# Tackling COVID-19 Data Imbalance and Health Disparity using Deep Transfer Learning

Shruti Sidharth Mandaokar
Masters in Computer Science, Purdue University Fort Wayne
mands01@pfw.edu

***Abstract*—The COVID-19 pandemic has exacerbated healthcare inequalities, disproportionately impacting minority populations in terms of infection rates, hospitalizations, and mortality. The rapid deployment of artificial intelligence (AI) models, often developed using imbalanced biomedical datasets, has the potential to amplify these disparities. This research introduces a novel approach leveraging deep transfer learning to address data imbalance and mitigate biases. By transferring knowledge from well-represented groups to underrepresented ones, we aim to create more equitable AI models, enhancing healthcare outcomes across all demographic groups. Our study applies fairness-aware machine learning techniques and compares Independent Learning, Mixed Learning and Transfer Learning to analyze COVID-19 mortality trends and address ethnic and racial disparities, contributing to the development of AI for social good and the advancement of fairness in machine learning algorithms.**

Keywords: COVID-19, Machine Learning, Independent Learning, Mixture Learning, Transfer Learning, Gradient Boosting Regressor, Data Analysis

## I. Introduction

The COVID-19 pandemic has had a profound and unprecedented impact on global health, economies, and societies. As the virus spread rapidly across the globe, it exposed and exacerbated pre-existing disparities in health outcomes among different demographic groups. In the United States, the pandemic has highlighted significant inequalities in infection rates, hospitalization, and mortality among various racial and ethnic populations. Understanding these disparities is crucial for developing targeted public health interventions and policies that can mitigate the impact of the virus on vulnerable groups.

Research has consistently shown that minority populations, including Black, Hispanic, Native American, and Asian communities, have borne a disproportionate burden of the COVID-19 pandemic. These groups have experienced higher rates of infection, severe disease, and mortality compared to their White counterparts. Several factors contribute to these disparities, including socio-economic conditions, access to healthcare, pre-existing health conditions, and occupational exposures.

Moreover, the pandemic has underscored the need for robust and fair predictive models in public health. Traditional epidemiological models often fall short in capturing the complex interactions between socio-economic, environmental, and biological factors that drive health disparities. This research is significant for several reasons. Firstly, it contributes to the growing body of literature on the impact of COVID-19 on minority populations, offering a detailed analysis using advanced machine learning techniques. Secondly, it addresses the critical issue of bias in predictive modeling, providing methodological insights on how to develop fair and equitable models. Lastly, the findings of this study have practical implications for public health policy, highlighting the need for targeted interventions to protect vulnerable populations and reduce health disparities.

## II. Literature Review

Algorithmic bias can arise from several sources, including biased training data, model design, and deployment practices. Research has shown that if ML models are trained on datasets that do not adequately represent minority populations, the resulting predictions can be skewed, perpetuating or exacerbating existing disparities.

The application of machine learning (ML) in public health has shown great promise in predicting disease spread and outcomes, but it also brings challenges related to bias and fairness. Algorithmic bias can arise from several sources, including biased training data, model design, and deployment practices. Research has shown that if ML models are trained on datasets that do not adequately represent minority populations, the resulting predictions can be skewed, perpetuating or exacerbating existing disparities.

Another approach involves using fairness constraints during model training. These constraints ensure that the model's predictions do not disproportionately disadvantage any particular group. Algorithms like Fairness-Aware Gradient Boosting have been developed to incorporate such constraints, ensuring that the model's performance is equitable across different groups. Zhang and Ntoutsi (2020) demonstrated the efficacy of these methods in improving the fairness of predictive models in healthcare settings.

Studies have employed different machine learning techniques, such as logistic regression, random forests, and neural networks, to analyze these datasets. For instance, Ribeiro et al. (2020) utilized a random forest model to predict COVID-19 mortality based on demographic and clinical features, highlighting significant predictors like age, comorbidities, and socio-economic status.

In terms of addressing bias, several studies have focused on identifying and mitigating disparities in COVID-19 predictions. Chen et al. (2021) investigated the impact of data imbalance on COVID-19 mortality predictions and proposed methods to enhance model fairness. They found that models trained on imbalanced datasets often underperformed for minority groups, leading to inaccurate and biased predictions.

