Jurnal ELTIKOM:
Jurnal Teknik Elektro, Teknologi Informasi dan Komputer

Vol. 8, No. 1, June 2024, page. 26-33
ISSN 2598-3245 (Print), ISSN 2598-3288 (Online)
DOI: http://doi.org/10.31961/eltikom.v8i1.1081
Available online at http://eltikom.poliban.ac.id

DETECTION OF BIAS IN MACHINE LEARNING MODELS FOR PREDICTING DEATHS CAUSED BY COVID-19

Fatimatus Zachra, Setio Basuki*
Department of Informatics, Universitas Muhammadiyah Malang, Malang, Indonesia
e-mail: fatimatuszachraa@gmail.com, setio_basuki@umm.ac.id

Received: 27 January 2024 – Revised: 12 March 2024 – Accepted: 12 March 2024

ABSTRACT

The COVID-19 pandemic has significantly impacted global health, resulting in numerous fatalities and presenting substantial challenges to national healthcare systems due to a sharp increase in cases. Key to managing this crisis is the rapid and accurate identification of COVID-19 infections, a task that can be enhanced with Machine Learning (ML) techniques. However, ML applications can also generate biased and potentially unfair outcomes for certain demographic groups. This paper introduces a ML model designed for detecting both COVID-19 cases and biases associated with specific patient attributes. The model employs Decision Tree and XGBoost algorithms for case detection, while bias analysis is performed using the DALEX library, which focuses on protected attributes such as age, gender, race, and ethnicity. DALEX works by creating an "explainer" object that represents the model, enabling exploration of the model's functions without requiring in-depth knowledge of its workings. This approach helps pinpoint influential attributes and uncover potential biases within the model. Model performance is assessed through accuracy metrics, with the Decision Tree algorithm achieving the highest accuracy at 99% following Bayesian hyperparameter optimization. However, high accuracy does not ensure fairness, as biases related to protected attributes may still persist.

Keywords*: bias, COVID-19, DALEX, machine learning, protected attributes.*

I. INTRODUCTION

IN March 2020, the World Health Organization (WHO) declared the COVID-19 outbreak a global public health emergency after the virus affected over 90,870 cases and resulted in 3,112 confirmed deaths across 72 countries [1]. Initially identified in Wuhan, Hubei Province, People's Republic of China, the virus is believed to have originated from the Rhinolophus sinicus bat, commonly sold at food markets in China [2]. COVID-19 is highly transmissible from wild animals to humans and among humans themselves, particularly through direct contact with an infected individual. The virus manifests with symptoms such as fever, dry cough, shortness of breath, fatigue, among others. The rapid increase in cases has overwhelmed medical personnel worldwide due to the disproportionate ratio of healthcare workers to patients [3]. To mitigate future outbreaks, it is critical to adopt measures such as employing Machine Learning (ML) models to predict potential high-risk cases, thereby enabling medical personnel to prioritize care and prevent further fatalities. ML models are increasingly applied in various sectors, including healthcare, to enhance decision-making processes. These models work by converting data inputs into actionable decisions based on chosen algorithms [4]. The application of ML not only improves decision accuracy but also boosts productivity and reduces costs [5].

Given the widespread integration of ML in various sectors, it is crucial to evaluate the fairness and justice of decisions made by these systems. Although fairness typically implies equal treatment, in the context of ML, it specifically refers to producing outcomes that are free from discriminatory biases against specific groups or populations. This concern is driven by significant attention from researchers, particularly following a ProPublica.com report that demonstrated judicial systems displaying biases against black individuals [6]. Sensitive attributes such as gender, race, religion, and ethnicity often contribute to biases [7]. Most Artificial Intelligence (AI) systems, including ML models, rely on data for training. If the training data is biased, the algorithms will inevitably learn and perpetuate these biases in

their predictions, leading to skewed outcomes [8]. Bias during data collection can arise if the data does not accurately represent the population intended for the model's application [9]. These biased results, once integrated into real-world systems, can significantly affect long-term decision-making. Therefore, it is essential for ML models to emphasize algorithmic transparency, allowing the processes and outcomes of decisions to be understandable and scrutinizable by humans [10]. This transparency includes providing detailed information about the algorithms used in decision-making [11].

