An Evaluation of the Air Quality in the Neighborhoods Surrounding the KMML Industrial Area in Chavara using Machine Learning Techniques

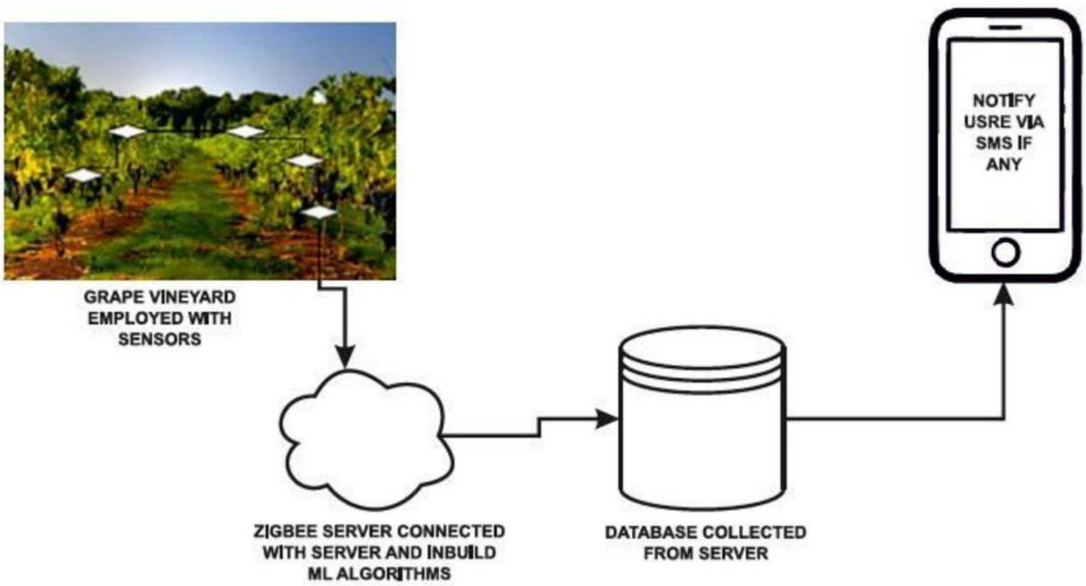
Susymary Johnson[1*], Deepalakshmi Perumalsamy[2]

[*1]Research Scholar, Department of Computer Applications, Kalasalingam Academy of Research and Education, Krishnankoil, Virdhunagar District, Tamil Nadu, 626126, India, susymaryjohnson8@gmail.com

[2]Professor, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Krishnankoil, Virudhunagar District, Tamil Nadu, 626126, India, deepalakshmiperumalsamy234@gmail.com

* Corresponding author E-mail address: susymaryjohnson8@gmail.com

Graphical Abstract



***Abstract*—** Researchers analyzed air quality data from residential areas near the Chavara manufacturing industry in southern India from 2010 to 2022. They compared air quality in the industrial area with surrounding residential areas, focusing on pollutants like SO2, NO2, RPM, Free Cl2, and TSPM. To predict air quality indicators accurately, they used a combination of artificial intelligence techniques. By comparing error metrics across different approaches, they identified the optimal method for accurate predictions. The study employed machine learning algorithms and SMOTE to predict Air Quality Index (AQI) levels in Indian cities. The ensemble method outperformed individual classifiers like K-nearest neighbor, Support Vector Machine, and Decision Tree, achieving higher accuracy, precision, recall, and F1-score, indicating its effectiveness in predicting AQI levels. The study also highlighted the importance of data preprocessing and balancing for improved prediction accuracy.

Keywords— Air Quality Evaluation, Machine Learning Techniques, Industry 4.0, and Chavara Kerala

I. Introduction

Outdoor as well as indoor air have been polluted with particulate matter, a recognized danger to health. The objective of this study was to contrast the amounts of particulate matter (PM) released by various goods and to evaluate the impact of outdoor as well as interior air quality caused by combust and/or non-combustion items utilized in the outdoors utilizing PM as a measure [1]. In order to evaluate the efficacy of air quality management policies, it is necessary to isolate as well as quantify the influence of meteorological elements, since these circumstances have a significant bearing on the dispersion as well as transportation of airborne pollutants. We used SO2, NO, NO2, NOx, CO, PM2.5, PM1, as well as PM10 observational data collected at the Chaoyang site, an urban location in Beijing, as well as atmospheric variables to conduct a multiple linear regression study [2] of the effect of the weather as well as control evaluates on the air quality in Beijing, China throughout APEC 2014. Clinical strategies for additional results from cardiac genes are addressed as well. Since cardiac is linked to 27% of all alleles included in the ACMG table for recurrence in additional results, this is important information. As no systematic method to study the literature has been used, the suggestions provided below are regarded as expert opinion in accordance with current ACMG policy. In order to better treat for patients as well as supervise those at risk in their families, cardiac testing for genes is recommended [3]. Quality of life (QOL) is significantly impaired as well as symptom load is high for patients with bronchiectasis. Utilizing baseline data from persons sequentially included in the planned German bronchiectasis registry PROGNOSIS [4], we offer psychological validity of a German version of a specific to the disease Quality of Life Questionnaire-Bronchiectasis (QOL-B), edition 3.1. There were a total of 904 patients whose QOL-B values could be considered. No flooring or ceiling impacts were detected that were meaningful. Cronbach's alpha for every level ranged between 0.73 and 1.00, indicating acceptable to outstanding internal coherence. Spontaneous sensorineural hearing loss (SSHL) has unknown causes. In Korea, there was no correlation between average levels of particulates with diameters of 2.5 m or fewer (PM2.5) and SSHL, but there was an inverse correlation between the highest levels of PM2.5 and SSHL. Two weeks of being exposed to NO_2 enhanced the likelihood of developing SSHL. There were not many long-term SSHL impacts from being exposed to air pollutants. We wanted to assess the link between SSHL and air pollution contact to ascertain whether or not contact to air pollutants resulted in long-term consequences [7]. Nail salon technicians are in great demand around the country, and the larger New York City region has the largest concentration of the profession. Professionals at nail shops are often exposed to harmful VOCs as well as phthalate [8]. A traumatic brain injury (TBI) might have far-reaching consequences, both for the victim and their loved ones. The term "health-related quality of life" (HRQoL) refers to an individual's level of contentment across several dimensions of health and well-being [9]. Across the globe, researchers and healthcare providers utilize the Quality of Living with trauma to the brain instrument to assess HRQoL following TBI. Clinicians meaningful assessments and choices regarding an afflicted person's health may be made by contrasting the individual's scores with standard deviations from the general population when assessing expressed HRQoL after TBI [10, 11]. Although the QOLIBRI has had worldwide use, only recently have value references been made accessible in the UK as well as Netherlands. The state of the air in the Fenwei Plains has worsened throughout the winter as well as holidays in the past few years due to numerous haze outbreaks. The haze was caused by a combination of factors, including coal burning, industrial pollution, as well as the unique environment and location of the Fenwei Plain. The specimens were taken from Feb 2-13, 2019, around the spring festival in Linfen, Fenwei Plain. Inductively linked plasma mass spectrometry(ICP-MS) was used to identify the 13 elements (Li, Be, Ti, Rb, Sc, Y, La, Ce, Zr, V, Tl, U, as well as Sn) present in PM2.5. Brief Synopsis The proof for innovative medicines' effects on survival rates has grown substantially in the past couple of decades [12]. Less is understood, however, about how innovative treatment combinations influence patients' quality of life. Rheumatoid arthritis, also known as (RA) is a form of autoimmune condition with a mysterious origin. Air pollution is frequently suggested as an indicator for illness, although this connection hasn't been sufficiently well investigated. The impact of environmental factors on RA symptoms was investigated. Patients with RA had their information culled from the KRRD (Kuwait Registry for Rheumatic Disorders). Throughout hospitalizations among 2013 and 2017 [13], patients' illness activity was evaluated utilizing the Disease Activity Score with 28 Considered Connections (DAS-28) as well as the Clinical Disease Activity Index (CDAI). Brief Synopsis Health-related quality of life (HRQoL) as well as sexual activity is two aspects of health that are affected by cancer treatment. Human Reproductive Quality of Life (HRQoL) as well as Sexual Function are Decreased in Individuals using Neuroendocrine Neoplasm (NEN) Due to Prolonged

Disease as well as High Tumor Burden Requiring Multiple Therapies with Various Mechanisms of Actions. Additionally, tumor generation and dissolution of peptides as well as biogenic amines are processes that induce numerous clinical diseases, such as the carcinoid condition, might be least partially responsible for the complexity of manifestations. Patients with NEN may benefit from peptide receptor radionuclide therapy (PRRT) [14]. Patients with NEN undergoing PRRT have shown a boost in HRQoL as measured by self-assessment instruments such the EORTC QLQ-C30 as well as QLQ-GINET21, according to observational investigations as well as randomized clinical studies. These surveys have merely scratched the surface of the effect on sex life. In this article, we'll look at what we know currently about how PRRT affects patients with NEN in terms of HRQoL as well as sexual activity. Radioligand treatment, or peptide receptor radionuclide therapy (PRRT), is a very successful antitumoral treatment for people suffering from neuroendocrine neoplasm (NEN). Health-related quality of life (HRQoL) is enhanced, as determined by patient-completed surveys. [15].

