Research on Green Machine Learning

A project Report submitted to

DEPARTMENT OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY

BACHELOR OF COMPUTER SCIENCE

(B.Sc.)

By

Nishant kumar Chaturvedi

Roll No: 22070150

Enrolment No.:

GGV/22/05150

Under the Guidance of

Dr. Rajwant

Singh Rao

DEPARTMENT OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY

GURU GHASIDAS VISHWAVIDYALAYA, BILASPUR

Session: 2024-2025

CERTIFICATE OF SUPERVISOR(S)/GUIDE

This is to certify that the work incorporated in the project **Research on Green Machine Learning** is a record of six-month project work assigned by our Institution, successfully carried out by **Nishant kumar Chaturvedi** bearing Enrolment No **GGV/22/05150** under my guidance and supervision for the award of Degree of Bachelor of Computer Science (BSc) of **DEPARTMENT OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY, GURU GHASIDAS VISHWAVIDYALAYA, BILASPUR C.G., INDIA**. To the best of my knowledge and belief the report embodies the work of the candidate him/herself and has duly been successfully completed.

Signature of the Supervisor/Guide

Name: .....

Signature of HOD

Name: …….

DECELERATION BY CANDIDATE

I Nishant Kumar Chaturvedi Student of VI Semester BSc, **DEPARTMENT OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY, GURU GHASIDAS VISHWAVIDYALAYA, BILASPUR**, bearing Enrolment Number **GGV/22/05150** here by declare that the project titled **Research on Green Machine learning** has been carried out by me under the Guidance/Supervision of **Dr. Rajwant Singh Rao, Assistant Professor,** submitted in partial fulfilment of the requirements for the award of the Degree of Bachelor of Computer Science (BSc) by the Department Of Computer Science & Information Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur during the academic year 2024-25. This report has not been submitted to any other University for any award of Degree.

Date: **(Signature of Candidate)**

Place:

ACKNOWLEDGEMENT

I have great pleasure in the submission of this project report entitled **Research on Green Machine Learning** in partial fulfilment of the degree of Bachelor of Computer Science. While Submitting this Project report, I take this opportunity to thank those directly or indirectly related to project work.

I would like to thank my guide **Dr. Rajwant Singh Rao,** who has provided the opportunity and organizing project for me. Without his active co-operation and guidance, it would have become very difficult to complete task in time.

I would like to express sincere thanks to **Dr. Ratnesh Prasad Srivastava**, Head of Department of Computer Science & Information Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur C.G

Acknowledgement is due to my parents, family members, friends and all those persons who have helped me directly or indirectly in the successful completion of the project work.

Nishant Kumar Chaturvedi

Table of Content

Green Machine Learning

1. Abstract –

This research is about Green Machine Learning and Green AI. It is basically a review paper of the research papers that I have read. As we see Artificial Intelligence and Machine learning technologies is becoming more integrated to modern technologies, in today's world the race of Integrating AI has been increased. Everyone is adding use of AI and ML to their technologies for more efficiency. This research paper presents a comprehensive study of existing papers on Green Machine Learning and Green AI, focusing on sustainable practices in model design, energy consumption, and real-world applications.

The study also highlights the role of anomaly detection in IoT and industrial settings, showing the importance of energy-efficient and real-time monitoring solutions. Moreover, it shows the paradigms of Red AI and Green AI, advocating for a shift toward more responsible and environmentally conscious AI development. This paper aims to provide actionable guidelines for researchers, developers, and institutions committed to creating AI systems that are both powerful and sustainable.

2. Keywords –

Green Artificial Intelligence (Green AI), Sustainable Machine Learning, Energy-Efficient AI, Carbon Emissions in AI, Anomaly Detection, AI Environmental Impact, Green-by-AI and Green-in-AI

3. Introduction –

Green Machine learning means making the ML model more optimised and environment friendly. When we talk about the term green machine learning how can we forgot Green AI, both terms are related to each other. Green AI means the Artificial Intelligence that are highly optimised which use less power to give accurate results. These green AI helps in making Green Machine Learning. Over the past decades Artificial Intelligence and Machine Learning have revolutionize multiple industries like healthcare, finance, agriculture, transportation, and

more. How ever this growth has come a greater environmental cost. The development, training and deployment of this models, especially large-scale deep learning system consumes a vast amount of energy that led to carbon emission. For example –

Training models like GPT3 has been reported to consume 1,287 megawatt-hour of electricity and generated 552 metric tons of Co_2 emissions, equivalent to the lifetime emissions of several cars or households, not to mention that the more recent GPT-4 is estimated to be 10 times larger[1]. [1]

This review-based research paper identifies the highly increasing development of Green Machine Learning and Green AI, which leads for the design and deployment of AI systems that are not only effective and scalable but also energy-efficient and environmentally responsible. Green AI aims to reduce the carbon footprint of AI technologies through algorithmic optimization.

This paper is structured as follows: Section 2 explains the review methodology. Section 3 provides a foundational overview of Green AI concepts. Section 4 analyses the eight reviewed papers thematically. Section 5 discusses challenges and limitations; Section 6 discuss future directions. Section 7 concludes the study with key takeaways and a call for environmentally responsible AI development.

4. Methodology –

The review paper is based on the detailed analysis of eight carefully selected research paper that are based on the topic Green Machine Learning (Green ML) and Green AI.

4.1 Selection Criteria –

The selected paper that are chosen are based on the following criteria –

- Focus on environmentally friendly AI/ML development, including energy-efficient algorithms, carbon emission tracking, anomaly detection in IoT, and green deployment strategies.
- Papers that offer systematic reviews, or practical implementations of Green AI techniques.
- Inclusion of both technical algorithm-level and non-technical policy, awareness, lifecycle analysis perspectives to provide a holistic view.

4.2 Time Period Covered –

The selected papers were published between 2018 and 2024, reflecting recent advancements in the field and growing awareness of the environmental implications of AI development.

