Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 256 (2025) 1081–1091

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2024

Assessing the Evolution of COVID-19 Risk Factors Post-Vaccine Rollout: A Comparative Study of Counties in States with High and Low Racial Diversity

Mingming Peng[a]*, Hongmei Chi[b]*, Askal Ayalew Ali[a]

[a]*Economic, Social and Administrative Pharmacy (ESAP), College of Pharmacy and Pharmaceutical Sciences, Florida A&M University, Tallahassee, FL, USA*

[b]*Department of Computer&Information Sciences, Florida A&M University, Tallahassee, FL, USA*

Abstract

This study investigates the changes in COVID-19 risk factors following the widespread distribution of vaccines, focusing on a comparative analysis between counties in states with varying levels of racial diversity. We analyzed data from counties in states with high racial diversity versus those with low racial diversity to identify disparities and trends in risk factors related to COVID-19. The research aims to understand how vaccine rollout has influenced these risk factors differently across diverse demographic settings. By examining variations in infection rates, vaccination coverage, and public health outcomes, the study highlights the impact of racial diversity on the effectiveness of COVID-19 mitigation strategies. It offers insights for tailored public health interventions. The findings contribute to a deeper understanding of how demographic factors shape the evolution of COVID-19 risk profiles in a post-vaccine era. This research lies in its potential to inform targeted interventions and policies to mitigate the ongoing impact of COVID-19 on vulnerable populations. By identifying the key risk factors in the post-vaccine era, public health authorities and policymakers can develop strategies to address the specific needs of communities most at risk, ultimately promoting health equity and resilience in future public health crises.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: mingming1.peng@famu.edu

10.1016/j.procs.2025.02.215

Keywords: COVID-19 outcomes, vulnerable populations, pandemic response policies, racial diversity, machine learning, public health crises

* Corresponding author. Tel.: 01 850-599-3050; fax:01 8505-99-3221
E-mail address: hongmei.chi@famu.edu

1. Introduction

The COVID-19 pandemic has profoundly impacted the United States and the world. Although vaccines have significantly changed the course of the pandemic, COVID-19 is still a significant public health challenge, even after their widespread use[1]. Data reveals that the United States experienced peak weekly COVID-19 hospitalizations and death rates in January 2022, one year after the vaccine rollout began[2]. It also reveals that COVID-19 community levels vary significantly across regions [3]. These disparities highlight the need to understand the social-demographic factors that keep affecting COVID-19 spread and severity in the post-vaccine era. This study aims to identify and analyze these evolving risk factors, with a focus on how they influence COVID-19 community levels across diverse regions of the United States from April 2021 to April 2022.

By the end of March 2024, COVID-19 had already killed 1,188,430 people in the United States. Additionally, the weekly death cases remained above 600, with the test-positive rate still around 3% [2]. This indicates that the disease is still present among us, and further genetic mutations may be possible [4], [5]. Therefore, it is necessary to re-examine the socioeconomic risk factors associated with COVID-19 in the post-vaccine era to address potential future risks.

The United States exhibits significant disparities among states in terms of population diversity, socioeconomic factors, and pandemic control policies[6], [7]. These diversities provide an excellent comparative perspective for study. Previous research has shown that racial and ethnic minorities have been disproportionately affected by the pandemic, with higher rates of infection, hospitalization, and death compared to white populations [8], [9]. Socioeconomic factors such as income, education, and employment have also been identified as key determinants of COVID-19 outcomes [10]. However, the extent to which these factors influence COVID-19 cases and deaths in the post-vaccine era remains unclear. Among studies exploring disease risk factors, machine learning has emerged as an ideal approach[11], [12], [13]. This is due to its ability to handle complex, high-dimensional data and nonlinear relationships, complementing traditional statistical methods.

Therefore, this study aims to explore the key risk factors influencing COVID-19 cases and deaths in different regions of the United States in the post-vaccine era by constructing models and comparing counties in the states with the highest and lowest racial diversity at four time points after the widespread use of vaccines in April 2021. By focusing on racial diversity, this study seeks to capture the complex interplay of socioeconomic factors and pandemic response policies that may contribute to disparities in COVID-19 outcomes.

The significance of this research lies in its potential to inform targeted interventions and policies to mitigate the ongoing impact of COVID-19 on vulnerable populations. By identifying the key risk factors in the post-vaccine era, public health authorities and policymakers can develop strategies to address the specific needs of communities most at risk, ultimately promoting health equity and resilience in the face of future public health crises.

This paper is organized as follows: Section 2 reviews the literature on COVID-19 risk factors and the various methodologies used in previous studies. Section 3 details the methodology employed in this study, including data collection, preprocessing, and model construction. Section 4 presents the analysis results, including model performance, feature importance analysis, temporal analysis, and comparisons between high and low racial diversity areas. Section 5 discusses the implications, interpretation, and limitations of the study. Finally, Section 6 concludes with key findings and future research directions.

2. Literature Review

This review examines how COVID-19 sociodemographic risk factors changed post-vaccine rollout, summarizes the literature, identifies key risk factors, and analyzes their study methods.

