Machine Learning-Enhanced Oaxaca-Blinder Decomposition for Analyzing Healthcare Test Price Variations Across Indian Cities

Anshuman Parida[1], Pranav Pant[2], Tarun Raj Singh[3]
Department of Computer Science & Engineering
Indian Institute of Technology Jodhpur
Jodhpur, India
b23es1008@iitj.ac.in[1], pant.4@iitj.ac.in[2], singh.188@iitj.ac.in[3]

***Abstract*—This study investigates the variations in healthcare test prices across Indian cities using an extended Oaxaca-Blinder decomposition method integrated with machine learning techniques. Despite the importance of affordable healthcare, significant price disparities exist in diagnostic tests across different regions, potentially limiting access for middle-class and economically disadvantaged populations. We create a comprehensive dataset *HEALTH-PRICE*, encompassing city demographics, economic indicators, and healthcare infrastructure by combining data from multiple sources and employing web scraping techniques. By analyzing the dataset, we decompose price disparities into explained and unexplained components. Our findings reveal significant roles played by factors such as city population, number of laboratories, per capita income, and labor density. The study offers actionable insights for policymakers aiming to reduce inequities in healthcare affordability and enhances understanding of structural inefficiencies in the healthcare system.**

***Index Terms*—Oaxaca-Blinder Decomposition, Healthcare Price Analysis, Machine Learning, Indian Cities, Policy Insights**

I. Introduction

Healthcare costs in India vary significantly across cities, influencing accessibility and equity in healthcare services. While previous studies have extensively examined disparities in healthcare utilization, gender-based spending, and insurance enrollment using Oaxaca-Blinder decomposition, little attention has been given to understanding the underlying causes of diagnostic test price variations.

This study extends the Oaxaca-Blinder decomposition method by developing an *Extended Multi-Group Model* to analyze test price variations across major Indian cities. Instead of limiting the analysis to two groups, our model uses a pooled baseline to systematically compare multiple centers. Furthermore, we integrate machine learning techniques to enhance decomposition accuracy, enabling the quantification of both observable and latent factors influencing price differences.

II. Literature Review

Some of the existing studies are:

Healthcare Inequality in India

- Acharya & Patra (2014) [6] applied the Oaxaca-Blinder decomposition method to analyze **inequality in healthcare access** among the poor and non-poor in India. However, this study primarily focused on healthcare utilization rather than price variations.

Economic Inequality in Eye Care Utilization

- Emamian et al. (2014) [7] used Oaxaca-Blinder decomposition to study **economic inequality in access to eye care services** in rural India.

Gender Discrimination in Healthcare Spending

- Kumar & Kumar (2022) [8] employed the decomposition approach to examine **gender discrimination in intra-household healthcare expenditures**.

Key contributions of this study:

- Unlike previous research, which focused on healthcare access and expenditure, this study specifically examines **healthcare test price variations**, offering insights into how market dynamics and regional factors contribute to cost disparities.
- The study identifies **clusters of cities with distinct pricing behaviors**, helping policymakers tailor interventions to specific urban healthcare markets.

III. HEALTH-PRICE Dataset

HEALTH-PRICE (**H**ealthcare **E**conomic **A**nalysis and **L**ongitudinal **T**est-price **H**ub for **P**rice **R**esearch and **I**nsights across **C**ities and **E**xaminations) is a comprehensive compilation of healthcare and economic indicators across 101 Indian cities. It encompasses data on 163 diseases and 1393 diagnostic test types.

It features crucial variables such as:

- **Population (2023 Approximation)** – Source: Census 2011 [1]
- **City Type** – Source: Department of Expenditure, Government of India (DOE) [9]

- **Zone and State Classification** – Source: DC MSME Report [10]
- **Disease Type & Diagnostic Test Name** – Source: Lal PathLabs [2]
- **Test Price (City-wise)** – Source: Lal PathLabs [2]
- **Number of Laboratories in Each City** – Source: Lal PathLabs [2]
- **Timestamp of Data Collection** – Machine-generated
- **Per Capita Income (City-Level)** – Source: Statista [3]
- **Labor Density** – Derived from multiple sources and approximated calculations

In total, the dataset has **126,477** data points, covering 101 cities, 163 diseases, and 1393 test types. The *final HEALTH-PRICE dataset*, containing the above variables and corresponding city-wise records, is available at [12].

