A Review on Predicts Emission of Greenhouse Gases Using ML

[1] Apeksha K Gowda , [2] Dr. S Geetha

[1]Student, [2]Associate Professor
[1]Department of CSE,
[1]BNM Institute of Technology, Bengaluru, India

Abstract: **The environment is under increased threat due to the rapid increase in GHG (Greenhouse Gas) emission. The effects of change of climate on the planet's inhabitants include an increase in heat and temperature at the surface of the earth, a rise in sea levels, and direct effects on human health such heat stress and other heat-related illnesses. Researchers have suggested a variety of strategies for predicting GHG emission; however machine learning-based approaches using gas feature input vectors appear to have the best software reliability, operating excellence, and prediction accuracy. We examined various greenhouse gases that are released from various sectors in this paper, examined various feature sources, reviewed machine learning approaches for forecasting the production of greenhouse gases, looked at the benefits and drawbacks of these approaches, and talked about the field's future directions. The numerous machine learning methods for predicting greenhouse gas emissions, which can then be used for analysis, will be discussed in this study.**

Index Terms Machine Learning, Prediction, Greenhouse gases (GHG);

I. INTRODUCTION

Greenhouse gas emissions are a major contributor to global warming, which has a direct impact on climate change, sea level rise, and other factors. It also has a direct impact on human health, including extreme heat as well as other thermally chronic conditions that can create quickly in very warm temperatures and high humidity, like heat stroke, and some people will experience severe heat illness. Significant environmental and health risks are posed by greenhouse gases like CO2 (Carbon dioxide), SO2(Sulphur dioxide), and N2O(Nitrous Oxide). With 53% of all greenhouse gases being CO2, it is one of the most significant greenhouse gases. Human respiration, industrialization, the burning of fossil fuels for energy, forest fires, and other factors all contribute to the release of CO2 into the atmosphere. The atmosphere contains sulphur dioxide, which has an impact on the local weather system and earth's temperature. Sulphur dioxide causes inflammation of the mucous membranes, the eyes, nose, throat, and lungs and is released by electrical production, commercial boilers, petrochemical industries, and hydrometallurgy. Nitrous oxide, sometimes known as laughing gas, seems to be an effective gas in the atmosphere that has a 350-times greater impact over atmospheric co2. The emission of toxic gases or hazardous greenhouse gases like Atmospheric the release of various greenhouse gases in to atmosphere as household, commercial, and human activity is one of the main drivers of global temperature rise. Using different toxic chemicals and herbicides, a producer engages in domestic agricultural activities to increase crop productivity. Human activities like deforestation and industrial activities like coal mining released CO2, CH4, Carbon monoxide (CO), and other harmful greenhouse gases. When people burn fossil fuels for transportation, they also release harmful chemicals into the environment, including CO2, CH4, and chlorofluorocarbons (CFCs). All of them contribute to climate change. The destruction of the living creators and the environment on earth will result if it all continues to get worse day by day.

As global industry and human civilization advance, fossil fuel consumption is increasing at an alarming rate, leading to serious environmental issues like the warming of the planet. The greenhouse effect is mostly caused by CO2, as well as other dangerous gases including Hydrocarbons and N2O. According to statistics, the global average annual Earth's surface temperature has been rising during the past two hundred years. This paper seeks to inform readers about the significance of the environmental challenges.

II. MACHINE LEARNING

Machine learning (ML) is a research area concerned with understanding and creating "learning" techniques, or technologies that allow use of data to improve performance on some set of tasks. It's thought of as a part of machine intelligence. Machine learning techniques create a method of statistical inference, also referred to as training examples, in addition to making decisions or judgments with out being expressly programmed to do it anyway. These models are used in a variety of applications, including object recognition, spam filtering, pharmacy, and voice recognition. Programs that use machine learning can do work without having them specifically coded. Computers use available data to learn in order to do specific jobs. For straightforward jobs given to systems, it really is able to construct techniques that instruct the device how to carry out all the steps necessary to address the issue at hand; no learning is required on the part of the computer. Instead of considering human developers describe each necessary measure, it may prove to be more reliable and effective to assist the machine in creating on its strategy. Machine learning uses a variety of techniques to train computers to complete jobs for which there isn't a totally suitable solution. One strategy is to declare a few of the right answers as valid when there are many possible replies. The computer can then use this as practice data to refine the algorithm(s) that employs to decide the proper responses.