Here are some common pollutants and their effects:

- Particulate Matter (PM10 and PM2.5): These particles can penetrate deep into the lungs and bloodstream, causing respiratory and cardiovascular issues.
- Sulfur Dioxide (SO2): SO2 can irritate the respiratory system and aggravate conditions like asthma.
- Nitrogen Oxides (NOx): NOx can contribute to the formation of ozone and fine particulate matter, leading to respiratory problems.
- Carbon Monoxide (CO): CO can reduce the blood's ability to carry oxygen, leading to fatigue, chest pain, and impaired vision.
- Ozone (O3): Ozone can irritate the respiratory system and worsen asthma and other lung diseases.

The proposed method offers a novel approach by combining SMOTE for class imbalance correction with ensemble learning techniques, enhancing the accuracy and robustness of Air Quality Index (AQI) prediction models.

Contribution of the work:

The work presents a novel approach that combines SMOTE for class imbalance correction with ensemble learning techniques for predicting Air Quality Index (AQI) levels. This approach improves prediction accuracy, robustness, and generalization performance compared to existing methods, offering a more effective solution for air quality monitoring and management.

II. Related Work

In [16], the authors existing measurements of particulate matter (PM) alongside lightweight diameters of 10, 4, 2.5, as well as 1 m taken at the same time inside (smoking ban) as well as outside of a college library on two separate the weekdays. One conventional cigarette (9920 g m3), one e-cigarette (9810 g m3), as well as three conventional cigarettes (8700 g m3) all contributed significantly to a decline in external quality of air, as did other combustion as well as non-combustion pollutants. Research has shown that indoor PM1 levels rise after and during outdoor smoking and vaping activities. The findings emphasized the necessity to revise smoke-free rules to include non-combustion merchandise, particularly for outdoor areas. Health promotion programs that raise public awareness of the risks of smoking in public places are also crucial.

In [17], the researchers present a linear regression model with multiple variables to forecast concentrations of pollutants using parameters of weather variables under the assumption of unchanged emissions circumstances; the simulation outcomes showed an estimation coefficient (R2) in a range of 0.494 to 0.783. According to our findings, air quality control measures lowered levels of SO2, NO, NO2, NOx, CO, PM2.5, PM1, as well as PM10 by 48.3%, 53.5%, 18.7%, 40.6%, 3.6%, 34.8%, 28.8%, as well as 40.6%, respectively, while the weather lowered levels by 1.7%, -2.8%, 18.7%, 4.5%, 18.6%, 27.5%, 30.6%, as well as 35.6%. Beijing's condition of air has risen dramatically thanks to a mix of favorable weather and effective pollution controls throughout the APEC summit. Reductions in sulfur dioxide as well as nitrogen oxides were driven mostly by control efforts, whereas reductions in carbon monoxide were driven primarily by climatic variables. The decrease in PM2.5 concentrations may be attributed nearly equally to weather-related factors as to management methods. The impact of climate on pollution levels was investigated using the weighting technique. The findings demonstrated that many climatic parameters acted as determinants of pollution concentration.

The authors of this article of reference [18] created this paper to serve as a resource for clinicians seeking up-to-date recommendations on how to conduct a genetic examination of cardiac as well as how to handle additional results involving cardiomyopathy-related genes. Every of the four primary cardiomyopathies—dilated, hypertrophic, arrhythmogenic just ventricular, as well as restrictive—has a known genetic cause and a medically actionable course of action to increase surviving, decrease morbidity, and boost the quality of life. Cooperation between the Heart Failure Society of America (HFSA) as well as the American College of Medical Genetics and Genomics (ACMG) resulted in revisions to recommendations initially published in 2009. Teams of experts with relevant experience were tasked with reviewing relevant material and deciding on final suggestions. The characteristic is used to support the HFSA document's guidelines for relatives, phenotypic screening of in danger relatives, referrals to specialist centers if necessary, genetic counseling, as well as cardiac treatments. When heart is suspected, a thorough genetic examination including testing is warranted.

Although this authors' Declaration of Conflicts of Interests had been approved for publication at the moment it was provided, the version that is shown in [19] is the source. In addition, we failed to mention in the Recognition section that this paper is a shorter version of "Genetic Assessment of Cardiomyopathy—a Cardiovascular Association American Management Guideline," which was first published in the journal of Heart Fail. Articles PDF as well as HTML editions were recently updated to reflect the change.

Other than for the ones related to Social Functioning, Vitality, as well as Emotional Working levels, investigators in [20] found that the QOL-B increases differentiated among patients according to factors such as the number of previous pulmonary exacerbations as well as hospital stays, breathlessness, bronchiectasis seriousness index, lung function, stool volume, bacterial status, as well as the requirement for regular pharmacotherapy. Except for the Social as well as Emotional Function scale, we found moderate to great congruence among indicators of illness severity as well as QOL-B scores. Each of the scales had an intraclass correlate value of 0.84 or above, indicating a high degree of test-retest reliability over a period of two weeks. The Respiratory Problems measure had an MCID of 8.5, whereas the Social Functions scale had an MCID of 14.1. In a nationally representative sample of adults with a condition called the German edition of the QOL-B (version 3.1) exhibits high validity and reliability across tests. However, more research into the flexible nature of QOL-B levels is needed throughout registries follow-up.

Utilizing patient's information gathered from electronic medical records at Tri-Service General Hospital in Chicago among 2011 and 2019, the writers of [21] suggest an observational study. All available information about SSHL patients was obtained. We utilized data on the air quality from the Songshan station between 2011 and 2019. The relative hazards (RRs) of SSHL related with exposures to PM2.5, O3, as well as NO2 within 1 month were the primary endpoints. Complex models of time series with a dispersed lag were used to look at the connections among these variables. 3. Outcomes For every 10-unit increase in exposure to PM2.5 with a 7-day latency, the relative risk (RR) of SSHL was 1.195 (95% C.I. : 1.047-1.363). For every 10 unit rise in O3 with a 9-day latency, the relative risk of SSHL was 1.14 (95% confidence interval [CI]: 1.003-1.3). For a 10 unit rise in NO2 with a latency of 23 days, the RR of SSHL were 1.284 (95% C.I. : 1.05-1.57). (4) We conclude that there is a lag effect between exposure to air pollutants and the development of SSHL. We spoke about potential avenues of inquiry into biological theories and ways to fund additional study in this area. Our results need to be confirmed by larger research incorporating people of different nationalities as well as investigating the causal correlations we observed.

To describe related to work variables of the air quality inside in nail parlors in the Greater NYC region, investigators in [22] set out to analyze a combination of phthalates as well as volatile organic compounds that were present in these establishments. From Feb to May 2021, we used silicone wristbands to assess phthalate in the air as well as passive samples to test volatile organic substances at

20 nail shops in the greater NYC region that cater to Asian clients. The features of nail salons were also analyzed. We detected as well as quantified 31 VOCs, including six phthalate. The greatest quantities were found for Di(2-ethylhexyl) phthalate as well as Diethyl phthalate, out of a total of six phthalates. Toluene, d-limonene, methacrylate of vinyl, as well as ethyl acrylate all had much greater concentrations than the others. The use of phthalate as well as volatile organic substances was shown to increase with the amount of clients, the frequency with which fake nails (acrylic) were applied, and the number of pedicure and nail polish stations in use. There are a lot of individuals who work in the nail business, as well as more people who go to nail salons, so it's possible that a lot of individuals get exposed to these hazardous substances.

In [23], researchers prepare reference values for use in validating the QOLIBRI throughout Italy's general population. A total of 3298 members of the general public in good health were surveyed using an online variant of the QOLIBRI questionnaire. The findings were compared to those of 298 people who participated in the international longitudinal observation cohort CENTER-TBI study, conducted at Italian institutions. The QOLIBRI was evaluated for its psychometric properties as well as measurement consistency. Predictors of interest for HRQoL in the community at large were determined by a regression analysis. Percentages were supplied as reference values. Based on invariance of measurements analysis, the QOLIBRI was shown to capture identical HRQoL components in both an accurate representation of the Italian population as a whole as well as the Italian TBI cohort included in the Center-TBI project. Higher QOLIBRI scores corresponded with older age, more years of schooling, as well as no evidence of an ongoing medical condition, all of which point to improved HRQoL. Age-, sex-, education-, and chronic illness-adjusted reference values for the Italian population as a whole were supplied. These are highly suggested for use in healthcare and research employing the QOLIBRI rating in Italy.

