Addressing Data Imbalance: Mitigating Racial Bias in Criminal Justice

Rishabh Kala, Neha Jagtap, Shyamal Gandhi, Shail Shah
rkala@ncsu.edu, njagtap@ncsu.edu, sgandhi6@ncsu.edu, sshah38@ncsu.edu
Department of Computer Science
North Carolina State University
Raleigh, North Carolina, USA

I. Problem Definition

The use of risk assessment tools like COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) in predicting inmate recidivism has revealed concerns about racial bias in the criminal justice system, particularly highlighted by ProPublica's analysis of COMPAS risk scores for Florida inmates. These disparities are compounded by imbalanced datasets within COMPAS, where one class (e.g., low-risk individuals) significantly outnumbers another (e.g., high-risk individuals). Such imbalances exacerbate the potential for biased predictions, undermining fairness and accuracy in risk assessment outcomes. The analysis done by ProPublica consists of a two-year follow-up study which gauges the true accuracy and metrics of the COMPAS system. Using the data from this analysis we have tried to study the cause for this bias. This research aims to investigate the intersection of racial bias and imbalanced datasets in COMPAS risk assessment. By understanding how imbalances contribute to disparities in risk scores across racial groups, this study seeks to identify strategies for mitigating bias and promoting fairness in risk assessment practices within the criminal justice system. We aim to assess if standard class-imbalance techniques can be applied to this data to mitigate this bias.

II. FACTUAL OVERVIEW

This section consists of the overview of related work and research papers written and published by researchers that focus on the problem of data imbalance and the finding solutions to them using variety of approaches.

Rawat. et. al investigates strategies for reducing class disparity in categorization issues [1]. It explores the difficulties caused by skewed datasets, in which one class greatly dominates the other, making conventional machine learning techniques useless. The survey examines a range of procedures, including cost-sensitive learning, hybrid approaches, algorithm-level approaches, resampling (under- and oversampling), and deep learning techniques. Every technique is carefully examined for its benefits, drawbacks, and suitability for use in various fields. Additionally, the study emphasizes the significance of assessment metrics in evaluating classifier performance, offering readers thorough insights to guide their decision-making in effectively managing class imbalance.

Wang. et. al examines the problem of class imbalance in medical big data categorization, which is a crucial component of artificial intelligence-based auxiliary diagnosis research [2]. It tackles the shortcomings of current methods caused by problems like marginalisation and parameter selection blindness, especially with the SMOTE algorithm. The study attempts to address these issues by distributing fresh sample points closer to the centre of the minority class, hence preventing data marginalisation, and by using an improved SMOTE algorithm based on Normal distribution. The experimental findings indicate better classification performance than the original SMOTE algorithm on datasets such as Pima, WDBC, WPBC, Ionosphere, and Breast-cancer-wisconsin. Interestingly, the effectiveness of the suggested algorithm is demonstrated by several assessment measures, including AUC, F-value, G-value, and OOB error.

In addition, parameter analysis emphasises how crucial it is to choose the right parameters in order to preserve the original data's distribution properties and maximise classification results. Overall, the study offers a fresh strategy for resolving class imbalance in the classification of medical data, with encouraging findings and suggestions for further development and use.

Applications such as smart home monitoring and healthcare depend on human activity detection, but developing reliable algorithms from unbalanced data is difficult [3]. Hamad et. al. tackles class imbalance in deep learning for smart home data, concentrating on binary sensor-based Activities of Daily Living (ADL) detection. To improve sensitivity to minority classes, the study use temporal window approaches in conjunction with a data-level perspective. In comparison to algorithm-level techniques, experimental results reveal considerable gains in classification performance, demonstrating the superiority of managing unbalanced data at the data level. Utilising techniques such as SMOTE in conjunction with CNN and LSTM models improves detection, especially for minority classes, highlighting the significance of data-level solutions in smart home activity recognition systems.

