Machine Learning Approaches for Pressure Injury Prediction

Muhammad Aurangzeb Ahmad
KenSci Inc
Department of Computer Science
University of Washington Bothell
Seattle, WA, USA
maahmad@uw.edu

Barrett Larson
Smith & Nephew
Stanford School of Medicine
Palo Alto, CA, USA
barrett.larson@smith-nephew.com

Steve Overman
KenSci inc
Seattle, WA, USA
steve@kensci.com

Vikas Kumar
KenSci Inc
Seattle, WA, USA
vikas@kensci.com

Jing Xie
Smith & Nephew
Memphis, TN, USA
jing.xie2@smith-nephew.com

Alan Rossington
Smith & Nephew
England, UK
alan@smith-nephew.com

Ankur Patel
Smith & Nephew
Dallas, TX, USA
ankur@smith-nephew.com

Ankur Teredesai
KenSci Inc
Department of Computer Science
University of Washington Tacoma
Seattle, WA, USA
ankurt@uw.edu

***Abstract*—Pressure Injuries are localized damages to the skin caused by sustained pressure. It is a common yet preventable disease affecting millions of patients. While there are multiple scales to determine if a patient has pressure injury, these methods suffer from high inter-rater subjectivity. To address this problem we create predictive models for pressure injury using Centers for Medicare & Medicaid Services claims data. The models show relatively good predictive performance, we also explore aspects of the model where they will be deployed in a real world clinical settings.**

***Index Terms*—Pressure Injury, HAPI, Braden Scale, CAPI**

I. Introduction

Pressure injuries (also called pressure ulcers or bedsores) are areas of localized skin and soft tissue damage caused by sustained pressure. Pressure injuries typically start as an area of redness over bony prominences, but they can quickly develop into full thickness wounds if left untreated. Just within the US, more than 2.5 million people develop pressure injuries annually, making them one of the most common complications occurring in hospitals. This problem impacts approximately one-in-thirty hospitalized patients, is often associated with higher risks of mortality and results in decreased quality of life [1]. Complications from pressure injuries can trigger other ailments like autonomic dysreflexia, bladder distension, pyarthroses, sepsis, amyloidosis, urethral fistula, and gangrene [2]. In some extreme cases complications from pressure injury can even be life threatening. However, despite being common, pressure injuries are considered largely preventable. Even a small reduction in pressure injury rates would result in significant improvement in population health.

To prevent pressure injuries, the National Pressure Injury Advisory Panel (NPIAP), the European Pressure Ulcer Advisory Panel (EPUAP) and the Pan Pacific Pressure Injury Alliance (PPPIA) regularly publish clinical practice guidelines for pressure injury prevention. These evidence-based guidelines recommend implementing a series of prevention measures based on a patient's unique clinical circumstance and risk factors. The definition of what constitutes pressure injury is standardized by the aforementioned organizations.

While reliable risk assessment is considered the cornerstone to a robust prevention strategy, traditional methods for identifying patients at risk for developing a pressure injury are limited. For example, the Braden Scale [3], which is one of the most widely used risk assessment tools, has been shown to have insufficient predictive validity and poor accuracy [4]. Traditional risk assessment methods are also subjective in nature and suffer from a high degree of inter-rater variability, with dependence on the skill of the examiner. Traditional methods are also typically based on a bedside assessment, which is conducted at a moment in time without necessarily considering a patient's detailed medical history or other risk factors that may not be directly observable. Additionally, traditional risk assessment scales are not designed to adapt to different patient populations, geographies, or care settings. Therefore, developing a comprehensive and objective risk assessment method, one that can accommodate diverse patient populations and integrate detailed medical history and risk factors, is warranted.

Pressure injuries are categorized into two broad categories with respect to how a patient acquires the condition: Community Acquired Pressure Injury (CAPI) and Hospital Acquired Pressure Injury (HAPI). In this study, we have focused on developing improved risk prediction algorithms for HAPI, given the clinical and financial burden associated with pressure injuries developed in the hospital. According to the Agency for Healthcare Research and Quality (AHRQ), HAPI is the only hospital-acquired condition with increasing prevalence over the past 5 years, suggesting an urgent need for improved prevention measures [5]. In this paper we not only focus on the problem of predicting pressure injury using claims data but also describe how such models could be used for triaging, patient risk stratification and improving care.