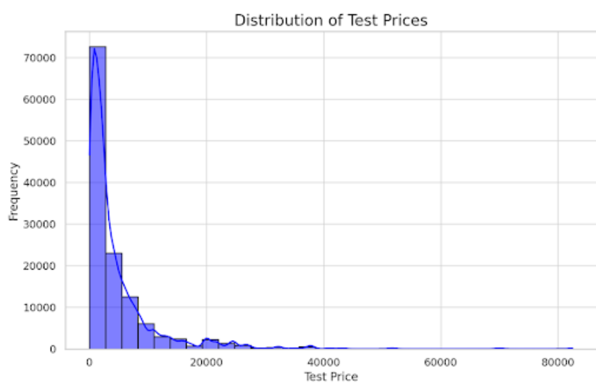

Fig. 1. Visualizing distribution of test prices of the HEALTH-PRICE dataset.

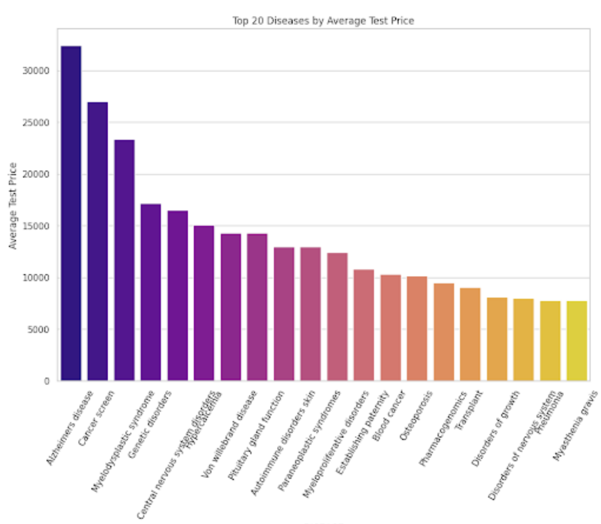

Fig. 2. Top 20 diseases by average test prices.

Fig. 3. City population vs average test prices.

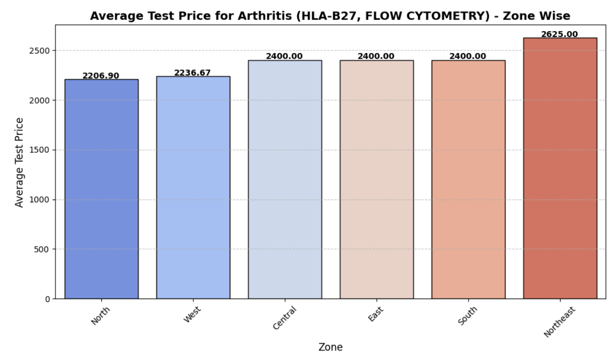

Fig. 4. Average test price for Arthritis (HLA-B27, FLOW CYTOMETRY) zone-wise.

IV. Data Collection Methodology

To systematically collect test price data across multiple Indian cities, we utilized web scraping techniques with Selenium and BeautifulSoup. The scraping process was designed to extract relevant details from Lal PathLabs' website, which provides information on diagnostic tests, pricing, and city-specific healthcare infrastructure.

A. Web Scraping Approach

1) **Automated Browsing with Selenium**
   - The Selenium WebDriver was used to open the website dynamically and interact with JavaScript-rendered elements.
   - A delay was introduced to ensure the page fully loads before extracting data.
2) **Parsing Web Content with BeautifulSoup**
   - The page source was retrieved after JavaScript execution to ensure accurate data extraction.
   - The relevant section containing city-wise test information was identified using CSS class selectors.
3) **Extracting Key Information**
   - All available city-wise test links were located within the 'quick-links' container.
   - The script iterated through anchor tags, retrieving test URLs for further processing.

4) **Automation & Data Storage**
   - The scraping process was executed iteratively to collect test details across different cities.
   - The extracted data was then structured into the dataset, aligning with other city-level economic and healthcare indicators.

By leveraging automated data extraction, we ensured up-to-date, large-scale data collection, minimizing manual errors and enhancing dataset completeness. This approach allowed for a comprehensive city-wise mapping of diagnostic test prices, forming the foundation for our decomposition analysis.

