ISSN : 2349-6495(P) | 2456-1908 (O)

International Journal of Advanced Engineering Research and Science

(IJAERS)

An Open Access Peer-Reviewed International Journal

Journal DOI: 10.22161/ijaers

Issue DOI: 10.22161/ijaers.113

AI PUBLICATIONS

Vol.- 11 | Issue - 3 | Mar 2024

editor.ijaers@gmail.com | editor@ijaers.com | https://www.ijaers.com/

International Journal of Advanced Engineering Research and Science (IJAERS)

(ISSN: 2349-6495(P)| 2456-1908(O))

DOI: 10.22161/ijaers

Vol-11, Issue-3

March, 2024

Publisher

AI Publication

Email: editor.ijaers@gmail.com; editor@ijaers.com

Web: www.ijaers.com

International Editorial/ Reviewer Board

Editor in Chief

- **Dr. Swapnesh Taterh (Chief-Editor),** Amity University, Jaipur, India

Cheif Executive Editor

- **S. Suman Rajest,** Vels Institute of Science, Technology & Advanced Studies, India chief-executive-editor@ijaers.com

Associate Editors

- **Dr. Ram Karan Singh,** King Khalid University, Guraiger, Abha 62529, Saudi Arabia
- **Dr. Shuai Li,** University of Cambridge, England, Great Britain

Editorial Member

- **Behrouz Takabi,** PhD, Texas A&M University, Texas, USA
- **Dr. Gamal Abd El-Nasser Ahmed Mohamed Said,** Port Training Institute (PTI), Arab Academy For Science, Technology and Maritime Transport, Egypt
- **Dr. Hou, Cheng-I ,** Chung Hua University, Hsinchu Taiwan
- **Dr. Ebrahim Nohani,** Islamic Azad University, Dezful, IRAN.
- **Dr. Ahmadad Nabih Zaki Rashed,** Menoufia University, EGYPT
- **Dr. Rabindra Kayastha ,** Kathmandu University, Nepal
- **Dr. Dinh Tran Ngoc Huy,** Banking and Finance, HCM, Viet Nam
- **Dr. Engin NAS,** Duzce University, Turkey
- **Dr. A. Heidari,** California South University (CSU), Irvine, California, USA
- **Dr. Uma Choudhary,** Mody University, Lakshmangarh, India
- **Dr. Varun Gupta,** National Informatic Center, Delhi, India
- **Dr. Ahmed Kadhim Hussein,** University of Babylon, Republic of Iraq
- **Dr. Vibhash Yadav,** Rajkiya Engineering College, Banda. UP, India
- **Dr. M. Kannan,** SCSVMV University, Kanchipuram, Tamil Nadu, India
- **José G. Vargas-Hernández,** University of Guadalajara Periférico Norte 799 Edif. G201-7, Núcleo Universitario Los Belenes, Zapopan, Jalisco, 45100, México
- **Dr. Sambit Kumar Mishra,** Gandhi Institute for Education and Technology, Baniatangi, Bhubaneswar, India
- **DR. C. M. Velu,** Datta Kala Group of Institutions, Pune, India
- **Dr. Deependra Pandey,** Amity University, Uttar Pradesh, India
- **Dr. K Ashok Reddy,** MLR Institute of Technology, Dundigal, Hyderabad, India
- **Dr. S.R.Boselin Prabhu,** SVS College of Engineering, Coimbatore, India
- **N. Balakumar,** Tamilnadu College of Engineering, Karumathampatti, Coimbatore, India
- **R. Poorvadevi,** SCSVMV University, Enathur, Kanchipuram, Tamil Nadu, India
- **Dr. Subha Ganguly,** Arawali Veterinary College, Sikar, India
- **Dr. P. Murali Krishna Prasad,** GVP College of Engineering for Women, Visakhapatnam, Andhra Pradesh, India
- **Anshul Singhal,** Bio Instrumentation Lab, MIT, USA
- **Mr. Lusekelo Kibona,** Ruaha Catholic University, Iringa, Tanzania
- **Sina Mahdavi,** Urmia Graduate Institute, Urmia, Iran
- **Dr. N. S. Mohan,** Manipal Institute of Technology, Manipal, India
- **Dr. Zafer Omer Ozdemir,** University of Health Sciences,Haydarpasa, Uskudar, Istanbul, TURKIYE
- **Bingxu Wang,** 2721 Patrick Henry St Apt 510, Auburn Hills, Michigan, United States

- **Dr. Jayashree Patil-Dake,** KPB Hinduja College of Commerce, Mumbai, India
- **Dr. Neel Kamal Purohit,** S.S. Jain Subodh P.G. College, Rambagh, Jaipur, India
- **Mohd Muntjir,** Taif University, Kingdom of Saudi Arabia
- **Xian Ming Meng,** China Automotive Technology & Research Center No.68, East Xianfeng Road, Dongli District, Tianjin, China
- **Herlandi de Souza Andrade,** FATEC Guaratingueta, State Center for Technological Education Paula Souza - CEETEPS
- **Dr. Payal Chadha,** University of Maryland University College Europe, Kuwait
- **Ahmed Moustafa Abd El-hamid Elmahalawy,** Menoufia University, Al Minufya, Egypt
- **Prof. Mark H. Rummeli,** University & Head of the characterisation center, Soochow Institute for Energy Materials Innovasions (SIEMES), Suzhou, Jiangsu Province, China
- **Dr. Eman Yaser Daraghmi,** Ptuk, Tulkarm, Palestine
- **Holmes Rajagukguk,** State University of Medan, Lecturer in Sisingamangaraja University North Tapanuli, Indonesia
- **Dr. Menderes KAM,** Dr. Engin PAK Cumayeri Vocational School, DÜZCE UNIVERSITY (University in Turkey), Turkey
- **Dr. Jatin Goyal,** Punjabi University, Patiala, Punjab, India | International Collaborator of GEITEC / UNIR / CNPq, Brazil
- **Ahmet İPEKÇİ,** Dr. Engin PAK Cumayeri Vocational School, DÜZCE UNIVERSITY, Turkey
- **Baarimah Abdullah Omar,** Universiti Malaysia Pahang (UMP), Gambang, 26300, Malaysia
- **Sabri UZUNER,** Dr. Engin PAK Cumayeri Vocational School Cumayeri/Duzce/Turkey
- **Ümit AĞBULUT,** Düzce University, Turkey
- **Dr. Mustafa ÖZKAN,** Trakya University, Edirne/ TURKEY
- **Dr. Indrani Bhattacharyya,** Dr. B.C. Roy College of Pharmacy and Allied Health Sciences, Durgapur, West Bengal, India
- **Egnon Kouakouc,** Nutrition/Health at University Felix Houphouet Boigny Abidjan, Ivory Coast
- **Dr. Suat SARIDEMİR,** Düzce University, Faculty of Technology, Turkey
- **Dr. Manvinder Singh Pahwa,** Director, Alumni Relations at Manipal University Jaipur, India
- **Omid Habibzadeh Bigdarvish,** University of Texas at Arlington, Texas, USA
- **Professor Dr. Ho Soon Min,** INTI International University, Jln BBN 12/1, Bandar, Baru Nilai, 71800 Negeri Sembilan, Malaysia
- **Ahmed Mohammed Morsy Hassan,** South Egypt Cancer Institute, Assiut University, Assiut, Egypt
- **Xian Ming Meng (Ph.D),** China Automotive Technology & Research Center, No.68, East Xianfeng Road, Tianjin, China
- **Ömer Erkan,** Konuralp Campus, Düzce-Turkey
- **Dr. Yousef Daradkeh,** Prince Sattam bin Abdulaziz University) PSAU), KSA
- **Peter JO,** IPB University, Indonesia
- **Nazmi Liana Binti Azmi,** Raja Perempuan Zainab II Hospital, 15586 Kota Bharu, Kelantan, Malaysia
- **Mr. Sagar Jamle,** Oriental University, Indore, India
- **Professor Grazione de Souza,** Applied Mathematics, Rio de Janeiro State University, Brazil
- **Kim Edward S. Santos,** Nueva Ecija University of Science and Technology, Philippines.

Detail with DOI (CrossRef)

International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-11, Issue-3; Mar, 2024
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.113.1

Analysis of mainstay discharge in the Way Batumerah Watershed for annual water allocation plans

Godfried Lewakabessy, Obednego D. Nara

Department of Civil Engineering Ambon State Polytechnic

Received: 15 Jan 2024,

Receive in revised form: 01 Mar 2024,

Accepted: 08 Mar 2024,

Available online: 15 Mar 2024

***Keywords*— Raw Water, Water Allocation, Mainstay Discharge, Maintenance Flow.**

***Abstract*—** *This water shortage has caused drought disasters whose impact is felt both in rural areas in decreasing food production and in urban areas experiencing raw water difficulties. In the implementation of the "Water Allocation Preparation" activity, several technical approaches will be carried out including: Data Collection; Data Analysis; Prioritization of water users; Calculation of water balance; Calculation of water allocation plan; and Implementation, Monitoring and Evaluation of the Implementation of the Annual Water Allocation Plan. In the watershed of Way Batumerah, the intake area has a watershed area of 6.97 Km^2 where the Batumerah Intake; Sedimentation occurs in the distribution pipe; There are two villages that are drained by intakes; There are community channels around the intake; and During the dry season, water is not enough to meet the needs of the community.The availability of water in the Batumerah River can be calculated in a period of half a month. In wet year conditions, the largest mainstay discharge (Q30%) of 1.83 m^3 /s occurred in October II. In normal year conditions, the largest mainstay discharge (Q50%) was obtained at 0.89 m^3/s in September II. In dry years, the largest mainstay discharge (Q80%) was obtained at 0.35 m^3/s in June I. The calculation of maintenance flow needs must follow KP 02/2013 which is 5% of the existing discharge in the relevant period or Q95 according to SE Director General of Natural Resources 05/2016.The water allocation unit and allocation limit for PDAMs in the Batumerah watershed are assumed to be in accordance with the value of PDAM's water needs for the dry condition half-month, which is 0.105 m^3/s. However, in certain months there is a deficit, so PDAM's water demand is the result of reducing water availability with minimum maintenance flow needs (0.001 m^3/s). Water use priority is a policy that regulates the order of water allocation based on applicable laws and regulations.*

I. INTRODUCTION

Water is the source of life for humans and other living things. The nature of water is very different compared to other resources, because water is a flowing resource, knows no administrative boundaries, and its needs depend on time, space, quantity and quality. Water problems can be broadly divided into three categories, namely: 1) Too much water, generally occurs in the rainy season, and often causes floods; 2) Too dirty water, that is, water pollution, occurs mostly due to industrial, household and agricultural waste; and 3) Too little water. This water shortage has caused drought disasters whose impact is felt both in rural areas in decreasing food production and in urban areas experiencing raw water difficulties. This also creates conflicts of interest between water users, as has often happened in the dry

season. With the increasing population and the development of socio-economic life of the community, it will also cause an increase in the need for water. This will cause water to increasingly become a scarce item at certain times for some locations that are prone to water shortages. This conflict of interest for water, which was originally only between individuals or community groups using water, with the spirit of decentralization and regional autonomy is feared to have the potential to escalate local conflicts into unwanted conflicts between districts / cities. Water allocation is a series of actions, which include: actions to regulate water rations/quotas in accordance with the type of water use, efforts to always be able to meet the amount and quality of water in accordance with the rights guaranteed by the state. The basic needs of daily life and people's agricultural business in the irrigation system are the top priority, as well as the rights of other water users obtained based on water use permits, where the amount / volume of water that can be taken from a network of water sources for business purposes has been determined. By allocating water, it is hoped that it can prevent violations that can result in interference with the human rights of other people or parties.

