HITACHI
Inspire the Next

A Data-Driven Approach for Region-wise Environmental Health and COVID-19 Risk Assessment Scores

Authors :
Sanjana Pai Nagarmat, Saiyed Kashif Shaukat
Hitachi India Pvt Ltd, R&D Centre

Presented by :
Sanjana Pai Nagarmat
Research Engineer
Hitachi India Private Ltd.
(sanjana@hitachi.co.in)

About Me

HITACHI
Inspire the Next

A passionate Research Engineer with over 5 years experience in research and development.

Work Experience	Hitachi India R&D (2015 - Present) • Smart city analytics • Cybersecurity (Vulnerability Risk Management) • Cloud Platform (Upstream technical contributor to Openstack, log monitoring, Federated authentication etc.)
Education	Bachelor of Engineering in Computer Science
Interests	Travelling, Cooking and food photography
Contact Information	sanjana@hitachi.co.in

HITACHI
Inspire the Next

Table of contents

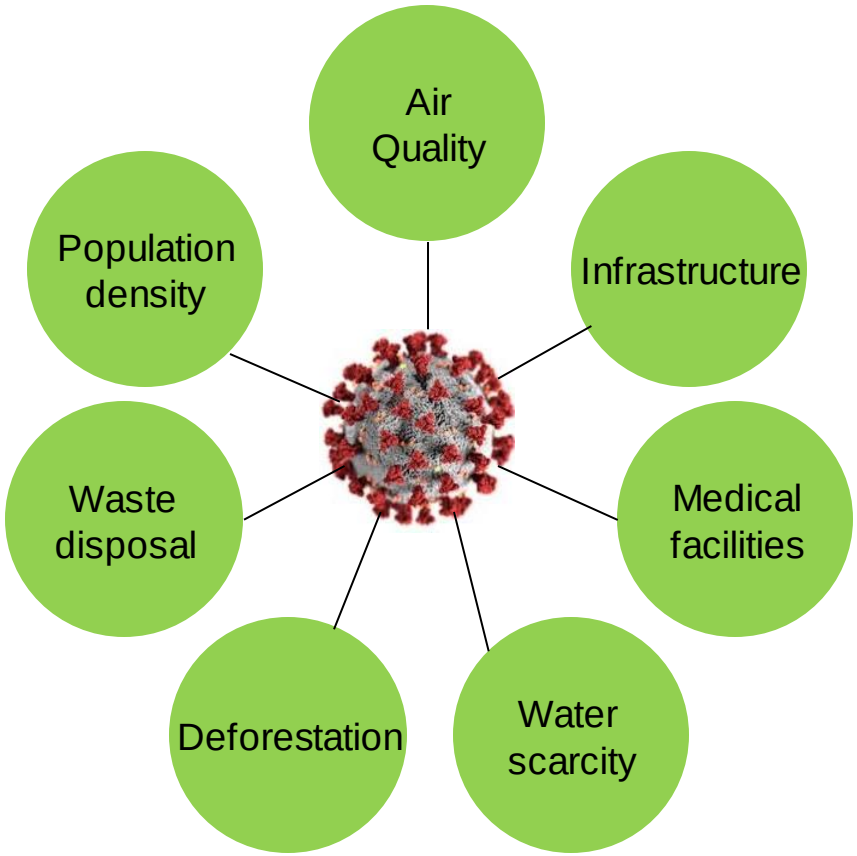
1.Introduction

HITACHI
Inspire the Next

Background

- Cities are facing massive issues related to environment, living conditions and medical facilities with rapid urbanization.
- India's urban population is expected to grow from 410 million in 2014 to **814 million** by 2050.*

- Smart cities need to solve the problems of Urbanization.
- **Environmental conditions** of the city need to be assessed and prioritized.
- **Medical facilities** in the cities has to be improved to provide better quality of life to citizens.
- Cities have to be **self sufficient** to handle situations like the pandemic.



*https://www.thehindubusinessline.com/opinion/mega-challenges-of-rural-urban-migration/article29577159.ece

2.Problem

HITACHI
Inspire the Next

Lack of data rich models that provide interpretable granular ward level solutions to aid and assist city planners.

2.1 Lack of granularity – Existing applications provide solution at a country, state or at a city level not at a ward level.

2.2 Interpretability – solutions do not provide quantified interpretable outputs.

2.3 Data segregation –data generated by the smart cities are in silos.

*https://www.thehindubusinessline.com/opinion/mega-challenges-of-rural-urban-migration/article29577159.ece

3. Related works

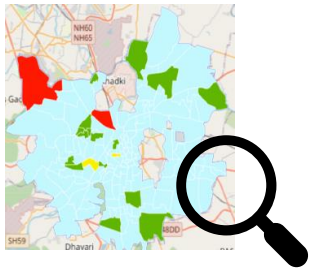
HITACHI
Inspire the Next

- Anita et al. studied the scope and opportunity in Pune to **understand the green cover** for maintaining a balanced floral diversity.
- National Research Development Corporation studied the **air quality** in Pune to develop an air information response plan.
- Morani et al. provide a **planting index** for a location by analyzing the associated tree cover. (Uses Geographic Information System (GIS) based visual approximation techniques to estimate the green tree cover.
- Shubham et al. studied the **effect of restricted emissions** during COVID-19 on air quality in India.
- Clement et al**.** studied and evaluated the **impact of some selected demographic and environmental variables** to identify potential risk areas and hotspots for COVID-19 transmission in Nigeria.