II. Related Work

Researchers have historically taken two general approaches to addressing the problem of identifying pressure injuries. One approach has been to use images of pressure injuries to identify the condition or its severity [6]. The other approach is using Electronic Health Records (EHR) to predict pressure injury or its severity. A number of studies using machine learning to predict pressure injuries and/or stages of pressure injury have been done with varying levels of success [7] [8] [9] [10] [11]. To the best of our knowledge only one of these have been applied in a real world clinical setting. Jin et al deployed an automated pressure injury risk assessment system (Auto-PIRAS) that can assess pressure injury risk using EHR data, without requiring nurses to collect or input additional data [8]

A number of studies have also been done to assess the predictive validity of different scales for assessing pressure injury risk, especially the Braden Scale. The Braden Scale has moderate predictive validity and low predictive specificity for pressure ulcers in long-term care residents [4] [12]. Several meta-analysis of risk assessment scales, especially the Braden Scale, reveal that a new and modified scale is needed [13] [12]. Machine learning analyses have also been done to determine which factors may be helpful in predicting pressure injury. A comprehensive study on pressure injury scales revealed that the use of a particular scale is less important as long as some validated scale is used in conjunction with additional risk factors not captured by the risk assessment system [14]. Multiple studies of factor analysis of the most predictive features for pressure injury reveal similar results (e.g., BMI, age, length of stay, mobility, friction/shear, norepinephrine infusion, peripheral vascular disease, pneumonia, cardiovascular disease, etc.) [15] [16] [7] [17] [18] [19].

III. Data

We used a subset of medical claims data from CMS (Centers for Medicare & Medicaid Services) to serve as a representative sample of the US population. This US federal database contains detailed information for more than 40 million members per year, comprising 400 million patient records in total (over one TB of data). This CMS data contains data from in-patient, out-patient and nursing home settings. The data subset that was employed consisted of 11,264,189 claims corresponding to 6,622,678 patients. The dataset encodes tens of thousands of diagnoses for the whole population via Diagnosis-related groups (DRG), which is a system used to classify hospital cases. DRGs enable a systematic way to encode and track diseases over time and even do population level analysis. The DRGs include diagnosis code for pressure injuries. After filtering for records with only PI codes we get 656,904 records, corresponding to 438,883 patients. A summary of differences and similarities between the characteristics of the HAPI cohort and the non-HAPI cohort is given in Table I. The most glaring difference between the two is the prevalence of previous pressure injury for the HAPI cohort and the preponderance of certain conditions in HAPI i.e., for CHF (Congestive Heart Failure) and two types of diabetes. This also implies that the HAPI cohort in general is sicker than the non-HAPI cohort.

Feature	HAPI	non-HAPI
Previous PI	18	2
CHF	41	30
DM No Complications	41	32
DM Complications	25	17
Hypertension	62	59
Recreational Drugs	1	2
Hypothroidism	19	18
Mechanical Ventilation 72 hours	<1	<1
Cardiography	49	49
Social Detriments of Health	2	2
Sex=Famale	46	47
Race=White	78	81