4.3Databases and Sources –

The papers were collected from reputable and widely recognized academic platforms, including:

- arXiv.org
- ScienceDirect
- IEEE Xplore
- MDPI
- ResearchGate
- HAL Open Science
- Official conference proceedings ESANN 2023

4.4 Review Process

To write this review, I followed a method similar to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), which is a well-known process used by researchers to make sure they choose and study papers in a fair and organized way. I didn't follow all the steps strictly, but used its main ideas to:

- Decide what kind of papers to include or leave out
- Make sure I didn't pick the same topic or paper more than once
- Keep the selection process honest and clear
- Organize the review in a neat and unbiased way

Using this process, I selected eight important research papers that helped me understand the current trends, problems, and future possibilities in Green Machine Learning and Green AI.

5. Overview of Green Machine Learning and Green AI Definition and core principles –

Green machine learning and green artificial intelligence refers to a new wave of development that aim to reduce the carbon emission, environmental impact of developing and using machine learning models. This approaches basically focus on reducing the energy consumption, carbon emission and wastage of computational power. The core principles of Green AI and Green ML Include:

- Developing energy efficient algorithm.
- Use of smaller and optimised models.
- Measuring the carbon footprint through AI system.
- Using the model that are pre-trained instead of starting from scratch.

The main goal is to create a model that is not only accurate and intelligent but also environment friendly.

5.1 Green-by-AI vs Green-in-AI [2]

Green-by-AI: This means using AI to make the environment sustainable. For example, AI can be used to improve energy efficiency in buildings, reduce water use in agriculture, and monitor climate change. In this approach, AI is a tool to solve environmental problems.

Green-in-AI: This refers to making AI more eco-friendly. It involves optimizing the design, training, and deployment of AI systems to use less electricity, produce less CO_2, and require limited resources. This includes practices like using renewable energy-powered data centres, and hardware-level improvements.

Together both of these approaches can help in making environment sustainable and safe.

Red AI: Red AI is the just opposite to Green AI. It mainly focusses on improving accuracy and performance at any cost, ignoring the environmental impact. Red AI relies heavily on training extremely large models with huge datasets, which requires massive computational power and energy. This leads to:

- Higher carbon emissions
- Increased water usage (for cooling data centres)
- Greater hardware demand, contributing to electronic waste

While Red AI may achieve accurate results that is best of all, but it raises serious concerns about environment sustainability and accessibility.

6. Analysis of reviewed paper –

The eight-research paper reviewed are combinedly covers a wide range of Green Machine learning and Green AI concepts. This section analyses them under four category which highlight their contribution, method and real-world implementations.

6.1 Anomaly Detection Techniques in Green ML-

Most of the paper focus on anomaly detection, particularly in time series data generated by Iot devices and industrial system. Anomalies represent the abnormal behaviour or faults and early detection of the problem leads to safety and efficiency.

Model used are –

LSTM (long Short-Term Memory): it is use in greenhouse system; this model can learn new pattern without needing anomaly labels for training. This makes the approach energy efficient. These are the RNN (recurrent neural network). [3]

CNN and U-Net: The use of Convolutional Neural Network and u-net architecture allow for anomaly detection in time series segmentation. CNN require less parameter and less computational power than traditional RNN. [4]

These were the two models that were used in time series anomaly detection, 4 experiment were conducted to check the CNN. The model achieved high accuracy in real world traffic and industrial scenario, missing some of the event due to data gaps.

Application –

- Iot based system
- Industrial environment
- Cybersecurity
- Healthcare monitoring

Green machine learning contribution has enabled the real time anomaly detection without much computational power and energy uses.

6.2 **Carbon emission and Environmental impact –**

Another thing that is been said in research paper is environmental cost of running ML model, particularly large-scale models that consumes a lot of energy and computational power to run and make prediction.

Key findings:

- Training models like chat GPT-3 uses more than 1200 MWh of electricity that releases CO_2 equivalent to the lifetime emissions of several cars. [5]
- Even the infrastructure like cooling system, severs routers etc consumes a lot of energy.
- PUE (power usage Effectiveness): It measures how effectively and efficiently a data centre uses energy. A PUE of 1.58 means 37% of the energy is used just for cooling the servers. [6]
- LCA (Life Cycle Assessment): It means that the carbon emission should be tracked form starting i.e. form hardware production to full deployment of the model. [7]
- Tracking tools: The paper talked about so many tools like code carbon carbonTracker, Green Algorithm, etc that helps the researchers to measure the carbon emission done by the ML model. [8]
- Offsetting: It means compensating for the pollution that can't be avoided. It basically means when you ml mode uses electricity that causes emission you can pay for that by investing in green projects like planting trees or renewable energy but it is not the solution one should focus on reducing the carbon emission rather than offsetting. Offsetting should be the second solution. [9]

6.3 Practical Use Cases of Green AI –

Several papers speak about real-life applications where Green AI principles are already making a great impact. These use cases demonstrate that sustainability and performance can go along with the development of new technologies.

- **Agriculture**: AI-powered drones with multiple cameras monitor crop health while using minimal data bandwidth.
- **Smart Energy Grids**: AI predicts energy demand and balances the supply to reduce wastage.

- **Industry**: ML helps in prediction the error before, reducing the downtime and optimizing energy uses.

7. Challenges and Limitation –

The green Machine Learning and Green AI helps in giving solution to reduce environmental impact. The research papers tell various challenges and limitation that are face during this process.

Many Green machine learning techniques require a high quality and relatively large amount of data to relate anomaly detection and model optimization.

Anomalies are rare in real world datasets making it very difficult for any model to train. It is very challenging to train a large model when you don't have a large amount of data. Noise and missing values in IoT or industrial datasets may affect model performance. publicly available datasets often lack diversity or real-time characteristics, which limits their use in real-world applications.

Implementing Green Machine Learning solution requires a great knowledge in this field and often great infrastructure. Techniques like MU-Net, CNN, LSTM, etc are complex and may not be easily adopted by small teams or institutions. Some of the carbon emission tracking tools are still developing and may give irregular results, making it difficult to measure and compare energy savings accurately. Smaller organizations or those in developing regions may not have access to green data centres or energy-efficient hardware, that may lead to creating a barrier to global adoption. Despite its many advantages, Green Machine Learning faces practical challenges related to data, accuracy and technical expertise.

