Google

COURSE
CERTIFICATE

Jun 17, 2022

Namit Shrivastava

has successfully completed

Foundations: Data, Data, Everywhere

an online non-credit course authorized by Google and offered through Coursera

Amanda Brophy
Global Director of Google Career Certificates

Verify at:
https://coursera.org/verify/68VZPGP78G8T

Coursera has confirmed the identity of this individual and their participation in the course.

DeepLearning.AI

COURSE CERTIFICATE

Jul 4, 2022

Namit Shrivastava

has successfully completed

AI For Everyone

an online non-credit course authorized by DeepLearning.AI and offered through Coursera

Andrew Ng
Founder, DeepLearning.AI
Co-founder, Coursera

EDUCATION FOR EVERYONE · coursera · COURSE CERTIFICATE

Verify at:
https://coursera.org/verify/Q99PM7ZMP2PR

Coursera has confirmed the identity of this individual and their participation in the course.

INTERNSHALA TRAININGS

Certificate of Training

NAMIT SHRIVASTAVA,

student of Birla Institute Of Technology And Science, Pilani, has successfully completed a six weeks online training on **Machine Learning**. The training consisted of Introduction to Machine Learning, Data, Introduction to Python, Data Exploration and Pre-processing, Linear Regression, Introduction to Dimensionality Reduction, Logistic Regression, Decision Tree, Ensemble Models and Clustering (Unsupervised Learning) modules. In the final assessment, NAMIT scored 48% marks.

We wish NAMIT all the best for the future.

Sarvesh Agarwal

FOUNDER & CEO, INTERNSHALA

Date of certification: 2021-05-31 Certificate no. : E3EF0F6F-4745-2A92-3EF8-F01EE3589AD5

For certificate authentication, please visit https://trainings.internshala.com/verify_certificate

INTERNSHALA TRAININGS

Certificate of Training

NAMIT SHRIVASTAVA,

student of Birla Institute Of Technology And Science, Pilani, has successfully completed a six weeks online training on **Data Science**. The training consisted of Introduction to Data Science, Python for Data Science, Understanding the Statistics for Data Science and Predictive Modeling and Basics of Machine Learning modules.
In the final assessment, NAMIT scored 73% marks.
We wish NAMIT all the best for the future.

Sarvesh Agarwal
FOUNDER & CEO, INTERNSHALA

Date of certification: 2021-03-27 Certificate no. : 041EC5B7-AE0C-29EB-605B-4A92B57EAF8A

For certificate authentication, please visit https://trainings.internshala.com/verify_certificate

INTERNSHALA TRAININGS

Certificate of Training

NAMIT SHRIVASTAVA

from Birla Institute Of Technology And Science, Pilani, has successfully completed a six weeks online training on **Deep Learning**. The training consisted of Getting Started with Deep Learning, Artificial Neural Networks, Convolutional Neural Networks, Image Recognition Project and Final Evaluation modules.
In the final assessment, NAMIT scored 86% marks.
We wish NAMIT all the best for the future endeavours.

Sarvesh Agarwal
FOUNDER & CEO, INTERNSHALA

Date of certification: 2021-09-02 Certificate no. : CC359827-9E38-7337-CA7B-12826FF9EB81

For certificate authentication, please visit https://trainings.internshala.com/verify_certificate

smartknower
STAY AHEAD OF THE CURVE

MOOD INDIGO

CERTIFICATE

OF COURSE COMPLETION

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Namit Shrivastava

Has successfully completed course on Data Science
from the period of 1/6/2022 to 31/7/2022.
During his/her training we found the student was dedicated and hard working.

06-09-2022
DATE

P. Naveen
ACADEMIC HEAD

Certificate ID : 941163141

smartknower
STAY AHEAD OF THE CURVE

CERTIFICATE
OF INTERNSHIP

Best AWARD

THIS CERTIFICATE IS PROUDLY PRESENTED TO:

Namit Shrivastava

Has successfully completed Internship On Data Science
from 1/6/2022 to 31/7/2022. During his/her internship,
the student was found to be dedicated, hardworking and intelligent

24-08-2022
DATE

P. Naveen
Academic Head

Certificate ID: 3852370708

Birla Institute of Technology & Science, Pilani
Pilani | Dubai | Goa | Hyderabad

Presents this

certificate of completion

To **Namit Shrivastava**, bearing I.D No. **2020A2PS1767P** of Pilani campus for successfully completing the online course on "**API based Programming**" offerred as part of **BITS Pilani Postman Classroom Program** held during Jan 25th to March 10th 2021.

C. Jagat Seesh

Dr. Jagat Sesh Challa
Program Coordinator

5th December 2021
Certificate Issue Date

POSTMAN

innovate achieve lead

This badge was issued to Namit Shrivastava on January 29, 2023

Verify | Celebrate

Microsoft Certified: Azure AI Fundamentals

Issued by Microsoft

Earners of the Azure AI Fundamentals certification have demonstrated foundational knowledge of machine learning (ML) and artificial intelligence (AI) concepts and related Microsoft Azure services.

Learn more

Skills

Azure Bot Services | Azure Machine Learning | Cognitive Services

Earning Criteria

Pass Exam AI-900: Microsoft Azure AI Fundamentals

Endorsements

American Council on Education

This credential has been successfully evaluated by the American Council on Education for college credit. It is recommended for a total of 1 college credits. For more information about ACE Learning Evaluations, visit www.acenet.edu.

DP Cube Learning

ETS TOEFL

CERTIFICATE

OF PARTICIPATION

ENGLISH OLYMPIAD

THIS CERTIFICATE IS PRESENTED TO

NAMIT SHRIVASTAVA

from BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

for participating in the English Olympiad,
Organised by DP Cube Learning & Powered by ETS TOEFL
held from 26th September 2022 to 1st October 2022.

Rashmi Gera
Head - Academics

Karthik Rajasekaran
Head - Business Development

Achieving Sustainability in Supply Chain during Disruption times: Role of Industry 4.0

Namit Shrivastava[1] and Manoj K Srivastava[2][0000-0002-8701-4522]

[1] Birla Institute of Technology and Science, Pilani
[2] Management Development Institute, Gurgaon
f20201767@pilani.bits-pilani.ac.in, mks@mdi.ac.in

Abstract. The demand to digitalize the automotive sector, which entails linking manufacturers to a larger supply chain, is increasing. In the automobile industry, there is a higher requirement for sustainable growth due to rising supply disruption and frequent technology changes. Industry 4.0 can speed up manufacturing, increase customizability, and cut down on setup and lead times. It may result in innovation. The study focuses on establishing link between Industry 4.0 technologies and green supply chain practices, which will help in achieving sustainability during disruption times. It follows a qualitative survey approach to identify the prominent Industry 4.0 technologies and green supply chain practices using a fuzzy-set analytical hierarchy process. The other section uses interpretive structural modelling with a multi-level hierarchical structure to study the cause-and-effect relationship between the final selected Industry 4.0 technologies and GSC practices. The study identifies a strong linkage between Industry 4.0 technologies and green supply chain practices to achieve overall sustainability in the supply chain. The future automotive supply chain should focus on driving Industry 4.0 technologies for effective implementation of green supply chain practices. Also in the Indian automotive sector, government regulation and policies and top management commitment are two key factors for driving sustainability in the Indian automotive supply chain.

Keywords: Industry 4.0, Green supply chain, Fuzzy set Analytic Hierarchy Process

1 Introduction

Over the past decade, organizations have been working continuously towards achieving a sustainable and resilient supply chain. Assuring a balance between economic, social, and environmental growth, sustainability is described as meeting the needs of the present without compromising those of future generations. However, with COVID-19 and issues like semiconductor shortages, sea container shortages, the Russia-Ukraine war, and natural disasters such as earthquakes and tsunamis, the automobile industry is currently facing many new challenges in managing their supply chain operations. Such supply chain disruptions result in shortages of critical components, resulting in a loss of production.

The supply chain in the automobile industry in India is under considerable pressure from the government of India and society to pursue a more sustainable model of growth. The automobile industry in India is going through tough times. The growth has slowed down due to a liquidity crunch in the market. COVID-19 has worsened the situation in the first half of the financial year 2020–21. The supply chain in automobiles is playing a crucial role and withstanding the pressure of changing regulatory norms of the government and political dynamics of India, especially with its neighboring countries [1].

The structure schema of the paper is as follows: First, an introduction highlights the need for going towards sustainability practices and efforts in the supply chains specially in the disruption times. Second, a literature review is presented on Industry 4.0 and green supply chain practices. The third section describes the methodology, including a qualitative survey, a fuzzy set analytic hierarchy process, and interpretive structural modelling. The fourth section discusses the results and cause-and-effect relationship between these technologies and practices. Finally, conclusions, limitations, and future research are explained.

2 Literature Review

Several research articles on “Sustainable Supply chain” and “Industry 4.0” from relevant web, academic and research sources are presented below.

2.1 Industry 4.0

Industry 4.0 technologies are considered as the disruptive ones, such as the Internet of Things (IOT), 3D printing/additive manufacturing, cloud computing, block chain, etc. come under the purview of Industry 4.0, which has a huge potential to bring drastic change in the way current manufacturing is done in India. Let us investigate these technologies in detail in a subsequent section.

Additive Manufacturing also known as 3D printing, meets all essential requirements for bringing the Industry 4.0 revolution to the manufacturing sector. As the word "additive" suggests, it is the addition of material over its one layer to another in contrast to conventional manufacturing processes where material is chipped off to obtain required dimensions [2, 3, 4].

Internet of things (IOT) is the connection between physical objects or machines and the Internet. It enables the implementation of smart connected products or embedded sensors, resulting in machine-to-machine connectivity through the Internet [5, 6, 7, 8].

Blockchain is a type of distributed ledger technology with blocks of records that are linked securely through cryptography. The major benefit of Blockchain technology is that it ensures secure data storage and transactions to happen in a transparent and secured manner, preventing any unauthorized interaction during whole process [9, 10, 11].

Big Data Analytics (BDA) is characterized by a large volume of data with a wide variety, which requires specific analytical methods to transform it into valuable data.

Many organizations are spending a lot of money on training their employees to manage big data using BDA tools. It helps in taking decisions in a structured manner with reliable and real-time analysis [12, 13].

Cloud Computing (CC) refers to the idea that data can be stored, collected, and accessed from specialized shared data centers all over the world. Many organizations are now shifting to cloud-based data storage services such as Office 365, large database solutions, etc. [8].

2.2 Green Supply Chain practices

Traditionally the idea of incorporating sustainable supply chain operations is referred to as "sustainable" or "green supply chain" (GSC).

Green Purchasing (GP) is defined as purchasing the product or selecting a supplier that has a lesser effect on the environment or human health as compared to products serving the same purpose. It is also known as "environmentally preferable purchasing." This includes sourcing recyclable products, reusable raw materials, and products that do not harm the environment [14].

Green Design. Achieving sustainability through a green design approach is better, as under this practice, at the product design stage itself, the design is optimized, taking into consideration the energy and material requirements for manufacturing the design into the final product [15].

Reverse Logistics (RL). It is a type of supply chain management in which the flow of goods occurs from the customer back to the seller or manufacturer. The purpose is to retrieve maximum value from products and material disposed.

Supplier and Customer Collaboration. Sustainability is not a one-time process; it is an act of continual improvement where an organization needs to work together with all its vendor partners to achieve overall excellence [15].

Government regulation & policies. Regardless of their form, regulations and policies are the primary drivers for companies to plan and execute sustainable practices of supply chain in their organizations [16].

Top Management Commitment. It is very essential that the top management of an organization be strongly committed to achieving sustainability. This should also be clearly visible in the company’s vision and mission statements [16].

3 Research Methodology

The paper uses two-stages of data collection through expert opinions and analysis. In the first stage, key Industry 4.0 technologies and Green Supply Chain practices are identified, which are ranked through Fuzzy- Analytic Hierarchy Process (FAHP). Thereafter, the last ranked Industry 4.0 technology and GSC practices is removed, and balance factors are taken for further analysis. In the second stage, Interpretive Structural Modelling (ISM) is used to identify different hierarchical levels. Later, a cross impact matrix multiplication analysis (MICMAC) is done to find respective power of variables

in two categories, namely dependence and driving. Finally, the cause-and-effect analysis is done on the structure obtained through ISM.

3.1 Expert's profile and Data Collection

For creating contextual links between the variables in the interpretive structural modelling, the expert opinions are used. The selected experts have extensive experience in the automobile and manufacturing domains in India. They are automobile supply chain professionals and have good industry exposure related to the practical application of Industry 4.0 (I4.0) technologies and sustainable practices in supply chains. The experts' profile is highlighted in Table 1 below:

Table 1. Experts' profile

S. No	Expert	Experience	Industry	Domain	Designation
1	Expert 1	15	Automobile Industry	Research & Development	Deputy General Manager
2	Expert 2	16	Automobile Industry	Supply Chain professional	Assistant General Manager
3	Expert 3	23	Automobile Industry	Supply Chain professional	Vice President
4	Expert 4	21	Automobile Industry	Product Development	Assistant General Manager
5	Expert 5	14	Manufacturing Domain	Research & Development	Senior Manager

4 Results and Analysis

Table 2 suggests a list of alternative technologies and GSC practices as per the literature review.

Table 2. Set of Alternative I4.0 technologies and GSC practices

	Alternative I4.0 technologies		Alternative GSC Practices
IOT	Internet of Things	GP	Green Purchasing
3D	3D printing	SCC	Supplier/Customer Collaboration
BC	Blockchain	GD	Green Design
BDA	Big Data Analytics	RL	Reverse Logistics
CC	Cloud Computing	GRP	Govt Regulation & Policies
		TMC	Top Management Commitment

Initially, the AHP matrix is constructed by taking inputs from various experts with large experience in the automotive industry in India. For selecting options among five experts' opinions, the average of all expert opinions is calculated, which is further rounded to the closest integer in the relative importance scale (1 to 7) to obtain the final AHP matrix. Table 3 to Table 6 are showing various steps involved in Fuzzy- Analytic Hierarchy Process (AHP).

Table 3. Initial I4.0 technologies Matrix (CR=.09<0.1)

I4.0 Technologies		IOT	3D	BC	BDA	CC
Internet of Things	IOT	1.00	3.00	5.00	1.00	3.00

3D printing	3D	0.33	1.00	4.00	1.00	3.00
Blockchain	BC	0.20	0.25	1.00	0.20	0.20
Big Data Analytics	BDA	1.00	1.00	5.00	1.00	5.00
Cloud Computing	CC	0.33	0.33	5.00	0.20	1.00

Table 4. Initial GSC Practices Matrix (CR=.094<0.1)

		GP	SCC	GD	RL	GRP	TMC
Green Purchasing	GP	1.00	0.33	0.50	2.00	2.00	2.00
Supplier/Customer Collaboration	SCC	3.03	1.00	3.00	4.00	4.00	3.00
Green Design	GD	2.00	0.33	1.00	2.00	2.00	1.00
Reverse Logistics	RL	0.50	0.25	0.50	1.00	0.50	0.17
Govt Regulation & Policies	GRP	0.50	0.25	0.50	2.00	1.00	2.00
Top Management Commitment	TMC	0.50	0.33	1.00	6.00	0.50	1.00

Table 5. I4.0 Technologies weighted Fuzzy set matrix.

I4.0 Technologies		**IOT**	**3D**	**BC**	**BDA**	**CC**	**AHP Weights**	**Fuzzy AHP Weights**	**Ranking**
Internet of Things	IOT	(1,1,1)	(2,3,4)	(4,5,6)	(1,1,1)	(2,3,4)	34%	34%	I
3D printing	3D	(1/4,1/3,1/2)	(1,1,1)	(3,4,5)	(1,1,1)	(2,3,4)	21%	21%	III
Blockchain	BC	(1/6,1/5,1/4)	(1/5,1/4,1/3)	(1,1,1)	(1/6,1/5,1/4)	(1/6,1/5,1/4)	5%	5%	V
Big Data Analytics	BDA	(1,1,1)	(1,1,1)	(4,5,6)	(1,1,1)	(4,5,6)	30%	30%	II
Cloud Computing	CC	(1/4,1/3,1/2)	(1/4,1/3,1/2)	(4,5,6)	(1/6,1/5,1/4)	(1,1,1)	11%	11%	IV

Table 6. GSC Practices weighted Fuzzy set matrix

Practices		**GP**	**SCC**	**GD**	**RL**	**GRP**	**TMC**	**AHP Weights**	**Fuzzy AHP Weights**	**Ranking**
Green Purchasing	GP	(1,1,1)	(1/4,1/3,1/2)	(1/3,1/2,1)	(1,2,3)	(1,2,3)	(1,2,3)	15%	15%	III
Supplier/Customer Collaboration	SCC	(2,3,4)	(1,1,1)	(2,3,4)	(3,4,5)	(3,4,5)	(2,3,4)	37%	37%	I
Green Design	GD	(1,2,3)	(1/4,1/3,1/2)	(1,1,1)	(1,2,3)	(1,2,3)	(1,1,1)	16%	16%	II
Reverse Logistics	RL	(1/3,1/2,1)	(1/5,1/4,1/3)	(1/3,1/2,1)	(1,1,1)	(1/3,1/2,1)	(1/7,1/6,1/5)	6%	6%	VI
Govt Regulation & Policies	GRP	(1/3,1/2,1)	(1/5,1/4,1/3)	(1/3,1/2,1)	(1,2,3)	(1,1,1)	(1,2,3)	11%	12%	V
Top Management Commitment	TMC	(1/3,1/2,1)	(1/4,1/3,1/2)	(1,1,1)	(5,6,7)	(1/3,1/2,1)	(1,1,1)	14%	13%	IV

There is a negligible difference in weights calculated from the normal AHP method and the fuzzy AHP method in both I4.0 technologies and GSC practices. It can be seen

from Table 7 that the Internet of Things (IOT), 3D-printing and others are the best choices among the alternative technologies for attaining supply chain sustainability. Blockchain is the least preferred choice among I4.0 technologies for achieving sustainability. Also, it can be seen from Table 8 that effective collaboration between supplier and customer is the most preferred choice to achieve sustainable automotive supply chain. However, reverse logistics is the least preferred choice among green supply chain practices.

Further below set of I4.0 technologies and GSC practices are taken to study the cause-and-effect diagram for achieving sustainability.

Table 7. Final selected I4.0 technologies and GSC practices for ISM modelling

	Alternative I4.0 technologies		Alternative GSC Practices
IOT	Internet of Things	SCC	Supplier/Customer Collaboration
3D	3D printing	GD	Green Design
BDA	Big Data Analytics	GRP	Govt Regulation & Policies
CC	Cloud Computing	TMC	Top Management Commitment
GP	Green Purchasing		

4.1 Interpretive Structural Modelling (ISM) for I4.0 technologies and GSC practices

In ISM methodology, the use of expert opinions is recommended, as they give their response based on practical experience in the subject domain. For selecting options among five experts' opinions, options with a clear majority are taken, and any arbitration with an option without a clear majority is pursued with revisiting the experts. SSIM was created with input from experts in two prestigious domains: I4.0 and GSC, as below. Table 8 to Table 10 are showing various steps involved in ISM modelling.

Table 8. Structural Self Interaction matrix (SSIM)

Structural Self-Interaction Matrix	Factors	3D	BDA	IOT	CC	SCC	GP	GD	TMC	GRP
3D Printing	3D	-	A	A	A	X	X	X	A	A
Big Data Analytics	BDA		-	X	X	X	V	V	A	O
Internet of Things	IOT			-	X	X	V	V	A	O
Cloud Computing	CC				-	A	V	O	A	O
Supplier/Customer Collaboration	SCC					-	X	X	A	A
Green Purchasing	GP						-	X	A	A
Green Design	GD							-	A	A
Top Management Commitment	TMC								-	A
Govt Regulation & Policies	GRP									-

Table 9. Final Reachability matrix

Final Reachability Matrix		Factors	1 3D	2 BDA	3 IOT	4 CC	5 SCC	6 GP	7 GD	8 TMC	9 GRP	Driving Power
1	3D Printing	3D	1	1*	1*	1*	1	1	1	0	0	7
2	Big Data Analytics	BDA	1	1	1	1	1	1	1	0	0	7
3	Internet of Things	IOT	1	1	1	1	1	1	1	0	0	7
4	Cloud Computing	CC	1	1	1	1	1*	1	1*	0	0	7
5	Supplier/Customer Coll.	SCC	1	1	1	1	1	1	1	0	0	7
6	Green Purchasing	GP	1	0	0	0	1	1	1	0	0	4
7	Green Design	GD	1	0	0	0	1	1	1	0	0	4
8	Top Management Commitment	TMC	1	1	1	1	1	1	1	1	0	8
9	Govt Regulation & Policies	GRP	1	1*	1*	1*	1	1	1	1	1	9
		Dep. Power	9	7	7	7	9	9	9	2	1	

In order to determine the levels of partitions, each factor's reachability, antecedent, and intersection set are found from the final reachability matrix. Following 3 iterations levels are obtained from final reachability matrix as shown in below Table 10.

Table 10. Levels of Partitions

Factors	Reachability Set	Antecedent Set	Intersection Set	Level
3D Printing	1,2,3,4,5,6,7	1,2,3,4,5,6,7,8,9	1,2,3,4,5,6,7	I
Big Data Analytics	1,2,3,4,5,6,7	1,2,3,4,5,8,9	1,2,3,4,5	II
Internet of Things	1,2,3,4,5,6,7	1,2,3,4,5,8,9	1,2,3,4,5	II
Cloud Computing	2,3,4,5,6,7	1,2,3,4,5,8,9	2,3,4,5	II
Supplier/Customer Collaboration	1,2,3,4,5,6,7	1,2,3,4,5,6,7,8,9	1,2,3,4,5,6,7	I
Green Purchasing	1,5,6,7	1,2,3,4,5,6,7,8,9	1,5,6,7	I
Green Design	1,5,6,7	1,2,3,4,5,6,7,8,9	1,5,6,7	I
Top Management Commitment	1,2,3,4,5,6,7,8	8,9	8	III
Govt Regulation & Policies	1,2,3,4,5,6,7,8,9	9	9	III

The final reachability matrix and levels of partitions are used to develop the ISM model, which incorporates selected I4.0 technologies and GSC practices from fuzzy AHP technique to ensure sustainability in the automotive supply chain. as shown in below Figure 1.

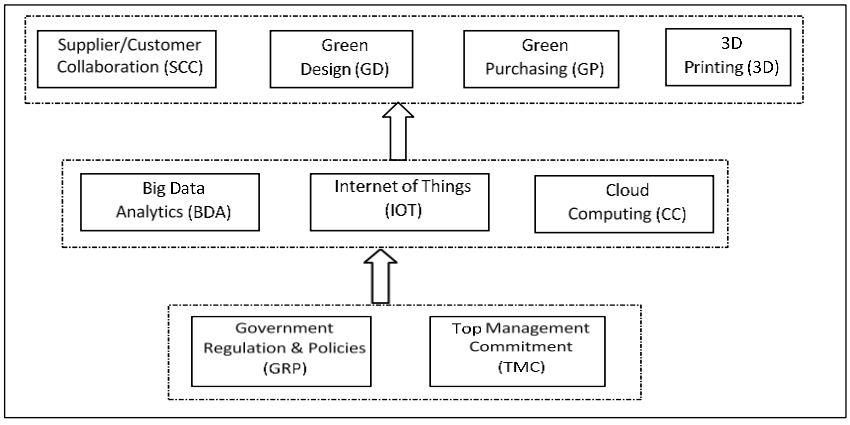


Fig. 1. ISM depicting levels of I4.0 technologies & GSC practices for achieving sustainability. (Source: Prepared by authors)

The above ISM model depicts that for achieving sustainability in the automotive supply chain in India, government regulation and policies and top management commitment towards sustainability are the most essential factors. It is followed by the Internet of Things, big data analytics, and cloud computing, which are directly or indirectly driving green supply chain practices to achieve overall sustainability. These factors are causing supplier and customer collaboration to be achieved in an effective manner.

4.2 Factor Classification using MICMAC Analysis

The driving power and dependence power of factors are further examined using a cross-impact matrix multiplication analysis (MICMAC). It is done to pinpoint the primary driving forces behind the system across different domains. There are four sorts of elements based on driving and dependence power.: autonomous factors (being less connected to the system and having weak driving & dependence power), linkage factors (having both strong driving & dependence power), dependent factors (having weak driving & strong dependence power) and driving factors (having strong driving power but weak dependence power). The results of the MICMAC analysis are shown in Figure 2 as below.

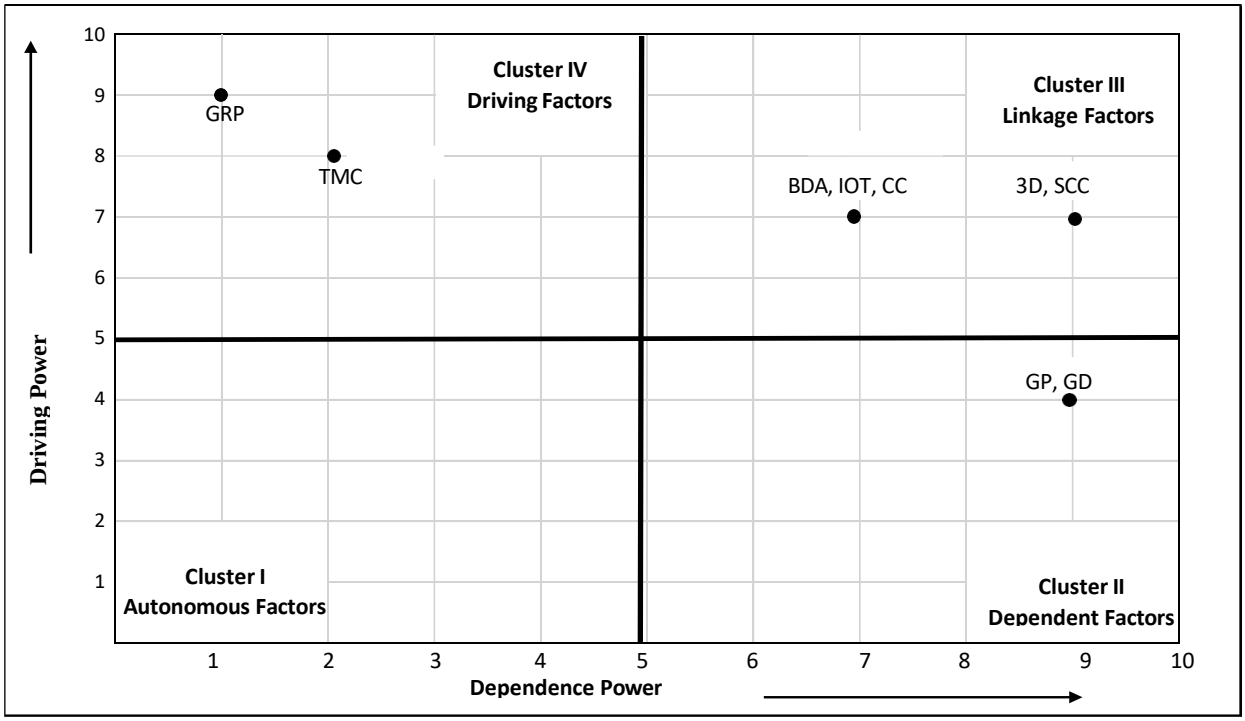


Fig. 2. Factor classification in MICMAC analysis (Source: Prepared by authors)

There are no autonomous factors present in the MICMAC analysis, as shown in Figure 2, hence all the Industry 4.0 technologies, considered and green supply chain practices are important. Top management commitment and government regulation and policies are driving factors that affect sustainability's performance because they are the driving dependent factors. Among them, government regulation and policies are key factors with highest driving power.

Big data analytics, the Internet of Things, cloud computing, and supplier/customer collaboration are linkage factors. They may have an impact on other system components for achieving sustainability in the supply chain. These factors holds both strong driving and dependent power. Dependent factors include green design and green purchasing, which have high dependence power. Thus, it can be interpreted from the analysis that industry technologies are essential to implementing GSC practices in the automotive supply chain.

5 Conclusion

There is a greater need to digitalize the automobile industry, which involves integrating manufacturers with extended and broad supply chain. The research aims to achieve the linkage between I4.0 technologies and green supply chain practices to achieve sustainability in the automotive supply chain. The study found that I4.0 technologies have a direct or indirect linkage with green supply chain practices.

The study uses a two-stage process wherein, in the first stage, prominent I4.0 technologies and green supply chain practices are selected using a fuzzy-set analytical hierarchy process. The expert opinions are validated using the AHP method consistency ratio, which is obtained within the range of 0 to 10%. The uncertainty in the expert's value preferences is further eliminated using a hesitant fuzzy AHP method to find the final set of I4.0 technologies and GSC practices for further cause-and-effect study. Thereafter, in stage 2, through interpretive structural modelling, the interrelationship between I4.0 technologies and GSC practices is studied for cause-and-effect analysis. The three-level hierarchical structure depicts that government regulation and policies, and top management commitment are two essential factors for driving sustainability in the automotive supply chain in India. Further MICMAC analysis supports the driving and dependence power matrix with no autonomous factors in selected variables. Also, it states that major I4.0 technologies such as the Internet of Things, big data analytics, cloud computing, etc. are linkage factors that could have an impact on other factors of the system to achieve supply chain sustainability. Thus, it can be interpreted from the analysis that in the automotive supply chain, I4.0 technologies are essential to strengthening GSC practices.

Further investigation can explore the specific challenges and barriers faced in implementing I4.0 technologies and green supply chain practices in the automotive sector. The study focuses on the Indian automotive supply chain, to examine the applicability of the findings in different geographical contexts. The paper highlights the importance of government regulations and top management commitment, but future research can

delve deeper into the specific policies and strategies that can drive sustainability in the supply chain.

References

1. Tseng, M. L., Tan, R. R., Chiu, A. S. F., Chien, C. F., & Kuo, T. C. (2018). Circular economy meets industry 4.0: Can big data drive industrial symbiosis? Resources, Conservation and Recycling, 131, 146–147. https://doi.org/10.1016/j.resconrec.2017.12.028
2. Frazier, W. E. (2014). Metal additive manufacturing: A review. Journal of Materials Engineering and Performance, 23(6), 1917–1928. https://doi.org/10.1007/s11665-014-0958-z
3. Chang, C.-Y., Pan, W., & Howard, R. (2017). Impact of Building Information Modeling Implementation on the Acceptance of Integrated Delivery Systems: Structural Equation Modeling Analysis. Journal of Construction Engineering and Management, 143(8), 4017044. https://doi.org/10.1061/(asce)co.1943-7862.0001335
4. Holmström, J., Liotta, G., & Chaudhuri, A. (2017). Sustainability outcomes through direct digital manufacturing-based operational practices: A design theory approach. Journal of Cleaner Production, 167, 951–961. https://doi.org/10.1016/j.jclepro.2017.03.092
5. Mastos, T. D., Nizamis, A., Vafeiadis, T., Alexopoulos, N., Ntinas, C., Gkortzis, D., Tzovaras, D. (2020). Industry 4.0 sustainable supply chains: An application of an IoT enabled scrap metal management solution. Journal of Cleaner Production, 269, 122, 377. https://doi.org/10.1016/j.jclepro.2020.122377
6. Rebelo, R. M. L., Pereira, S. C. F., & Queiroz, M. M. (2022). The interplay between the Internet of things and supply chain management: challenges and opportunities based on a systematic literature review. Benchmarking, 29(2), 683–711. https://doi.org/10.1108/BIJ-02-2021-0085
7. De Vass, T., Shee, H., & Miah, S. J. (2021). IoT in Supply Chain Management: Opportunities and Challenges for Businesses in Early Industry 4.0 Context. Operations and Supply Chain Management: An International Journal, 14(2), 148–161. https://doi.org/10.31387/oscm0450293
8. Huang, F., Chen, J., Sun, L., Zhang, Y., & Yao, S. (2020). Value-based contract for smart operation and maintenance service based on equitable entropy. International Journal of Production Research, 58(4), 1271–1284. https://doi.org/10.1080/00207543.2019.1617450
9. Agrawal, P., & Narain, R. (2023). Analysis of enablers for the digitalization of supply chain using an interpretive structural modelling approach. International Journal of Productivity and Performance Management, 72(2), 410–439. https://doi.org/10.1108/IJPPM-09-2020-0481
10. Esmaeilian, B., Sarkis, J., Lewis, K., & Behdad, S. (2020). Blockchain for the future of sustainable supply chain management in Industry 4.0. Resources, Conservation and Recycling, 163, 64, 105. https://doi.org/10.1016/j.resconrec.2020.105064
11. Jabbar, S., Lloyd, H., Hammoudeh, M., Adebisi, B., & Raza, U. (2021). Blockchain-enabled supply chain: analysis, challenges, and future directions. Multimedia Systems, 27(4), 787–806. https://doi.org/10.1007/s00530-020-00687-0
12. Kache, F., & Seuring, S. (2017). Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. International Journal of Operations and Production Management, 37(1), 10–36. https://doi.org/10.1108/IJOPM-02-2015-0078

13. Gravili, G., Benvenuto, M., Avram, A., & Viola, C. (2018). The influence of the Digital Divide on Big Data generation within supply chain management. International Journal of Logistics Management, 29(2), 592–628. https://doi.org/10.1108/IJLM-06-2017-0175
14. Min, H., & Galle, W. P. (2001). Green purchasing practices of US firms. International Journal of Operations and Production Management, 21(9), 1222–1238. https://doi.org/10.1108/EUM0000000005923.
15. Vachon, S., & Klassen, R. D. (2008). Environmental management and manufacturing performance: The role of collaboration in the supply chain. International Journal of Production Economics, 111(2), 299–315. https://doi.org/10.1016/j.ijpe.2006.11.030
16. Centobelli, P., Cerchione, R., Chiaroni, D., Del Vecchio, P., & Urbinati, A. (2020). Designing business models in circular economy: A systematic literature review and research agenda. Business Strategy and the Environment, 29(4), 1734–1749. https://doi.org/10.1002/bse.2466

Licence to Publish
Proceedings Papers

SPRINGER NATURE

Licensee	Springer Nature Singapore Pte Ltd.	(the 'Licensee')
Title of the Proceedings Volume/Edited Book or Conference Name:	Advances in Data-driven Computing and Intelligent Systems: Selected Papers from ADCIS 2023, Volume 2	(the 'Volume')
Volume Editor(s) Name(s):	Swagatam Das, Snehanshu Saha, Carlos A. Coello Coello and Jagdish C. Bansal	
Proposed Title of the Contribution:	Achieving Sustainability in Supply Chain during Disruption times: Role of Industry 4.0	(the 'Contribution')
Series: The Contribution may be published in the following series	A Springer book series Lecture Notes in Networks and Systems	
Author(s) Full Name(s):	Namit Shrivastava, Manoj K Srivastava	(the 'Author')

When Author is more than one person the expression "Author" as used in this Agreement will apply collectively unless otherwise indicated.