In order to develop AI systems while maintaining patient privacy, Barbara provides a thorough investigation of the detection and correction of data bias in healthcare using the BayesBoost algorithm [4]. The technique systematically describes the procedures for uncertainty analysis and the creation of synthetic data, enabling the identification of underrepresented groups and the creation of synthetic datasets that are representative. During the experimental stage, which used artificial datasets from the Clinical Practice Research Datalink (CPRD), BayesBoost's ability to identify biases, rectify them, and enhance classification accuracy is shown. The study indicates that BayesBoost is superior to other strategies like SMOTE and AdaSyn in terms of getting greater prediction performance while maintaining data fairness.

Through the lens of class-imbalanced few-shot learning (CIFSL), Ochal et. al. examines the effects of class imbalance on few-shot learning (FSL) [5]. It explores the behaviour of FSL algorithms on datasets with notable differences in class distribution. The study challenges previous assumptions by showing that FSL algorithms perform less well on imbalanced datasets than on balanced datasets. Furthermore, the research investigates the effectiveness of random-shot meta-training in resolving imbalanced tasks and concludes that it is not always beneficial and may even be detrimental. In addition, the paper suggests using supervised learning adaptations such reweighting losses and random oversampling (ROS) to lessen the effects of class imbalance, with encouraging outcomes.

III. Proposed Approach

A. Novel Aspects

This research addresses the pervasive issue of racial bias in the criminal justice system, focusing on risk assessment practices. By exploring the intersectionality of racial bias and dataset imbalance within the COMPAS framework, it uncovers how imbalanced datasets contribute to racial disparities in risk scores. Through rigorous empirical analysis of COMPAS data and demographic information, the study offers data-driven insights into this complex interplay. Additionally, it proposes mitigation strategies to promote fairness in risk assessment, providing actionable recommendations for policymakers and stakeholders involved in criminal justice reform efforts.

B. Methodology

Our methodology encompasses a systematic approach that follows several key steps to address the challenges posed by data imbalance and model biases in predicting inmate recidivism. Beginning with data cleaning and processing, we then conduct an exploratory data analysis to understand the dataset's characteristics, ensuring data consistency and to select a baseline model for our data. Subsequently, we construct a diverse set of models, including Naive Bayes, Random Forests, Decision Trees, MultiLayer Neural Network (MLP) classifier, Logistic Regression, Support Vector Machine (SVM), Voting Ensemble, and Stacking Ensemble, using pre-built models from the sklearn library on the data to predict recidivism. These models are then evaluated using comprehensive performance metrics to assess their effectiveness in mitigating bias and enhancing fairness and accuracy in predicting recidivism. Based on these metrics we pick a baseline model

for our research. Then to tackle data imbalance, we employ techniques such as random sampling and Synthetic Minority Over-sampling Technique (SMOTE) to create a more balanced dataset. Additionally, we leverage synthetic data generation using LLM and introduce model biasing techniques with custom loss functions to further refine our models. Through this rigorous methodology, we aim to provide actionable insights for policymakers and stakeholders involved in criminal justice reform efforts.

C. *Rationale*

Our research focuses on the critical issue of data imbalance and its impact on model bias in criminal justice risk assessment. We aim to understand if the unbalanced nature of the data is one of the causes for biased predictions in models and identify effective strategies for mitigating bias and improving model fairness. By assessing the extent of bias, investigating underlying mechanisms, and comparing different mitigation strategies, we seek to provide insightful recommendations for promoting fairness, equity, and transparency in predictive modeling within the criminal justice system. Through empirical analysis and stakeholder engagement, our research aims to advance understanding and contribute to the development of fairer and more equitable predictive models. We are using the data from a two-year follow-up study to compare the prediction of the old model with the follow-up data and then further apply imbalance mitigating techniques to gauge their effect on the bias.

IV. DATASET

Our research utilizes the COMPAS dataset, sourced from Kaggle [6], to investigate the impact of data imbalance on model bias in criminal justice risk assessment. The dataset contains information on individuals subjected to the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) risk assessment tool, widely used in the criminal justice system for predicting recidivism risk.

The dataset comprises a comprehensive set of features, including demographic information (such as age, race, and gender), criminal history, and COMPAS risk scores, among others. Notably, the dataset also includes ground truth labels indicating whether individuals were re-arrested within a certain timeframe, providing valuable data for model training and evaluation.