III. MACHINE LEARNING ALGORITHMS AND RELATED PERFORMANCE MATRICES.

Machine learning algorithms that are appropriate for estimating GHG emission include Logistic regression, Gaussian Naive Bayes (GaussianNB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT), linear regression, and Random Forest (RF).

A generalized linear regression analytic methodology for assessing a specific thing's probability is logistic regression [2] [7]. The objective of logistic regression is to identify the better model that explains the relationship among the variables as well as an array of independent variables. (2) GaussianNB [2] is based on the Bayes theory as well as assumes that the necessary features are independent of one another. (3) SVM uses supervised learning, a generalized linear classifier can categorizes data [12]. It is the popular one kernel learning techniques is the SVM, which can perform nonlinear classification. (4) The goal of the KNN [12] algorithm is to locate a sample's k closest neighbors. (5) The DT [13] is a a nonlinear trained classifier that may also create decision rules can be used from a data set with feature representation and it present these rules using the tree's structure to address classification and regression issues. (6) A classifier with several decision trees is called the Random forest [4] [12] [13]. The majority of decision trees' output categories determine this tree's output category. (7)Linear Regression [2] [4] 12] [13] [15] is a supervised ML model it identifies the dependent and independent variables' linear connection and determining the best fit linear line between them.

Researchers created a number of evaluation measures to assess the effectiveness of machine learning systems. The following formula displays the point of comparison:

$$\text{Accuracy score} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FN} + \text{TN} + \text{FP})}$$

$$\text{Precision Score} = \frac{\text{TP}}{(\text{TP} + \text{FP})}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})}$$

$$\text{F1} - \text{Score} = \frac{2 * \text{Precision Score} * \text{Recall Score}}{(\text{Preecision Score} + \text{Recall Score})}$$

Where, A model's ability to forecast the positive class is gauged by the statistic known as TP (True Positive). As a result, the model predicted a positive outcome, and the outcome was favourable. When a model predicts that an instance belongs to a class when it doesn't, this is known as a false positive (FP). False positives pose a danger since they could lead to poor decisions. The model predicts that certain outcomes will be negative, or TN (True Negative). One of the criteria used to assess a classification model's effectiveness is true negatives. False Negative (FN) results whenever a model predicts a negative outcome when the actual result is positive.

IV. LITERATURE REVIEW

The GHG emissions released by the ship during fuel use are predicted by ZHIHUI HU et al. [1]. GPR (Gaussian Process Regression) and BPNN (Back-Propagation Neural Network) approaches are used by machines to forecast gas consumption in ships. The results demonstrate that CO2 was emitted from the ship during the fuel consumption process. Finally, machine learning models such as the BPNN and GPR approach aid in the prediction of GHG emissions from ships.

Xiangyong Lu et al., [2] set a restriction on the amount of carbon that can be released via transportation. Their findings in the city were utilized to anticipate actual time and fine-grained emission records of carbon during transportation for the entire city, including taxi Location data, transportation carbon pollution data and street layouts, factors of pursuits (POIs), and weather data. To examine the properties of accumulated records and infer transportation carbon emissions, a 3-layerP NN (three-layer perceptron neural network) is designed. We build our technique on large-scale tests based on 5 real-world data sources received in Zhuhai, China. Our methodology outperforms three major computer learning techniques (Logistic Regression, Linear Regression, Gaussian Naive Bayes) and two deep learning techniques, according to the findings (Deep Belief Networks, Stacked Denoising Autoencoder).