4. Methodology

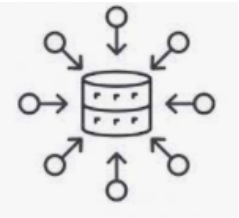
HITACHI
Inspire the Next

A novel data-driven approach which **aggregates** relevant data sources, creates data rich models to provide **granular ward level** insights and interpretable **scores** to improve the overall **health and living conditions in the cities.**



4.1 Combining and utilizing rich feature set to provide **granular** ward-wise environmental health scores and recommendations.

4.2 Investigating the environmental attributes, their trends and variations during the COVID-19 pandemic.



4.3 Aggregated data rich models that provides dynamic ward level risk scores.

*https://www.thehindubusinessline.com/opinion/mega-challenges-of-rural-urban-migration/article29577159.ece

4.1. Environmental Health scores and Recommendations

HITACHI
Inspire the Next

PROBLEM

Increased urbanization is causing **deterioration of environment** and living conditions.

PROPOSED SOLUTION

Provides an **insight** and **recommendation** system that helps city planner take **data driven decisions** to improve cities and to enforce smarter policies. Our system provides,

- Insights related to the trees, population and air quality at the ward level.
- Ward wise environmental health score.
- Recommendations for creating green buffer zones around user specified locations to promote planting trees and help counter the urbanization effects.

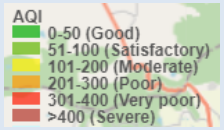
4.1.1 Health Score formulation

HITACHI
Inspire the Next

Data

Tree Census

Air Quality Index



Population Census

$$Health\_Score = (Tree\_Score * 30) + (AQI\_Score * 40) + (Population\_Score * 30)^{*} \quad \text{...[1]}$$

$$Tree\_Score = \frac{\sum_{i=1}^{n} Tree\_Value(i)}{Ideal\_Tree\_Score} \quad \text{...[2]}$$

$$Ideal\_Tree\_Score = Max\_Tree\_Score * Ward\_Population * 7 \quad \text{...[3]}$$

$$Tree\_Value = w1(Height) + w2(Canopy) + w3(Phenology) + w4(Condition) \quad \text{...[4]}$$

* https://www.fs.usda.gov/treesearch/pubs/38074

4.1.1 Health Score formulation cont…

HITACHI
Inspire the Next

TABLE I
WEIGHTS FOR TREE ATTRIBUTES

w1		w2		w3		w4	
Height	***Score***	***Canopy***	***Score***	***Condition***	***Score***	***Phenology***	***Score***
0-3m	0.2	0-1m	0.2	Poor	0.25	Seasonal	0.3
3-5m	0.4	1-3m	0.4	Average	0.5	Evergreen	0.7
5-8m	0.6	3-5m	0.6	Healthy	0.75		
8m+	0.8	5m+	0.8				

$$AQI\_Score = \frac{(Max\_AQI\_Value - AQI\_Value)}{Max\_AQI\_Value} \quad \text{…[5]}$$

$$Population\_Score = \frac{(Max\_Density - Ward\_Density)}{Max\_Density} \quad \text{…[6]}$$

4.1.1 Health Score formulation cont…

HITACHI
Inspire the Next

TABLE II
HEALTH SCORES OF PMC-PUNE WARDS

Ward Name	Health Score	AQI	#Trees	Population
Baner-Balewadi	94.8	54.1	207671	150190
Ved Bhavan	92.2	80.4	115324	171906
Lohagaon Vimantal	91.9	111.2	293974	163064
Mundhvagaon	91.0	94.4	160422	90815
V Mahavidhyalaya	90.5	98.0	81776	68110
S Mahavidhyalaya	90.0	106.3	19766	73811
Fergussion College	89.0	114.9	50105	76911
Magarpatta Hadapsar	89.0	83.8	25203	193003
Sadhana Vidhyalaya	88.5	93.5	22666	135103
Kothrud Gaon	86.8	103.2	72851	90671
RajBhavan	85.6	123.6	5084	88484
Shanivarwada	82.3	133.8	54499	61042
PhuleNagar Yerwada	81.7	160.2	125254	78268
Koregaon Park	81.1	111.7	5458	71299
AundhGaon	68.2	93.2	7663	83859
D.P Dattawadi	65.8	83.5	42050	83026
S.G Rugnalya	63.7	103.3	3071	86535
Tingre Station	62.2	83.2	4336	111015