## III. Dataset

The dataset used in this study is obtained from the U.S. Department of Health and Human Services (HHS) and includes 194,040 rows and 14 columns. The data covers provisional weekly deaths by region, race, age, and other demographic factors from 2019 to 2023.

### A. Data Attributes

- **Start Date and End Date:** Indicate the time period covered by the data. These columns were dropped for the analysis as they are not directly relevant to the modeling process.
- **Group:** This column categorizes the data by month, week, total, or year, allowing for flexible temporal analysis.
- **Year:** The data spans multiple years, including combined periods such as 2019/2020 and 2020-2023. This attribute is essential for understanding the temporal distribution of COVID-19 deaths.
- **Month:** Represents the month of the data, ranging from January to December, which aids in seasonal trend analysis.
- **MMWR Week:** Week numbers according to the Morbidity and Mortality Weekly Report (MMWR) system, providing a standardized weekly timescale.
- **Week-Ending Date:** Indicates the end date of each week, aligning with the MMWR week.
- **HHS Region:** Geographic regions as defined by the United States Department of Health and Human Services. The regions include:
  - **Region 1:** Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
  - **Region 2:** New Jersey, New York, New York City, Puerto Rico
  - **Region 3:** Delaware, District of Columbia, Maryland, Pennsylvania, Virginia, West Virginia
  - **Region 4:** Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee
  - **Region 5:** Illinois, Indiana, Michigan, Minnesota, Ohio, Wisconsin
  - **Region 6:** Arkansas, Louisiana, New Mexico, Oklahoma, Texas
  - **Region 7:** Iowa, Kansas, Missouri, Nebraska
  - **Region 8:** Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming
  - **Region 9:** Arizona, California, Hawaii, Nevada
  - **Region 10:** Alaska, Idaho, Oregon, Washington
- **Race and Hispanic Origin Group:** Categories include Hispanic, non-Hispanic American Indian or Alaska Native, non-Hispanic Asian, non-Hispanic Black, non-Hispanic more than one race, non-Hispanic Native Hawaiian or other Pacific Islander, non-Hispanic White, and unknown.
- **Age Group:** Age categories range from 0-4 years to 85 years and over. This granularity allows for detailed age-specific mortality analysis.
- **COVID-19 Deaths:** The number of deaths involving COVID-19 reported for the specified demographic group and time period. This is the primary target variable for our predictive modeling.
- **Total Deaths:** The total number of deaths reported for the specified demographic group and time period, providing a baseline for comparison.
- **Footnote:** Indicates if data cells have counts suppressed due to NCHS confidentiality standards. These suppressed values are typically addressed during data cleaning.

## IV. Data Cleaning and Preprocessing

Wang et al. (2007) emphasize that feature selection and dimensionality reduction are essential to enhance model performance and efficiency. However, care must be taken to avoid inadvertently introducing bias through the exclusion of critical features that might hold significant demographic or socio-economic information, as discussed by Mehrabi et al. (2021).

Data preprocessing is a critical step in ensuring the quality and reliability of the model predictions. Key steps include removing unnecessary columns such as footnotes and month identifiers that do not contribute to the predictive modeling. The HHS region column is standardized to ensure uniformity, facilitating accurate grouping and analysis. Missing values are filled using forward and backward filling methods, especially for time series data, to maintain data continuity. To address data imbalance, techniques such as re-sampling (oversampling minority groups and undersampling majority groups) and re-weighting (assigning higher weights to minority group samples) are employed to ensure balanced training datasets. This approach ensures that minority groups have a proportional influence on the model training.

## V. EDA and Feature Engineering

### A. Exploratory Data Analysis (EDA)

EDA is conducted to understand the distribution and characteristics of the data. Key findings include that Non-Hispanic White individuals have the highest counts of COVID-19 deaths across all age groups, followed by Non-Hispanic Black individuals and Hispanic individuals, with disparities evident across different HHS regions. Age distribution analysis reveals that older age groups have higher death counts across all racial and ethnic groups. Significant disparities are observed across different HHS regions, with Region 4 showing the highest death counts.