II. LITERATURE REVIEW

2.1 Water Allocation System

Effective water allocation requires an approach that can be applied to three key areas: 1) policies and regulations, 2) management strategies, and 3) institutional capacity. Policies and regulations constitute the legal framework for water allocation. Management strategies include plans, tools, methodologies, processes, and approaches to implement legal and policy frameworks. While institutional capacity is the ability of organizations and stakeholders in shaping policies and strategies (WWF, 2007). These three elements determine the functioning of the water allocation system and the protection of related resources in the watershed. If one of the three elements is inconsistent with the other two, then the implementation of the system will be dangerous. Without a policy level, there will be no clear legal understanding of water use. Without a good management strategy, the implementation of laws to achieve sustainable water allocation will experience obstacles. Without institutional capabilities, the implementation of water allocation will be weak. To be able to implement a fair and sustainable allocation of water requires intervention from one or more of the above aspects.

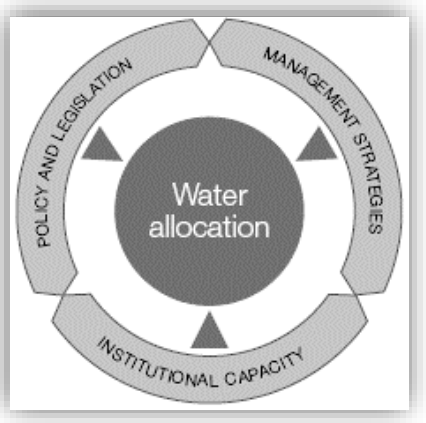


Fig.1 Water Allocation System according to WWF (2007)

2.2 Basic principles of Water Allocation

Water allocation is the process of rationing water for various types of uses whose amount is adjusted to the availability of water contained in a source / location, and in accordance with Law No.11/1974, article 11 paragraph 1. In accordance with the development of water allocation management, water allocation planning must be comprehensive from upstream to downstream and based on river basins (WS) or at least on a watershed scale, and with the basic principles of water allocation management must be based on PUPR Regulation No. 06 / PRT / M / 2015, article 9 namely in allocating water and water sources for water resources infrastructure operations must be carried out based on the principles of: a. prioritizing the allocation of water for the fulfillment of basic daily needs and irrigation for people's agriculture in the existing irrigation system; b. maintain the continuity of water allocation for other existing water users; c. pay attention to the allocation of water to meet basic daily needs for residents who live near water sources and / or around water-carrying networks.

Planning water allocation at the watershed level or watershed level is a comprehensive effort that

Based on policies that aim to uphold the following five principles:

1) Justice; i.e. allocating water fairly and proportionately among type groups use, justice between administrative regions, and justice between upstream and downstream regions.
2) Environmental protection; That is to allocate a certain amount of fresh water for needs ecosystem and including to accommodate the needs of sediment transport, water recharge soil, waste decomposition and ecosystem sustainability in estuaries.
3) Development priorities; allocate water to support development needs economic and social, among others,

to support strategic priorities and protect dependencies of existing needs.

4) Balance between water supply and demand; i.e. balancing the water supply with dynamic needs demands, especially to manage variability natural supply of water, and to avoid or prevent frequent water deficits.
5) Unexpected.Promote efficient use of water; i.e. promote continuously to water users so that they are moved and able to take initiatives to make efficiency in water use

2.3 Preparation for Water Allocation

Division of Government Affairs field of public works and spatial planning, sub affairs: Water Resources, then the authority in

The arrangement/management of water allocation activities is as follows:

1. Water sources located in river basins whose management authority is located
2. on Government implemented by the Directorate General of Water Resources of the Ministry Public Works and Public Housing through BBWS/BWS and/or Legal Entities.
3. Water sources located in river areas whose management authority is located in the Provincial Government is carried out by the BPSDA WS concerned.
4. Water sources located in river areas whose management authority is on: The Regency/City Government is implemented by SKPD formed by the Government The District/City concerned

2.4 Water Availability and Demand

Water availability basically consists of three forms, namely rainwater, surface water, and groundwater. The main water source in water allocation management is surface water sources in the form of water in rivers, channels, lakes, and other reservoirs. The use of groundwater is in fact very helpful in meeting the needs of raw water and irrigation water in areas that are difficult to get surface water, but its sustainability needs to be maintained with controlled intake under safe yield. In water allocation management, rainwater contributes to reducing the need for irrigation water, namely in the form of effective rain. In some areas with inadequate surface water quality, rain harvesting is carried out, where rainwater is collected into a source of water for domestic use.

To express water availability using only a number, the number is the average of existing discharge data. This method does not provide information about data variability. Presenting data as 12 numbers that express a monthly average provides more information about data variability in a year, but does not provide information about how reliable debits are. A number that shows the variability of water availability as well as shows how much reliable discharge is the mainstay discharge.

2.5 Mainstay Debit

A mainstay discharge is a discharge that can be relied upon for a certain level of reliability or reliability. For irrigation purposes, mainstay discharge with 80% reliability is usually used as stipulated in the Irrigation Planning Criteria (Directorate General of Irrigation, 1985). This means that with an 80% chance that the discharge that occurs is greater or equal to the discharge, or in other words the irrigation system can fail once in five years. For drinking water purposes and industry, higher reliability is required, which is around 90% to 95%. If river water is used for hydroelectric power generation, it requires very high reliability, which is between 95% to 99% (Goodman, 1984).

2.6 Household and Urban Water Needs

Household and urban water needs (domestic and municipal) are often referred to as raw water if the water has not been treated, and clean water or drinking water if the water has been treated using a Water Treatment Plant. This need is very important to always be met, because failure to meet household and urban water needs can cause disease outbreaks and public unrest. The amount of water demand depends on the population, consumption patterns that are in line with the increase in welfare levels, and the size of the city, or village which can be assumed to depend on the population.

2.7 Industrial Water Requirements

Industrial water requirements are generally constant relative to time. With the increase in industry, so does the need for industrial water. Industrial water demand surveys are needed to determine the average water use in certain types of industries. This index number can then be related to the size of the industry, for example through the number of products produced, or the number of labor. For industries located in an industrial area, a rough estimate of water needs per hectare can be used between 0.5 to 2 liters / s.

2.8 Agricultural Water Requirements

In water allocation management in river basins, irrigation water demand data can be obtained from river basin managers, such as the District/City Irrigation Public Works Office (DPUP), or the Provincial Water Resources Office, or River Basin Centers and Halls, as input for water allocation management. The amount of irrigation water demand in this field can be checked with the help of a computer model to calculate irrigation water needs, based on parameters that influence, including planting patterns and schedules, effective rainfall, percolation, efficiency,

groups, and so on based on KP01 irrigation network planning criteria from the Directorate General of Irrigation (1985). Water requirements in rice fields depend on factors: land preparation, consumptive use, percolation and seepage, water layer change, effective rainfall, and irrigation efficiency. Water needs in this rice field can be expressed in units of mm / day or liter / s / ha.

2.9 Water Requirements for Maintenance Flow

Based on Government Regulation Number 38 of 2011 concerning Rivers, the amount of river maintenance flow is the mainstay discharge of 95%. The amount of river maintenance flow is equivalent to 20 years of dry discharge, a relatively small number, but in river areas with maximum resource utilization is considered to cause conflicts of interest with other water users. The following discussion is about how river maintenance flows in various countries are calculated and applied.

III. METHODOLOGY

In the implementation of the "Water Allocation Preparation" activity, several technical approaches will be carried out including: Systematically the technical approach and activities to be carried out for the work of "Water Allocation Preparation", can be seen in the Work Implementation Flow Chart in **Figure 2**

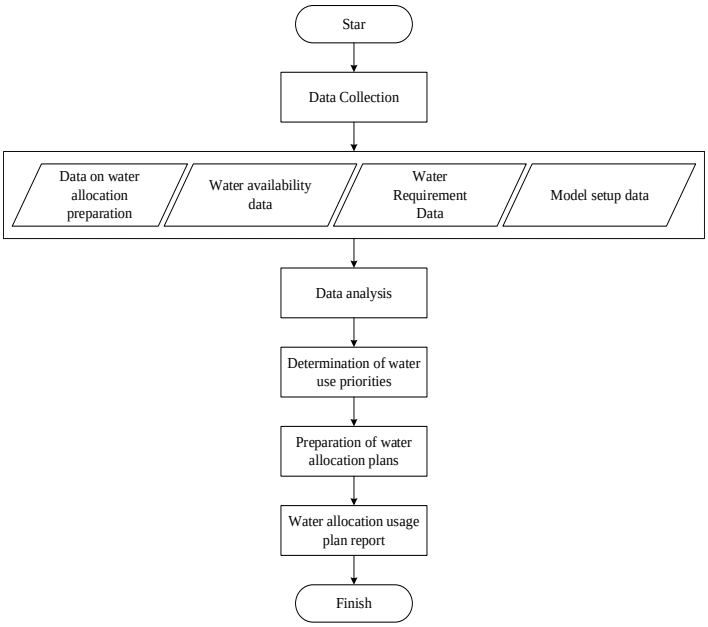


Fig.2 Research Flow Chart

IV. DISCUSSION OF RESULTS

4.1 Ambon-Seram WS Administrative Region

Water resources management in the Ambon-Seram River Basin is carried out within the administrative area of the Maluku Provincial government. Administratively, Ambon Seram River Basin is included in three regencies and one city, namely Central Maluku Regency, West Seram, East Seram and Ambon City.The topographic data used in this study was generated through digital data in the form of a Digital Elevation Map (DEM).

Data taken from SRTM (Shuttle Radar Topography Mission) satellite captures. which can be accessed openly through the USGS (*Unites States Geological Survey*) on https://earthexplorer.usgs.gov/ online site. DEM can be accessed in raster form in GeoTiff format with the accuracy of the data taken is 1 arc second or 30 m. In this study, the object of analysis is in the Way Batumerah watershed.

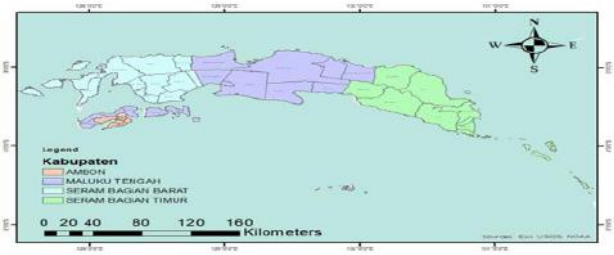


Fig.3 Ambon-Seram WS Administrative Region

4.2 Batumerah Watershed

The Batumerah watershed is located in Ambon City and covers Sirimau District. Geographically, Batumerah watershed is located at 3°42'21.204"S - 3°40'37.425"S and 128°10'57.66"E - 128°13'28.693"E. The total area of the watershed is 6.97 km^2. The Batumerah watershed area is shown on Figure 4.