Corresponding Author Name:	Namit Shrivastava	
Instructions for Authors	https://www.springer.com/gp/authors-editors/conference-proceedings/conference-proceedings-guidelines	(the 'Instructions for Authors')

1 Grant of Rights

a) For good and valuable consideration, the Author hereby grants to the Licensee the perpetual, exclusive, world-wide, assignable, sublicensable and unlimited right to: publish, reproduce, copy, distribute, communicate, display publicly, sell, rent and/or otherwise make available the contribution identified above, including any supplementary information and graphic elements therein (e.g. illustrations, charts, moving images) (the 'Contribution') in any language, in any versions or editions in any and all forms and/or media of expression (including without limitation in connection with any and all end-user devices), whether now known or developed in the future. Without limitation, the above grant includes: (i) the right to edit, alter, adapt, adjust and prepare derivative works; (ii) all advertising and marketing rights including without limitation in relation to social media; (iii) rights for any training, educational and/or instructional purposes; (iv) the right to add and/or remove links or combinations with other media/works; and (v) the right to create, use and/or license and/or sublicense content data or metadata of any kind in relation to the Contribution (including abstracts and summaries) without restriction. The above rights are granted in relation to the Contribution as a whole or any part and with or in relation to any other works.

b) Without limiting the rights granted above, Licensee is granted the rights to use the Contribution for the purposes of analysis, testing, and development of publishing- and research-related workflows, systems, products, projects, and services; to confidentially share the Contribution with select third parties to do the same; and to retain and store the Contribution and any associated correspondence/files/forms to maintain the historical record, and to facilitate research integrity investigations. The grant of rights set forth in

this clause (b) is irrevocable.

c) If the Licensee elects not to publish the Contribution for any reason, all publishing rights under this Agreement as set forth in clause 1a above will revert to the Author.

2 Copyright

Ownership of copyright in the Contribution will be vested in the name of the Author. When reproducing the Contribution or extracts from it, the Author will acknowledge and reference first publication in the Volume.

3 Use of Contribution Versions

a) For purposes of this Agreement: (i) references to the "Contribution" include all versions of the Contribution; (ii) "Submitted Manuscript" means the version of the Contribution as first submitted by the Author prior to peer review; (iii) "Accepted Manuscript" means the version of the Contribution accepted for publication, but prior to copy-editing and typesetting; and (iv) "Version of Record" means the version of the Contribution published by the Licensee, after copy-editing and typesetting. Rights to all versions of the Manuscript are granted on an exclusive basis, except for the Submitted Manuscript, to which rights are granted on a non-exclusive basis.

b) The Author may make the Submitted Manuscript available at any time and under any terms (including, but not limited to, under a CC BY licence), at the Author's discretion. Once the Contribution has been published, the Author will include an acknowledgement and provide a link to the Version of Record on the publisher's website: "This preprint has not undergone peer review (when applicable) or any post-submission improvements or corrections. The Version of Record of this contribution is published in [insert volume title], and is available online at https://doi.org/[insert DOI]".

c) The Licensee grants to the Author (i) the right to make the Accepted Manuscript available on their own personal, self-maintained website immediately on acceptance, (ii) the right to make the Accepted Manuscript available for public release on any of the following twelve (12) months after first publication (the "Embargo Period"): their employer's internal website; their institutional and/or funder repositories. Accepted Manuscripts may be deposited in such repositories immediately upon acceptance, provided they are not made publicly available until after the Embargo Period.
The rights granted to the Author with respect to the Accepted Manuscript are subject to the conditions that (i) the Accepted Manuscript is not enhanced or substantially reformatted by the Author or any third party, and (ii) the Author includes on the Accepted Manuscript an acknowledgement in the following form, together with a link to the published version on the publisher's website: "This version of the contribution has been accepted for publication, after peer review (when applicable) but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/[insert DOI]. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms".
Under no circumstances may an Accepted Manuscript be shared or distributed under a Creative Commons or other form of open access licence.
Any use of the Accepted Manuscript not expressly permitted under this subclause (c) is

subject to the Licensee's prior consent.

d) The Licensee grants to Author the following non-exclusive rights to the Version of Record, provided that, when reproducing the Version of Record or extracts from it, the Author acknowledges and references first publication in the Volume according to current citation standards. As a minimum, the acknowledgement must state: "First published in [Volume, page number, year] by Springer Nature".

i. to reuse graphic elements created by the Author and contained in the Contribution, in presentations and other works created by them;

ii. the Author and any academic institution where they work at the time may reproduce the Contribution for the purpose of course teaching (but not for inclusion in course pack material for onward sale by libraries and institutions);

iii. to reuse the Version of Record or any part in a thesis written by the same Author, and to make a copy of that thesis available in a repository of the Author(s)' awarding academic institution, or other repository required by the awarding academic institution. An acknowledgement should be included in the citation: "Reproduced with permission from Springer Nature";

iv. to reproduce, or to allow a third party to reproduce the Contribution, in whole or in part, in any other type of work (other than thesis) written by the Author for distribution by a publisher after an embargo period of 12 months; and

v. to publish an expanded version of their Contribution provided the expanded version (i) includes at least 30% new material (ii) includes an express statement specifying the incremental change in the expanded version (e.g., new results, better description of materials, etc.).

4 Warranties & Representations

Author warrants and represents that:

a)

i. the Author is the sole copyright owner or has been authorised by any additional copyright owner(s) to grant the rights defined in clause 1,

ii. the Contribution does not infringe any intellectual property rights (including without limitation copyright, database rights or trade mark rights) or other third party rights and no licence from or payments to a third party are required to publish the Contribution,

iii. the Contribution has not been previously published or licensed, nor has the Author committed to licensing any version of the Contribution under a licence inconsistent with the terms of this Agreement,

iv. if the Contribution contains materials from other sources (e.g. illustrations, tables, text quotations), Author has obtained written permissions to the extent necessary from the copyright holder(s), to license to the Licensee the same rights as set out in clause 1 but on a non-exclusive basis and without the right to use any graphic

elements on a stand-alone basis and has cited any such materials correctly;

b) all of the facts contained in the Contribution are according to the current body of research true and accurate;

c) nothing in the Contribution is obscene, defamatory, violates any right of privacy or publicity, infringes any other human, personal or other rights of any person or entity or is otherwise unlawful and that informed consent to publish has been obtained for any research participants;

d) nothing in the Contribution infringes any duty of confidentiality owed to any third party or violates any contract, express or implied, of the Author;

e) all institutional, governmental, and/or other approvals which may be required in connection with the research reflected in the Contribution have been obtained and continue in effect;

f) all statements and declarations made by the Author in connection with the Contribution are true and correct;

g) the signatory who has signed this Agreement has full right, power and authority to enter into this Agreement on behalf of all of the Authors; and

h) the Author complies in full with: i. all instructions and policies in the Instructions for Authors, ii. the Licensee's ethics rules (available at https://www.springernature.com/gp/authors/book-authors-code-of-conduct), as may be updated by the Licensee at any time in its sole discretion.

5 Cooperation

a) The Author will cooperate fully with the Licensee in relation to any legal action that might arise from the publication of the Contribution, and the Author will give the Licensee access at reasonable times to any relevant accounts, documents and records within the power or control of the Author. The Author agrees that any Licensee affiliate through which the Licensee exercises any rights or performs any obligations under this Agreement is intended to have the benefit of and will have the right to enforce the terms of this Agreement.

b) Author authorises the Licensee to take such steps as it considers necessary at its own expense in the Author's name(s) and on their behalf if the Licensee believes that a third party is infringing or is likely to infringe copyright in the Contribution including but not limited to initiating legal proceedings.

6 Author List

Changes of authorship, including, but not limited to, changes in the corresponding author or the sequence of authors, are not permitted after acceptance of a manuscript.

7 Post Publication Actions

The Author agrees that the Licensee may remove or retract the Contribution or publish a correction or other notice in relation to the Contribution if the Licensee determines that such

actions are appropriate from an editorial, research integrity, or legal perspective.

8 Controlling Terms

The terms of this Agreement will supersede any other terms that the Author or any third party may assert apply to any version of the Contribution.

9 Governing Law

This Agreement shall be governed by, and shall be construed in accordance with, the laws of the Republic of Singapore. The courts of Singapore, Singapore shall have the exclusive jurisdiction.

Signed for and on behalf of the Author	Print Name:	Date:
	**Namit Shrivastava**	**25 Oct 2023**

Address:	**8001, Samriddhi Building, MDI Campus, Management Development Institute, Mehrauli Road, Sukhrali, Gurgaon, Haryana, India, PIN 122001**
Email:	**f20201767@pilani.bits-pilani.ac.in, namit507@gmail.com**

STATISTICAL ANALYSIS AND FORECASTING OF SOLAR ENERGY

Prepared for:

Dr. Sumanta Pasari

Instructor In-charge, Applied Statistical Methods (MATH F432)

Department of Mathematics, BITS Pilani

Prepared by:

Namit Shrivastava	**2020A2PS1767P**
Aryan Raj	**2020A2PS1777P**
Yashvardhan Prasad	**2020A2PS1775P**
Vippul Jangira	**2019B4A20751P**
Chaitanya Balani	**2020ABPS1846P**
Prem Shankar	**2020A4PS0557P**
Mohit Kumar	**2020A1PS1721P**

Table of Contents

Introduction

We can employ cutting-edge technology, such as thermal power plants, solar power systems, artificial photosynthesis, etc., to harness solar energy, which is the radiant light and heat energy from the sun. It is a type of renewable energy that is regarded as the most trustworthy energy source at the moment because it is abundant, pollution-free, and renewable.

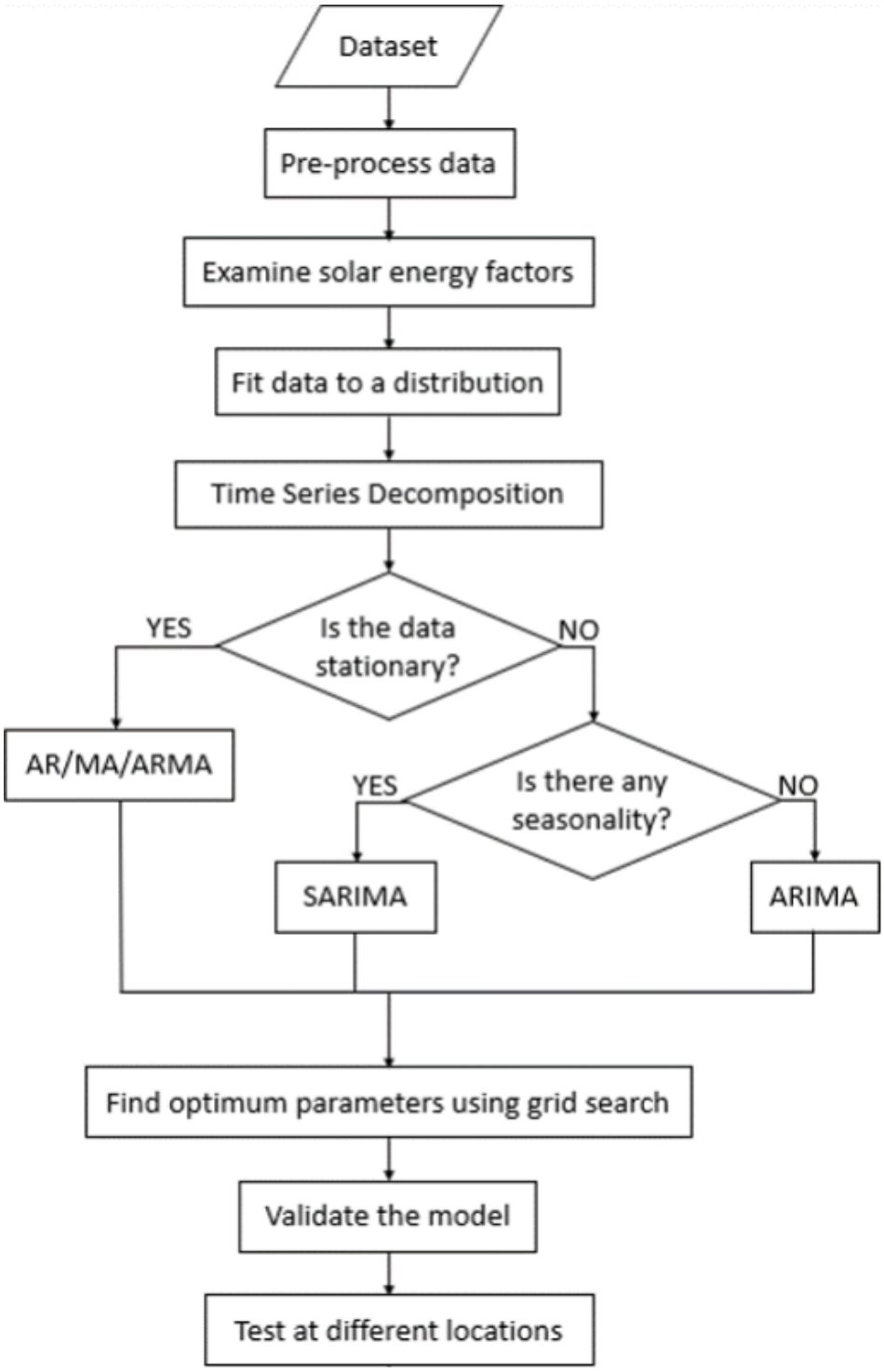
In response to environmental issues including pollution, the adoption of renewable energy sources is rising in popularity. India's National Time series value Change effort now includes solar power as a crucial element, and the National Solar Mission is one of the nation's most significant missions.

For the creation of tools that can harvest solar energy more effectively, forecasting and analysis of solar energy are consequently crucial. The main focus of this report is the intrastate examination of solar energy statistics. Datasets acquired from a range of solar farms in Rajasthan are analysed using a number of statistical techniques.

The report's approach to its methodology is broken down into the following components for the most part:

- Review of Solar Energy Parameters
- Methods for selecting an appropriate distribution fit for the GHI data
- Analysis and breakdown of time series and stationary data analysis
- Forecasting using a variety of different models, including AR, MA, ARMA, ARIMA, and SARIMA
- Conclusion

Methodology

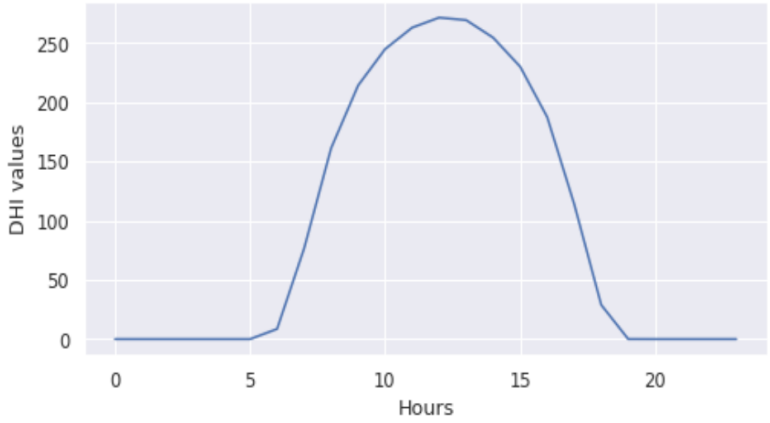


Pre-processing

The findings indicate that a number of variables influence how much solar radiation reaches the Earth's surface. To better understand the situation and to see how the various factors affect the solar radiation value as well as one another (via correlation), let's first define these words.

DHI (Diffuse Horizontal Irradiance)

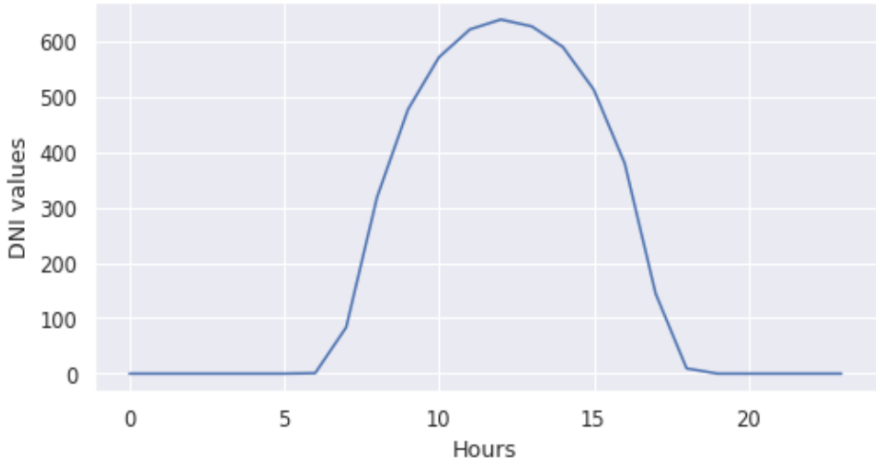
A surface's Diffuse Horizontal Irradiance (DHI) is the amount of radiation it receives per unit area that does not arrive via a straight channel from the sun but is instead scattered by particles and molecules in the atmosphere.
We have charted the DHI daily pattern (fig 2.1) and found that DHI is reaching its peak around



midday, or the 12th hour.

DNI (Direct Normal Irradiance)

The quantity of solar radiation that a surface receives per unit area when it is always held normal (or perpendicular) to the rays that come in a straight line from the direction of the sun at its current position in the sky is known as the Direct Normal Irradiance (DNI).
Daily DNI (Fig. 2.2) followed the same peak-at-12-noon pattern as DHI.

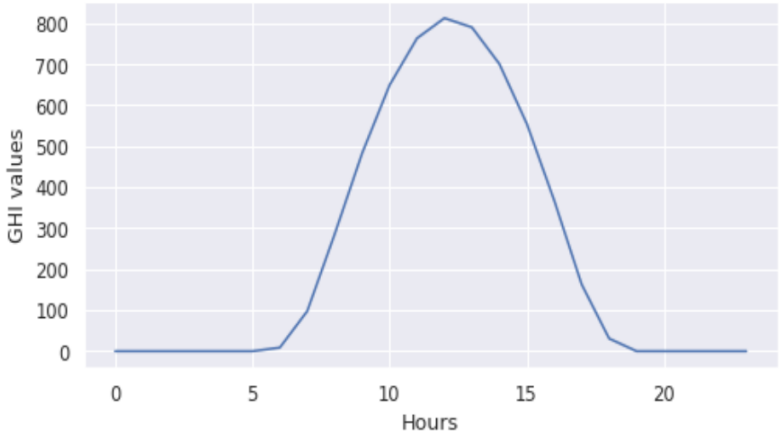


Global Horizontal Irradiance(GHI)

Shortwave radiation from the sun that reaches Earth's surface in a horizontal direction is measured as global horizontal irradiance (GHI). The number is a combination of the Direct Normal Irradiance (DNI) and the Diffuse Horizontal Irradiance (DHI) at the given Solar Zenith Angle(θ)

$$\text{Global Horizontal Irradiance (GHI)} = ((\text{DNI}) \times \cos(\theta)) + (\text{DHI})$$

While DHI and DNI follow roughly the same pattern throughout the day, peaking about noon, GHI's daily trend is slightly less steep.



Dew point

The Dew Point is the temperature at which water droplets begin to condense and dew begins to form in the atmosphere.

Relative humidity

It is the amount of water vapor in the air as a percentage of the amount required for saturation at the same temperature.

Solar zenith angle

It is the angle formed by the vertical and sun’s ray.

Snow depth

It is the vertical height of snow at the ground at the normal observation time.

Wind speed

It is the rate at which wind moves at the observation point

Conclusion

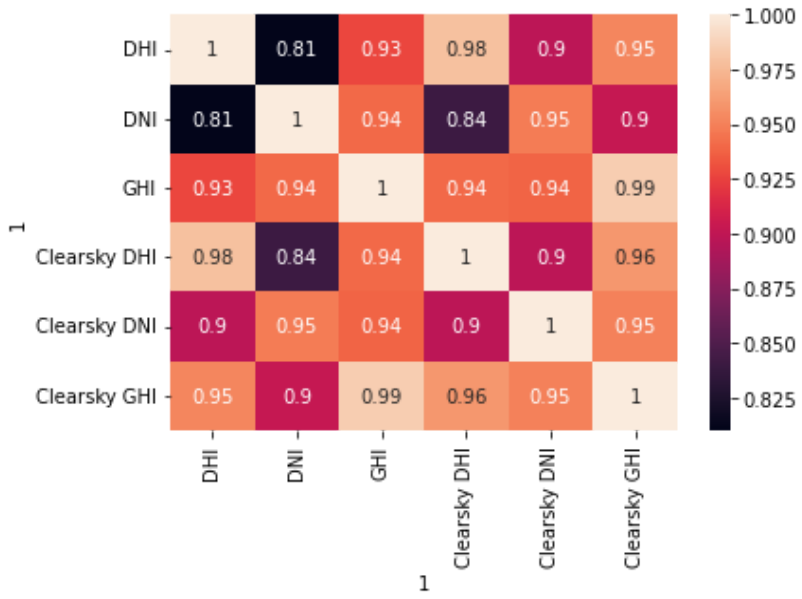
DHI, DNI, GHI, temperature, and relative humidity are relevant to the measuring of solar energy. Dew point is dependent on temperature and relative humidity, making it significant to solar energy as well.

Correlation matrix

	DHI	DNI	GHI	Clearsky DHI	Clearsky DNI	Clearsky GHI
DHI	1.000000	0.810126	0.927168	0.978921	0.897983	0.952093
DNI	0.810126	1.000000	0.940300	0.840339	0.947746	0.902440
GHI	0.927168	0.940300	1.000000	0.940370	0.938249	0.985022
Clearsky DHI	0.978921	0.840339	0.940370	1.000000	0.898374	0.959677
Clearsky DNI	0.897983	0.947746	0.938249	0.898374	1.000000	0.949993
Clearsky GHI	0.952093	0.902440	0.985022	0.959677	0.949993	1.000000

Rajasthan1 - Correlation Matrix

Making a heatmap (Fig. 2.1) of the correlations between all of the variables is the first step in making sense of our data. However, as we have only utilised the Rajasthan 1 dataset from 2000 to 2014 in this paper, it is probable that the data for other solar parks would also exhibit similar pattern for the different parameters.



Rajasthan1 - Correlation Heatmap

Using the heatmap, we can observe that the GHI, DHI, and DNI for Clearsky, respectively, have strong positive relationships. Every feasible pair of these factor-based pairs has a correlation of at least 0.80. We can confidently select one to analyse given the close ties between the aforementioned parameters. Since it considers both the diffuse and direct irradiance from the sun, as well as the solar zenith angle, the Global Horizontal Irradiance (GHI) is utilised.

GHI Data Analysis

We used R software to determine which model was most appropriate for the data, and to check for evidence of a specific distribution pattern in the datasets in question. Specifically, we get the following findings:

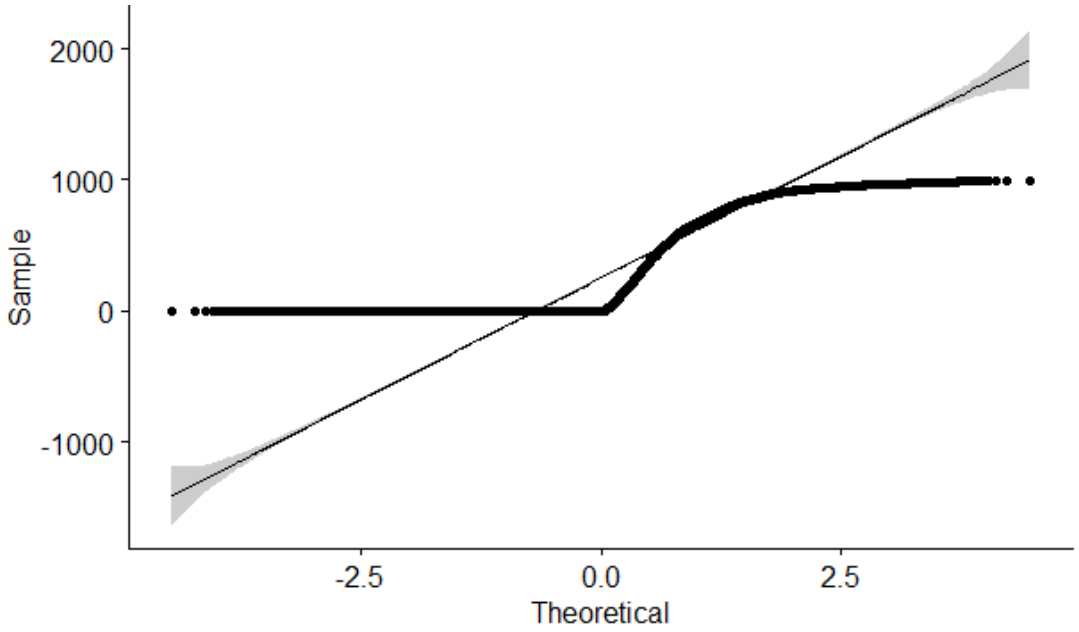


Fig1:- QQ Plots

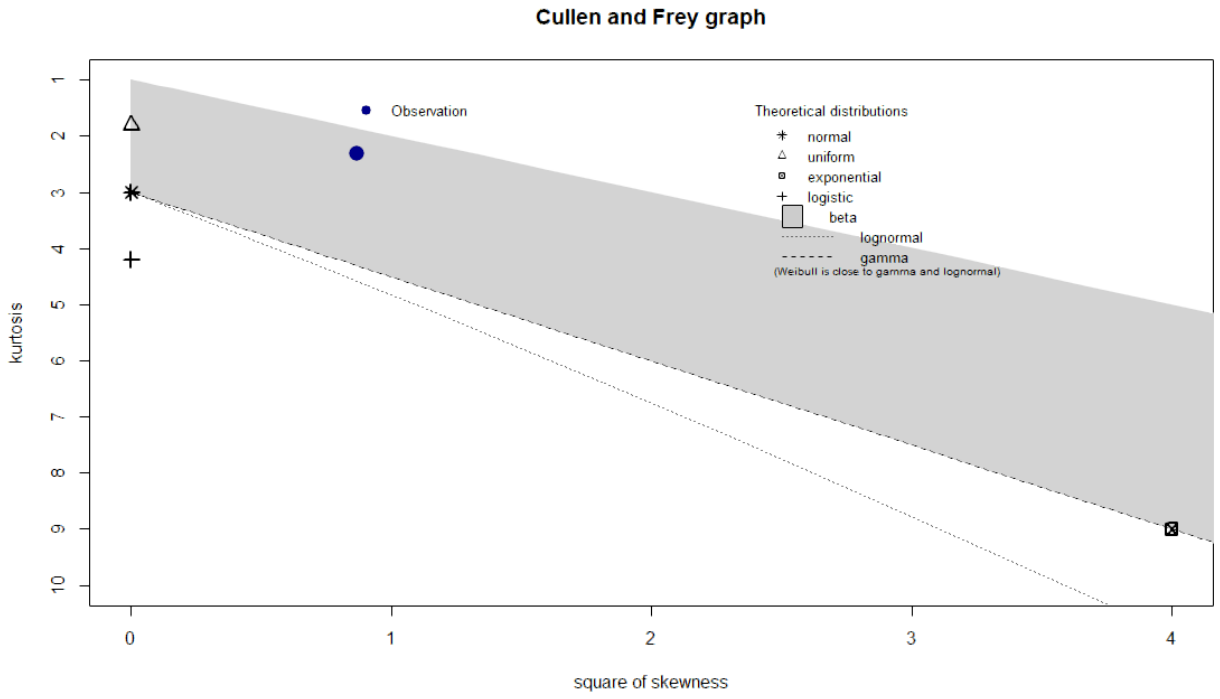
The same result can also be obtained by carrying out (Kolmogorov–Smirnov) or KS Test. It is a non-parametric test of the equality of continuous one-dimensional probability distributions. Its test is defined as:

H0: The data follows a specified distribution(Null Hypothesis)
H1: The data doesn't follow a specified distribution(Alternate Hypothesis)

Alpha, representing the significance level, is compared to the KS statistic value. When the fit is strong, the KS statistic value should be close to 1 (Max = 1.0), while when the fit is poor, it should be closer to 0 (Min = 0.0). The resulting information is:

```
> ks.test(Combined$GHI, "pnorm")

        Asymptotic one-sample Kolmogorov-Smirnov test

data:  Combined$GHI
D = 0.5, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Here, p-value< α(level of significance), thus indicating that the given data does not follow normal distribution.
We tried fitting the GHI Data to several probability distributions. For this purpose, we have used the Cullen and Frey Graph which displays the relation between Kurtosis and square of skewness. The graph plotted is as follows:-




```
summary statistics
------
min:  0   max:  995
median:  0
mean:  237.8901
estimated sd:  315.3451
estimated skewness:  0.9314054
estimated kurtosis:  2.311441
```

From the above graph, it can be seen that beta distribution is the distribution which would yield the least sum of squares, although still the goodness of fit tests does not approve of it.

FORECASTING

Autoregressive (AR) Model

The Autoregressive Model, often known as the AR model, uses solely previous period data to forecast future ones. It is a linear model, where current period data are the result of multiplying the total of previous results by a certain number. We write it as AR(p), where "p" stands for the model's order and the number of lag values we wish to take into account.
For instance, if we take X as time-series variable, then an AR(1), also known as a simple autoregressive model, would look something like this:
$X_t = C + \phi_1 X_{t-1} + \epsilon_t$
X_{t-1} represents the value of X during the previous period.
The coefficient ϕ_1 is a numeric constant by which we multiply the lagged variable (X_{t-1}).
ϵ_t = Residual

In order to predict using AR Model, we first fit the AR model to the GHI values of the dataset using the ARIMA function in R. The results obtained are as follows:-

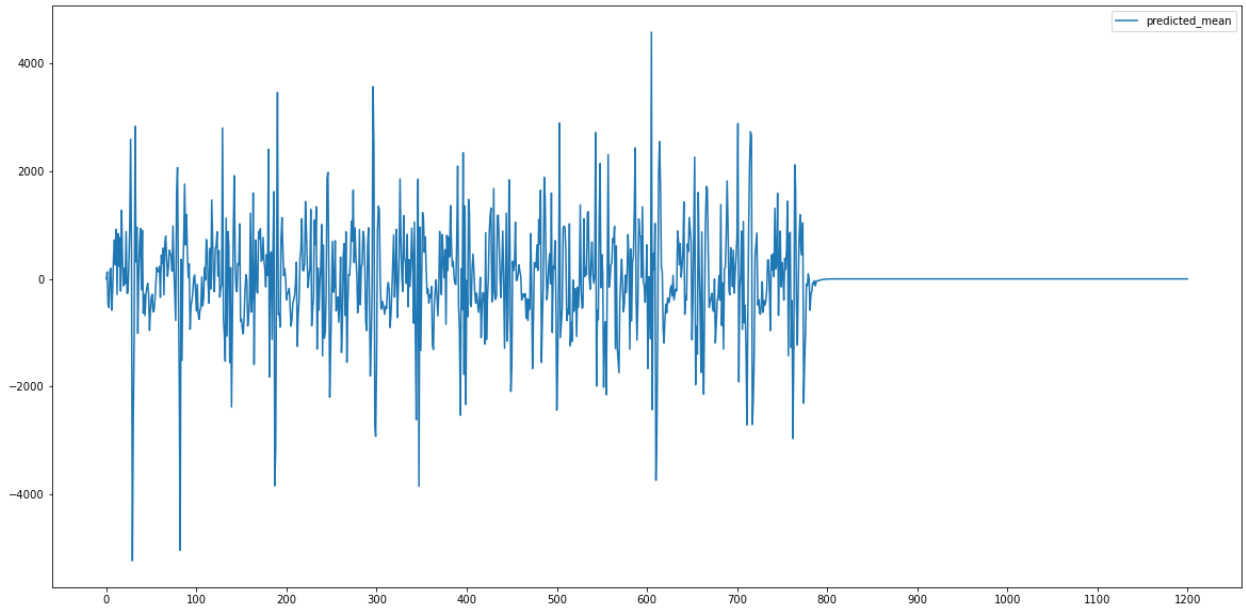
```
> AR<- arima(Combined$GHI, order=c(1,0,0))
> print(AR)