Some early studies explored the relationship between COVID-19-related outcomes and social determinants. Several studies have found associations between race or ethnicity, socioeconomic factors, and increased COVID-19 incidence and mortality rates. For instance, several studies and CDC consistently show that racial and ethnic minorities are disproportionately affected by COVID-19 [14], [15], [16], [17]. The identified SDoH indicators associated with COVID-19 outcomes include race/ethnicity, poverty, median income level, housing density and insecurity, healthcare access, occupation, transportation patterns, education, air quality, food insecurity, and old age. Another study using NHANES data also found that COVID-19 mortality rates were significantly higher among individuals of nonwhite race/ethnicity, those with incomes below the median, and those with less than a high school education[18]. A cross-sectional study analyzed 4.29 million COVID-19 cases and 147,074 deaths in the U.S. in 2020 and explored the relationship between county-level sociodemographic risk factors and COVID-19 incidence and mortality rates [19]. They found that Social Vulnerability Index (SVI), socioeconomic status, racial/ethnic minority status, household composition, and environmental factors were significantly associated with COVID-19 incidence and mortality rates.

What's more, some studies focus on exploring the impact of controlling policies on COVID-19 outcomes. They use methods such as cost-effectiveness analysis, microsimulation models, and data envelopment analysis to evaluate the effectiveness of different policies (e.g., lockdown, testing) in various population densities, sizes, and stages of development. Their finding revealed the importance of controlling policies in influencing COVID-19-related outcomes[20], [21].

In the later stages of the pandemic, several studies explored the influence of social factors on COVID-19-related health outcomes in the post-vaccine era. Kerschbaumer et al. conducted a study in Austria and revealed that factors such as poverty, job insecurity, limited social participation, and digital exclusion posed direct threats to COVID-19-related individual health[22].

Various methods have been employed to explore risk factors for COVID-19-related health outcomes. A study compared multivariate logistic regression and machine learning methods to predict severe COVID-19 in hospitalized children with Omicron variant infection. They found that machine-learning models demonstrated higher accuracy[23].

Another study compared six geostatistical models with two machine learning methods in exploring socioexposomic associations with COVID-19 outcomes across New Jersey[24].

While previous studies have provided valuable insights into the sociodemographic factors influencing COVID-19 outcomes, several areas warrant further investigation. First, most of the existing research focuses on the pre-vaccine phase of the pandemic, and there is a need to explore how these risk factors have evolved in the post-vaccine era. Second, although some studies have employed machine learning methods to predict COVID-19 outcomes, there is still potential to leverage a broader range of algorithms and compare their performance in capturing complex, nonlinear relationships between sociodemographic factors and COVID-19 community levels. Third, while prior research has identified various risk factors, there is a need to analyze how the importance of these factors changes over time and under different conditions, such as vaccine coverage and virus variants. Fourth, previous studies have not adequately explored whether there are differences in risk factors between areas with high and low racial diversity.

3. Methodology

This section details the method used to examine the impact of sociodemographic factors on COVID-19 community levels in the U.S. post-vaccination. Data were collected from multiple sources between April 2021 and April 2022. Three machine learning models—XGBoost, Random Forest, and LightGBM—were implemented for predictive analysis, with feature importance assessed using SHAP values. Sub-analyses were performed to examine temporal changes in risk factors and to compare areas of high and low racial diversity. Fig. 1 illustrates the study steps.

3.1. Data Collection and Preprocessing

The data for this study was integrated from multiple sources. COVID-19 health outcomes were obtained from the CDC's COVID-19 Community Profile Report[3]. County-level social-demographic information came from County Health Rankings & Roadmaps [25]. Policy data was sourced from the COVID-19 US State Policy Database and Ballotpedia.org [26], [27]. Unemployment rates were from the US Bureau of Labor Statistics and state-level Racial and Ethnic Diversity index data from the US Census[28].

Features from the CDC's COVID-19 Community Profile Report include vaccination rates, percentages of uninsured and impoverished populations, demographic breakdowns, Social Vulnerability Index (SVI), Community Vulnerability Index (CCVI), healthcare metrics, COVID-19 cases and deaths per 100,000, and county classifications. The features included in the County Health Rankings are socioeconomic indicators, health outcomes, household characteristics, broadband access, and demographic details. Policy data from databases and Ballotpedia.org cover public mask mandates, travel/gathering bans, and economic support policies.

The data preprocessing focused on handling missing values, converting character variables to categorical variables, and ensuring data consistency. Missing values were filled with zeros, medians, or based on rural-urban classifications, depending on the feature. All character-type variables were converted to numeric variables for efficient training. A time feature was added to record the time point of each data entry.