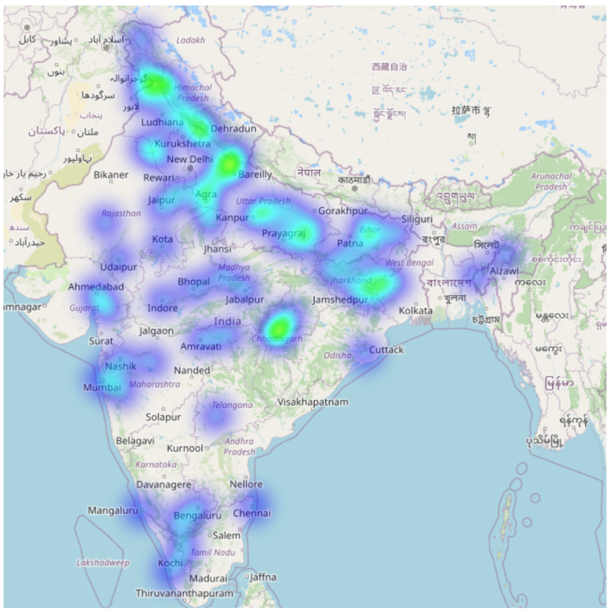

Fig. 5. Visualizing distribution of test prices of Alzheimers.

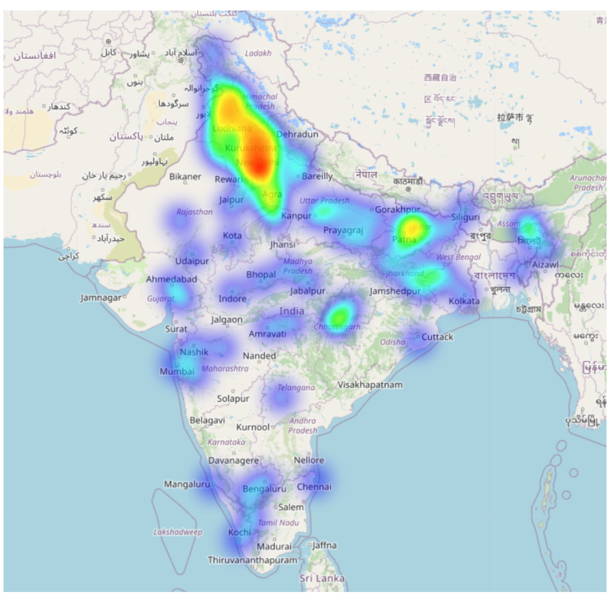

Fig. 6. Visualizing distribution of test prices of Diabetes.

V. Methodology

Our methodology employs a multi-faceted approach, combining Ordinary Least Squares (OLS) regression and Oaxaca-Blinder decomposition to analyze healthcare test price variations across Indian cities.

A. Mathematical Foundations and Justification

The Oaxaca-Blinder decomposition is a statistical technique used to analyze differences in an outcome variable (e.g., wages, healthcare costs) between two groups by separating the differences into explained and unexplained components. Traditionally, this method compares two groups (e.g., men vs. women) and attributes disparities to differences in predictor variables (e.g., education, experience) and differences in how these predictors influence the outcome. In this study, we extend the binary-group Oaxaca-Blinder decomposition to a multi-group setting, where healthcare test price variations across multiple cities are analyzed against a pooled reference model. Our mathematical justification is based on:

1) **Ordinary Least Squares (OLS)** regression models to estimate mean test prices.
2) **Pooled regression modeling** to create a baseline reference for comparison.
3) **Decomposition analysis** to break down the observed differences into explained and unexplained components.

B. OLS Regression Model

We begin with an OLS regression model to estimate the relationship between test prices (the dependent variable) and city-specific predictor variables such as population, number of labs, and per capita income:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \epsilon_i \tag{1}$$

2.2.1 Mathematical Justification for Regression Models

The **ordinary least squares (OLS) regression** is a standard approach to estimate the relationship between a dependent variable y (mean test price) and a set of predictor variables X (e.g., population, number of labs). The regression equation takes the form:

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \epsilon_i$$

where:

- y_i = test price in city i,
- $X_{i1}, X_{i2}, \ldots, X_{ik}$ are predictor variables (e.g., population, number of labs, per capita income),
- β_0 is the intercept,
- β_k are the estimated regression coefficients,
- $\epsilon_i \sim N(0, \sigma^2)$ is the error term.

The OLS regression minimizes the sum of squared residuals:

$$\min_{\beta} \sum_{i=1}^{n} \left(y_i - \left(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} \right) \right)^2$$

By estimating β, we quantify how each predictor influences test prices across different cities.