However, one key challenge of the dataset is its inherent imbalance, where the distribution of individuals across risk categories (e.g., low, medium, high) may be skewed, potentially leading to biased predictions in predictive models. This imbalance is particularly salient in the distribution of racial and ethnic groups within the dataset, raising concerns about the fairness and equity of risk assessment outcomes.

By leveraging this dataset, our research aims to explore the relationship between data imbalance, model bias, and demographic factors such as race and ethnicity. Through rigorous analysis and experimentation, we seek to uncover insights into the drivers of bias in COMPAS risk assessment and identify strategies for mitigating bias and improving model fairness. Ultimately, our research endeavors to contribute to the development of fairer and more equitable risk assessment practices within the criminal justice system

A. *Dataset Analysis*

In our analysis of the COMPAS dataset, we observe pronounced data imbalances both in the distribution of recidivism risk categories and across racial groups. The dataset reveals a significant disproportion among risk categories, with a majority of individuals classified as ”Low” risk, totaling 6258 individuals, while ”Medium” and ”High” risk categories exhibit notably lower counts of 2845 and 2183 individuals, respectively. Similarly, racial disparities are evident, with African-American individuals comprising the largest group at 5641, followed by Caucasian individuals at 3902, Hispanic individuals at 1044, and smaller representation for ”Other” at 617, Asian at 58, and Native American at 38 individuals. These imbalances pose challenges for predictive modeling, potentially leading to biased predictions and inaccuracies in risk assessment outcomes, affecting decision-making within the criminal justice system. Addressing these imbalances is crucial for ensuring fairness and equity. Thus, our research seeks to investigate their influence on model bias and develop strategies to mitigate bias,

fostering more accurate and equitable predictive models for criminal justice risk assessment.

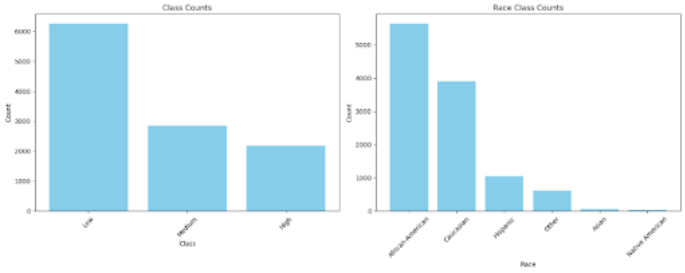


Fig. 1. Class and race counts

V. Model Predictions before applying techniques to address data imbalance

We will compare our results before and after implementing mitigation strategies to assess their effectiveness in addressing bias and improving the accuracy and fairness of our predictive models. Before implementing strategies for mitigation, we initially used various classifiers for prediction. These classifiers include Naive Bayes, Decision Tree, Random Forest, Logistic Regression, Multilayer Neural Network, and Support Vector Machine (SVM)

A. *Analysis of model*

Here's a detailed analysis of the model performance metrics and the impact of the imbalanced dataset on the predictions:

1) Naive Bayes: Despite achieving a moderate accuracy of 0.70, Naive Bayes exhibits low precision (0.63) and recall (0.24) for predicting individuals at risk of recidivism. The F1 score, which balances precision and recall, is also relatively low at 0.34. This indicates that the model struggles to correctly identify individuals at high risk of recidivism, likely due to the imbalance in the dataset. The low recall suggests that the model misses a significant number of true positive instances, leading to biased predictions favoring the majority class.

2) Decision Tree and Random Forest: Both Decision Tree and Random Forest models demonstrate similar challenges in predicting individuals at high risk of recidivism. While they achieve reasonable accuracy scores (0.70), their precision and recall for the positive class are suboptimal. This suggests that the models may have difficulty accurately predicting individuals at high risk of recidivism due to the imbalance in the dataset. The imbalance likely leads to biased predictions, with the models favoring the majority class and exhibiting lower sensitivity to the minority class.