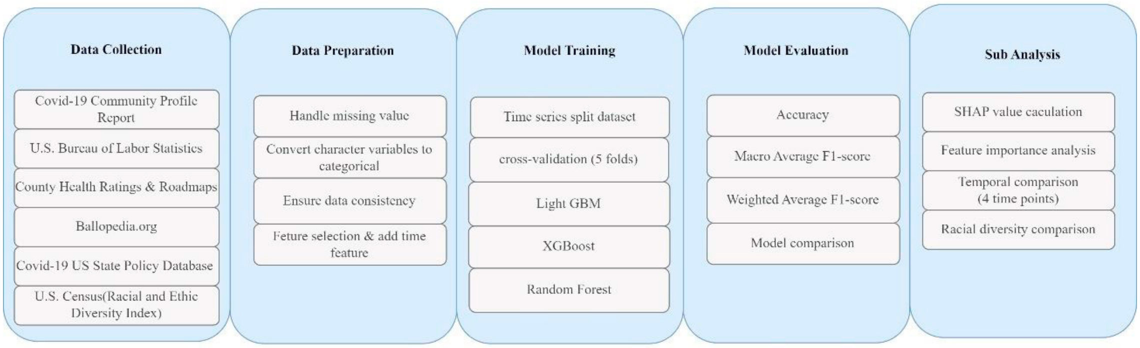

Fig. 1. Research Methodology Diagram.

3.2. Model Training

In order to examine the impact of sociodemographic factors on community infection levels on post-COVID-19 vaccine rollout, this study selects four time points between April 2021 and April 2022. These time points were selected to reflect different stages of the pandemic post-vaccination, representing various phases of vaccine rollout, emergence of new variants (such as Delta and Omicron), and varying levels of disease prevalence. The chosen time points are April 2021 (when priority groups in all states had completed vaccination and mass vaccination for adults began), September 2021 (peak of the Delta variant), January 2022 (peak of the Omicron variant), and April 2022 (one year after mass vaccination began, with no prevalent high-risk variants).

Three machine-learning models were developed to predict COVID-19 community levels using multiple sociodemographic features to ensure a robust analysis. The algorithms used to train these models were XGBoost, Random Forest, and Light GBM. XGBoost was chosen for its high performance and ability to handle complex interactions between features. Random Forest was selected for its capability to mitigate overfitting and provide reliable feature importance rankings. Light GBM was included due to its efficiency with large datasets and effectiveness in capturing nonlinear relationships.

Data was split into training (80%) and testing (20%) sets using time series split cross-validation with five consecutive folds. Model performance was evaluated using Accuracy, Weighted Average F1-score, and Macro Average F1-score.

The study utilized feature importance scores provided by the models and SHAP (SHapley Additive exPlanations) values to identify the most crucial risk factors. SHAP values were chosen as the primary method for interpreting model results due to their ability to provide a more granular and personalized understanding of feature importance.

After establishing and evaluating the performances of those predictive models, sub-analyses were conducted for each of the four time points to gain insights into how risk factors evolved. The following comparisons were performed: (i)Temporal comparison: Analyzing changes in the importance of risk factors across different time points. (ii) Racial diversity comparison: Comparing risk factors in counties with high racial diversity to those with low racial diversity. (iii) Model comparison: Evaluating the performance and identifying risk factors across the three models.

This approach allows for informing targeted public health interventions and policies to mitigate the ongoing impact of the pandemic, taking into account the changing dynamics of the disease over time and across different population groups.

4. Preliminary Results

This section presents the analysis results. It compares the performance of three machine learning models, identifies key risk factors using SHAP values, examines their evolution from April 2021 to April 2022, and compares feature importance between high and low racial diversity areas.

4.1. Model Performance

To evaluate the performance of three machine learning models, we compared their Accuracy, Macro Average F1-score, and Weighted Average F1-score. LightGBM showed the best overall performance, with the highest accuracy of 94%, a macro average F1-score of 0.92, and a weighted average F1-score of 0.95. This indicates LightGBM's effectiveness in capturing the complex relationships between risk factors and COVID-19 community levels. Random Forest also performed well, achieving an accuracy of 92% a macro average F1-score of 0.88, and a weighted average F1-score of 0.92. This suggests Random Forest handled potential class imbalances effectively. XGBoost, while less effective than the other two, still reached an accuracy of 86% and a weighted average F1-score of 0.86. Its lower macro average F1-score of 0.78 indicates some difficulty with certain classes.

The macro average F1-scores show that LightGBM (0.92) performed more consistently across all classes compared to Random Forest (0.88) and XGBoost (0.78). The weighted average F1-scores further confirm LightGBM's superior performance, closely followed by Random Forest. Table 1 summarizes these results.

Table 1. Model Performance Comparison.

Model	Accuracy	Macro Avg F1-score	Weighted Avg F1-score
XGBoost	0.86	0.78	0.86
Random Forest	0.92	0.88	0.92
LightGBM	0.94	0.92	0.95

4.2. Feature Importance Analysis

We used SHAP values calculated with TreeExplainer for each model to identify the most important risk factors. Based on the SHAP value analysis, we identified the following key features for predicting COVID-19 community levels over one year:

Population and Healthcare Infrastructure: "Population" and "Total # of hospital CCNs" consistently appear as top features across all models, suggesting that the size of the community and its healthcare capacity are crucial factors.

Vaccination Rates: "People who are fully vaccinated as % of the total population" is a significant feature in all models, highlighting the importance of vaccination in determining community COVID-19 levels.