TABLE I
Relative Percentage of Some Features

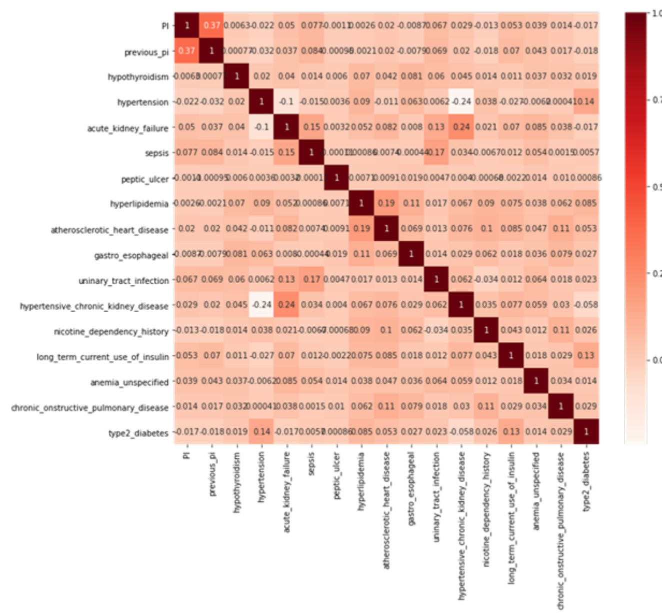

Fig. 1. Correlation between Pressure Injury and major comorbidities

Quite often, pressure injuries are accompanied by other comorbidities which may make a patient more vulnerable to PI. Figure 1 shows correlations between pressure injury and comorbidities that a patient may have. In general, we observe that the correlations between different conditions are as expected (e.g., relatively high correlation between hypertensive chronic kidney disease and acute kidney failure). One thing that stands out is that there is a high level of correlation (correlation = 0.37) between a patient developing a pressure injury after being admitted to a hospital and the patient having a prior history of diagnosis of pressure injury. This also makes sense from a domain perspective since a patient who has had pressure injury in the past is more vulnerable to get pressure injury again [1].

IV. Experiments

We pose the problem of predicting HAPI as a binary classification problem where the two classes are patients with HAPI vs. non-HAPI patients. The two classes are defined as follows:

An encounter is considered to belong to the positive class if the following conditions are true:

- Presence of PI DRG codes as specified by CMS in the primary or secondary diagnosis
- 12 months of continuous enrollment in Medicare
- The PI diagnosis not in POA (present on arrival)
- The length of stay for the encounter is more than 1 day

Similarly, an encountered is considered to belong to the negative class if the following conditions are true:

- 12 months of continuous enrollment in Medicare
- The length of stay for the encounter is more than 1 day
- The PI diagnosis not in POA (present on arrival)
- No current diagnosis of PI

Given these definitions, it should be noted that we are predicting HAPI post-admission (i.e., predicting if a patient will develop pressure injury after they have been admitted to the hospital). While there is a relatively small number of encounters that satisfy the condition of being in the positive class i.e., developing PI post-admission, there are tens of millions of encounters where the opposite is true. To ensure that we are able to compare the two classes, we randomly sampled encounters from the negative class. Multiple random samples were employed and machine learning models were built on these samples to ensure the validity of the results. The sample that was employed had equal representation of the two classes.

For baseline results we compared our model with two baselines: One baseline was based on random predictions with respect to label distribution and the other baseline was based on a logistic regression model. The problem was set up as a standard binary class classification problem with 10-fold cross validation. After employing the filtering criteria described above, the final set that was used for the experiments consisted of 44,136 instances of the positive class, and An approximately equal number of negative instances are randomly sampled.

Feature selection for the models was done based on a combination of domain-driven and data-driven techniques. For the domain-driven method, we had three different domain experts go over a set of more than 500 different features to determine the set of features that made the most sense from a clinical perspective. In the data-driven method, we employed various feature selection methods, like univariate feature selection and recursive feature selection, to narrow down the set of features. The final list of features that was used for model building consisted of 188 features. A summary of various feature categories and their corresponding examples is given in Table II.

Category	Number of Features	Example
Diagnosis	68	CHF
Behavioral	8	Recreational Drugs
Social	10	Homelessness
Procedure	42	Cardiography
Labs	42	Troponin
Medication	4	Vesopresser
Utilization	8	Previous Length of Stay
Demographics	6	Age

TABLE II
DISTRIBUTION OF FEATURE CATEGORIES

Metric	Stratified Baseline	Logistic Regression	XGB
Accuracy	0.50	0.61	0.74
Precision	0.52	0.48	0.70
Recall	0.52	0.63	0.70
F-Score	0.52	0.54	0.70
AUC	0.50	0.66	0.74

TABLE III
SUMMARY OF PREDICTION RESULTS

V. Results

In addition to the two baselines described above, we also built the classification models using six different machine learning algorithms (Naive Bayes, Extreme Gradient Boosting, Decision Trees, Random Forest, AdaBoost with Decision Stump). Here we report the results for the model corresponding to the best predictive performance, which was the Extreme Gradient Boosting (XGB) algorithm. The results are given in Table III. The results show a remarked improvement of performance over the two baselines. Additionally, we note that the results are comparable to what has been reported previously in the literature, but with one major difference - most high performing models were reported on specialized populations e.g., ICU, geriatric populations etc. For the current set of experiments, we used the general population without filtering for any specialized criteria. It has been previously reported in the literature that PI prediction models perform better for specialized populations. We note that one reason for not using any specialized filtering criteria is that filtering for ICU populations is non-trivial and not always reliable in this dataset. This also implies that the model is likely to be more robust, as it is able to perform well across multiple sub-populations.