8. Results –

By reading all research paper and experiment I came to know the following things –

1. Offsetting should be the second option to opt one should focus on optimising the algorithm and model first
2. Anomaly detect is difficult because real life anomalies are very rare making it difficult to train the model for anomaly detection.
3. Two terms were given Green in AI and Green by AI. We must opt both of them for a sustainable development.

4. The model should not start its development form scratch in fact it should use the pre trained model which will save much of energy.
5. The data centre and servers should be established in countries that use green and renewable source of energy.
6. The development of AI and ML model should be under proper surveillance and rules so it is developed for low carbon emission.

9. Conclusion –

The review explores the growing field of Green Machine learning and Green AI that helps in making a balance in the nature. The analysis reveals that while AI and ML have transformed industries with their powerful abilities, they also contribute significantly to energy consumption, carbon emissions, and environmental damage. The paper proposed two solution one is green by ai and green in ai and another one is offsetting. Offsetting should be the second or last option because one cannot restore the balance in nature just by investing the same amount equivalent to damage. The primary focus should be on developing a model that is optimised to give accurate result will less energy consumption and less computational power. The development of sustainable AI systems is not just a technical challenge but a moral and environmental duty. It demands a combine effort from researchers, developers, institutions, and policymakers. Encouraging open-source carbon tracking tools, establishing green computing policies, and investing in clean energy infrastructure are all necessary steps.

In conclusion, this review says that the future of AI must be smart, efficient, and environmentally friendly. Only through responsible innovation we can make sure that AI serves humanity without compromising the planet.

10. Future direction –

As Green machine learning and green ai continue to grow, there are so may aspect that have a significant impact while addressing environmental challenges that are due to traditional AI system. Based on research paper I read following are the area that can lead to future research and development.

Development of Light weight Models: one of the best ways is to develop lightweight models that deliver strong performance with less computational power. This models:

- use few parameters and require less memory

- Train faster and consumes less power
- They are suitable for low-resource environment, including mobiles.

Green AI policy: The research papers clearly implies that there should be some standard set of rules for developing AI and ML models that can lead to a guide for developing models that are sustainable and environment friendly.

- Encouraging organizations to use green-certified data centres
- Supporting open-access tools for carbon tracking and energy monitoring.
- Establishing the data centres and server in area that uses green energy.

The future of Green Machine Learning lies in designing lighter and policy-driven AI systems. By focusing on energy-efficient models, responsible governance, and research, the AI community can lead a more sustainable and best technological future.

11. References –

1. Research paper 4: - Green Machine Learning (ESANN 2023 special session) - Introduction section.

Authors – Veronica Bolon-Canedo, Laura Moran-Fernandez, Brais Cancela and Amparo Alonso-Betanzos Published –

https://www.esann.org/sites/default/files/proceedings/2023/ES2023-3.pdf

2. Research paper 5: - A review of green artificial intelligence: Towards a more sustainable future. Authors – Veronica Bolon-Canedo, Laura Moran-Fernandez, Brais Cancela and Amparo Alonso-Betanzos

Published –

https://www.sciencedirect.com/science/article/pii/S0925231224008671

3. Research paper 1: - Greenhouse: A Zero-Positive Machine Learning System for Time-Series Anomaly Detection. Introduction section

Authors - Tae Jun Lee(Microsoft),

Justin Gottschlich, Nesime Tatbul (Intel Labs)

Eric Metcalf, Stan Zdonik (Brown University)

Published at - https://arxiv.org/abs/1801.03168

4.Research paper 6: - Time Series Anomaly Detection Using Convolutional Neural Networks and Transfer Learning. Section 3 CNN-based time series segmentation

Authors – Tailai Wen, Roy Keyes

Published –

https://arxiv.org/pdf/1905.13628

5. Research paper 4: - Green Machine Learning (ESANN 2023 special session). Introduction section

Authors – Veronica Bolon-Canedo, Laura Moran-Fernandez, Brais Cancela and Amparo Alonso-Betanzos

Published –

https://www.esann.org/sites/default/files/proceedings/2023/ES2023-3.pdf

6. Research paper 2 -: A Practical Guide to Quantifying Carbon Emissions for Machine Learning researchers and practitioners. Carbon Tracking section

Authors – Anne-Laure Ligozat, Sasha Luccioni

Published -

https://hal.science/hal-03376391/file/A_Practical_Guide_to_Quantifying_Carbon_Emissions.pdf

7.Research paper 2 -: A Practical Guide to Quantifying Carbon Emissions for Machine Learning researchers and practitioners. Carbon Tracking - life cycle Assessment section

Authors – Anne-Laure Ligozat, Sasha Luccioni

Published -

https://hal.science/hal-03376391/file/A_Practical_Guide_to_Quantifying_Carbon_Emissions.pdf

8.Research paper 5: - A review of green artificial intelligence: Towards a more sustainable future. Energy consumption calculation tools section

Authors – Veronica Bolon-Canedo, Laura Moran-Fernandez, Brais Cancela and Amparo Alonso-Betanzos

Published –

https://www.sciencedirect.com/science/article/pii/S0925231224008671

9.Research paper 2 -: A Practical Guide to Quantifying Carbon Emissions for Machine Learning researchers and practitioners. Offsetting section

Authors – Anne-Laure Ligozat, Sasha Luccioni

Published -

https://hal.science/hal-03376391/file/A_Practical_Guide_to_Quantifying_Carbon_Emissions.pdf

Review of Research papers: -

Research paper 1: - Greenhouse: A Zero-Positive Machine Learning System for Time-Series Anomaly Detection

Summary –

The first research paper talks about a system called Greenhouse that basically detects anomaly or any unexpected behaviour in the time series data that is being collected over time.

In the world of Internet of things (IoT) data is being collected every second resulting in collection of millions of data. In various devices like self-driving car, smart factories or mobile tracking we have tons of sensor that is collecting data every second.

- This data is usually numbers changing over time (like temperature readings changing over time)
- Sometime, something strange happens like a sudden spike in temperature and we want to catch those spikes.

Anomaly Detection – Anomaly detection means spotting when something unusual happens.

It is basically done to detect environmental issues that need attention.

Ignoring boring/useless data to save computing power.