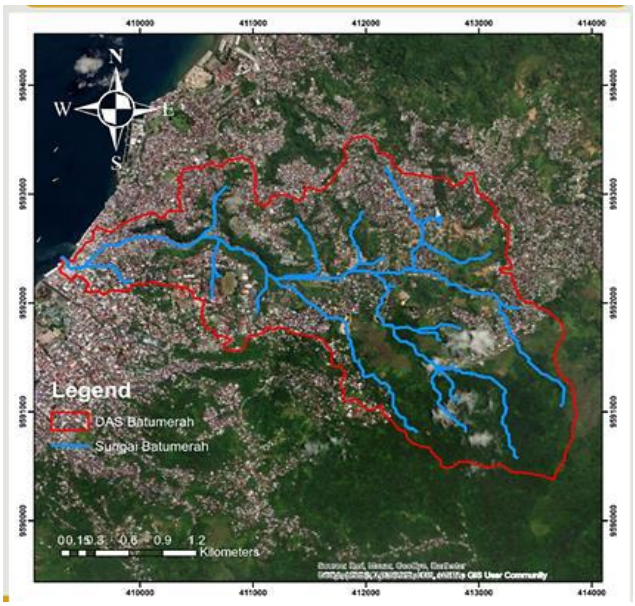


Fig.4 Batumerah watershed

4.3 Water Needs Retrieval Building

Human use of water can basically be divided into water extraction and on-site use. In the Way Batumerah watershed, the intake area has a watershed area of 6.97 km^2 where Batumerah Intake;

- Sedimentation occurs in the distribution pipe;
- There are two villages that are drained by intakes;
- There are community channels around the intake; and
- During the dry season, water is not enough to meet the needs of the community.

4.4 Water balance plan

In hydrological calculations, rain data is needed provided by rainfall posts around the study area. In the Batumerah watershed, there is one nearby rainfall post, namely the Telaga Kodok rainfall post and the IAIN Rainfall Post. Due to the lack of data on the rainfall post network in Maluku Province, the amount of hydrological and climatological data is very limited. Therefore, the data at the rainfall post is filled with satellite rainfall data Global Precipitation Mission (GPM) GPM is a satellite rainfall and snow calculation project by NASA and JAXA. Through GPM Core Observatory will be received from two types of satellites, namely GPM Microwave Imager (GMI) and Dual-frequency Precipitation Radar (DPR).

The availability of water in the Batumerah watershed is a discharge that can be provided by the Batumerah River to meet water needs in the Batumerah watershed. The calculation of water availability is carried out by hydrological calculations where rain will be simulated into surface flow. The rainfall data obtained is rain correction data from GPM satellites with a span of 11 years. Due to limited discharge data, a simulation of rain calculation into surface flow was carried out using the NRECA model. The calculation was carried out using daily rainfall data and climate data that had been obtained previously to calculate surface flow in the Batumerah River. The resulting discharge is calibrated with discharge data obtained from field research with the smallest error. Observation discharge in the study watershed and its surroundings is not available, so only two-day measurement data projected from the Ruapa watershed is used as calibration data.

Table 1 Batumerah Intake Calibration

Calibration					
Date	R	Qobs	Qsin	NSET	NSEB
4/11/2019	2.590	0.150	0.160	0.000	0.002
23/03/2021	0.390	0.070	0.070	0.000	0.002

Calibration	
NSE	1.00
Correl R:Qsin	78.57%

From the modeling results, daily discharge data in a span of 11 years is shown in Graph 5 comparison of the results of discharge modeling and rainfall data. From the comparison results, a fairly good correlation was obtained visually between rainfall and surface flow.

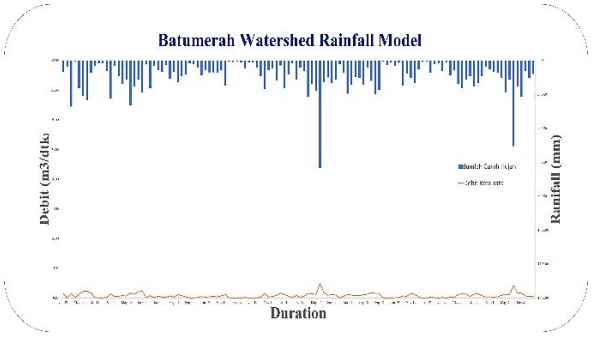


Fig.5 Comparison of Discharge and Rainfall in Batumerah Watershed

The availability of water in the Batumerah River can be calculated in a period of half a month as shown in Table 2 Half-Month Mainstay Discharge in Batumerah Watershed . In wet year conditions, the largest mainstay discharge (Q30%) of 1.83 m^3 / second occurred in October II. In normal year conditions, the largest mainstay discharge (Q50%) of 0.89 m was obtained^3/sec in September II. In dry years, the largest mainstay discharge (Q80%) of 0.35 m^3/s was obtained in June I. The graph of the mainstay discharge for the half-month period is shown in Figure 6.

Table 2 Half-Month Mainstay Discharge in Batumerah Watershed

Year	Month											
	Jan		Feb		Mar		Apr		May		Jun	
	1	2	1	2	1	2	1	2	1	2	1	2
	Jan		Feb		Mar		Apr		Mei		Jun	
Q30	0.298	0.262	0.21	0.346	0.397	0.29	0.454	0.386	0.583	0.836	0.795	1.446
Q50	0.18	0.131	0.084	0.195	0.229	0.165	0.282	0.26	0.386	0.441	0.543	0.838
Q80	0.07	0.031	0.016	0.052	0.11	0.056	0.085	0.109	0.162	0.194	0.275	0.265
Q95	0.011	0.006	0.003	0.005	0.052	0.019	0.01	0.034	0.044	0.103	0.063	0.051

Year	Month											
	Jul		Aug		Sep		Oct		Nov		Dec	
	1	2	1	2	1	2	1	2	1	2	1	2
	Jul		Agst		Sep		Oct		Nov		Dec	
Q30	1.128	1.734	0.703	0.37	0.598	0.315	0.276	0.19	0.133	0.204	0.226	0.348
Q50	0.592	0.561	0.398	0.169	0.267	0.192	0.133	0.102	0.083	0.056	0.105	0.186
Q80	0.185	0.038	0.018	0.041	0.022	0.009	0.016	0.015	0.028	0.006	0.008	0.047
Q95	0.054	0.008	0.002	0.002	2E-04	8E-06	0.003	6E-04	0.007	0.001	0.001	0.002

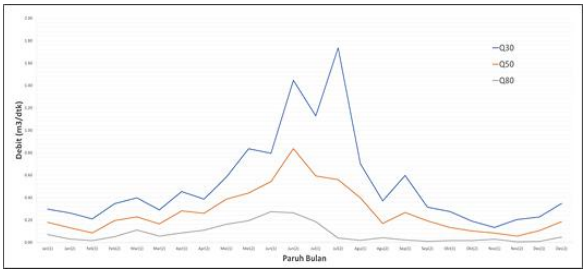


Fig.6 Half-Month Mainstay Discharge in Batumerah Watershed

4.5 Domestic and Non-domestic Water Needs

Domestic water demand is the amount of water needed by domestic consumers for household water purposes. Domestic water demand can be estimated by population, people's lifestyle, and socioeconomic conditions. Water sources in the Batumerah watershed come from river intakes and pumps. The water requirements of the Batumerah watershed from the river intakes are shown **Error! Reference source not found.**in figure 7.

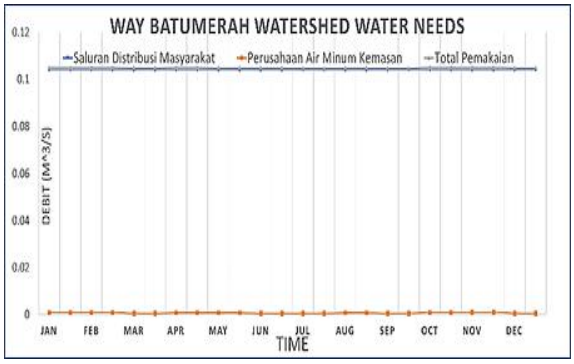


Fig.7 Water Needs of Batumerah Watershed

4.6 River Conservancy Water Needs

The calculation of maintenance flow needs must follow KP (*planning criteria*) 02/2013, which is 5% of the existing discharge in the relevant period or Q95 according to SE (*form letter*) Director General of Natural Resources 05/2016. So that the maintenance flow in the Batumerah watershed.

Table 3 Maintenance Flow Needs of the Batumerah Watershed Half-Month Maintenance

Year	Month											
	Jan		Feb		Mar		Apr		May		Jun	
	1	2	1	2	1	2	1	2	1	2	1	2
	Jan		Feb		Mar		Apr		Mei		Jun	
Q95	0.011	0.006	0.003	0.005	0.052	0.019	0.01	0.034	0.044	0.103	0.063	0.051

Year	Month											
	Jul		Aug		Sep		Oct		Nov		Dec	
	1	2	1	2	1	2	1	2	1	2	1	2
	Jul		Agst		Sep		Oct		Nov		Dec	
Q95	0.054	0.008	0.002	0.002	2E-04	8E-06	0.003	6E-04	0.007	0.001	0.001	0.002

4.7 Water Balance Calculation

The calculation of the water balance in the Batumerah watershed is carried out by finding the difference between the availability of water in the weir and the water demand in the Batumerah watershed. Water availability in the Batumerah

watershed can be predicted from the characteristic pattern of periodic rainfall in the dry, normal and wet year rain pattern groups so that the flow discharge of dry, normal and wet year conditions is obtained. Water requirements can be calculated based on irrigation, domestic, non-domestic water requirements and river maintenance.

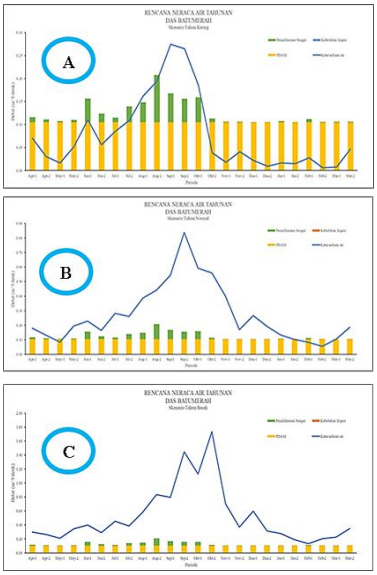


Fig.8. Water Balance of Batumerah Watershed Dry, Normal, Wet Year Pattern

V. CONCLUSION

1. The water allocation unit and allocation limit for PDAMs (*Local water company*) in the Batumerah watershed are assumed to be in accordance with the value of PDAM's water needs for the dry condition half-month, which is 0.105 m^3/sec. However, in certain months there is a deficit, so PDAM's water demand is the result of reducing water availability with minimum maintenance flow needs (0.001 m^3/sec).
2. Water use priority is a policy that regulates the order of water allocation based on applicable laws and regulations.