Call:
arima(x = Combined$GHI, order = c(1, 0, 0))

Coefficients:
          ar1  intercept
      0.9358   237.8901
s.e.  0.0010     4.6366

sigma^2 estimated as 12351:  log likelihood = -805441.2,  aic = 1610888
```

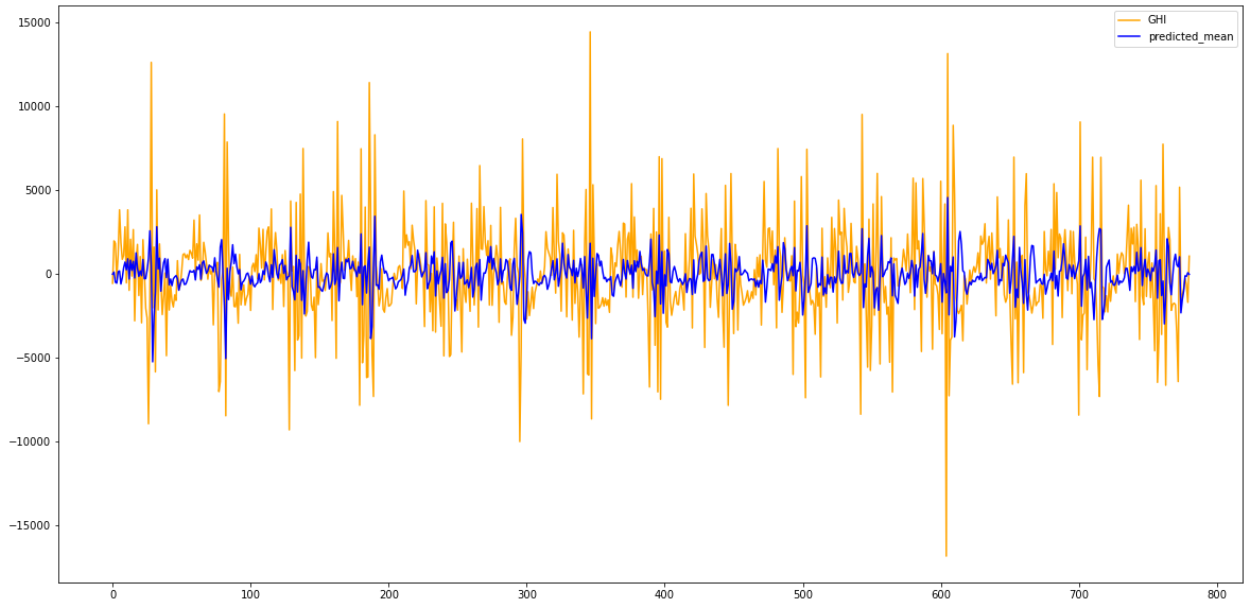
An estimated AR model may be used to predict using the predict() method. The graphs obtained using the AR prediction method is :-

The plot for forecasting using an auto regressive (AR) model is as follows:

Minimum AIC value was obtained to be 14706.814 for the model of order p=6.

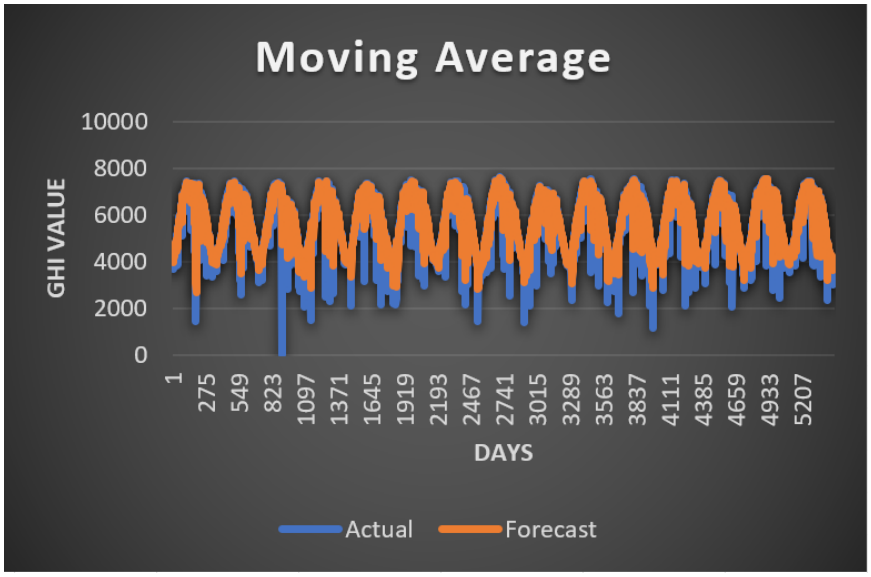
```
MAE = 2137.136
MAPE = 3.572
MSE = 8663585.29
```




GHI
predicted_mean
15000
10000
5000
0
−5000
−10000
−15000
0
100
200
300
400
500
600
700
800

Moving Average Model

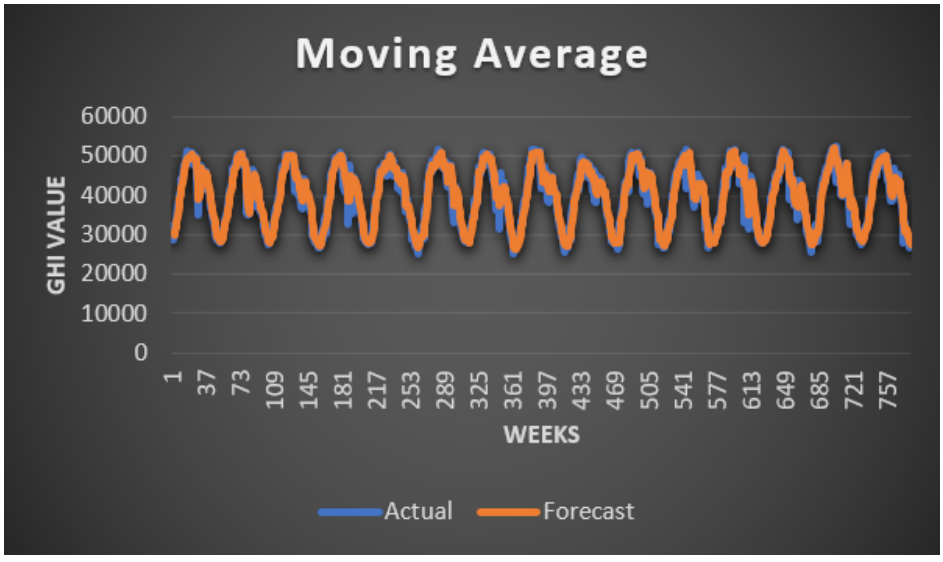
The Moving Average approach allows us to make predictions based on a predetermined window of historical data. By applying the Moving Average model with a window size of 3



, we are able to predict future GHI values with high accuracy(using the daily data)

MAE	MSE	MAPE
221.9051	176540.2	4.524782

For weekly data,



MAE	MSE	MAPE
1584.16	4355122	4.06918

The following is an expression for the MA(q) process:

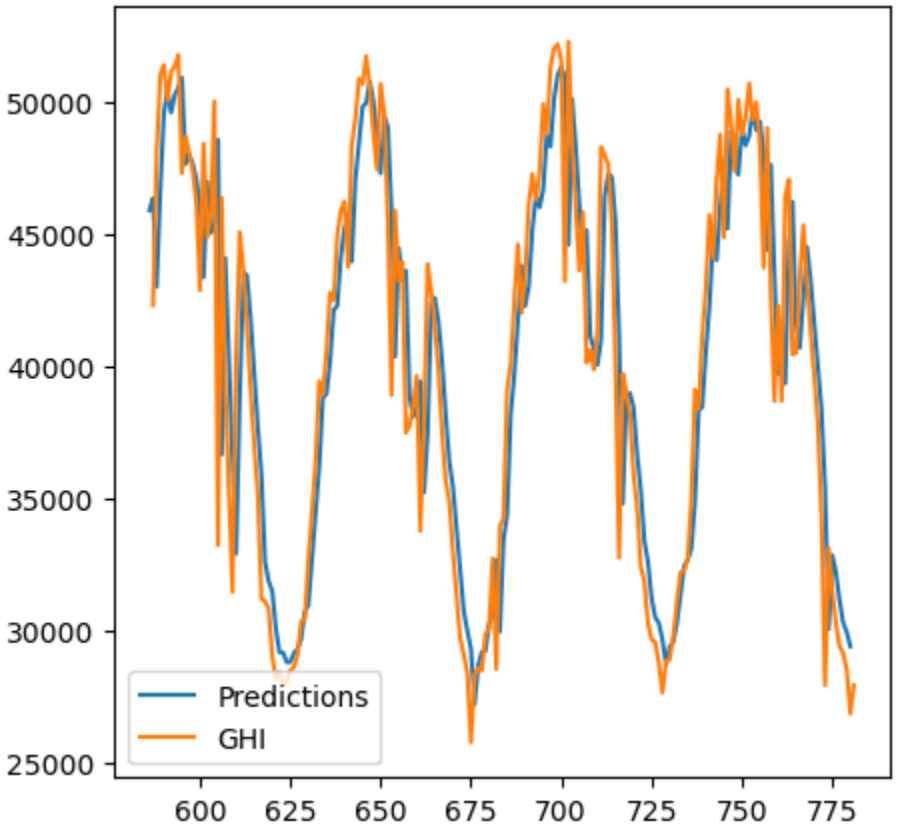
$x_t = (1 - \beta_1 B - \beta_2 B^2 - ... - \beta_p B^p)wt = \phi_q(B)w_t$

where B is a backshift operator and q is a polynomial of degree q. Here, each x_t number represents a weighted moving average of the prior q forecast errors.

Unfortunately, this results in extremely lengthy training times and a high computational cost. The choice of window of moving average also depends on the data. If there is high dependence of new values on previous values, the larger window can be considered for forecasting. As seen from the results of the model, the Moving average model gives quite high forecasting error and hence, is not best suited.Although, with a smaller window gives better results.Following results are obtained when we use a window of interval 3.

Auto Regressive Moving Average (ARMA) Model

ARMA model uses both moving average and auto regression for forecasting future values . The model is defined by two parameters p and q where p is the order of AR model and q is the order of MA model .Here we have taken both p and q to be 1



Where X-axis is the index of the GHI data and Y-axis is the GHI value observed weekly for years till 2014 The AIC value obtained by forecasting using ARMA was 14747.251.

.

If a time series has the form x_t;t = 0, ±1, ±2, . . then it has the form ARMA(p, q) if it is stationary and if it has the given parameters.

$$X_t = w_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j w_{t-j},$$

where $\phi_p \neq 0$, $\theta_q \neq 0$, and $\sigma_w^2 > 0$, $w_t \sim wn(0, \sigma_w^2)$.

The model may be recast in a more succinct manner with the assistance of the AR operator and the MA operator that we established before as

$$\phi(B)X_t = \theta(B)w_t$$

It's possible that you've noticed that the result of multiplying the same factor by both sides of the equation yields the same result.

$$\eta(B)\phi(B)X_t = \eta(B)\theta(B)w_t$$

If we didn't know about parameter redundancy, we may conclude that the data are associated when, in reality, they aren't at all.

Auto Regressive Integrated Moving Average (ARIMA) Model

The ARIMA model combines autoregression, moving average, and preprocessing "difference." ARMA model cannot work with non-stationary data. The Autoregressive Integrated Moving Average model overcomes this limitation by adding "differencing" to the ARMA model. Each ARIMA model employs three hyperparameters (p,d,q), the meanings of which are comparable to those of the ARMA model's p and q, while d represents the number of times the data must be differenced to produce a stationary output.

ARIMA has the following parameters:

- p: Trend autoregression order
- d: Trend difference order
- q: Trend moving average order

ARIMA model is also written as ARIMA(p,d,q)

P-value from Augmented Dickey Fuller (ADF) Test was obtained as 3.67e-16.
H0: Time series data is non-stationary.
H1: Time Series data is stationary.
As the p-value is very less than significance level 0.05, we can conclude that our data is stationary. Therefore, the ARIMA model will be similar to the ARMA model.

```
ad_fuller_result = adfuller(data['GHI'])
print(f'ADF Statistic: {ad_fuller_result[0]}')
print(f'p-value: {ad_fuller_result[1]}')
```

✓ 0.8s

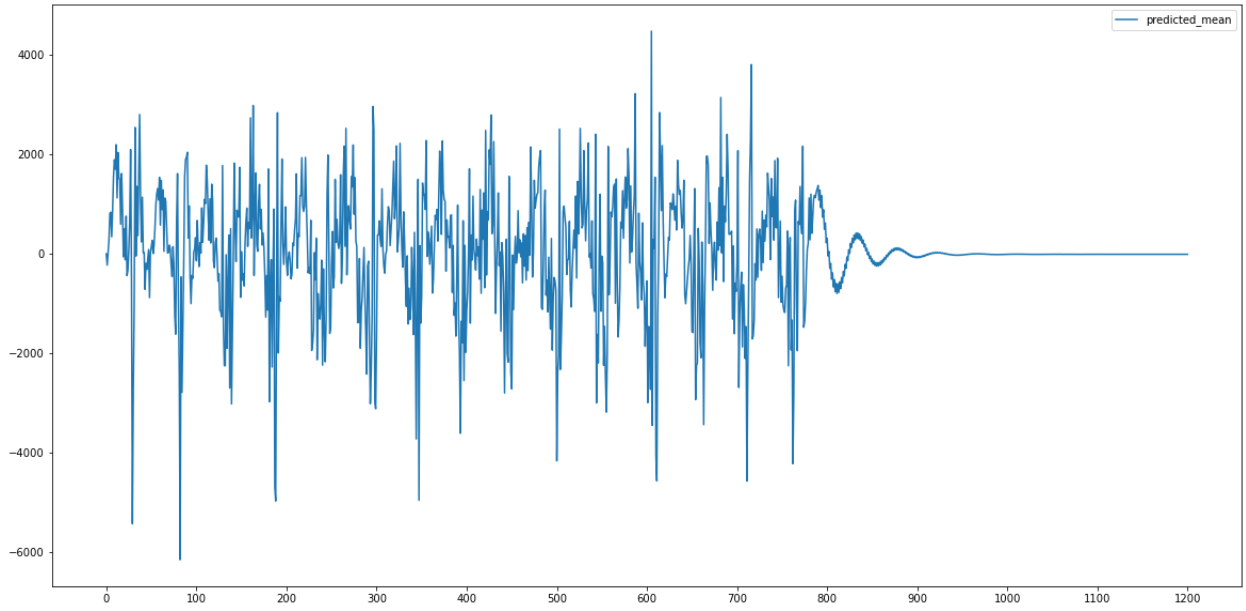
```
ADF Statistic: -9.489838688135503
p-value: 3.6779324671181 82e-16
```

Despite this, a differencing order of 1 was applied on the data prepared by combining the data for all years and converting to weekly form by summing GHI values for a week. We find that the combination (12,1,4) yields the minimum AIC value.

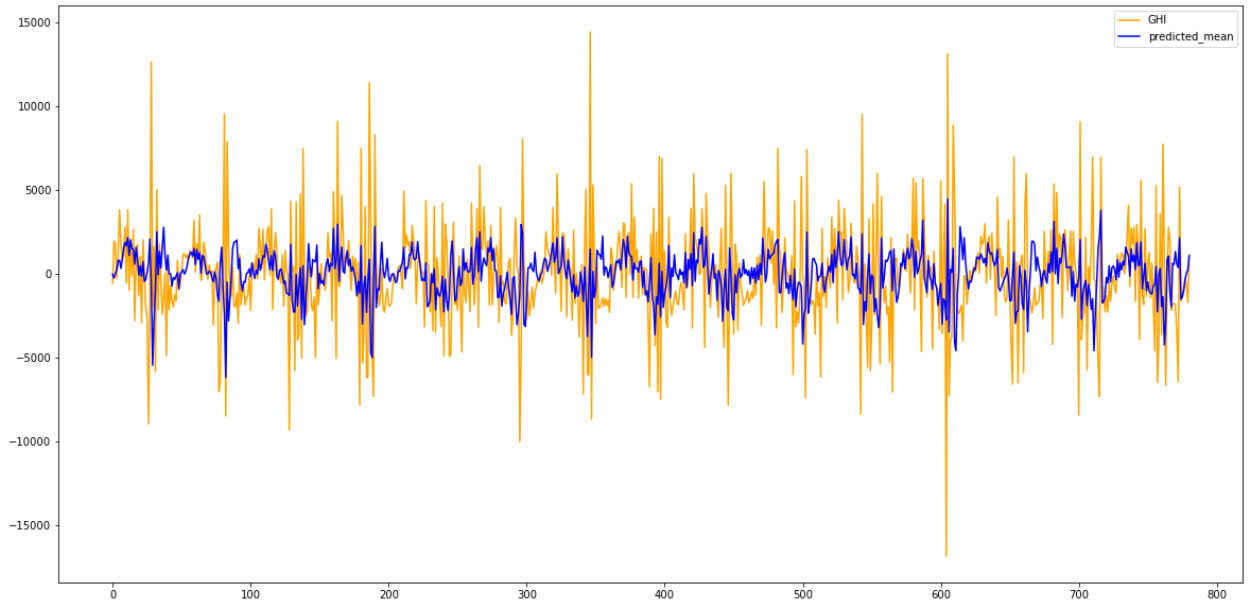
	(p, d, q)	AIC
0	(12, 1, 4)	14664.345554
1	(14, 1, 6)	14667.653330
2	(11, 1, 2)	14672.635652
3	(2, 1, 11)	14683.994771
4	(4, 1, 5)	14684.898595
...	...	...
220	(4, 1, 0)	14782.342682
221	(3, 1, 0)	14835.839571
222	(2, 1, 0)	14955.315523
223	(1, 1, 0)	15120.482980
224	(0, 1, 0)	15475.485213

225 rows × 2 columns

```
MAE = 2041.609
MAPE = 4.449
MSE = 8086807.355
```



The values on the X-axis represent the weeks. The values on the Y-axis represent GHI values predicted by the ARIMA model. The total values in the dataset were 781.



The mathematical form for d-order differencing is: $(1 - B)^d x_t$

for which B stands for the backshift operator. If white noise is produced by differencing by order d, then the series is integrated of order d. The integration by order d is the meaning of the extra I in ARIMA(p, d, q), so the function performs ARMA(p, q) on integrated data. A mathematical expression for this model is:

$$\theta_p(B)(1 - B)^d x_t = \phi_q(B) w_t$$

Seasonal Auto Regressive Integrated Moving Average (SARIMA) Model

- Seasonal Autoregressive Integrated Moving Average, SARIMA or Seasonal ARIMA , is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component.
- It adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality.
- A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA.
- The seasonal part of the model consists of terms that are very similar to the non-seasonal components of the model, but they involve backshifts of the seasonal period.

It adds 3 seasonal hyperparameters (P,D,Q) and a hyperparameter for seasonality m to the three hyperparameters of ARIMA. Thus each SARIMA model is characterized by the hyperparameters: (p,d,q)(P,D,Q,m) where p,d and q have meanings similar to the ARIMA model and the other hyperparameters are used as follows:

• P: seasonal autoregressive order

D: seasonal difference order

• Q: seasonal moving average order

• m: number of time steps in seasonal data

Assumptions for SARIMA -

- Seasonal variation needs to be constant.
- The time series under consideration must be marginally stationary or can be merged to generate a stationary series apart from the seasonal component.
- The error terms are thought to be independent, identically distributed variables that were randomly picked from a normal distribution with a mean of 0.

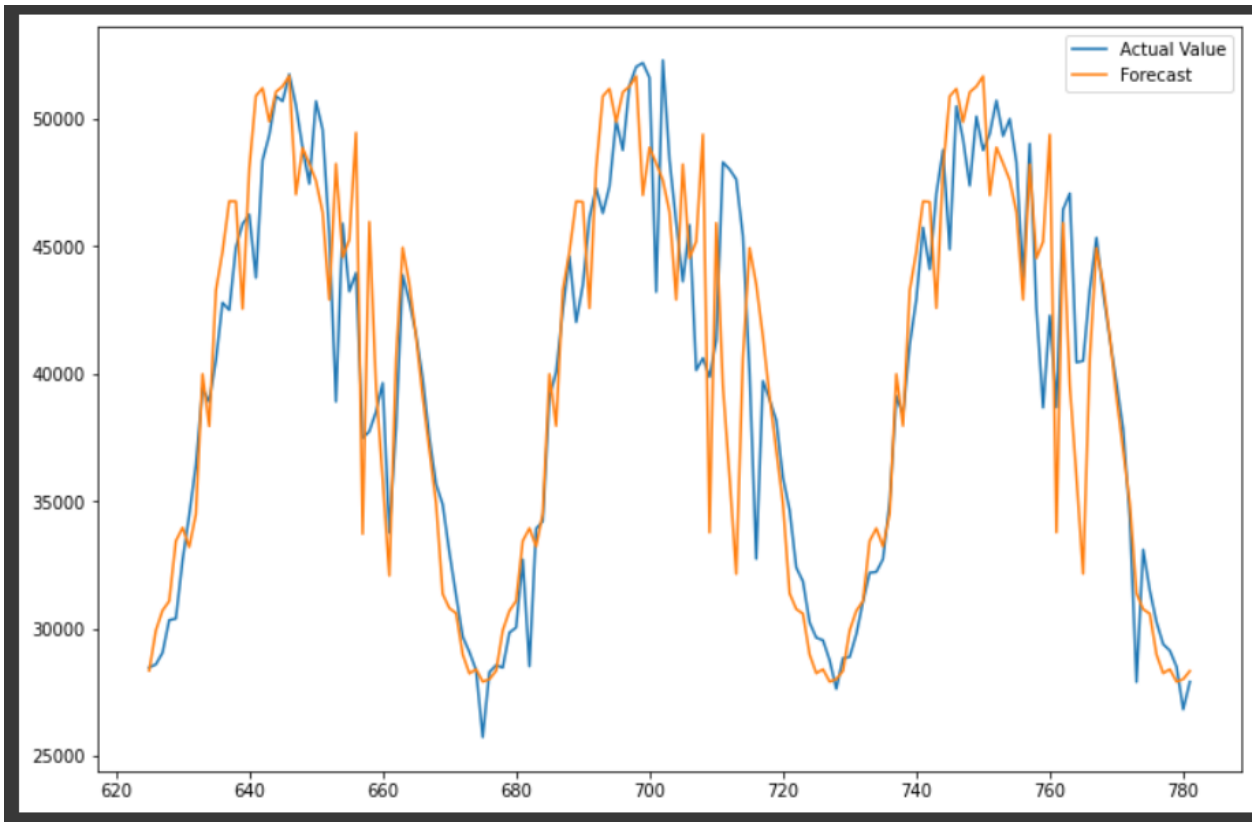


Fig: Rajasthan-1 SARIMA Model Weekly Forecasts

- Parameters - (1,0,1)(1,1,1,52)
- MAPE = 5.927 (weekly)
- MAE =2397.875 (weekly)

```
mape = mean_absolute_percentage_error(test , predictions)
mae = mean_absolute_error(test ,predictions)
print(mape)
print(mae)
```