3) Logistic Regression and SVM: Logistic Regression and SVM models also face challenges in effectively capturing individuals at high risk of recidivism. Despite achieving comparable accuracy scores (0.70), their precision and recall for the positive class are relatively low. This indicates that these models struggle to accurately identify individuals at high risk, likely due to the imbalance in the dataset. The imbalance introduces bias in the predictions, leading to lower sensitivity to the minority class and potentially inaccurate risk assessments.

4) Multilayer Neural Network: The Multilayer Neural Network outperforms other models with the highest accuracy score of 0.73. Additionally, it demonstrates relatively higher precision and recall for the positive class. However, the model still exhibits challenges in accurately predicting individuals at high risk of recidivism, as indicated by the moderate F1 score of 0.49. While the neural network shows improvement in handling imbalanced data compared to other models, there is still room for enhancement in effectively capturing individuals at high risk.

Overall, the analysis highlights the significant impact of imbalanced dataset on model predictions. The imbalance leads to biased predictions, with the models showing lower sensitivity to the minority class and potentially inaccurate risk assessments for individuals at high risk of recidivism. Addressing these challenges through mitigation strategies is crucial to improving the fairness and accuracy of predictive models in criminal justice risk assessment.

VI. Experiments

A. *Hypothesis*

The observation that Northpointe's assessment tool correctly predicts recidivism 61 percent of the time, but exhibits differential error rates between racial groups, suggests potential biases within the criminal justice system. Blacks are almost twice as likely as whites to be labeled a higher risk without re-offending, while whites are more likely

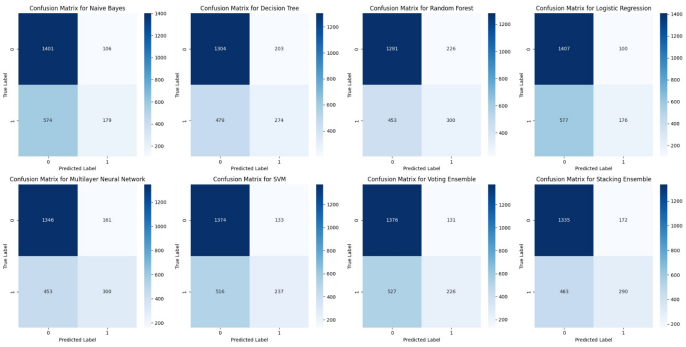


Fig. 2. Confusion Matrix for all applied algorithms

to be labeled lower risk but go on to commit other crimes. This disparity underscores the presence of systematic biases in predictive models, potentially exacerbated by feature imbalances in the dataset. Our hypotheses aim to investigate these biases and their relationship to feature imbalance, seeking to ensure fairness and equity in risk assessments and decision-making processes within the criminal justice system.

B. *Experimental Design*

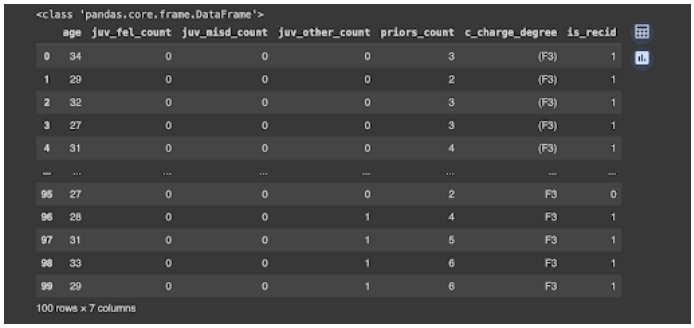
1) Data Cleaning and Processing: With the goal of creating AI systems that preserve patient privacy, the paper provides a thorough analysis of the identification and rectification of data bias in healthcare using the BayesBoost technique. The process of creating representative synthetic datasets and doing uncertainty analysis are described in detail in the approach, which also successfully identifies underrepresented groups. Using synthetic datasets from the Clinical Practice Research Datalink (CPRD), the testing phase shows how effective BayesBoost is at identifying biases, fixing them, and enhancing classification accuracy. The study demonstrates the advantage of BayesBoost in achieving greater prediction performance while maintaining data fairness by comparing findings with other techniques such as SMOTE and AdaSyn.