This study focuses on detecting bias in Machine Learning (ML) models that predict COVID-19 mortality rates. It utilizes the 'SyntheaCovid100k' dataset, available through the SyntheticMass website [12], which has been previously explored in a study [13]. For bias detection, this research employs the DALEX library (moDel Agnostic Language for Exploration and eXplanation). DALEX is chosen for its capability to thoroughly analyze existing biases in ML models, supported by robust visualization tools and precise analytic methods. The specific attributes examined in this study include protected attributes such as gender, age, race, and ethnicity.

II. Related Work

This section reviews existing literature to summarize current methodologies in constructing ML models and detecting biases within them, particularly in the context of predicting COVID-19 mortality. ML models have been extensively applied to predict the mortality of COVID-19 patients. One such study by Zoabi et al. [14] utilized data from 51,831 individuals, including 4,769 COVID-19 cases, and employed 8 binary features such as gender, age, infected contacts, and five clinical symptoms in their analysis. The study reported a ROC curve accuracy of 90%. Another study by Zakariaee et al. [15] assessed various ML algorithms using a dataset from Ayatollah Talleghani Hospital in Abadan, Iran, and evaluated model performance using metrics like accuracy, precision, sensitivity, specificity, and ROC (AUC), with Random Forest achieving the highest accuracy of 97.2%. Elshennawy et al. [16] explored deep learning models to predict COVID-19 mortality, using a dataset of 12,020 cases, and found that the IMG-CNN model outperformed others with an accuracy of 94.14%. Additionally, Nishant et al. [17] compared the effectiveness of the Random Forest, XGBoost, and Extra Trees Classifier models using data from 4,711 hospitalized patients, noting that the Random Forest model had the best performance with an accuracy of 85.64%.

Research on bias detection in ML has been actively pursued by several researchers. Estiri et al. [18] discuss a framework for the objective evaluation of medical AI, focusing on binary classification models. This study utilized data from over 56,000 patients at Mass General Brigham (MGB) and evaluated unrecognized biases in four AI models developed early in the pandemic in Boston. The evaluation metrics included discrimination, accuracy, and overall model performance. The findings revealed that the models displayed inconsistencies in their predictions, with most showing higher error rates for older patients. Afrose et al. [19] developed a bias correction technique called double prioritization (DP) to mitigate representation biases in ML-based prognosis, specifically training ML models for distinct ethnic or age groups. This method successfully produced more representative prognostication models for underrepresented racial and age groups. Meanwhile, Allen et al. [20] assessed ML algorithms to reduce bias in predicting in-hospital mortality across white and non-white patient groups, using data from the Medical Information Mart for Intensive Care-III database. They compared the bias and accuracy of these ML models against traditional scores such as the Modified Early Warning Score (MEWS), Simplified Acute Physiology Score II (SAPS II), and Acute Physiologic Assessment and Chronic Health Evaluation (APACHE). The study found biases in the SAPS II and MEWS scores, indicating a need for improved fairness in predictive models.

Wiśniewski et al. [7] conducted a study using the DALEX library to identify bias in binary classification outcomes, specifically utilizing the German Credit Data dataset. The study found that biases adversely affected female subjects. This research serves as a benchmark for comparative studies. However, it did not extensively explore the different types of bias that can occur. Consequently, this study aims to build upon previous work by developing a comprehensive approach to investigating biases related to protected attributes.