Socioeconomic Factors: (i) "Unemployment" appears as an important feature in LightGBM and Random Forest models, indicating that economic conditions play a role in community transmission. (ii) "% Broadband Access" (LightGBM) and "% Rural" (Random Forest) suggest that connectivity and urbanization levels are relevant factors.

Environmental Factors: "Average Daily PM2.5" is consistently important across models, implying that air quality may influence COVID-19 community levels.

Behavioral and Policy Factors: (i) "travel/gatherings ban" (Random Forest and XGBoost) indicates that policy interventions significantly predict community levels. (ii) "% Long Commute - Drives Alone" (LightGBM) suggests that commuting patterns may affect transmission.

Demographic Factors: Racial and ethnic factors (e.g., "% Hispanic", "% Non-Hispanic Asian", "% Non-Hispanic Black") appear in the top features, indicating that demographic factors play a role in community COVID-19 levels.

Living Conditions: "Overcrowding" (LightGBM) and "Food Insecure" (XGBoost) suggest that living conditions and food security are relevant factors.

Education: "High School Graduation Rate" (XGBoost) implies that education levels may influence community COVID-19 levels. Fig. 2 shows the top 10 SHAP values identified by each model.

4.3. Time Point Analysis

To know how social-demographic factors for COVID-19 community levels evolved, sub-analyses were conducted for four key time points: April 2021, September 2021, January 2022, and April 2022.

April 2021 - Initial stage of mass vaccination: The LightGBM model identified racial diversity (importance 0.760) as the most significant predictor, followed by population size (0.421) and unemployment rate (0.127). The XGBoost model identified population size (0.091) and vaccination rate (0.029) as the most important features, also emphasizing average daily PM2.5 (0.022) and overcrowding (0.020). The Random Forest model emphasized vaccination rate (0.032) and travel/gathering bans (0.030) as important factors.

September 2021 - Peak of Delta variant: In LightGBM, racial diversity (0.622) and population size (0.364) remained important, but the unemployment rate (0.149) and average daily PM2.5 (0.130) increased in importance. In the XGBoost model, population size (0.058) remained the most important, with the unemployment rate (0.023) rising in significance. The Random Forest model showed an increase in the importance of the number of hospitals (0.025).

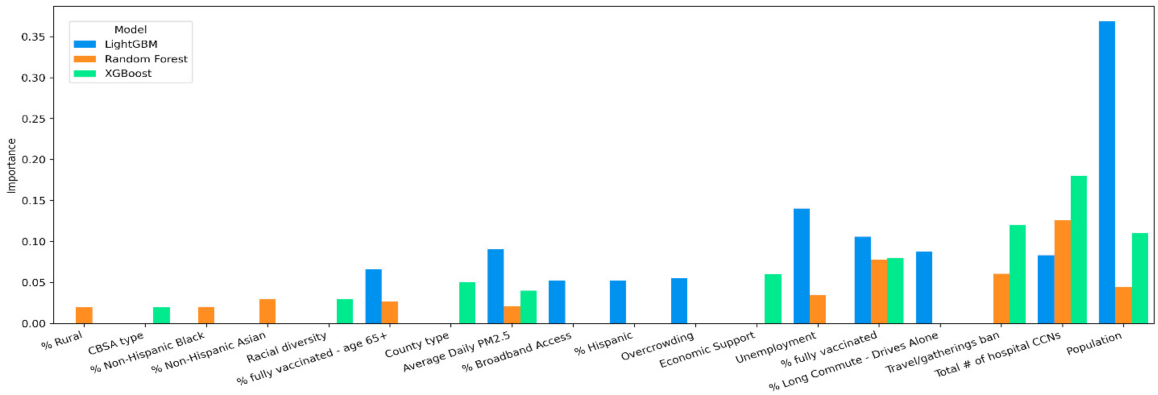

Fig. 2. Comparative importance of key features across models.

January 2022 - Peak of Omicron variant: In the LightGBM model, the importance of racial diversity increased significantly (1.189), while population size (0.324) and unemployment rate (0.141) remained important. The XGBoost model results were similar to those in September, with a slight increase in the importance of hospital numbers (0.017). In the Random Forest model, the number of hospitals (0.026) remained the most important factor.

April 2022 - One year after mass vaccination began: In the LightGBM model, racial diversity (1.146) and population size (0.365) remained the most important, but the number of hospitals (0.184) and vaccination rate (0.169) increased significantly in importance. The XGBoost model showed population size (0.087) and vaccination rate (0.084) as equally important, with hospital numbers (0.047) also increasing in importance. In the Random Forest model, the importance of hospital numbers (0.061) increased significantly. Fig. 3 shows the top 10 critical features of the three models at four time points.

4.3 Feature Importance Analysis in High and Low Racial Diversity Areas

Overall Feature Importance Comparison

Across all models, several features consistently ranked as highly important in both high and low-diversity areas. Table 2 shows the top five common features and their importance scores.

Table 2. Top common features in high and low racial diversity areas.