```
0.059270533562983355
2397.875106934906
```

Comparing the MAPE values for both ARIMA and SARIMA models , we can say that ARIMA is preferred over SARIMA , which infers that there is negligible seasonality in our data.

Machine learning for Time Series Analysis

In recent years, as the availability of data and computer power has increased, Machine Learning has become an integral component of the new generation of Time Series Forecasting models, producing great results.
In traditional models such as AR, MA, etc., feature engineering is conducted manually and needs human parameter optimization. Machine Learning models require just characteristics and dynamics derived directly from the data. This allows them to expedite data preparation and learn complex data patterns more quickly and thoroughly.
In recent years, numerous new architectures have been created as diverse time series problems are investigated in a variety of domains. This has also been facilitated by the increasing availability of open-source frameworks, which has made the development of new bespoke network components easier and quicker.

Some of the well known methods in ML are:

- Recurrent Neural Networks (RNNs): They are the most classical and used architecture for Time Series Forecasting problems.
- Long Short-Term Memory (LSTM): They are an evolution of RNNs developed to overcome the vanishing gradient problem.
- Gated Recurrent Unit (GRU): They are another evolution of RNNs, like LSTM.
- Encoder-Decoder Model: This is a model for RNNs introduced to address the problems where input sequences differ in length from output sequences.
- Attention Mechanism: This is an evolution of the Encoder-Decoder Model, developed in order to avoid forgetting of the earlier parts of the sequence.

The decision is mostly based on whether the target variable has a strong correlation between its past, present, and future values. If strong correlations exist then time series maybe appropriate. Modeling and implementing time series analysis is simpler, but its empirical character is reliant on the assumption of a correlated goal variable. The disadvantage of time series is that prediction modeling does not account for root causes and influences.

This assignment involves univariate forecasting, and research indicates that classical time series models are more accurate for univariate, straightforward time series forecasting. In addition, classical time series methods are easily extensible, but machine learning models are more objectively focused. While machine learning may bring benefit in datasets with complex irregular data, missing observations, excessive noise, complex connection between various variates. But in our situation, for simple univariate forecasting we advise classical approaches because of better accuracy and lesser calculation cost.

Conclusion

- Using the Augmented Dicky Fuller test, we concluded that our data is stationary. So ARMA is best suited for forecasting.
- Although, ARIMA, SARIMA and ARMA work in similar manner for the stationary data, any methodology can be adopted for forecasting.
- We can also see that there is a high correlation between all the GHI,DHI and DNI factors and this is intra-state data for Rajasthan. So we can expect similar results from other solar parks of Rajasthan as well.
- Using the Kolmogorov Smirnoff test, we found that the most probable underlying probability distribution of the data is the beta distribution as it gives the minimum least square error among all the possible theoretical distributions.

AI Project

Automated Plagiarism Detector

Group 7

Namit Shrivastava	2020A2PS1767P
Mohit Gupta	2020A3PS1021P
Sahil Shah	2020A7PS0140P

Problem Statement

Electronic submissions have become increasingly common for student assignments, however this opens the door to more instances of plagiarism. It is relatively simple to copy data from a variety of sources and then paste it into a single work without giving any credit to the original authors of the material. This is made possible by the proliferation of information across the globe through the internet. It’s a loss on both fronts as students who engage in such behaviour will not learn as much as they could and authors who put hours of effort into their work are not given due credit. Consequently, there is a requirement for identifying instances of plagiarism in order to heighten and enhance the quality of a student's educational experience and have uniformity in evaluating different students by professors. Detecting instances of plagiarism manually is a challenging and time-consuming endeavour. Because of this, there is a pressing need to develop an automated system that can both detect instances of plagiarism and enhance the overall quality of the educational experience for the student.

PLAGIARISM PLUG-IN PROTOTYPE

Selecting the algorithm for document comparison and incorporating it into a new framework combining local and global search is necessary for developing the prototype for anti-plagiarism plug-in. The goal of this project is to create an anti-plagiarism module that may be used with a variety of assignment types, including but not limited to free-form essays, worksheet-based projects, and even audio podcasts. The following characteristics of the algorithm were chosen for incorporation into an anti-plagiarism add-on in light of this need.

- The programme needs to process various inputs that can be transformed into text.
- The algorithm's one-versus-all check performance needs to be robust even while processing massive data sets.

The Winnowing Algorithm developed by Schleimer, Wilkerson, and Aiken was chosen for use because it best meets these criteria. The approach relies on generating fingerprints (hashes) from the input texts, which are then used to compare the texts. In particular, Winnowing's benefits include:

- One, it takes any form of input and works with it. All that's required is some sort of plain text (like code, an essay, or a podcast script) to feed into the system.
- The second benefit is how quickly it processes massive data volumes. By adjusting the size of the grammar, a fast comparison can be made, and then a more in-depth one can be performed.
- Student work submitted for grading can have the "template fingerprint" removed without any effort.

Proposed System

As part of the proposed system, we plan to create software capable of identifying instances of plagiarism within academic assignments containing visual content. This will reduce instances of students plagiarising the work of their peers, boost the standard of education, and give each student an opportunity to develop his or her own abilities. The similarity between photos is used by the system to identify instances of plagiarism. The suggested approach employs reverse image search to identify plagiarised content.
The integration with Moodle CMS is a primary goal of this add-on (MOODLE is an acronym for "Modular Object-Oriented Dynamic Learning Environment,"). The Figure 1 loosely connected architecture was created to test how well Winnowing and global search function together.

We've built a prototype of the system's software on this architecture. The prototype evaluates student work in light of both a local document database and the results produced by a global search engine. Two main steps are required to complete a local search, while a third is recommended:

1. The system performs tokenization to strip the content of any meaningless characters (such as punctuation, whitespace, etc.)
2. The system computes the fingerprints of the provided document and saves the results to a database.
3. The fingerprint of the worksheet is being scrubbed from the submission if necessary.

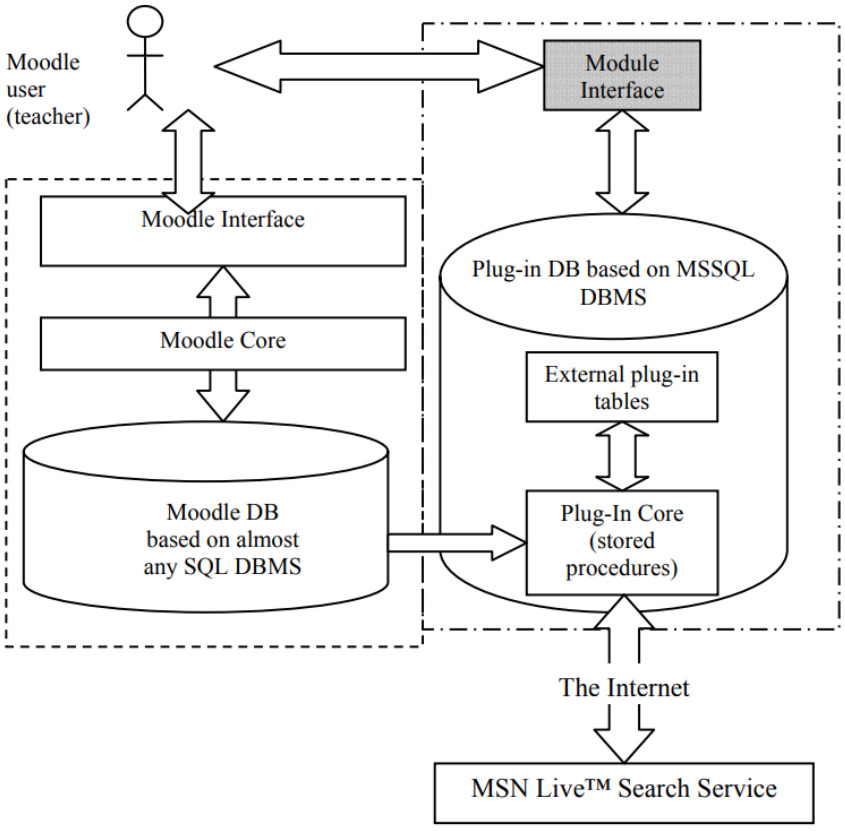


Figure 1. The simplified data flows for software prototype.

The global search requires an additional four steps to complete:

1. Conduct an online search for key terms (or even individual sentences) from the content. At this point, the system establishes a free connection to MSN Live Search (the service developed by Microsoft).
2. The system uses the search results to generate fingerprints of the most relevant documents.
3. Check the input against "local" and "global" document fingerprints.
4. Determine whether or not the work contains "local" or "global" plagiarised content. This choice determines how the supplementary materials are displayed to the educator (like fragments of original and copied paper). Figures 2 and 3 show screenshots containing this data for local and global searches, respectively.

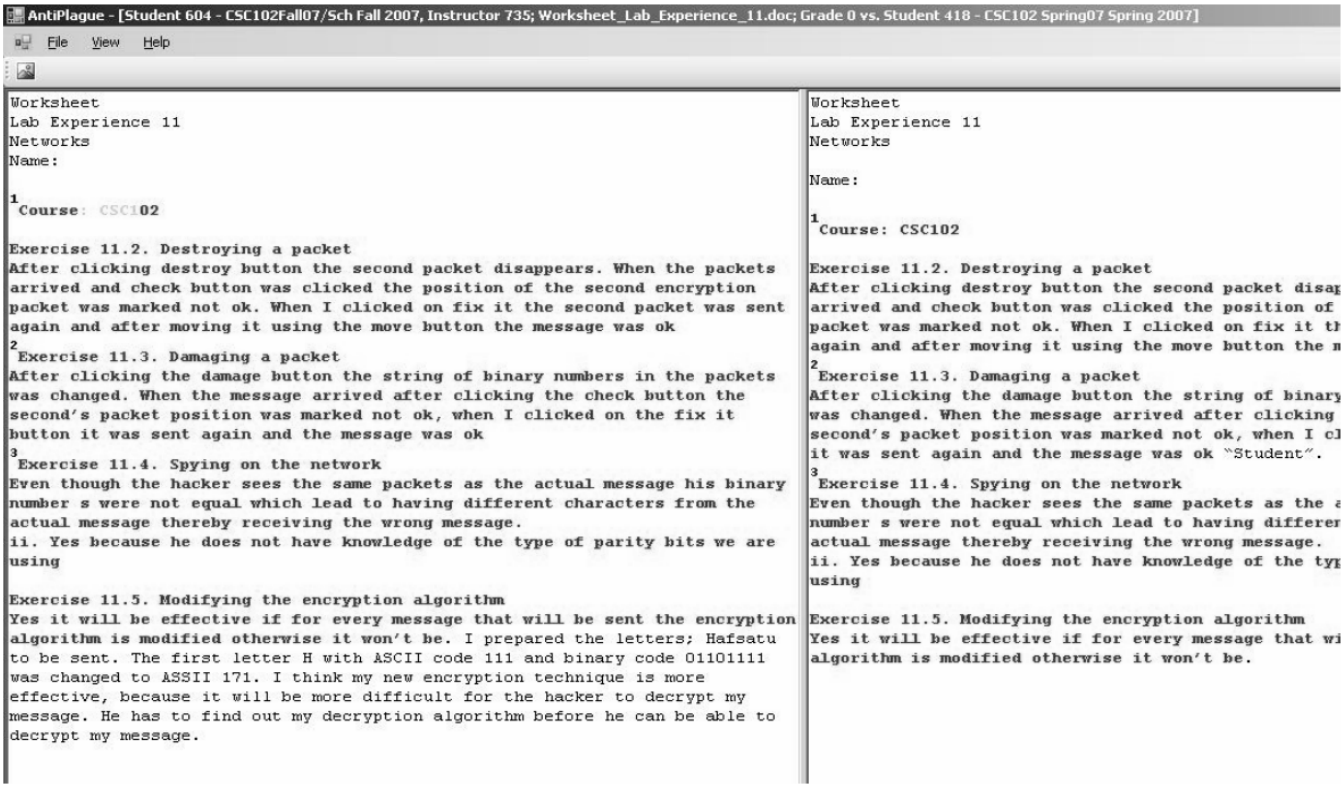


Figure 2. Screenshot comparing the submission with similar document from local database

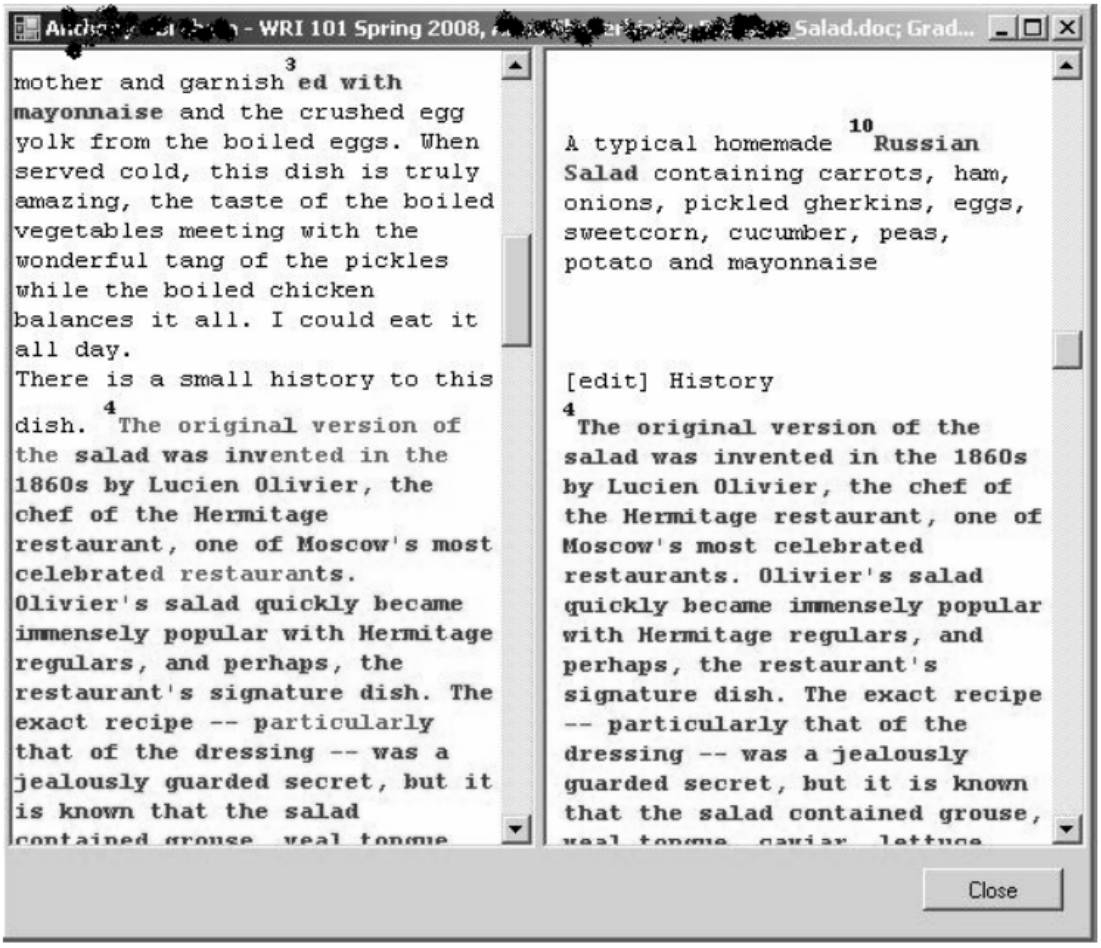


Figure 3. Screenshot comparing the submission with similar document from the Internet.

The prototype that was made doesn't meet all the requirements for Moodle plug-ins because it uses some Microsoft technologies. However, the same algorithms could be easily put into place using PHP, which is what Moodle needs.

The new open architecture for an anti-plagiarism plug-in was made after prototyping showed that it worked well. It can be seen in Figure 2. The main idea behind this architecture is that the anti-plagiarism plug-in should have a tokenization request broker added to it. With this kind of addition, the system will be able to handle almost any kind of soft submission. The broker will figure out what kind of file was sent and then call the right server to tokenize the submission.

As a result of the tokenization process, the system will return a set of symbols (words, sentences, etc.). After this point, it doesn't matter to the anti-plagiarism algorithm what

kind of assignment a student originally turned in, because the algorithm will now work with the plain tokenized text.

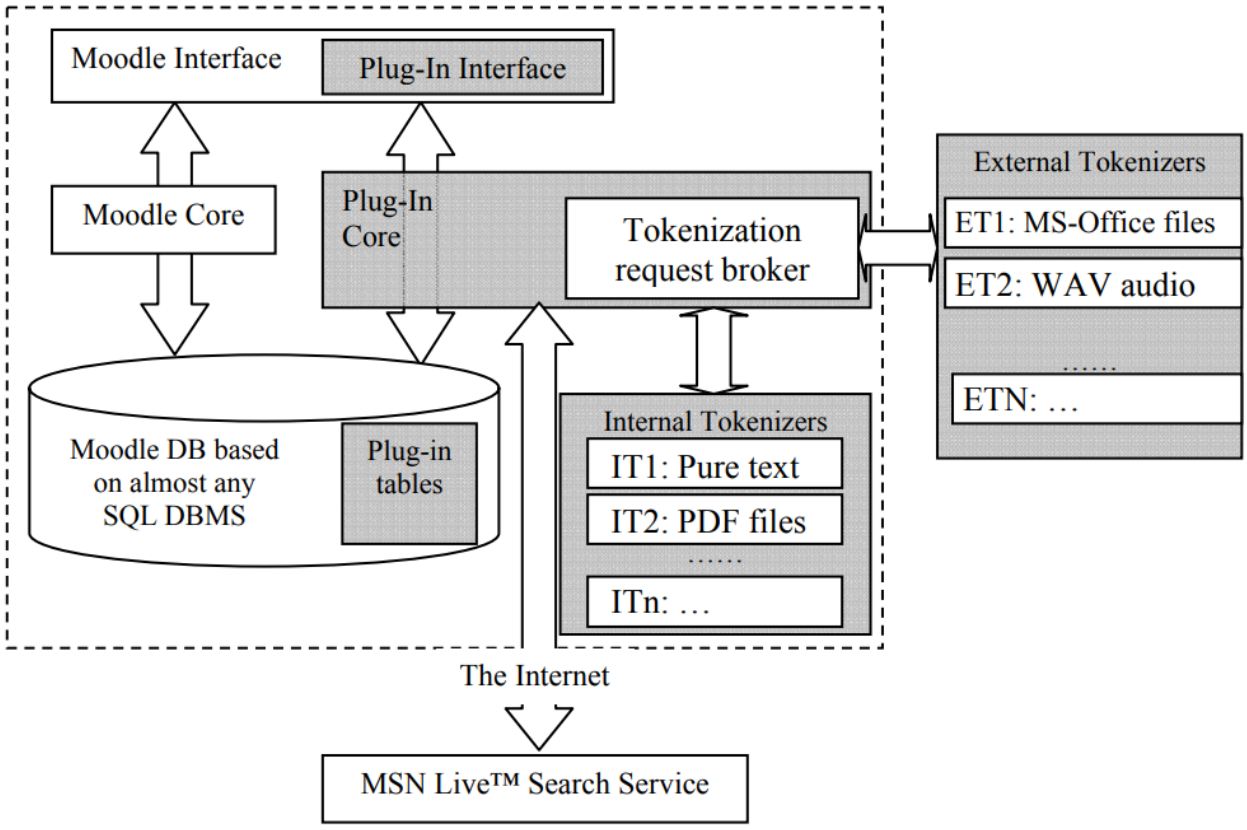


Figure 4. The preliminary architecture with tokenization request broker

Figure 2 shows that something like this will take the project to the next level. If the right tokenizer is connected to the system, the anti-plagiarism module will be able to be used to evaluate audio podcasts. Using speech recognition software, the audio information could be turned into text, which could then be used with the above algorithm.

Working Flow For Image Plagiarism

Collection of assignments

Each student's name and ID number will be gathered digitally along with any assignments or papers they may have submitted. In order to effectively detect instances

of plagiarism. A client-side and server-side validation UI for files (pdf, docs.) is developed. The submitted assignments are temporarily stored on local storage until they can be processed, at which point they are erased.

Pre-processing

In the first, crucial stage, called "pre-processing," all submitted assignments are transformed into the correct format. The format of all assignments has been standardised. After being extracted from the text, the photos are then stored in a database for further use. Only image files in the.jpeg,.png, or.jpg format are gathered; text, numbers, and other symbols are ignored.

Pseudocode for pre-processing –

```
1. OPEN THE FILE PREPROCESSED USING fitz.open
2. FOR EVERY PAGE IN OPENED PDF :
      a. EXTRACT THE LIST OF IMAGES USING getImageList()
      b. FOR EVERY IMAGE IN THE EXTRACTED LIST :
          i.   GET THE IMAGE BYTES USING extractImage()
         ii.   GET THE IMAGE EXTENSION AND LOAD IT TO PIL
        iii.   SAVE THE IMAGE TO LOCAL DISK
```

Hash value computation

Find the hash values of images that are already in the database and save those values. Do the same thing for any new images that are added. The new image is compared to every other image in the database, and it is put into a category based on the percentage of hash values that match.

A-Hash Value Computation

This algorithm works very quickly, but it is not sensitive to changes like scaling the original image, compressing and stretching it, or changing its brightness and contrast. It is based on the average value, so it is sensitive to operations that change this average value (like changing the levels or color balance).

The instructions below detail how to construct an A-Hash:

Reducing the size of the image: The original picture is shrunk down to a more manageable size (usually it is 8x8 or 16x16 points which will show a 64 or 256 bytes respectively).

Image grey-scaling: This change aids in decreasing the hash size by a factor of 3 by describing the number of components from three values of RGB to one level of grey.

Computing the average value: The next step is to determine the average value of all of the picture points.

Simplifying the image: In this system, each pixel reports a value of 0 if its value is below the average and a value of 1 if it is above the mean. As a result, we can think of the image as a collection of bits. The hash is calculated by reading each line sequentially and combining the values.

Pseudocode for hash value computation

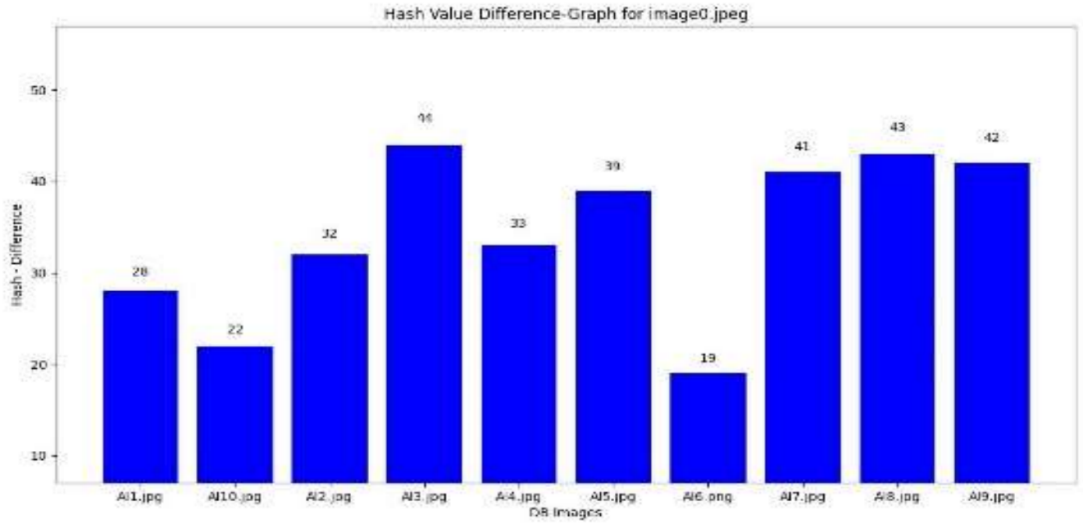
```
1. FUNCTION LOAD_IMAGES(FOLDER) - TO EXTRACT THE LIST OF IMAGES IN THE
   GIVEN FOLDER
2. ANSWER LIST = []
      a. FOR EVERY FILE IN THE FOLDER :
          i.   IF FILE HAS EXTENSION FROM (.JPEG , .JPG , .PNG)
                  1. ADD TO THE ANSWER LIST
      b. RETURN ANSWER LIST

1. EXTRACT THE IMAGES IN THE DATABASE AND USER INPUT USING LOAD_IMAGES
2. ANSWER LIST = []
3. FOR EVERY IMAGE IN USER INPUT :
      a. CALCULATE THE HASH VALUE USING imagehash.average_hash() AS HASH1
      b. FOR EVERY IMAGE IN DATABASE :
          i.   CALCULATE THE HASH VALUE USING imagehash.average_hash() AS
               HASH2
         ii.   IF HASH1 == HASH2
                  1. ADD THE IMAGE IN USER INPUT TO THE ANSWER LIST
```

Comparison and classification: After a hash is computed, it is compared to the hash values of all photos in the database. All photos are checked for similarities using their hash values, and those that match are marked as plagiarised.

Result: Both the pirated text and the plagiarised photos are stored in a Word document and an Excel sheet, respectively. A bar chart can also be used to depict the same set of data.

Excel Sheet: Differences in image hashes between uploaded photographs and those already in the database are recorded in an Excel spreadsheet for further use. A hash difference value for every image in the database is stored in its own Excel sheet.

Bar graph: The difference in hash values and their impact on database pictures is depicted in the following graph. Database image on the x-axis and hash value difference for image0 on the y-axis.



Distinction in hash values for picture 0 displayed as a bar chart

Resultant document: These are the results of a plagiarism scan done on the supplied assignment document, which revealed instances of copied photos. This refers to duplicate photos, which are visually similar to preexisting photographs in a database but have a different hash value. The system has now produced its final result.

Conclusion and Results: Accuracy Scores of Different Classifiers

The code we used to implement our model is here. Its most important snippets help us conclude our findings. In it's essence, our code accepts strings in small tokens and compares these small tokens with each other and then with a source file for similarity. The following results show how often the final verdict of the plagiarism checker, based on similarity between tokens, is correct.

The dataset we are using has taken 2 documents and made all possible pairs of sentences. It then makes a moving window of size 3 words and compares the sentences in each pair and based on a threshold, decides if they are plagiarism or not, assigning values 1 and 0 for the same in the column labeled '0'.

We used multiple standard accuracy testing functions such as SVM Classifier, Random Forest Classifier, Decision Classifier, Neural Network Classifier to test aspects of methodology in multiple ways. Its results are given below.

SVM Classifier

```
#SVM CLassifier

from sklearn.svm import LinearSVC

clf_SVM = LinearSVC(random_state=0, tol=1e-5, dual = False)
clf_SVM.fit(X_train, y_train)
```

```
LinearSVC(dual=False, random_state=0, tol=1e-05)
```

```
[79] y_pred = clf_SVM.predict(X_test)
```

```
[80] print('Accuracy: ', np.sum(y_pred == y_test)/X_test.shape[0])
```

```
Accuracy:  0.6805219079765178
```

Random Forest Classifier

```
[19] #Random Forest Classifier

     from sklearn.ensemble import RandomForestClassifier


     clf_Rf = RandomForestClassifier(max_depth=2, random_state=0)
     clf_Rf.fit(X_train, y_train)
```

```
RandomForestClassifier(max_depth=2, random_state=0)
```

```
[20] y_pred = clf_Rf.predict(X_test)
```

```
[21] print('Accuracy: ', np.sum(y_pred == y_test)/X_test.shape[0])
```

```
Accuracy:  0.61015134468116
```

Decision Trees Classifier

```
[38] #Decision Trees Classifier

     from sklearn import tree

     clf_DT = tree.DecisionTreeClassifier()
     clf_DT = clf_DT.fit(X_train, y_train)
```

```
y_pred = clf_DT.predict(X_test)
print('Accuracy: ', np.sum(y_pred == y_test)/X_test.shape[0])
```

```
Accuracy:  0.6339962980437702
```

Neural Network Based Classifier

```
[42] #NN Classifier

     from sklearn.neural_network import MLPClassifier

     clf_NN = MLPClassifier(solver='lbfgs', alpha=1e-5,hidden_layer_sizes=(5, 2), random_state=1)
     clf_NN = clf_NN.fit(X_train, y_train)
```

```
y_pred = clf_NN.predict(X_test)
print('Accuracy: ', np.sum(y_pred == y_test)/X_test.shape[0])
```

```
Accuracy:  0.5025587050411934
```

Nearest Neighbours

```
[44] #Nearest Neighbours Classifier

     from sklearn.neighbors import NearestCentroid

     clf_neighbour = NearestCentroid()
     clf_neighbour.fit(X_train, y_train)
```

```
NearestCentroid()
```

```
y_pred = clf_neighbour.predict(X_test)
print('Accuracy: ', np.sum(y_pred == y_test)/X_test.shape[0])
```

```
Accuracy:  0.5981744274670635
```

Stochastic Gradient Descent

```
[46] #Stochastic Gradient Descent Classifier

     from sklearn.linear_model import SGDClassifier

     clf_SG = SGDClassifier(loss="hinge", penalty="l2", max_iter=5)
     clf_SG.fit(X_train, y_train)
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/linear_model/_stochastic_g
  warnings.warn(
SGDClassifier(max_iter=5)
```

```
y_pred = clf_SG.predict(X_test)
print('Accuracy: ', np.sum(y_pred == y_test)/X_test.shape[0])
```

```
Accuracy:  0.6325808441911952
```

Naive Bayes Classifier

```
[49] # Naive Bayes Classifier - Took too long to run on our dataset

     # Naive Bayes Classifier

     from sklearn.naive_bayes import GaussianNB
     nb = GaussianNB()
     nb.fit(X_train, y_train)
```

```
GaussianNB()
```

```
y_pred = nb.predict(X_test)
print('Accuracy: ', np.sum(y_pred == y_test)/X_test.shape[0])
```