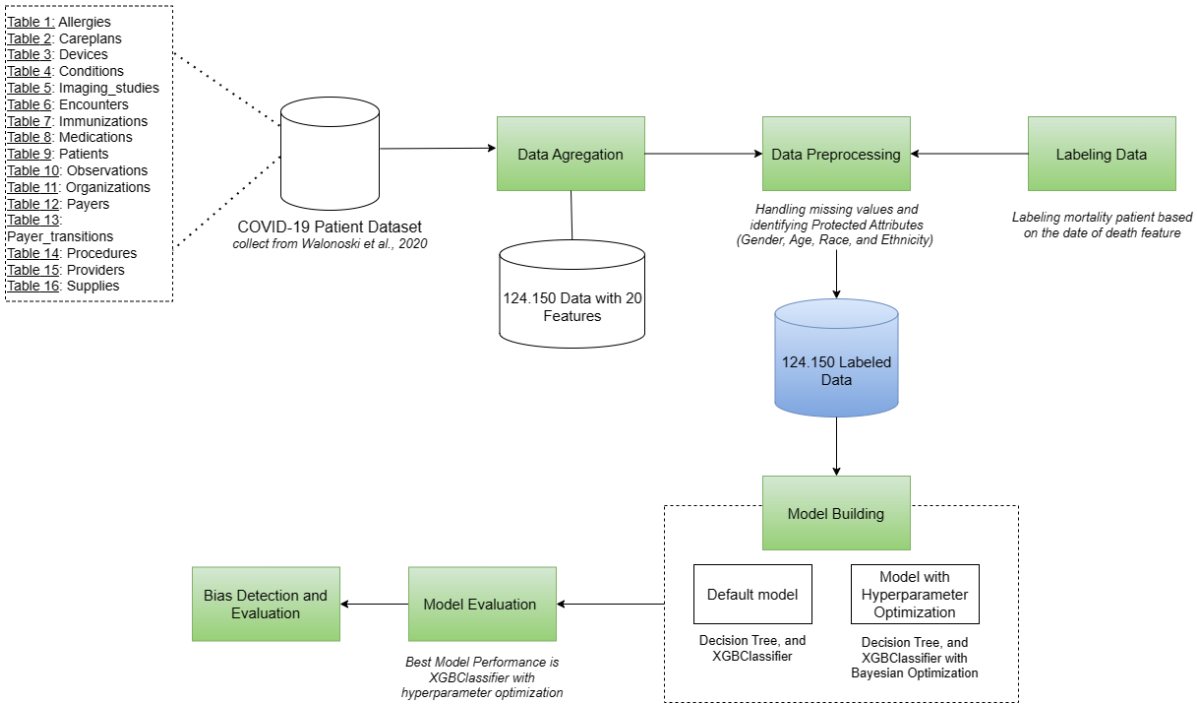

Figure 2. The system architecture for developing an ML model to predict mortality among COVID-19 patients and to detect model bias

III. RESEARCH METHOD

This section outlines the research methods employed in the study, which are divided into several stages: data aggregation, data preprocessing, formation of classification models using the Decision Tree and XGBoost classifiers, and evaluation and interpretation of the developed model. Figure 1 illustrates the research flow, starting with the collection of COVID-19 patient data via the SyntheticMass website.

A. Dataset

The dataset utilized in this study is available through the SyntheticMass website, which was previously investigated by Walonoski et al. [13]. They employed this dataset for the development of algorithms and prototypes aimed at addressing current or potential future pandemics. Comprising 16 CSV-format tables, each file in this dataset contains patient medical record data. These 16 tables were consolidated into one aggregated table containing 124,150 entries and 21 attributes. Table 1 provides a detailed description of the attributes used in this study.

B. Data preprocessing

This section describes the preprocessing methods used to clean the data. The patient medical data utilized in this study contained numerous missing values, which are critical to address as they can impact model performance. The chosen technique for handling missing values involves filling them with the mode of the respective values. This method helps preserve the integrity and completeness of the data, thereby enhancing its contribution to analysis and model building. By substituting null values with the mode, we maintain the optimal amount of data. Rather than eliminating null data—which reduces the sample size available for analysis—filling in with the mode retains most of the existing information. The next step in preprocessing is label encoding. Generally, machine learning models perform best with integer (int) and floating-point (float) values. As some attributes in this dataset are of the 'object' data type, it is necessary to apply label encoding to convert these attributes to integer types. Subsequently, the dataset is divided into 80% training data and 20% testing data for experimental analysis. This split allows the model to learn patterns from the majority of the data (80%) while using the remaining portion (20%) to evaluate model performance. Such an approach ensures the model generalizes well to new data and helps prevent overfitting, where a model performs well on training data but poorly on testing data [21]. The training set is crucial for the model's learning phase, whereas the testing set acts as an independent dataset to validate the model's effectiveness.