By Zhengbing He et al., "Estimating CO2 Emissions from Freeway Traffic," [3]. The carbon dioxide (CO2) emissions of a vehicle are forecasted using a spatial-temporal (ST) cell-based model. The findings demonstrate that accurate CO2 emission from a traffic dataset is made possible by the construction of the spatial-temporal map and its accompanying traffic flow. The model opens a gate for predicting emissions of CO2 from widely available low-fidelity visitor data because the model input, i.e., spatial-temporal diagram, can generated with more than a few site visitor waft data, particularly using loop detector facts and low frequency floating automobile data.

For SF6 substitutions, HAO SUN et al. [4] anticipate the boiling temperature and electrical strength by using computer learning techniques, such as random forest and ANN (artificial neural network), to forecast the boiling temperature and electrical power. Using a set of descriptors derived from the molecular shape of 74 molecules, the performance of three different methods—more than one linear regression, synthetic neural network, and random forest—is compared and assessed in terms of sensitivity to sample size, predict stability and accuracy, and predictor interpretability. Random forest displays the greatest results in efficiency and robustness given the limited number of records provided. The boiling temperature was calculated using the same techniques, with random forest providing even better results.

Khoa N. Le et al. [5] devised a model to help engineers and technicians process sustainable projects using inexpensive equipment. The model was created using MS Excel and Visual Basic; using the equipment as effective software, designers and practitioners may learn about the various chemicals' contributions to greenhouse gas emissions throughout a concrete's useful life. A proposed model was developed to analyze Green Star concrete credits, but it could also be utilized to predicts and evaluate the other Green Star credits if specified and extra features are provided.

Ana Carolina Riekstin et al., [6] have developed a time sequence approach for predicting GHG emissions that may be linked to smart home control systems. The Long Short-Term Memory (LSTM). recurrent neural network was utilized by the researchers. When the area under investigation has an electricity matrix that is primarily based on fossil fuels and is less intermittent, the prediction outcomes get close to two % mean absolute %age error (MAPE). The MAPE is around 12 % in locations where there are more renewable sources.

Jeeven Ravi et al., [7] developed a technique to prevent global warming from occurring as a result of the production of certain harmful GHGs. They want to use big data applications to talk about the current state of the climate. The 5 unique graphs on the basis of greenhouse gas emissions were created using real-time Power-Bi forecasting technology to predict GHG from extraordinary sectors. Finally, using the Power-Bi forecasting device, The damaging consequences of local weather change can be lessened by the use of a number of mitigating strategies. However, the planned approach will no longer recognize which greenhouse gas emits the most.

Sonali Kangralkar et al., [8], forecasting real-time CO2 emissions from a car. Machine Learning algorithms and concepts are used. KNN 28%, Linear Regression 82.5%, SVM 80%, Random Forest 99%, and Decision Tree 96% of ML algorithms were employed for estimating CO2 emissions and function a comparison analysis. The results show that K-Nearest Neighbor has a 28% success rate, Support Vector Machine has an 80% success rate, Linear Regression has an 82.5% success rate, Decision Tree has a 96% success rate, and Random Forest has a 99% success rate. Finally, these algorithms were used to forecast CO2 emissions and comparative research was conducted.

Using soil variables such as moisture of soil, temperature of soil, humidity, and pH values, Pranali K et al. [9] attempted to determine the greenhouse gas CH4 and CO2 emissions from agriculture. For evaluation, we used one-of-a-kind Machine learning algorithms like Zero R, Random Forest, Multilayer Perception, and Multiple Linear Regression. The R squared fee of 0.786 implies that 78 % of the projected values are comparable to proper CH4eq emission research. Finally, the suggested gadget will be used to capture real-time data on soil properties.

V. Tanania et al., [10], using an Indian dataset from 1995 to 2018, evaluate and anticipates carbon dioxide (CO2) emissions. The motivation of this paper also serves to inform readers on the seriousness of the pollution challenges we currently face.. For predicting and inspecting the same, the statistical technique of multiple linear regression is applied. CO2 emissions are considered the structural variable, whereas India's population, year, and electricity usage are considered the independent variables. The usage of several linear regression models yielded a look score of 96.40%.

