2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Identification of Bias in Pediatric Surgical Fellowship Letters of Recommendation Using Natural Language Processing

Sacha A. Williams, Geoffrey M. Gray, Bryce Bludevich, Iris Irby, Henry Chang, Paul D. Danielson, Raquel Gonzalez, Christopher W. Snyder, Luis M. Ahumada, Nicole M. Chandler

***Abstract*— Bias in letters of recommendation can influence applications to surgical fellowships. Machine learning can be used to quantify differences in letters of recommendation by race and gender. We find differences in polarity and intensity of emotions, with differences in anger being statistically significant.**

***Clinical Relevance*— This work identifies bias present in letters for recommendation for pediatric surgical fellowships using machine learning.**

I. INTRODUCTION

Application to pediatric surgical fellowships is a highly competitive process. One component of these applications is the addition of letters of recommendation (LORs) for the applicant. Bias in LORs can be subtle and difficult to detect by the reader, but it can have an outsized influence on the decision-making process. To better understand the prevalence and type of bias, we use machine learning to assess bias in LORs for pediatric surgical fellowship applications from 2016-2021 at a mid-sized quaternary care academic hospital. Findings indicate disparities in the both the polarity (positive vs. negative statements), as well as the types and intensity of the emotions expressed.

II. METHODS

The corpus was produced using the natural language toolkit. The Valence Aware Dictionary for sEntiment Reasoning (VADER) model was used to calculate polarity scores.[1] While VADER was originally designed for tweets, it outperforms traditional models. To keep performance high, we calculated polarity by sentence and averaged the values. The National Research Council dataset was used for emotion and intensity analysis.[2] Demographics were taken from surgical fellowship application. The Kruskal-Wallis H-test was used to test for statistical significance.[3]

III. RESULTS

Polarity scores are shown in Fig. 1. These scores range in value from -1 (most negative statement) to 1 (most positive statement), with a value of 0 indicating a neutral statement. Black applicants have the highest mean polarity (most positive average), while Hispanic/LatinX applicants have the lowest. Distributions for Black and Hispanic men as well as Asian women, have a right-ward skew. White and Asian men, as well as Other women, has a left-ward skew. White and Hispanic women had a nearly uniform distribution. Differences between the polarity distributions were not found to be statistically significant (data not shown).

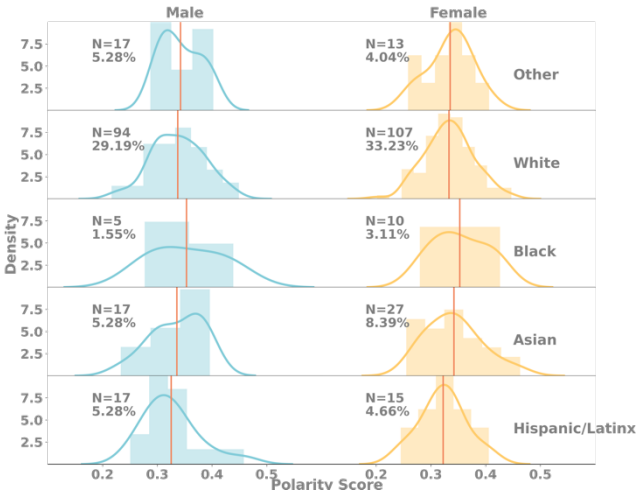

Figure 1. Polarity score distributions by race and gender.

The types of emotions and the intensity of these emotions were also calculated (data not shown). Differences in the intensity of anger within the letters was found to be statistically significant.

IV. DISCUSSION & CONCLUSION

This work identifies racial and gender bias present in LORs submitted as part of a pediatric surgical fellowship. Differences in polarity are identified as well as emotional intensity. While the types of emotions used are highly similar between race and gender groups, the intensity of these emotions reveals differences with anger being statistically significant. From this work, it can be concluded that bias in LORs if reflected as differences in polarity, which are likely a result of the intensity of the emotions being used and not the types of emotions being expressed.

REFERENCES

1. Hutto CJ, Gilbert E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Conference: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.* 2015.
2. Mohammad SM, Turney PD. CROWDSOURCING A WORD-EMOTION ASSOCIATION LEXICON. *Computational Intelligence.* 2013;29(3):436-465.
3. Kruskal WH, Wallis WA. USE OF RANKS IN ONE-CRITERION VARIANCE ANALYSIS. *Journal of the American Statistical Association.* 1952;47(260):583-621.

*Research supported by Johns Hopkins All Children's Hospital.

Luis M. Ahumada is with the *Center for Pediatric Data Science and Analytics Methodology, Johns Hopkins All Children's Hospital, St. Petersburg, FL* 33701 USA, phone: 727-767-2349; e-mail: ahumada@jhmi.edu.