TABLE 1
LIST OF ATTRIBUTES ON THE 100K COVID-19 DATASET USED IN BUILDING CLASSIFICATION AND BIAS DETECTION MODELS.

Attributes	Data type	Description
ID	categorical	Unique markers for each patient
Marital	categorical	Markers in terms of patient status
Race	categorical	Markers of the patient's race
Ethnicity	categorical	Markers of the patient's ethnicity
Gender	categorical	Gender markers
Healthcare_expenses	numerical	Health service costs
Healthcare_coverage	numerical	Costs covered by the organization
NumImmunizations	numerical	Number of patient immunization history
NumPayer_transitions	numerical	Number of patient payment transition history
NumDevices	numerical	Number of medical equipment
NumSupplies	numerical	Number of health care providers
NumAllergies	numerical	Number of patient allergy history
NumCareplans	numerical	Number of patient treatment plans
NumImages	numerical	Number of image studies of disease
NumEncounters	numerical	Number of patient consultation history
NumObservations	numerical	Number of medical observations
NumMedications	numerical	Number of patient medications
NumProcedures	numerical	Number of medical procedures
NumConditions	numerical	Number of patient conditions
Age	numerical	Marker in terms of age
Dead	Boolean	Marker of patient death

C. Methodology

Machine learning algorithms have inherent limitations, and building models with high accuracy across diverse datasets is challenging. Consequently, using multiple models may enhance overall prediction accuracy. This research proposes two classifier models: decision tree and XGBoost.

A decision tree is a machine learning algorithm that operates as a decision-making tool. It constructs a classification model in the form of a tree structure, where each node represents a test on an attribute, and branches illustrate the outcomes of these tests [22]. During each iteration, the algorithm partitions the training set based on the outcomes of discrete functions and input attributes. Nodes are continuously split following these criteria until predefined stopping conditions are met. While decision trees are straightforward and easy to understand, their decision-making can sometimes be suboptimal, leading to incorrect conclusions [23].

XGBoost, or eXtreme Gradient Boosting, is a popular algorithm in supervised learning tasks such as classification, regression, and ranking. It builds on the principles of gradient boosting and is recognized for its high performance, scalability, and efficiency. The XGBClassifier module, specifically designed for classification, offers several adjustable hyperparameters to enhance model performance. XGBoost excels in processing large-scale data and has been effectively applied in various sectors including healthcare, finance, and industrial settings. Its exceptional performance and versatility have led to widespread adoption within the machine learning community.

D. Hyperparameter Optimization

Hyperparameter optimization involves identifying the best combination of hyperparameters in a machine learning (ML) model to enhance its performance. Hyperparameters, distinct from model parameters, are not learned from the training data but set prior to the learning process. Optimizing these hyperparameters can significantly affect the model's efficacy. In this research, Bayesian optimization [24], is employed to streamline this process. Bayesian optimization aims to minimize the number of evaluations needed to identify the optimal hyperparameters. It begins by constructing a probabilistic model that predicts the behavior of the objective function. This model is progressively refined through sampling and uses Bayes' rule to update the posterior distribution, incorporating new data to improve the estimation of the objective function. The optimization process also involves an acquisition function, which guides the selection of the next sample point by determining the most promising hyperparameters to evaluate next.