REFERENCES

[1] Alberta Environmental Protection, 1993. WRMM (Water Resources Management Model), Program Description, Calgary, Canada.

[2] Bird, J., Arriens, W. L., D. Von Custodio, 2009. Water Rights and Water Allocation, Issues and Challenges for Asia, Asia Development Bank, Mandaluyong City, Philippines.
[3] Collischonna , W., Reinaldo Haasb, Ivanilto Andreollia, and Carlos Eduardo Morelli Tuccia. 2005. Forecasting River Uruguay flow using rainfall forecasts from a regional weatherprediction model. Journal of Hydrology, 305 (2005), 87–98.
[4] Febriamansyah, R., 2006. The Use of AHP (The Analytic Hierarchy Process) Method For Irrigation Water Allocation In A Small River Basin (Case Study In Tampo River Basin In West Sumatra, Indonesia), Eleventh biennial global conference of IASCP, Survival of the commons: Mounting challenges and new realities, Bali 19-23 June 2006.
[5] Gany, A. Hafied A., Hatmoko, W., I. A.Yusuf. A General Overview of Decision Support System for Water Resources Planning and Management In Indonesia, Journal of Irrigation Research and Development, December 2001.
[6] Goodman, Alvin S., 1984. Principles of Water Resources Planning, Prentice-Hall, Englewood Cliffs.
[7] Hatmoko, W., 2007. The Decreasing Trend of Dependable Flow in Some Rivers In Java. Proceedings International Seminar on River and Development, Bali, 25-27 April 2007. ISBN 978-979-17093-0-9.
[8] MILT, 2007. Guidebook of Normal Flow Rate Investigation (Draft). River Environment Division, River Bureau, Ministry of Land, Infrastructure, Transport and Tourism.
[9] Quanxi Shao, Heung Wong, Ming Li, Wai-Cheung Ip, 2009. Streamflow forecasting using functional-coefficient time series model with periodic variation, Journal of Hydrology 368 (2009) 88–95.
[10] The Houw Liong, P.M Siregar, R. Gernowo, and F. Heru Widodo, (2007) Long-Term Climate Prediction in Indonesia Based on Solar Activity with ANFIS, Proceedings of the National Scientific Meeting – Indonesian Hydrological Society, Water Resources Conditions for the Achievement of Increasing Rice Production by 2 Million Tons in 2007, Jakarta February 28, 2007.
[11] Van der Krogt, W., 2010. Operational management of the Cidurian irrigation scheme, Indonesia, Exercise using the simulation model RIBASIM, Deltares, Delft.
[12] Young, W.J., 2004. Water Allocation and Environmental Flows in Lake Basin Management, Lake Basin Management Initiative Thematic Paper.

International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-11, Issue-3; Mar, 2024
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.113.2

Nuclear law: Scientometric study of the scientific literature indexed by the Scopus database between 1970 and 2023 using VOSviewer

Almeida dos Santos. A[1], Souza Lima. R.[2], Araripe Marinho.C[3], Assumpção Silva. E[4]

Department of Administration and Logistics, Brazilian Nuclear Energy Commision
Email: alberto.santos@cnen.gov.br
[2]Department of Nuclear Law, Brazilian Nuclear Energy Commision
Email: romulo.lima@cnen.gov.br
[3]Department of Technology and Innovation, Institute of Nuclear Engineering
Email : cristovao.marinho@ien.gov.br
[4]Department of Administration and Logistics, Brazilian Nuclear Energy Commision
Email: eduardo.silva@cnen.gov.br

Received: 15 Jan 2024,

Receive in revised form: 01 Mar 2024,

Accepted: 08 Mar 2024,

Available online: 18 Mar 2024

Keywords *— **Nuclear law, Scientometric study, Scientific literature, Scopus***

Abstract *— Nuclear law refers to the legal framework that governs the development, regulation, and use of nuclear energy and technology. The gap that this study seeks to fill concerns the use of the scientometric method, based on an information structure found in indexed scientific studies on the evolution of the theme of nuclear law. The present study aimed to map out the themes of scientific production between 1970 and the first semester of 2023. Data were collected from research activities indexed by the SCOPUS database and analyzed with the aid of the scientometric VOSviewer software. Considering the limitations of this study, the main scientometric results related to the theme of nuclear law are the following: the United Kingdom and the United States of America present the greatest collaboration in co-authorship between countries; the subject has a broader scope in the field of social science and no significant impact or frequency of Brazilian publications or authors was identified. It was concluded that nuclear law, in line with legal and ethical principles, can contribute to the wellbeing of humanity and the preservation of the environment.*

I. INTRODUCTION

Nuclear law refers to the legal framework that governs the development, regulation, and use of nuclear energy and technology. It covers national and international laws, regulations, and treaties that aim to ensure the safe, secure, and peaceful use of nuclear energy, minimizing the risks associated with nuclear materials and facilities (TANTER, 2014; AMPOVSKA, 2013; SZIEBIG, 2021).

Nuclear law covers various aspects related to the nuclear industry, including the licensing and operation of nuclear power plants, nuclear safety and security measures, management and disposal of radioactive waste, nuclear non-proliferation, and liability for nuclear accidents. It also involves regulating the mining, enrichment, and transport of uranium, as well as the import and export of nuclear materials and technology (ARBOUSSET, 2021; SANDS, 1996; BOWDEN, 2021).

The legal framework for nuclear energy varies from country to country, but often includes specific legislation and regulatory bodies responsible for overseeing the nuclear industry. These bodies are typically tasked with enforcing safety standards, carrying out inspections, granting licenses, and ensuring compliance with

international obligations (BUGOS, 2022; VIKTORIYA, 2019).

At the international level, nuclear law is governed by various treaties and conventions, such as the Treaty on the Non-Proliferation of Nuclear Weapons (NPT), the Safeguards Agreements of the International Atomic Energy Agency (IAEA), the Convention on Nuclear Safety and the Convention for the Prevention of Marine Pollution by Discharge of Wastes and Other Materials, known as the "London Convention", 1972, which entered into force in 1975. In this Convention, at its seventh Consultative Meeting of the Contracting Parties in 1983, an amendment was made to ban the dumping of all radioactive waste at sea. These agreements establish international standards of nuclear safety, protection, non-proliferation, and cooperation between nations (BURNS, 2022; KOVUDHIKRULRUNGSRI, NAKSEEHARACH, 2011; CALMET, 1989).

Given the potentially catastrophic consequences of nuclear accidents and the dual-use nature of nuclear technology, nuclear law plays a crucial role in ensuring the responsible and peaceful use of nuclear energy, while also addressing issues of safety, security, the environment, and non-proliferation (HANDRLICA, 2018; KECSKÉS, 2008; EKARDT, VON BREDOW, 2011).

Considering the relevance of the theme, The present work aimed to conduct a scientometric study of the scientific production indexed by the Scopus database between 1970 and 2023, which answers the following questions regarding the theme of nuclear law: (1) Which countries collaborated with each other with the co-authorship of documents? (2) Which organizations have the most public documents? (3) Who are the most published authors? (4) Which journals concentrate publications? (5) What are the most used keywords in publications? (6) In which field of science is this subject the broadest? (7) What are the impacts and frequency of Brazilian publications or authors?

The scientometric study, as a research method, is a research field that applies quantitative methods to analyze scientific publications and their characteristics. It involves the systematic collection and analysis of data related to scientific literature, such as citation counts, authorship patterns, publication trends, and collaborations (VAN ECK; WALTMAN, 2010).

Based on the limitations of this study, the main scientometric results related to the theme of nuclear law are the following: (1) the United Kingdom and the United States of America have the greatest collaboration in co-authorship between countries; (2) Charles University in Prague is the organization with the largest number of published documents; (3) Jakub Handrlica is the most published author on the subject; (4) the Journal of World Energy Law and Business has concentrated a significant number of publications on the subject over the last 5 years; (5) nuclear energy and international law are the two most used keywords when the topic of the article is nuclear law; (6) the subject has a broader scope in the field of social sciences; and (7) no significant impact and frequency of Brazilian publications or authors was identified.

II. METHODOLOGY

2.1 Scientometric studies

The main objective of scientometric studies is to provide insights into the structure, dynamics, and impact of scientific research. These studies often examine various aspects of scientific output, including research productivity, citation patterns, co-authorship networks, and scientific collaboration at individual, institutional, and national levels. By analyzing such data, scientometricians can identify trends, patterns, and relationships within scientific communities and across disciplines.

Scientometric analyses can be conducted using a variety of methods, including citation, co-citation, analysis, and content analyses. These methods enable researchers to examine the impact and influence of scientific publications, identify key contributors and research fronts, and assess the development and evolution of scientific fields.

Scientometric studies are valuable for multiple purposes, including the evaluation of individual researchers' productivity and impact, assessing the performance of research institutions, tracking the progress of scientific fields over time, and identifying emerging research areas. They also provide insights into knowledge dissemination, the flow of information within scientific communities, and the interconnectedness of research (BONYADI NAEINI, MOGHISEH, 2023).

By applying scientometric approaches, researchers can gain a deeper understanding of scientific endeavors, identify areas of collaboration and potential interdisciplinary research, and inform policy decisions related to scientific funding, research evaluation, and resource allocation.

Scientometric studies are based on the analysis of published literature and may not capture the entire landscape of scientific research, including unpublished or non-indexed works. Nonetheless, they provide a quantitative foundation for understanding the

characteristics and dynamics of scientific knowledge and contribute to evidence-based decision-making in academia and research policy (VAN ECK; WALTMAN, 2010).

2.2 Confirmation of the research gap through bibliometrics

Considering that bibliometrics is a valid instrument for scientometric studies, within the scope of this qualitative study, the context and the preliminary gap of the investigation were established based on the preliminary research. The core question of the study was then defined, which seeks to understand scientific interest and concern through causal relationships with the context in order to justify and explain it (TREINTA ET AL., 2014).

The objective of the work was defined, and through this objective, the method allowed the researchers to define the thematic areas associated with the core question of the study, seeking to guide the application of operators using the Boolean architecture for extracting bibliometric data from the SCOPUS base.

From the systematic analysis of the returned documents, observing the framework of the study, the understanding that there is a research gap was corroborated, that is, the absence of works that sought to map the development of "research trends on nuclear law" in the scientific literature, considering the period from 1970 to the first half of 2023, using scientometric methodology, and based on research data published and indexed by the SCOPUS database.

To confirm the preliminary research gap, in the first round of filter application, and using Boolean architecture, it was possible to identify only one study collected in the databases indexed by SCOPUS, using keywords associated with the thematic areas: scientometrics, nuclear legislation, and nuclear law.

