MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231113

Uncertainty in Breast Cancer Risk Prediction: A Conformal Prediction Study of Race Stratification

Alexander S. Millar [a], John Arnn [a], Sam Himes [a], and Julio C. Facelli [a, 1]
[a] *Department of Biomedical Informatics and Clinical and Translational Science Institute, The University of Utah, Salt Lake City, UT 84108, USA*
ORCID ID: Julio Facelli https://orcid.org/0000-0003-1449-477X

Abstract. The use of Artificial Intelligence (AI) in medicine has attracted a great deal of attention in the medical literature, but less is known about how to assess the uncertainty of individual predictions in clinical applications. This paper demonstrates the use of Conformal Prediction (CP) to provide insight on racial stratification of uncertainty quantification for breast cancer risk prediction. The results presented here show that CP methods provide important information about the diminished quality of predictions for individuals of minority racial backgrounds.

Keywords. Artificial Intelligence in Medicine, Conformal Prediction, Uncertainty Quantification, Breast Cancer Risk

1. Introduction

Uncertainty plays a crucial role in the interpretation of individual risk predictions, as may be the case for a medical practitioner or a patient assessing the high/low breast cancer risk prediction of an individual patient based on a machine learning (ML) model [6]. The more "unusual" an individual's set of risk factors are compared to those that were used in the parameterization of the ML model, the more likely that the prediction will have a larger uncertainty. The level of uncertainty of predictions can be quantified by nonconformity measurements [1]. Nonconformity measures provide a means to quantify how unusual an individual is compared to the observed population. Conformal Prediction (CP) is an emerging method to compute non-conformity and its applications to clinical medical problems have been reviewed in a recent paper [7]. This paper utilizes a synthetic breast cancer risk prediction (Gail model) dataset from the literature [6] to demonstrate the use of CP methods to evaluate the differential uncertainty for different patient classes, modeling race-differential risk.

2. Methods

This work utilizes synthetic data derived from the work of Ming et al. [6]. The dataset was developed toward comparing Machine Learning (ML) techniques to the traditional

[1] Corresponding Author: Julio Facelli, julio.facelli@utah.edu.

Gail models for breast cancer risk prediction [3; 4]. Ming et al. [6] demonstrated that the synthetic data was sufficiently representative for our purposes, especially considering the ability it provides to represent the underlying relationships among the data elements. The ability to modify sample size allowed exploring the impact of under-representation on traditionally underserved populations. The ability to modify the underlying risk function allowed exploring scenarios where there is population-specific risk that may contribute to uncertainty.

The National Cancer Institute's Breast Cancer Risk Assessment Tool (BCRAT), based on the Gail model, [4] considers the following elements when predicting a patient's risk of developing invasive breast cancer: age, age to predict through, age at the start of menstruation, age at first live birth of a child, number of first-degree relatives with breast cancer, number of previous breast biopsies, presence of atypical hyperplasia in a biopsy, and race/ethnic group. This work follows Ming et al. [6] setting the age to predict through value constant at 90, which is considered lifetime risk and leaving the underlying distribution of risk factors and the race distribution unchanged. To simulate the race-differential the risk was scaled from the original breast cancer risk—set at 50% in the original dataset—by the arbitrary factors reported in Table 1. Note that these race-differential risk factors are arbitrary and used here only as exemplars, because the underlying risk factors by race are still not fully established [8] and the goal of this paper is to demonstrate a methodology that can be used under any differential risk distribution. The original sample size investigated by Ming et al. [6] was 1,200 individuals. This work explores a sample size of 12,000 individuals, which was generated using the code provided in Ref. [6]. This allowed us to explore the impact of sample size on uncertainty in predictions corresponding to less represented racial groups.

The Orange Conformal Prediction module [2; 5] was used for all the work reported here. For each dataset, four different underlying classification models were considered: logistic regression (LR), random forest (RF), k-nearest neighbors (KNN), and AdaBoost. Conformity was calculated using the inverse probability measure and a race conditional inductive prediction model [1]. This work also investigated how the confidence and credibility ranking of predictions varied over the prediction distribution and the relationship between the Lower Decile Range (LDR) of confidence/credibility (low confidence predictions) and race groups.

Table 1: Racial composition and scaling factor used to consider race differential risk of breast cancer.

Race	Percentage of the Population	Risk Scaling Factor
0	0.50	-0.0025
1	0.20	0.00375
2	0.20	-0.00125
3	0.08	0.00125
4	0.015	0.00875
5	0.003	0.0175
6	0.002	0.005

3. Results

Figure 1 presents the distribution of predictions ranked by confidence for the original sample size of N=1200 and an increased sample size of N=12,000, respectively. The Lower Decile Range (LDR) is highlighted, showing predictions in the LDR are consistently low confidence. It is apparent that for both data sets the number of predictions in the low-credibility region is approximately 25% for both samples.