E. Model Agnostic Language for Exploration and Explanation (DALEX)

DALEX is an open-source software package that facilitates the understanding and explanation of prediction results from machine learning (ML) models. It uses a "model agnostic" approach, meaning it

operates independently of the specific details of any ML model, such as classification, regression, or deep learning models [25]. DALEX achieves this through an object called an explainer, which represents the model and allows users to explore it without knowledge of its internal workings. One key feature is the measurement of variable contributions, helping to pinpoint attributes significantly impacting predictions and uncovering potential biases within the model [25]. Model performance is commonly evaluated by accuracy, aiding in ranking and selecting the best models. However, accuracy alone does not indicate the absence of bias or guarantee that the model performs fairly. The trade-off between accuracy and fairness is critical, as reducing discrimination often means compromising performance. To address these issues, DALEX introduces the dalex.explainer class, which provides methods for both explaining and assessing the fairness of ML models [26]. It includes the fairness_check() function, which assesses fairness metrics and returns a fairness_object. This object integrates models with metrics in a structured format, enhancing the understanding of fairness through visualizations. The fairness_check() function also examines sensitive attributes—denoted as "protected" parameters like gender, race, and ethnicity—and identifies "privileged" parameters within these attributes. Additionally, DALEX functions produce numeric summaries in a tabular format, which can be visualized using its plot function to compare outcomes from multiple ML models.

IV. Result and Discussion

A. Model Performance of Mortality Patient

The classification model designed to detect COVID-19 related deaths has been tested using an aggregated dataset consisting of 124,150 data points, evaluated primarily on accuracy. This dataset