```
Accuracy:  0.609062534025333
```

From the above accuracy (and other trials with different parameter values), it is clear that for our model's training, the best classifier model will be **SVM Classifier.**

Despite having low overall accuracies, we can only assume that this would get better with time invested into this model and makes want to invest more time into this field.

References

https://www.degruyter.com/document/doi/10.1515/jisys-2014-0146/html?lang=en
https://dl.acm.org/doi/abs/10.1145/3345317
https://www.irjet.net/archives/V9/i3/IRJET-V9I3197.pdf
https://www.ijcaonline.org/research/volume125/number11/naik-2015-ijca-906113.pdf

CONVOLUTIONAL NEURAL NETWORKS

Prepared for:

Dr. Navneet Goyal

Instructor In-charge, Machine Learning (BITS F464)

Department of Computer Science & Information Systems, BITS Pilani

Prepared by:

Namit Shrivastava	**2020A2PS1767P**
Aryan Raj	**2020A2PS1777P**
Yashvardhan Prasad	**2020A2PS1775P**

Convolutional Neural Network

Neural Network

Inspired by the human brain, a neural network is a branch of machine learning designed such that it simulates the way biological neurons signal to each other.
They consist of node layers—an input layer, multiple hidden layers, and an output layer. Each node is linked to another node, and each connection has a weight and a threshold associated with it. If the output of a node is above the threshold value, that node is activated and data is sent to the next layer of the network. Else, no data is forwarded from that node.
There are several types of neural networks, each with their own uses and data types.

- Convolutional Neural Network (CNN): often used for computer vision, image processing and image classification tasks.
- Feed Forward Neural Network: used for pattern recognition and computer vision tasks. A Multi-Layer Perceptron model: used for complex classifications and machine translation tasks.
- Recurrent neural networks (RNNs): used for speech recognition, text to speech conversion and other natural language processing (NLP) tasks. Recently, RNNs have been developed to learn longer-term dependencies and are called Long Short-Term Memory Networks (LSTM)
- Modular Neural Networks: used for stock market predictions and compression of input data.

Structure of a CNN

Convolutional neural networks do better than other neural networks when they are fed images, speech, or audio signals. They comprise of three layers:

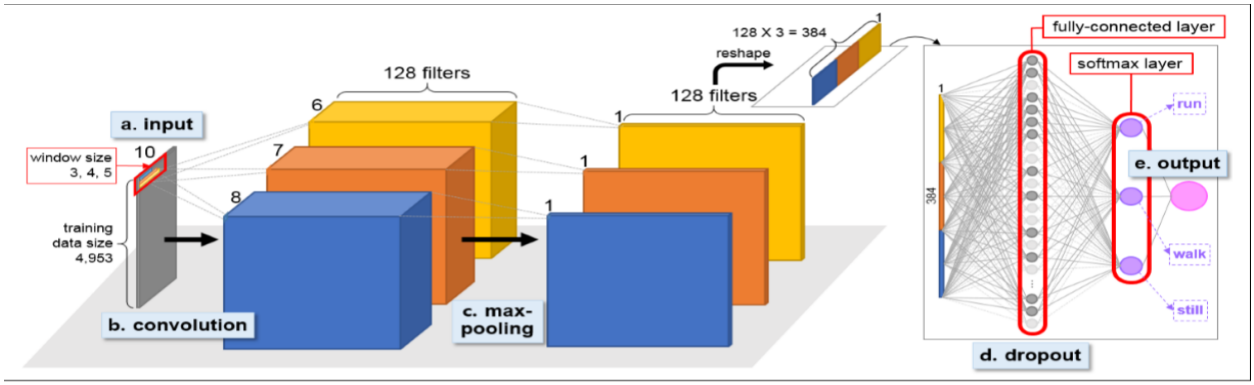
1. Convolutional layer
   Foundation of CNN. Major part of the processing takes place in this layer. A convolutional layer requires input data, a filter, and a feature map.

2. Pooling layer
   also called "downsampling," reduces the number of parameters in the input (dimensionality reduction).

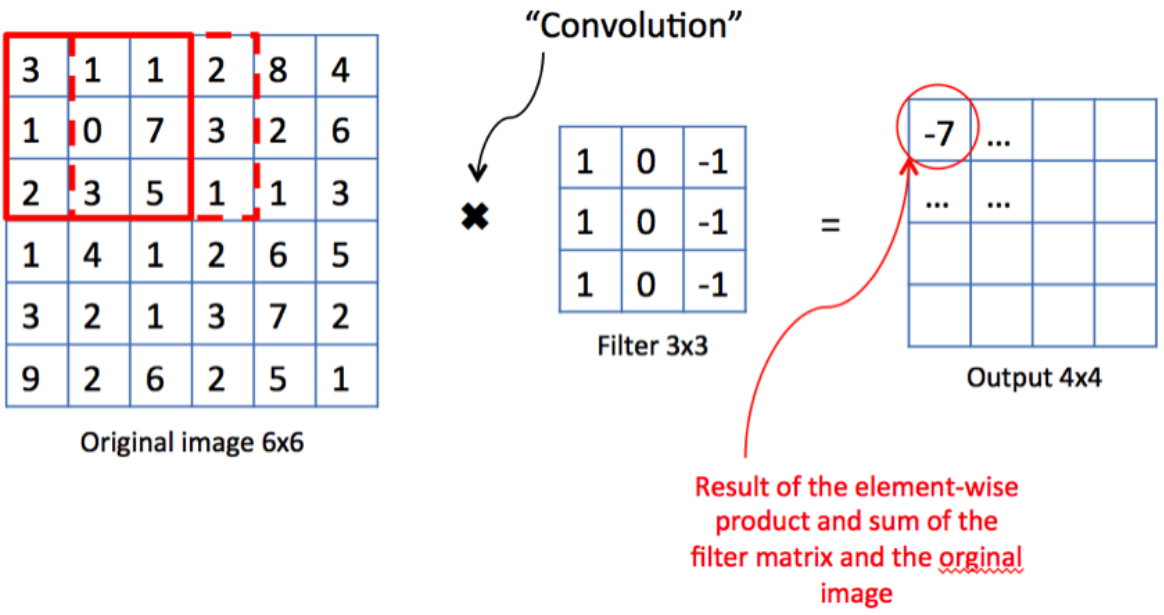
3. Fully-connected layer
   Also referred to as the FC layer. This layer is in charge of classifying the data based on the features that were taken from the previous layers and the filters that were used in them.

A neural network can have multiple convolutional layers, but the fully connected layer is the final layer. With each new layer, the CNN gets more complicated and can recognize more parts of the image. As the image data moves forward through the CNN's layers, it starts to identify larger parts of the objects in the image until it identifies the right one.

Working of CNN



An image can be considered to be a matrix of pixel values. A CNN is able to successfully capture the spatial and temporal characteristics of an image through the application of relevant filters. The architecture provides a better fit to the image dataset due to the reduction in the number of parameters involved and the reusability of weights.



A) CONVOLUTIONAL LAYER

Let the input be an image. Images are composed of pixels. Consequently, a 3D image can be viewed as a matrix of pixels. It will possess three dimensions—height, width, and depth. CNN is equipped with a feature detector, also known as a kernel or a filter, which moves across the image's fields to determine whether the feature is present (the convolution process).

The feature detector is made up of a 2D array of weights, each of which represents a different part of the image. Typically, the filter size is a 3x3 matrix, but the size can vary depending on the circumstance. In addition, this determines the size of the output field. The filter is then applied to a portion of the matrix, and a dot product between the input pixels and the filter is calculated. The resultant dot product is then fed to an output array. The filter is then moved by one stride, and this procedure is repeated until the filter has traversed the entire matrix or image. The final output of this process is known as a feature map. Each output value in the feature map is not required to correspond to each pixel in the input image. It must only connect to the filter being applied (also known as the "receptive field"). In convolutional and pooling layers the output array does not have to correspond directly to each input value. Due to this property, these layers are often called "partially connected" layers.

The feature detector does not change its weights as it traverses the image; this is known as parameter sharing. During training, certain parameters, such as the weight values, are modified by backpropagation and gradient descent. But the size of the output volume is affected by three hyperparameters that must be set before training the neural network.

1. The number of filters influences the output's depth. For instance, three distinct filters would generate three unique feature maps, creating a three-dimensional depth.

2. The number of pixels that the kernel passes over in the input matrix in each step is called a stride. Although stride values of two or greater are uncommon, a larger stride reduces output.

3. There are three types of padding:
    - Valid padding: Also known as no padding, in this case the last convolution is discarded if the dimensions do not align.
    - Same padding: This padding ensures that the output layer has the same dimensions as the input layer.
    - Full padding: This type of padding increases the size of the output by appending zeros to the input's border.

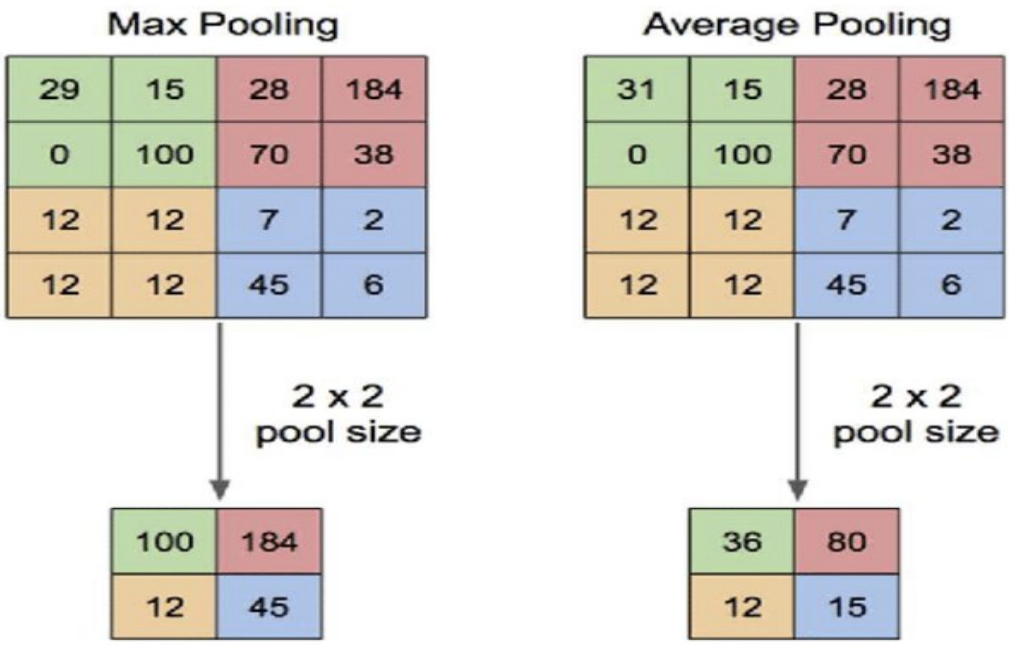
    When the input image exceeds the filter size, zero-padding is used . All of the elements that don't fit in the input matrix are set to zero, so the output is either bigger or the same as the input.

CNN introduces non-linearity into the model by applying a Rectified Linear Unit (ReLU) transformation to the feature map. The convolutional layer transforms the image into numbers, enabling the neural network to interpret and extract relevant patterns.

B) POOLING LAYER

Also referred to as "downsampling," it conducts dimensionality reduction by decreasing the number of input parameters. In the pooling operation a weightless filter is passed across the entire input, similar to convolution. The kernel takes the values in the receptive field and uses an aggregation function on them to fill the output array. There are two main types of pooling:

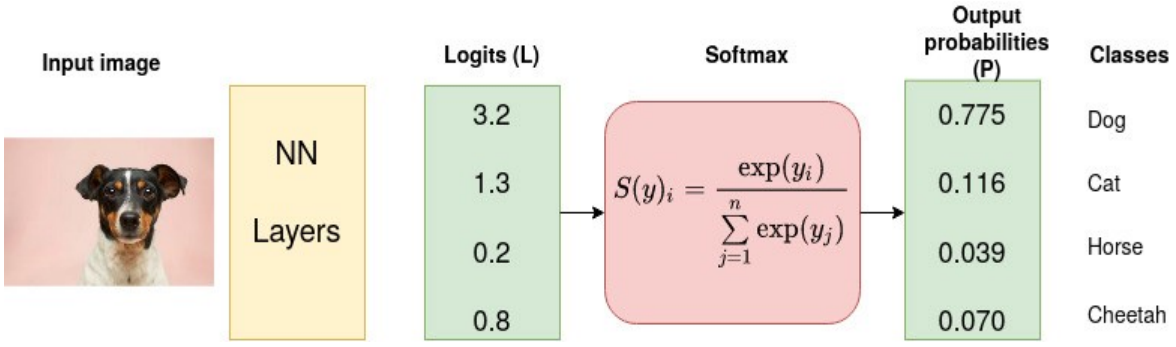
- **Max Pooling**: Pixel with the greatest value is selected from the receptive field and sent to the output array. This method is utilized more frequently than average pooling.
- **Average Pooling**: average value within the receptive field is calculated and sent to the output array.

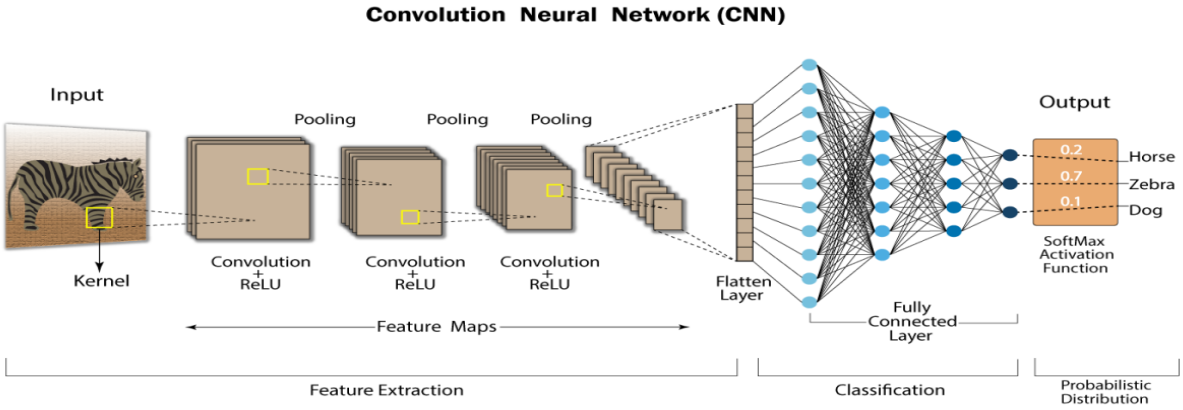


CNN receives a number of benefits from the pooling layer, despite the fact that a significant amount of information is lost. They help in reducing complexity, enhancing efficiency, and reducing the risk of overfitting.

C) FULLY-CONNECTED LAYER

Each node in the output layer is connected directly to a node in the previous layer. This layer is in charge of classifying the data based on the features that were taken from the previous layers and the filters that were used in them. The FC layer uses a softmax activation function for classification. A Softmax activation function scales the logits into probabilities. Logits are the unnormalized final scores of the model, representing the score for each possible output class. A Softmax activation function takes a vector of logits as input, applies the softmax function to each value, and returns the normalized scores (calculating a probability ranging from 0 to 1) for each possible output class. The class with the greatest probability is our final result.


Input image
NN
Layers
Logits (L)
3.2
1.3
0.2
0.8
Softmax
$S(y)_i = \frac{\exp(y_i)}{\sum_{j=1}^{n} \exp(y_j)}$
Output probabilities (P)
0.775
0.116
0.039
0.070
Classes
Dog
Cat
Horse
Cheetah


Convolution Neural Network (CNN)
Input
Kernel
Pooling
Pooling
Pooling
Convolution + ReLU
Convolution + ReLU
Convolution + ReLU
Feature Maps
Flatten Layer
Fully Connected Layer
Output
0.2
0.7
0.1
Horse
Zebra
Dog
SoftMax Activation Function
Feature Extraction
Classification
Probabilistic Distribution

References

1) Education, I. C. (2020, October 20). What are Convolutional Neural Networks? Retrieved November 28, 2022, from

   https://www.ibm.com/cloud/learn/convolutional-neural-networks

2) G. (2021, July 11). Basic Overviews on Convolutional Neural Networks. Retrieved November 28, 2022, from

   https://guymonahan.medium.com/basic-overviews-on-convolutional-neural-networks-f205180116be

Problem 2: Design of Experiments

Table of Contents

INTRODUCTION

Experiment design is the process of planning and carrying out an experiment to get a reliable, useful, and cost-effective answer to a question by measuring the results. If an experiment is well-designed, it will produce valid data, and if the data is analysed correctly, we can draw reliable statistical conclusions.

It is a systematic and efficient way for scientists and engineers to study the relationship between multiple input variables (also called "factors") and key output variables (also called "responses"). It is a structured approach for collecting data and making discoveries. It is a good way to plan experiments so that the data can be analysed in a way that leads to valid and objective conclusions. Well chosen experimental designs maximise the amount of "information" that can be obtained for a given amount of experimental effort.

A proper experimental design serves as a road map for the study methods, allowing readers to understand how the data was obtained more clearly and, as a result, assist them in properly analysing the results.

When to use Experimental Design?

1) To determine whether a factor, or a collection of factors, has an effect on the response.
2) To determine whether factors interact in their effect on the response.
3) To model the behaviour of the response as a function of the factors.
4) To optimise the response.

Principles of Experimental Design

Every experiment adheres to these three fundamental guidelines:

a) Randomisation

Randomization is the key to a good experiment. This means giving treatments to experimental units at random so that every possible treatment assignment has the same chance.

- Obtaining a demographic sample that is representative
- Dividing up the experimental units into the various treatments at random, eliminates systematic bias.
- Contributes to the distribution of the unobserved variance caused by variables influencing both the independent and dependent variables throughout the experiment. Consequently, the errors become random and independent, which renders the observations random as well.

b) Replication

The validity of the experiment can be checked by duplicating the experimental unit so that the same conditions can be repeated multiple times. This gives a more accurate estimate of the experimental error.

- Experimentation is repeated by subjecting the same group of organisms to the same treatment a predetermined number of times in order to acquire a more robust and trustworthy estimate than that obtained from a single observation.
- Increasing the number of observations boosts the experiment's precision effectively. Suppose, for instance, the variance of a random variable x is σ^2. The variance of the sample mean, $\bar{\mathbf{x}}$, based on n observations is therefore (σ^2/n). Consequently, the variance of $\bar{\mathbf{x}}$ diminishes as n grows.

c) Local Control

Local control entails the grouping of identical experimental units into groups or blocks with the elimination of variation within the blocks in an effort to minimise experimental error.

- Aids in a more precise and efficient experiment
- Handles the extraneous sources of variation that are a factor in experimental error but are beyond the control of randomization and replication.
- This means that the error component should ideally only include the variation that can be attributed to the treatments themselves, namely the variance between blocks.

Methods of Experimental Design

1) Completely Randomised Design

Treatments are assigned randomly to experimental units. This ensures that each experimental unit has an equal chance of receiving a treatment. Any variation between experimental units receiving the same treatment is considered an experimental error. Consequently, this design only applies to experiments with homogeneous experimental units, such as laboratory experiments, where it is assumed that there are no uncontrolled variables.

Advantages

1) Simple layout
2) Complete flexibility. Any number of treatments and replications for each treatment can be tried.
3) The design provides the greatest number of degrees of freedom for experimental error.

Disadvantages

1) Rarely suitable for experiments because it is extremely challenging to find identical experimental scenarios. This design does not use the Principle of Local Control.
2) Comparatively less accurate than other designs

Example

Let's say a company that makes car wax has three different formulas and wants to find the best one. In order to test the quality of these waxes (Wax 1, Wax 2 and Wax 3), each wax is applied to a sample of 10 cars of the same model (total 30 cars). Now, each of these cars is subjected to multiple tests such as repeated car washes, scratch test, dust test, etc.

In this scenario, randomization was used to assign a wax to a car. Each car had an equal chance of getting one of the three waxes. Replication was employed by using cars of the same model to maintain homogeneity. The results of the tests would be analyzed collectively for 10 cars based on the wax applied. This highlights the use of local control as analyzing collective results can help in some elimination of variation from the results.

2) Randomised Block Design

To eliminate the effects of a few of the most important extraneous / nuisance factors in this situation, the idea of 'Blocking' is applied. The fundamental idea is to construct homogeneous blocks such that the extraneous factors are constant in them while the factor of interest is allowed to vary.

Advantages

1) Effectively handles non-homogeneous experimental material.
2) It is adaptable enough to support countless treatments, blocks, and replications.
3) The sample sizes for the various treatments do not have to be equal.
4) Smaller error variance as the Local Control principle makes that sure because of the homogeneous blocks and because of parting away some variance from the error variance due to the difference among blocks. Thus, this dominates over the Complete Block Design which has high experimental error due to high variability among experimental units.
5) Relatively easy statistical analysis even with the missing data.
6) If an entire treatment or a block needs to be dropped from the analysis for some reason, such as spoiled results, the analysis is not complicated thereby.

Disadvantages

1) Not suitable for a large number of treatments because the block size becomes too large. Because the prima facie idea of Randomised Block Design is based on the fact of reducing the variability within blocks, but with the increase of block size, we deviate from our basic setup.
2) It requires some strong assumptions more than that for a completely randomised design - like no interactions between treatments and blocks and constant variance from block to block. So, interactions between block and treatment effects increase error.

3) Generalised Randomised Block Design

When the nature of the interactions between blocks and treatments is of interest, more than one replicate is required in each treatment within a block. This design is called a Generalised Randomised Block Design.

Each treatment only occurs once in each block in RBD, making it impossible to test for a treatment-by-block interaction. However, GRBD (Generalised Randomised Block Design) permits replications of each treatment level within a block. The two factors, treatment, and block are also interchangeable in GRBD.

Smaller groups or blocks of experimental units often provide better homogeneity when they represent physical things. Because of this, we do not advocate employing a block design with more experimental units per block than the minimal x, where x is the number of levels of the treatment factor. However, in situations where the experimental runs represent trials rather than actual physical objects, greater block sizes may not always result in an increase in the variability of experimental units within a block, allowing for the speedy creation of experimental runs.

<u>Advantages</u>

1) When compared to RCBDs(Randomised complete block design) with more blocks, GRBD model designs provide more degrees of freedom for investigating treatment-effects.
2) Replication of the treatment inside each block, which enables the estimate and testing of an interaction term within a linear model without assuming any mistake.
3) Instead of using an RCBD with additional blocks while doing an experiment, if the experimenter wants to boost power, they may use a GRBD instead.

<u>Disadvantages</u>

1) Because there are numerous units in each block and each treatment must be administered to a number of units in each block, this process is quite tedious.
2) Because replications are used, experiment costs are higher.

4) Optimal Design

An optimal design optimises a numerical criterion that is usually related to variability or other statistical design properties and uses as input the number of runs, their causes and possible levels, the structure of the block (if any), and the assumed form of the relationship between the answer and factors.

Advantages

1) lowers the cost of experiments by figuring out the needed statistical model with fewer trials than before.
2) It can handle many different kinds of factors, like processes, mixtures, and discrete factors.
3) It can take care of treatments, i.e., factor levels that are continuous rather than discrete.
4) The design can be made better when the design space is limited, such as when the mathematical process space has settings of factors that are not possible in real life.

Disadvantages

Complexity is high because deciding on a good model and its corresponding criterion function requires a good understanding of both statistical theory and designing experiments in the real world.

5) Bayesian Experimental Design

Bayesian experimental design is based on Bayesian inference to interpret the observations and data collected during the experiment, which uses Bayes' theorem to update the probability for a hypothesis as new data or information becomes available. This makes it possible to take into account both any prior knowledge about the parameters to be determined and observational uncertainty.

The aim is to design an experiment so that the expected utility of outcome is maximised. The utility criterion chosen will decide the optimal experiment to conduct.

In simple terms, the Bayesian approach says that we can't be sure about what we believe, so we try to figure out the probability distribution of our

belief/parameter. We try to estimate the probability distribution of the belief by taking into account what we already know and what we have seen. Simply put, we don't look at a single point estimate. Instead, we look at the probability distribution of the hypothesis given the data. The name for this is the "posterior probability."

Steps taken are:

1. Collect information about how the prior distribution looks like (not a necessity; we can use uninformed priors)
2. Deciding the decision boundary (for e.g. we want to 90% sure that version A is better than B)
3. Collecting the data
4. Calculating the posterior distribution for conversion rates of both the variants
5. Inference

Advantages

1) It lets us utilise prior information.
2) Utility functions can be tailored to specific experiments.
3) Which experiment design will inform the most about the model can be predicted before experiments in a laboratory are conducted.
4) It provides a convenient setting for a wide range of models, such as hierarchical models and missing data problems.
5) It provides interpretable answers, such as “the true parameter has a probability of 0.95 of falling in a 95% credible interval.”
6) It obeys the likelihood principle.
7) BED is also useful when the model raises specific physical questions that are answered by some experiments but not by others.

Disadvantages

1) This design can’t be used in real problems without using software as it involves various integrations and numerical optimization.
2) Choosing the prior requires a lot of skill, expertise and experience.
3) Posterior distributions can be heavily influenced by the priors.

4) It doesn't tell us how to select the prior. There is no correct way to choose.
5) High computation cost for models with large number of parameters

Example

We have a population of trees in a forest, from past experience it is known that 40% of trees are eucalyptus. A person goes into the forest to test this hypothesis and collects data for 300 trees, out of which he found only 20 trees of eucalyptus so p^ is just 0.067. In general following the frequentist approach statistician would say we reject the hypothesis, but the Bayesian approach would suggest, taking into consideration that trees may be in clusters and not spread homogeneously, that our result may not be correct, he would recommend a probability distribution of the population parameter with mean of 0.4.

6) Quasi-Experimental Design

A quasi-experiment is an empirical interventional research design in which randomization is not used to determine the causal effects of an intervention on a target population. The absence of randomization renders the experiment quasi-natural. This distinguishes between quasi-experimental designs and randomised experimental designs. Due to this, the internal validity, that is, the confidence you have that the causality is not due to outside factors, is not quite as strong as in an actual experiment as the confounding variables are inherent in the preselected elements.

Advantages

1) High external validity
2) High control over targeted hypotheses
3) Ability to be combined with other methodologies

Disadvantages

1) Low internal validity
2) Risk of inaccurate data
3) Risk of bias

Example

Consider a psychological health survey that you wish to conduct among mentally unstable patients for research of a drug that you are developing. But as per the ethics of the organisation, you cannot directly convey a survey among the patients about the treatment. So in such a scenario, we conduct a quasi experimental design approach. In this approach, we look at already collected data from the psychologists and do our survey according to that.

ANOVA Procedure for Randomised Block Design

Consider the following case where we apply ANOVA to an experiment conducted in the Randomised block design setting.

In this case, some students were first segmented into six categories based on their IQ, and each of these groups was further divided into three more groups corresponding to three different types of teaching methods. These students then were evaluated and the average scores of the tests corresponding to the IQ range and teaching method have been given to us in a table as shown below. Our aim is to check if there is a difference in the effectiveness of the method of teaching. In this test, we would take the alpha as 0.05.

	Teaching Method		
IQ	A	B	C
91-95	84	85	85
96-100	86	86	88
101-105	86	87	88
106-110	89	88	89
111-115	88	89	89
116-120	91	90	91

Our analysis starts by declaring the null hypothesis and the negation of the null hypothesis.

$$H0 = \mu1 = \mu2 = \mu3$$

H1 = All means are not equal

Factors of Interest- Teaching methods (3)
Blocks- IQ groups (Six)

Observations are obtained from K independent groups. Within the i th group, there are ni observations of a single variable Y. We use Yij to represent the value of Y from the j th individual of the i th group.

For this, we are going to use the F-test and the test statistic is as follows.

$$F_{K-1,N-K} = \frac{MSB}{MSE}$$

Here MSB and MSE are given by the following equations.

$$MSB = \frac{1}{K-1}\sum_{i=1}^{K} n_i(\bar{Y}_i - \bar{Y})^2$$

$$MSE = \frac{1}{N-K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_i\right)^2$$

The quantities MSB and MSE are often called the mean square between and mean square error, respectively.

There are other two types of mean which would use here, one is the overall mean, and the other is the block mean, and they are given by the following formulae.

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{K}\sum_{j=1}^{n_i} Y_{ij} \qquad \bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$$

The one on the left is the overall mean and the one on the right is the block mean.

The block means can be calculated very easily using a spreadsheet and the means are as follows.

Groups	Count	Sum	Average
Row 1	3	254	84.66666667
Row 2	3	260	86.66666667
Row 3	3	261	87
Row 4	3	266	88.66666667
Row 5	3	266	88.66666667
Row 6	3	272	90.66666667

Now we can calculate MSD and it is as following:
MSD = ⅕ * (3*9.3228 +3* 1.1095 + 3*0.5184 + 3*0.8961 + 3*0.8961 + 3*8.682)

MSD = 12.8555

MSE for this data can be also calculated, and I am directly writing down the calculated value below.
MSE = 0.6111

Hence, the F value would be;

F = 12.855/0.6111 = 21.0363

The corresponding p-value is 0.00001473449482, and it is less than 0.05.
Thus, we fail to reject H1.
Thus, there is no significant difference in the effectiveness of the three teaching methods.

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit

Between Groups	64.27777778	5	12.85555556	21.03636364	0.000014734 49482	3.1058752 36
Within Groups	7.333333333	12	0.6111111111			
Total	71.61111111	17				

References

1. https://en.wikipedia.org/wiki/Bayesian_experimental_design
2. https://staffblogs.le.ac.uk/bayeswithstata/2015/08/07/bayesian-experimental-design-part-ii/
3. https://staffblogs.le.ac.uk/bayeswithstata/2015/07/31/bayesian-experimental-design/
4. https://www.sciencedirect.com/science/article/abs/pii/S0894177711001944
5. https://towardsdatascience.com/exploring-experimentation-in-bayesian-territory-8e2c40df77bd
6. https://cxl.com/blog/bayesian-frequentist-ab-testing/

TAJ MAHAL: SAVING THE CORRODING BEAUTY

Prepared for:

Dr. Anupam Singhal

Instructor In-charge, Environmental Impact Assessment (CE F434) Department of Civil Engineering, BITS Pilani

Prepared by:

Namit Shrivastava	**2020A2PS1767P**
Kavya Mishra	**2020A2PS1731P**
Prashansa Mamgain	**2020A2PS1762P**
Palak Solanki	**2019B2A21053P**

TABLE OF CONTENT

ACKNOWLEDGEMENT

We'd like to thank Dr. Anupam Singhal for directing us to his informative articles and for prompting us to research a proposed explanation for the marble discoloration in the Taj Mahal. The talks and conversations got us to think in new ways and from different points of view, so we could write a full and fair review.

OBJECTIVE

Millions of tourists flock to India each year to see the Taj Mahal and its famous white marble domes. Over the past few decades, the marble on the outside of the Taj Mahal has changed color and needs to be cleaned carefully every few years to keep its original beauty.

The building's location and design make for a number of exposure scenarios that can be used to study how surface moisture makes it easier for pollutants to move from the ground to the air. up to the surface and into the air. At sunrise, the Taj is oriented so that the sun's rays will cast a shadow across the building's northern face. When the patterns of discoloration in the Taj's marble are looked at, they support the idea that the building is most exposed to air pollution in the morning, when dew is on the building and the air is moving enough to bring pollutants into contact with the building's surface.

People know that discoloration is caused by pollution in the air in the Agra area, but the exact sources of this pollution have not been found. In light of this, samples of ambient particulate matter (PM) were taken over the course of a year and found to contain relatively high concentrations of light-absorbing particles like black carbon (BC), light-absorbing organic carbon (brown carbon, BrC), and dust, all of which could potentially discolor the Taj Mahal's marble surfaces. Particles deposited on marble surrogate surfaces at the Taj Mahal have been analyzed. Based on the results, it looks like a lot of the outside surfaces are covered with both dust and carbonaceous particles.

Although industrial pollutants, vehicle exhaust, and biomass burning have all been blamed for the soiling, emissions from the burning of open municipal solid waste (MSW) may also have a significant impact. Biomass burning emissions, which would include MSW emissions, accounted for nearly 40% of organic matter (OM)—a component of PM—deposition to the surface of the Taj Mahal in a recent source apportionment study of fine particulate matter (PM2.5); dung cake burning, used extensively in the region for cooking, was suggested as the culprit and is now banned within the city limits, but the burning of MSW, a ubiquitous practice in the area, may play a more important role.

Open municipal solid waste burning results in approximately 150 (130) mg m^{-2} yr^{-1} of PM2.5 being deposited to the surface of the Taj Mahal, while dung cake burning results in approximately 12 (3.2) mg m^{-2} yr^{-1}. Trash burning in lower socioeconomic class neighborhoods is the primary contributor to an estimated 713 (377-1050) premature deaths in Agra each year. The Taj Mahal, human health, and the built environment as a whole would all benefit from an efficient MSW management plan.

Ongoing investigations into these accounts have converged on the hypothesis that acid rain is the cause of an illness dubbed "Stone Cancer." Reports obtained from all across the world also documented such deterioration of stone structures, acid rain was revealed to be the root cause of this problem. Harmful gasses like NOx and SOx, together with chromium and other heavy metals, damage the atmosphere. In addition to industrial processes, automotive exhausts and even domestic processes can contribute to the accumulation of these pollutants in the atmosphere. These nitrogen and sulfur compounds emit acids when they combine with water. Hydrogen ions are found in greater concentrations in such precipitation, making it unusually acidic (pH is low). This is known as acid rain, and it is extremely hazardous to the health of vegetation, animals, and man-made structures alike.

As a result of our research, we have come up with a new way to figure out how these deposited particles affect the way visible light is reflected off of a surface and, in turn, how the human eye sees color. Carbonaceous particles (both BC and BrC from burning fossil fuels and biomass) and light-absorbing dust that have settled on the building may be what is making the outside of the Taj Mahal darker. In general, the results show that the deposition of light-absorbing particulate matter in areas with high aerosol loading is changing the look of both natural and man-made surfaces.

A contrasting image showcasing the different shades of Taj Mahal

LITERATURE REVIEW

Studies have shown that poor air quality is to blame for the soiling and discoloration, so steps have been taken to mitigate the effects of local air pollution near the Taj Mahal. These steps include limiting vehicle access to the complex, shutting down over 200 businesses in Agra, mandating that iron foundries install scrubbers and filters on their smokestacks, preventing the construction of new polluting businesses within a defined buffer zone around the mausoleum, and more. Biomass burning is responsible for roughly 40% of all organic matter (OM) deposition to the surface of the Taj Mahal, according to a recent source apportionment analysis of fine particulate matter (PM2.5, whose particles are less than 2.5 m in aerodynamic diameter).

The open combustion of municipal solid waste (MSW) and dung cake burning are two sources of biomass burning PM2.5 in Agra, both of which would be included in the assessment of deposited OM. As a result of the high concentrations of particulate matter (PM) in the air in Agra, visibility is reduced, further diminishing the Taj Mahal's aesthetic attractiveness.

The degradation of vision and the coloring of the Taj Mahal are likely the most apparent results of municipal solid waste (MSW) and dung cake burning in the vicinity; however, the health of the local population is also at risk. Among the 67 environmental factors linked to premature mortality, the Global Burden of Disease (GBD) ranks exposure to ambient PM pollution as the fifth leading cause of premature mortality in India, after high blood pressure, indoor air pollution (which is also affected by dung cake burning), smoking, and dietary risks. The greatest effect on mortality in India due to outdoor air pollution is due to energy use from homes and businesses, especially biomass burning for heating and cooking.

Agra's rapid population growth and inadequate municipal solid waste (MSW) management infrastructure have combined to produce inefficient waste management, leading to the accumulation of excessive amounts of trash on the city's streets. As a result of inefficient combustion and increased pollutant emissions, garbage is often burned in the open on roadsides and in residential and business locations in Agra and throughout India. According to the Central Pollution Control Board of India, municipal solid waste incineration accounts for between 5 and 11 percent of main PM emissions from urban sources. Chlorinated organics, dioxins, polyaromatic hydrocarbons (PAHs), various volatile organic compounds (VOCs), and