2) Addressing Data Imbalance: To address the imbalance of the dataset, we first tackle both feature and target imbalances. Feature imbalance refers to the unequal distribution of racial attributes within the dataset, while target imbalance indicates an unequal distribution of outcomes (e.g., recidivism rates) across different racial groups. To mitigate these imbalances, we employ two techniques: random sampling and Synthetic Minority Oversampling Technique (SMOTE).

Random sampling involves selecting a subset of data points from each racial group randomly, ensuring a more balanced representation of racial attributes in the dataset. This approach helps to alleviate feature imbalance by equalizing the number of samples across different racial groups.

SMOTE is a data augmentation technique specifically designed to address target imbalance by oversampling minority class instances. It works by generating synthetic examples of the minority class using interpolation between existing minority class samples. By creating synthetic data points, SMOTE increases the representation of the minority class, thereby mitigating target imbalance and improving the overall balance of the dataset.

We are constructing a diverse set of models to compare their performance both before and after applying sampling techniques. The models include Naive Bayes, Random Forests, Decision Trees, MultiLayer Neural Network (MLP) classifier, Logistic Regression, Support Vector Machine (SVM), Voting Ensemble, and Stacking Ensemble. Each of these models represents a distinct approach to classification, ranging from simple probabilistic methods like Naive Bayes to more complex ensemble techniques such as Voting and Stacking. By evaluating these models before and after sampling, we aim to gauge the effectiveness of the sampling techniques in mitigating bias and enhancing the fairness and accuracy of predictions, particularly concerning the association of race with the risk of recidivism within the dataset.



<class 'pandas.core.frame.DataFrame'>

	age	juv_fel_count	juv_misd_count	juv_other_count	priors_count	c_charge_degree	is_recid
0	34	0	0	0	3	(F3)	1
1	29	0	0	0	2	(F3)	1
2	32	0	0	0	3	(F3)	1
3	27	0	0	0	3	(F3)	1
4	31	0	0	0	4	(F3)	1
...	...	...	...	...	...	...	...
95	27	0	0	0	2	F3	0
96	28	0	0	1	4	F3	1
97	31	0	0	1	5	F3	1
98	33	0	0	1	6	F3	1
99	29	0	0	1	6	F3	1

100 rows × 7 columns

Fig. 3. Augmented Data Frame

3) Synthetic Data generation using LLM: To generate synthetic data, we utilize the LLM (Large Language Model) framework using LangChain and LangChain experimental libraries from OpenAI. Specifically, we employ the FewShotPromptTemplate from LangChain prompts, which facilitates generating synthetic data based on few-shot learning techniques. The FewShotPromptTemplate is a prompt-based approach that allows the model to learn from a small number of examples or shots provided as input. It leverages the capabilities of the LLM to understand and extrapolate patterns from the provided examples, enabling the generation of realistic synthetic data. By utilizing FewShotPromptTemplate, we enable the model to learn and generate synthetic data based on a limited set of input examples, making it suitable for scenarios where only a small amount of labeled data is available for training. This approach enhances the scalability and adaptability of the synthetic data generation process, enabling the creation of diverse and representative datasets for various applications.

4) Model Biasing with Custom Loss Function: In this phase, we introduce model biasing techniques using a custom loss function to address inherent biases present in the dataset. Specifically, our goal is to mitigate the false association of African-Americans with a high risk of recidivism. The custom loss function is designed to penalize the model for predicting a high risk of recidivism specifically for individuals of African-American race. By adjusting class weights within the loss function, we aim to counteract biased predictions and promote fairness in model outcomes.

5) Model Training: In our model training phase, we utilize pre-built models from the scikit-learn (sklearn) library (add sklearn documentation link if possible), a widely-used machine learning toolkit in Python. Leveraging the functionalities offered by sklearn, we train the following models:

1. Naive Bayes: We employ a Naive Bayes classifier, utilizing its simplicity and efficiency in handling categorical data. The classifier is trained on the preprocessed and balanced dataset to model the conditional independence between features, ensuring robust performance in classification tasks.

2. Random Forests: Our approach involves utilizing the Random Forest algorithm, renowned for capturing complex relationships and effectively handling imbalanced datasets. Multiple decision trees are trained on bootstrapped samples of the dataset, and their predictions are aggregated to enhance accuracy and generalization, thus bolstering the model's robustness.