Now people started using Deep learning like **LSTM (Long short-term memory)** networks, a type of **RNN (Recurrent Neural Network)** to detect anomalies.

LSTM is great in understanding the patterns in sequence data (data over time)

Greenhouse is a new system that is designed to use smart machine learning and data techniques to anomalies quickly and accurately. It can handle larger amount of time-series data.

What's in the paper –

- How machine learning system work.
- Some early result using **TensorFlow** (a popular ML library).

TensorFlow is defined as an open-source platform and framework for machine learning, which includes libraries and tools based on Python and Java — designed with the objective of training machine learning and deep learning models on data.

Greenhous follow **sliding windows approach**, meaning it keeps moving forward and predicting based on previous chunks of data.

There are basically two phases of greenhouse –

1. Training phase

   During this phase it trains though LSTM model and set a threshold to detect anomaly.

2. Inference phase (Detecting Anomalies)

   When we feed new data, it checks how far those errors are form the normal pattern if the error is above the threshold, it marks is anomaly.

Greenhouse was tested against another system called LSTM-AD using two real-world datasets:

- Twitter AAPL (Apple stock-related data from Twitter)
- Nyc taxi (Taxi trip data from New York City)

key findings –

- Greenhouse had higher precision than LSTM-AD (meaning it made fewer false alarms).
- Greenhouse learns from normal data it doesn't require any anomaly data to train, this feature make is unique and special.
- This system is specially designed for IoT devices where Realtime data monitoring is require

Advantages –

1. Greenhouse can be trained using only normal data, these feature makes is very effective as anomalies are rare and hard to label in real life.
2. Because of these zero positive learning approaches it makes is more practical and easier to deploy in real world IoT system.

Disadvantage –

1. Greenhous sometimes misses actual anomalies especially when compared with LSTM-AD.
2. This could limit its use in applications with irregular data collection intervals.

Ongoing Research

The researchers are working to improve and expand Greenhouse so that it can become a complete, real-time anomaly detection system for IoT.

1. Detecting Range-Based Anomalies
2. Real-Time (Online) Anomaly Detection
3. Using Human Feedback

Conclusion –

Greenhouse is a smart system that can detect unusual patterns (anomalies) in time-series data without needing any examples of those anomalies during training. It uses an LSTM-based model and works well even with smaller training datasets. The system is efficient, accurate, and suitable for IoT applications where real-time monitoring is important. Ongoing work is focused on making it faster, more flexible, and ready for real-world deployment across devices and networks.

Reference –

Published at - https://arxiv.org/abs/1801.03168

Authors - Tae Jun Lee(Microsoft),

Justin Gottschlich, Nesime Tatbul (Intel Labs)

Eric Metcalf, Stan Zdonik (Brown University)

Research paper 2 -: A Practical Guide to Quantifying Carbon Emissions for Machine Learning researchers and practitioners

Summary

This research paper talks about how machine learning contributes to the Carbon emission and climate change. Not only during training the model but even the machine is in still or stationary state it produces carbon emission. Just like car and factories releases carbon dioxide (Co2). For training and using machine learning model require a lot of electricity, which comes from source that pollute the environment.

The paper says Co2 is harmful in excess it basically traps the heat and causes global warming.

The machine learning model needs electricity to run and function even for training the model, they use powerful computer that consume a lot of energy. Most of the research paper focus on training phase's carbon emissions, but:

- **Training** uses high energy when running big models and datasets.
- **Inference** (when the model is used daily life, like in Alexa, Google Search, or Netflix recommendations) also adds up over time.

Not only the model but the server running on ML system use energy even when it is stationary. The whole infrastructure requiring a lot of energy like cooling system routers and storage devices. Data centres are measured using a term called PUE (Power Usage Effectiveness) this tells how much extra energy is spent on things like cooling.

PUE = Total Facility Energy ÷ IT Equipment Energy

The average reported data centre PUE is 1.58, meaning that around 37% of the energy consumed is used for things like data centre cooling, lighting and distribution.

LCA (Life cycle Assessment) is important as it says we should also count environmental cost of: (Production to disposal)

- Making servers and chips
- Transporting them
- Maintaining them
- Disposing of them

Next thing paper talks about offsetting – It means compensating for the pollution that can't be avoided. It basically means when you ml mode uses electricity that causes emission you can pay for that by investing in green projects like planting trees or renewable energy but it is not the solution one should focus on reducing the carbon emission rather than offsetting. Offsetting should be the second solution.

As it is better to reduce your emissions in the first place, rather than pollute and then try to fix it later.

As a Machine Learning practitioner:

- Use smaller datasets when testing your model to avoid unnecessary energy use.
- Avoid saving multiple copies of the same data — share files within your team.
- Choose data centre that run on green energy (like using Electricity Map to pick the least polluting region).
- Reuse pre-trained models instead of training from scratch every time.

As an institution:

- Run ML jobs in areas with clean energy.
- Enable carbon tracking tools by default on your servers.
- Limit the time of training to reduce the wating of energy.
- Run awareness campaigns so everyone understands ML's impact on the environment.

Don't rely on carbon offsetting as a first solution it's only for emissions you can't avoid. Instead, focus on reducing waste, using clean energy, and tracking your emissions to make Machine Learning more eco-friendly.

Tools are suggested to measure ML emissions are - CodeCarbon, Carbon Tracker, Experiment Impact Tracker, Green Algorithms

Advantage –

1. Raises awareness and educates ML practitioners and institutions about the hidden carbon emissions.
2. Provides actionable tips, gives practical steps to reduce emissions like choosing green data centres, using tracking tools.

Disadvantage-

1. Carbon emission values are approximate and may vary depending on region, hardware, and usage pattern.
2. No deep technical model discussion. The paper focus on awareness and ethics not on proposing a new model or architecture.