A pooled regression provides baseline coefficients β^p for comparison, while city-specific models capture local pricing heterogeneity through group-specific coefficients β^g.

2.2.2 Pooled vs. Group-Specific Regression Models

We consider two models:

1) **Pooled Regression Model (Baseline for Comparison)**
   - Combines data from all cities to estimate an **overall trend** in test pricing.

   $$y = \beta_p^0 + \beta_p^1 X_1 + \beta_p^2 X_2 + \beta_p^3 X_3 + \beta_p^4 X_4 + \epsilon$$

   - This provides **pooled coefficients** β_p, which serve as a reference for comparison.
2) **City-Specific Regression Models**
   - Separate models are estimated for each city g, allowing for heterogeneity in pricing.

   $$y_g = \beta_g^0 + \beta_g^1 X_1 + \beta_g^2 X_2 + \beta_g^3 X_3 + \beta_g^4 X_4 + \epsilon_g$$

   - These provide **group-specific coefficients** β_g for each city.

1. Explained Component (Differences Due to Observable Factors)

This component measures the impact of **differences in predictor values** (e.g., population, income) between city g and the pooled average:

$$E_g = \left(\bar{X}_g - \bar{X}_p\right) \cdot \beta_p$$

where:

- $\bar{X}_g$ and $\bar{X}_p$ are the mean predictor values for city g and all cities combined,
- β_p are the pooled model coefficients.

This term represents the **price variation explained by differences in city characteristics**.

2. Unexplained Component (Differences Due to Structural and Latent Factors)

The unexplained component accounts for differences in **coefficients** (pricing structures) between city g and the pooled model:

$$U_g = X_p \cdot (\beta_g - \beta_p)$$

where:

- β_g are the city-specific coefficients,
- β_p are the pooled coefficients.

This term captures **pricing disparities not explained by observable city-level variables**, often due to **market inefficiencies, regulatory differences, or unobserved heterogeneity**.

3. Adjusting for Intercept Differences

The **base-level pricing variations** between cities and the pooled model are captured by intercept differences:

$$\text{Intercept Difference} = \beta_g^0 - \beta_p^0$$

Thus, the revised unexplained component is:

$$U_g = (X_p \cdot (\beta_g - \beta_p)) + (\beta_g^0 - \beta_p^0)$$

where the second term accounts for **city-specific fixed effects**.

By estimating β, we quantify how each predictor influences test prices across different cities. Additionally, we integrate machine learning models to refine coefficient estimates, improving the robustness and interpretability of both the explained and unexplained components.

Implementation

Step 1: Loading and Preprocessing the Dataset

1. Dataset Structure: The dataset includes columns such as city name (*CITY*), city type (*CityType(HRA)*), test prices (*TEST PRICE*), and predictor variables: population (*City Population*), number of labs (*Number of Labs*), income level (*PerCapitaIncome*), and labor density (*LaborDensity*).

- **Test price variations are averaged across cities:**

  $$P_c = \frac{1}{n_c} \sum_{i=1}^{n_c} P_i$$

  where P_c is the mean test price for city c, and n_c is the number of test prices observed for that city.

2. Encoding Variables: Categorical variables like *PerCapitaIncome* and *LaborDensity* are ordinally encoded:

- *Low = 0, Medium = 1, High = 2*

Step 2: Building the Regression Models

1. Pooled Regression Model: The pooled model estimates a baseline trend across all city types, ignoring group-specific differences:

$$\overline{P} = \beta_{p0} + \sum_{j=1}^{k} \beta_{pj} \cdot X_j$$

where:

- $\overline{P}$: Mean test price (dependent variable).
- X_j: Predictor variables (e.g. population, labs, income).
- β_{pj}: Coefficients of the pooled model, which reflect the influence of each predictor across all cities.

2. Group-Specific Regression Models: For each city type g, the code fits separate OLS models to account for heterogeneity:

$$P_g = \beta_{g0} + \sum_{j=1}^{k} \beta_{gj} \cdot X_j$$

where:

- β_{gj} are the group-specific coefficients, reflecting the unique relationship between predictors and test prices in city type g.