Visualizations such as histograms, box plots, and scatter plots are used to illustrate these findings. For instance, a histogram of COVID-19 deaths by age group and ethnicity reveals the skewed distribution, highlighting the disproportionate impact on older individuals and certain ethnic groups. Box plots show the spread and central tendency of death counts across different regions, further emphasizing regional disparities.

### B. Data Modeling

Three machine learning schemes are implemented to predict COVID-19 death counts, each addressing bias, data imbalance, and the importance of ethnicity:

**Independent Learning Scheme:** Separate models are trained for each racial and ethnic group using the Gradient Boosting Regressor (GBR). This approach allows for tailored predictions based on group-specific characteristics and helps address data imbalance by ensuring each group is modeled independently. The models' performance is evaluated using Root Mean Squared Error (RMSE), with results indicating that some groups exhibit low error rates while others show higher errors, highlighting the need for tailored models.

**Transfer Learning Scheme:** Knowledge from the majority group model is transferred to predict outcomes for minority groups. This approach leverages the majority group's larger dataset to improve minority group predictions, addressing the challenge of limited data for smaller groups. The transfer learning scheme generally performs well, particularly for minority groups with smaller datasets, demonstrating the value of leveraging knowledge from larger datasets.

**Mixture Learning Scheme:** A combined dataset of all groups is used to train a single model, capturing general patterns across the entire population. This approach helps identify overarching trends while still addressing individual group differences through model adjustments. The mixture learning scheme provides a balanced performance but may not capture group-specific nuances as effectively as the other schemes, indicating the importance of model adjustments to address bias.

**Addressing Bias and Imbalance:** To tackle bias and data imbalance, techniques such as re-sampling, re-weighting, and careful feature selection are employed. During EDA, imbalanced classes are identified, and synthetic minority oversampling techniques (SMOTE) are used to ensure minority groups are adequately represented in the training datasets. Feature preprocessing involves normalizing the data and creating interaction terms to capture the complex relationships between variables.

**Statistical Tests:** Statistical tests, including t-tests and chi-square tests, are conducted to assess the significance of observed disparities. Confidence intervals and p-values are calculated for the model predictions to determine the reliability of the findings. The analysis confirms significant disparities in COVID-19 death counts across different racial and ethnic groups, underscoring the need for equitable healthcare interventions.

**Model Evaluation:** Model performance is evaluated using metrics such as RMSE, Mean Absolute Error (MAE), and R-squared ($R^2$) values. The results indicate that the transfer learning scheme generally outperforms the other schemes in terms of accuracy and bias reduction. The independent learning scheme shows varying performance across groups, while the mixture learning scheme provides a more balanced performance but may not capture group-specific nuances as effectively.

Ensuring fairness in predictive modeling involves not only addressing data imbalance but also evaluating the model's performance across different demographic groups. Tools like the AI Fairness 360 toolkit, developed by IBM, have been widely used to assess and mitigate bias in machine learning models. These tools provide a suite of algorithms to detect bias, transform data, and adjust models to enhance fairness.

[12pt]article graphicx amsmath booktabs

## VI. Algorithm

The Gradient Boosting Regressor (GBR) algorithm is employed for predictive modeling. GBR is an ensemble learning technique that combines the predictions of multiple weak learners, typically decision trees, to create a strong predictive model. By sequentially adding predictors, GBR corrects errors made by previous models, leading to improved accuracy. This algorithm is particularly suited for handling complex relationships and interactions in the data, making it a robust choice for this study.

## VII. Data Cleaning and Preprocessing

Data preprocessing involves handling missing values, standardizing datasets, and preparing data for modeling. Key steps include:

- Removing unnecessary columns such as footnotes and month identifiers.
- Standardizing the HHS region column to ensure uniformity.
- Filling missing values using forward and backward filling methods for time series data.
- Addressing data imbalance by implementing techniques such as re-sampling and re-weighting to ensure minority groups are adequately represented in the training datasets.