Sadah A. T. Muawad et al., [11] purpose this work was to determine the amount of CH4 released during the decomposition of organic material in landfills, from a cell in a landfill in Khartoum, Sudan. Total landfill gas was estimated using an emissions model. The 50-year projection horizon was chosen. Using the suggested model, generating power from waste disposal cells might well aid in lowering emissions of GHG. Reducing the use of fossil fuels and reducing emissions from landfills fuels that lessen the effects of global warming contribution; the national grid can get the power generated.

Tan Chun Ho et al., [12] Describe a regression model using GDP, power consumption and production, FDI revenue, net inflow, and electricity consumption as the predictors for CO2 emissions from design and production. The relationship between the prediction model's output and the measurement results was strong (modified R2=96.65 percentage; R2=94.18 percentage), demonstrating the model's efficacy and accuracy. The people can be forewarned early to comply with air quality requirements using these data, which are reassuring and reliable.

Greenhouse gases emission mainly discussed about the review of the literature which has been published by the authors. The literature survey mainly deals with the existing system for their model, analysis of the existing model is done and implementation done accordingly for the existing system. The prediction of different greenhouse gases that are emitted from different sectors based on dataset was researched, designed and presented around the concept of machine learning. The innovation in this paper lies in the design of the prediction model. As indicated in Table 1, all of the techniques that we have researched are systematically matched.

V. LIMITATIONS AND FUTURE WORK

A. Limitations of the ML Models.

However many algorithms have been used in earlier studies. When it comes to GHG emission, those approaches perform wonderfully overall, but they do have serious downsides, most notably those listed below:

- There aren't any standard datasets available. In 12 Publications that estimate different GHG based on machine learning from January 2019 to November 2021, 9 GHG emission datasets are utilized to verify the practical effect. Figure 1 displays the statistics of these datasets. It is challenging to determine whether an identification strategy is good or bad because there are no best classification datasets.
- A classifier model created using a current dataset does not assure that it would work well with another dataset. On some datasets, various methods may initially yield promising detection results. As time goes on, the learning model may no longer be able to classify new emission samples.

- The resistance to the targeted prediction is frequently poor. Systems are visible, which makes debugging challenging.

[9]	Decision Tree, SOM regression, ZeroR, RF, Multilayer Perception and Multiple Linear regression	soil sample using different sensors from the farm.	1638	Mean absolute error	0.0005 0.128 0.002 0.0008	CH4 and CO2
				Root mean squared error	0.0006 0.1589 0.0185 0.0047 0.0012	
[10]	multiple linear regression	Collected by the researchers	-	Accuracy	96.40	CO2
[11]	Landfill gas emissions model	Landfill data in Khartoum State-Sudan	-	-	-	CH4
[12]	Multiple linear regression	CO2 emissions in Malaysia from 1971 to 2011	-	correlation coefficient	adjusted R^2 of 96.75 predicted R^2 of 94.19	CO2 emission

B. Future Directions

GHG emission statistics have consistently been a popular area of research, according to articles published between January 2019 and November 2021. The static feature-based prediction method has always been unquestionably superior. It seems certain that the machine learning estimation technique will continue to be popular for some time to come.

- The effectiveness of ML techniques will improve. The Accuracy measure has been used 7 times in the research that employed machine learning to predict GHG emission levels between January 2016 and November 2012. The majority of Accuracy measures are over 90%, as seen in Figure 2. It is anticipated that the detection techniques would keep this critical Accuracy metric while becoming more effective and quick.
- Large-scale prediction will be more likely to be supported by the detection approach. Many algorithms were tested on small datasets in earlier research studies. The scalability of these techniques was not tested on big datasets, despite the fact that high performance metrics were attained. Future systems will have a greater demand for quick prediction techniques to handle enormous systems due to the expanding number of systems.
- The predictive algorithm must be able to distinguishbetween low and high threat levels. For known greenhouse gases, the prediction model based on machine learning offers high categorization capabilities. The threat level or imposition of the threat level of the various greenhouse gases released from various sectors will be predicted using future prediction models.