It is important to note that prediction of a particular outcome is not necessarily sufficient to justify usage of a model in a clinical setting. Model transparency may be required to foster trust from clinicians and other healthcare personnel [20]. To ensure that our models are transparent, we determined the top explanation factors for the prediction models using the SHAP framework [21]. The top factors are given in Figure 2, which shows not only the top factors, but also their relative importance. One factor that consistently dominates other factors is previous history of pressure injury. This makes sense from a clinical perspective (i.e., if a patient has previous pressure injury in the past, then they are indeed more susceptible to pressure injury). Additional factors that stand out are age and LOS (length of stay). These results are also consistent with what we know from the domain regarding pressure injury risk e.g., older patients are more at risk, male patients are also at greater risk, and patients with longer lengths of stays are more likely to be associated with greater risk for PI.

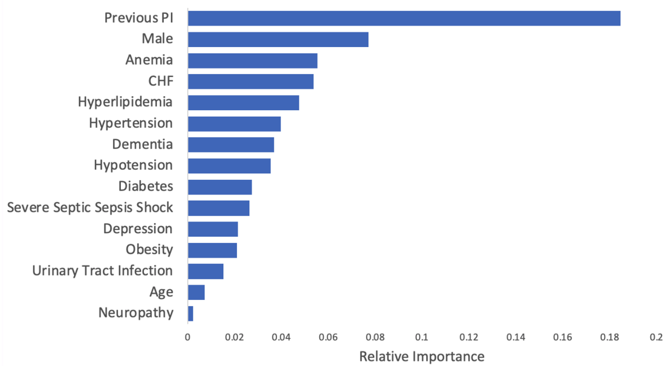

Fig. 2. Top contributing features for PI prediction as computed by SHAP

Given that pressure injuries can occur for a variety of reasons, we also explored model explanations for individual patients (i.e., why is a particular patient predicted to have a higher risk of pressure injury as compared to others). One such peculiar instance is given in Figure 3 for illustration purposes. Here, while the prior history of PI is the top factor driving risk prediction, the rest of the factors are somewhat different. A combination of multiple comorbidities unique to these patients are driving the risk prediction. It is also interesting to note that the The risk factors are for a female patient and thus gender does not appear as a top factor for this high risk patient. The example also illustrates how the risk prediction is localized to the individual patient, based on their unique combination of risk factors.

Given the number of features analyzed by this model, it is best suited for integration with an electronic health record system. However, the model can be adapted to accommodate manual entry of features. For this model to be useful in the context of manually inputting data, the number of features analyzed may have to be reduced. In a real-world use case, a clinician may need to input values for a patient in a dashboard to compute the relevant risk score for a patient. In order to ensure that this is feasible and to build well-informed features for data, we reduced the number of features based upon: (i) Common well-known features such as diagnosis and procedures) (ii) Clinical domain knowledge from PI (such as prior cases of PI), and the data itself (such as history of chronic conditions). This resulted in a model of just 34 features with some trade-off in performance, a summary of the previous best results and the minimal model is given in Table IV. The decline in performance is expected given that aa subset of the features is being used.

Even with the reduction in performance the output of such models can still be used to inform decision making in a clinical setting.

Metric	General Model	Minimal Model
Accuracy	0.74	0.61
Precision	0.70	0.66
Recall	0.70	0.45
F-Score	0.70	0.54
AUC	0.74	0.62