3. Decision Tree: We implement a Decision Tree classifier to model nonlinear relationships and feature interactions. Through recursive partitioning, the dataset is split into homogeneous subsets based on feature conditions, optimizing splits to minimize impurity and ensuring both interpretability and computational efficiency.

4. MultiLayer Neural Network (MLP Classifier): Our strategy includes deploying a MultiLayer Perceptron (MLP) classifier to learn intricate patterns and capture nonlinear decision boundaries. Multiple layers of interconnected neurons are trained using backpropagation to minimize prediction error, enhancing the model's ability to grasp complex relationships within the data.

5. Logistic Regression: Logistic Regression is applied to estimate the probability of binary outcomes. We estimate the coefficients of the logistic function using maximum likelihood estimation and apply regularization techniques to prevent overfitting, ensuring interpretable and reliable results in our classification tasks.

6. Support Vector Machine (SVM): We train an SVM classifier to find the optimal hyperplane that separates data points of different classes with the maximum margin. Kernel functions are utilized to transform input data into a higher-dimensional space, maximizing the margin between support vectors and ensuring robust performance, especially in high-dimensional datasets.

7. Voting Ensemble: Our methodology involves constructing a Voting Ensemble by combining predictions from multiple base classifiers using majority voting. Diverse base classifiers with different algorithms and hyperparameters are trained, and their predictions are aggregated to improve overall predictive performance and robustness, thereby enhancing the model's generalization capability.

8. Stacking Ensemble: We build a Stacking Ensemble by training a meta-classifier on the predictions of multiple base classifiers. Leveraging cross-validation, we generate meta-features and train the

meta-classifier on these features to make final predictions, harnessing the strengths of diverse base classifiers to enhance predictive accuracy in our classification tasks.

9. Logistic Classifier (Logistic Regression): In our methodology, we incorporate a logistic classifier, also referred to as logistic regression, which is a widely-used statistical model for binary classification tasks. This classifier estimates the probability of a binary outcome based on input features by fitting a logistic function to the data. Recognized for its efficiency, interpretability, and suitability for scenarios where the relationship between input features and the target variable is linear, the logistic classifier plays a crucial role in our classification tasks.

10. MLP Classifier (Multilayer Perceptron): Our approach integrates an MLP classifier, also known as a multilayer perceptron, which is a type of artificial neural network comprising multiple layers of nodes (neurons). These neurons process input data through nonlinear transformations, allowing the MLP classifier to learn complex patterns and relationships in the data. Well-suited for tasks with nonlinear decision boundaries, the MLP classifier excels in capturing intricate feature interactions and can achieve high accuracy on diverse datasets, albeit at the expense of increased computational complexity.

6) Evaluation Models: Performance Metrics: Evaluate the trained models using a comprehensive set of performance metrics, including:

1. Accuracy: Proportion of correctly classified instances.

2. Precision: Proportion of true positive predictions among all positive predictions.

3. Recall (Sensitivity): Proportion of true positive predictions among all actual positive instances.

4. F1-score: Harmonic mean of precision and recall, balancing between precision and recall.

5. Area under the ROC curve (AUC-ROC): Measure of the classifier's ability to distinguish between classes.

6. Confusion Matrix: Tabulation of true positive, true negative, false positive, and false negative predictions.

In the evaluation phase, we assess the influence of dataset imbalances on model performance by conducting a comparative analysis of performance metrics before and after the implementation of sampling techniques. This analysis focuses on evaluating changes in precision, recall, and F1-score for both minority and majority classes to ascertain the effectiveness of the sampling methods in mitigating bias. Additionally, we scrutinize the Confusion Matrix to detect any alterations in the distribution of true positive, true negative, false positive, and false negative predictions. These observations serve to identify improvements or exacerbations of imbalances resulting from the sampling techniques.

VII. Results and Discussion

A. Model Performance

After employing random sampling and SMOTE techniques to address the data imbalance issue, we found that there was generally no substantial deviation in accuracy, recall, precision, and F1 score compared to the old models. However, there were certain exceptions observed, particularly in the case of the multilayer neural network and stacking ensembling methods, where notable differences were identified.