Step 3: Oaxaca-Blinder Decomposition

1. Observed Differences in Mean Prices: The difference between a specific city type g and the pooled model is defined as:

$$\Delta P_g = P_g - \overline{P}$$

2. Decomposing the Difference into Components:

a) Explained Component: This reflects differences due to variations in the means of the predictor variables (X) between city type g and the pooled reference:

$$E_g = (\overline{X}_g - \overline{X}_p) \cdot \beta_p$$

where:

- $\overline{X}_g$: Mean predictors for city type g.
- $\overline{X}_p$: Mean predictors for the pooled model.
- β_p: Pooled regression coefficients.
- E_g: Variation explained by observable characteristics.

b) Unexplained Component: This arises from differences in the coefficients (β) themselves:

$$U_g = \overline{X}_p \cdot (\beta_g - \beta_p)$$

where:

- β_g: Coefficients of the group-specific model.
- $(\beta_g - \beta_p)$: Difference in how predictors influence test prices between group g and the pooled reference.
- U_g: Captures latent effects, unobserved factors, or structural differences in pricing.

c) Adjustment for Intercept Differences: The unexplained component also adjusts for variations in the intercept:

$$U_g = \overline{X}_p \cdot (\beta_g - \beta_p) + (\beta_{g0} - \beta_{p0})$$

where:

- $(\beta_{g0} - \beta_{p0})$: Base-level price differences attributable to city-specific fixed effects.

Now many of you might think for capturing relationships neural networks might be better as they can capture non linear relationships,but there is a drawback to that ,now below i will explain the method for incorportating neural networks into this and after that the problems in this method.

VI. Neural Oaxaca-Blinder Decomposition

A. Motivation for Using Neural Networks

While OLS regression provides interpretability, it imposes a linear structure that may not accurately capture the relationships between test prices and city characteristics. Neural networks allow us to:

- **Model nonlinear interactions** (e.g., the effect of income on test prices may not be constant).
- **Improve counterfactual estimation** (i.e., predict what a city's test price would be under different conditions).
- **Handle high-dimensional feature spaces** without extensive manual feature engineering.

B. Neural Network Regression Model

We replace the OLS model with a **feedforward neural network** (MLP) that predicts test prices:

$$y_i = f_\theta(X_i) + \epsilon_i \tag{2}$$

where f_θ is a neural network with parameters θ. The network is trained to minimize:

$$\min_\theta \sum_{i=1}^{n} (y_i - f_\theta(X_i))^2 \tag{3}$$

using ReLU activation functions and Adam optimization.

C. Counterfactual Estimation with Neural Networks

Instead of using OLS to estimate counterfactual prices $\hat{Y}_A^{(B)}$, we use the trained neural network:

$$\hat{Y}_A^{(B)} = f_{\theta_B}(X_A) \tag{4}$$

This allows us to compute:

- **Explained Component (Using NN Predictions):**

$$E_g = \left(\bar{X}_g - \bar{X}_p\right) \cdot \theta_p \tag{5}$$

- **Unexplained Component (Using NN Predictions):**

$$U_g = \left(f_{\theta_p}(X_p) - f_{\theta_g}(X_p)\right) \tag{6}$$

D. Interpretability with SHAP Values

Since neural networks lack explicit coefficients, we use SHAP (Shapley Additive Explanations) to decompose price differences into contributions from individual predictors:

$$y_i = \sum_{j=1}^{k} SHAP_j + \epsilon_i \tag{7}$$

This provides feature-level attribution while maintaining the Oaxaca-Blinder decomposition framework.

So one of the major issue here is: NN Limitation: SHAP values are not equal to regression coefficients

- **SHAP values are instance-specific explanations**
- **Regression provides population-level parameters**
- **No statistical theory for using SHAP means in decomposition**

The traditional Blinder-Oaxaca decomposition was developed with linear regression models in mind, where each predictor has a clear, fixed coefficient that directly measures its marginal effect on the outcome. This framework makes it possible to cleanly decompose differences between groups into parts explained by differences in the predictors (the "explained" part) and parts due to differences in the returns to those predictors (the "unexplained" part).

In contrast, neural networks are nonlinear models that do not have fixed, globally interpretable coefficients. Instead, when you apply methods like SHAP (SHapley Additive exPlanations) to a neural network, you obtain a set of values that reflect each feature's contribution to a particular prediction. These SHAP values are:

- **Local and Instance-Specific:** They explain how much each feature contributed to a specific prediction rather than serving as a global, constant effect across all observations.
- **Post-Hoc Explanations:** They are derived after the model has been trained, summarizing the effect of each input in a complex, often non-additive way.
- **Not Structural Parameters:** Unlike regression coefficients, they do not have a direct interpretation as the marginal effect of a predictor on the outcome in a structural sense.