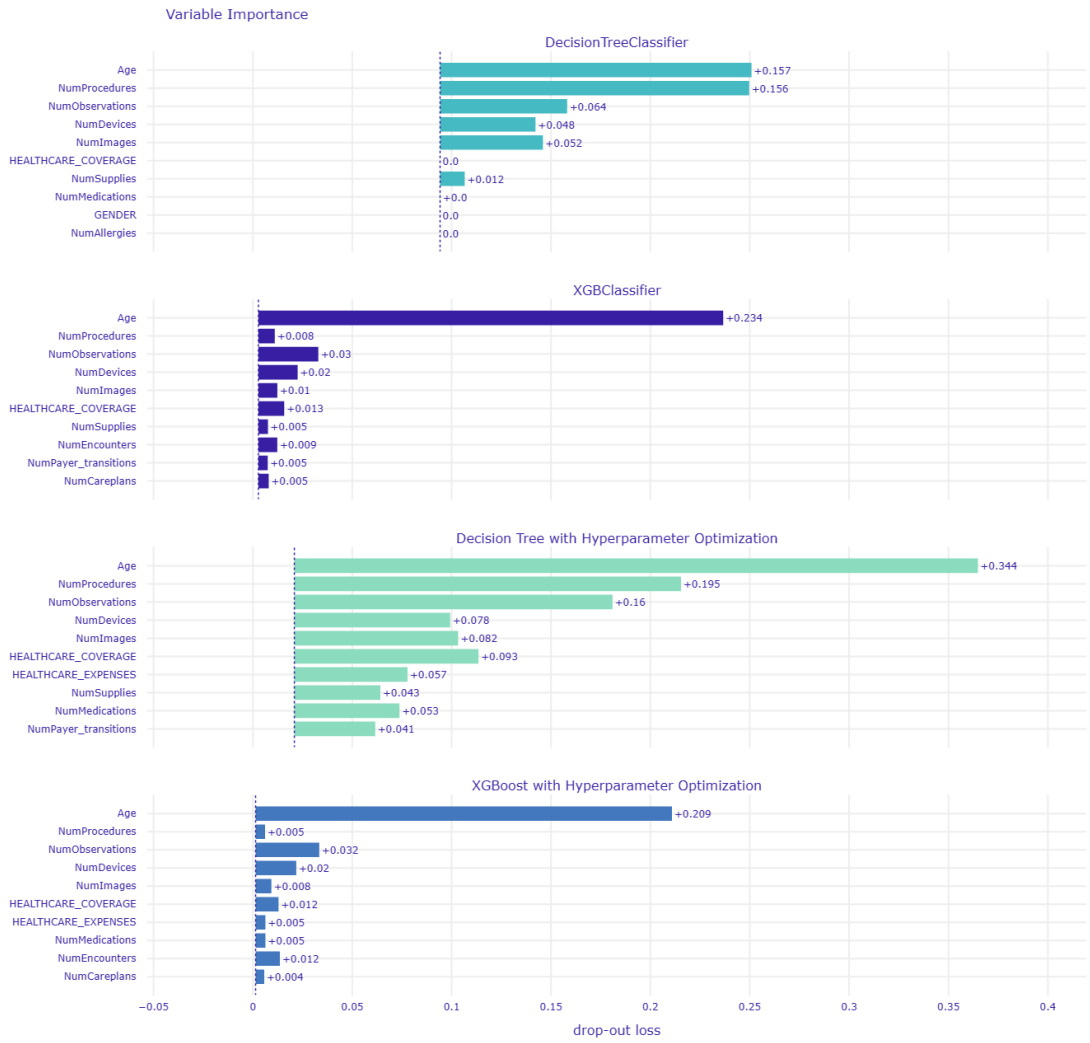

Figure 3. Features contribution of models in predicting mortality patient COVID-19

TABLE 2
MODEL PERFORMANCE METRICS FOR CLASSIFICATION RESULT OF MORTALITY PATIENT

Algorithm	Before Optimization	After Optimization
XGBClassifier	Prediction accuracy: 98%	Prediction accuracy: 98%
	Default hyperparameters: max_depth = 5, min_samples_split = 2, min_sample_leaf = 1	Hyperparameters optimized: max_depth = 47, min_samples_split = 2, min_sample_leaf = 1
DecisionTreeClassifier	Prediction accuracy: 93%	Prediction accuracy: 99%
	Default hyperparameters: colsample_bytree = 1, gamma = 0, learning_rate = 0.3, n = 100, subsample = 1	Hyperparameters optimized: colsample_bytree = 0.8, gamma = 0, learning_rate = 0.29, max_depth = 7, n = 152, subsample = 1

TABLE 3
RATIOS METRICS IN FAIRNESS CHECK ON EACH COMBINATION OF PROTECTED ATTRIBUTES

Algorithm	Combination of attributes	Protected Attributes	TPR	ACC	PPV	STP
Decision Tree with Parameter Optimization	Gender + Age	Female Old	2.71	0.96	0.93	36.8
		Female Young	0.86	1.004	1.01	0.4
		Male Old	2.75	0.95	0.95	51.6
	Gender + Race	Female Asian	1.004	0.99	0.99	1.04
		Female black	0.99	1	1	0.95
		Female native	0.99	0.97	0.99	1.42
		Female other	0.97	0.97	0.89	1.09
		Male Asian	0.99	0.98	1.03	1.36
		Male black	1.01	0.99	1.03	1.35
		Male native	1.10	1	1.03	1.94
		Male other	0.76	0.97	1.09	0.75
		Male white	1.01	0.99	1.02	1.43
	Gender + Ethnicity	Female Hispanic	0.98	1.01	0.98	0.66
		Female non-Hispanic	0.98	1.008	0.97	0.70
		Male Hispanic	0.99	1.003	1.008	0.93

underwent preprocessing to enhance its quality for testing with the ML model, including filling missing values and applying label encoding to categorical attributes. The preprocessed data was then tested using Decision Tree and XGBoost algorithms, both of which utilized Bayesian optimization to refine hyperparameters and enhance model performance. Best performance model was produced by Decision Tree with Bayesian hyperparamter optimization which the accuracy is 99%. However, this high level of accuracy should be approached with caution in decision-making processes, as it does not necessarily confirm the absence of bias within the model.

Bias detection in DALEX is utilized to verify that the accuracy achieved by a model also supports fairness in decision-making. DALEX offers a function for exploring ML models, which helps to clarify how various variables or features influence model predictions. The outcomes of this exploration are illustrated in Figure 2. The plot reveals that among the features examined across both models, the age feature has the most substantial influence on the prediction outcomes, indicating its significant impact.

B. Bias Detection

This research focuses on detecting bias in previously developed ML models, specifically classification models that employ Decision Tree and XGBoost algorithms. The bias detection process targets the model with the best performance, which, in this case, is the Decision Tree optimized through hyperparameter adjustments. Bias detection will be conducted on protected attributes such as gender, age, race, and ethnicity. These attributes will be examined in combinations—gender with age, gender with race, and gender with ethnicity—to enhance the detection of potential biases. The outcomes of this analysis will be documented in a table displaying various fairness metrics.