Quantitative analyses of modifications in PM2.5 level of exposure as well as associated economic and health advantages were made using air quality computational modeling, health risk evaluation, as well as ecological valuation methods by the investigators in [24], who sought to comprehend the public health advantages of the Air Quality Action Plan enacted in Shanghai from 2013-2017. The percentage of people whose exposure to yearly PM2.5 concentrations less or greater than 35 g•m-3 has grown between 1.62% in the starting period to 34.06% in a control period. Ambient PM2.5 exposure was associated with a 13% lower mortality risk in the control year compared to the starting year. The positive effects on healthcare are estimated to be worth around 11.841 billion RMB, or about 0.55 percent (0.23 percent to 0.82 percent) of Shanghai's GDP in 2013. Protecting the wellness of individuals is aided by the action plan's execution. Health improvements in high-population, high-PM2.5 locations are most noticeable in neighborhoods on the outskirts of Shanghai City.

Clustering as well as backwards trajectory were employed to assess the combined weather information, regional and chronological distributions of pollutants, as well as potential source evaluation presented in [25]. During the time of collection, the mean amount of SO2 was 58.39 g•m-3, which is over the 24 h average mass content limit(50.00 g•m-3) of the national overall air quality standard (GB 3095-2012). O3, NO2, as well as CO averaged 52.15, 29.02, and 2.29 micrograms per cubic meter, correspondingly. The findings highlighted SO2 as the predominant contaminant. Dispersion from cities mostly affects NO2 and CO levels. Based on the retraced path, it seems that the Fenwei Plain's basin morphology is a major contributor of the haze. Based on an examination of the PSCF of soil reports, we know that northern Shaanxi, southern Gansu, as well as southern Ningxia—all which are heavily influenced by the rainy season climate—are the prospective dominating regions.

We explain the relevance to standard of life to individuals with severe prostate cancer and offer some of the most frequently employed measures to assess comfort of life in current randomized studies, as described by researchers in [26]. Furthermore, we talk about how simple it is for physicians to utilize these standardized questionnaires, as well as concentrate on how to enhance quality of life measurement in such clinical settings with advanced prostate cancer. Abstract Health-related aspects of life (HRQOL)-associated data is currently lacking, despite the reality that the treatment environment of terminal prostate cancer is fast changing as well as oncological advantages have been demonstrated for a multitude of novel medicines as well as indications. We highlight the most common instruments used to assess HRQOL in recent randomized trials, as well as talk about the significance of HRQOL for patients suffering from advanced PC (metastatic hormone-sensitive cancer of the prostate [mHSPC], spread castration-resistant prostate cancer [mCRPC], and non-metastatic castration-resistant prostate cancer [nmCRPC]). In addition, we talk about how simple it is for doctors to utilize these standardized questionnaires, how they should be used in the clinical setting, and how to enhance HRQOL assessment in various clinical settings of serious prostate cancer.

Utilizing pollutants from the air constituents (PM10, NO2, SO2, O3, as well as CO), the writers of [27] suggest an assessment of pollution in the air. Six surveillance sites in Kuwait gave information on pollutants in the air throughout the identical time frame, which was collected by the Kuwait Environmental Public Authority (K-EPA). For chained equations (MICE) approach was used to perform several accusations to approximate the missing atmospheric pollution information. Dates as well as patient governorate address were used to match up air quality information with patient records. STATA was used for the analysis of correlations, descriptive statistics, as well as regression analysis. A total of 1651 rheumatoid arthritis patients were analyzed, including 9875 follow-up visits. Using the DAS-28, we found that those who had been subjected to NO2 and SO2 had a higher risk of developing RA (=0.003, 95% CI: 0.0004-0005, p0.01), but those who were exposed to PM10, O3, as well as CO had no greater risk of developing the disease. In conclusion, we found a robust correlation among exposure to polluted air and the severity of RA symptoms. This research provides evidence that air pollution is associated with an increased risk of RA, and it implies that authorities should take more steps to address this public health issue.

The purpose of the narrative overview presented in [28] was to describe the state of research on the effects of PRRT on HRQoL as well as sexual activity in patients with NEN. We searched PubMed, Embase, as well as the American Psychological Association's PsycInfo databases for relevant articles. 15 studies met our criteria (12 for HRQoL & 3 for sexual activity). Global health status, disease-related anxieties, societal and mental health, as well as cancer-related symptoms including tiredness and diarrhea all improved significantly for individuals with NEN after PRRT therapy. Radioligand treatment had no effect on the severity or frequency of additional symptoms, including nausea, vomiting, dyspnea, constipation, or financial burden. Not as encouraging were the results on sexual functioning; only a small number of studies employing suitable questionnaires examined this problem in people treated with radioligand treatment. As a result, further research is required before any firm conclusions can be made on the effects of PRRT on sex behaviors.

Patients' specimens may be compared to reactive, as well as leukemia/lymphoma normal, references datasets via the use of unique data analysis techniques presented by the investigators in [29]. Such reference datasets need rigorous quality assurance (QA) measures during construction. In this study, we assembled a data collection as well as established a quality assurance (QA) approach to validate the EuroFlow Acute Myeloid Leukemia (AML) database structure, which is predicated on the EuroFlow AML eight-color panel, which consists of six distinct antibodies pairings as well as four backbone indicators. It included an overall of 1142 cases with AML and 42 samples from healthy bone marrow were used in this study. Multi-dimensional analysis of both backbone and tube-specific indicators were used for quality assurance on 803 instances of AML, as well as classical analysis comparing median as well as peak levels of expression was employed to contrast the results. Re-analysis of over 300 cases as well as a separate sample of 339 AML patients were used to validate the QA approach. The final group was evaluated, and the results showed the presence of distinct immunophenotypic tendencies across AML subgroups; this set of data therefore proves suitable for further investigation into the immunophenotypic diversity of AML. Our findings highlight challenges in developing large flow cytometric datasets and offer approaches for overcoming them. Furthermore, the offered method may aid in the creation of additional data bases, which in turn might aid in the creation of innovative instruments for (semi-) automated quality assurance and subsequently analysis of data.

Clinical care of individuals who have various illnesses as well as pharmacological prescription is complicated, as introduced by the authors in [30]. The research aims to determine whether tendencies of having many diseases are linked to higher rates of PIP and ADRs as measures of drug quality. Seven hundred and forty individuals over the age of 65 who were hospitalized owing to an aggravation of a chronic disease were included in a multicenter retrospective observational study. Data on demographics, health history, as well as adverse drug reactions (ADRs) or polypharmacy (many medications) were gathered. The existence, quantity, or particular kinds of PIP or ADRs were compared to multiple illnesses groups that had already been established (osteoarticular, psychogeriatric, mild chronic illness, cardiorespiratory). All of the clusters had statistically significant correlations. Both PIP (94.9%) as well as ADRs (48.2%), mainly resulting from anxiolytics as well as anti hypertensive were most common in the osteoarticular group, preceded by the mild chronic illness cluster, which was linked to ADRs from antihypertensives as well as insulin. Overall, signs from the cardiorespiratory cluster were superior than those from the psychogeriatric group, which revealed PIP as well as ADRs of neuroleptics. In the identified relationships provide credence to multiple medical conditions patterns as well as lend support to targeted medication review measures based on individual patient profiles. Therefore, it may be beneficial to hospitals to learn the connection among multimorbidity characteristics as well as drug quality measures. Number of Clinical Trial: NCT02830425.

The literature review provided insights into various aspects related to air quality prediction and management. It covered the use of artificial intelligence (AI) techniques, such as machine learning algorithms, for predicting air quality indicators, including the Air Quality Index (AQI). The review highlighted the importance of data preprocessing and balancing techniques, such as SMOTE, to improve the accuracy of prediction. The proposed system offers several advantages over existing systems. It integrates SMOTE for class imbalance correction, which improves the model's ability to handle imbalanced data, leading to more accurate predictions. Additionally, the use of ensemble learning techniques enhances the model's robustness and generalization performance. Overall, these improvements result in a more effective and reliable system for predicting Air Quality Index (AQI) levels compared to existing systems.

III. PROPOSED WORK

A. *Study Area Used*

Approximately 285 acres make up the Kerala Mineral as well as Metals Ltd (KMML business), which is located in Chavara close to NH-47. Sankaramangalam in Chavara, who taluk, approximately 32 kilometers from Kollam Main Railways Station, as well as Neendakara Port, which is approximately five kilometers to the southwest, make up the subject matter area, which also includes the immediate surroundings of the KMML industrial area. Chavara as well as Panama panchayats are both part of the Chavara taluk. Kerala Minerals as well as Metals Ltd. (KMML) as well as Indian Rare Earth Ltd. are two of the major employers in the region. Study stations S1, S2, S3, S4, S5, as well as S6 can be seen on an outline of their vicinity to the KMML the manufacturing industry in Chavara taluk, Kollam region (Fig. 1).

