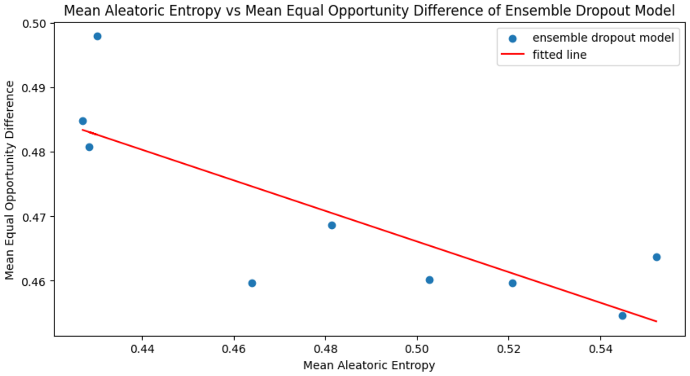


Fig 7. Mean Equal Opp. Difference vs Mean Aleatoric Uncertainty

6. Future Work

This research area offers immense opportunities for exploration and further study. Based on the insights we have gained from our experiments, there is potential to develop models that address the trade-off between fairness and uncertainty, and minimize both simultaneously.

To gain a deeper understanding of the nuances of the dataset, test runs on synthetic data can be carried out. This will allow for a more controlled and systematic study of the effects of different levels of uncertainties and their impact on fairness.

In addition, further exploration can be carried out on image and text data, as these types of data have their own unique characteristics and challenges. Test runs on these types of data can provide a strong basis for better interpretation of uncertainty and its relationship with fairness.

Overall, there is a lot of room for further investigation in this field, and we hope that our work will inspire and guide future research in the pursuit of fair and accurate machine learning models.

7. Conclusion

In conclusion, our research focused on understanding the relationship between uncertainty and fairness in machine learning models. We utilized the aif360 adult dataset and trained Neural Network models to quantify aleatoric and epistemic uncertainties, as well as measure the fairness of the dataset using the Equalized Opportunity Difference (EOD) metric. Our results showed that epistemic uncertainty decreases with an increase in dataset knowledge, but there was no clear relationship between epistemic uncertainty and predicted fairness. On the other hand, our findings indicated that adding random noise to the dataset increases aleatoric uncertainty and leads to a decrease in inherent biases, resulting in a more fair dataset.

Future work in this area could explore the development of models that minimize both uncertainty and bias to achieve a fairer and more reliable machine learning model. Additionally, conducting test runs on synthetic data and other types of data, such as image and text data, can provide a better understanding of the intricacies of uncertainty in various datasets. Overall, this research provides valuable insights into the complex relationship between uncertainty and fairness in machine learning, and opens up opportunities for further exploration and development.

8. References

1. https://github.com/zohairhashmi/CS483-uncertainty-quantification
2. https://github.com/mvaldenegro/keras-uncertainty/tree/master/keras_uncertainty
3. https://github.com/Trusted-AI/AIF360
4. M.Valdenegro-Toro and D.S.Mori (2022) *A Deeper Look into Aleatoric and Epistemic Uncertainty Disentanglement*
5. Abhin Shah, Maohao Shen, Jongha Jon Ryu, Subhro Das, Prasanna Sattigeri, Yuheng Bu, and Gregory W. Wornell (2023) *Group Fairness with Uncertainty in Sensitive Attributes*
6. Zhen Guo, Zelin Wan, Qisheng Zhang, Xujiang Zhao, Feng Chen, Jin-Hee Cho, Qi Zhang, Lance M. Kaplan, Dong H. Jeong and Audun Jøsanga. (2022) *Survey on Uncertainty Reasoning and Quantification for Decision Making: Belief Theory Meets Deep Learning*
7. Siddartha Devic, David Kempe, Vatsal Sharan and Aleksandra Korolova. (2023) *Fairness in Matching under Uncertainty*
8. Yijun Xiao and William Yang Wang. (2019) *Quantifying Uncertainties in Natural Language Processing Tasks.*
9. Wenchong He and Zhe Jiang (2023) *A Survey on Uncertainty Quantification Methods for Deep Neural Networks: An Uncertainty Source's Perspective.*