The DALEX analysis reveals that the model tends to make unbalanced predictions for certain attribute groups. As detailed in Table 3, the fairness metrics generated using the DALEX library indicate bias in the classification model. These metrics are calculated through comparative measurements, using an epsilon value as a benchmark. The epsilon value measures the degree of fairness or imbalance in the model; it sets a threshold or tolerance for imbalance or unfairness. In this study, the epsilon value is set at 0.8, aligning with the four-fifths rule. According to this rule, if the selection rate for a particular group is less than 80% of the rate for the group with the highest selection rate, it signifies adverse impact on the former group [27]. A difference between groups, with one considered privileged and another not,

that falls below one per epsilon value, indicates significant imbalance or injustice. Hence, a fairness metric is considered free from discrimination if it falls within the range of 0.8 to 1.25.

Combining the attributes of gender and age, researchers identified young men as the privileged group. However, the fairness metrics indicate bias against young women, as evidenced by their Statistical Parity (STP) value falling below the acceptable range. In another variable combination of gender and race, women of white race are considered privileged. The fairness metrics from Table 3 show that men of other races experience bias, with their TPR being below the reasonable range. The final combination, involving gender and ethnicity, designates non-Hispanic men as the privileged group. Here, non-Hispanic women face discrimination, indicated by their STP value being below the normal range. These results underscore the importance of identifying and addressing biases in model predictions, particularly for specific groups. Such imbalances may stem from uneven data distribution or underlying societal factors. The presence of bias can significantly affect clinical decision-making and the formulation of public health policies. Therefore, efforts to mitigate bias and enhance model fairness should be prioritized, especially in healthcare settings, to improve the overall utility of the models.

V. CONCLUSION

This research aims to detect bias in machine learning (ML) models, specifically focusing on the Decision Tree algorithm enhanced by Bayesian hyperparameter optimization to elevate model performance. The algorithm was applied to the SyntheaCOVID10k dataset, classifying mortality among COVID-19 patients. The DALEX library, used for both model evaluation and bias detection, facilitated this process. Although the Decision Tree model with hyperparameter optimization achieved an impressive 99% accuracy, this high accuracy does not inherently ensure fairness. Subsequent investigations revealed bias against certain protected attributes including gender, race, and ethnicity. Notably, biases were identified affecting young women, men of other races, and non-Hispanic women, illustrating inherent injustices in the model outcomes. The presence of such biases can significantly skew the decisions made by the model. Therefore, this research not only highlights current biases but also suggests potential future studies aimed at mitigating these biases to develop fairer ML models.