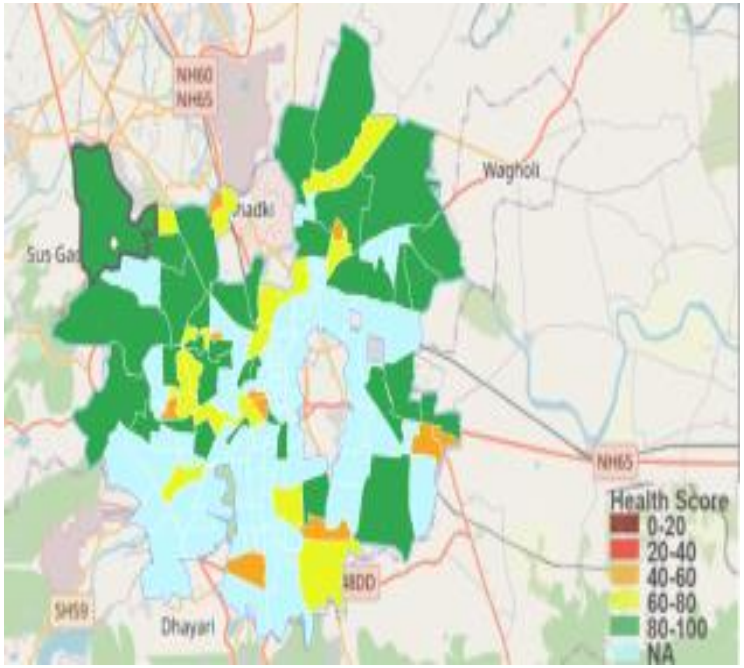


Figure 1. Ward Health Scores

4.1.2 Recommendation

Green Buffer Zone Recommendation to improve the green cover of a selected place based on its calculated health score.

1. Select location and radius.
2. Compute Health Score.
3. Provide **Recommendations:**

(Threshold : 80)

if (health score < 80)

then: provide recommendation to plant trees.

else,

Ward is healthy.

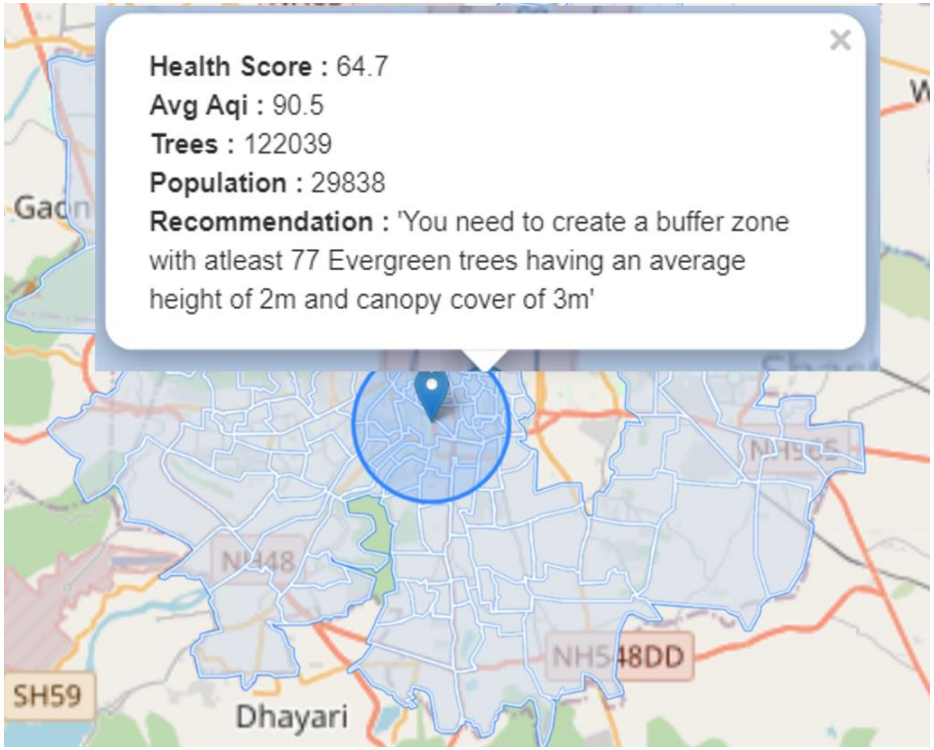


Figure 2. Green Buffer Zone Recommendation

4.2. Attribute Variations and Trends

HITACHI
Inspire the Next

Studied variations and trends in AQI, sound levels and individual pollutants across wards in Pune (Period : Jan 2020 to April 2021)

- AQI levels decreased by 41.0%
- Sound levels (dB) decreased by 3.0%
- Ozone levels increased by 77.5%
- PM2.5 levels decrease by 83.7%
- PM10 levels decrease by 70.3%
- No2 levels decrease by 82.5%