Because the Blinder-Oaxaca decomposition relies on the linear and additive structure of coefficients to partition differences between groups, substituting these with SHAP values from a neural network is not straightforward or theoretically equivalent. SHAP values, while very useful for interpreting individual predictions in complex models, do not provide the consistent, aggregate parameter estimates that the traditional decomposition requires.

VII. Results and Analysis

Our analysis reveals significant variations in healthcare diagnostic test prices across Indian cities, with both observable and latent factors contributing to these disparities. The extended Oaxaca-Blinder decomposition, enhanced by machine learning techniques, shows that while factors like population size, laboratory density, and per capita income explain a portion of these variations, a substantial component remains unexplained.

Decomposition Results for CityType(HRA) Z:

	Factor	Explained Component	Unexplained Component
0	Intercept	0.000000	38.548509
1	City Population	42.917583	-120.884636
2	Number of Labs	-14.031057	-120.332750
3	PerCapitaIncome	-49.790790	-181.124324
4	LaborDensity	1.586061	333.174593
5	Total	-19.318204	-50.618608

Decomposition Results for CityType(HRA) Y:

	Factor	Explained Component	Unexplained Component
0	Intercept	0.000000	32.971213
1	City Population	20.730791	114.340648
2	Number of Labs	8.807432	-7.119449
3	PerCapitaIncome	22.163633	55.517810
4	LaborDensity	-0.868992	-147.662059
5	Total	50.832865	48.048163

Decomposition Results for CityType(HRA) X:

	Factor	Explained Component	Unexplained Component
0	Intercept	0.000000	-4368.283447
1	City Population	-349.156055	NaN
2	Number of Labs	8.951237	1248.579411
3	PerCapitaIncome	93.125582	433.647352
4	LaborDensity	-1.866357	855.599561
5	Total	-248.945594	-1830.457123

Fig. 7. Illustration of the decomposition of average healthcare diagnostic test price variations based on city types.

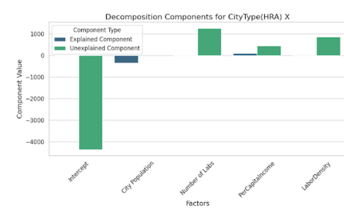

Fig. 8. Illustration of the decomposition component for citytype X.

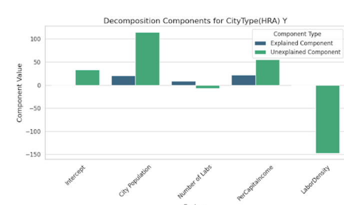

Fig. 9. Illustration of the decomposition component for citytype Y.

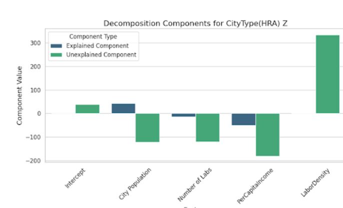

Fig. 10. Illustration of the decomposition component for citytype Z.

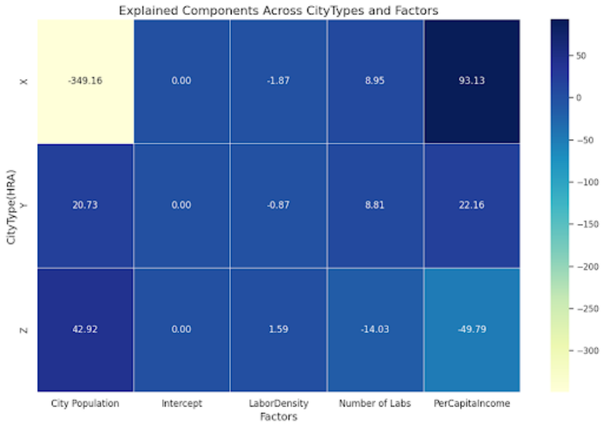

Fig. 11. Heatmap Explained.