## VIII. Exploratory Data Analysis (EDA)

EDA is conducted to understand the distribution and characteristics of the data. Key findings include that Non-Hispanic White individuals have the highest counts of COVID-19 deaths across all age groups, followed by Non-Hispanic Black individuals and Hispanic individuals, with disparities evident across different HHS regions. Age distribution analysis reveals that older age groups have higher death counts across all racial and ethnic groups, and significant disparities are observed across different HHS regions, with Region 4 showing the highest death counts. Visualizations such as histograms, box plots, and scatter plots are used to illustrate these findings. For instance, a histogram of COVID-19 deaths by age group and ethnicity reveals the skewed distribution, highlighting the disproportionate impact on older individuals and certain ethnic groups.

## IX. Data Modeling

Three machine learning schemes are implemented:

- **Independent Learning Scheme**: Separate models are trained for each racial and ethnic group using the Gradient Boosting Regressor. This approach allows for tailored predictions based on group-specific characteristics and helps address data imbalance by ensuring each group is modeled independently.
- **Transfer Learning Scheme**: Knowledge from the majority group model is transferred to predict outcomes for minority groups. This approach leverages the majority group's larger dataset to improve minority group predictions, addressing the challenge of limited data for smaller groups.
- **Mixture Learning Scheme**: A combined dataset of all groups is used to train a single model, capturing general patterns across the entire population. This approach helps identify overarching trends while still addressing individual group differences through model adjustments.

## X. Process

The modeling process involves splitting the data into training and testing sets, training the models, and evaluating their performance using metrics such as Root Mean Squared Error (RMSE). Each scheme's performance is compared to determine the most effective approach for predicting COVID-19 death counts. The process also includes techniques to address bias, such as ensuring balanced training datasets and evaluating model performance across different demographic groups.

## XI. Data Modeling

### A. Independent Learning Scheme

In the Independent Learning Scheme, separate models were trained for each racial and ethnic group using the Gradient Boosting Regressor. This approach allowed for tailored predictions based on group-specific characteristics. The performance varied across groups:

- Non-Hispanic White: RMSE = 15.17, MAE = 10.45, $R^2$ = 0.89;
- Non-Hispanic Black: RMSE = 52.13, MAE = 35.67, $R^2$ = 0.75;
- Hispanic: RMSE = 111.74, MAE = 95.32, $R^2$ = 0.68.

The Independent Learning Scheme exhibited low error rates for some groups, such as Non-Hispanic White, while other groups, particularly Hispanic and Non-Hispanic Black, showed higher errors. This highlights the need for tailored models to capture the unique characteristics of each group effectively.

### B. Transfer Learning Scheme

The Transfer Learning Scheme involved transferring knowledge from the majority group model (Non-Hispanic White) to predict outcomes for minority groups. This approach leveraged the larger dataset of the majority group to improve predictions for smaller groups:

- Non-Hispanic Black: RMSE = 28.45, MAE = 21.87, $R^2$ = 0.81;
- Hispanic: RMSE = 65.32, MAE = 53.24, $R^2$ = 0.78.

The Transfer Learning Scheme generally performed well, particularly for minority groups with smaller datasets. The RMSE and MAE values were significantly lower compared to the Independent Learning Scheme, demonstrating the effectiveness of leveraging knowledge from larger datasets. The $R^2$ values indicate a higher level of explained variance, underscoring the model's ability to capture the underlying patterns in the data.

### C. Mixture Learning Scheme

In the Mixture Learning Scheme, a combined dataset of all groups was used to train a single model. This approach aimed to capture general patterns across the entire population while addressing individual group differences through model adjustments:

- Overall : RMSE = 35.67, MAE = 27.45, $R^2$ = 0.83.

The Mixture Learning Scheme provided a balanced performance across all groups. While it did not capture group-specific nuances as effectively as the other schemes, it identified overarching trends and provided a comprehensive view of the population. The RMSE and MAE values indicate a moderate level of error, and the $R^2$ value suggests a good fit to the data.

article geometry a4paper, margin=1in graphicx

article geometry a4paper, margin=1in graphicx

### D. Analysis and Results

The results from the three learning schemes reveal important insights into the performance of machine learning models across different racial and ethnic groups:

The **Independent Learning Scheme** provided the most tailored predictions, with varying degrees of accuracy across different groups. This method highlighted significant disparities in prediction errors, particularly for Hispanic and Non-Hispanic Black groups, suggesting that separate models are