Six separate sites were analyzed over two separate seasons (summer as well as winter) among 2022 and 2010. One control facility (Station 6) in a harmless surroundings, i.e., 14 km separate of the KMML industrial region in the north westward guidance was chosen as well after an exploratory survey, and five test sampling locations (Train Station 1, Train Station 2, Train Station 3, Train Station 4, Train Station 5, as well as Train Station 6) in the vicinity were selected radially in the south, east, and north. S1 is 0.5 kilometers south of the KMML business district, S2 is 1 kilometers north, S3 is 0.5 kilometers east, S4 is 3 kilometers north, as well as S5 is 1 kilometers south-east. The Arabian Sea may be seen off the coast of the industry's western side. Throughout the research duration, the outside environment was monitored for the presence of several pollutants in both the summertime (May) as well as winter (December) months. These pollutants included sulfur dioxide (SO2), nitrogen oxides (NOx), as well as chlorine (Cl2). The mean values for every variable were calculated after triple collection of samples was carried out at every investigation site throughout the investigation's seasons. During the summer and winter months, a Portable Weather Tracker (Kestrel 4500NV, U.S.A.) measured air temperature, relative humidity, Heat index, Dew point, Barometric pressure, and wind speed in the research region. The environmental information has been collected at the specified stations hourly. Utilizing a High-Volume Airflow analyzer (Envirotech, Model APM 460 BL, Respirable Dust Sampler) with gases sample connection, at a rate of flow of 1.2 m3/min, air was sampled on a daily basis (every 8 hours) at all of the chosen stations in the summer (May) as well as winter (December) periods. The Air Sampling with Respirable Dust collecting adapter was used to collect samples via a Glass Fibre Filters (Whatman, GF/A, 20.3 25.5 cm) of the dispersed particulate matter, breathable particle matter (size 10 m). The impingers of the gaseous sample unit were filled with suitable absorbing solutions to capture the gases airborne chemicals (SO2, NOx, and Free Chlorine). Particularly important for human exposure, the entrance tube was set about 1.5 m above the surface of the ground.

The research area's typical summer (May) as well as winter (December) wind speed, air temperatures, relative humidity, warmth index, dew point, as well as atmospheric pressure have been collected utilizing a portable conditions tracker as well as are presented in Table 1. The average summer as well as wintertime temperatures of the surrounding air was recorded. The mean temperature of the air in the research region was 31.89 degrees Celsius in the summer and 30 degrees Celsius in the winter. In the summer, the average percentage of humidity was 77.09%, but in the winter it was just 76.4%. In the studied winter and summer seasons, the breeze blows from east to west.

Table 1. Weather data of the study area during Summer & Winter seasons.

S. No.	Measurement	Average of Recorded values	
		Summer	Winter
1	Speed of Wind	4mph	2mph
2	Heat	31.89°C	30°C
3	Humidity Relativeness	77.09°	76.4%
4	Index of Heat	43.2°C	37.4°C
5	Point of Dew	27.39°C	25.39°C
6	Pressure of Barometric	29.85inHg	29.85inHg

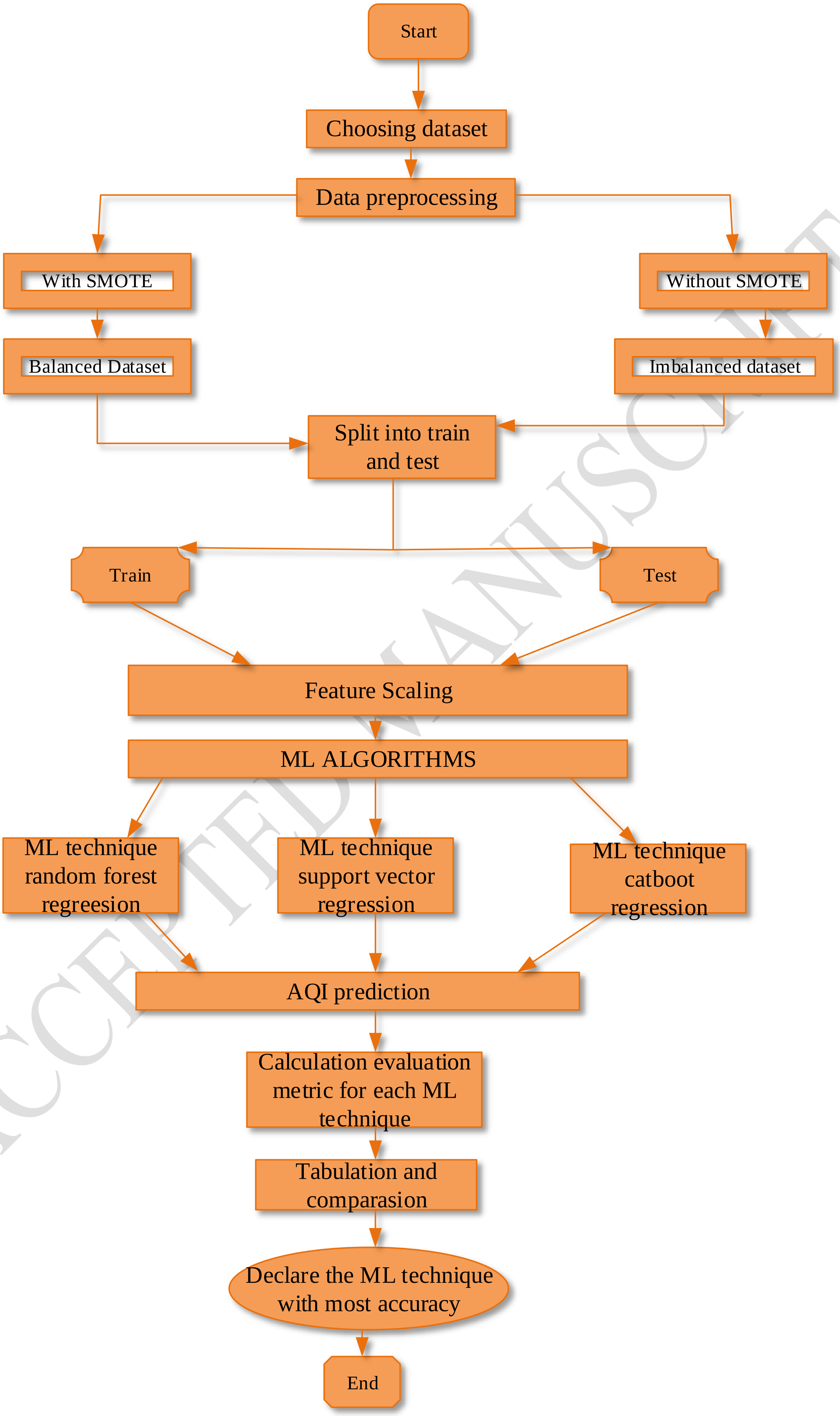

Start
Choosing dataset
Data preprocessing
With SMOTE
Without SMOTE
Balanced Dataset
Imbalanced dataset
Split into train and test
Train
Test
Feature Scaling
ML ALGORITHMS
ML technique random forest regreesion
ML technique support vector regression
ML technique catboot regression
AQI prediction
Calculation evaluation metric for each ML technique
Tabulation and comparasion
Declare the ML technique with most accuracy
End

Fig.1. Proposed Flowchart

B. *SMOTE Algorithm*

There is asymmetry in the original data set. The procedure for the simulated minority over sampling method (SMOTE) is utilized to rectify the unbalanced data collection. The technique uses over sampling to improve accuracy. To guarantee that every label for each class gets the same amount of rows in a set or the same number of columns as necessary, additional rows are added to the database if necessary. An unbalanced data collection contains asymmetry. A skew range of classes results from an unbalanced dataset, which has several effects on the reliability of the model's predictions.

This means that a balancing of the data is essential. The process of overs using a positive label is one way to increase precision. In this study, oversampling is accomplished with the help of SMOTE. Using a neighbor-based approach, the SMOTE method boosts the representation of underrepresented groups within a dataset.

After that point, it doubles the frequency with which the underrepresented group (the positives) appears, reaching six to twelve times. It helps balance out data sets to boost the efficiency of algorithms while preventing problems with overfitting. Typical implementations of SMOTE involve locating a feature vector as well as its nearest neighbor, computing the distinction among the two, increasing it by an arbitrary number from zero to one, locating an additional point on the line section after adding the selected number to the vector of features, and so on. SMOTE is preferable to making copies that are slightly off than the original information since it generates completely new data pieces.