heavy metals including lead, cadmium, and mercury can all be found in the combustion byproducts of plastics and other garbage, which are included in MSW emissions with biomass. The GBD strategy does not take into account the unique dangers posed by these harmful chemicals.

FIGURE 1. A clean surface and an area being prepared for cleaning on a marble Mosque dome at the Taj Mahal

In Indian cities, dung cake combustion as a cooking fuel has been extensively explored; 11% of rural Indian households rely on cow dung as their major cooking fuel. Burning of open MSW and dung cake appears to be more prevalent in locations with poorer inhabitants, hence worsening the exposures of more sensitive groups. Emissions of municipal solid waste and dung cake can also affect the radiative balance and result in regional and global change.

Contributions of MSW and dung cake burning to ambient OM and BC (pollutants known to discolor surfaces) concentrations in Agra, deposition to and soiling of the Taj Mahal, and health impacts are evaluated by quantifying location-specific MSW and dung cake burning emissions, performing air quality and deposition modeling, and conducting a health impact assessment. This data can be used to analyze the possible advantages of policy measures, such as improved MSW collection management techniques and the accompanying infrastructure in and around Agra.

Several variables combine to make the Taj Mahal an attractive and enlightening research topic. Particularly, its location causes regular dewfall, and the widespread use of high-sulfur coal in local lime kilns ensures high ambient concentrations of sulfur dioxide.
Moreover, discoloration of the Taj is noticeable in some areas, particularly those that are protected from falling rain.

Figure 1, which was previously exhibited, depicts localized staining of marble at the Taj Mahal, in an area where it appears that direct sunshine in the morning can evaporate dew from certain regions but not others. The view is gazing upwards, late in the afternoon, at the rain-protected underside of a niche in a north-facing wall. The right side of the niche receives the direct sunlight beam in the morning and so dries quickly after dew deposition at night; it is less discolored than the left side, which is obscured by the morning sun.

Vegetation is known to reduce pollution in an unexpected manner and can therefore be utilized in urban environments; social forestry is one method for combating pollution monsters. The government adopted this strategy, and soon a "Buffer Zone" was established around the Taj Mahal. This lush vegetation not only increased the attractiveness of the location, but also offered a viable solution for conserving the location and the marble marvel.

Our experiments on another part of the strategy revealed, to our surprise, that the trees that were supposed to defend the monument were in danger. The combination of a high SPM level and heavy metal pollution increased the infection potential of fungal pathogens. The buffer zone is predominantly populated by two species of trees, out of a number of other species. One was Mahogany and the other was Mimusops elangi, and their occurrence rate was approximately 50-55%.

METHODOLOGIES

Collecting and Analyzing Ambient Particles

Beginning on November 5, 2011, and ending in June of 2012, right before the start of the monsoon season, ambient aerosol samples were taken around the Taj Mahal to ascertain the impact that particulate matter (PM) had on the building. Every six days of the month, filters were collected to test for major anions, organic carbon, elemental carbon, and trace elements in PM2.5 (fine particulate matter with diameters less than 2.5 m) and total suspended particulate matter (TSP).

Upstream cyclone data was used to determine the PM2.5 threshold, whereas ambient air was measured directly by the TSP. Monthly composites were created by combining samples from each filter, and these were then extracted and evaluated by GCMS for source-specific trace organic chemicals. Chemical mass balance (CMB) modeling was employed to determine source contributions to particulate organic carbon based on quantities of trace organics.

Target Sampling and Analysis of Marble Deposits

Several precleaned marble deposition targets (with dimensions of 2X2X0.5 cm) were placed outside within about 300 m of the main dome of the Taj Mahal, in addition to ambient samples. Targets and air sampling equipment were placed in a less-visited area of the Taj Mahal that was only accessible to ASI employees. From April through June of 2012, precleaned marble cuboids were affixed to the Taj Mahal's superstructure with double-sided tape at various points. A few of the marble slabs were laid on their sides, while the others were set up vertically.

The marble samples were maintained in the freezer both before and after exposure to prevent the decomposition of the deposited particles.

Two parallel marble targets were subjected to scanning electron microscopy (LEO 1530, Carl Zeiss Microscopy) and energy dispersive X-ray spectroscopy (Oxford Instruments Xmax detector). To catch particles from 100 nm to 100 m in size, numerous images were taken at various magnifications. With the help of image processing in Matlab, we were able to extract information about particle sizes and shapes from SEM images. Around a thousand particles

were analyzed using EDX on the identical marble targets. Chemical composition and particle concentration as a function of target area might be estimated using data from SEM/EDX studies of the marble target. Next, we'll explain how we used this data to calculate an approximation of the marble's surface color shift.

Relating Marble Surface Color to Deposition Particles

We devised a method that evaluates the impact of deposited particles on wavelength-dependent surface reflectance to predict the effect of particle deposition on the apparent colour change of the marble substrate. This technique expands upon earlier efforts to calculate the impact of particle deposition on plant leaves on the amount of light accessible for photosynthesis. To begin, we estimate the optical depth of the deposited particles as a function of wavelength (τ_λ) using SEM/EDX studies of the particles deposited to the marble targets.

$$\tau_\lambda = \frac{\pi}{4} \sum_{i=1}^{n} \mathrm{Ac}_i D_{\mathrm{p},i}^2 \left[Q_{\mathrm{s}\lambda} + Q_{\mathrm{a}\lambda} \right]_i \quad \textbf{1}$$

$Q_{s\lambda}$ and $Q_{a\lambda}$ are the wavelength-dependent Mie scattering and absorption efficiencies, respectively, determined based on particle size and composition, where Ac is the areal particle number concentration (number of particles deposited per area of the marble surface) for each size bin, i and D_p is the particle diameter for deposited particles.

Following the determination of the optical depth, the wavelength-dependent single scattering albedo, ω_λ (ratio of light scattering to extinction), is calculated as

$$\omega_\lambda = \frac{\sum_{i=1}^{n} \mathrm{Ac}_i D_{\mathrm{p},i}^2 Q_{\mathrm{s}\lambda,i}}{\sum_{i=1}^{n} \mathrm{Ac}_i D_{p,i}^2 \left[Q_{\mathrm{s}\lambda} + Q_{\mathrm{a}\lambda} \right]_i} \quad \textbf{2}$$

The single scattering albedo is a crucial element that governs the proportionate quantity of light absorption on the white marble's surface. The single scattering albedo for white, scattering-only particles is close to 1, and the surface reflectance of a white surface will not change.

The change in surface reflectance of the white marble surface is computed using SBDART, a radiative transfer model with input values including and estimated from **eqs. 1 and 2,** and the asymmetry parameter (relative amount of light scattering in the forward direction) as a function of wavelength using Mie theory.

To determine the perceived colour change of the white marble surface due to particle deposition and the corresponding change in spectral surface reflectance, we utilised the model described by D'Andrade and Romney to convert spectral reflectance to perceived colour according to the Munsell colour system. Value (lightness/darkness), hue (colour), and chroma (purity/saturation) are the three components that comprise the Munsell colour system. The model used to assess the perceived colour of the marble surface with deposited particles estimates the Munsell colour using the spectral reflectance from the radiative transfer model of the marble surface loaded with particles. The Munsell colour estimation also accounts for the human eye's reaction as a function of incident light wavelength.

The garbage and dung cake stockpiles can now be viewed

To measure the regional and temporal developments of open MSW burning, waste burn rate inventories were compiled in Agra using a recently established field transect method. Researchers walk the transect (route/line) and keep track of MSW burning episodes, estimated weight, and composition at a set distance from the transect line (usually the distance that can be seen). Next, we calculate the MSW burning occurrence density using the total number of incidents and the surveyed region.

This study used two different transect routes in Agra to assess the density, composition, and tonnage of rubbish burn over the course of three days for each route between May 30 and June 2, 2015. Based on census data from the tract level, these surveys evaluated MSW burning in 14 communities of varying socioeconomic status (SES). Since the very high levels of garbage burning found in neighbourhoods or near roads are not captured by satellite-driven studies at the global scale, the on-ground field technique is crucial to generating an enhanced PM emission inventory from MSW burning.

Using the SES-based trash burning rates, one may calculate the open waste burn rate, TWB_i (g-MSW day^{-1}), in a given electoral ward, i.

$$TWB_i = WBR_{lowSES} * POP_{i,lowSES} + WBR_{highSES} * (1 - POP_{i,lowSES}) \quad (1)$$

in which $POP_{i,lowSES}$=illiterate population in the ward as recorded in the 2011 census and $WBR_{highSES}$=daily per capita trash burn rate of the high SES. In this research, literacy was used as the key indicator of social status, and 64 percent of the population of Agra was found to be literate. Electoral districts were used to compile waste burn inventories, with each district simulated using its own dedicated emission grid, and five more regions were also simulated.

Cow dung cake usage as a cooking fuel was evaluated using census data.
The percentage of homes utilizing various cooking fuels was reported by the census at the ward/precinct level. The number of houses using cow dung as a fuel for cooking in each ward/precinct was multiplied by the yearly per-household consumption to get the annual burning inventory, which was then converted to the daily average emission rate. Additionally, the effects of firewood and crop residue on air quality were modeled using the same approach.

Inventory of municipal solid waste and dung cake burns for use in AERMOD dispersion modelling

Open MSW and dung cake burn rates were incorporated into AERMOD, a Gaussian plume dispersion model, in order to spatially characterize the yearly mean ambient PM2.5 concentrations from MSW and dung cake burning. AERMOD is a recommended regulatory air pollution dispersion model, but it does not account for atmospheric chemical processes or secondary pollution production, so it has limits. The results provided here are source-specific consequences from emissions within the study domain; background transport is not taken into account. AERMET, a meteorological input to AERMOD, utilized integrated hourly surface data from the National Climatic Data Center (NCDC) at the Agra Station of the National Oceanic and Atmospheric Administration (NOAA) and upper air data from the US National Weather Service

(NWS) at the Delhi Station. The Global 30 Arc-Second Elevation (GTOPO30) Digital Elevation Models were utilised in AERMAP, a terrain processing input to AERMOD.

Using published emission factors, we calculated the OM and BC source emission rates for both MSW and dung cake burning. In this work, PM2.5 component-specific emission parameters for MSW burning are based on measurements of waste burning in peri-urban settlements near Mexico City at different combustion phases. OC(CO_2) and black carbon (BC) emission factors were 5.3 (±4.9) and 0.65 (±0.27) g kg^{-1} respectively. In recent studies of garbage burning in Nepal, several samples were enriched for specific compositions of plastic and foil; these emission factors are within the stated range of 0.04-9.97 g BC kg^{-1} burned, but are less than the reported range of 8.4-73.9 g OC kg^{-1} burned. As a result of differences in MSW composition and combustion stage, estimates of MSW emissions are often subject to large margins of error. Households all throughout the Indo-Gangetic Plain had their emissions from dung cake burning measured. The emission parameters for OC were multiplied by a factor of OM/OC of 2.1 to take into account the presence of components other than carbon in organic compounds.

Dry deposition and contamination of the Taj Mahal

Pollutant deposition on the Taj Mahal's surface contributes to its browning, hence the impacts of wet and dry deposition from municipal solid waste and dung cake emissions were assessed. Using predicted concentrations, measured size distributions, and size-dependent deposition velocities, dry deposition rates were estimated. The deposition velocity is a parameter that combines the aerodynamic transport across the atmospheric surface layer, the transport across the quasi-laminar sublayer, and the uptake at the surface. The average particle size of carbonaceous PM species at the surface of the Taj Mahal was determined using scanning electron microscopy (SEM) (LEO 1530, Carl Zeiss Microscopy) and energy dispersive x-ray spectroscopy (Oxford Instruments Xmax detectors). The average particle size was less than 1 micron.

The PM2.5 component specific mass fluxes (g m^{-2} s^{-1}) of OM and BC to the surface of the Taj Mahal due to dry deposition were determined to be as follows:

$$F_i(t) = -V_{D,i}(d_{p,\text{ave}}) * [C_i(t)],$$

where V_D is the velocity of size-specific surface deposition (m s^{-1}) and $d_{p,ave}$ is the mean particle diameter. The pollutant concentration utilised here, $[C_i(t)]$, is the yearly average ambient pollutant concentration from open garbage and dung cake burning at the Taj Mahal, as measured by AERMOD. In this research, wet deposition was included to account for precipitation; nevertheless, wet deposition loadings were minor compared to dry deposition loadings.

From the modelled number of particles deposited per area of the surface and the total surface area of the aerosol deposited per area of the surface, we were able to calculate the percentage of the Taj Mahal's surface that is covered by pollutant deposition from MSW and dung cake burning emissions. N, the number of particles per square metre, was calculated using the following sources and pollutants:

$$N_i = \frac{\sigma_i}{\rho_i {d_{p,\mathrm{ave}}}^3/6},$$

where σ_i (mg m^{-2} yr^{-1}) is the source-specific pollutant loading, ρ_i is the pollutant (OM or BC) density, and $d_{p,ave}$ is the measured mean particle size (1 m) at the monitoring site.
Wi, the fractional coverage of PM2.5 emissions in a year due to MSW and dung cake burning, was then calculated using the average particle surface area as follows:

$$\Omega_i = \frac{6\sigma_i}{\rho_i d_{p,\mathrm{ave}}}.$$

The origin of the infectious agent

Mahogany and Mimusops elangi trees were planted in the buffer zone around the Taj Mahal, and the infected leaves of both species' leaves were gathered from those trees.

The growth media

Each and every growth medium came from Merck Specialties Pvt. Ltd., which was the vendor (Mumbai, India). The potato dextrose agar medium was used for the cultivation of all of the

cultures (PDA). In order to determine the degree of germination inhibition caused by the fungal spore, the widely used fungicide blitox was tested. In order to provide an acidic environment around the leaves, nitric acid was utilized. Absolute ethanol was obtained from Changshu Chemical for this particular application (Yangyuan, China).

Microscopy

All of the leaf slices and spores were observed with the use of a compound light microscope, and a cotton blue stain made of lactophenol was utilized.

Determination of the SPM count and pH of the leaves gathered from the Taj Mahal

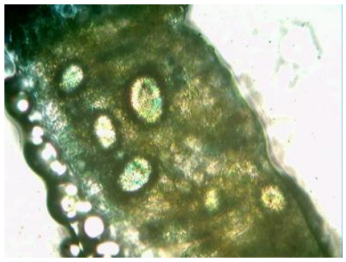
The leaves were washed in sterile, distilled water before the SPM count was calculated. Following the washing of the leaves, this water was put through a pH meter, and the pH level of the water was calculated. All of these measurements—dissolved oxygen concentration, conductivity, and number of SPMs—were taken with the same instrument.

Tables and figures:

Table1: showing the following parameters determined for the leaves collected from Taj Mahal

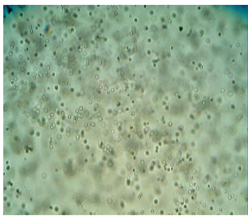
Name	Electrical conductivity (µs)	SPM (in mg)	pH
Mahogany	4.302	67.3	5.568
Mimusops elangi	35.80	120.6	5.965

(a)

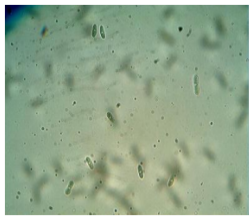


(b)

Fig1.(a) and (b) shows growth of fungal pathogen inside the cells of the leaf



(a)



(b)

Fig 2: (a) shows *Cladosporium* sp. And (b) shows *Mycospharella* sp.

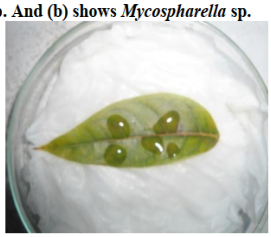


Fig 3: low infection in absence of SPM and acid vapors

Fig 7: Absence of germination in presence of fungicide Blitox

RESULTS AND IMPACT ANALYSIS

Ambient particulate concentrations

For both TSP and $PM_{2.5}$ during the sample period, Figure 2a displays the average mass concentrations of particulate organic carbon mass (OM), ions, dust, and elemental carbon (EC). TSP and PM2.5 have respective mean daily concentrations (and standard deviations) of 135 (55) and 60 (39)μgm^{-3}. The results show that the area has poor air quality since they are much higher than the yearly World Health Organization (WHO) PM recommendations for PM_{10} and $PM_{2.5}$ of 20 μgm^{-3} and 10 μgm^{-3}, respectively. The percentage of particles larger than 2.5 μm is around 60%, and it is mostly caused by coarse mode dust, which goes from making up 15% of the mass of $PM_{2.5}$ to 30% of the total suspended particle mass. In addition to dust, elemental carbon (EC), which makes up 2% of the TSP mass, and OM, which makes up 39% of the TSP mass, are other PM components that absorb light in the visible spectrum and may, thus, affect the color of the outer white marble surfaces.

According to estimates of the sources of OM in the mass fraction of $PM_{2.5}$ illustrated in Figure 2b, biomass burning, a well-known source of BrC, is responsible for almost half of the OM, with considerable contributions from vehicle emissions.

It should be noted that biomass burning OM can result from a range of processes, including the burning of trash and other waste that is common in the area, wood and dung combustion, crop residue burning, and the combustion of crop residue. Since elemental and organic carbon, as well as dust, are common light-absorbing aerosols in Agra, it is possible that PM deposition to the white marble surfaces is what is causing the famous Taj Mahal dome and other exterior Taj Mahal buildings to discolor.

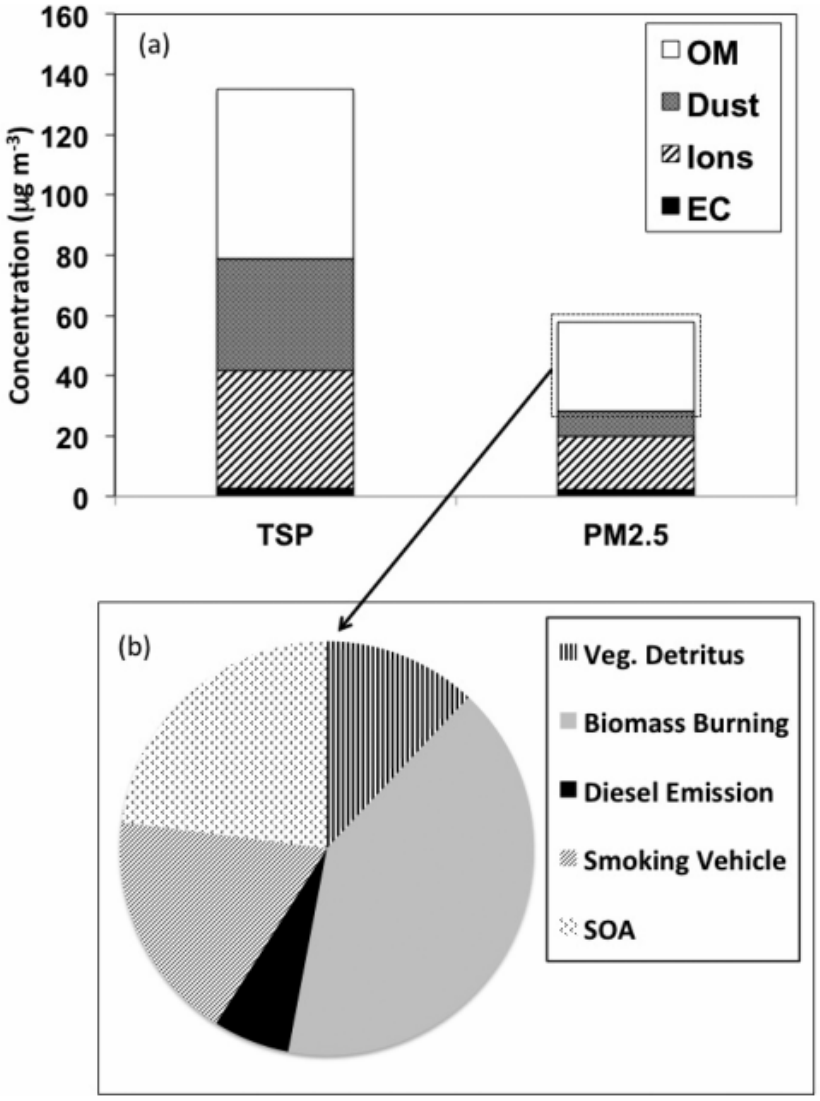


Figure 2. General chemical composition of (a) total suspended particulate matter (TSP) and fine particulate matter ($PM_{2.5}$), and (b) source apportionment of $PM_{2.5}$ organic mass (OM) based on filter sampling at the Taj Mahal.

Size and Composition of Particles Deposited to Marble Targets -

Figure 3a displays a scanning electron microscope (SEM) image of a marble deposition target that was horizontally positioned at the Taj Mahal and exposed for approximately two months in the 2012 pre-monsoon season. The distribution of particles as a function of particle size deposited to the marble target (Figure 3b) shows peaks in particle size at approximately 12 μm and another mode at 4–5 μm. Surface area concentration (surface area of particles per unit

marble surface area per micron) is a measure of the density of particles on surfaces. About 30% of the surface is thought to be coated by particles.

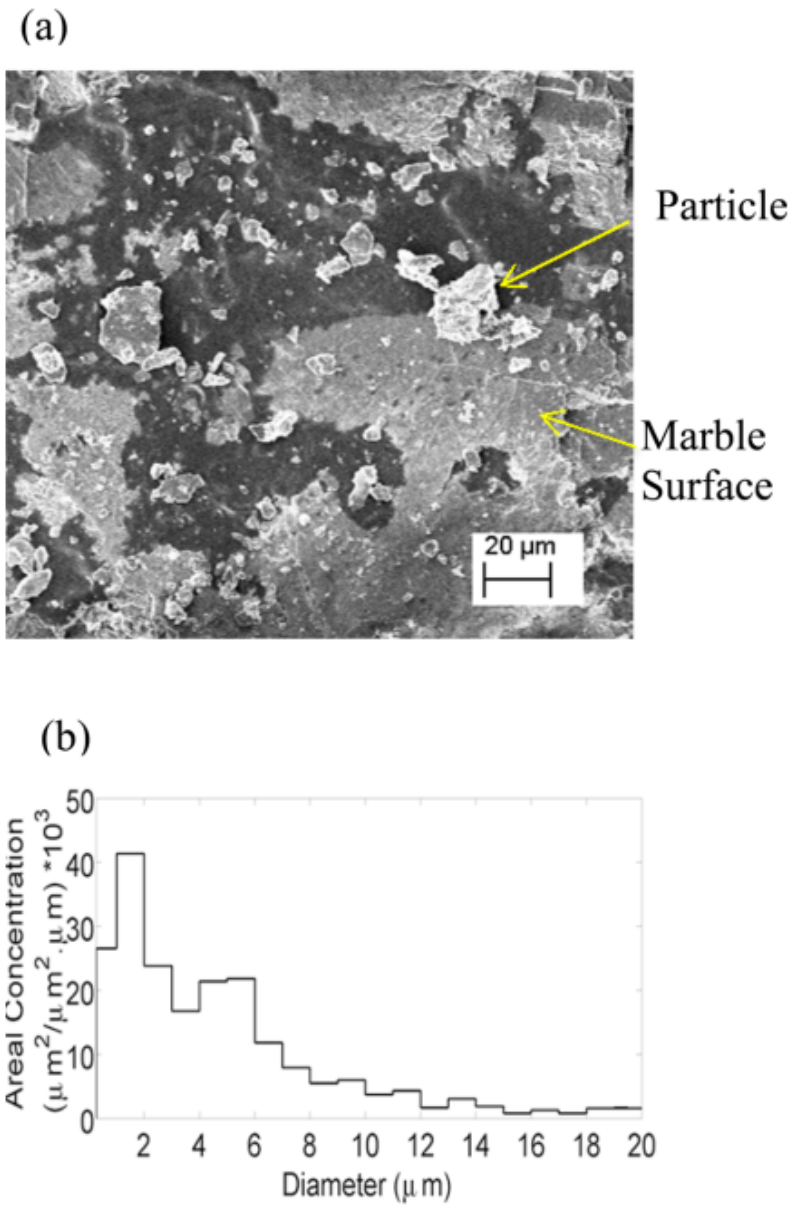


Figure 3. (a) SEM image of marble target from Taj Mahal indicating deposited particles and (b) surface area concentration of deposited particles on the marble target as a function of particle diameter.

A significant portion of the particle surface area concentration is caused by the deposition of coarse particles, as evidenced by the fact that roughly 70% of the deposited particle surface area is for particles with diameters greater than 2 µm. The relative high concentration of dust particles found in Agra and the fact that the dry deposition velocity of coarse (5 µm) particles is

nearly 100 times larger than that of accumulation mode (1.0 μm) aerosol particles are both contributing factors to the dominance of coarse particles.

Given the comparatively low precipitation in Agra throughout the autumn through spring, when particle loadings are high and the summer monsoon rainfall is not there, it is likely that dry deposition is the predominant mode of particle transport to the Taj surface. It is also crucial to keep in mind that once deposited, water insoluble particles (including dust, BC, and a small amount of OM) are probably difficult to remove from the Taj Mahal surface by precipitation wash-off. This is based on parallel observations of the accumulation of water-insoluble particles on leaf surfaces made in China's Yangtze delta, a region with significant PM loadings.

The EDX tests show that more than 70% of the particles are largely of crustal origin, with crustal components predominating in the spectrum. The majority of the crustal particles belonged to the coarse size fraction, with diameters greater than ~2 to 3 μm. Significant levels of carbon were also present in particles with diameters less than 2 μm, which make up about 30% of the deposited particle surface area and are most likely from the OM sources shown in Figure 2b's CMB data. SI section S2 provides more thorough information on the EDX analyses, including sample particle EDX analysis spectra.

Deposited Particles and Perceived Color -

To estimate τ^{λ}, we assume that particles with diameters smaller than 3 μm are predominantly made of light-absorbing organic carbon (BrC) with wavelength-dependent refractive indices as reported by Liu et al. This assumption is based on the fact that ambient filters indicated that roughly half of the $PM_{2.5}$ was carbonaceous, as well as the fact that we did not observe the presence of major ion-related elements (i.e., S) deposed to the marble targets, but did observe a predominance of carbon particles in the less than 3 μm particle size range. In addition, we assume that 10% of the particles smaller than 3 μm are black carbon (BC) particles with a refractive index comparable to that of soot. We probably overestimate the effect of BrC and BC absorbing light because we think that all particles smaller than 3 m are made of carbon and not of elements or ions.

The contributions of dust, BrC, and BC to aerosol optical depth at 400 nm are estimated to be 0.222, 0.144, and 0.016, respectively, demonstrating the significance of light extinction by all three components. The single scattering albedo at 400 and 700 nm is estimated to be 0.64 and

0.95, respectively, demonstrating that near-ultraviolet wavelengths absorb more light than the longer 700 nm wavelength.

Figure 4a depicts the surface reflectance of a pure marble surface in addition to the estimated surface reflectance for multiple situations, including the influence of each light-absorbing particle component independently (BC, BrC, and dust), and the case when all components are combined. As seen in the graph, BC absorbs uniformly across all wavelengths, whereas brown carbon exhibits preferential absorption near 400 nm for shorter wavelengths. Because hematite absorbs light at shorter wavelengths, dust greatly reduces the surface's ability to reflect light at all wavelengths, but especially at shorter wavelengths. Dust, BrC, and BC, when combined, are anticipated to significantly affect the surface reflectance with increased absorption (i.e., lower surface reflectance values) at shorter wavelengths.

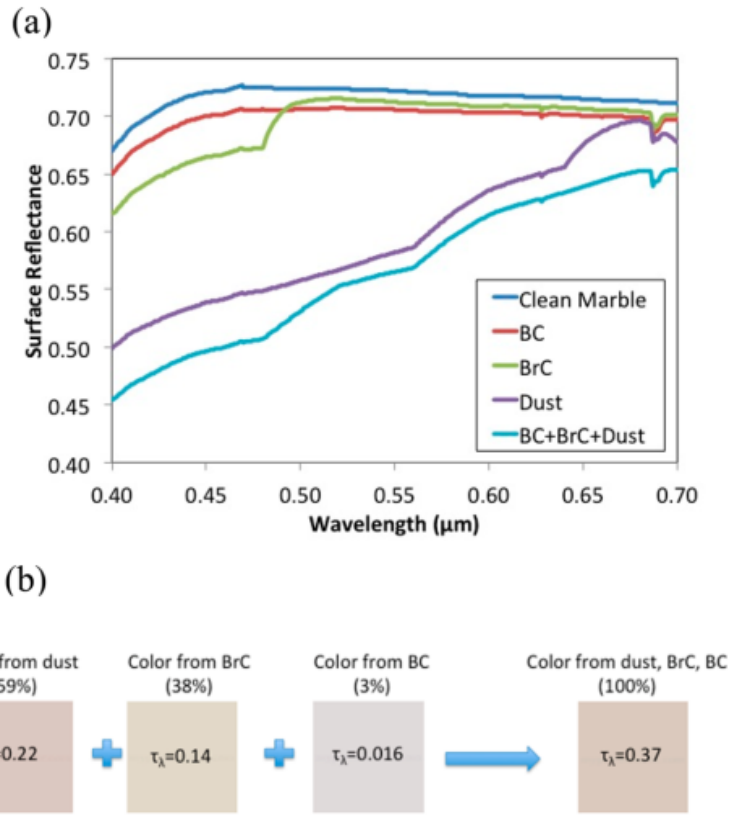


Figure 4. (a) Estimated marble target surface reflectance for a clean surface and surface area coverage of particles based on Figure 3 for black carbon (BC), brown carbon (BrC), dust and all particles (BC +BrC+dust) (b) Change in color of white marble surface for dust, BrC, BC separately and combined. Values in parentheses represent fraction of total surface area concentration contributed by each component with AOD values estimated by eq 1 at 400 nm.

Figure 4b depicts the estimated change in color of the white surface due to the deposition of light-absorbing particles, as observed in Figure 3b, over the two-month exposure time of the targets. The results suggest that each component individually contributes to the discoloration of the Taj Mahal's white marble surfaces. Due to the proportional change in surface reflectance at each wavelength, the color change for BC alone (which we think makes up 3% of the total particle surface area) is a grayish color. BrC (~30% of the total particle surface area) and dust have an effect on color, with UV absorption resulting in yellowish-brown colors. When mixed, the observed surface color moves toward darker yellow-brown tones. It should be noted that our sample targets were mounted for a relatively short period of time (~2 months) compared to the typical time between cleanings of the outer Taj Mahal surfaces (several years), so it is reasonable to expect that the perceived color of the marble target would be less intense than that of the white surfaces of the Taj Mahal.

In fact, the marble target surface appeared lighter in hue than the color estimates in Figure 4b, despite being qualitatively identical. There are numerous uncertainties involved in predicting perceived color, such as the loading, particle size, and optical characteristics. Analyses (included in SI section S3) indicate that the results are fairly sensitive to both particle loading and particle size. Assuming a 50% uncertainty for aerosol loading, for instance, has a moderate effect on the perceived color and does not alter the conclusion that dust and carbonaceous particles contribute to the observed color change of the Taj Mahal.
Overall, the results indicate that light-absorbing particles play a significant role in the discoloration of the Taj Mahal's surface, and that dust, as well as BC and BrC primarily derived from biomass combustion, trash/refuse burning, and mobile sources, all contribute significantly to the discoloration.

In places of high aerosol loading, the deposition of light-absorbing particulate matter on both natural and man-made surfaces results in a significant discoloration, according to this study. Through the changing of surface albedo and consequently perceived color, the deterioration affects not only cultural relics, but also the aesthetics of the surroundings. The measurement/modeling approach developed in this paper permits the estimation of surface color changes based on the relative amounts of light-absorbing particles deposited on surfaces, and can be used to develop future control strategies to prevent the discoloration of the environment by particle deposition, which will also improve air quality.

Open MSW and dung cake burning emissions to modeled concentrations throughout Agra and model evaluation -

Using the field transect method devised by Nagpure et al., the total average trash burn rate in Agra was calculated to be 130 g MSW per capita per day, with greater per capita burn rates seen in low socioeconomic status (SES) areas. In the city, burn rates were higher in the morning than in the evening, while the difference was smaller in rural areas (areas outside the city).

If Agra's per capita average garbage burn rate is extended to the total population of India, the annual national burn rate would be 68,000 Gg yr^{-1}, which is consistent with the model results of 35,000–75,000 Gg yr^{-1} for India by Wiedinmyer et al. The total cow dung cake burning emissions by ward in Agra were computed from household fuel consumption data and ranged from 0 to 9100 kg day^{-1} ward-1 within the study domain, whereas open waste burning emissions ranged from 490 to 25000 kg day^{-1} $ward^{-1}$. According to a survey on sustainable solid waste management in India, the average daily trash creation rate in Agra is 580 g $capita^{-1}$. Using this MSW generation rate, the average MSW combustion rate in Agra is 23%, which is greater than the 5%–10% estimates from earlier studies of trash combustion in Indian cities.

Using emission parameters from the literature and actual burn rates, the annual cumulative emissions in Agra from open waste and dung burning were estimated to be 2500 (2200) kg yr^{-1} for the OM and BC components of $PM_{2.5}$, respectively. The Taj Mahal had concentrations of 4.1 (3.8) and 0.24 (0.10) µg m3 for OM and BC from MSW burning and 0.32 (9.1 x 10^{-2}) and 0.019 (9.7 x 10^{-4}) µg m3 for OM and BC from dung cake burning, according to AERMOD simulations (figure 1 and SI figure 5). Due to changes in the types of waste and the stage at which they are burned, uncertainty was only looked at for the emission variables, which are where most of the uncertainty lies. The calculation doesn't take into account how these sources' gaseous emissions also cause PM2.5 to be made.

Using measurements from a recent $PM_{2.5}$ source apportionment research at the Taj Mahal, it was shown that biomass burning emissions contribute 12 µg m3 to OM (which can come from a range of combustion activities, including wood, crop, dung, and MSW burning) at the Taj Mahal. MSW is the largest contribution of the modeled biomass burning sources.
Maximum combined yearly average impacts on $PM_{2.5}$ in Agra were 33 (30) µg m3 from burning municipal solid waste and 3.3 (0.90) µg m3 from burning dung cake.

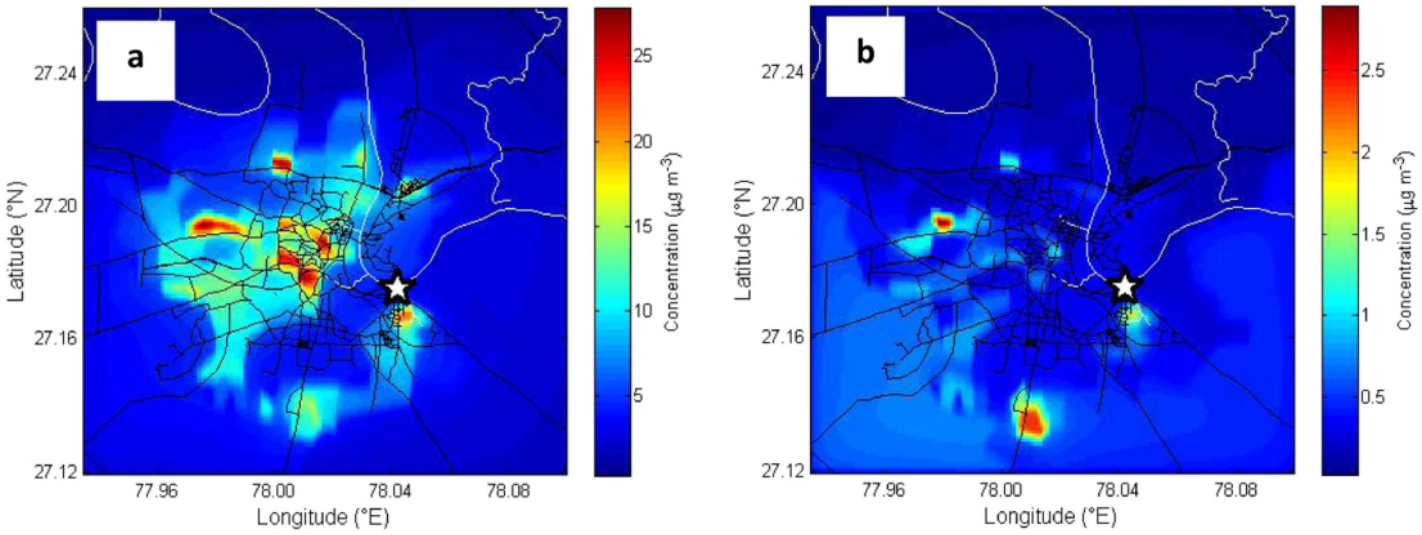


Figure 1. Annual average fine particulate matter ($PM_{2.5}$) concentrations in Agra from: (a), open MSW burning (b), dung cake burning. Modeled [$PM_{2.5}$] at the Taj Mahal (depicted by the white star) was 4.3 ($\pm$3.9) μg m^{-3} from MSW emissions and 0.34 ($\pm 9.1 \times 10^{-2}$) μg m^{-3} from dung cake burning emissions. These concentration profiles generated in AERMOD showed higher pollution from both forms of biomass burning concentrated in areas of lower socioeconomic status. Organic matter (OM) and black carbon (BC), the $PM_{2.5}$ components modeled, concentration profiles show the same spatial variation, but OM concentrations contribute more than BC to ambient $PM_{2.5}$ (SI figure 5).

High levels were discovered in low-income areas where MSW and dung cakes are commonly burned. In the city of Agra as a whole, burning trash in the open is more of a problem than burning dung cakes, except in rural areas where dung cakes are the main cooking fuel. In Agra, the combined annual average $PM_{2.5}$ concentration from open trash and dung cake burning was 4.3 (3.8) µg m3 for OM and 0.25 (0.10) µg m3 for BC. Recent ambient OC and elemental carbon concentration measurements throughout Agra ranged from 10.2 (7.2) to 30 (13) µg m3 and 1.3 (0.8) to 4.0 (1.5) µg m3, respectively, indicating that the source effect modeling results averaged across the study domain are consistent with ambient data.

Adverse health and premature mortality assessments -

According to estimates of premature mortality linked to $PM_{2.5}$ (BC OM +) emissions from burning MSW and dung cake, there are 713 (377-1050) cases of premature mortality from outdoor exposure each year in Agra, with 380 (247-540) cases attributed to IHD, 231 (98-362) cases attributed to stroke, 94 (31-170) cases attributed to COPD, and 7 (1-12) cases attributed to LC for adults (age 25). An extra 0–2 cases (age 5 years) of premature mortality from ALRI from MSW and cow dung cake burning occur each year in Agra. The total human YLL is estimated at 10087 years (5480-14 520) from one year of exposure for all-cause mortality (i.e., ALRI, COPD,

IHD, stroke, and LC) attributable to $PM_{2.5}$ emissions from MSW and cow dung cake burning, where IHD (56%) is the highest contributor followed by stroke (32%), COPD (11%), and LC(1%).

Deposition and soiling of the Taj Mahal -

Using the simulated concentrations, observed size distributions, and rainfall data, the deposition of MSW and dung cake burning emissions to the Taj Mahal via dry and wet deposition was quantified. The Taj Mahal's average surface area median diameter of the carbonaceous particles that were deposited to outdoor surfaces was determined to be ~1 μm by detailed size distributions measured on-site, which was used in conjunction with deposition velocity relationships to derive a deposition velocity of 0.11 cm s^{-1}. In earlier studies in metropolitan areas, similar deposition velocities for particles of similar size and content were measured. The Taj Mahal receives an estimated total yearly combined $PM_{2.5}$ dry deposition of 150 (130) mg m^{-2} from open waste burning and 12 (3.2) mg m^{-2} from burning dung cake (table 2).

Wet deposition loadings were lower than dry deposition. Although brown carbon (BrC), which absorbs light, makes up approximately eight times as much of an organism's mass as black carbon (BC), which is a powerful light absorber. Since both sources also produce gaseous emissions, this approach is probably underestimating the overall OM deposition from the two sources because emission factor measurements do not take secondary production into account. In order to more accurately assess discoloration, the Taj Mahal's surface pollution was assessed. If the Taj Mahal's perceived colour is affected by pollution, the fractional surface area coverage will likely surpass.

Dung cake burning emissions generated an additional 5.7x 10^{-2} (1.6 x 10^{-2}) annually, while MSW burning emissions exhibited a fractional cover of 0.73 (0.67). Since 1994, there have been four treatment cleanings. Given the intervals between cleanings, the influence of burning emissions from MSW and dung cake is expected to be more than a fractional coverage of 1, indicating that the combined deposition of both substances will result in surface discoloration.

Possible pollution threat to the Green Buffer Zone around the Taj Mahal -

