DNA Analysis Visualization Guide

1. Model Accuracy Comparison (Bar Chart)

- **Type**: Bar Chart
- **Data Source**: modelAccuracy array
- **Processing**:

    - Shows performance metrics for different ML models (SVM, KNN, CNN, Hybrid)
    - Each model has overall accuracy and disease-specific performance
    - Hybrid model (SVM+CNN+HMM) shows highest accuracy (97-98%)
    - Disease-specific variations:

        - Cancer: Higher accuracy (94-98%)
        - Alzheimer's: Slightly lower (91-97%)
        - Other diseases: Progressively lower accuracy

2. Region-Wise Disease Risk Heatmap (Bar Chart)

- **Type**: Bar Chart with Risk Levels
- **Data Source**: locationRisk array
- **Processing**:

    - Calculates risk scores based on:

        - Disease severity (Cancer: 0.9, Alzheimer's: 0.85, etc.)
        - Parameter weights (1.2, 0.8, 1.5)
        - Location-specific factors

    - Risk normalization: Divided by 3.5 for 0-1 scale
    - Shows geographical distribution of disease risks across Indian cities

3. Feature Importance Analysis (Vertical Bar Chart)

- **Type**: Vertical Bar Chart
- **Data Source**: parameterAverages array
- **Features**:

    - DNA Mutation Rate (weighted by 1.3 for Cancer)
    - Gene Expression (weighted by 1.2)
    - Biomarker Level (weighted by 0.9)

- **Processing**:

    - Disease-specific multipliers applied
    - Values normalized and averaged

4. Disease Trend Over Time (Line Chart)

- **Type**: Line Chart
- **Data Source**: timeSeriesData array

- **Processing**:

    - Maps disease distribution over years (2020 onwards)
    - Weighted distribution:

        - Cancer: 2.5x weight
        - Alzheimer's: 1.8x weight
        - Cystic Fibrosis: 1.2x weight

    - Shows disease progression patterns

5. Clustering & Anomaly Detection (Scatter Plot)

- **Type**: Scatter Chart
- **Data Source**: Location and risk data
- **Processing**:

    - X-axis: Risk scores
    - Y-axis: Random distribution for visualization
    - Clusters formed based on:

        - Disease types
        - Risk levels
        - Geographical proximity

6. Region-Wise Case Distribution (Pie Chart)

- **Type**: Pie Chart
- **Data Source**: Disease distribution per location
- **Processing**:

    - Shows proportion of cases by region
    - Includes percentage calculations
    - Color-coded by disease type
    - Dynamic labeling with percentages

7. Prediction Confidence Distribution (Histogram)

- **Type**: Bar Chart (Histogram)
- **Data Source**: Risk scores and confidence levels
- **Processing**:

    - Shows distribution of prediction confidence
    - Based on:

        - Model accuracy
        - Parameter reliability
        - Disease severity

8. Model Performance Metrics (Radar Chart)

- **Type**: Radar Chart
- **Data Source**: Various performance metrics

- **Metrics Displayed**:
  - Accuracy
  - Precision
  - Recall
  - Disease-specific performance
  - Parameter reliability

Data Processing Details

Disease Weighting

```
const weight = {
 "Cancer": 2.5,
 "Alzheimer's Disease": 1.8,
 "Cystic Fibrosis": 1.2,
 "Others": 1.0
}
```

Risk Calculation

```
risk = (param1 * 1.2 + param2 * 0.8 + param3 * 1.5) * baseSeverity / 3.5
```

Parameter Variations

- DNA Mutation Rate: Disease-specific multiplier (1.3 for Cancer)
- Gene Expression: Standard multiplier (1.2)
- Biomarker Level: Reduced weight (0.9)

Visualization Updates

- All graphs update automatically when:

  1. New data is processed
  2. Different model is selected
  3. New file is uploaded
  4. Parameters are adjusted

Chart Interaction Features

- Zoom capabilities
- Hover tooltips
- Click for detailed view
- Export options
- Dynamic resizing