Comprehensive Asthma Prediction Model Analysis and Documentation

Executive Summary

This report documents our development of predictive models for asthma cases across multiple hospitals in Nigeria from 2018 to 2022. Our analysis began with traditional regression models using environmental factors (SO_2 and PM2.5) as predictors, which proved inadequate. We then developed a neural network approach that incorporated hospital-specific information, achieving significant improvement in predictive accuracy. The final model can explain approximately 60.6% of the variance in asthma cases, demonstrating that hospital-specific characteristics are crucial factors in predicting asthma incidence.

Data Overview

The dataset contains 60 records from 12 different hospitals over 5 years (2018-2022) with the following key information:

- Hospital location
- Year
- Number of asthma cases
- SO_2 levels
- PM2.5 levels
- Demographic breakdowns (gender and age groups)

Key Statistics:

- Average asthma cases per hospital per year: 137
- Range of asthma cases: 0 to 563
- SO_2 range: 0.42 to 0.83
- PM2.5 range: 4.02e-08 to 5.95e-08

Hospital Distribution:

The dataset includes hospitals of varying sizes and characteristics:

- Large regional centers (e.g., OAUTHC IFE with up to 563 cases per year)
- Medium-sized facilities (e.g., FMC IDO with around 200 cases per year)
- Smaller local hospitals (e.g., GEN. HOSPITAL, IJU with as few as 2 cases per year)

Analysis Journey

Phase 1: Traditional Models

We initially tested five different prediction models using only year, SO_2, and PM2.5 as predictors:

Model	RMSE	R^2 Score
Linear Regression	54.99	-0.0895
Ridge Regression	55.12	-0.0943
Lasso Regression	55.05	-0.0918
Random Forest	122.25	-4.3836
Gradient Boosting	128.07	-4.9092

Phase 1 Insights:

The negative R^2 scores indicated that all models performed worse than simply using the mean of asthma cases as a predictor. This suggested that:

1. The relationship between the predictors (YEAR, SO_2, PM2.5) and asthma cases was not well captured by these models
2. Significant factors influencing asthma cases were missing from our dataset
3. Hospital-specific factors were likely dominating the patterns

Phase 2: Neural Network with Hospital Information

Based on our initial findings, we developed an enhanced approach:

1. **Including Hospital Identity**: We added hospital-specific identifiers to allow the model to learn baseline differences between facilities
2. **Advanced Architecture**: We implemented a neural network with multiple layers and dropout for regularization
3. **Feature Engineering**: We created polynomial terms and interaction features to capture non-linear relationships

Neural Network Architecture:

- Input layer (matching feature dimensions)
- First hidden layer: 64 neurons with ReLU activation
- Dropout layer (30% dropout rate)
- Second hidden layer: 32 neurons with ReLU activation

- Dropout layer (20% dropout rate)
- Third hidden layer: 16 neurons with ReLU activation
- Output layer: 1 neuron (linear activation for regression)

Phase 2 Results:

The neural network achieved an R^2 score of 0.6057, indicating it explains approximately 60.6% of the variance in asthma cases. This represents a dramatic improvement over the previous models.

Key Findings

1. **Hospital Identity is Critical**: The dramatic improvement in performance confirms that hospital-specific factors are dominant in determining asthma case numbers. These likely include hospital size, local population served, specialization, and reporting practices.
2. **Environmental Factors Alone Are Insufficient**: While air quality measures (SO_2 and PM2.5) may contribute to asthma cases, they cannot predict cases accurately without accounting for hospital context.
3. **Non-Linear Relationships**: The success of the neural network suggests complex, non-linear relationships between predictors and asthma cases that simpler models couldn't capture.
4. **Hospital-Specific Predictions**: The model produces significantly different predictions for different hospitals given the same environmental conditions, reflecting real-world variation.