After a 14-day incubation period, the isolated leaf patches gave birth to cottony fungal pathogendevelopment. Similar leaf spot disease was produced in fresh leaves by spore suspension made from the spores mentioned above that were seen. The pathogen was isolated as a result of this supporting Koch's Postulates. According to investigations done under a

microscope, the fungal growth on the underlying mesophyll cells and the infection thread that runs along the leaf lamina. The two different leaves were determined to be infected by two distinct species. While a pathogen that produced long, rod-shaped spores was found to be infecting mahogany leaves, a pathogen that also produced small, circular spores was found to be infecting Mimusops elangi.

On the agar slant, it was also noticed that the spores appeared crimson. According to studies, mahogany leaves have a similar spot disease that was caused by organisms that looked like Mycospharella. The other organism resembled Cladosporium in appearance quite a bit.

(a)

(b)

Fig 4: comparison showing greater infection(b) in presence of SPM & acid vapors

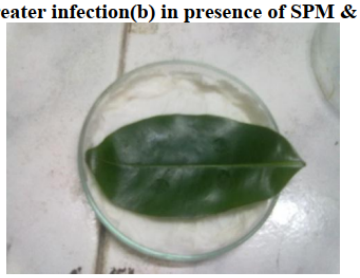
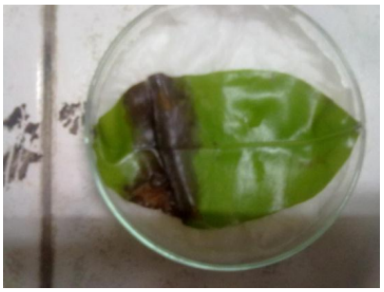
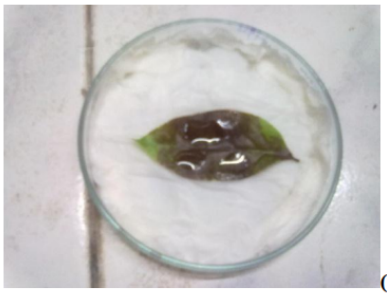


Fig 5: Absence in infection in absence of SPM and acid vapors



(a)



(b)

Fig 6:comparison showing greater infection (b) in presence of SPM & acid vapors

The average length and width of the elongated spores were measured to be 1.5 µm and 0.25 µm, respectively. Interesting findings came from the three-setup. While the one with SPM content displayed roughly 45% infection, the one without SPM and acid vapours only produced 15% infection. Further observation revealed that practically the entire leaf of the plant producing both SPM and acid vapours was infected. In one investigation, Blitox was found to entirely

prevent spore germination in the slide bio-assay, which was found to completely inhibit the percentage of infection.

Studies done on different plants have shown that SO4 can cause foliar harm that progresses into necrotic lesions and can impede a plant's net assimilation.
Studies done in the United States on perennial weeds that were fumigated showed that the plants were hurt in ways like browning of the leaves, cell death, and necrosis.
Mimusops elangi was exposed to various levels of NO2 in one experiment, and the effects included a reduction in shoot length, fresh and dry weights, and ascorbic acid levels, which led to a greater vulnerability to fungus infections.

The phyto-toxicity of SO2 in Cicer arietinum led to a variety of morphological, physiological, and biochemical reactions, including a decrease in metabolic rate and cellular functioning.
When comparing the results from the aforementioned study to those from other studies, it can be concluded that a high SPM count combined with atmospheric NOX and SOX levels causes the observed leaf damage. The infection process in these leaves may have been started by this injury.

In an experiment using field applications of seedling plants, it was discovered that blitox (copper oxychloride) inhibits the growth of Cladosporium oxysporum. In Queensland, Australia, Mycospharella musicola-caused banana leaf spot disease was effectively controlled with a mixture of white oil emulsion, malachite green, and copper oxychloride (BLITOX).
In laboratory trials, Copper oxychloride was found to have a considerable impact on reducing the germination of both pathogens. Roadside plants may therefore promote urban forestry if protected with antifungal dusting, if this result is compared to the previously reported research. Also, it could be said that trees that produce less wood but could be treated with antifungal spray could be used to restore the social forest belt around the Taj Mahal and protect the monument in the future.

However, today's antagonists are black and organic carbon particles generated by cars and other polluting devices.

CONCLUSION

Recent research shows that pollution is still hurting the Taj Mahal's white marble, which is a safe conclusion to draw. In the 1980s, people thought that sulfur dioxide was the only thing that made the sparkling facade turn yellow.

In 1996, the Supreme Court noted that "a yellow hue permeates the entire monument." According to the petitioner, the Taj's yellow hue is being accentuated by unsightly brown and black blotches, and the monument is on the road to deterioration owing to atmospheric pollution. Despite investigations attributing the deterioration to different causes, such as microscopic algae, the buildup of dirt, and the application of a resin designed to maintain the monument, nothing but pollution was found.

The 1996 court order prompted a variety of anti-pollution initiatives. The ongoing discoloration, on the other hand, shows that the Court's orders haven't been followed or that "something new and different" needs to be done.

Here are some of the main reasons why the Taj Mahal is in the shape it is in now:

- **Natural Causes**:
  Minerals that aren't pure enough and are in the marble oxidize and leave brown stains. Rain and snow can also cause chips and cracks in marble, in addition to wearing it down. The iron dowels used to secure the marble slabs to the building corrode, and the rust is deposited on the marble by the rain.

- **Tourists:**
  The number of tourists, which can reach more than 50,000 on some days, is the biggest threat to the Taj Mahal. Constant foot traffic degrades the marble floors. The building gets more humid when people come in, and the grease on their hands causes dirt to build up on the walls. The red sandstone used in the Mehman on either side of the Taj faces the greatest threat. In contrast to marble, red sandstone is porous.

- **Air Pollution:**
  A recent Indo-American study found that air pollution is making dust and carbon-containing particles fall on the Taj Mahal and change its color. Before, the "yellowing" of the monument was blamed on the Mathura refinery and small factories. This led the Supreme Court to order that these units use cleaner fuels.

- **Receding, polluted Yamuna:**
  According to news reports from 2011, the Yamuna is shrinking, which is making the sal wood in the Taj's foundations weaker. The wood needs constant moisture to keep from cracking, which is why the Yamuna is shrinking. The round wells that were cut into the project, consisting of food plazas and shopping malls planned between the Agra Fort and the Taj, were shelved on the Supreme Court's orders because it posed a threat to the monument. The ASI says that sandstorms can damage the marble surface when sand builds up in the reclaimed riverbed. The plot has not yet been converted into a green belt as the court ordered.

RECOMMENDATIONS

To preserve its unique appearance, the Taj Mahal is a global icon that requires immediate attention. In light of their grave worries, the Archaeological Survey of India (ASI) has taken a variety of measures to make the Taj Mahal more functional. The American Society of Internal Medicine (ASI) approves using mud packs as a therapy method. All the surfaces that weren't marble were polished until they shone like new. Dirt, grease, and even insect and bird droppings can be removed from marble by spreading a thick paste of clay (preferably Multani Mitti) and letting it sit for a time, then washing it with purified water. Even though this method has been tried and tested, it is hard to use and takes a lot of time. Therefore, this occurs every few years on average (1924, 2001, 2008, 2014).

ASI also gave the government a "Site Management Plan" that explains most of the steps they plan to take. This includes maintaining regular landscape upkeep, using shoe covers for guests, Taj and having roads automatically washed with water within a 5 km radius of the area.

It is the factory exhaust that poses the most threat. So, if we make it illegal for these factories to run, we might be able to cut down on or get rid of both waste in waterways and air pollution. There may also be a change in how cars are mostly used in urban areas. Gasoline-powered and electric automobiles are also viable options. Because the area has naturally low pollution, NBC should keep a close eye on it. There is a lot of smog and dust because of the crematory that is conveniently located nearby. As a result, we need to switch to electric crematoriums and shut down the factories in historic districts.

As shown by the numbers, the companies and plants along the Yamuna have wasted a total of 2,000 metric tons of water. Cleaning up the area within a 500-meter radius of the TTZ will keep it looking as good as new from any angle. In and around Agra, there is room to decrease plastic use.

Today, India is a model nation, drawing visitors from all around the world who help boost the economy. The current administration makes every effort to protect this Mughal masterpiece. However, we urbanites have a duty to treat this landmark with the awe and reverence it

deserves. It's possible that the original skin on the outside of the building has been fixed up and changed many times. Its worth transcends the centuries and remains constant.

- Cleaner fuels should be utilised such as compressed natural gas (CNG) and liquid petroleum gas (LPG).
- Changing to unleaded gas in cars.
- Relocating polluting factories and businesses to less populated areas.
- Vehicles and nearby factories contribute to Agra's air pollution by creating harmful substances including sulphur dioxide (SO2), nitrogen dioxide (NO2), smoke, dust, soot, etc. Unleaded gasoline, which has no lead and does not emit toxic compounds, is one example of a cleaner fuel that reduces pollution. This is why the Indian Supreme Court has proposed these measures to preserve the Taj Mahal.
- Agrichemical runoff and pesticide use are major causes of environmental degradation.

After the Uttar Pradesh government gave a rough plan for protecting the Taj Mahal, the Indian Supreme Court told them to do better. When the state's lawyer tried to pretend that he was surprised that the state government hadn't consulted the Archeological Survey of India (ASI) before writing the report, the Supreme Court blasted him. People all around the country are anxious about the Taj Mahal's alleged yellowing due to pollution.

The judges even asked Attorney General KK Venugopal what would happen if UNESCO took the Taj Mahal off the list of World Heritage sites. In addition, the Supreme Court urged Venugopal to clarify who exactly is in charge of the upkeep of the 10,400-square-kilometer Taj Trapezium Zone (TTZ). While the next hearing in the case has been scheduled for August 28, we must have answers as to how we can help protect India's most treasured monument from environmental degradation immediately.

- Keep India's crown jewel in pristine condition from the time it was erected in 1653 by Mughal emperor Shah Jahan by keeping the area around the TTZ litter-free.
- Reducing the usage of single-use plastics in and around Agra is another crucial step we can take to safeguard the 17th-century tomb. While a statewide ban on plastics has been enacted by the government of Uttar Pradesh, enforcement of the prohibition remains spotty at best.

- Another step we're doing to protect what has been called "the jewel of Muslim art in India and one of the globally adored marvels of the world's heritage" is prohibiting the use of gasoline and diesel vehicles within a 500-meter radius of the mediaeval dome.
- The Supreme Court heard testimony from an environmental attorney who said that insects from the River Yamuna creep into the Taj Mahal and leave their faeces on the marble floors of the massive, 42-acre structure. We need to know that the New7Wonders of the World (2000-2007) project winner is being negatively impacted by the sewage that the city of Agra is dumping into the river.
- The world's most lavish symbol of love, the Taj Mahal, has been severely damaged by acid rain created by smoke from industrial chimneys. While much of such efforts consist of just adhering to rules which are already in place, the Supreme Court's comment of what we would do if UNESCO decided to revoke the world heritage classification from a site that gets approximately 8 million tourists annually prompts some serious thought.
- Building a new dam to assist restore the flow of water to the Yamuna river, turning off some of the 52 discharge pipes spilling waste into the water, and strengthening local sewage treatment plants are all current initiatives to save the Taj.
- For many years, the ASI has relied on mud packs as one of its primary tools for cleaning the yellow stains that have developed on the Taj Mahal's white marble exterior. The technique is commonly used to clean marble and it is believed that it would help bring back the monument's original luster and hue.
- The Indian government has established the Taj Trapezium Zone (TTZ), a 10,400-square-kilometer (4,000 sq mi) area around the monument where stringent emissions limits are in place, to assist reduce air pollution.

Here are things one can do to help preserve the Taj Mahal's splendor for future generations: For starters, the government should impose a limit on the number of tourists who can enter the country in a given time period. The Archaeological Survey of India has fixed up the Mehtab Bagh, which is right behind the Taj Mahal. Along the other bank of the river, the state government has planted a thick green barrier. The authorities are using a mud therapy treatment to keep the monument at its best. The damage that insects cause to the monument is also being mitigated.

REFERENCES

https://www.iasparliament.com/current-affairs/threats-to-taj-mahal#:~:text=The%20Yamuna%20water%2C%20contaminated%20with,a%20loss%20of%20aquatic%20life.

https://www.nextias.com/current-affairs/29-09-2022/taj-mahal--pollution-warning

https://home.iitk.ac.in/~snt/pdf/Bergin_EST_2015.pdf

https://www.researchgate.net/publication/269189363_The_Discoloration_of_the_Taj_Mahal_due_to_Particulate_Carbon_and_Dust_Deposition

http://envis.tropmet.res.in/Abstract_Servicing/TajDiscoloration.pdf

https://europepmc.org/article/med/25470755

https://iopscience.iop.org/article/10.1088/1748-9326/11/10/104009

https://openarchive.icomos.org/717/1/wash44.pdf

Madhav Gadgil and Kasturirangan reports

Prepared for:

Dr. Selva Balaji M

Instructor In-charge, Introduction to Environmental Engineering

(CE F323) Department of Civil Engineering, BITS Pilani

Prepared by:

Namit Shrivastava	**2020A2PS1767P**
Prasast Singh	**2020A2PS1740P**
Sachin Shankar	**2020A2PS1744P**
Palak Solanki	**2019B2A21053P**

TABLE OF CONTENTS

ACKNOWLEDGEMENT

We'd like to thank Dr. Selva Balaji M for directing us to his informative articles and for prompting us to research a proposed explanation for Madhav Gadgil – Kasturi Rangan reports. The talks and conversations got us to think in new ways and from different points of view, so we could write a full and fair review.

INTRODUCTION

The environmental research commission Gadgil Commission is named after its chairman, Madhav Gadgil. The official name of the commission is Western Ghats Ecology Expert Panel (WGEEP). On 31 August 2011, the commission submitted its report to the Government of India.

The Western Ghats are a vast region encompassing six states, 44 districts, and 142 taluks. It is home to numerous endangered plant and animal species. The Western Ghats are home to thirteen national parks and a number of sanctuaries, which comprise the most extensive wilderness in India. These forested hills, designated by UNESCO as one of the world's eight most significant biodiversity regions, are also the source of numerous rivers, including the Godavari, Krishna, and Cauvery.

The Western Ghats serve as a massive water reservoir for six states. Now there are numerous leaks and a water shortage. Now, all the rivers have dried up. And wherever water exists, it is severely contaminated. The Western Ghats require a great deal of consideration in terms of India's sustainability, particularly in the South. The Ministry of Environment and Forests of India established the Gadgil commission in March 2010 to devise a plan for conserving these Ghats.

The territory was divided into three ESA zones by the Gadgil committee. Sixty percent of the land is designated as Ecologically Sensitive Area (ESA)-1, the highest designation possible. According to the findings of the study, this area is off-limits to any sort of construction. About a quarter of the land is classified as Ecologically Sensitive Area 3 (ESA-3), the lowest priority category, where construction is permitted. ESA-2 covers the remaining 15% by area.

As an example, in ESA-1, mining is prohibited, in ESA-2, existing mining activities are allowed to continue, and in ESA-3, new mining activities are permitted.

Why was the Gadgil Committee Constituted?

- By order dated 2010, the Ministry of Environment and Forests, Government of India, established the Western Ghats Ecology Expert Panel (WGEEP), headed by Professor Madhav Gadgil.

- To determine the ecological status of the Western Ghats, it is necessary to identify ecologically sensitive zones for notification under the Environment (Protection) Act of 1986.

- Propose modalities for the conservation, preservation, and revitalization of the Western Ghats Region through a thorough consultation with all stakeholders.

- Recommend modalities for the establishment of the Western Ghats Ecology Authority under the Environment (Protection) Act, 1986, a professional body to manage the region's ecology and promote sustainable development.

Western Ghats

RECOMMENDATIONS

Gadgil Committee Recommendation: The report also reflected that the Gadgil committee included prominent ecologists. The report was regarded as being more in favor of environmental protection and activists than of economic growth.

- The entire mountain range of the Western Ghats was declared as an Ecologically Sensitive Area (ESA) by the Western Ghats Ecology Expert Panel (WGEEP).

- Ecologically Sensitive Zones (ESZ) 1, 2, and 3 have been assigned to each of the 142 taluks that make up the Western Ghats boundaries, as described in the panel's report. Due to ESZ-1's strategic importance, nearly all industrial development (mine, thermal power plants, etc.) was forbidden within its borders.

- No new dams with large-scale storage should be allowed in Ecologically Sensitive Zone 1, according to the Gadgil report. It said that the locations for the hydel projects in Athirappilly, Kerala, and Gundia, Karnataka, should not be given environmental clearance because they are located in Ecologically Sensitive Zone 1.

- The current structure of environmental governance is called into question in the Gadgil Committee report. It preferred a bottom-up strategy (starting with Gram sabhas) over a top-down one. More autonomy for subnational governments was another demand.

- It was recommended that the Ministry of Environment and Forests establish a Western Ghats Ecology Authority (WGEA) using the authority granted by Section 3 of the Environment (Protection) Act, 1986.
-

Kasturirangan Committee Report Recommendations:

- According to the Kasturirangan assessment, only 37% (about 60,000 sq. km.) of the Western Ghats should be included in ESA.

- No mining, quarrying, or sand mining is permitted in the ESa.

- The group determined that 90% of the Western Ghats' natural landscape should be protected by ESA, while the remaining 8% should be reserved for cultural uses such as human villages, agricultural areas, and plantations.

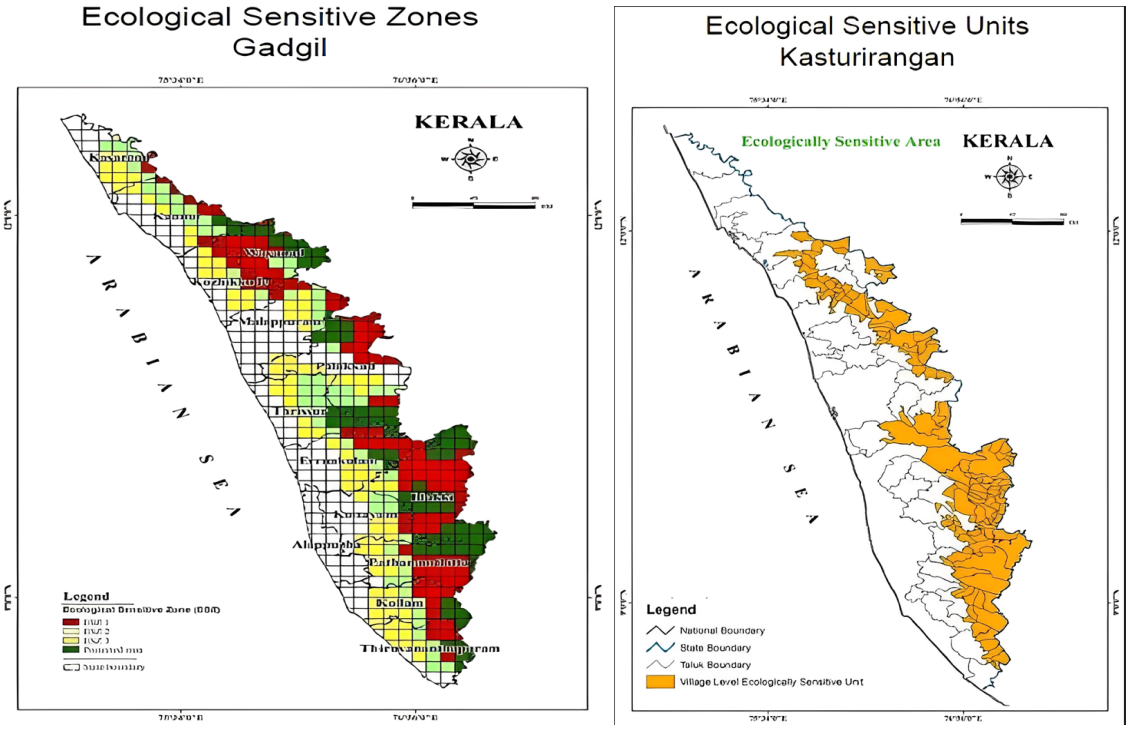
- In the next five years, or at the expiration of mining leases, whichever comes first, mining operations in the ESA should be phased out.

- We must prohibit all forms of thermal electricity and approve hydropower projects only after careful consideration.

- Polluting "red" industry should be outlawed completely in certain regions.

- The environmentally sensitive areas (ESAs) should not include populated areas or plantations, one of the many farmer-friendly suggestions in the Kasturirangan report on the Western Ghats.

- According to the Kasturirangan report, the ESA has jurisdiction over 123 communities.



CRITICISMS

Gadgil Report

- The Gadgil Committee report received the most backlash because it was seen as too environmentally friendly and out of touch with reality.
- Some others said that it would be impossible to follow the recommendations.
- The Western Ghats, which are a hindrance on the energy and development fronts for a variety of states, are the subject of a recommendation in a report by Gadgil.
- The establishment of the WGEA was met with opposition. The states maintain that they have the ability to provide protection under current statutes.
- Revenue losses caused by enacting the recommendations in the Gadgil report are not addressed.
- The power industry took a major hit with the release of the Gadgil study, which argues against building dams in the Western Ghats. Critics believe that this advice should not be implemented because India's energy demands are increasing.

Kasturirangan Report

- The area in the Western Ghats was divided into zones with the help of aerial surveys and remote sensing by the Kasturirangan panel. Many mistakes were made in the report because these methods were utilised without first investigating the situation on the ground.
- Instead of gramme sabhas having authority, bureaucrats and forest officials do. Many people are worried that if the Kasturirangan Committee recommendation is put into action, the farmers will be forced to leave their land. The mining and quarrying industries stand to benefit greatly from the recommendations made in this research. It will be very bad for the ecology if these lobbying groups and tourism boom. There will be a lack of clean water and an increase in pollution. Farmers will be forced to leave the area eventually. They can't farm there because of the climate.

- The implementation of a "erroneous method" led to many settlements being classified as Ecologically Sensitive Areas (ESA), despite the fact that the area contained solely rubber plantations and no forested land at all.

- Many environmentally sensitive places were omitted from the Kasturirangan study yet listed under ESA.

Calamities hit Kerala one after the other, but Gadgil report dissidents firm on stance

STAGES OF IMPLEMENTATION

Gadgil Report

- For the sake of ecological management, the Western Ghats need to be delimited and divided into three zones: Zone 1 (very sensitive), Zone 2 (somewhat sensitive), and Zone 3 (moderately sensitive).

- Suggesting that the entire region be labelled as an environmentally sensitive area (ESA), and suggesting that any activities that have a significant damaging or interventionist influence on the environment be prohibited in Environmentally Sensitive Zones 1 and 2.

- Suggesting the establishment of a Western Ghats Ecology Authority in order to control and oversee these activities in the region, as well as a governance structure that works from the bottom up for the environment.

- On August 31, 2011, the report was turned in to the Ministry of Environment and Forests.

- Encountering opposition from the six states that are of concern, all of which are against the suggestions made in the study.

Kasturirangan Report

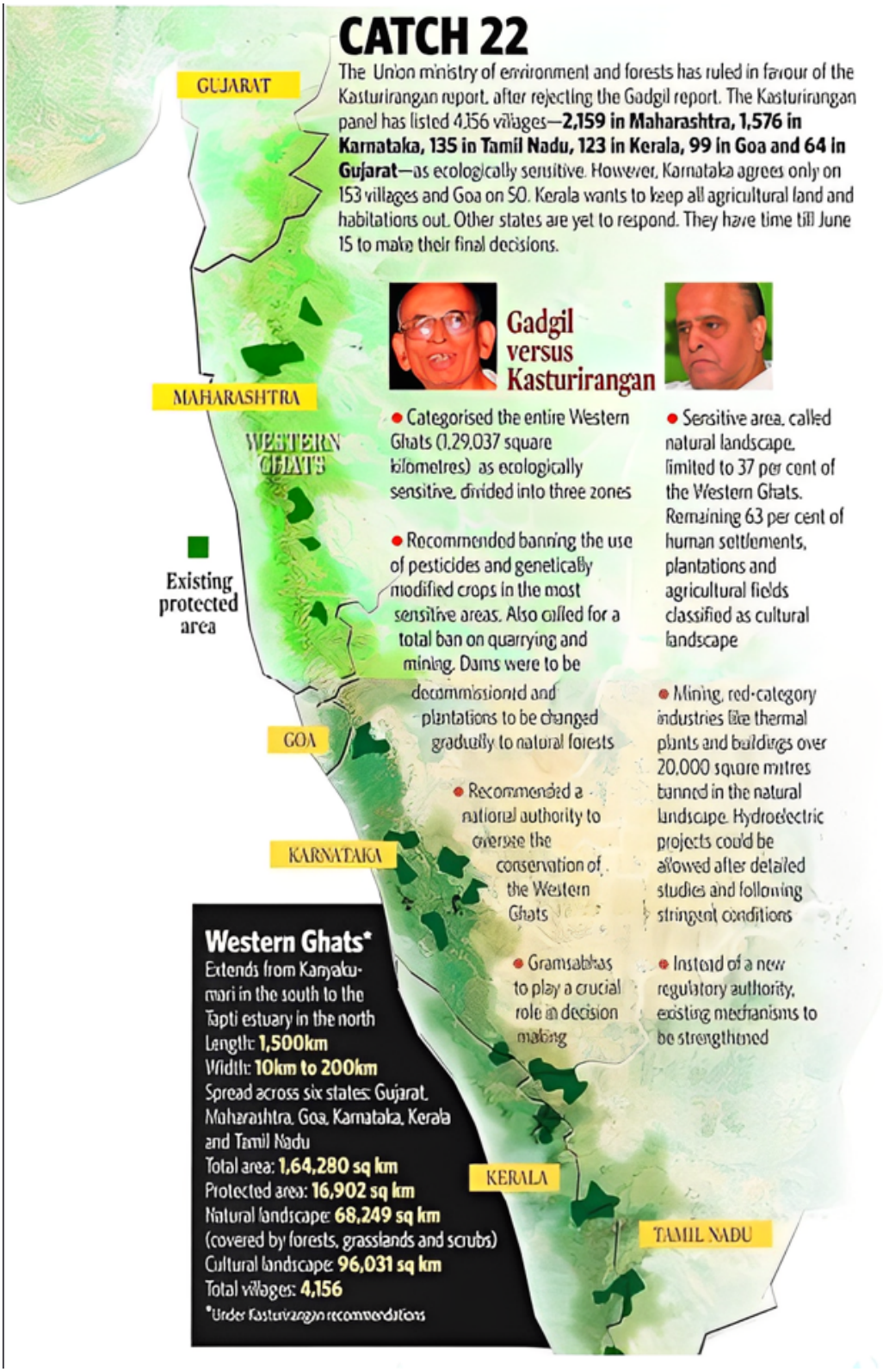
- Analysing the report of the Gadgil Committee in a comprehensive and multidisciplinary manner, taking into consideration the comments obtained from various states, central ministries, and other entities.

- Making a proposal to designate 37% of the Western Ghats as an eco-sensitive area (ESA), as well as mapping and demarcating that area using satellite imagery.

- Distributing a draught notification that is based on the report and soliciting feedback from the general public in the form of objections and suggestions.

- The formation of a powerful committee with the responsibilities of conducting a physical verification of ESA limits and addressing the concerns of the states.

- Attempting to gain the active support of the local populace as well as consulting with state administrations on a more local level in order to save the Western Ghats.



Highlighting features of the reports

COMPARATIVE REVIEW

The Madhav Gadgil Report and the Kasturirangan Report are two exhaustive studies on the conservation and sustainable development of India's Western Ghats region. Both reports provide conservation and sustainable development recommendations for the region, but their approaches and recommendations differ. The following is a comparison of the two reports:

Both reports advocate for the identification and designation of ecologically sensitive zones (ESZs) in the Western Ghats region. However, the Madhav Gadgil Report proposes a more comprehensive approach, identifying nearly 75 percent of the Western Ghats as ecologically fragile. In contrast, the Kasturirangan Report identifies only 37% of the Western Ghats as ecologically sensitive, and its approach to ESZ designation is more flexible.

Land Use Planning: The Madhav Gadgil Report suggests a more decentralised approach to land use planning, with local communities participating in decision-making processes. In contrast, the Kasturirangan Report proposes a more centralised approach, with high-level working groups preparing zonal master plans.

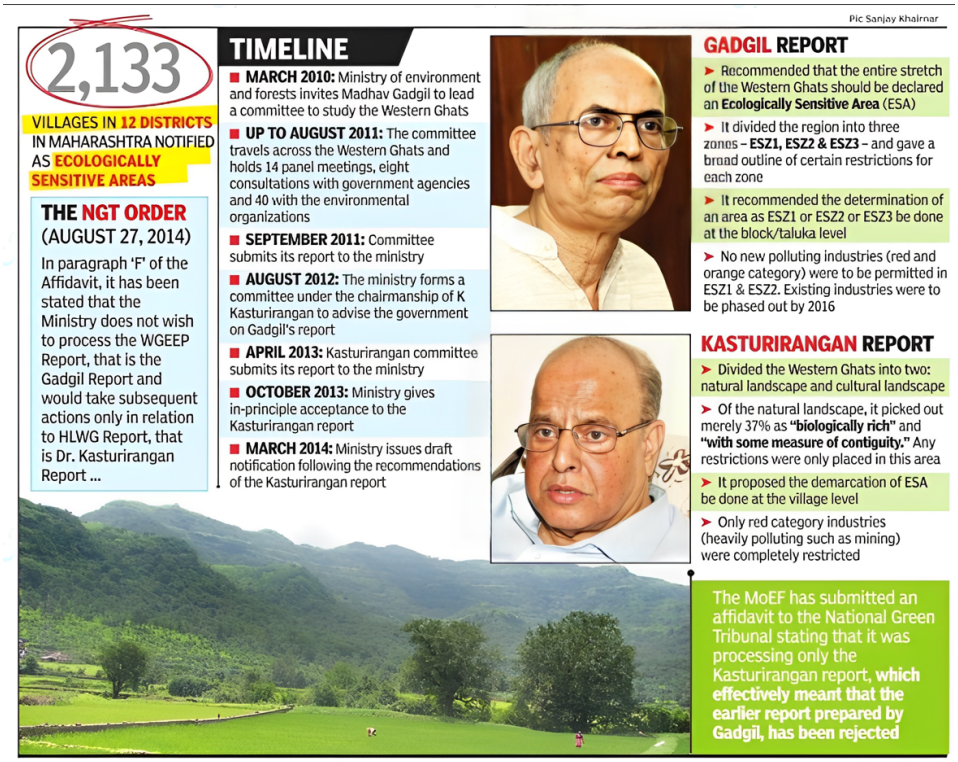
Conservation Measures: Both reports recommend conservation measures such as forest protection and mining and quarrying regulation. However, the Kasturirangan Report contains more specific recommendations regarding the promotion of sustainable agriculture practises, renewable energy sources, and environmental governance.

The **Madhav Gadgil Report** proposes a more comprehensive and stringent approach to implementation, with a **four-step process** that includes the designation of ESZs, the development of zonal master plans, notification and regulation, and the participation of local communities. In contrast, the **Kasturirangan Report** proposes a **three-step procedure** consisting of the designation of ESZs, the development of zonal master plans, and the approval and notification of plans.

Gadgil Report	Kasturirangan Report
The entire WG needs to be considered an ESZ	The ESZ has been cleared of all cash crops, farmland, and human habitation.
Identified three distinct types of security measures, and compiled a list of relevant practices.	defined the cultural landscape as different from the natural landscape.

The area recommended for ESZ designation is 137,000 hectares.	An ESZ of 60,000 hectares is warranted.
Eliminate the usage of chemical pesticides and genetically modified crops.	curb what it considered overly interventionist and environmentally destructive actions in the ESZ.
Decommissioning of Hydro Power Projects	There would be a complete ban on mining, any industries classified as "red" (such as thermal power), and any construction larger than 20,000 square meters.
Grasslands are slowly being replaced by trees.	In order to guarantee sufficient river flow and a sufficient amount of space between hydropower projects, the panel imposed strict restrictions.

In conclusion, both reports offer valuable insights and suggestions for the preservation and sustainable development of the Western Ghats region. However, the Madhav Gadgil Report is more comprehensive and decentralized, whereas the Kasturirangan Report is more centralized and provides more specific recommendations for conservation measures.



News clip analysing and summarizing the report

Why did the Gadgil Committee come into existence?

By order dated 2010, the Ministry of Environment and Forests, Government of India, established the Western Ghats Ecology Expert Panel (WGEEP), under the leadership of Prof. Madhav Gadgil, to assess the status of ecology and demarcate areas within the Western Ghats Region for notifying ecologically sensitive zones under the Environment (Protection) Act, 1986.

In addition, it proposes methods for the conservation, preservation, and revitalization of the Western Ghats Region through a comprehensive consultation with all stakeholders.Recommends modalities for the establishment of the Western Ghats Ecology Authority under the Environment (Protection) Act, 1986, a professional body to administer the region's ecology and promote sustainable development.

CONCLUSION

The Gadgil Committee report put an emphasis on the environment and suggested steps to protect it, whereas the Kasturirangan Committee report sought to strike a balance between development and protection. However, both reports have been met with strong opposition from influential groups and ordinary residents.

Some have argued that the Kasturirangan report watered down the provisions of the first study, while others have questioned the feasibility of the first report due to worries about people's livelihoods and the impact on small farming communities. Nonetheless, it is true that both stories have attracted public attention and inquiry, luring both nature lovers and business owners. Thus, effective conservation of the Ecologically Sensitive Western Ghats will arise from a frank debate with all stakeholders, addressing concerns of all those affected.

REFERENCES

- ClearIAS. (2014). Gadgil Report and Kasturirangan Report on Western Ghats. [online] Available at: https://www.clearias.com/gadgil-report-kasturirangan-report-western-ghats/.
- The Indian Express. (2021). Ecologist Madhav Gadgil calls Kasturirangan panel report 'faulty and unscientific'. [online] Available at: https://indianexpress.com/article/cities/bangalore/ecologist-madhav-gadgil-calls-kasturirangan-panel-report-faulty-and-unscientific-7662823/ [Accessed 29 Apr. 2023].
- www.slideshare.net. (n.d.). Critical Comparison b/w WGEEP Report & HLWG Report on Western Ghats. [online] Available at: https://www.slideshare.net/A-jay/kasthuri-rangan-vs-madhav-gadgil [Accessed 29 Apr. 2023].
- Desikachar, N. (n.d.). A COMPARITIVE STUDY OF Prof. MADHAV GADGIL AND Dr. KASTURIRANGAN REPORT on THE WESTERN GHATS. www.academia.edu. [online] Available at: https://www.academia.edu/9649437/A_COMPARITIVE_STUDY_OF_Prof_MADHAV_GADGIL_AND_Dr_KASTURIRANGAN_REPORT_on_THE_WESTERN_GHATS [Accessed 29 Apr. 2023].
- Gadgil slams Kasturirangan report on Western Ghats. (2013). The Hindu. [online] 6 Nov. Available at: https://www.thehindu.com/news/cities/Kochi/gadgil-slams-kasturirangan-report-on-western-ghats/article5319939.ece [Accessed 29 Apr. 2023].
- BYJUS. (n.d.). *Gadgil Report & Kasturirangan Report on Western Ghats - Features & Controversies [UPSC Notes]*. [online] Available at: https://byjus.com/free-ias-prep/gadgil-report-kasturirangan-report-western-ghats/ [Accessed 29 Apr. 2023].
- Drishti IAS. (n.d.). *Plea on Western Ghats*. [online] Available at: https://www.drishtiias.com/daily-updates/daily-news-analysis/plea-on-western-ghats [Accessed 29 Apr. 2023].

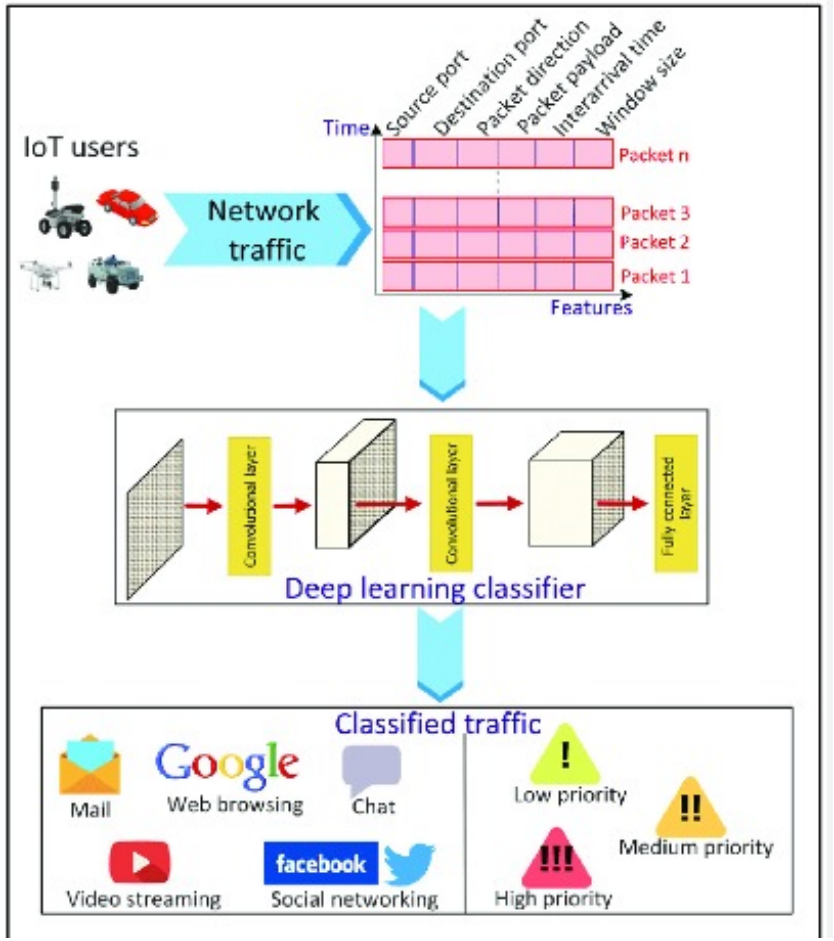
LEXNet(CNN) for Internet traffic classification

Data Mining CS F415

Prof. Kamlesh Tiwari
Prof. Yashvardhan sharma

Madhav Vohra
Dhruv Kamra
Tauqeer Akthar
Namit Srivastava
Ishaan Shankar
Deepanshu Choudhry

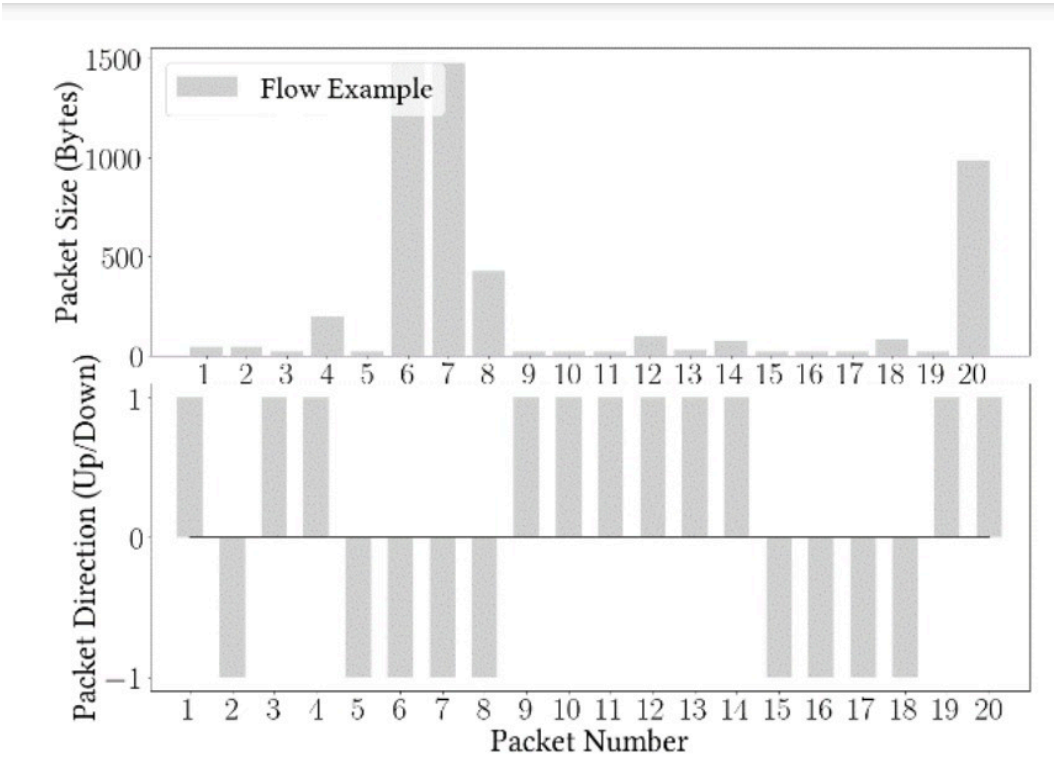
Understanding the data



- Internet network traffic data refers to the information generated by the communication and exchange of digital packets between devices connected to the internet.
- This data encompasses a wide array of activities, ranging from simple web browsing and email communication to more complex processes like file transfers, video streaming, and online gaming.
- A packet is a unit of data routed on the network, and a flow is a set of packets that shares the same 5-tuple (i.e., source IP, source port, destination IP, destination port, and transport-layer protocol

*

Using the data



Input (20*2)

- Classification Task: Sorting internet traffic by analyzing packet-level data like size and direction, treating each flow as a time series.
- From Single to Double: Unlike previous methods, which only considered packet size, this approach adds direction as a variable. This doesn't compromise performance but adds more depth to traffic analysis..
- Rich Explanations: The new method gives a more detailed understanding by including both packet size and direction as factors. Importantly, this doesn't sacrifice efficiency; the system maintains its performance in classifying applications.

*

Current Challenges

- Hardware Constraints Oversight:
  - Challenge: Deep learning approaches often ignore the limited computational resources of networking hardware like routers.
  - Implication: Designs may be impractical for real-world deployment in resource-constrained environments.

- Shortcoming in Explainability:
  - Challenge: Current approaches fall short in providing faithful explainability, a key requirement emphasized by regulatory bodies.
  - Issue: Despite attempts to address explainability using methods like SHAP, these may not meet the required level of faithfulness. This can pose challenges in understanding and justifying the decisions made by the traffic classifiers.

- Dataset Limitations:
  - Challenge: Evaluation on small, less diverse datasets hampers generalizability to real-world scenarios.
  - Consequence: Models may not accurately reflect the complexity and variety of applications in actual commercial networks.

Why LEXNet

- Efficiency and Size:
  - Advantage: LEXNet is designed to be lightweight and efficient, meaning it operates faster and with a smaller model size compared to other advanced neural networks.
  - Benefit: This makes LEXNet more practical for real-world use, especially in situations where computational resources are limited.
- Explainability Feature:
  - Distinctive Trait: LEXNet introduces a prototype layer for explainability, allowing users to understand why a classification decision was made.
  - Importance: This feature enhances transparency, making it easier for users to trust and comprehend the model's predictions, which is crucial in applications subject to regulatory scrutiny.
- Performance:
  - LEXNet achieves the same high accuracy as the best-performing state-of-the-art neural networks.
  - Advantage: It provides accuracy comparable to top-performing models while offering the benefits of efficiency and explainability, making it a well-rounded choice for internet traffic classification tasks.

Data preprocessing and validation set

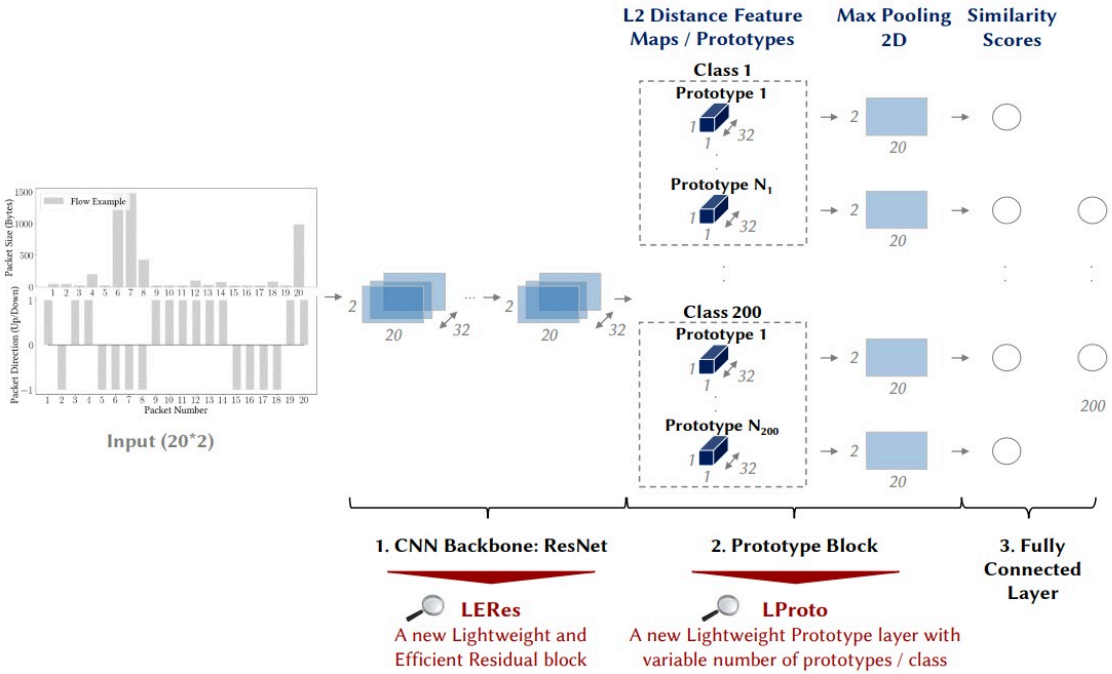
```
# UTS to MTS (format of x: tabular [n_samples, n_packets], features: +/- packet size)
X_train = np.load("./data/" + dataset + "/train_x.npy")
X1 = abs(X_train)
X2 = np.sign(X_train)