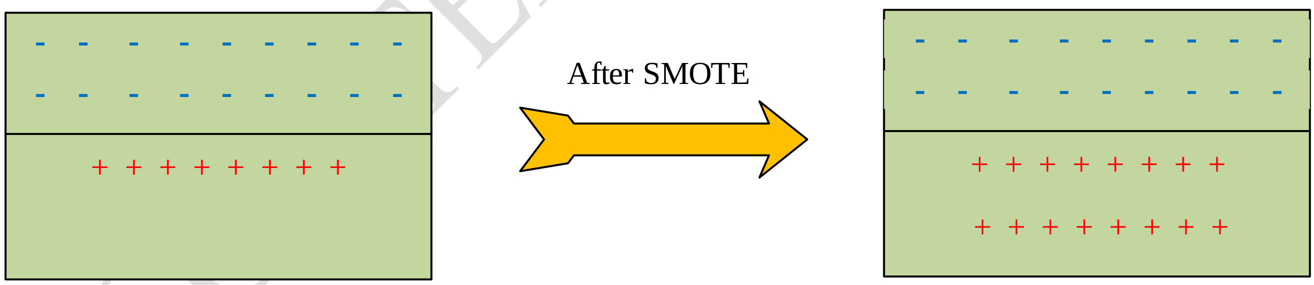


Fig.2 SMOTE Model

In this research, we use the SMOTE method to balance dataset in order to increase model accuracy. A skewed distribution of classes due to an uneven dataset will lead to inaccurate predictions. Balanced datasets lead to more accurate models, more balanced reliability, as well as a better balanced detection rate. Thus, SMOTE is used to achieve this goal and enhance precision. SMOTE is advantageous since it does not generate identical numbers instead making synthetic data points that deviate from the real data points by a small amount. This approach helps get over the over fitting problem brought on by random sampling by generating instances that are comparable to the minority instances that already exist.

Classifiers' generalization skills are enhanced by SMOTE because it generates limits on decisions that are both less specific and more expansive.

C. *Ensemble Machine Learning Classification*

In artificial intelligence, a combination framework is a technique where numerous models are combined to provide a superior result than that of just one model. It is the goal of the ensembles predictive framework to either reduce variance (through bagging) or bias (by boosting) or increase prediction quality (via stack).

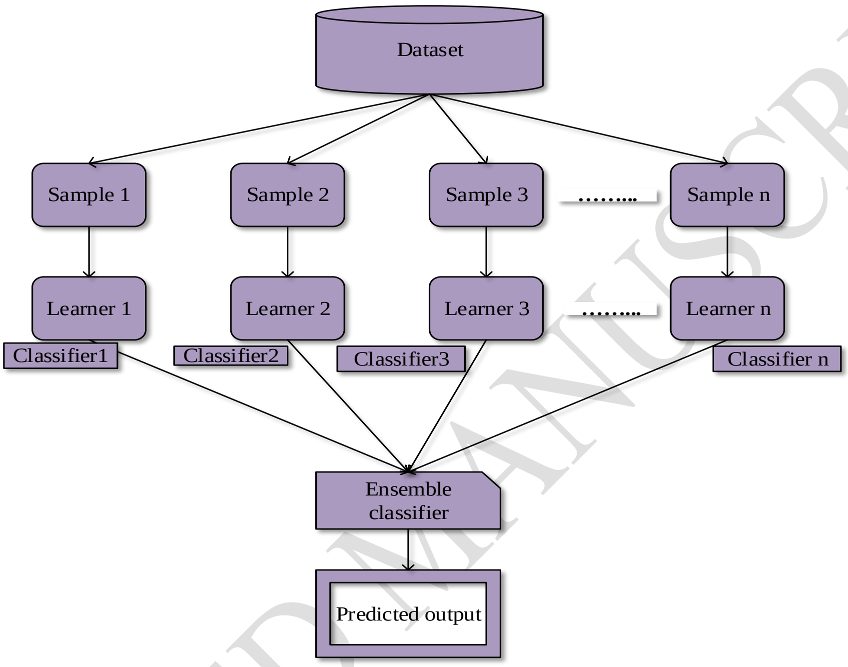


Fig.3 Ensemble Machine Learning Model

A. *SVM*

The goal of a support vector machine (SVM) is to form a function with an objective, then locate a partition hyper plane that may fulfill the class criterion, all while adhering to the notion of reducing structural risk.

(x_i, y_i) = Separable linear dataset.

$i = 1, \dots, n, x \in R^d$ and $y \in \{+1, -1\}$ = Label of class.

$\omega \cdot x + b = 0$ = Partition hyper plane.

Where

ω =Partition of normal vector hyper plane.

b =Offset hyperplane.

A partitioning hyper plane to create the bilateral blank region that is as far away from the location in the training data set as feasible is optimal, i.e., $2/\| \omega \|$, The search for the maximum, which may be described as,

$$\text{Minimize } \phi(\omega) = \frac{1}{2} \| \omega \|^2 \qquad (1)$$

The following is an example of a constraint condition.

$$y_i(\omega \cdot x_i + b) \geq 1 \qquad (2)$$

In this case, the Lagrange function is defined as:

$$L(\omega, b, \alpha) = \frac{1}{2}(\omega \cdot \omega) - \sum_{i=1}^{n} \alpha_i(y_i(\omega \cdot x_i + b) - 1) \qquad (3)$$

Two conditions for the following subject, i.e., $\sum_{i=1}^{n} y_i \alpha_i = 0$ as well as $\alpha_1 \geq 0$, Consequently, the formula below may be used to locate the lagrange function's minimum.

$$\max Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \qquad (4)$$

The best definition of the class function is as follows.

$$f(x) = \text{sgn}\left((\omega^* \cdot x) + b^*\right) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x) + b^*\right) \qquad (5)$$

Nonlinear mapping is used in the case of nonlinear reparability $\kappa(x)$ may be used to transform x into a linear separable feature space in a higher dimension.

B. *Decision Tree*

The choice tree's base node is constructed first, followed by its child nodes. The knowledge is organized into categories, with the nodes standing in for the defining qualities and attributes that serve as choice points. Nodes are connected to one another at various levels via branches, which stand for the various judgments reached after checking the node's properties. Using a tree data arrangement as well as an if-then expression, it is a guided machine learning approach.

The Desmond Density measure for data abundance is the basis of the choice tree technique. If P is the probability distribution, then $P = (p1, p2, P_3, \ldots, Pn)$, In the equation that follows, where (Pi) is the chance that the number (i) would emerge throughout the process, we can calculate the information contained by this distribution, which we name the Entropy of P with a sample data set S.

$$\mathrm{P} = (\mathrm{p1}, \mathrm{p2}, \mathrm{P3}, \ldots, \mathrm{Pn})\ \mathrm{E}_{13} \qquad (6)$$

$$\text{Entropy } (P) = \sum_{i=1}^{n} PiX\log(Pi) \qquad (7)$$

E14

$$\text{Gain } (P, T) = \text{ Entropy } (P) - \sum_{j=1}^{n} (PjX \text{ Entropy } (Pj)) \qquad (8)$$

E15

Where Pj is a complete list of all possible values for (T)

C. KNN

An alternative way to think about the kNN method is as a system of votes, where a fresh data point's classification is decided by the majority of the class labelling of its closest 'k' (in which k is an integers) neighbours in the space of features. Consider a hypothetical election in a tiny town of a few hundred people, in which you must choose a party to represent you. The easiest way to find out which party in politics your neighbors favor is to just ask them. You are more likely to cast a vote for the Republican Party if your 'k' closest neighbors are members of that party. This is analogous to the way the kNN method works, where the information point with the fewest neighbors is assigned a class identity with the highest probability.

Let's go in further with a different illustration. Just pretend that you have some information regarding fruits like pears as well as grapes. The size and the shape spherical the fruit is also factor towards the final rating. To see this, you choose to create a chart. You can also use this method to identify an unknown fruit by plotting it on the graph as well as calculating its distance to the k (some number) closest spots. Using the coordinate system shown below, I have absolute certainty that we are dealing with a pear since the three closest coordinates all represent pears. Using the four closest points, we can be 75% certain that this is a pear since three of them are pears and one is a grape. Later in this post, we'll discuss the various distance measurement techniques as well as how to get the ideal value for k.

The total amount of neighbours used by the KNN method is defined by the value of k, making it a pivotal variable. In the k-nearest neighbors (k-NN) technique, the quantity of k must be determined by analyzing the data at hand. A greater number of k is preferable if the provided data is more prone to anomalies or noise. To prevent categorization draws, it is preferable to use an odd number for k. To determine the optimal k for a certain data set, cross-valid techniques may be used.

D. Classification

Computation of Air Quality Index: The five levels of the established Indian the air quality index reflect varying degrees of air pollution:

- **Good:** The AQI for a given area might range from zero to one hundred. There is minimal to no danger from pollution in the air, as well as the air condition is good.
- **Moderate:** Among 101 to 200, that's a community's AQI. The air is generally safe to breathe, although a tiny proportion of individuals may have some mild health effects from a few specific contaminants. Some persons may be extremely susceptible to ozone and have respiratory problems as a result. Sensitive populations might experience health impacts with AQI levels among 201 and 300.
- This suggests that they will have milder effects than the overall population. For instance, exposure to ozone poses a larger threat to persons who already have lung illness, and particle pollution poses a bigger threat to those who already have respiratory illnesses or heart disease. Whenever the AQI falls within this range, there is little risk of harm to the general population.
- **Very Poor:** When the Air Quality Index (AQI) ranges from 301 as well as 400, it is possible for everybody to feel the consequences on their health. Health impacts may be more severe for people in vulnerable populations.
- **Severe:** A health warning is issued for the whole population if the value is between 401 and 500. The following Eqs may be used to determine the individual sub-indices that make up the IND-AQI. (1) and (2)

$$q = 100(V/Vs) \qquad (9)$$

Where,

q =Rating Quality;

V = Values of observed parameters and

$V\,S$ = Recommended value of standard parameter (CPCB, MoEF, 1998, 2009).