First, Boolean operators "OR"; "AND" and "AND"; "AND" were used with the code "ALL", which returned documents as long as the word researched appeared in one of the following variables: in the title of the paper, in the title of the source, language, author, editor, affiliation, abstract, key words, references, DOI, ISBN, ISSN, CODEN, subjects, volume fields, year of publication, sequence bank, number of sequence bank, number, chemical name, number of CAS registration, manufacturer, editor or conferences.

In the following rounds, the use of operators "OR"; "AND" and "AND"; "AND" was kept, combined with more specific search codes, namely, TITLE-ABS-KEY, which returned documents as long as the word searched appeared in variables: abstract, title of the paper, or key word.

During the search in the base, using Boolean architecture, it was observed that, as the uses of operators with codes form more specific search scripts, as is the case of the code TITLE-ABS-KEY, the number of returned documents decreased and reached zero, as demonstrated in Table 1, confirming the research gap.

Table 1: Boolean architeture to verify the reserarch gap

Round	Operators	Architecture with codes	Number of documents returned
1ª	OR, AND	(ALL ("nuclear law")) OR (ALL ("nuclear legislation")) AND (ALL (scientometric))	1
2ª	AND, AND	(ALL ("nuclear law")) AND (ALL ("nuclear legislation")) AND (ALL (scientometric))	0
3ª	OR, AND	(TITLE-ABS-KEY ("nuclear law")) OR (TITLE-ABS-KEY ("nuclear legislation")) AND (TITLE-ABS-KEY (scientometric))	0
4ª	AND, AND	(TITLE-ABS-KEY ("nuclear law")) AND (TITLE-ABS-KEY ("nuclear legislation")) AND (TITLE-ABS-KEY (scientometric))	0

Source: Developed by the Authors

2.3 Mapping of the development of studies indexed in the Scopus database

After the research gap had been confirmed, in the second phase of the application of filters using the Boolean architecture, 502 national and international studies were identified and collected from the databases indexed by SCOPUS. The details of the Boolean research application and the collection product are presented in Table 2.

Table 2: Boolean architecture linked to the study's core subject matter

Architecture with Boolean operators	Amount of documents
(ALL ("NUCLEAR LAW") AND (LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012) OR LIMIT-TO (PUBYEAR , 2011) OR LIMIT-TO (PUBYEAR , 2010) OR LIMIT-TO (PUBYEAR , 2009) OR LIMIT-TO (PUBYEAR , 2008) OR LIMIT-TO (PUBYEAR , 2007) OR LIMIT-TO (PUBYEAR , 2006) OR LIMIT-TO (PUBYEAR , 2005) OR LIMIT-TO (PUBYEAR , 2004) OR LIMIT-TO (PUBYEAR , 2003) OR LIMIT-TO (PUBYEAR , 2002) OR LIMIT-TO (PUBYEAR , 1999) OR LIMIT-TO (PUBYEAR , 1998) OR LIMIT-TO (PUBYEAR , 1997) OR LIMIT-TO (PUBYEAR , 1996) OR LIMIT-TO (PUBYEAR , 1995) OR LIMIT-TO (PUBYEAR , 1993) OR LIMIT-TO (PUBYEAR , 1991) OR LIMIT-TO (PUBYEAR , 1990) OR LIMIT-TO (PUBYEAR , 1988) OR LIMIT-TO (PUBYEAR , 1986) OR LIMIT-TO (PUBYEAR , 1985) OR LIMIT-TO (PUBYEAR , 1982) OR LIMIT-TO (PUBYEAR , 1981) OR LIMIT-TO (PUBYEAR , 1980) OR LIMIT-TO (PUBYEAR , 1979) OR LIMIT-TO (PUBYEAR , 1978) OR LIMIT-TO (PUBYEAR , 1977) OR LIMIT-TO (PUBYEAR , 1976) OR LIMIT-TO (PUBYEAR , 1975) OR LIMIT-TO (PUBYEAR , 1972) OR LIMIT-TO (PUBYEAR , 1971) OR LIMIT-TO (PUBYEAR , 1970))	502

Source: Developed by the Authors

Excel software was used to conduct data analysis and the presentation of results, as follows: (1) comparison among the research organizations with the highest number of documents published on nuclear law between 1970 and the first half of 2023; (2) comparison among authors with the highest number of documents published on nuclear law between 1970 and the first half of 2023; (3) frequency of documents published between 1970 and the first half of 2023; documents organized by country and area of knowledge; and (4) comparison among the types of areas indexed by the Scopus database on nuclear law. Data analysis and the presentation of clusters included: (1) countries co-authorship; (2) citation of documents by organizations; (3) citation of authors; (4) bibliographic cupping of documents by sources; (5) comparison among the three journals with the highest number of documents published on nuclear law between 1986 and the first half of 2023; and (6) keywords co-occurrences, the collected data were entered into the VOSviewer software database for scientometric data processing, available at www.vosviewer.com.

The program identifies the core subject matter in documents in a low dimension space so that the space between any two elements mirrors the equivalence or affinity of items with the best possible precision. For each pair of items, i and j, the VOSviewer requires a similar ij (Sij ≥0) as entry. It treats Sij similarities as measurements in a ratio scale. Thus, the device minimizes a weighted sum of square distances among all pairs of items.

The square of the distance between a pair of studies is calculated by the similarity among items. To avoid trivial solutions where all elements have the same position, the restriction imposed is that the average distance between two items must be equal to 1.

Considering the core objective of this study, for each of the 502 documents exported from the SCOPUS base, in CSV (Excel) format, were considered for the VOSviewer to analyze the database, fields, types of data, and counting methods, as presented in Tables 3 and 4 (VAN ECK; WALTMAN, 2010).

Table 3: Types of fields for scientometric analysis

Data and fields from the Scopus database for migration to the VOSviewer software		
Data	**Types of fields**	**Number of fields**
Citation information	Authors, document title, year, source title, volume, pages, citation count, source, and types of	9

	documents, doi	
Bibliographic information	Affiliations, editor, idiom of the original document, corresponding address, abbreviated source title	5
Abstract and key words	Abstract, author key words, key words index	3
Other information	Information and conference, references	2

Source: Developed by the Authors

Table 4: Types of counting methods for scientometric analysis

Counting methods used in the VOSviewer software.		
Type	**Consolidated description**	**Number of consolidated descriptions**
Full	Only the presence or absence of a term in a document is considered.	1
Fragmented	The weight of a connection is fragmented, so that each reference, citation, or document makes the same global contribution.	1

Source: Developed by the Authors

Tables 5 and 6 present a consolidated description of the types of analysis used in this phase of research and the criteria used in the software parameterization, including the counting of terms, the bibliographic coupling and co-citation, the respective analysis units present in the database, and the number of analysis units selected to calculate the total strength of the links for the preparation of cluster maps.

Table 5: Types of analysis used in the VOSviewer software.

Types of analysis used in the VOSviewer software.		
Type	**Consolidated description**	**Number of Consolidated descriptions**
Counting of terms	Frequency of term appearing in the main body of the document (title and abstract)	1
Bibliographic coupling	The list of items is determined based on the number of references that they share.	1
Co-citation	The list of items is determined based on the number of times they are cited together.	1

Source: Developed by the Authors

Table 6: Types of parameterization used in the VOSviewer software.

Criteria used in software the VOSviewer parameterization.			
Type of analysis	**Unit of analysis**	**Minimum number of analysis units in the database**	**Number of analysis units selected for calculation**
Co-authorship	Countries	01	49
Co-occurrences	Keywords	02	459
Citation	Authors	01	235
(Citation) Documents	Organizations	01	395
Citation	Countries	01	49
Bibliographic Coupling	Sources	01	194

Source: Developed by the Authors

Specifically to create the maps, based on the main body of both the title and the abstract of the documents, the punctuation of the terms was created based on the year of publication in the field, using the binary counting method, where the number of occurrences of the term was at least equal to 10. The number of analyzed units selected to calculate the total form of links was 37,302 terms. After processing the data in Microsoft Excel and VOSviewer software, the results analysis phase began.

Figure 1 illustrates the macro view of the methodological process used in this scientometric study.

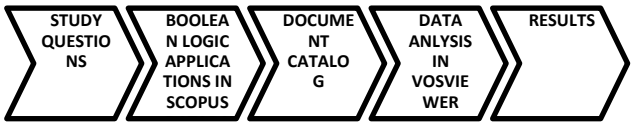


Figure 1: macro view of the methodological process.

Source: Developed by the Authors

III. RESULTS AND DISCUSSIONS

3.1 Co-authorship of countries

Co-authorship of countries refers to the collaboration between researchers from different countries in publishing scientific papers. When researchers from multiple countries collaborate on a research project and publish a paper together, it is considered a co-authored publication between those countries.

Analyzing the co-authorship patterns of countries can provide insights into the extent and nature of international scientific collaboration. This helps to identify trends, patterns, and networks in scientific research, and can provide valuable information for policymakers, funding agencies, and researchers themselves.

Co-authoring networks view the connections between countries based on their collaborative relationships. Figure 1 shows that the United Kingdom and the United States of America have the greatest collaborative connection of co-authorship between countries on the subject of nuclear law.

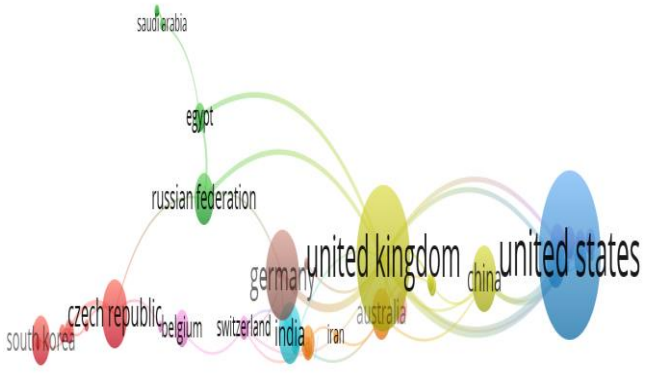


Figure 2: Countries co-authorship

Source: Developed by the Authors

The field of nuclear law encompasses regulations and legal frameworks related to the peaceful use of nuclear energy, nuclear safety, nuclear liability, non-proliferation and other legal aspects related to nuclear activities. Collaboration between US and UK authors in this field can contribute to the exchange of knowledge, the sharing of expertise, and the development of international legal frameworks in the nuclear sector. The results presented in Figure 2 demonstrate that US and UK researchers frequently collaborate on research papers and publications related to nuclear law, reflecting a shared interest in specific topics or legal issues within the field. In collaborating, US and UK authors bring diverse perspectives, legal frameworks, and experiences to tackle complex challenges in nuclear law.

3.2 Citation and documents published by organizations

Figures 3 and 4 below show the results related to the citation of documents by organizations and the comparison between the research organizations with the highest number of documents published on nuclear law between 1970 and the first half of 2023.