Conclusion

The paper says that the impact of carbon emission is not just on machine learning or not during the training but across the entire ML lifecycle. Form hardware use to daily usage. It shows that ML can indirectly harm the environment through high energy usage, especially when large models are trained frequently

Instead of relying on carbon offsetting, the paper encourages practitioners and institutions to reduce emissions directly by making smarter and greener choices. They should focus on optimising the model so that it uses less energy for during the working and traning of the model

Reference –

Authors – Anne-Laure Ligozat, Sasha Luccioni

Published -

https://hal.science/hal-03376391/file/A_Practical_Guide_to_Quantifying_Carbon_Emissions.pdf

Research paper 3: - A Systematic Review of Anomaly Detection Using Machine and Deep Learning Techniques

Summary –

This paper is an overview of various method used to identify the anomalies, unusual behaviour and pattern in data that is know as anomalies.

Anomaly detection –

Anomaly detection basically means finding the unusual pattern or behaviour in the data. These anomalies can indicate critical issue such as fraud detection system failure or security breaches. It's basically finding any unusual behaviour in the data.

Techniques discussed in the paper –

Machine learning models -

- KNN (k-Nearest Neighbour)
- SVM (Support vector Machine)
- Decision tree

Deep Learning models –

- Neural network (Mimic the human brain)
- Autoencoder (useful in anomaly detection)
- Generative Adversarial Network GANs (It consist of generator and discriminator they work together to detect anomalies)

The paper talks about some of the good ways to check how good the model is on finding the anomalies

1. Accuracy
2. Precision
3. Recall
4. F1-score (A balance between precision and recall)

The paper basically reviews the existing machine learning model and deep learning that is used to detect anomalies. In this paper the datasets discussed are publicly available and commonly used in the research community. They can be accessed through academic repositories or official websites associated with each dataset.

Comparing to the other research paper following this make this paper special

It includes both ML and DL approaches. It basically Summarizes datasets used in research. It covers almost all the real-world application such as :

Surveillance, IoT (Internet of Things) Finance, Healthcare, Traffic monitoring

It also talks about the key challenges in the Anomaly detection like is there is poor dataset or the data contain Noise and complexity (Real world data is messy) then it will be very difficult for the model to detect anomaly.

In some of the cases system demand to the anomaly detection to be in real-time that require continuous power supply. Some cases like fraud detection or in real-time monitoring of the patient health it will require the model to detect the anomaly instantly as it occurs otherwise it may cause some serious issue.

Advantages:

1. The paper explains many different ML and DL methods for detecting anomalies.
2. It talks about how anomaly detection is used in real life, like in cybersecurity, finance, and healthcare making the research practical and relevant.

Disadvantages:

1. The paper mostly summarizes existing methods but doesn't test or compare them with real data.
2. It doesn't deeply explore recent and advanced topics like AI or energy-efficient models.

Future Direction -

The future of anomaly detection can be improved in many ways. First, collecting more detailed data can help models learn better. Using smarter machine learning techniques can boost performance. Cross-Checking should be done to test models more accurately and reduce false alarms. Lastly, working together with government agencies can help test these models in real-life situations that can lead to better solutions.

Conclusion –

This paper gives a clear summary of how machine learning and deep learning are used for finding unusual patterns (anomalies) in data. It focuses on two main goals: first, reviewing recent research from 2019 to 2021, and second, discussing the accuracy methods used in those studies. The paper also explains where anomaly detection is used in real life and highlights the newest techniques used in the field.

It includes a list of datasets used in experiments, many of which come from real-world situations. The authors found that some parts of this field are still very new and need more research.

Reference –

Authors –

Sarfaraz Natha, Department of Software Engineering, SSUET, Karachi, Pakistan

Mehwish Leghari, Department of Artificial Intelligence, QUEST, Nawabshah, Pakistan.

Muhammad Awais Rajput, Department of Artificial Intelligence, QUEST, Nawabshah, Pakistan.

Syed Saood Zia, Department of Artificial Intelligence, QUEST, Nawabshah, Pakistan.

Jawaid Shabir, Department of Computer Engineering, SSUET, Karachi, Pakistan.

Published – https://www.researchgate.net/publication/365193314_A_Systematic_Review_of_Anomaly_detection_using_Machine_and_Deep_Learning_Techniques

Research paper 4: - Green Machine Learning (ESANN 2023 special session)

Summary –

The paper is of ESANN (European symposium on Artificial Neural Network) that occurred in 2023. It tells about Green Machine learning, an approach that aims in making machine learning more environment friendly, sustainable, accessible and efficient by Reducing energy consumption, making high power performance without requiring heavy computational resource.

In past few years the use of Artificial intelligence and machine learning has been increased they have brought a revolution in the industries, the efficiency and accuracy have been achieved in past few years in various field like healthcare, finance transportation education and entertainment.

In the next coming years, this energy consumption is projected to multiply, potentially reaching over 30% of the world's total energy consumption by 2030. GPT-3 consumed 700,000+ litters of water, highlighting non-obvious environmental impacts. ChatGPT with GPT-3.5 allegedly consumed 1,287 megawatts and generated 552 metric tons of carbon dioxide emissions during its training, as reported by various sources. Not to mention that the more recent GPT-4 is estimated to be 10 times larger.

The paper also talks about –

Green by AI: Using AI to solve environmental problems

Green in AI: Making AI systems themselves more energy-efficient and environmentally friendly.

Researchers have identified that a harmful trend know as Red AI is growing so fast which is not good for environment. This refers to AI that gets better results by huge amount of computing power ignoring the energy efficiency. A study in 2018 showed that the computing power needed for training AI models has been doubling every 3 - 4 months that is much faster than what was expected. Red AI usually focuses only on accuracy, ignoring energy efficiency. Green AI tries to balance both.

The paper also talks about many ways or several practice to make AI more eco-friendly –

- Algorithm optimisation
- Hardware optimisation
- Edge computing - instead of sending data to large data centre processing it directly on device
- Data center choices - Establishing data center in countries where clean energy is used

The paper also tells the tools to track carbon footprint, the tools like carbonTracker and green algorithm helps researchers to measure how much energy and Co2 their AI model is using.

Unique in the paper –

The ESANN 2023 special session on Green Machine Learning highlighted several innovative approaches aimed at enhancing the sustainability and efficiency of machine learning (ML) systems. Researchers are finding ways to make machine learning (ML) more energy-efficient and eco-friendly.