Table1: Continued.

Table 1: Machine learning techniques that are frequently used to forecast greenhouse gas levels.

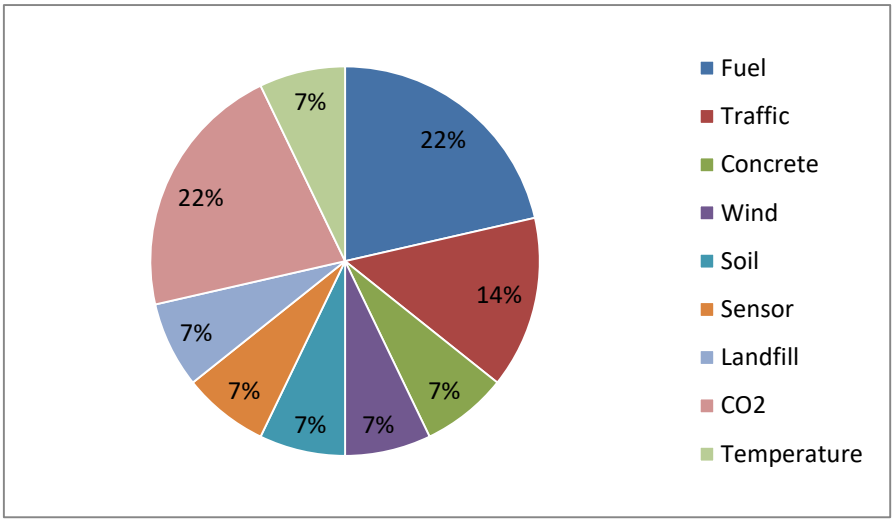

Figure 1: Statistics of several datasets used in works that predicted GHG emission levels between January 2019 and November 2021.

Reference	Algorithm	Datasets	Number of Samples	Metrics	Values in %	Predicted GHG
[1]	GPR, BPNN	Fuel Consumption Data	24386	Mean Value	2236.4 14.7	CO2
[2]	DBN, GaussianNB, Linear, 3-Layer PNN, LogisticR, SDA.	road network dataset	30277	Accuracy	32.306 81.72 81.38 90.86 88.31 32.348	CO2
[3]	First- and second-order regression models	low-fidelity traffic data	34073	Accuracy	86.05	CO2
[4]	Multiple Linear Regression, ANN, RF	Electrical Strength and Boiling Temperature dataset	67	R 2 -value	95.1 96.67	SF6
[5]	computer-aided model	concrete mixture dataset	-	Accuracy	Accuracy of the model dependent on specific application	GHG emissions
[6]	LSTM, RNN	wind and solar source from IESO.	Real time data	-	-	GHG
[7]	logistic regression model	dataset was downloaded from Kaggle	1889	Accuracy	97.40	GHG from the electricity generation sector
[8]	KNN, SVM, Linear Regression, RF	sensor data from an automobile	500+	Accuracy	28 80 82.5 96 99	CO2 emissions

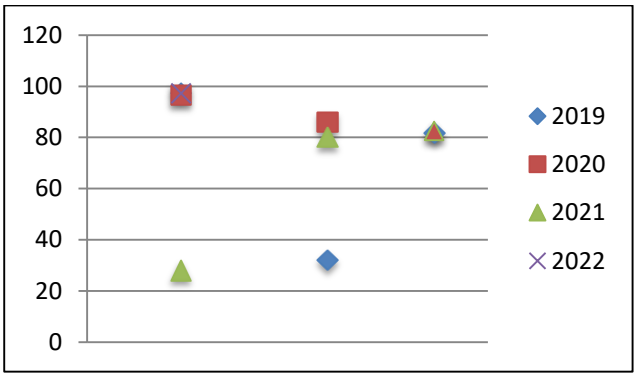

Figure 2: Statistics of the Accuracy metric of works predicting GHG emissions using machine learning from January 2019 to November 2021.