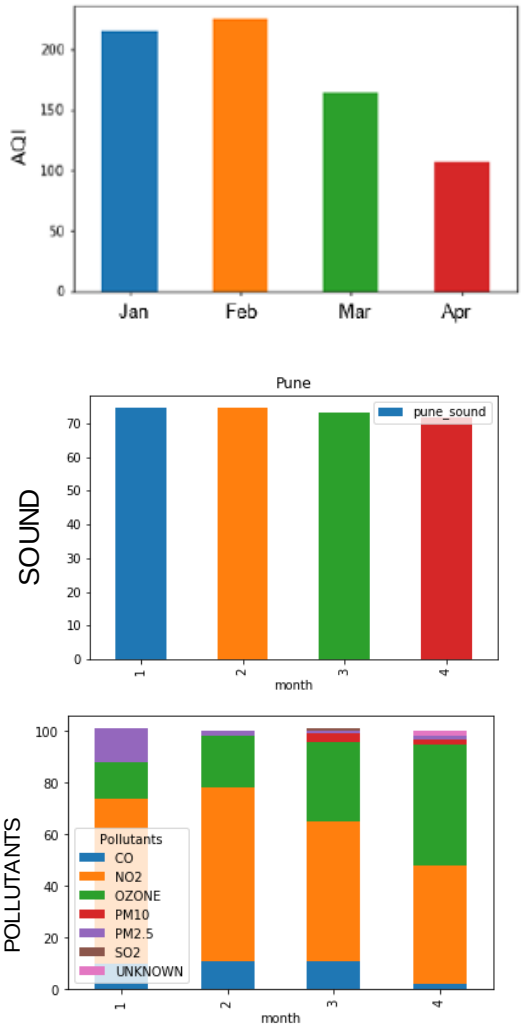


Figure 3. Attribute Variations

4.2.1 Visualization – Environmental Parameter Trend

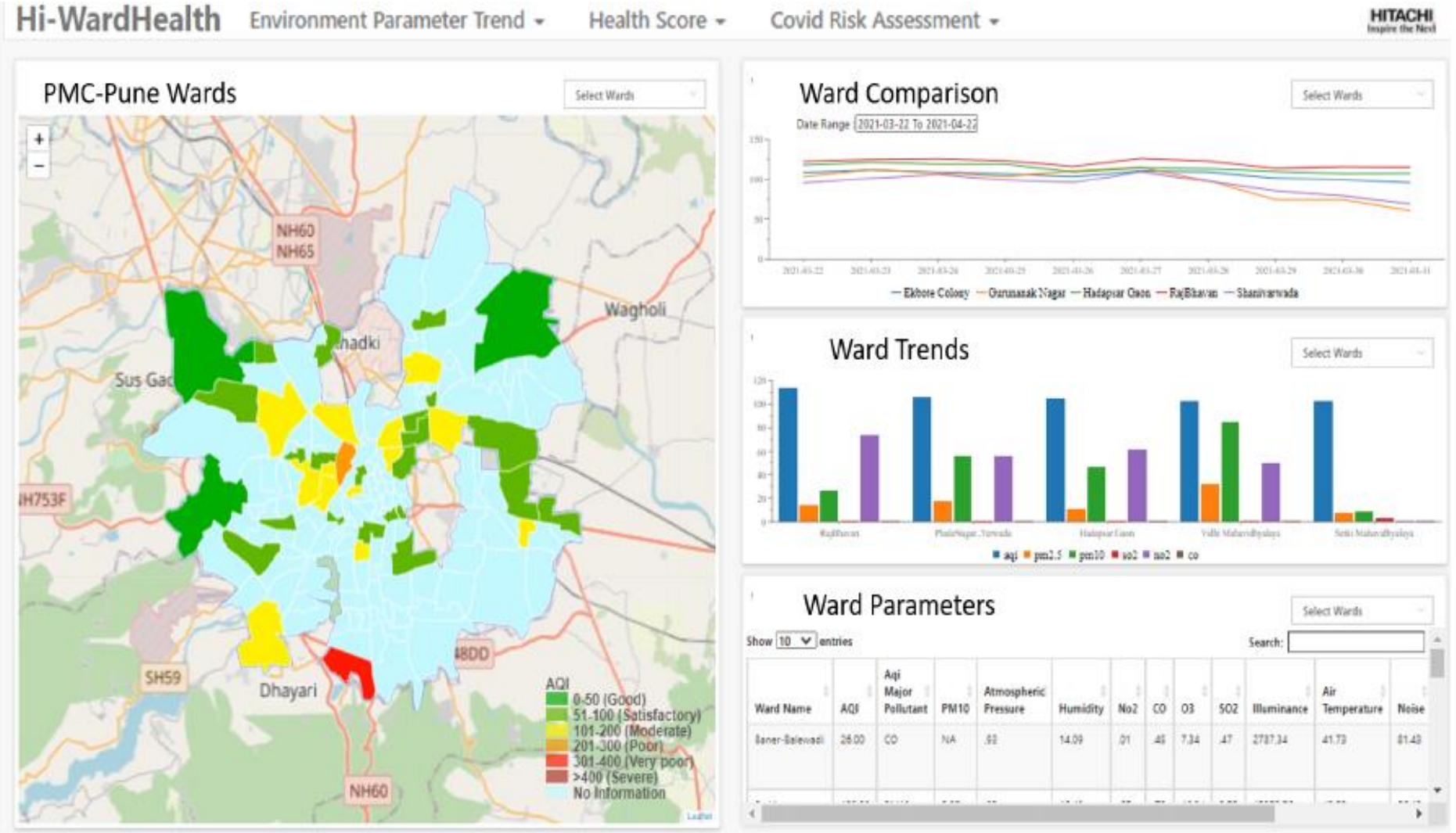


Figure 4. Pune-PMC Environmental parameter trend

4.2.1 Visualization –Ward level environmental parameter trend

HITACHI
Inspire the Next

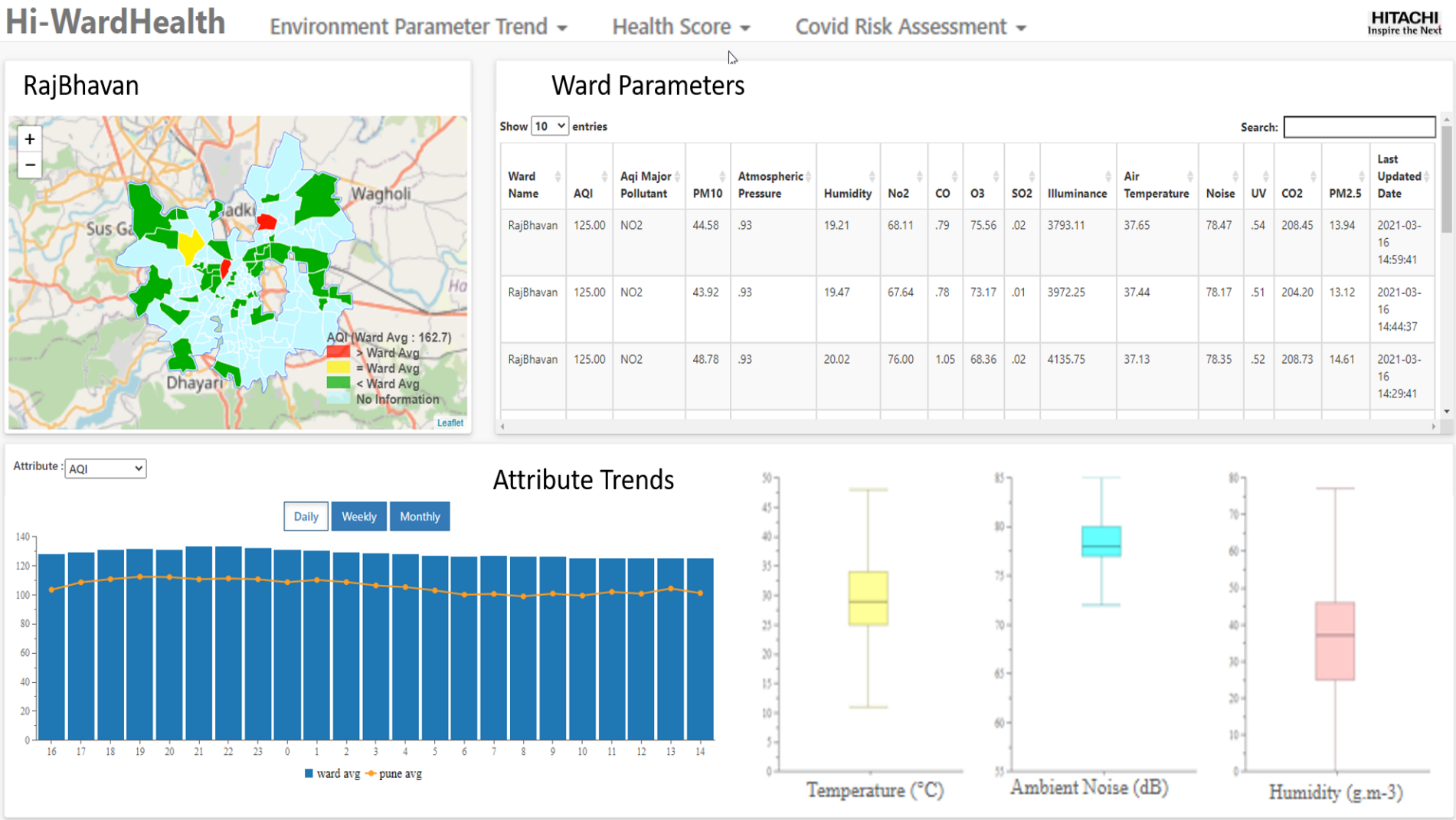


Figure 5. Pune-PMC Ward level Environmental parameter trend

4.3 COVID-19 ward risk assessment

HITACHI
Inspire the Next

Background:

The unprecedented event of COVID 19 has put enormous amount of responsibility and pressure on city/town administrators and utilities to **improve** the existing **healthcare facilities**.