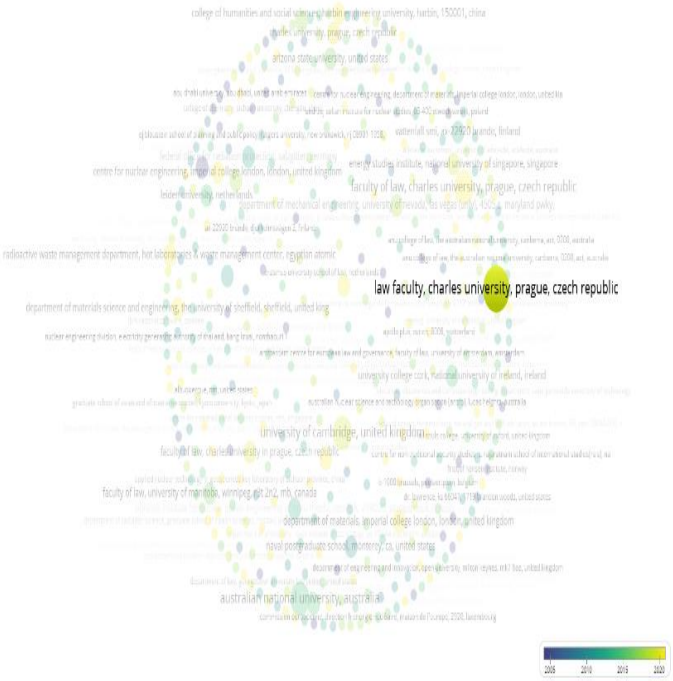


Fig.3: Citation of documents by organizations

Source: Developed by the Authors

Although the Czech Republic is not ranked number one among the most cited countries in nuclear law, the organization that has the highest frequency of citations is the School of Law at Charles University in Prague, which is an organization that plays a prominent role in nuclear law studies. The university offers academic programs and

conducts cutting-edge research in this particular area. The Charles University School of Law is known internationally for its academic tradition.

The college houses the Center for Nuclear and Environmental Law (CENEA), which is one of the leading research and teaching centers in nuclear law. CENEA collaborates with other national and international institutions and contributes to the development of nuclear and environmental law through research, seminars and conferences (HANDRLICA, 2019).

Nuclear law studies at Charles University cover a wide range of resources related to regulatiosn, governance, and legal issues of nuclear energy. Courses offered may include International Nuclear Law: Study of international treaties, conventions, and regulations related to nuclear energy, including the Czech Republic's participation in international bodies, such as the IAEA.

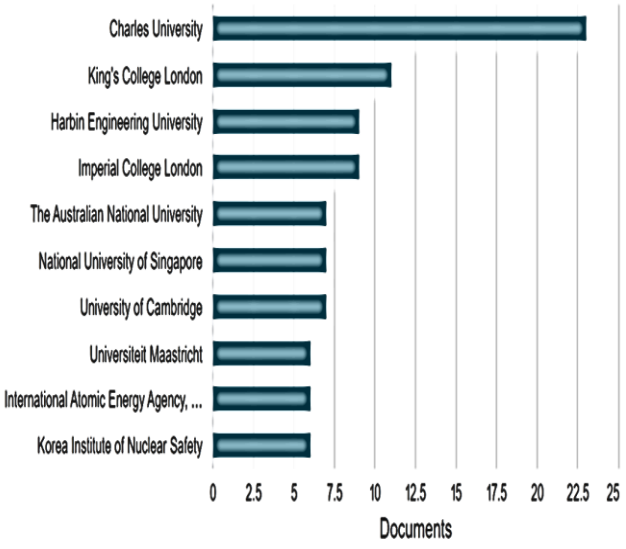


Fig.4: Comparison between the research organizations with the highest number of documents published on nuclear law between 1970 and the first half of 2023

Source: Developed by the Authors

3.3 Citation of authors and number of documents published

Figures 5 and 6 below show the results referring to the comparison among the authors with the highest number of documents published on nuclear law between 1970 and the first half of 2023, and the authors with the highest frequency of citations in indexed scientific works.

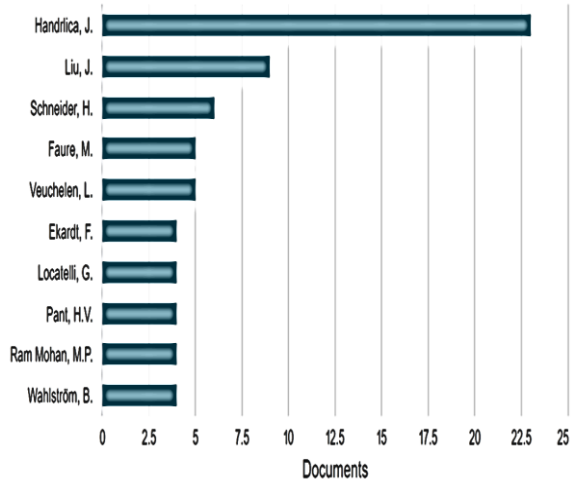


Fig.5: Comparison among the authors with the highest number of documents published on nuclear law between 1970 and the first half of 2023

Source: Developed by the Authors

In the area of scientometrics, the comparison of authors by the number of published documents is a common analysis to assess the scientific productivity of researchers. This metric is useful for identifying authors who have a significant contribution in terms of the number of publications.

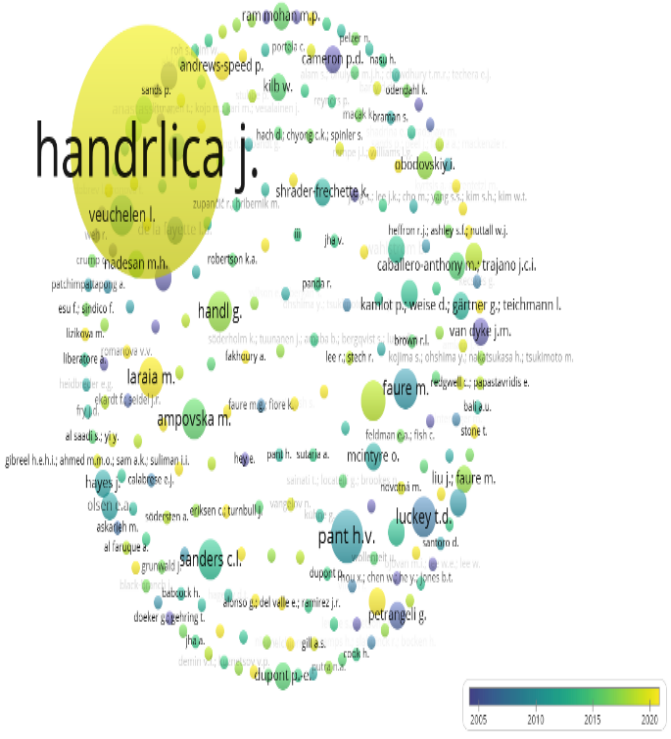


Figure.6: Citation of authors

Source: Developed by the Auth

In scientometrics, one citation per author refers to counting the number of times a given author is cited in academic or scientific works. This metric is often used to assess an author's influence and impact upon the academic community.

From the content analysis of the articles, 4 authors and 9 articles stand out, where the object and thematic centrality of the scientific study is nuclear law in a more specific and defined perspective, as shown in Table 7.

Table 7: Authors of articles, where the core object and thematic of the study is nuclear law – defined perspective

Document Title	Authors	Source	Year
Nemesis of New Technologies in Nuclear Law	Handrlica, J.	Pravnik	2022
The Feast of Insignificance of Small Modular Reactors in International Nuclear Law	Handrlica, J.; Novotná, M	Czech Yearbook of Public and Private International Law	2021
Whither the Future of International Nuclear Law? A Survey in Legal Futurism	Handrlica, J.	Czech Yearbook of Public and Private International Law	2020
From the Front Lines of International Nuclear Law: Looking Back at the 'Nuclear Inter Jura' Congress, Held in Abu Dhabi, United Arab Emirates	Handrlica, J.	Journal of World Energy Law and Business	2019
Nuclear Law Revisited as an Academic Discipline	Handrlica, J.	Journal of World Energy Law and Business	2019
"Atomic Law" or "Nuclear Law"? An Academic Discussion Revisited	Handrlica, J.	Brics Law Journal	2018
"Exclusivism" in International Nuclear Law: The Concept Revisited	Handrlica, J.	Lawyer Quarterly	2018
International Cooperation and Guarantee of Nuclear Safety in the Implementation of National Nuclear Law	Schneider, H.	Atw – Internationale Zeitschrift Fur Kernenergie	2011
The Justification and Optimisation Principle in International Nuclear Law: Theory and Practice	Veuchelen, L.	Atoms For Peace	2005

Source: Developed by the Authors

Figures 5 and 6, and Table 7, show that the articles on nuclear law presented by Jakub Handrlica, from Charles University, in the Czech Republic, are those that produced the greatest impact on the Scopus database, mainly from 2019 onwards. The author's main areas of interest include energy and nuclear law and international administrative law (HANDRLICA, 2021).

3.4 Frequency of documents published, bibliographic coupling, and journals

The Figure 7 below shows the frequency of documents published on the subject of nuclear law, and indexed by the Spocus database, between 1970 and the first half of 2023.

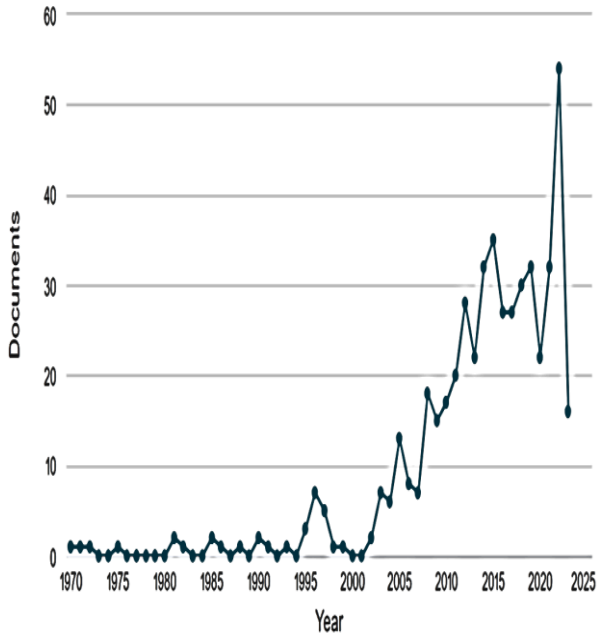


Fig.7: Frequency of documents published between 1970 and the first half of 2023

Source: Developed by the Authors

The growth of scientific publications on nuclear law is related to the growing importance of the subject and the development of nuclear energy and energy technologies worldwide. In recent years, there has been a significant increase in interest in and awareness of the legal aspects of nuclear power and energy activities. This is due to several factors, such as the growing demand for nuclear energy in some countries, the development of technologies, the need for effective regulation and governance, and concerns about safety at nuclear facilities. As a result of this increased interest, it is expected that the number of scientific publications on nuclear law has also increased. This includes articles in peer-reviewed journals, books, book chapters, and other types of scholarly publications.

Figure 8 presents the bibliographic coupling of documents by sources, which refers to the practice of identifying and analyzing the connections between scientific documents through the cited bibliographic sources. It is a technique used in scientometric analysis to examine the interconnectivity and mutual influence between scientific articles.