- One study used mutual information to pick only the most important data features, saving time and power.
- Another used a clever math trick (logarithmic division) to simplify calculations.
- A third project focused on train maintenance, using ML to spot wheel problems quickly and efficiently.
- The fourth study used drones to monitor crops using multispectral images (which capture more than just visible light).

All these approaches aimed to reduce computational waste and power usage. They also focused on making models faster. Overall, the session showed how ML can be both smart and sustainable.

Advantage –

1. The paper tells the way to reduce the energy consumption and helping the machine learning model to become more sustainable and eco-friendlier.
2. It shows the practical use cases (like agriculture) proving that machine learning model can be both smart and efficient.

Disadvantage –

1. Some method suggested in the paper might work well in some of the specific domain but may not perform well with other areas or large application.
2. The techniques mentioned in the paper like mutual information optimization or spectral band selection using neural networks can be hard to implement without technical expertise.

Conclusion –

The paper shows that Machine Learning can be made more eco-friendly by designing energy-efficient models and processes. It highlights smart techniques like feature selection, early exits in neural networks, and optimized sensor placement to reduce computational load. These methods not only help in saving energy but also improve speed and efficiency in real-world applications like agriculture and predictive maintenance.

The authors tell how important it is to have the balance of performance with sustainability especially as AI system grows larger. They called a collaboration between researcher and engineers to create a ML system that are both powerful and environment friendly.

Reference –

Authors – Veronica Bolon-Canedo, Laura Moran-Fernandez, Brais Cancela and Amparo Alonso-Betanzos

Published –

https://www.esann.org/sites/default/files/proceedings/2023/ES2023-3.pdf

Research paper 5: - A review of green artificial intelligence: Towards a more sustainable future.

Summary –

This paper discusses that green AI as the one of the most important approaches to improve the environmental sustainability of AI systems. They described the AI solution for eco-friendly practices in another field that is Green-by AI and Green-In AI.

Some authors have estimated that training GPT-3 on a database of 500 billion words required 1287 MWh of electricity and 10,000 computer chips, equivalent to the energy needed to power around 121 homes for a year in the USA. As an example, GPT-3 was accessed 590 million times in January 2023, leading to energy consumption equivalent to that of 175,000 persons.4 Moreover, in inference time, each ChatGPT query consumes energy equivalent to running a 5 W LED bulb for 1hr 20 min, representing 260.42 MWh per day.

This basically tells that over past few years the use of AI and ML has been increased and they have brought a revolutionary in the field by giving accurate prediction in various sector like healthcare, finance, education etc.

It's like AI is powerful, but it pollutes the environment, Modern AI models, like GPT-2 or Gopher, are very large language model and need huge amount of computing power to train and training these models release a lot of co2. For example, training GPT-3 emits as much CO2 as driving a car for its entire life or the average emissions of several human lives in a year.

The high energy usage form training big model, water usage to cool down the data centres, Greenhouse gas emission form electricity used these all thing makes AI harmful for environment. The solution is also proposed in the paper that is –

- Reducing the Environment impact of AI.
- Using optimised algorithm that require less energy and computational power to give output.
- Using cloud and edge devices like smartphone instant of giant server.

Low carbon footprint, smaller and faster models, clear and understandable decision-making, better data quality and less complexity etc are some of the features of Green AI. There are basically two types of AI –

- Green-By AI
- Green-In AI

Green-By AI – This basically means using AI to make the world more sustainable by reducing pollution, conserving energy etc. Energy Efficiency means using AI for making smart grid to balance the power supply and demand in real time to reduce the wastage of energy. AI helps to predict how much solar and wind energy is required or will be available form the coming

source. Using AI in Agriculture like smart irrigation system supplying only that much amount of water to the field as require, predicts the crops yields and diseases early.

Using AI in climate change and environment change. Predicting this thing in advance of minimise the destruction that will be caused by this thing.

Green-in AI – This section talks about making the AI itself more ecofriendly because training big data model can use a lot of energy. Like replacing red AI with Green AI, optimising the algorithm more and more so that I can produce accurate result with less power consumption. Hardware optimisation not only software but one should focus on optimising the hardware also. Making the chips that help the AI to perform its task more efficiently.

In Simple Words AI can help the planet by improving energy use, transportation, farming, and climate actions that's Green-by AI. But we also need to make sure AI itself doesn't waste too much energy that's Green-in AI.

The ML model grows larger and they use a lot of energy. So, it's important for researcher and developer to measure how much energy their models use and how much carbon they emit. Some of the tools to track energy use are –

CarbonTracker (Tracks CO_2 emissions using real-time data and models)

CodeCarbon (Helps developers track how much carbon their code generates)

Green Algorithm (Helps make computations more efficient and less wasteful)

The paper talks about the lightweight model used in the green AI.

Advantage –

1. It reviews the AI in both aspect theoretical and practical implementation the AI model that is energy used form development to deployment.
2. IT cover the practical tools like carbonTracker CodeCarbon, etc making is useful for researchers to track carbon emission.

Disadvantage –

1. Since its just a review on Green AI it does not present any new experiments only discuss the past results.
2. It talks about the current tools for measuring carbon footprint often give inconsistent results.

Conclusion –

Artificial Intelligence (AI) has huge potential to help the world—but it also uses a lot of energy, especially with large models. This paper shows that if we want AI to be helpful and eco-friendly we need to replace Red AI with Green AI, try to optimise the algorithms and large language model more and more in order to produce effective result in less power. In the end, the paper says that we need to care about the environment as much as we care about performance.

Measure and reduce the carbon footprint of AI systems using specialized tools Researchers, companies, and governments must work together to make AI more sustainable for the planet and future generations.

Reference –

Authors – Veronica Bolon-Canedo, Laura Moran-Fernandez, Brais Cancela and Amparo Alonso-Betanzos

Published –

https://www.sciencedirect.com/science/article/pii/S0925231224008671

Research paper 6: - Time Series Anomaly Detection Using Convolutional Neural Networks and Transfer Learning

Summary –

Time series anomaly plays an important role in automated monitoring system. In Most of the deep learning model the time anomaly detection used were based on RNN (recurrent neural network). In this paper convolutional neural network approach is used for anomaly detection. It is an increasingly important topic today, because of its wider application in the context of the Internet of Things (IoT), especially in industrial environments.