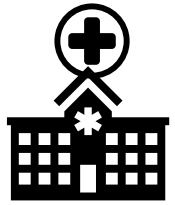
Proposed solution

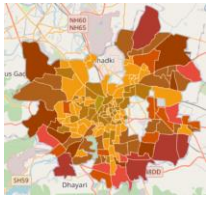
- Assess the overall health and infrastructure facilities available at ward level to **indicate** the **current risk**.
- **Formulate** it as **risk** score and provide **ward specific insights**.
- **Categorize wards** into critical, moderate and low risk zones.
- Help take **prioritized actions** for **ward planning** and further development.

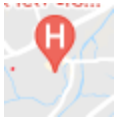
4.3.1 Data sources

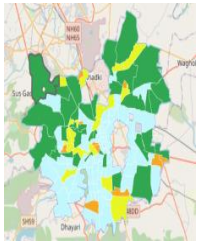
HITACHI
Inspire the Next

- **COVID hospitals and bed details**: list of hospitals and #beds allocated to treat COVID-19 patients updated on a day-to-day basis.
- **Additional hospitals**: list of additional healthcare facilities and beds available in PMC-Pune region.
- **Demographics**: Publicly available information like - number of literates, number of children below age 6, working population and average family size.
- **Hotspots**: COVID-19 containment zones in PMC-Pune.
- **Environmental health scores**: AQI, tree and population count based Health Score.









4.3.2 Risk Assessment score formulation

HITACHI
Inspire the Next

- Number of active cases in the ward can be an indicator of ward risk.
- Risk score of a ward is highly correlated to Active Patients in a ward.

$$\begin{aligned} Risk\_Sum = {} & 0.1 * (Patients\ without\ oxygen) \\ & + 0.2 * (Patients\ with\ oxygen) \\ & + 0.3 * (Patients\ without\ ventilator) \\ & + 0.4 * (Patients\ with\ ventilator) \end{aligned} \quad \ldots[6]$$

$$Initial\_RAS = \frac{Risk\_Sum}{Ward\_Population} \quad \ldots[7]$$

- Studied feature correlation with active patients and selected relevant features. Selected features and historical data was used to train the model.

Feature#	Feature Name
1	Health Score
2	Houses
3	Literate Population
4	Population under 6
5	Family Size
6	Working Population
7	Hospitals
8	Oxygen beds
9	Beds without oxygen
10	ICU ventilator beds
11	ICU beds without ventilator
12	Additional beds
13	Hotspots

4.3.2 Risk Assessment score formulation cont..

HITACHI
Inspire the Next

- Different algorithms were used to predict the Risk Assessment Score (RAS) and Gradient boosting was selected.

TABLE IV

PREDICTION ERRORS WITH MACHINE LEARNING MODELS: LINEAR REGRESSION (LR), K-NEAREST NEIGHBORS (KNN), RANDOM FOREST (RF), GRADIENT TREE BOOSTING (GTB)

Dates	LR	KNN	RF	GTB
12-01-2021	23.46	5.49	6.23	5.55
13-01-2021	23.69	3.68	4.38	3.56
14-01-2021	26.11	5.98	6.57	5.87
15-01-2021	23.31	4.72	5.23	4.39
16-01-2021	22.38	4.96	5.53	4.66
17-01-2021	23.70	5.44	6.06	5.35
27-01-2021	32.51	12.75	11.77	10.93
28-01-2021	31.55	11.38	10.34	9.57
29-01-2021	19.67	6.48	4.45	4.18
30-01-2021	19.75	6.79	5.3	4.78
31-01-2021	20.14	7.60	6.11	5.6
01-02-2021	19.52	6.22	4.73	4.19
02-02-2021	16.21	2.82	3.97	4.1
03-02-2021	16.92	3.43	4.57	5.02
04-02-2021	17.11	2.49	3.63	4.07
Average error	**22.40**	**6.01**	**5.92**	**5.45**

4.3.2 Risk Assessment score formulation cont..

HITACHI
Inspire the Next

TABLE V
WARD LEVEL COVID-19 RISK ASSESSMENT SCORES

Ward Name	Risk Score	Risk Zone
Dhanori	1.9	Very Low
Bopodi	3.3	Very Low
Yerwada Prison Press	6.7	Very Low
Bibvewadi	14.5	Very Low
Renuka Swarup Prashala	16.1	Very Low
Kharadigaon	19.2	Very Low
Yashwantrao Chavan Natyagraha	23.2	Low
Bharti Vidhyapeeth	23.5	Low
Dinanath Mangeshkar Rugnalaya	42.7	Moderate
Baner-Balewadi	81.9	Severe

- Categorized wards risk zones based on risk scores.
- Wards can be further prioritized based on the scores.