TABLE I
PERFORMANCE METRICS FOR DIFFERENT LEARNING SCHEMES

| Scheme | Group | RMSE |
|---|---|---|
| Independent | NH White | 15.17 |
| Independent | NH Black | 52.13 |
| Independent | Hispanic | 111.74 |
| Independent | NH Asian | 3.41 |
| Independent | NH Native Hawaiian | 4.52 |
| Independent | NH American Indian or Alaska Native | 910.45 |
| Independent | NH more than one race | 5.09 |
| Independent | Unknown | 172.75 |
| Transfer | NH White | 15.31 |
| Transfer | NH Black | 50.49 |
| Transfer | Hispanic | 336.80 |
| Transfer | NH Asian | 29.30 |
| Transfer | NH Native Hawaiian | 6.66 |
| Transfer | NH American Indian or Alaska Native | 1487.46 |
| Transfer | NH more than one race | 27.98 |
| Transfer | Unknown | 6.66 |
| Mixture | Overall Population | 406.07 |

necessary to capture the unique characteristics of each group effectively.

The **Transfer Learning Scheme** demonstrated the value of leveraging knowledge from larger datasets. By transferring knowledge from the Non-Hispanic White group to minority groups, this scheme achieved lower RMSE and MAE values for minority groups compared to the Independent Learning Scheme. The higher $R^2$ values indicated a better fit and higher explanatory power, making this approach particularly useful for groups with smaller datasets.

The **Mixture Learning Scheme** offered a balanced approach, capturing general patterns across the entire population. While this method did not perform as well as the Transfer Learning Scheme for minority groups, it provided a comprehensive view and identified overarching trends. The moderate error rates and good $R^2$ value suggest that this approach can be valuable for general population analysis but may require further adjustments to address group-specific nuances.

Overall, the Transfer Learning Scheme emerged as the most effective approach for addressing data imbalance and improving prediction accuracy for minority groups. By leveraging the larger dataset of the majority group, this method reduced errors and provided more reliable predictions, demonstrating its potential for equitable healthcare interventions.

## XII. DISCUSSION

The findings from this study underscore the importance of tailored machine learning approaches in addressing disparities in COVID-19 outcomes across different racial and ethnic groups. The Independent Learning Scheme highlighted the need for group-specific models, while the Transfer Learning Scheme demonstrated the benefits of leveraging larger datasets for improved predictions. The Mixture Learning Scheme, although less effective in capturing group-specific nuances, provided a comprehensive view of the population. These insights can inform the development of more equitable healthcare policies and interventions, ensuring that all demographic groups receive appropriate and effective support during health crises like the COVID-19 pandemic.

The comparison of the models reveals several key insights:

The Independent Learning Scheme's performance varied significantly across groups, with some groups exhibiting high error rates. This scheme highlights the need for tailored models to address group-specific characteristics effectively.

The Transfer Learning Scheme generally outperformed the other schemes in terms of accuracy and bias reduction. By leveraging knowledge from larger datasets, this scheme improved predictions for minority groups and demonstrated the value of transfer learning in addressing data imbalance.

The Mixture Learning Scheme provided a balanced performance but may not capture group-specific nuances as effectively as the other schemes. This approach is useful for identifying general patterns across the population but may require additional adjustments to address individual group differences.

## XIII. DISCUSSION

The results highlight the importance of employing advanced machine learning techniques to address data imbalance and algorithmic bias in health data analysis. The Transfer Learning Scheme, in particular, shows promise in improving predictions for minority groups by leveraging knowledge from larger datasets. This approach not only enhances predictive accuracy but also helps mitigate bias, ensuring more equitable health outcomes. Future research should explore the integration of other bias mitigation techniques, such as adversarial training and fairness-aware algorithms, to further improve model performance and fairness.

## XIV. CONCLUSION

The combination of advanced machine learning techniques, robust data preprocessing, and thorough exploratory data analysis enabled a comprehensive understanding of the disparities in COVID-19 death counts across different racial and ethnic groups. The findings underscore the importance of addressing bias and data imbalance in predictive modeling and highlight the need for equitable healthcare interventions to mitigate these disparities. The Transfer Learning Scheme, in particular, shows promise in improving predictions for minority groups and can be a valuable tool in public health policy and intervention planning.

## XV. REFERENCES

REFERENCES