Feature	High Diversity Area	Low Diversity Area
Population	0.382989	0.347388
Unemployment	0.144485	0.134059
fully vaccinated % of total Population	0.115969	0.091122
Average Daily PM2.5	0.091148	0.090727
Total # of hospital CCNs	0.082783	0.083762

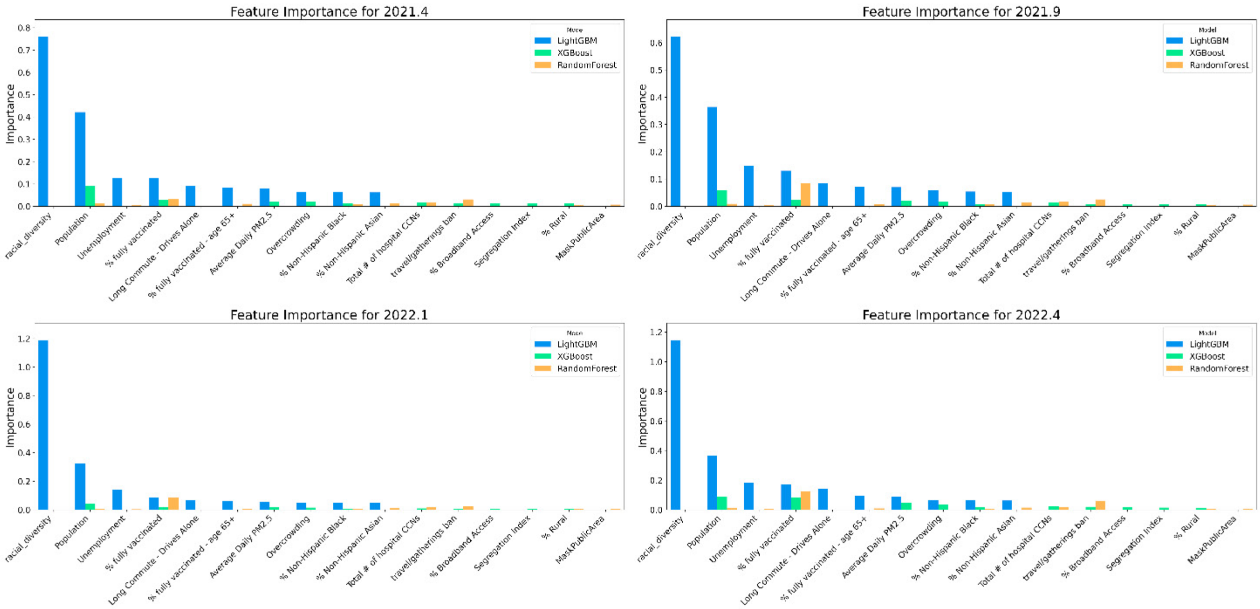

Fig. 3. Comparative feature importance across models and time points.

Distinctive Features in High and Low-Diversity Areas

In high-diversity areas, certain features showed higher importance across models. The LightGBM model identified "% Long Commute - Drives Alone" (importance: 0.088829), "Overcrowding" (0.058337), and "% Broadband Access" (0.056593) as particularly important. The Random Forest model emphasized, "People who are fully vaccinated as % of total population" (0.021174) and "travel/gatherings ban" (0.016036) in high diversity areas. XGBoost identified "CCVI score" (0.010542) and "High School Graduation Rate" (0.008112) as more important in these areas.

Low-diversity areas exhibited a different set of important features. The LightGBM model found "% Hispanic" (0.053640), "% Non-Hispanic Black" (0.051400), and "% Uninsured" (0.050059) to be more significant in these areas. The Random Forest model identified "% Non-Hispanic Asian" (0.007690) and "% Rural" (importance not available for high diversity areas) as more important in low diversity regions. XGBoost identified the "Social Association Rate" (0.006931) as having increased importance in low diversity areas.

Across models, some features showed consistent differences in importance between high- and low-diversity areas. For instance, the Random Forest model found "Total # of hospital CCNs" to be slightly more important in low diversity areas (0.033273) compared to high diversity areas (0.030421). Similarly, XGBoost identified "Population" as more important in high diversity areas (0.072216) compared to low diversity areas (0.064956), a trend consistent across all models but with varying magnitudes.

These differences in feature importance between high and low-diversity areas, as identified by different models, suggest that the factors influencing COVID-19 community levels may vary based on the racial diversity of the area. Fig. 4 shows the comparison of feature importance between high and low racial diversity counties for the three models.

Other Key Factors

Environmental, healthcare, infrastructure, and socioeconomic factors showed varying importance:

- Average Daily PM2.5 was consistently important across all models and diversity levels (importance range: 0.019970 - 0.091148).
- % Broadband Access showed higher importance in high diversity areas (0.056593) compared to low diversity areas (0.045320).
- XGBoost identified the CCVI score as important in both high (0.010542) and low (0.009963) diversity areas.

Policy-related factors also showed varying importance in high- and low-diverse areas. The Random Forest model found "travel/gatherings ban" to be more important in high diversity areas (0.016036) compared to low diversity areas (0.013848). Similarly, XGBoost identified "Economic Support" as a more significant factor in high-diversity regions.