TABLE VI
CORRELATION BETWEEN RISK SCORE AND ACTIVE PATIENTS

Ward Name	Predicted_RAS	Active Patients*
Dhanori	1.9	7
Bopodi	3.3	23
Bibvewadi	14.5	43
Yashwantrao Chavan Natyagraha	23.2	76
Dinanath Mangeshkar Rugnalaya	42.7	145
Baner-Balewadi	81.9	599

*per 10000 population.

- The predicted risk scores of previous day highly correlate to indicate the active patients of the current day.

4.3.3 Visualization – Risk Assessment

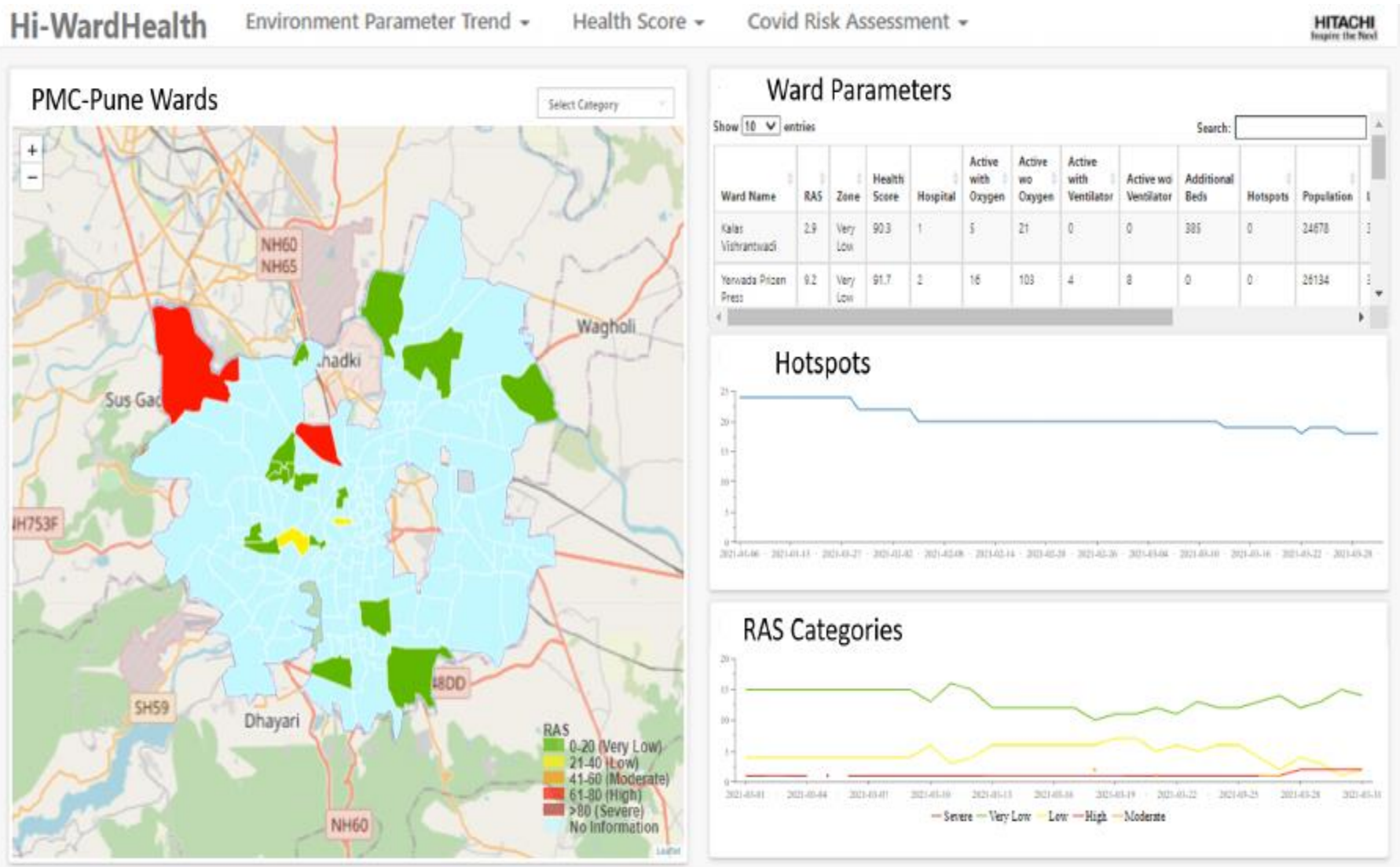


Figure 6. Pune-PMC COVID-19 Risk Assessment Scores

4.3.3 Visualization – Ward level Risk Assessment

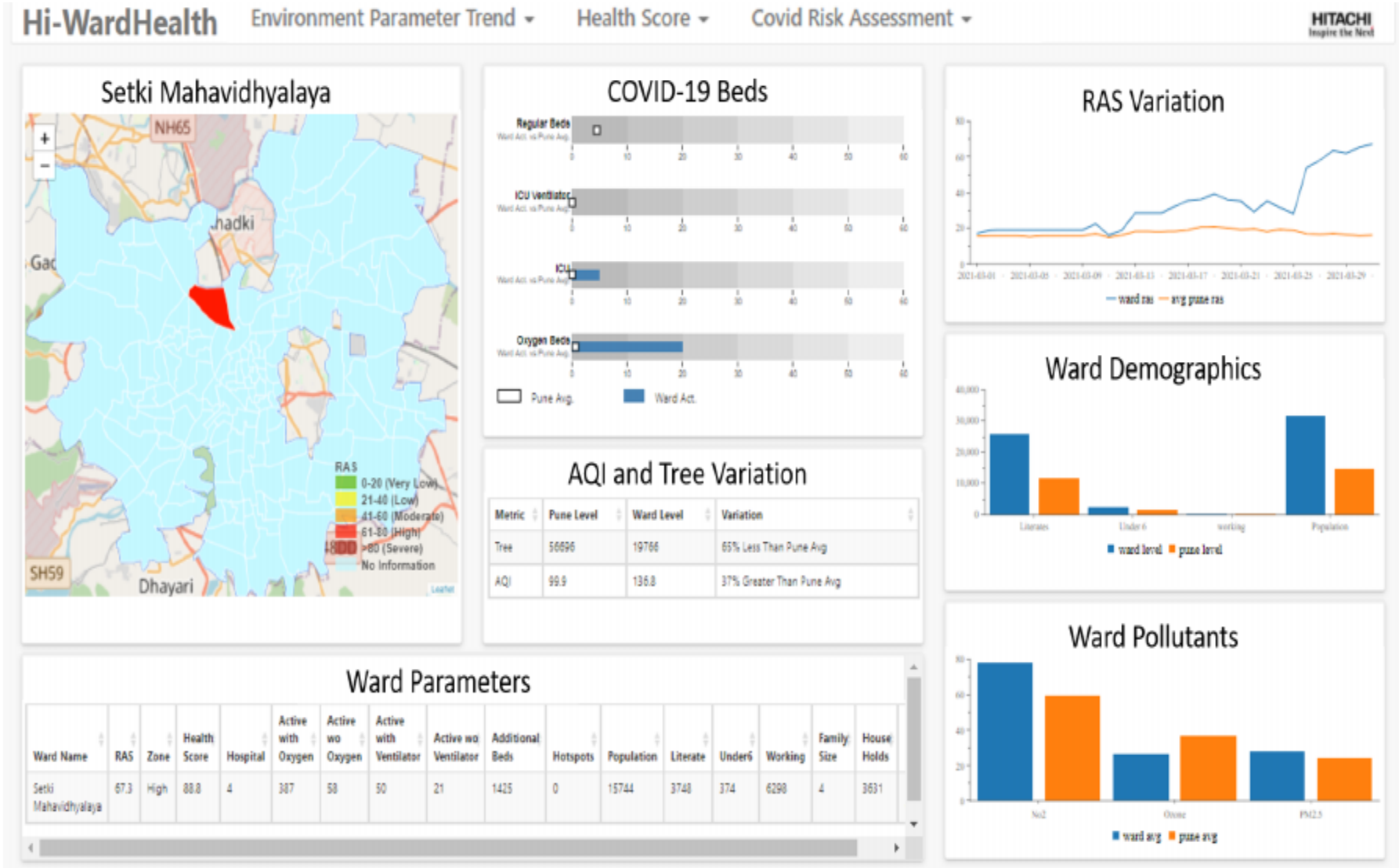


Figure 7. Pune-PMC Ward level COVID-19 Risk Assessment Score

5. Results and Discussion

HITACHI
Inspire the Next

- Health scores were derived for 63 wards.
  (AQI: 33 sensors, Tree data: 73 wards, Population census:144 wards.)
- Out of the 63 wards, 12 showed moderate, 15 satisfactory and 36 showed good health scores.
- Recommendations to improve green cover of an area is provided
- Environmental parameters and pollutant concentrations changed significantly during the lockdown.
- COVID-19 risk assessment scores were calculated for over 20 wards and categorized into risk zones.
- Detailed ward level analysis for the environmental and medical infrastructure facilities is provided.

6. Conclusion and Future work

HITACHI
Inspire the Next

- Dynamic numerical scores help understand the relative ward behaviour.
- Environmental health scores help improve the green cover of a place in the long run.
- Risk assessment scores can help authorities prioritize wards and plan actions. (Improving #beds, planning vaccination drive so on.)
- When the overall condition of a ward improves, its ability to handle situations like the pandemic increases.
- Additional features can improve the risk score formulation. (Average hospitalization duration of patients, number of direct contacts, COVID-19 transmission related information etc.)
- Interpretations of visualizations can be further enhanced with Human Computer Interaction techniques.
- Work can be extended to other smart cities where similar data is available.