if preprocess:
    scaler1 = preprocessing.MinMaxScaler()
    X1 = scaler1.fit_transform(X1)

    scaler2 = preprocessing.MinMaxScaler()
    X2 = scaler2.fit_transform(X2)

X3 = np.concatenate(np.array([X1, X2]), axis=1)
X_train = np.reshape(X3, (X_train.shape[0], 1, X_train.shape[1], 2), order="F")

# Labels encoding
y_train = np.load("./data/" + dataset + "/train_y.npy")
le = preprocessing.LabelEncoder()
y_train = le.fit_transform(y_train)
```

```
# Train/validation split
skf = StratifiedKFold(n_splits=5, shuffle=False)
train_index, val_index = list(skf.split(X_train, y_train))[val_split[1] - 1]
X_train, X_val = X_train[train_index], X_train[val_index]
y_train, y_val = y_train[train_index], y_train[val_index]
np.save(xp_dir + "train_index.npy", train_index)
np.save(xp_dir + "validation_index.npy", val_index)
```

The validation set is essential to assess the performance and generalization ability of a machine learning model on unseen data, helping to prevent overfitting

*

LEXNet Structure



- **CNN Backbone** (LERes Block):
- Innovation: LEXNet revamps the widely used residual block with cost-efficient operations, reducing parameters (-19%) and CPU inference time (-41%) while maintaining network accuracy (-0.7%). This redesign, named LERes block, enhances efficiency.
- **Prototype Layer** (LProto):
- Transformation: LEXNet transforms the state-of-the-art prototype block into a unique prototype layer (LProto). This modification significantly cuts parameters (-36%) and CPU inference time (-24%), with a 4% accuracy boost. Unlike the original block, LProto dynamically learns the number of prototypes per class during training for improved feature characterizationCorporate Social Responsibility / Stakeholders

*

LERes(CNN backbone)

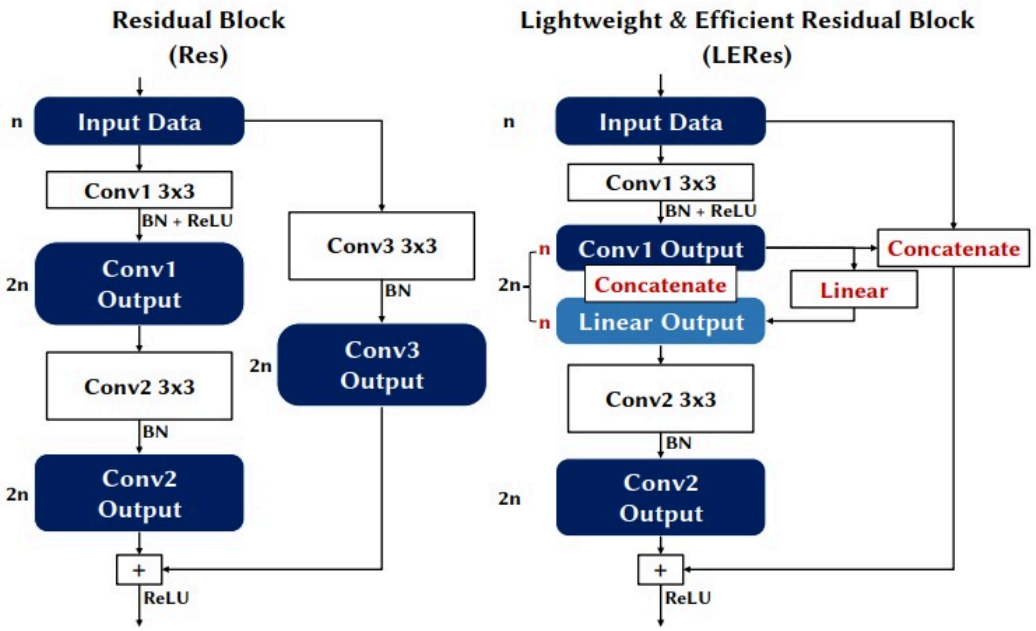


Figure 2: ResNet original residual block (Res) and our new Lightweight and Efficient Residual block (LERes) with its contributions in red.

1. Channel Width Equality:

Enhancement: LERes Block ensures equal channel width in convolutions, departing from ResNet's original design with varying channel widths. This modification simplifies memory access and improves efficiency.

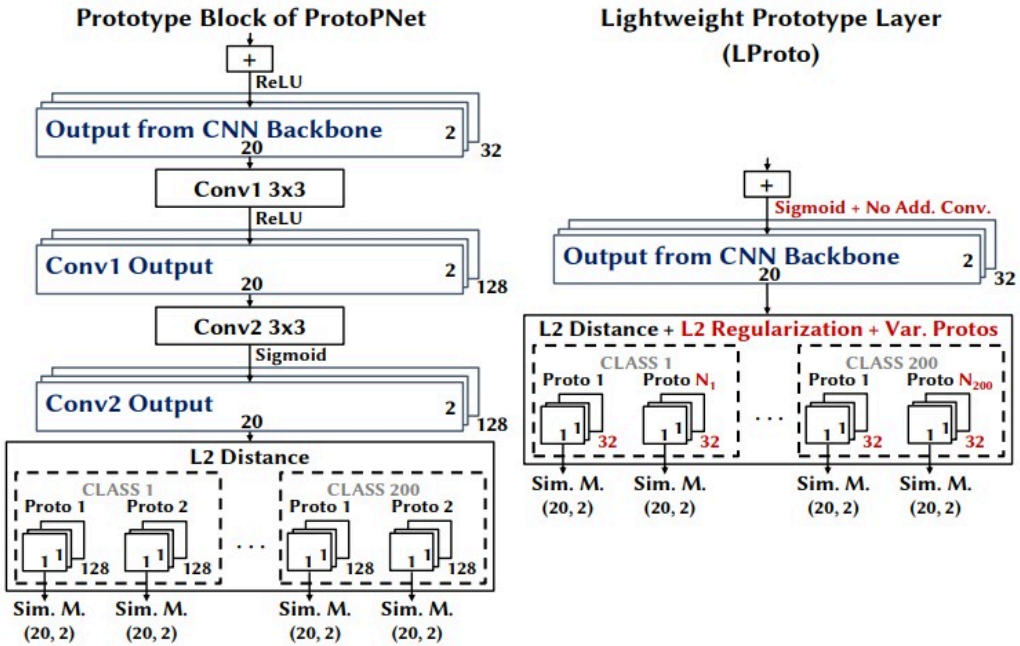
2. Cost-Efficient Feature Expansion:

Innovation: LERes Block employs a cost-efficient method to double feature maps, involving linear transformations of the output from the first convolution. This approach maintains prediction performance while reducing redundancy in feature maps.

3. Shortcut Link Optimization:

Streamlined Connectivity: LERes Block optimizes the shortcut link by avoiding a third convolution to match input and output dimensions. Instead, it retains the input data, concatenating it with the output from the first convolution. This simplification reduces computational load while preserving accuracy in traffic classification tasks.

LProto



1. Reduced Prototype Depth:

Modification: LProto Block eliminates two additional convolutional layers present in the original ProtoPNet prototype block, resulting in significantly shallower prototypes (32 depth versus 128).

Advantage: This reduction in depth not only decreases the number of weights but also contributes to a more efficient computation

2. Activation Function Replacement:

Optimization: LProto Block replaces the last activation function of the CNN backbone with a Sigmoid, based on findings from experiments showing improved accuracy.

Impact: This adjustment enhances the network's predictive capabilities and aligns with best practices for the final layer activation function.

3. Variable Prototypes per Class:

Innovation: LProto Block introduces the capability for LEXNet to learn a variable number of prototypes per class during training, unlike ProtoPNet, which requires a fixed number set as a hyperparameter.

Significance: This dynamic feature allows LEXNet to tailor the representation of each class with an appropriate number of prototypes, providing more informative insights to end-users while optimizing the total number of prototypes.

*

LEXNet Algorithm and Final structure

Algorithm 1 LEXNet

Initialize: $w_{back} \leftarrow$ Kaiming uniform initialization; $\forall j$: prototype $\mathbf{p}_j \leftarrow$ Uniform$([0,1]^{H \times W \times D})$; $\forall k, j : w_h^{(k,j)} \leftarrow 1$ if $\mathbf{p}_j \in \mathbf{P}_k$, $w_h^{(k,j)} \leftarrow -0.5$ if $\mathbf{p}_j \notin \mathbf{P}_k$

for epoch $t = 1, ..., N_{epochs}$ **do**
  /* Stage 1: SGD of layers before the last */
  for SGD training epoch $t' = t+1, ..., t+N_{SGD}$ **do**
    if $t' < 5$ **then** /* Warm-up of the backbone */
      $w_{back} \leftarrow$ update weights of the backbone;
    $w_{back} \leftarrow$ update weights of the backbone;
    $\mathbf{P} \leftarrow$ update weights of the prototypes;
  $t \leftarrow t + N_{SGD}$
  /* Stage 2: update of the prototypes */
  for each prototype $\mathbf{p}_j$ **do**
    $\mathbf{p}_j \leftarrow$ project prototype onto the nearest latent training patch from the same class;
  $dists \leftarrow$ min distance of each sample to the prototypes of its class;
  $avg\_dists \leftarrow$ average $dist$ per class across all samples;
  if kurtosis$(avg_dists) > 0$ **then**
    $classes \leftarrow$ classes in 25-th percentile of avg_dists;
    for each k in $classes$ **do**
      $\mathbf{P}_k \leftarrow$ add 1 prototype;
  /* Stage 3: training of the last layer */
  for training epoch $t'' = 1, ..., N_{last}$ **do**
    $w_h \leftarrow$ update weights of the last layer

Table 1: Overall architecture of LEXNet.

Input Dims	Operator	Stride	Out Channels	Cum. Params
1×20×2	Conv3x3+BN	1	8	88
8×20×2	LERes Block	1	16	3,088
16×20×2	LERes Block	1	16	7,760
16×20×2	LERes Block	1	32	19,520
32×20×2	LERes Block	1	32	38,080
32×20×2	LProto Layer	-	340	50,880
340×20×2	Max Pooling	-	340	50,880
340×1×1	FC	-	200	118,880

*

```
# Calculate cluster cost
prototypes_of_correct_class = torch.t(
    model.module.prototype_class_identity[:, target.cpu()])
# ).cuda()
inverted_distances, _ = torch.max(
    (max_dist - min_distances) * prototypes_of_correct_class, dim=1
)
cluster_cost = torch.mean(max_dist - inverted_distances)

# Calculate separation cost
prototypes_of_wrong_class = 1 - prototypes_of_correct_class
inverted_distances_to_nontarget_prototypes, _ = torch.max(
    (max_dist - min_distances) * prototypes_of_wrong_class, dim=1
)
separation_cost = torch.mean(
    max_dist - inverted_distances_to_nontarget_prototypes
)
```

```
# Calculate avg cluster cost
avg_separation_cost = torch.sum(
    min_distances * prototypes_of_wrong_class, dim=1
) / torch.sum(prototypes_of_wrong_class, dim=1)
avg_separation_cost = torch.mean(avg_separation_cost)

# Prototypes L2 regularization
l21 = model.module.prototype_vectors.norm(p=2)

# Regularization
if use_l_mask:
    # l2_mask = 1 - torch.t(model.module.prototype_class_identity).cuda()
    l2_mask = 1 - torch.t(model.module.prototype_class_identity)
    l22 = (model.module.last_layer.weight * l2_mask).norm(p=2)
else:
    l22 = model.module.last_layer.weight.norm(p=2)
```

```
super(LEResidualBlock, self).__init__()
self.conv1 = conv3x3(in_channels, in_channels)
self.linear1 = linear3x3(in_channels, in_channels)
self.bn1 = nn.BatchNorm2d(in_channels)
self.relu = nn.ReLU(inplace=True)
self.conv2 = conv3x3(out_channels, out_channels)
self.bn2 = nn.BatchNorm2d(out_channels)
self.linear_transformation = linear_transformation
if baseline_activation_function == "sigmoid":
    self.last_activation = nn.Sigmoid()
else:
    self.last_activation = nn.ReLU()
```

```
def forward(self, x):
    residual = x
    out = self.conv1(x)
    out = self.bn1(out)
    out = self.relu(out)
    if self.linear_transformation:
        residual = torch.cat([residual, out], dim=1)
        x1 = self.linear1(out)
        out = torch.cat([out, x1], dim=1)
    out = self.conv2(out)
    out = self.bn2(out)
    out += residual
    out = self.last_activation(out)
    return out
```

Results on MNIST Dataset

```
        Accuracy:                63.65833333333333%
        p dist pair:     6.038164138793945
training set size: 38400
validation set size: 9600
test set size: 12000
number of classes: 10
batch size: 1024
  Dataset Architecture          Train_Size  ... Accuracy_Train Accuracy_Validation  Accuracy_Test
0   MNIST       lexnet  (38400, 1, 784, 2)  ...       0.663359            0.637396       0.636583
```

Thank you.

Birla Institute of Technology & Science, Pilani

Pilani | Dubai | Goa | Hyderabad

TRANSCRIPT

ID NO 2020A2PS1767P

NAME NAMIT SHRIVASTAVA

Degree(s) Working for:
B.E.(Civil)

	COURSE NO		COURSE TITLE	UNITS	GRADE	
FIRST SEMESTER 2020-2021	BIO	F110	BIOLOGY LABORATORY	1	A	
	BITS	F110	ENGINEERING GRAPHICS	2	C	
	BITS	F111	THERMODYNAMICS	3	C-	
	CHEM	F110	CHEMISTRY LABORATORY	1	A	
	CHEM	F111	GENERAL CHEMISTRY	3	C	
	EEE	F111	ELECTRICAL SCIENCES	3	C-	
	MATH	F111	MATHEMATICS I	3	C-	
	PHY	F110	PHYSICS LABORATORY	1	B	
	PHY	F111	MECH OSCILLATIONS & WAVE	3	C	
				CGPA...6.05		
SECOND SEMESTER 2020-2021	BIO	F111	GENERAL BIOLOGY	3	B	
	BITS	F112	TECHNICAL REPORT WRITING	2	A	
	CS	F111	COMPUTER PROGRAMMING	4	C	
	MATH	F112	MATHEMATICS II	3	C-	
	MATH	F113	PROBABILITY & STATISTICS	3	C	
	ME	F112	WORKSHOP PRACTICE	2	B	
				CGPA...6.43		
FIRST SEMESTER 2021-2022	CE	F211	MECHANICS OF SOLIDS	3	B	
	CE	F213	SURVEYING	4	B	
	CE	F230	CIVIL ENGINEERING MATERIALS	4	A-	
	CE	F231	FLUID MECHANICS	3	B-	
	GS	F221	BUSINESS COMMUNICATION	3	B	HEL
	HSS	F325	CINEMATIC ADAPTATION	3	A-	HEL
	MATH	F211	MATHEMATICS III	3	C	
				CGPA...7.00		

(Continued..)

REGISTRAR

TRANSCRIPT

ID NO 2020A2PS1767P

NAME NAMIT SHRIVASTAVA

	COURSE NO		COURSE TITLE	UNITS	GRADE	
SECOND SEMESTER 2021-2022	BITS	F225	ENVIRONMENTAL STUDIES	3	GD	
	CE	F241	ANALYSIS OF STRUCTURES	3	D	
	CE	F242	CONSTRUCTION PLAN & TECH	3	B	
	CE	F243	SOIL MECHANICS	4	C	
	CE	F244	HIGHWAY ENGINEERING	4	B	
	CE	F266	STUDY PROJECT	3	A	DEL
	MGTS	F211	PRINCIPLES OF MANAGEMENT	3	B	
				CGPA...7.08		
SUMMER TERM 2021-2022	BITS	F221	PRACTICE SCHOOL I	5	A	
				CGPA...7.25		
FIRST SEMESTER 2022-2023	BITS	F464	MACHINE LEARNING	3	B	EL
	CE	F312	HYDRAULIC ENGINEERING	4	B	
	CE	F313	FOUNDATION ENGINEERING	3	B	
	CE	F320	DESIGN OF RE CONCRETE STRU	3	C	
	CE	F434	ENV IMPACT ASSESSMENT	3	B	DEL
	CS	F407	ARTIFICIAL INTELLIGENCE	3	C	EL
	HSS	F328	HUMAN RESOURCE DEVELOP	3	A	HEL
	MATH	F432	APPLIED STATISTICAL METHODS	3	B	EL
				CGPA...7.36		
SECOND SEMESTER 2022-2023	CE	F321	ENGINEERING HYDROLOGY	3	B	
	CE	F323	INTRO TO ENVIRN ENGG	3	B	DEL
	CE	F342	WATER & WASTEWATER TREAT	4	A	
	CE	F343	DES OF STEEL STRUCTURES	3	A-	
	CS	F213	OBJECT ORIENTED PROG	4	C	EL
	CS	F320	FOUNDATIONS OF DATA SCIENCE	3	A-	EL
	CS	F372	OPERATING SYSTEMS	3	C	EL
				CGPA...7.47		

(Continued..)

REGISTRAR

Birla Institute of Technology & Science, Pilani

Pilani | Dubai | Goa | Hyderabad

TRANSCRIPT

ID NO 2020A2PS1767P

NAME NAMIT SHRIVASTAVA

	COURSE NO		COURSE TITLE	UNITS	GRADE	
FIRST SEMESTER 2023-2024	CE	F366	LABORATORY PROJECT	3	A	DEL
	CE	F491	SPECIAL PROJECT	3	A	EL
	CS	F415	DATA MINING	3	B	EL
	CS	F441	SEL TOPICS FROM COMP SC	3	A	EL

CGPA...7.64

(Continued..)

REGISTRAR

Birla Institute of Technology & Science, Pilani

Pilani | Dubai | Goa | Hyderabad

TRANSCRIPT

ID NO 2020A2PS1767P

NAME NAMIT SHRIVASTAVA

SUMMARY

Admitted in: FIRST SEMESTER 2020-2021
To: B.E.(Civil)(With Practice School)

Last Registered: SECOND SEMESTER 2023-2024

*Non-Letter Grade 'CLR' indicates that the course is cleared to fulfill graduation requirements. This grade was not used to compute CGPA and was awarded on meeting the minimum criteria set for clearing the said course. This was a one-time provision in view of the COVID-19 pandemic crisis.

Degree(s) Working for:

B.E.(Civil)

Date of Issue: 15-Feb-24

REGISTRAR

Birla Institute of Technology & Science, Pilani

Pilani | Dubai | Goa | Hyderabad

1. This transcript contains complete record of academic performance of the student given in a chronological order. For details the Academic Regulations as well as the Bulletin of the institute should be consulted.
2. The medium of instruction is English.
3. The academic year consists of two semesters and a summer term when required. The courses and the grades along with units are shown against each semester/term in which the student registered.
4. **Evaluation:**
   (a) The performance in most courses is spelt out in terms of letter grades A, A-, B, B-, C, C-, D, E. Each letter grade has a qualitative meaning and a grade point value as given below:

Letter Grade *	A	A-	B	B-	C	C-	D	E
Qualitative Meaning	Excellent	Very Good	Good	Above Average	Fair / Average	Below Average	Poor	Exposed
Grade Point	10	9	8	7	6	5	4	2

   **Note: Among the letter grades mentioned above, the grades A-, B- and C- were introduced with effect from First Semester 2011-2012.*

   (b) In some courses, descriptive non-letter grades (Excellent; Good; Fair; Poor; Acceptable; Unacceptable; Satisfactory(S); Unsatisfactory(U); Above Average; Average; Below Average; Outstanding; Very Good; Continuing) are awarded which carry no grade point.
5. **CGPA:**
   The up-to date overall performance is reported by the Cumulative Grade Point Average (CGPA), which is a weighted average calculated as below:

$$CGPA = (u_1g_1 + u_2g_2 + u_3g_3 + \ldots\ldots) / (u_1 + u_2 + u_3 + \ldots\ldots)$$

   where u_1, u_2, u_3,... denote units associated with the courses taken by the student and g_1, g_2, g_3,.... denote grade points of the letter grades awarded in the respective courses. Whenever a student repeats a course and gets a new letter grade the new grade replaces the earlier grade in the calculation of the CGPA.
6. The other symbols & Reports used in the transcript are:

AU	-	Audit	XR	-	Previous grade 'X' repeated
DP	-	Discontinued from the Programme	RC	-	Registration Cancelled
EL	-	Elective	RRA	-	Required to Register Again
GA	-	Grade Awaited	S	-	Satisfactory
I	-	Incomplete	U	-	Unsatisfactory
NC	-	Not Cleared	W	-	Withdrawn
TGA	-	Thesis Grade Awaited	NA	-	Not Applicable
DEL	-	Discipline Elective	HEL	-	Humanities Elective

7. Optional elective (OE) is not a required component but if it is taken, the letter grade obtained is included in the CGPA. If the student is reported as NC in any one of the courses taken in this category, no further action is necessary.
8. **Flexibilities:**
   The system permits many academic and other flexibilities like: (a) Admission with marginal deficiency. (Additional remedial courses are prescribed, and when necessary programme duration is extended); (b) Admission with advanced standing (The courses for which exemption was given are listed in the transcript; such courses are not included in the CGPA. The CGPA and division, if applicable are awarded on the basis of only courses taken in the institute); (c) Transfer from one programme to another, before the completion of the first. (The accumulated units and CGPA in the previous programme are carried over as the input to the new programme); (d) working concurrently for two degrees out to the integrated first degree programmes (dual degree scheme); The requirements for both the degrees in the dual degree scheme are concurrently met. Consequently, the CGPA and the division awarded for the two degrees would be the same (e) Each programme in the first degree level or higher degree level offers a choice between Practice School and Thesis/Dissertation streams.
9. **Academic Counselling:**
   (a) The educational philosophy interlinks and at the same time distinguishes between the performance of a student in a single course and his overall cumulative performance. His progress and performance is monitored at the end of every semester/term by noting whether (i) he has secured more than one E grade in that semester/terms; (ii) he has obtained a CGPA less than 4.50 in case of integrated first degree programme and less than 5.50 in case of higher degree programme upto that semester/term; (iii) he has spent more than 50% extra time than what is prescribed for him up to that semester/terms in his programme.
   (b) Whenever a student's performance comes under the clauses (i), (ii) or (iii) in (a) above, the student comes under the purview of Academic Counselling Board (ACB) which would counsel him, temporarily restrict his options, may require him to transfer to another suitable programme or leave the institute if he is unable to meet the probationary conditions laid down by it.
10. **Eligibility Requirements:**
    A student has to obtain a minimum CGPA of 4.50 in case of integrated first degree programmes, a minimum CGPA of 5.50 in case of higher degree programmes and a minimum CGPA of 5.50 wherever applicable, in case of Ph.D. programmes to be eligible for the degree.
11. A long gap between the last semester of registration and the semester in which the student becomes eligible for the degree, may be due to delayed award of a grade because of late submission of some components of evaluation in a course/dissertation/thesis by the student.
12. **Division:** It is a classification based on CGPA as follows:

Distinction	:	CGPA 9.00 or more
I Division	:	CGPA 7.00 or more but less than 9.00
II Division	:	CGPA 4.50 or more but less than 7.00

    (No division is awarded for diploma, higher degrees or Ph.D. programmes)