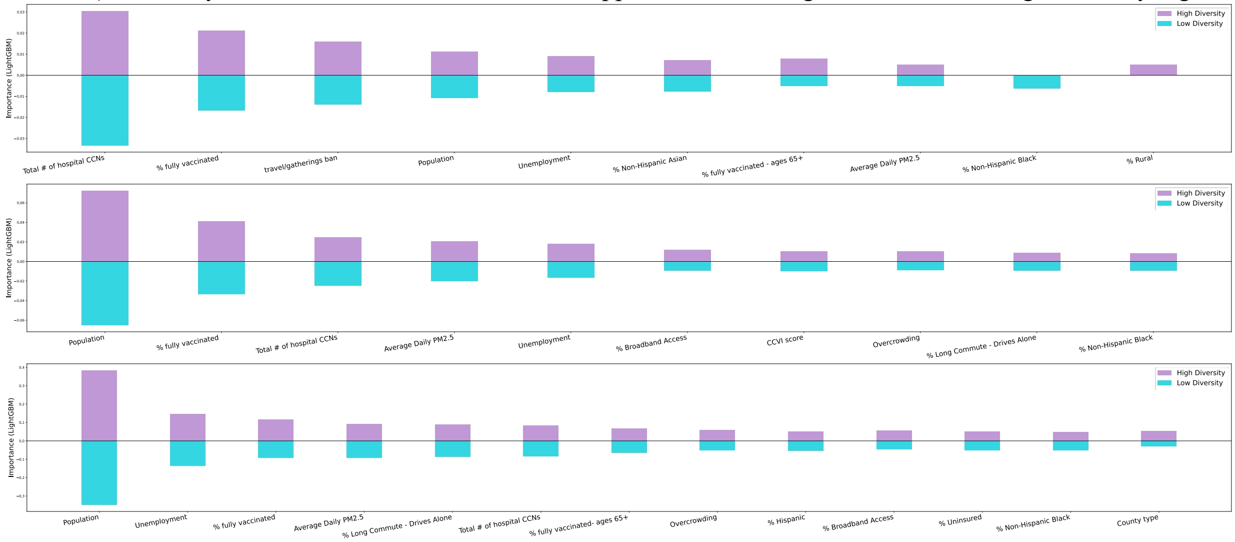

Fig. 4. Feature Importance Comparison Across Models: High vs. Low Racial Diversity.

5. Discussion

This study employed three machine learning models - LightGBM, Random Forest, and XGBoost - to explore the sociodemographic factors influencing COVID-19 community levels in the year following the widespread vaccine rollout in the United States. By analyzing data from four key time points, the research revealed the evolution of risk factors over time and differences between areas of high and low racial diversity.

5.1. Changes in Risk Factors Over Time

The study revealed significant changes in key factors influencing COVID-19 community levels over time:

- Population Size and Healthcare Infrastructure: All models found population size and the number of hospitals remained key factors throughout the study period, highlighting their impact on COVID-19 community levels.
- Vaccination Rate: LightGBM and XGBoost showed that their importance generally increased, peaking in April 2022. Random Forest did not display this increasing trend, but it consistently ranked it among the top 10 features. This indicates its persistent influence on COVID-19 community levels despite changing relative importance.
- Socioeconomic Factors: All models consistently identified the unemployment rate as an important factor, maintaining relatively stable importance throughout the study period. This suggests a persistent influence of economic conditions on pandemic outcomes.
- Environmental Factors: Average daily PM2.5 was significant across multiple time points, especially in XGBoost and LightGBM models. This aligns with previous studies [19], emphasizing air quality's potential influence on COVID-19 transmission.

- Racial Diversity: In the best-performing model, racial diversity remained the top predictor, peaking in January 2022, consistent with earlier research [14], [16].
- Policy Factors: The importance of travel/gathering bans decreased over time, potentially differing from previous findings on control policy importance[20]. This may reflect varying policy effectiveness across pandemic stages.

5.2. Differences Between High and Low Racial Diversity Areas

The study found significant differences in factors affecting COVID-19 community levels between high and low racial diversity areas:

- Common Important Factors: Across all models, population size, unemployment rate, vaccination rate, PM2.5, and number of hospitals were important in both types of areas, though to varying degrees.
- Factors Unique to High Diversity Areas: LightGBM and Random Forest models identified factors such as long commute times, overcrowding, and broadband access as more critical in high-diversity areas. XGBoost identified the importance of the CCVI score in these areas. This finding aligns with previous research indicating that during the COVID-19 pandemic, lack of internet access significantly impacted the ability of minority communities to access health information and medical services[29].
- Factors Unique to Low Diversity Areas: LightGBM and Random Forest models found factors like the percentage of Hispanic and non-Hispanic Black population and the percentage of the uninsured population more significant in low diversity areas. XGBoost emphasized the importance of the Social Association Rate in these areas.
- Policy Impact: Random Forest and XGBoost models suggested that travel/gathering bans and economic support policies were more critical in high-diversity areas, possibly reflecting different sensitivities to policy interventions.