If' 'n' parameters are taken into account, the Air Quality Index (AQI) is calculated as the Arithmetic Mean of these 'n' Quality Ratings.

Geometric mean, $\mathrm{g} = \text{anti} \log\{(\log \mathrm{a} + \log \mathrm{b} + \cdots \log \mathrm{x})/\mathrm{n}$

Where a, b, c, d, x = air quality ratings of varying levels; and n = rating scale for air quality n.

The sum of all the sub-AQIs is presented as the IND-AQI, the overall AQI.

Oversampling, particularly using techniques like SMOTE, involves creating synthetic samples for the minority class by interpolating between existing samples. This helps balance class distribution and prevents the model from being biased towards the majority class, improving its ability to generalize and make accurate predictions for the minority class.

IV. RESULTS & DISCUSSION

Training and evaluation sets will be created when the dataset is cleaned, reduced, and set up to our specifications. To allow for the methods to be readily applicable in a real-world use case, the aim is to employ the simplest, most simplistic implementations possible. To determine which of these three techniques is more precise, we'll use a variety of criteria to make comparisons among them. By comparing several approaches to predicting the AQI, we may learn more about our options as well as choose the one that best fits our needs. We will also use the SMOTE method to compare the precision levels achieved using an uneven as well as an even database. Therefore, the technique is an ordered procedure, with the first stage being the identification and preparation of an appropriate dataset. After that, SMOTE is used for some more data preparation to get the dataset closer to parity. In order to highlight any modifications for efficiency that may develop owing to weighing, both balanced and unbalanced datasets will be retained and utilized. After that, the training set is often divided into training sets in order to train the predictions and assess their accuracy against actual data, as is customary in automated learning procedures. Normalization as well as scaling of features are performed.

Step 1. Choosing a dataset
We met our needs by selecting a large database from Kaggle as well as obtaining its corresponding CSV file.

Step 2. Data preprocessing
Bangalore, Kolkata, Hyderabad, as well as New Delhi cities were pulled from the original database after it was cleaned up in data processing. It is crucial to examine the levels of pollution in various metropolitan centers in India due to the fact that they are the primary sources of pollutants in the country. The pollution levels may be roughly estimated in these larger cities due to their substantial population densities. The data sets were cleaned by deleting rows with values that were null, as well as the property xylene was eliminated since its column entries were blank in all of the four cities considered. Unneeded, insignificant, or incorrect information is purged using Microsoft Excel.

Step 3. Applying the SMOTE algorithm

The category imbalance in the AQI_Bucket results is corrected using a synthetic minority over sampling method (SMOTE) once the dataset has been cleaned. It took 3, 11, 9, as well as 24 manual iterations, respectively, to establish a good degree of balance in Delhi, Bangalore, Kolkata, as well as Hyderabad. This is done so that the final dataset is more representative of the whole population.

Step 4. Not applying the SMOTE algorithm

Here, the artificial minority over sampling method (SMOTE) is utilized right after the dataset has been cleaned of extraneous, irrelevant, and incorrect information.

Step 5. Splitting of the dataset

A ratio of 80:20 separates the datasets into test and training data. These will be put to use in both the training and testing phases of the model's development. Machine learning systems' predictions are validated against the raw data to improve accuracy.

Step 6. Training the dataset

Using 80 percent of the information to train with has been shown to be optimal in empirical studies. To create the test and instruction sets, the data is sampled at random. It has achieved widespread popularity and acceptance.

Step 7. Testing the dataset

The best outcomes may be achieved, according to empirical research, by using the remaining twenty percent of the information for the purpose of testing. To create the test as well as training sets, the data is sampled at random. It has achieved widespread popularity and acceptance.

Step 8. Feature scaling

In an effort to render the information usable as well as consistent, the information has been standardized. The Scikit-Learn Library's Standard Scaler was utilized for this purpose. The characteristics are normalized by dropping the mean and adjusting the variance down to the unit level.

Step 9. Applying machine learning (ML) techniques

Different algorithms, including random forests regression, support vector regression, as well as CatBoost regression analysis, are utilized to predict the index of air quality after standardizing the variety of characteristics in the data sets. These methods are subsequently contrasted to determine which one provides the greatest precision level for every city.

Step 10. Applying ML technique-random forest regression

To solve problems with regression and classification, guided machine learning algorithms like random forests are used. Using the average for regression and the overwhelming vote for categorization, it generates trees of decision from multiple data points. Accurate and straightforward forecasts are generated by a forest of randomness. Big data sets can be managed efficiently.

Step 11. Applying ML technique-support vector regression

Regression issues may be solved with the help of support vector regression, a machine learning method that is supervised approach. It may be used for forecasting discrete values. Finding the optimal regression line is crucial to SVR. The hyper plane with the greatest number of points is the most appropriate one for the SVR. Because of SVR's adaptability, we can choose our own tolerance for modeling error.

Step 12. Applying ML technique-CatBoost regression

CatBoost regress is a guided artificial intelligence technique that uses gradient-boosted trees of choices. Several decision chains are gradually built throughout training. The primary goal of boost is to produce a strong, competitive prediction model via greedy search by gradually integrating an abundance of weak models or algorithms that just slightly outperform chance. It employs symmetrical trees, which makes the inference process fast, and improving methods, which help reduce overestimation while increasing the validity of the model.

Step 13. AQI prediction

The AQI quality for every town is assessed with the use of artificial intelligence algorithms. The levels of precision in each of the four cities are tallied and shown graphically.

Step 14. Calculation of evaluation metric for each ML technique

support vector regression, random forest regression, and CatBoost regression are assessed using accuracy, precision, recall and F1-score.

1334x12 table

	1 SAMP_DAY	2 MONITOR_ST	3 SO2_preprocesse	4 NO2_preprocessed	5 PM10_preprocessed	6 SPM_preprocessed	7 PM10_SubIndex	8 SO2_SubIndex	9 NO2_SubInd	10 SPM_SubIndex	11 AQI	12 AQI_BUCKET
1	NaT	'Chavara'	80	24	225	565	183.3333	100	30	417.8571	418	3
2	NaT	'Chavara'	40	24	230	600	186.6667	50	30	442.8571	443	3
3	NaT	'Chavara'	60	32	145	495	130	75	40	367.8571	368	4
4	01/13/2010	'Chavara'	60	52	285	610	235	75	65	450	450	3
5	01/15/2010	'Chavara'	160	30	355	420	306.2500	126.6667	37.5000	314.2857	314	4
6	01/18/2010	'Chavara'	40	26	250	250	200	50	32.5000	150	200	1
7	01/21/2010	'Chavara'	120	18	190	360	160	113.3333	22.5000	366.6667	367	4
8	01/24/2010	'Chavara'	120	14	330	470	280	113.3333	17.5000	350	350	4
9	07/22/2010	'Chavara'	160	34	305	490	255	126.6667	42.5000	364.2857	364	4
10	07/26/2010	'Chavara'	80	22	130	615	120	100	27.5000	453.5714	454	3
11	07/28/2010	'Chavara'	80	30	340	520	290	100	37.5000	385.7143	386	4
12	NaT	'Chavara'	120	26	110	530	106.6667	113.3333	32.5000	392.8571	393	4
13	NaT	'Chavara'	40	32	130	500	120	50	40	371.4286	371	4
14	NaT	'Chavara'	60	34	115	385	110	75	42.5000	408.3333	408	3
15	NaT	'Chavara'	60	22	115	565	110	75	27.5000	417.8571	418	3
16	NaT	'Chavara'	240	30	155	600	136.6667	153.3333	37.5000	442.8571	443	3
17	08/15/2010	'Chavara'	40	26	130	495	120	50	32.5000	367.8571	368	4
18	08/24/2010	'Chavara'	120	18	90	610	90	113.3333	22.5000	450	450	3
19	08/26/2010	'Chavara'	140	14	255	420	205	120	17.5000	314.2857	314	4

Fig.4. Input Dataset

Fig 4 shows the sample values of the input dataset. In this, the chavara dataset is used for predicting the Air Quality Index values. The dataset comprises of 12 variables and 1334 instances as shown in the above figure. The first two variables denote the date and place of the recording. The next 8 variables show the chemical component in that place at that time. The 11th variable gives the Air Quality index value. The final 12th variable denotes the AQ level like Moderate (1), poor (2), Severe (3) and very poor (4).