The use of AI and ML has been widely grown in past few years and in most of the cases the anomaly detection is very important. In this research, we created a CNN-based deep network for time series anomaly detection. In particular, we were inspired by a successful image segmentation network, U-Net, and applied a time series version of U-Net to detect anomalous segments in time series. Before deep learning became popular, people used traditional math-based methods or basic machine learning. Then came deep learning. **RNNs** (Recurrent Neural Networks) became the popular choice because they naturally handle sequences

CNN based Time series segmentation means understanding the time series data using convolutional neural network. It basically works on image but here we will use time series data that is 1D.

In the paper U-Net is used for time series which was originally made for medical image segmentation, but here we have modified it to use for time series segmentation. Here we have used 1D version of U-Net to predict weather there is an anomaly in the data or not. It has a U-shaped structure that's why it is called U-Net. It basically consists of two main parts an encoder and a decoder. U-Net basically helps in labelling each time point like detecting if it's an anomaly or not. In U-Net, skip connections are like shortcuts between the encoder and decoder. Normally, when data passes through the encoder, some small details can get lost as it gets compressed. Skip connections help by directly passing those details from the encoder to the decoder at the same level.

The paper also talked about MU-Net (Multivariate U-Net) When you train a U-Net model on univariate time series data and later try to use it on multivariate data there's a problem:

The first layer of U-Net was made for 1 channel. Now you have many channels. so, if you just copy-paste the weights from the old model to the new one, the model gets confused and here comes the concept of MU-Net (Multivariate U-Net). One solution was proposed that why can't just add up all the channels but it was not practically possible as adding them doesn't make sense and loses important information.

It starts by splitting each channel separately. Each channel goes through its own encoder, reusing the same pretrained univariate model weights. Then, the outputs from all channels are combined. After that, the model continues with shared encoding, decoding, and output layers just like in regular U-Net.

Experiments conducted in the paper are –

1. Dodgers Loop Sensor Dataset (Univariate with sufficient data)

- Description: Traffic data near Dodger Stadium across 28 weeks.
- Goal: Detect anomalies during baseball games.
- Test - Total 81 events occurred out of which 42 were used for training and 39 were used for testing the model.
- Result: Detected 36/39 events correctly. Few false positives, often due to missing data.

2. Gasoil Plant Heating Loop Dataset (Multivariate with sufficient data)

- Dataset Source: Secure Water Treatment (SWaT)
- Description: 48 sequences of gasoil plant control simulations with cyber-attack labels.
- Goal: Detect cyber-attacks.
- Result: Detected 21/22 attacks with 3 false positives.

3. Synthetic Curves with Unusual Shapes (Univariate with insufficient data, used transfer learning)

- Dataset Source: Custom synthetic dataset with diverse curve shapes.
- Description: Detect unusual shapes in time series.
- Goal: More complex anomaly detection than in pretraining.
- Significance: Demonstrates value of transfer learning in low-data situations.

4. Electromyography (EMG) Dataset (Multivariate with transfer learning)

- Dataset Source: UCI EMG Dataset
- Description: 8-channel signals for 7 hand gestures across 36 subjects.
- Goal: Gesture segmentation (multi-class single-label).
- Significance: Validates MU-Net's effectiveness for multivariate transfer learning.

Advantage –

1. Introduces MU-Net architecture for multivariate transfer.
2. Proposes CNN-based segmentation instead of RNNs for time series.

Disadvantage –

1. Focused only on CNNs; no performance comparison with RNNs or classical methods.
2. MU-Net increases model complexity.

Conclusion –

In this study, the authors proposed a U-Net-based deep learning model for anomaly detection in time series data, which is the first to use CNNs for segmenting time series in this context. The model was tested on both univariate and multivariate data and showed good performance. The authors suggest that future work should include comparing their approach with other models and improving the previously trained dataset to make the model even more transferable.

Reference –

Authors – Tailai Wen, Roy Keyes

Published –

https://arxiv.org/pdf/1905.13628

Research paper 7: - Machine Learning for Anomaly Detection in Industrial Environments

Summary –

The paper tells anomaly detection is one of the important aspects as the part of safety and production management. This work aims to deliver an overview of the use of machine learning for anomaly detection. It also discusses the challenges for future research. The research finds that despite of such advance technologies there is no scope for improvement in anomaly detection through machine learning.

As industries keep updating their technologies it became very important to quickly detect any un usual patter or error called anomaly in the production process in order to prevent accidents protect workers and save money on fixing it. Machine learning is good at spotting error or unusual pattern in data that might mean something is wrong. But there are some challenges that is anomalies are rare so algorithm might miss them or raise too many false alarms. One of the big areas where anomaly detection is used is cybersecurity. ML helps to detect any unusual pattern or any strange activity on computer or network, like virus, hacker attacks etc. The paper looks at many research studied and summarize the types of ML used for anomaly detection in industry.

The paper focuses on 3 main questions:

1. What are the most recent studies on anomaly detection in industry?
2. Which ML models work best for detecting these anomalies?
3. How does ML improve safety in industrial environments?

While all these studies cover different aspects of anomaly detection, none of them focus only on machine learning in industrial environments, that's what this paper is about –

The goal is to give a complete overview that how ML is used for Anomaly detection in the industry setting only.

In this paper a systematic method was used called PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) PRISMA is a well-known method used in research to make sure reviews are structured and unbiased. It helps researchers to set a clear rule for including and excluding the studies, avoid duplicate studies, colleting accurate data. They searched four trusted online research databases: IEEE Xplore, Scopus, Science Direct, Google Scholar. To find all relevant research, they searched using terms like:

- Anomaly detection
- Industrial environment
- Machine learning
- Fault detection

The studies that were published between 2015 to 2023 were included in this research. They mainly focused on anomaly detection through machine learning. After applying all these rules,

they found 10 high-quality articles that matched their criteria and used them in their review. After reviewing the 10 articles they found that –

There was no single best model each one of them preformed differently depending on type of data. Random Forest models performed well in multiple studies. Isolation Forest and other unsupervised methods are effective for certain types of data like network intrusion. There's a clear need to balance accuracy with speed/training time, depending on the industrial use case.