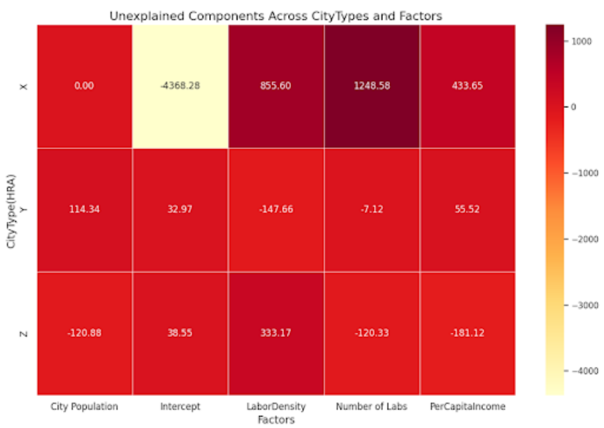

Fig. 12. Heatmap Unexplained.

Decomposition Results for CityType(HRA) Z:

	Factor	Explained Component	Unexplained Component
0	Intercept	0.000000	-4.122204
1	City Population	1.282586	-1.746383
2	Number of Labs	0.092787	3.043371
3	PerCapitaIncome	2.676964	1.315415
4	LaborDensity	-2.288355	2.835900
5	Total	1.763982	1.326100

Decomposition Results for CityType(HRA) Y:

	Factor	Explained Component	Unexplained Component
0	Intercept	0.000000	-1.418852
1	City Population	0.268422	1.409349
2	Number of Labs	-0.054393	0.031750
3	PerCapitaIncome	-1.197914	4.578763
4	LaborDensity	1.209816	-2.719897
5	Total	0.225932	1.881113

Decomposition Results for CityType(HRA) X:

	Factor	Explained Component	Unexplained Component
0	Intercept	0.000000	-518.659430
1	City Population	-8.943096	NaN
2	Number of Labs	-0.068788	137.628350
3	PerCapitaIncome	-4.893484	-17.941634
4	LaborDensity	2.727667	111.112304
5	Total	-11.177700	-287.860411

Fig. 13. Illustration of the decomposition of healthcare diagnostic test price for fructosamine test of Diabetes based on city types.

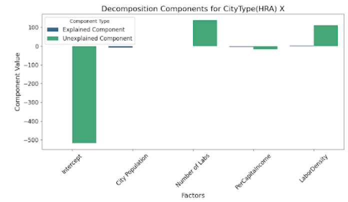

Fig. 14. Citytype X.

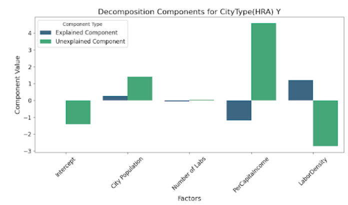

Fig. 15. Citytype Y.

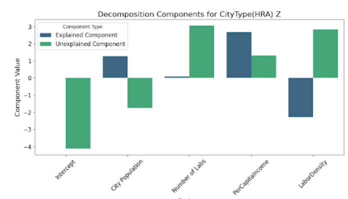

Fig. 16. Citytype Z.

Similarly we have found out results for groupings based on disease type,test type,zone and other independent variables. Detailed decomposition results (including specific disease/test prices and groupings by zone, disease type, number of parameters, and test types) can be accessed by the link in references.

Key Observations:

- **Observable Factors:** Population size, per capita income, and laboratory density have notable explanatory power in accounting for price variations. These are quantifiable variables that policymakers can influence directly.
- **Unexplained Disparities:** The unexplained portion is large, suggesting that market inefficiencies, regulatory inconsistencies, and provider-specific pricing strategies significantly shape test cost structures.
- **Impact on Vulnerable Populations:** Middle-class and economically disadvantaged groups are disproportionately affected by price disparities, underscoring the policy relevance of these findings.

The integration of machine learning methods has provided unprecedented precision in estimating these components, offering a more nuanced understanding of how structural factors influence healthcare costs across different urban centers.

VIII. Conclusion and Future Work

This study provides a comprehensive analysis of healthcare diagnostic test price variations across Indian cities by leveraging an extended Oaxaca-Blinder decomposition model integrated with machine learning techniques. By decomposing price disparities into explained and unexplained components, we identified key factors driving these variations, including population size, number of laboratories, per capita income, and labor density. The integration of machine learning enhanced the precision of these estimations, enabling a nuanced understanding of observable and latent influences on test prices. The findings highlight that a significant portion of the price disparities remains unexplained, likely attributed to market inefficiencies, regulatory inconsistencies, and pricing strategies employed by healthcare providers. These insights are particularly critical in a country like India, where the majority of the population belongs to middle-class or economically disadvantaged groups, who are most affected by healthcare cost inequities.