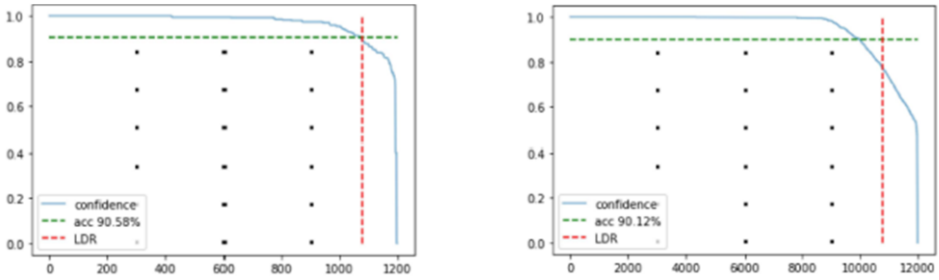

Figure 1. Prediction confidence distribution for the two data sets used in this study.

Figures 2 presents the corresponding prevalence of individuals belonging to each race/class group in the data sets considered here and the prevalence of individuals belonging to each population among predictions in the LDR of confidence values.

4. Discussion

Figure 1 shows that the confidence of predictions tends to fall off sharply at the end of the distribution, but this figure does not provide any information about the populations that are more affected by this decline of prediction quality. This decline is independent of the population size and corresponds to approximately 25% of the population. This work explored whether there appeared to be any relationship between the less represented race groups and predictions in the Lower Decile Range (LDR) of confidence distribution. Figure 2 shows that for the smaller sample (N=1,200) there is an overall increase in the proportion of the less represented races in the LDR region, as denoted by the larger orange bars when compared to the blue bars representing the race distribution in the overall population. In general, it is observed that the majority race is underrepresented in the LDR. From the right-side plots, it is apparent that these trends are less prominent for the larger sample with N=12,000, where disproportional representation between distribution in the overall population and the prevalence in the LDR is much less apparent. This can be rationalized by arguing that for N=1,200 the representation of the majority race is complete, but it is not for the less represented races. This suggest that to have a uniform distribution of uncertainties that is independent of race the sample size needs to be dictated by the number of individuals in the minority races and not by the overall population.