TABLE IV
SUMMARY OF PREDICTION RESULTS

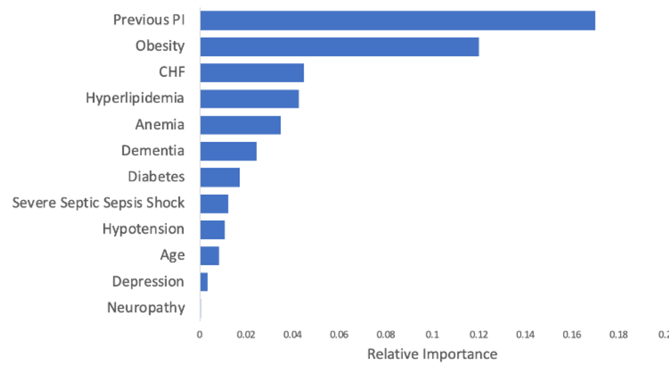

Fig. 3. Top contributing features for a particular patient

VI. Conclusion and Future Work

Reliably identifying patients at risk for pressure injury will significantly aid prevention efforts. With more reliable risk stratification, prevention efforts and resources can be more appropriately allocated. Given the constraints on the healthcare system, we need to improve our ability to successfully manage large populations of patients in a clinically and financially efficient manner. In this paper we addressed the problem of predicting if a patient would develop a PI as a binary classification problem. The results obtained were comparable to the state of the art, with the additional benefit of being more generalizable and amenable to a use-case that allows manual data entry. A machine learning approach to assessing pressure injury also carries the promise of adaptability of pressure injury scales to different patient populations, geographies, and care settings. We presented two models for prediction of pressure injury, a 'complete' model where input from 188 features was used and a minimal model consisting of only 34 features. Analysis of the models using the SHAP post-hoc explanation model revealed that the top factors responsible for prediction was in accordance with domain knowledge related to pressure injury.

Developing these prediction models were part of a technology demonstration project which is work in progress currently. Our plan is to test these models in a real world clinical setting. Our next step is to develop a dashboard that allows testing of the minimal model. This dashboard will allow clinicians to input relevant variables into a model to help risk stratify patients. It should be noted that while the minimal model consists of 34 features, 31 of these features are related to previous diagnosis. We can thus group these into a single category from which the end user can select the ones which are applicable to the patient. The risk assessment could then be used to determine how much a patient should be prioritized for preventive care in order to reduce the risk of pressure injury. A prototype of such a dashboard is given in Figure 4, here the healthcare personnel has the option to input data from a limited number of variables as described by the minimal model.

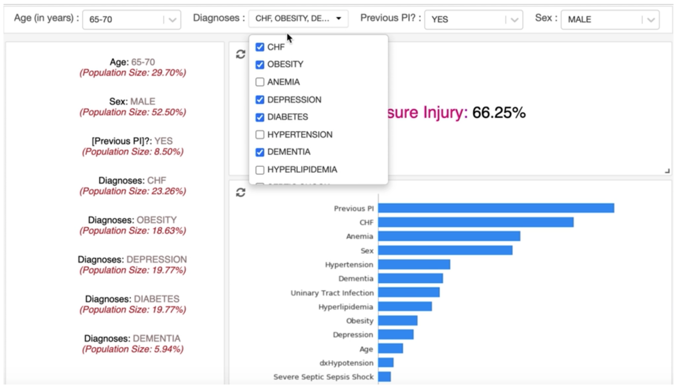

Fig. 4. A Prototype of the Clinical tool for predicting Pressure Injury

The output of the model will be a risk score corresponding to the prediction probability for that instance. Such a dashboard would also enable the end user to determine how differences in the presence or absence of certain conditions would impact the prediction of the outcome. As an example consider that assuming all factors are equal, the risk score for men is likely to be higher as compared to men. This is also in concordance with what we know about the prevalence of pressure injury across genders in the vast medical literature [2]. As part of future work, we are also working on improving the predictive model for the general model as well as for the minimal model. The long term plan is to integrate the dashboard in the clinical workflow so that it can be adopted readily by clinical staff. Lastly, once the dashboard has been tested among a group of users in a clinical setting via user acceptance tests in various hospital systems, we hope that it could then be deployed.

References

[1] S. Bergquist-Beringer, L. Dong, J. He, and N. Dunton, "Pressure ulcers and prevention among acute care hospitals in the united states," *The Joint Commission Journal on Quality and Patient Safety*, vol. 39, no. 9, pp. 404–414, 2013.

[2] L. E. Edsberg, D. Langemo, M. M. Baharestani, M. E. Posthauer, and M. Goldberg, "Unavoidable pressure injury: state of the science and consensus outcomes," *Journal of Wound Ostomy & Continence Nursing*, vol. 41, no. 4, pp. 313–334, 2014.

[3] N. Bergstrom, P. Demuth, and B. Braden, "A clinical trial of the braden scale for predicting pressure sore risk." *The Nursing Clinics of North America*, vol. 22, no. 2, pp. 417–428, 1987.

[4] S. Hyun, B. Vermillion, C. Newton, M. Fall, X. Li, P. Kaewprag, S. Moffatt-Bruce, and E. R. Lenz, "Predictive validity of the braden scale for patients in intensive care units," *American Journal of Critical Care*, vol. 22, no. 6, pp. 514–520, 2013.