Implications for Middle-Class and Economically Disadvantaged Populations

- **Enhanced Affordability:** The study underscores the urgent need to address unexplained price disparities, which disproportionately affect lower-income households. Reducing such disparities can directly make healthcare more affordable for middle-class and poor families.
- **Targeted Subsidies:** Policymakers can use the explained components (e.g., population, income levels, number of laboratories) to identify cities or regions where test prices are unjustifiably high relative to the local economic conditions. Targeted subsidies or price caps can be implemented in these areas to ensure affordability.
- **Transparency in Pricing:** Unexplained components often arise from opaque pricing mechanisms. Encouraging transparency and mandating price disclosures for diagnostic services can empower consumers and discourage exploitative practices.
- **Improved Healthcare Accessibility:** By reducing cost barriers, diagnostic tests become more accessible to underserved populations, enhancing overall healthcare access and utilization.
- **Subsidized Diagnostics for the Poor:** Implementing schemes such as free or subsidized diagnostic services for low-income households can reduce the financial burden of healthcare costs. These programs can be tailored to regions with high unexplained price disparities.

Policy Implications:

- **Regulatory Mechanisms:** Establish transparent guidelines to ensure that diagnostic test pricing remains fair and accessible.
- **Infrastructure Investment:** Improve laboratory facilities and public healthcare centers in under-served cities to reduce monopoly-driven pricing.
- **Price Transparency:** Promote open data and comparison platforms to empower consumers with clear information on test costs
- **Standardized Pricing Regulations:** The government can introduce standardized pricing frameworks for diagnostic tests based on key predictors. This ensures uniformity and prevents arbitrary price hikes.
- **Public Healthcare Investments:** Strengthening public healthcare infrastructure, particularly in underserved areas, can increase competition and lower test prices. Investments in government laboratories and diagnostic centers can provide affordable alternatives to private services..

Future Directions:

- **Expanded Dataset:** Incorporate additional cities and newer diagnostic test types to validate the generalizability of our findings.
- **Refined ML Models:** Utilize advanced machine learning methods for even more robust decomposition estimates.
- **Real-Time Monitoring:** Develop dashboards and API integrations that update test prices dynamically, offering policymakers near real-time insights.

By spotlighting the roles of both quantifiable city characteristics and latent market inefficiencies, our approach paves the way for evidence-based strategies to make healthcare more affordable and equitable across India's diverse urban landscapes.

References

[1] Census 2011, "Census of India: 2011," [Online]. Available: https://www.census2011.co.in/

[2] Lal PathLabs, "Diagnostic and Pathology Services," [Online]. Available: https://www.lalpathlabs.com/

[3] Statista, "Global Statistics and Industry Data," [Online]. Available: https://www.statista.com/

[4] R. Oaxaca, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, vol. 14, no. 3, pp. 693–709, 1973.

[5] A. S. Blinder, "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources*, vol. 8, no. 4, pp. 436–455, 1973.

[6] A. Acharya and N. Patra, "Inequality in Healthcare Access among the Poor and Non-Poor in India: A Oaxaca-Blinder Decomposition Approach," *ResearchGate*, 2014. [Online]. Available: https://www.researchgate.net

[7] M. H. Emamian *et al.*, "Economic Inequality in Access to Eye Care Services in Rural India: A Oaxaca-Blinder Decomposition Study," *NIH*, 2014. [Online]. Available: https://www.nih.gov

[8] R. Kumar and P. Kumar, "Gender Discrimination in Intra-Household Healthcare Expenditures: Evidence from India," *Taylor & Francis*, 2022. [Online]. Available: https://www.taylorandfrancis.com

[9] Department of Expenditure (DOE), Government of India, "Classification of Cities for House Rent Allowance," [Online]. Available: https://www.doe.gov.in/

[10] Development Commissioner, Ministry of Micro, Small & Medium Enterprises (DC MSME), "Zonal and State Classification Reports," [Online]. Available: https://dcmsme.gov.in/

[11] "Detailed Results," [Online]. Available: https://docs.google.com/document/d/1EjU-BFO0hPZB-kFMHNXzLFi8hkjskds7/

[12] "HEALTH-PRICE Final Dataset," [Online]. Available: https://drive.google.com/file/d/1AnsBxfH_G2bzNmGcgKCPogZ7gdQij8aL/view?usp=sharing