The first stage of pre-processing is performed by removing the first two variables, as it has lower impact in AQ prediction. These two variables removed and is then used for further processing.

_preprocessed	NO2_preprocessed	PM10_preprocessed	SPM_preprocessed	PM10_SubIndex	SO2_SubIndex	NO2_SubIndex	SPM_SubIndex	AQI	AQI_BUCK
80	24	225	565	183.33	100	30	417.86	418	3
40	24	230	600	186.67	50	30	442.86	443	3
60	32	145	495	130	75	40	367.86	368	4
60	52	285	610	235	75	65	450	450	3
160	30	355	420	306.25	126.67	37.5	314.29	314	4

Fig 5. First Stage pre-processed

The first stage pre-process is followed by the second stage pre-processing steps called SMOTE to balance the data. As in original data, the number of samples in each class are not equally distributed as shown in the below Fig 6.

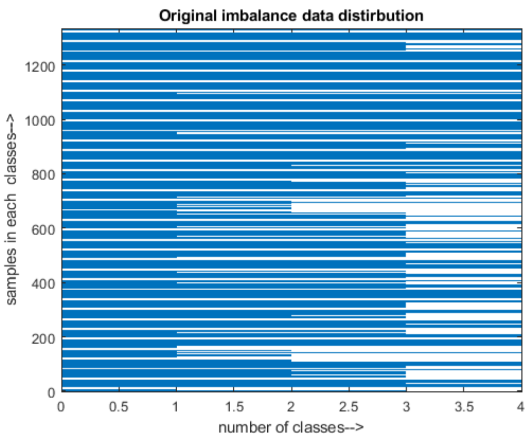


Fig 6. Before SMOTE processing

Fig 6 shows that the samples are unevenly distributed between the classes. Then, these data are processed using SMOTE to evenly distribute the classes as shown in the below Fig 7.

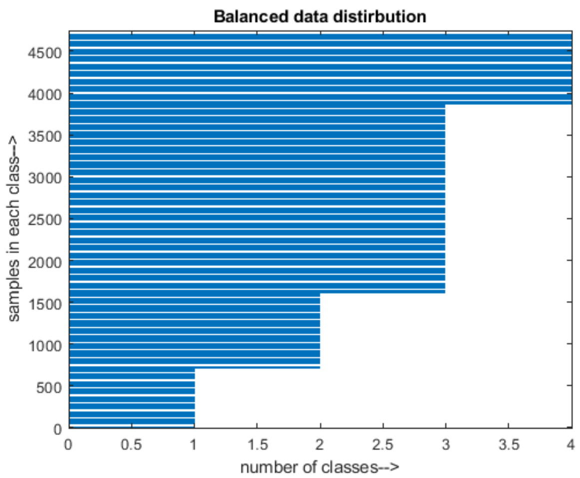


Fig 7. SMOTE data

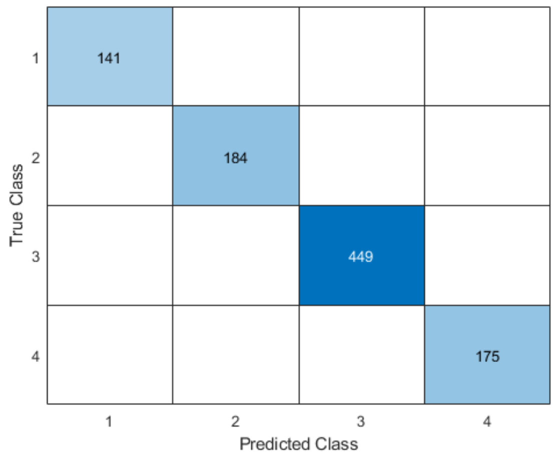
Fig 7 shows that the number of samples in each class is evenly distributed and the number of samples in the set is also increased to 4587. This increased dataset is then split into training and testing using hold-out approach with 80% as training and 20 % testing.

Then, the individual models like K-nearest neighbor, Support vector Machine, Decision tree algorithm were trained using 80% of data and then tested with 20% of data.

```
acc_knn =

    0.9913


acc_svm =

     1


acc_tree =

     1
```

Fig 8. Evaluation of individual models

From the Fig 8, it can be observed that the K-NN model accuracy is minimum as compared to other models. This also enhanced by combining all the model result and the corresponding output is shown in the below figure.




```
ans =

     1
```

Fig 9. Ensemble model Confusion Matrix & Accuracy.

From the Fig 9, it can be observed that the proposed ensemble model can achieve the higher accuracy as compared to the K-nearest neighbor. Based on this, the proposed method performance is compared with the KNN, Decision tree and SVM in terms of Precision, recall, F1-Score and accuracy.

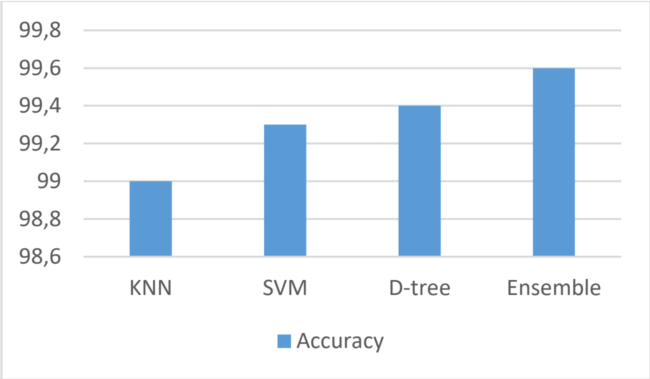


Fig 10. Accuracy comparison

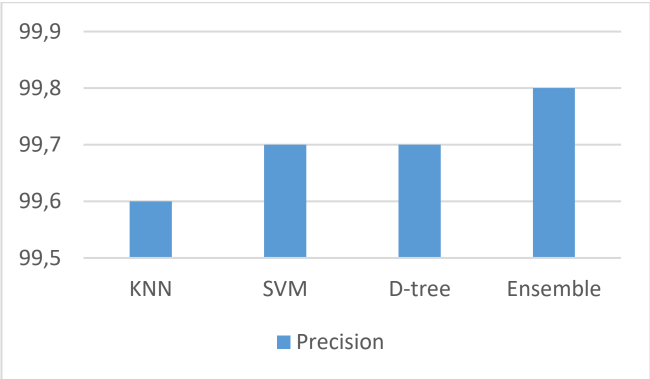


Fig 11. Precision Comparison

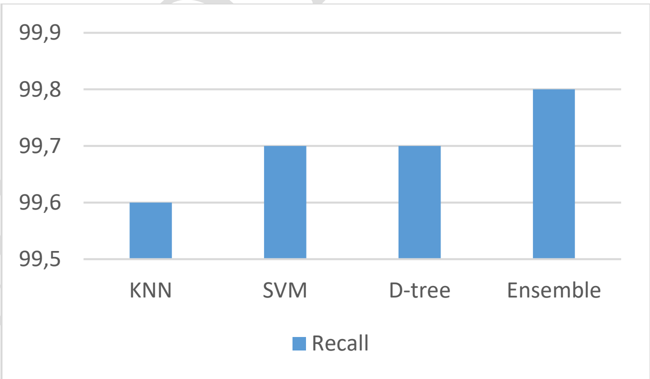


Fig 12. Recall comparison

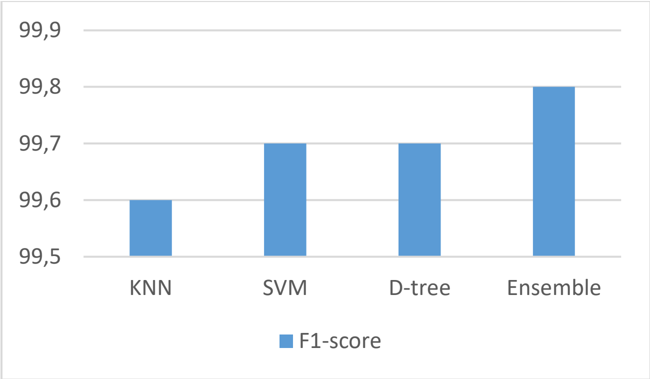


Fig 13. F1-score comparison

Fig 10- Fig 13 shows the comparison charts of the proposed ensemble approach with the individual classifiers. From the results, it can be observed that the proposed approach achieved better results as compared to the existing and individual classifiers. The study employed machine learning algorithms and SMOTE to predict Air Quality Index (AQI) levels in Indian cities. The ensemble method outperformed individual classifiers like K-nearest neighbor, Support Vector Machine, and Decision Tree, achieving higher accuracy, precision, recall, and F1-score, indicating its effectiveness in predicting AQI levels. The study also highlighted the importance of data preprocessing and balancing for improved prediction accuracy.