Prediction Capabilities

For a specific scenario (Year=2023, SO_2=0.65, PM2.5=5.5e-08):

- **OAUTHC IFE (Hospital ID 4)**: 74 asthma cases predicted
- **Unknown/New Hospital**: 122 asthma cases predicted

Hospital Reference Guide

For future predictions, hospital IDs are as follows:

ID	Hospital Name
0	EKSUTH
1	FMC IDO
2	GENERAL HOSPITAL, IKERE-EKITI
3	GENERAL HOSPITAL OTUN
4	OAUTHC IFE
5	SDA HOSPITALS, ILE IFE
6	FMC OWO
7	UNIMEDTH ONDO
8	GEN HOSPITAL, ORE
9	GEN.HOSPITAL, ILE-OLUJI
10	GEN HOSPITAL, IDANRE
11	GEN. HOSPITAL, IJU

Limitations

Despite the improvements, some limitations remain:

1. **Limited Dataset Size**: With only 60 records across 12 hospitals, the model has relatively few training examples.
2. **Missing Contextual Factors**: Information about seasonal variations, local population demographics, and specific environmental conditions near each hospital could further improve predictions.
3. **Unknown Hospital Generalization**: Predictions for new/unknown hospitals rely on patterns learned from existing facilities and may be less accurate.
4. **Limited Temporal Coverage**: The dataset only covers 5 years, which may not be sufficient to capture long-term trends or cyclical patterns.

Recommendations for Improved Predictions

1. **Normalize by Population**: Convert raw case numbers to rates based on local population or hospital capacity to make comparisons more meaningful.
2. **Include Additional Environmental Variables**: Add other air quality measures, seasonal factors, or weather patterns that might influence asthma incidence.
3. **Collect More Data**: Additional years of data would enhance the model's ability to recognize patterns.
4. **Specialized Models**: Consider developing separate models for different hospital types (large vs. small, urban vs. rural).

5. **Regular Retraining**: Update the model annually with new data to capture changing patterns.

Practical Applications

The neural network model with hospital-specific information can be used to:

1. **Resource Planning**: Help hospitals anticipate staffing and resource needs based on predicted asthma cases.
2. **Public Health Monitoring**: Track the relationship between environmental factors and asthma cases at different locations.
3. **Intervention Assessment**: Evaluate the impact of environmental interventions on asthma case numbers.
4. **Regional Analysis**: Compare expected vs. actual case numbers across different hospitals to identify anomalies.

Non-Technical Explanation

For stakeholders without technical expertise, here's a simplified explanation:

What We Found

Our analysis discovered that predicting asthma cases isn't simply about air quality measurements. Each hospital has its own unique pattern of cases influenced by many factors including its size, the community it serves, and its specialization.

Why Early Models Failed

Our first attempts used only air quality data (SO_2 and PM2.5) and year to predict asthma cases. This would be like trying to predict restaurant customer numbers using only the outside temperature - it's a factor, but not the main one!

How We Fixed It

By teaching our model to recognize each hospital's unique patterns and using more sophisticated mathematical techniques (neural networks), we created a much more accurate prediction system. This is like adding information about each restaurant's size, menu, and location to our prediction.

What This Means

The model can now make reasonable predictions about expected asthma cases for each hospital given certain conditions. For example, with SO_2 levels of 0.65 and PM2.5 levels of 5.5e-08 in 2023:

- OAUTHC IFE would expect around 74 cases

- A typical hospital in the region would expect around 122 cases

Next Steps

To make the model even better, we would need:

- More years of data
- Information about local populations
- More detailed environmental measurements
- Hospital capacity information

Conclusion

Our analysis demonstrates that predicting asthma cases requires accounting for hospital-specific characteristics alongside environmental factors. The neural network model with hospital identifiers represents a significant improvement in prediction capability and provides a foundation for more sophisticated predictive tools in healthcare resource planning and environmental health monitoring.

With further refinement and additional data, this approach could become an important planning tool for healthcare facilities managing asthma cases across Nigeria and potentially other regions with similar healthcare systems.