When carrying out the bibliographic coupling, the bibliographic references cited in a given article were identified, and these references were tracked to find the documents that contain them.

This allows one to map reference networks and discover which studies are closely related to or have influenced the original work.

Bibliographic coupling by sources is a valuable tool to understand the dynamics of scientific research and the interaction between academic journals. It also helps to identify key contributions to a field of study, track the spread of ideas, and identify gaps in scientific knowledge.

Figure 8 presents the main cluster formed from the bibliographic coupling of documents by source, identifying the Journal of World Energy Law and Business (JWELB) as that with the greatest connectivity and interaction with other journals.

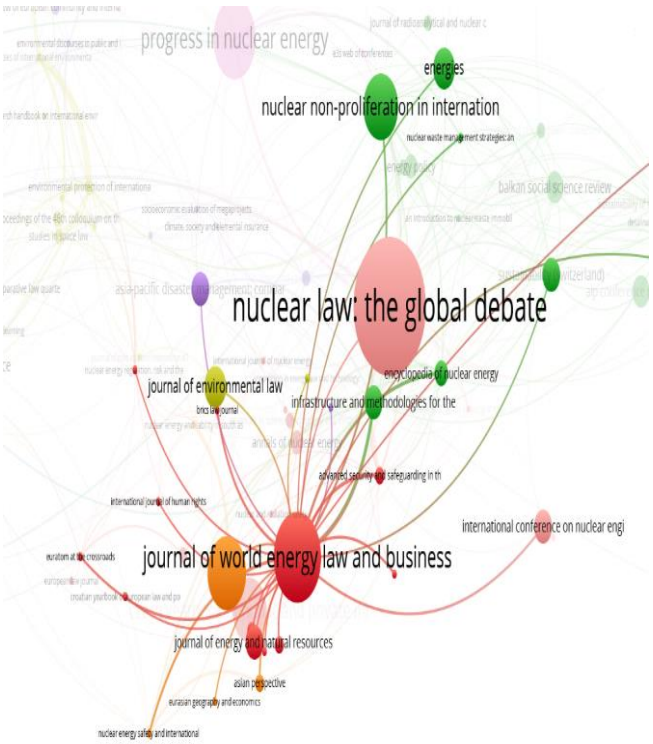


Fig.8: Bibliographic Coupling of Documents by Sources

Source: Developed by the Authors

The JWELB is a specialist academic publication that focuses on energy law and business at a global level. It covers a wide range of energy-related subjects, including nuclear law, regulation, public policy, energy investment, environmental issues, and sustainability.

JWELB is a peer-reviewed journal that publishes high-quality articles accredited by academics, practitioners, and industry experts. The journal is published in collaboration with the Association of International Petroleum Negotiators (AIPN) and is widely recognized as a leading publication in the field of nuclear law and energy business.

JWELB's mission is to promote understanding and advance knowledge in the operational interdisciplinary areas of nuclear law and energy business on a global scale. It also provides a forum for debate and dissemination of original and innovative research, as well as an in-depth analysis of legal and business issues relevant to the energy sector.

Articles published on JWELB cover a wide variety of topics such as energy contracts, arbitration and dispute resolution, oil contract negotiation, regulatory issues in different jurisdictions, the development of renewable energy resources, sustainability policies, among many other aspects related to energy and its legal and business context.

The journal is an important reference for academics, industry professionals, and policymakers interested in issues related to nuclear law and energy business around the world. Through the publication of high-quality research, JWELB contributes to the advancement of knowledge and promotes discussion on the challenges and opportunities facing the energy sector on a global scale.

Figure 9 shows the comparison among the three journals with the highest number of documents published on nuclear law between 1986 and the first half of 2023. It was observed that the Journal of World Energy Law and Business, from 2018 onwards, has concentrated a significant number of publications on the theme (HANDRLICA, 2021; HANDRLICA, 2019).

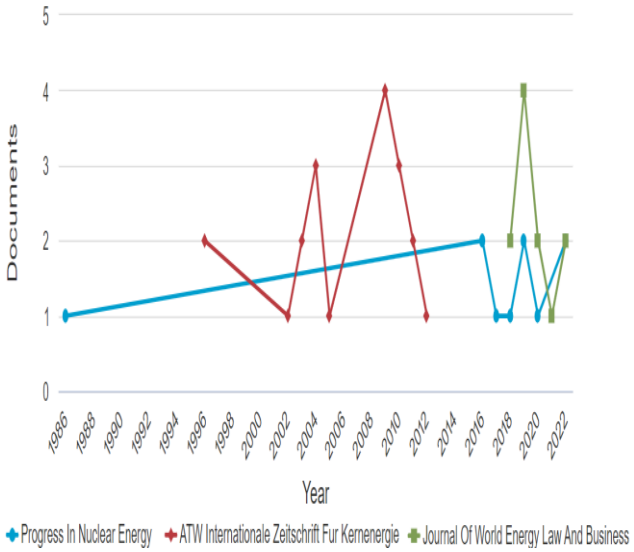


Fig.9: Comparison among the three journals with the highest number of documents published on nuclear law between 1986 and the first half of 2023

Source: Developed by the Authors

It was found that the Atw – Internationale Zeitschrift Fur Kernenergie, between 1996 and 2012, contained the largest number of publications on the subject of nuclear law. Progress in Nuclear Energy, on the other hand, has shown stability in a number of publications on nuclear law over the last 27 years.

Internationale Zeitschrift Fur Kernenergie is a scientific journal specialized in nuclear energy, covering a wide range of topics related to this specific field. The journal aims to provide a platform for the publication and dissemination of high-quality research on science, technology, applications, and regulatory aspects related to nuclear energy.

The journal is published in German and is one of the leading specialist magazines on nuclear energy in Germany and other German-speaking countries. It also has a long history, spanning over a 70-year period, and is known for its relevance and contributions to the nuclear community.

The journal publishes original scientific articles, reviews, short articles, and technical reports covering a wide range of topics related to nuclear energy, including regulations, safety standards, nuclear licensing, radiation protection, nuclear legislation, and safety management.

Progress in Nuclear Energy is an international scientific journal covering a wide range of resources related to nuclear energy. It publishes original research, reviews, case studies, and short papers on many aspects of nuclear power, including reactor technology, nuclear safety, toxic waste management, energy policy, lifecycle analysis, modeling and simulation, among others.

The journal aims to provide a platform to share knowledge and advance the field of nuclear energy. It promotes collaboration among researchers, academics, engineers, and professionals in the nuclear industry, and is well-known in the development of innovative solutions and improvement of nuclear technology.

Articles published in Progress in Nuclear Energy cover a variety of areas, such as energy policy and regulation. In this field, articles can explore government policies and regulations related to nuclear energy, including issues of non-perspectives, international agreements, nuclear law, and safety policies.

3.5 Keywords co-occurrences

Figure 10 shows the co-occurrence of keywords in the documents that were examined in this study. The co-occurrence of keywords in scientometric studies is a technique used to analyze the frequency and relationship between keywords in scientific documents. This analysis

makes it possible to identify patterns, trends, and research themes in a given field of study.

By performing a keyword co-occurrence analysis, researchers identify the keywords assigned to a set of documents and calculate how often those keywords appear together. This helps to understand the interconnectivity between keywords, and reveals the structure and thematic relationships of the scientific literature.

There are several ways to perform keyword co-occurrence analysis. A common approach is to use network analysis methods, as in this study, of keyword co-occurrence networks, which represent keywords as nodes and co-occurrence relations as edges. These networks can be viewed through graphs or network diagrams, which highlight the most frequent keywords and their interconnections.

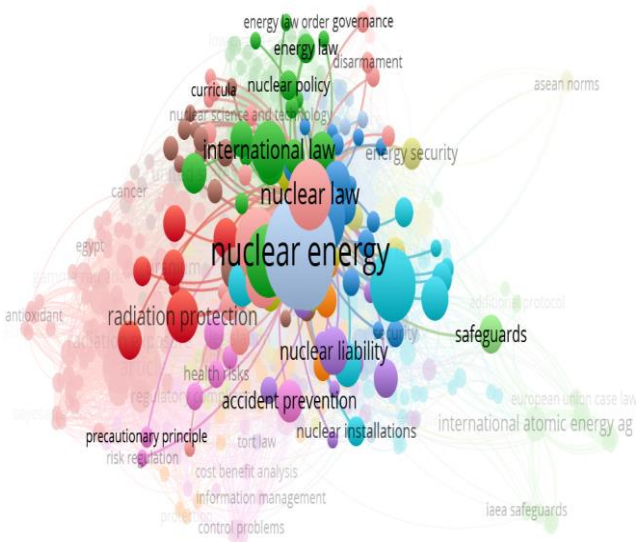


Fig.10: Keywords co-occurrences

Source: Developed by the Authors

Figure 10 shows that the keyword "nuclear law" nearly as frequently as the terms nuclear energy and international law. The strength of this relationship is natural, considering that international law seeks to regulate and govern activities related to the use of nuclear energy, nuclear weapons and nuclear non-proliferation, offering a set of norms and principles that aim to prevent the undue and uncontrolled use of nuclear energy, as well as promoting peaceful cooperation in this field.

Since the development of the first nuclear weapons during World War II, international law has played a key role in trying to control and regulate the use of these devastating weapons. The central landmark in this regard is the Treaty on the Non-Proliferation of Nuclear Weapons (NPT), which entered into force in 1970 and has been a pillar of international nuclear law. The NPT seeks to prevent the spread of nuclear weapons by promoting nuclear disarmament and facilitating access to nuclear energy for peaceful purposes (HANDRLICA, 2018).

In short, the relationship between international law and nuclear law is essential for regulating the peaceful use of nuclear energy and controlling the proliferation of nuclear weapons. While challenges remain, international law plays a key role in the quest for a safer, nuclear-weapon-free world.

3.6 Comparison among the types of areas indexed by the Scopus database on nuclear law

Figure 11 shows the comparison among the types of areas indexed by the Scopus database on nuclear law. It is interesting to note that more than half of the documents published on the subject were in the field of engineering, energy, and social sciences.

The highest concentration of indexed documents associated with the field of social sciences can be explained through four aspects: (1) social and political impacts; (2) public participation and decision-making; (3) global regulation and governance; and (4) historical perspective and normative evolution.

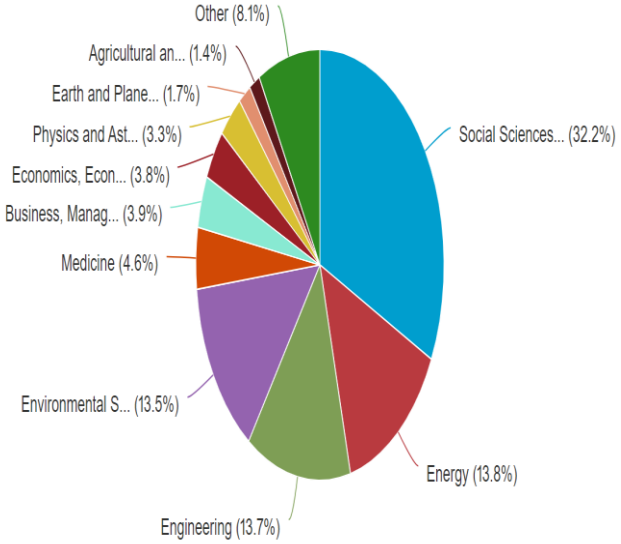


Fig.11: Comparison among the types of areas indexed by the Scopus database on nuclear law

Source: Developed by the Authors

In short, nuclear law is a multidisciplinary field that involves social, political, ethical, and legal issues. Social science researchers can explore nuclear law to understand the social and political implications of nuclear energy, analyze decision-making processes, examine global

governance, and investigate the field's normative evolution.