[5] K. H. Sheetz, J. B. Dimick, M. J. Englesbe, and A. M. Ryan, "Hospital-acquired condition reduction program is not associated with additional patient safety improvement," *Health Affairs*, vol. 38, no. 11, pp. 1858–1865, 2019.

[6] S. Zahia, M. B. G. Zapirain, X. Sevillano, A. González, P. J. Kim, and A. Elmaghraby, "Pressure injury image analysis with machine learning techniques: A systematic review on previous and possible future methods," *Artificial intelligence in medicine*, vol. 102, p. 101742, 2020.

[7] X. Deng, T. Yu, and A. Hu, "Predicting the risk for hospital-acquired pressure ulcers in critical care patients," *Critical care nurse*, vol. 37, no. 4, pp. e1–e11, 2017.

[8] Y. Jin, T. Jin, and S.-M. Lee, "Automated pressure injury risk assessment system incorporated into an electronic health record system," *Nursing research*, vol. 66, no. 6, pp. 462–472, 2017.

[9] P. Kaewprag, C. Newton, B. Vermillion, S. Hyun, K. Huang, and R. Machiraju, "Predictive models for pressure ulcers from intensive care unit electronic health records using bayesian networks," *BMC medical informatics and decision making*, vol. 17, no. 2, pp. 81–91, 2017.

[10] J. Alderden, G. A. Pepper, A. Wilson, J. D. Whitney, S. Richardson, R. Butcher, Y. Jo, and M. R. Cummins, "Predicting pressure injury in critical care patients: a machine-learning model," *American Journal of Critical Care*, vol. 27, no. 6, pp. 461–468, 2018.

[11] J. Levy, J. F. Lima, M. W. Miller, G. L. Freed, and J. O'Malley, "Investigating the potential for machine learning prediction of patient outcomes: a retrospective study of hospital acquired pressure injuries," *medRxiv*, 2020.

[12] M. Wei, L. Wu, Y. Chen, Q. Fu, W. Chen, and D. Yang, "Predictive validity of the braden scale for pressure ulcer risk in critical care: A meta-analysis," *Nursing in critical care*, vol. 25, no. 3, pp. 165–170, 2020.

[13] H.-L. Chen, W.-Q. Shen, and P. Liu, "A meta-analysis to evaluate the predictive validity of the braden scale for pressure ulcer risk assessment in long-term care." *Ostomy/wound management*, vol. 62, no. 9, pp. 20–28, 2016.

[14] C. V. Bouten, C. W. Oomens, F. P. Baaijens, and D. L. Bader, "The etiology of pressure ulcers: skin deep or muscle bound?" *Archives of physical medicine and rehabilitation*, vol. 84, no. 4, pp. 616–619, 2003.

[15] J. Cox, "Predictors of pressure ulcers in adult critical care patients," *American journal of critical care*, vol. 20, no. 5, pp. 364–375, 2011.

[16] C.-T. Su, P.-C. Wang, Y.-C. Chen, and L.-F. Chen, "Data mining techniques for assisting the diagnosis of pressure ulcer development in surgical patients," *Journal of medical systems*, vol. 36, no. 4, pp. 2387–2399, 2012.

[17] F. Aloweni, S. Y. Ang, S. Fook-Chong, N. Agus, P. Yong, M. M. Goh, L. Tucker-Kellogg, and R. C. Soh, "A prediction tool for hospital-acquired pressure ulcers among surgical patients: Surgical pressure ulcer risk score," *International wound journal*, vol. 16, no. 1, pp. 164–175, 2019.

[18] J. Pittman, T. Beeson, J. Dillon, Z. Yang, and J. Cuddigan, "Hospital-acquired pressure injuries in critical and progressive care: avoidable versus unavoidable," *American journal of critical care*, vol. 28, no. 5, pp. 338–350, 2019.

[19] J. Cox, M. Schallom, and C. Jung, "Identifying risk factors for pressure injury in adult critical care patients," *American Journal of Critical Care*, vol. 29, no. 3, pp. 204–213, 2020.

[20] M. A. Ahmad, A. Patel, C. Eckert, V. Kumar, and A. Teredesai, "Fairness in machine learning for healthcare," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3529–3530.

[21] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.