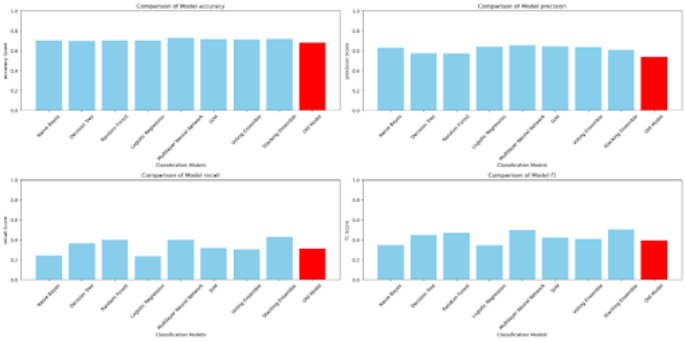


Fig. 4. Model Performance Comparison

B. Model Analysis

Despite implementing fairness-aware classifiers Multilayer Neural Network (MLP) and logistic regression models—with custom loss functions tailored to address fairness concerns, our evaluation did not reveal any significant deviation in performance compared to previous models. Despite efforts to mitigate biases and promote fairness by incorporating penalties for bias and demographic parity considerations, the accuracy, precision, recall, and F1 score remained largely consistent. This suggests that while our approach aimed to enhance fairness

in predictions, it did not result in discernible improvements in model performance.

Base Logistic Classifier
Accuracy: 0.69

	precision	recall	f1-score	support
0	0.7028	0.9222	0.7976	1246
1	0.6460	0.2670	0.3778	663
accuracy			0.6946	1909
macro avg	0.6744	0.5946	0.5877	1909
weighted avg	0.6830	0.6946	0.6518	1909

Biased Logistic Classifier
Accuracy: 0.70

	precision	recall	f1-score	support
0	0.7039	0.9213	0.7981	1246
1	0.6475	0.2715	0.3826	663
accuracy			0.6957	1909
macro avg	0.6757	0.5964	0.5903	1909
weighted avg	0.6843	0.6957	0.6538	1909

Fig. 5. Base and Biased Logistic Classifier

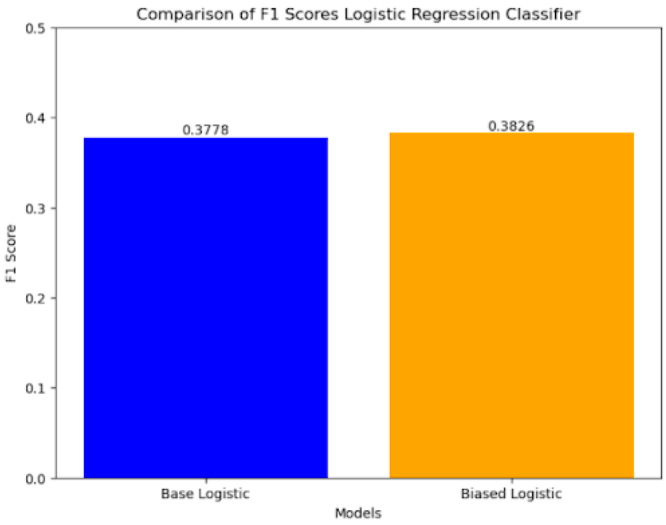


Fig. 6. Comparsion of teo classifiers based on F1 scores

VIII. Conclusion

After applying various techniques to address data imbalance, including random sampling and SMOTE, along with custom cost functions in model training, we observed minimal improvement in the model's predictive performance. Despite our efforts,