The paper reflects on the research findings and offers a broader view of - Which ML models are commonly used and why, what factors affect their performance What challenges still exist in applying ML to industrial anomaly detection. The most widely used ML models across the reviewed studies include:

- Random Forest (RF)
- k-Nearest Neighbours (k-NN)
- Decision Trees (DT)
- Support Vector Machines (SVM)
- Artificial Neural Networks (ANN)

These models are generally good performers in detecting anomalies in various industrial scenarios.

Despite of having good result the field is not yet perfect. The research paper implies that more research needs to be done in the field like Enhancing Accuracy, Reducing the training time. The paper says - *"We looked at a lot of studies, and while some models work very well, the real-world performance depends on the specific situation. Choosing the right ML model isn't just about picking the most accurate one—it's about finding what fits best for the problem".* It means it is not necessary to choose the most accurate model but one should choose the best fit one.

Advantage –

1. This paper talks about Early and Accurate Anomaly Detection
2. Choosing the right ML model isn't just about picking the most accurate one, it's about finding what fits best for the problem

Disadvantage –

1. The effectiveness of ML models varies significantly with different datasets. A model that performs well on one dataset may fail on another due to data imbalance.
2. Some models especially deep learning requires significant computing resources and time for training

Conclusion –

This review highlights the growing importance of machine learning-based anomaly detection in industrial environments. Various models such as Random Forests, Decision Trees, SVMs, k-NNs, and Neural Networks have shown promising results in identifying faults and unusual

pattern across various industrial datasets. However, model performance is highly dependent on the dataset's characteristics, such as size, quality, and domain specificity etc. Overall, machine learning presents an important approach for intelligent anomaly detection.

Reference –

Authors – Denitsa Grunova, Vasiliki Bakratsi, Eleni Vrochidou and George A. Papakostas

Published –

https://www.mdpi.com/2673-4591/70/1/25

Research paper 8: - Development and Deployment of Green Artificial intelligence

Summary –

Green AI is a new field of technology that focus on developing or creating artificial intelligence system that are eco friendly and good for environment. The main aim to create such AI is to reduce the negative impact of AI, making an AI that works faster and more accurate even keeping the environmental factor. Green AI can help in many ways following are some about which the paper talks about –

- Running smart electricity grids more efficiently
- Using less energy in buildings
- Helping farmers grow more crops
- Improving public transport
- Managing waste better

All these uses can help in making the life more sustainable, reducing the pollution and saving the natural resources.

However, there is a downgrade of this, creating AI that helps in various part of our day-to-day life and training it require a lot of energy and computational power that cause pollution, carbon emission which is not good for environment. So, its important to find ways to reduce the energy usage and make sure we have good policies and rules for making AI.

Making AI is not just about saving energy but it is also important to create a better faster and more sustainable world. In this paper we talked about -

- How Green AI is being developed and used
- The methods and tools used to make AI more eco-friendly
- Some of the problems and challenges in doing this
- Real-life examples of companies using Green AI

Regular AI system uses a lot of electricity to work and produce accurate result. If this continues then one day it will lead to end of power source. It basically leads to more and more carbon emission, which is getting worst for our environment. Also, there is a lot of power is needed to store the data, move data and keeping the serves and machine cools. The big data centres that power AI and other digital technologies are already responsible for about 2% of the world's carbon emissions, and this number is expected to grow.

The use of Green AI brings many benefits like It is less harmful for environment. It predicts accurate results with low computational power and electricity. Green AI is not all about saving the money or environment, it's a social responsibility that we need to develop and use technologies that help fight climate change and protect the planet. Developing Green AI is not helping save environment but is also improve the Green ML which has so many positive aspects in environment like –

- Lower Environmental impact - Machine Learning (ML) models, especially large ones like deep learning or LLMs, use a lot of electricity during training and deployment.
- Reduced Energy Costs - Training ML models on massive datasets takes a lot of computing power, which costs money, but green because of green ML it is reduce as

  Small datasets = Reduce energy use = lower electricity bill

- Wider Access to ML - If ML models are made lighter and more efficient. They can run on cheaper or low-power devices it can be used by more people especially of developing countries.
- Green ML encourages the creation of new algorithms that require fewer resources.

Green Machine Learning has huge potential to make AI more sustainable and fair but there are real technical, financial, and policy challenges that need to be tackled with collaboration from researchers, industry leaders, and governments.

Future of Green Machine learning is bright as the need of ai is increasing day-by-day and power supply is limited so every one will start shifting toward green machine learning and green ai. Following thing will be done as future of green machine learning-

Using Renewable Energy with ML, In the future, Machine Learning systems can be designed to run on solar, wind, or other green energy sources.

creating New Energy-Saving ML Algorithms We can build new types of ML algorithms and models that, use less data and fewer computing resources to predict the accurate output and results.

Improving Current ML systems, we don't always need to start from scratch. We can make existing ML models smaller, faster, and more efficient by optimising them up to there best in every scenario.

Advantages –

1. Environment Friendly, Green AI helps to reduces the amount of energy used by AI systems, which directly helps lowering the carbon emissions and protects the environment.
2. Cost Saving, it basically helps companies and organizations save money by cutting down on electricity and hardware costs, making AI more affordable that can be used by many people.

Disadvantages –

1. Energy vs. Accuracy - Making an ML model energy-efficient sometimes leads to reduce in its performance or accuracy.
2. Training efficient models still often requires large datasets.

Conclusion –

Green Artificial Intelligence is a smart and necessary step toward a cleaner and more sustainable future. It helps reduce the harmful impact of traditional AI by using less energy and making systems more efficient. This not only helps protect the environment but also saves money and resources. The paper discuss how Green AI is important and leads to make more and more green machine learning model so that they can run on minimal supply of electricity and produce less carbon emission which helps in reducing its adverse impact on environment. Green AI makes technology more accessible for smaller companies and developing countries which directly makes the Machine learning cheaper for many countries to afford it and use it.

Reference –

Author - Dr. Anil Kumar Pandey Programmer, Computer Centre Banaras Hindu University, Varanasi -221 005

Published –

https://ijmcr.in/index.php/ijmcr/article/view/545/451