V. Conclusion

Pollution levels around the KMML facility in the Kollam area of southern India were found to be very high after an evaluation of PM10 (respirable fine particles) concentration at several sampling locations. S4 had the highest TSPM content compared to the other research sites, followed by S1, S2, S3, S4, S5, and S6. The National Ambient Air Quality Standards (NAAQS) criteria set by India's Central Pollution Control Board (CPCB) as well as Ministry of Environment & Forests (MoEF) were met for both sulfur dioxide (SO2) and nitrogen oxides (NOx). Free chlorinated pollution was found at survey sites in the industry's northern, eastern, as well as southern directions. Sensor II, located in the northern part of the business district, has the greatest concentration of chlorine free. Therefore, the research found those persons, particularly youngsters and the elderly, in neighborhoods close to the KMML company experience a variety of health concerns owing to air pollution caused by PM10 as well as chlorine. Data collected for criterion pollutants, including particulate matter, sulphur dioxide, as well as nitrogen oxides, were used to evaluate the research area's quality of air. The state of the air at the study sites S1, S3, S4, and S5 was considered to be around average for the two seasons. In each season analyzed, the pollution level at the monitoring station (S6) was high. The current investigation found free chlorine pollution in the research stations S1, S2, S3, S4, and S5, indicating poor air quality in the area around the KMML facility. People with respiratory issues in the study region may be more vulnerable to ozone and chlorine. Therefore, it is suggested that the KMML plant install high-efficiency pollutants management systems in order to regulate air pollutants in the research region. The industrial administration can also institute measures to mitigate air quality by planting air pollution resistant plants in the region around factories. The federal government of India's Ministry of Environment and Forests (MoEF) needs to publish Ambient Air Quality Guidelines for chlorine in industrial regions since there are now no such limitations in place. The current research also indicates the need of routine air quality monitoring for chlorine and its health consequences at particular locations within the study region.

Future efforts focus on implementing long-term air quality monitoring programs and evaluating the effectiveness of mitigation strategies, including the installation of high-efficiency pollutants management systems and the use of air pollution-resistant plants. These measures are essential for enhancing air quality in the vicinity of industrial facilities like KMML.

Nomenclature

SMOTE	Synthetic Minority Over-sampling Technique
SO2	Sulfur Dioxide
NOx	Nitrogen Oxides
AQI	Air Quality Index
PM10	Particulate Matter 10 micrometers or less in diameter
TSPM	Total Suspended Particulate Matter

Declaration:

Ethics Approval and Consent to Participate:

No participation of humans takes place in this implementation process

Human and Animal Rights:

No violation of Human and Animal Rights is involved.

Funding:

No funding is involved in this work.

Data availability statement:

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study

Conflict of Interest:

Conflict of Interest is not applicable in this work.

Authorship contributions:

All authors are contributed equally to this work

Acknowledgement:

There is no acknowledgement involved in this work.

References

1. Cammalleri, V., Marotta, D., Protano, C., Vitali, M., Villari, P., Cattaruzza, M.S., & Gro, O.B. (2020). How Do Combustion and Non-Combustion Products Used Outdoors Affect Outdoor and Indoor Particulate Matter Levels? A Field Evaluation Near the Entrance of an Italian University Library. International Journal of Environmental Research and Public Health, 17.
2. Li, Y., Wang, J., Han, T., Wang, Y., He, D., Quan, W., & Ma, Z. (2019). [Using Multiple Linear Regression Method to Evaluate the Impact of Meteorological Conditions and Control Measures on Air Quality in Beijing During APEC 2014]. Huan jing ke xue= Huanjing kexue, 40 3, 1024-1034 .
3. Hershberger, R., Givertz, M.M., Ho, C.Y., Judge, D.P., Kantor, P.F., McBride, K.L., Morales, A., Taylor, M.R., Vatta, M., Ware, S.M., & Committee, O.B. (2018). Genetic evaluation of cardiomyopathy: a clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). Genetics in Medicine, 20, 899-909.
4. Hershberger, R., Givertz, M.M., Ho, C.Y., Judge, D.P., Kantor, P.F., McBride, K.L., Morales, A., Taylor, M.R., Vatta, M., Ware, S.M., & Committee, O.B. (2019). Correction: Genetic evaluation of cardiomyopathy: a clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). Genetics in Medicine, 1.
5. Quellhorst, L., Barten-Neiner, G., de Roux, A., Diel, R., Mertsch, P., Pink, I., Rademacher, J., Sutharsan, S., Welte, T., Zurawski, A., Ringshausen, F.C., & Group, O.B. (2022). Psychometric Validation of the German Translation of the Quality of Life Questionnaire-Bronchiectasis (QOL-B)—Data from the German Bronchiectasis Registry PROGNOSIS. Journal of Clinical Medicine, 11.

6. Cheng, C., Chen, Y., Yen, S., Lin, H., Lin, H., Chou, K., & Cheng, C. (2022). Air Pollution Exposure and the Relative Risk of Sudden Sensorineural Hearing Loss in Taipei. International Journal of Environmental Research and Public Health, 19.
7. Han, I., Seo, J.Y., Barr, D.B., Panuwet, P., Yakimavets, V., D'Souza, P.E., An-Han, H., Afshar, M., & Chao, Y. (2022). Evaluating Indoor Air Phthalates and Volatile Organic Compounds in Nail Salons in the Greater New York City Area: A Pilot Study. International Journal of Environmental Research and Public Health, 19.
8. Krenz, U., Greving, S., Zeldovich, M., Haagsma, J.A., Polinder, S., von Steinbüchel, N., & Inves, O.B. (2023). Reference Values of the Quality of Life after Brain Injury (QOLIBRI) from a General Population Sample in Italy. Journal of Clinical Medicine, 12.
9. Dai, H., An, J., Li, L., Huang, C., Yan, R., Zhu, S., Ma, Y., Song, W., & Kan, H. (2019). [Health Benefit Analyses of the Clean Air Action Plan Implementation in Shanghai]. Huan jing ke xue= Huanjing kexue, 40 1, 24-32 .
10. Liu, W., Hu, T., Mao, Y., Xu, A., Su, Y., Li, X., Cheng, C., Shi, M., Liang, L., Xing, X., Zhang, J., & Qi, S. (2021). [Characteristics and Origin Analysis of Air Pollution During the Spring Festival in Linfen, Fenwei Plain]. Huan jing ke xue= Huanjing kexue, 42 11, 5122-5130 .
11. Kretschmer, A., van den Bergh, R.C., Martini, A., Marra, G., Valerio, M., Tsaur, I., Heidegger, I., Kasivisvanathan, V., Kesch, C., Preisser, F., Fankhauser, C.D., Zattoni, F., Ceci, F., Olivier, J., Chiu, P., Puche-Sanz, I., Thibault, C., Gandaglia, G., Tilki, D., & Group, O.B. (2021). Assessment of Health-Related Quality of Life in Patients with Advanced Prostate Cancer—Current State and Future Perspectives. Cancers, 14.
12. Alsaber, A.R., Pan, J., Al-Herz, A., Alkandary, D.S., Al-Hurban, A.E., Setiya, P., & Group, O.B. (2020). Influence of Ambient Air Pollution on Rheumatoid Arthritis Disease Activity Score Index. International Journal of Environmental Research and Public Health, 17.
13. Malandrino, P., Mazzilli, R., Puliani, G., Di Molfetta, S., Pugliese, G., Olana, S., Colao, A., Faggiano, A., & Group, O.B. (2022). The Effects of Radioligand Therapy on Quality of Life and Sexual Function in Patients with Neuroendocrine Neoplasms. Cancers, 15.
14. Bras, A.E., Matarraz, S., Nierkens, S., Fernandez, P., Philippé, J., Aanei, C.M., de Mello, F.V., Burgos, L., van der Sluijs-Gelling, A.J., Grigore, G.E., van Dongen, J.J., Órfão, A., van der Velden, V.H., & Consortium, O.B. (2022). Quality Assessment of a Large Multi-Center Flow Cytometric Dataset of Acute Myeloid Leukemia Patients—A EuroFlow Study. Cancers, 14.
15. Lleal, M., Baré, M., Ortonobes, S., Sevilla-Sánchez, D., Jordana, R., Herranz, S., Gorgas, M.Q., Espaulella-Ferrer, M., Arellano, M., de Antonio, M., Nazco, G.J., Hernández-Luis, R., & Group, O.B. (2022). Comprehensive Multimorbidity Patterns in Older Patients Are Associated with Quality

Indicators of Medication—MoPIM Cohort Study. International Journal of Environmental Research and Public Health, 19.

16. Chiranjeevi, P., & Rajaram, A. (2023). A lightweight deep learning model based recommender system by sentiment analysis. Journal of Intelligent & Fuzzy Systems, (Preprint), 1-14.
17. Chiranjeevi, P., & Rajaram, A. (2022). Twitter Sentiment Analysis for Environmental Weather Conditions in Recommendation of Tourism. JOURNAL OF ENVIRONMENTAL PROTECTION AND ECOLOGY, 23(5), 2113-2123.