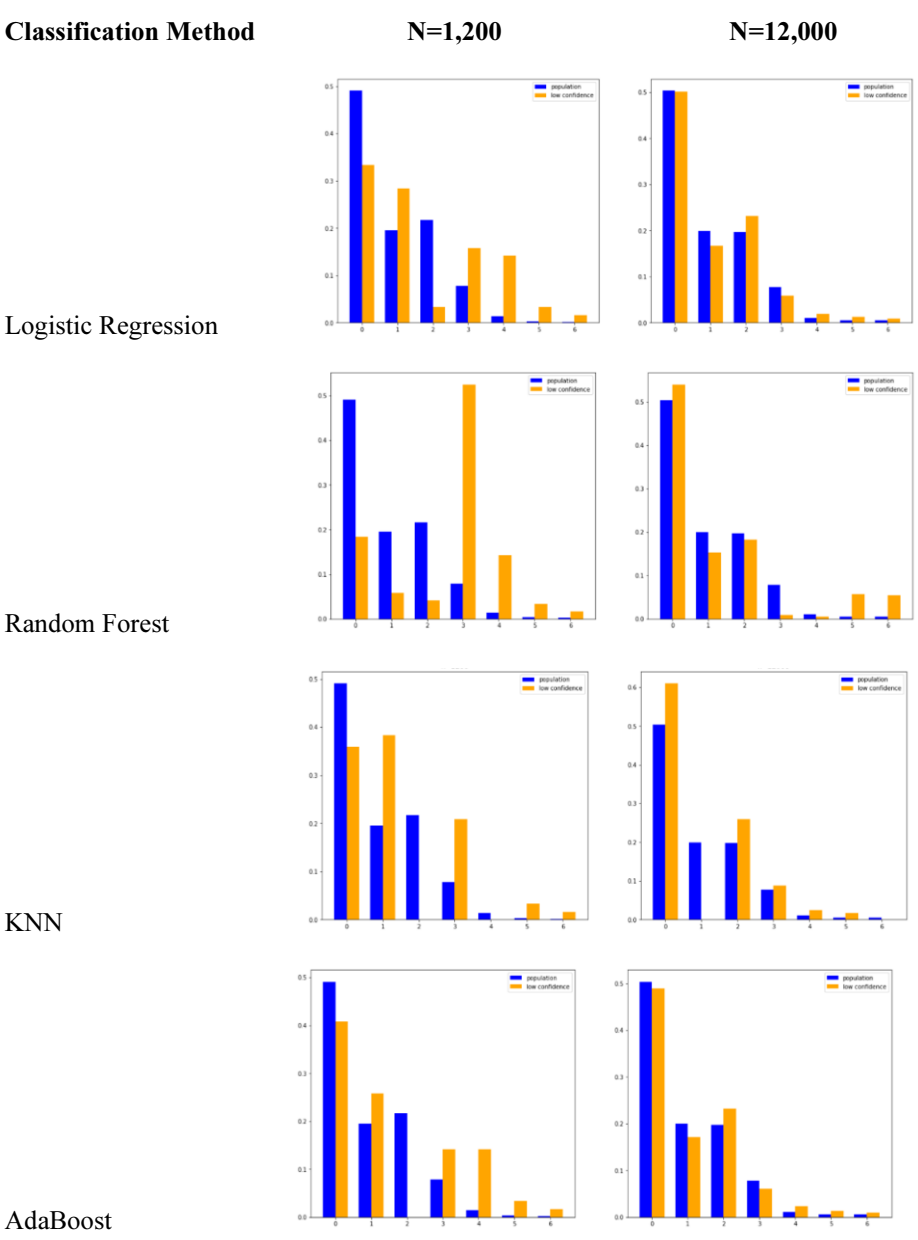

Figure 2. Comparison of the distribution of race groups in the population (blue) and the corresponding prevalence among those in the LDR of the confidence distribution (orange).

5. Conclusions

The method presented here show that CP can be used effectively to find the race stratification of individuals in the LDR of a predictive model. The results presented here cannot be used for clinical applications because the race-differential risks were arbitrarily assigned for the purpose of these demonstrations, but the methods demonstrated here can be used with any risk model derived from ML approaches. This work also shows that the undesirable stratification associating larger uncertainties to minority races can be corrected when using larger overall samples. Future work will include investigations to better understand the sensitivity of under/over representation of races in the LDR; classes upon their differential risks; and to find out if increased target, instead of overall, sample size could remediate stratification with a smaller enlargement of the sample size.

The synthetic datasets, data generation protocol, experiment results, and related code have been made freely accessible and can be found in the archived repository available on GitHub: (https://github.com/illato/vigilant-computing-machine/).

Acknowledgments
This work has been partially supported by National Center for Advancing Translational Sciences (UL1TR002538). Computer resources were provided by the University of Utah Center for High-Performance Computing, which has been partially funded by the NIH Shared Instrumentation Grant 1S10OD02164401A1.

References

[1] in: *Conformal Prediction for Reliable Machine Learning,* V.N. Balasubramanian, S.-S. Ho, and V. Vovk, eds., Morgan Kaufmann, Boston, 2014, p. v.
[2] J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik, and B. Zupan, Orange: data mining toolbox in python, *J. Mach. Learn. Res.* **14** (2013), 2349–2353.
[3] M.H. Gail, L.A. Brinton, D.P. Byar, D.K. Corle, S.B. Green, C. Schairer, and J.J. Mulvihill, Projecting individualized probabilities of developing breast cancer for white females who are being examined annually, *Journal of the National Cancer Institute* **81** (1989), 1879-1886.
[4] M.H. Gail and P.L. Mai, Comparing Breast Cancer Risk Assessment Models, *JNCI: Journal of the National Cancer Institute* **102** (2010), 665-668.
[5] T. Hočevar, B. Zupan, and J. Stålring, Conformal Prediction with Orange, *Journal of Statistical Software* **98** (2021), 1 - 22.
[6] C. Ming, V. Viassolo, N. Probst-Hensch, P.O. Chappuis, I.D. Dinov, and M.C. Katapodi, Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models, *Breast Cancer Research* **21** (2019), 75.
[7] J. Vazquez and J.C. Facelli, Conformal Prediction in Clinical Medical Sciences, *Journal of Healthcare Informatics Research* **6** (2022), 241-252.
[8] C.G. Yedjou, J.N. Sims, L. Miele, F. Noubissi, L. Lowe, D.D. Fonseca, R.A. Alo, M. Payton, and P.B. Tchounwou, Health and Racial Disparity in Breast Cancer, *Adv Exp Med Biol* **1152** (2019), 31-49.