REFERENCES

[1] C. W. Morfi *et al.*, "Kajian terkini Coronavirus disease 2019 (COVID-19)," *J. Ilmu Kesehat. Indones.*, vol. 1, no. 1, 2020.
[2] C. Long *et al.*, "Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT?," *Eur. J. Radiol.*, vol. 126, p. 108961, 2020.
[3] M. R. H. Mondal, S. Bharati, and P. Podder, "Diagnosis of COVID-19 using machine learning and deep learning: a review," *Curr. Med. Imaging*, vol. 17, no. 12, pp. 1403–1418, 2021.
[4] J. Ammar, "Cyber Gremlin: social networking, machine learning and the global war on Al-Qaida-and IS-inspired terrorism," *Int. J. Law Inf. Technol.*, vol. 27, no. 3, pp. 238–265, 2019.
[5] M. Hardt *et al.*, "Amazon sagemaker clarify: Machine learning bias detection and explainability in the cloud," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 2974–2983.
[6] S. Alelyani, "Detection and evaluation of machine learning bias," *Appl. Sci.*, vol. 11, no. 14, p. 6271, 2021.
[7] J. Wisniewski and P. Biecek, "fairmodels: a Flexible Tool for Bias Detection, Visualization, and Mitigation in Binary Classification Models.," *R J.*, vol. 14, no. 1, pp. 227–243, 2022.
[8] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–35, 2021.
[9] S. Mitchell, E. Potash, S. Barocas, A. D'Amour, and K. Lum, "Algorithmic fairness: Choices, assumptions, and definitions," *Annu. Rev. Stat. its Appl.*, vol. 8, pp. 141–163, 2021.
[10] Y. K. Dwivedi *et al.*, "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *Int. J. Inf. Manage.*, vol. 57, p. 101994, 2021.
[11] W. Seymour, "Detecting bias: does an algorithm have to be transparent in order to Be Fair?," *BIAS 2018*, 2018.
[12] Synthea Development Team, "SyntheticMass." Accessed: Sep. 09, 2023. [Online]. Available: https://synthea.mitre.org/
[13] J. Walonoski *et al.*, "Synthea™ Novel coronavirus (COVID-19) model and synthetic data set," *Intell. Med.*, vol. 1, p. 100007, 2020.
[14] Y. Zoabi, S. Deri-Rozov, and N. Shomron, "Machine learning-based prediction of COVID-19 diagnosis based on symptoms," *npj Digit. Med.*, vol. 4, no. 1, pp. 1–5, 2021.
[15] S. S. Zakariaee, N. Naderi, M. Ebrahimi, and H. Kazemi-Arpanahi, "Comparing machine learning algorithms to predict COVID-19 mortality using a dataset including chest computed tomography severity score data," *Sci. Rep.*, vol. 13, no. 1, p. 11343, 2023.
[16] N. M. Elshennawy, D. M. Ibrahim, A. M. Sarhan, and M. Arafa, "Deep-Risk: Deep Learning-Based Mortality Risk Predictive Models for COVID-19," *Diagnostics*, vol. 12, no. 8, p. 1847, 2022.
[17] N. Rai, N. Kaushik, D. Kumar, C. Raj, and A. Ali, "Mortality prediction of COVID-19 patients using soft voting classifier," *Int. J. Cogn. Comput. Eng.*, vol. 3, pp. 172–179, 2022.
[18] H. Estiri *et al.*, "An objective framework for evaluating unrecognized bias in medical AI models predicting COVID-19 outcomes," *J. Am. Med. Informatics Assoc.*, vol. 29, no. 8, pp. 1334–1341, 2022.
[19] S. Afrose, W. Song, C. B. Nemeroff, C. Lu, and D. Yao, "Subpopulation-specific machine learning prognosis for underrepresented patients with double prioritized bias correction," *Commun. Med.*, vol. 2, no. 1, p. 111, 2022.
[20] A. Allen *et al.*, "A racially unbiased, machine learning approach to prediction of mortality: algorithm development study," *JMIR public Heal. Surveill.*, vol. 6, no. 4, p. e22400, 2020.

[21] H. J. Seltman, “Experimental design and analysis,” *B. is World Wide Web*, 2018.
[22] Priyanka and D. Kumar, “Decision tree classifier: a detailed survey,” *Int. J. Inf. Decis. Sci.*, vol. 12, no. 3, pp. 246–269, 2020.
[23] B. Charbuty and A. Abdulazeez, “Classification based on decision tree algorithm for machine learning,” *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021.
[24] X. Wang, Y. Jin, S. Schmitt, and M. Olhofer, “Recent advances in Bayesian optimization,” *ACM Comput. Surv.*, vol. 55, no. 13s, pp. 1–36, 2023.
[25] P. Biecek, “DALEX: Explainers for complex predictive models in R,” *J. Mach. Learn. Res.*, vol. 19, no. 84, pp. 1–5, 2018.
[26] H. Baniecki, W. Kretowicz, P. PiÄ, and J. WiĹ, “Dalex: responsible machine learning with interactive explainability and fairness in python,” *J. Mach. Learn. Res.*, vol. 22, no. 214, pp. 1–7, 2021.
[27] N. Mondragon, “What is Adverse Impact? And Why Measuring It Matters,” HireVue. Accessed: Sep. 09, 2023. [Online]. Available: https://www.hirevue.com/blog/hiring/what-is-adverse-impact-and-why-measuring-it-matters