VI. CONCLUSION

The experiments give good separation for the lungs sound from the heart sound. These algorithms use the sound which is recorded and then processed to remove the heart sounds along with other sounds like abdominal sounds. The digestive sounds also interfere as the diaphragm acts as an amplifier for certain frequencies. More fine tuning is required to get accuracy around 95% while separating the sounds. Better separation ensures better analysis for the smart stethoscope.

REFERENCES

1. Z. Hu, Y. Jin, Q. Hu, S. Sen, T. Zhou and M. T. Osman, "Prediction of Fuel Consumption for Enroute Ship Based on Machine Learning," in IEEE Access, vol. 7, pp. 119497-119505, 2019, doi: 10.1109/ACCESS.2019.2933630.
2. X. Lu, K. Ota, M. Dong, C. Yu and H. Jin, "Predicting Transportation Carbon Emission with Urban Big Data," in IEEE Transactions on Sustainable Computing, vol. 2, no. 4, pp. 333-344, 1 Oct.-Dec. 2017, doi: 10.1109/TSUSC.2017.2728805.
3. Z. He, W. Zhang and N. Jia, "Estimating Carbon Dioxide Emissions of Freeway Traffic: A Spatiotemporal Cell-Based Model," in IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 5, pp. 1976-1986, May 2020, doi: 10.1109/TITS.2019.2909316.
4. H. Sun, L. Liang, C. Wang, Y. Wu, F. Yang and M. Rong, "Prediction of the Electrical Strength and Boiling Temperature of the Substitutes for Greenhouse Gas SF_6 Using Neural Network and Random Forest," in IEEE Access, vol. 8, pp. 124204-124216, 2020, doi: 10.1109/ACCESS.2020.3004519.
5. K. N. Le, V. W. Y. Tam, C. N. N. Tran, J. Wang and B. Goggins, "Life-Cycle Greenhouse Gas Emission Analyses for Green Star's Concrete Credits in Australia," in IEEE Transactions on Engineering Management, vol. 66, no. 3, pp. 286-298, Aug. 2019, doi: 10.1109/TEM.2018.2832094.
6. A. C. Riekstin, A. Langevin, T. Dandres, G. Gagnon and M. Cheriet, "Time Series-Based GHG Emissions Prediction for Smart Homes," in IEEE Transactions on Sustainable Computing, vol. 5, no. 1, pp. 134-146, 1 Jan.-March 2020, doi: 10.1109/TSUSC.2018.2886164.
7. J. Ravi, "Identification of Greenhouse Gases Emission Thorough Exploration of The Emission from Different Sectors", International Conference on Computing Engineering and Design (ICCED), 2020, pp. 1-7, doi: 10.1109/ICCED51276.2020.9415811.
8. S. Kangralkar and R. Khanai, "Machine Learning Application for Automotive Emission Prediction," International Conference for Convergence in Technology (I2CT), 2021, pp. 1-5, doi: 10.1109/I2CT51068.2021.9418152.
9. P. K. Kosamkar and V. Y. Kulkarni, "Effect of Soil Attribute on Greenhouse Gas Emission from Agriculture using IoT," 2019 IEEE Pune Section International Conference (PuneCon), 2019, pp. 1-6, doi: 10.1109/PuneCon46936.2019.9105819.
10. V. Tanania, S. Shukla and S. Singh, "Time Series Data Analysis and Prediction of CO2 Emissions," International Conference on Cloud Computing, Data Science & Engineering, 2020, pp. 665-669, doi: 10.1109/Confluence47617.2020.9058001.
11. Muawad, Sadah. (2019). Prediction of Landfill Gas Emissions and Energy Recovery Potential from an Engineered Landfill, in Khartoum State-Sudan. 1-5. 10.1109/ICCCEEE46830.2019.9071020.
12. Tan, Chunho & Sim, Chong Keat & Lim, Hwee San. (2015). A prediction model for CO2 emission from manufacturing industry and construction in Malaysia. 469-472. 10.1109/IconSpace.2015.7283771.