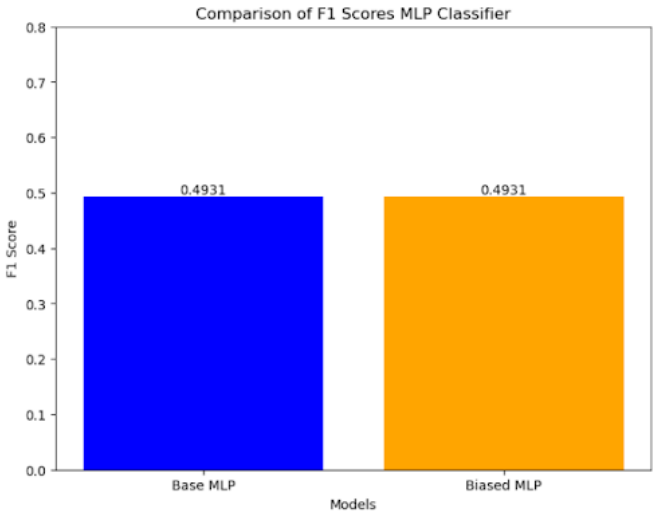


Fig. 7. Comparison of MLP classifiers based on F1 scores

Base MLP Classifier
Accuracy: 0.71

	precision	recall	f1-score	support
0	0.7340	0.8748	0.7982	1246
1	0.6321	0.4042	0.4931	663
accuracy			0.7114	1909
macro avg	0.6830	0.6395	0.6457	1909
weighted avg	0.6986	0.7114	0.6923	1909

Biased MLP Classifier
Accuracy: 0.71

	precision	recall	f1-score	support
0	0.7340	0.8748	0.7982	1246
1	0.6321	0.4042	0.4931	663
accuracy			0.7114	1909
macro avg	0.6830	0.6395	0.6457	1909
weighted avg	0.6986	0.7114	0.6923	1909

Fig. 8. Base and Biased MLP Classifier

there was no significant change in model accuracy, precision, recall, or F1 score Random sampling, while intended to rectify imbalance by randomly removing instances from the majority class, failed to capture crucial data patterns, potentially leading to suboptimal model outcomes. Similarly, SMOTE, aimed at rebalancing class distribution through synthetic sample generation, faced limitations in accurately representing the underlying minority class distribution. Moreover, neither method effectively tackled biases inherent in dataset features, par-

ticularly concerning sensitive attributes like race or gender, highlighting the persistence of bias in model predictions even after balancing class distribution. Furthermore, the success of these techniques relies heavily on data quality and quantity, with sparse or homogenous datasets posing challenges in generating representative samples. Additionally, the computational demands of SMOTE, particularly in high-dimensional datasets, hinder scalability. Thus, while random sampling and SMOTE offer valuable strategies for addressing data imbalance, their shortcomings emphasize the need for more nuanced approaches, such as data augmentation with Language Models, to cultivate fair and robust predictive models.

Furthermore, when utilizing data augmentation with the Language Model (LLM), we encountered limitations due to the cost and feasibility constraints, restricting us from generating a substantial amount of synthetic data. However, it's noteworthy that leveraging LLM for data augmentation has the potential to mitigate bias in model predictions significantly if sufficient data can be generated. Thus, while our current endeavors have not yielded substantial improvements, further exploration of LLM-based data augmentation could hold promise for enhancing model fairness and predictive accuracy in the future.

IX. Meeting Attendance

References

[1] S. Rawat and A. Mishra, "Review of methods for handling class-imbalanced in classification problems," 11 2022.

[2] S. Wang, Y. Dai, J. Shen, and J. Xuan, "Research on expansion and classification of imbalanced data based on smote algorithm," *Scientific Reports*, vol. 11, no. 1, p. 24039, 2021.

[3] R. A. Hamad, M. Kimura, and J. Lundström, "Efficacy of imbalanced data handling methods on deep learning for smart homes environments," *SN Computer Science*, vol. 1, no. 4, p. 204, 2020.

[4] B. Draghi, Z. Wang, P. Myles, and A. Tucker, "Identifying and handling data bias within primary healthcare data using synthetic data generators," *Heliyon*, vol. 10, no. 2, p. e24164, 2024.

[5] M. Ochal, M. Patacchiola, J. Vazquez, A. Storkey, and S. Wang, "Few-shot learning with class imbalance," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 5, pp. 1348–1358, 2023.

[6] Danofer, "Compass dataset." https://www.kaggle.com/datasets/danofer/compass.