5.3. Model Insights and Study Limitations

Regarding model selection, machine learning helps handle high-dimensional data and make predictions. However, it's less interpretable than traditional models. In this study, while the models often agreed on the importance of certain factors, there were also significant differences. For example, the models showed varying insights into the importance of features such as racial diversity and high school graduation rates. However, these differences also demonstrated the value of using multiple models in complex predictive tasks, as each model may capture different aspects of the underlying data relationships.

Regarding the research features included, while we innovatively included policy factors, their effectiveness was difficult to assess fully. Because many areas relaxed or stopped control measures later in our study period, unmeasured or unidentified factors may influence COVID-19 community levels, such as public adherence to guidelines. These limitations revealed the complexity of evaluating the long-term impacts of policies and capturing all relevant variables in COVID-19 research.

Our study’s focus on high and low racial diversity states over one year may limit the generalizability of findings to broader areas and longer timeframes. County-level data might overlook finer local differences, suggesting that future studies could benefit from more granular data, such as neighborhood or census tract information, to reveal more detailed COVID-19 spread patterns and risk factors.

6. Conclusions and Future Work

This study used multiple machine learning models to analyze COVID-19 risk factors in the U.S. during the first year of vaccine rollout. Particularly, it focuses on counties with varying racial diversity. Findings reveal key sociodemographic factors influencing COVID-19 community levels and their changes over time. Population size and healthcare infrastructure remained significant predictors while vaccination rates' importance increased. The unemployment rate remained stable, and environmental factors like PM2.5 levels were crucial. Racial diversity emerged as a significant predictor, especially in the LightGBM model, while policy factors decreased in importance over time. The study found risk factor differences between high and low-racial diversity areas, emphasizing the need for tailored interventions.

Future research could include more states and extended periods, improving results' generality and understanding of long-term effects. Studies can also incorporate additional variables not measured or identified in this study, such as public adherence to guidelines, local healthcare capacity, and viral variants. Future studies could also utilize other machine learning techniques to capture complex nonlinear relationships and enhance our understanding of COVID-19 risk factors, assess the effectiveness of COVID-19 policies [32] and their impact on minority groups[33], and provide valuable insights for future decision-making.

Acknowledgments:

M. Peng's work is supported by the Title III program at the FAMU College of Pharmacy and Pharmaceutical Sciences, Institute of Public Health (CoPPS, IPH) for essential departmental support. H. Chi's work is partially supported by the National Science Foundation under Grant No. 2131164. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] J. Li and P. Giabbanelli, "Returning to a Normal Life via COVID-19 Vaccines in the United States: A Large-scale Agent-Based Simulation Study," *JMIR Med Inform*, vol. 9, no. 4, p. e27419, Apr. 2021, doi: 10.2196/27419.
[2] [CDC, "COVID Data Tracker."
[3] CDC, "COVID-19 Community Profile Report," Jul. 2023.
[4] P. Satapathy *et al.*, "Global spread of COVID-19's JN.1 variant: Implications and public health responses," *New Microbes New Infect*, vol. 57, p. 101225, Mar. 2024, doi: 10.1016/j.nmni.2024.101225.
[5] D. Wang *et al.*, "EFCAB4B (CRACR2A/Rab46) Genetic Variants Associated with COVID-19 Fatality," *COVID*, vol. 4, no. 7, pp. 1087–1099, Jul. 2024, doi: 10.3390/covid4070075.
[6] K. N. C. Duong *et al.*, "Disparities in COVID-19 related outcomes in the United States by race and ethnicity pre-vaccination era: an umbrella review of meta-analyses," *Front Public Health*, vol. 11, Sep. 2023, doi: 10.3389/fpubh.2023.1206988.
[7] D. Hawkins, "Differential occupational risk for COVID-19 and other infection exposure according to race and ethnicity," *Am J Ind Med*, vol. 63, no. 9, pp. 817–820, Sep. 2020, doi: 10.1002/ajim.23145.
[8] D. B. G. Tai, A. Shah, C. A. Doubeni, I. G. Sia, and M. L. Wieland, "The Disproportionate Impact of COVID-19 on Racial and Ethnic Minorities in the United States," *Clinical Infectious Diseases*, vol. 72, no. 4, pp. 703–706, Feb. 2021, doi: 10.1093/cid/ciaa815.
[9] T. M. Holden, M. A. Simon, D. T. Arnold, V. Halloway, and J. Gerardin, "Structural racism and COVID-19 response: higher risk of exposure drives disparate COVID-19 deaths among Black and Hispanic/Latinx residents of Illinois, USA," *BMC Public Health*, vol. 22, no. 1, p. 312, Dec. 2022, doi: 10.1186/s12889-022-12698-9.
[10] V. Abedi *et al.*, "Racial, Economic, and Health Inequality and COVID-19 Infection in the United States," *J Racial Ethn Health Disparities*, vol. 8, no. 3, pp. 732–742, Jun. 2021, doi: 10.1007/s40615-020-00833-4.
[11] M. Seyedtabib, R. Najafi-Vosough, and N. Kamyari, "The predictive power of data: machine learning analysis for Covid-19 mortality based on personal, clinical, preclinical, and laboratory variables in a case–control study," *BMC Infect Dis*, vol. 24, no. 1, p. 411, Apr. 2024, doi: 10.1186/s12879-024-09298-w.
[12] G. Drouard *et al.*, "Exploring machine learning strategies for predicting cardiovascular disease risk factors from multi-omic data," *BMC Med Inform Decis Mak*, vol. 24, no. 1, p. 116, May 2024, doi: 10.1186/s12911-024-02521-3.
[13] M. Naseem, R. Akhund, H. Arshad, and M. T. Ibrahim, "Exploring the Potential of Artificial Intelligence and Machine Learning to Combat COVID-19 and Existing Opportunities for LMIC: A Scoping Review," *J Prim Care Community Health*, vol. 11, p. 215013272096363, Jan. 2020, doi: 10.1177/2150132720963634.
[14] T. L. Upshaw, C. Brown, R. Smith, M. Perri, C. Ziegler, and A. D. Pinto, "Social determinants of COVID-19 incidence and outcomes: A rapid review," *PLoS One*, vol. 16, no. 3, p. e0248336, Mar. 2021, doi: 10.1371/journal.pone.0248336.
[15] [15] M. 2; L. M. R. P. F. B. A. M. P. S. P. Z. L. M. P. P. S. P. 2; J. P. C. P. J. D. M. P. 2; B. N. B. M. T. M. D. P. Jeremy A.W. Gold, "Race, Ethnicity, and Age Trends in Persons Who Died from COVID-19 — United States, May–August 2020," 2020.
[16] A. Khanijahani, S. Iezadi, K. Gholipour, S. Azami-Aghdash, and D. Naghibi, "A systematic review of racial/ethnic and socioeconomic disparities in COVID-19," *Int J Equity Health*, vol. 20, no. 1, p. 248, Nov. 2021, doi: 10.1186/s12939-021-01582-4.
[17] W. S. Brakefield, O. A. Olusanya, B. White, and A. Shaban-Nejad, "Social Determinants and Indicators of COVID-19 Among Marginalized Communities: A Scientific Review and Call to Action for Pandemic Response and Recovery," *Disaster Med Public Health Prep*, vol. 17, p. e193, May 2023, doi: 10.1017/dmp.2022.104.
[18] B. Seligman, M. Ferranna, and D. E. Bloom, "Social determinants of mortality from COVID-19: A simulation study using NHANES," *PLoS Med*, vol. 18, no. 1, p. e1003490, Jan. 2021, doi: 10.1371/journal.pmed.1003490.
[19] M. Karmakar, P. M. Lantz, and R. Tipirneni, "Association of Social and Demographic Factors With COVID-19 Incidence and Death Rates in the US," *JAMA Netw Open*, vol. 4, no. 1, p. e2036462, Jan. 2021, doi: 10.1001/jamanetworkopen.2020.36462.
[20] M. Klumpp, D. Loske, and S. Bicciato, "COVID-19 health policy evaluation: integrating health and economic perspectives with a data envelopment analysis approach," *The European Journal of Health Economics*, vol. 23, no. 8, pp. 1263–1285, Nov. 2022, doi: 10.1007/s10198-021-01425-7.
[21] A. Boloori and S. Saghafian, "Health and Economic Impacts of Lockdown Policies in the Early Stage of COVID-19 in the United States," *Serv Sci*, vol. 15, no. 3, pp. 188–211, 2023, doi: 10.1287/serv.2023.0321.

[22] L. Kerschbaumer, L. Crossett, M. Holaus, and U. Costa, “COVID-19 and health inequalities: The impact of social determinants of health on individuals affected by poverty,” *Health Policy Technol*, vol. 13, no. 1, p. 100803, Mar. 2024, doi: 10.1016/j.hlpt.2023.100803.
[23] P. Liu *et al.*, “Machine learning versus multivariate logistic regression for predicting severe COVID-19 in hospitalized children with Omicron variant infection,” *J Med Virol*, vol. 96, no. 2, Feb. 2024, doi: 10.1002/jmv.29447.
[24] X. Ren, Z. Mi, and P. G. Georgopoulos, “Socioexposomics of COVID-19 across New Jersey: a comparison of geostatistical and machine learning approaches,” *J Expo Sci Environ Epidemiol*, Feb. 2023, doi: 10.1038/s41370-023-00518-0.
[25] K. L. Kingery, “County health rankings & roadmaps,” *Journal of Youth Development*, pp. 259–263, Sep. 2018.
[26] Ballotpedia, “Documenting America’s Path to Recovery,” https://ballotpedia.org/Documenting_America%27s_Path_to_Recovery.
[27] J. et al. Raifman, “COVID-19 US state policy database,” 2020, *https://docs.google.com/spreadsheets/d/1zu9qEWI8PsOI_i8nI_S29HDGHlIp2lfVMsGxpQ5tvAQ/edit#gid=973655443*.
[28] U. S. C. Bureau, “Racial and Ethnic Diversity in the United States: 2010 Census and 2020 Census,” Census. gov.
[29] K. M. Krawchenko, “ Increasing and Equalizing Internet Connectivity in American Homes to Improve Health Outcomes,” Georgetown University, 2024.