IV. CONCLUSIONS

Considering the limitations of this study, the main scientometric results related to the theme of nuclear law are the following: (1) the United Kingdom and the United States of America have the greatest collaboration in co-authorship between countries; (2) Charles University in Prague is the organization with the largest number of published documents; (3) Jakub Handrlica is the most published author on the subject; (4) the Journal of World Energy Law and Business has published a significant number of articles on the subject over the last 5 years; (5) nuclear energy and international law are the two most used keywords when the topic of the article is nuclear law; (6) the subject has a broader scope in the field of social sciences; and (7) no significant impact and frequency of Brazilian publications or authors was identified.

From the results, it is possible to conclude that the present scientometric study, based on the analysis of the scientific production on Nuclear Law, indexed by the Scopus database, between the years 1970 and 2023, using the VOSviewer tool, provided a comprehensive and detailed view of the evolution and trends of this area of knowledge over time.

The results revealed a notable growth in interest and scientific production in relation to Nuclear Law, especially in recent decades, following the accelerated development of the nuclear sector and the legal challenges that arose with this technological advance. This increase demonstrates the growing importance that the subject has acquired in the academic community and in society as a whole, highlighting its critical role in the formulation of policies, regulations and legal approaches to deal with complex and multifaceted core issues (FIALKOFF, MAN, 2022).

In addition, the analysis of citation networks and collaborations between researchers has enabled the identification of the main actors and institutions involved in the study of Nuclear Law, as well as the most relevant and interconnected themes that have been addressed in the scientific literature.

However, this study also highlighted the continuing need to invest in research in this area and encourage cooperation between researchers and institutions from different countries. Looking specifically at the discipline of Brazilian Nuclear Law, the scenario is worse than one can imagine. In fact, the literature in this field is quote sparse. There is no regular training in Nuclear Law in Brazil. Law academies do not offer knowledge of Nuclear Law, except for general classes in Regulatory Law, Administrative Law, and Constitutional Law, primarily because the core framework on nuclear activities is inserted in the Federal Constitution, which establishes the precept of monopoly, as well as the rules and principles related to nuclear activities. On the other hand, training in such areas as physics, chemistry, and nuclear engineering does not offer any type of approach to Nuclear Law.

Considering that Nuclear Law regulates nuclear activities in a broad sense, specifically from a legal point of view, this scenario reveals an inadequate and undesirable general internal framework, in view of the legal aspects of civil liability for nuclear damage, safeguards, the licensing process of nuclear energy, administrative sanctions, and legal proceedings in general (BARROS, 2011).

However, in 2022, during the 1st International Conference on Nuclear Law, held in Vienna, Austria, the International Atomic Energy Agency (IAEA) signed practical agreements with some countries, including Brazil, to support training in Nuclear Law. Due to this initiative, the IAEA and the National Nuclear Energy Commission of Brazil (CNEN) successfully produced the first training on Nuclear Law in May/June 2023. Today, CNEN, through its Institute of Nuclear Engineering, is preparing a postgraduate course in Nuclear Law, which is expected to begin in early 2024.

It is an important issue to understand whether this type of training will improve the scientific scenario of Nuclear Law in Brazil. Thus, it is recommended that future studies investigate the evolution of the impact and frequency of publications by Brazilian authors and the international cooperation of national institutions on the subject.

The expansion of knowledge in this field will require the involvement of experts from different disciplines, promoting a multidisciplinary approach that covers not only legal aspects, but also technical, ethical, and social issues related to nuclear energy that can be explored in future research: international responsibility in the case of nuclear accidents, cybersecurity, and the protection of nuclear infrastructure, safe transport of nuclear materials, ethical and social aspects of nuclear energy, comparative legislation on nuclear energy, protection of human rights in nuclear areas, and legal aspects of nuclear energy in contexts of conflict and international cooperation (GROSSI, 2022; JADALHAQ, ALQODSI, 2021).

In conclusion, the scientometric study of the scientific production on Nuclear Law, conducted using the

VOSviewer tool, brought a valuable contribution to the understanding of the evolution, geographic distribution, and collaboration of researchers in the area. The findings of this work are essential in supporting informed decision-making by legislators, government officials, and other actors involved in the regulation and governance of the nuclear sector. Technological progress will continue to generate complex challenges, and the science of Nuclear Law will play a vital role so that nuclear energy can be used in a safe and responsible manner, in line with legal and ethical principles, contributing to the wellbeing of humanity and the preservation of the environment.

REFERENCES

[1] Ampovska, M. (2013). Nuclear Energy and Nuclear Law in Macedonia and Neighbor Countries-Bulgaria, Serbia and Albania. *Balkan Social Science Review*, (1), 3-23.Ampovska, Marija et al. Nuclear Insurance Pools Worldwide: The Role In The Nuclear Law. Balkan Social Science Review, v. 9, n. 9, p. 7-21, 2017

[2] Ampovska, M. (2017). Nuclear insurance pools worldwide: the role in the nuclear law. *Balkan Social Science Review*, *9*(9), 7-21.

[3] Arbousset, H., Rambour, M., Schellenberger, T., Lahorgue, M. B., Leger, M., & Rubercy, G. D. (2021). Nuclear law January-December 2020. Synthesis. *Droit de l'Environnement*, 42-48.

[4] Barros, A. C. R. (2011). Ensaio jurídico sobre o dano nuclear no direito brasileiro.

[5] Bonyadi Naeini, A., & Moghiseh, Z. (2023). Open Access Scientific Outputs Published by Iranian Researchers: Scientometrics and Altmetrics Study. *Scientometrics Research Journal*, *9*(1), 125-150.

[6] Bowden, P. (2021). Principles and Practice of International Nuclear Law. *Nuclear L. Bull.*, *107*, 83.

[7] Giacomo, C., & Bugos, S. (2022). Putin Calls Up Reservists, Renews Nuclear Threat. *Arms Control Today*, *52*(8), 28-29.

[8] Burns, S. (2022). Milestones in Nuclear Law: A Journey in Nuclear Regulation. *Nuclear Law*, 55.

[9] Calmet, D. P. (1989). Ocean disposal of radioactive waste. Status report. *IAEA bulletin*, *31*(4), 47-50.

[10] Ekardt, F., & von Bredow, H. (2011). *Managing the ecological and social ambivalences of bioenergy: sustainability criteria versus extended carbon markets* (pp. 455-480). Springer Berlin Heidelberg.

[11] Herbach, J. (2021). *International Arms Control Law and the Prevention of Nuclear Terrorism*. Edward Elgar Publishing.

[12] Grossi, R. M. (2022). Nuclear Law: The Global Debate. *Nuclear Law*, 1.

[13] Handrlica, J. (2018). "atomic law" or "nuclear law"? An academic discussion Revisited. *BRICS Law Journal*, *5*(3), 135-151.

[14] Handrlica, J. (2021). Black swans, dragon kings and the uncertainty in international nuclear law. *The Journal of World Energy Law & Business*, *14*(1), 25-37.

[15] Handrlica, J. (2021). Book review: Handbook on European Nuclear Law: Competences of the Euratom Community under the Euratom Treaty, by Rasa Engstedt.(Alphen aan den Rijn: Kluwer Law International, 2021). *Common Market Law Review*, *58*(3).

[16] Handrlica, J. (2018). "EXCLUSIVISM" IN INTERNATIONAL NUCLEAR LAW: THE CONCEPT REVISITED. *The Lawyer Quarterly*, *8*(3).

[17] Handrlica, J. (2019). From the front lines of international nuclear law: looking back at the 'Nuclear Inter Jura'Congress, held in Abu Dhabi, United Arab Emirates. *The Journal of World Energy Law & Business*, *12*(3), 271-274.

[18] Handrlica, J. (2019). Nuclear law revisited as an academic discipline. *The Journal of World Energy Law & Business*, *12*(1), 52-68.

[19] Handrlica, J. (2021). The mirage of universalism in international nuclear liability law: A critical assessment 10 years after Fukushima. *Review of European, Comparative & International Environmental Law*, *30*(3), 375-386.

[20] Jadalhaq, I. M., & Alqodsi, E. M. (2021). Tort law makes a quantum leap: a review of the civil liability regime for nuclear operators in UAE law. *Journal of Property, Planning and Environmental Law*, *13*(1), 17-30.

[21] Kecskés, G. (2008). The concepts of state responsibility and liability in nuclear law. *Acta Juridica Hungarica*, *49*(2), 221-252.

[22] Kovudhikrulrungsri, L., & Nakseeharach, D. (2011). Liability Regime of International Space Law: Some Lessons from International Nuclear Law. *JE Asia & Int'l L.*, *4*, 291.

[23] Romanova, V., & Handrlica, J. (2019). Problems and Tendencies of Nuclear Law. *Russ. LJ*, *7*, 194.

[24] Sands, P. (1996). Observations on International Nuclear Law Ten Years after Chernobyl. *Rev. Eur. Comp. & Int'l Envtl. L.*, *5*, 199.

[25] Sziebig, O. J. (2021). The Challenges Of Nuclear Waste Disposal–International Environmental Law Perspective. *Curentul Juridic*, *86*(3), 88-100.

[26] Tanter, R. (2021). Hope Becomes Law: The Treaty on the Prohibition of Nuclear Weapons in the Asia-Pacific Region. *Journal for Peace and Nuclear Disarmament*, *4*(sup1), 234-275.

[27] Treinta, F. T., Farias Filho, J. R., Sant'Anna, A. P., & Rabelo, L. M. (2014). Methodology of bibliographical research using multicriteria decision-making methods. *Production*, *24*, 508-520.

[28] Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *scientometrics*, *84*(2), 523-538.

International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-11, Issue-3; Mar, 2024
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.113.3

Hunters of Joy Extension Project: An Experience Report

Projeto de Extensão Caçadores da Alegria: Um Relato de Experiência

Amanda Lima Franco[1], Adalberto Bandeira Pinheiro Júnior[2], Alicia Pinheiro[3], Ana Carolina Maués de Souza[4], Ana Carolina Franco Menezes[5], Ana Clara Valente de Alencar[6], Anabelle Guedes Macedo[7], Ariane Rebelo Viana[8], Arthur Armindo Silveira Consul[9], Bárbara Ferreira Pereira[10], Bárbara Giuliana Mendonça Góes[11], Dayane Pricila Silva de Oliveira[12], Diego Teles Borges Leal[13], Giovanna May Nogami[14], Jully Jamile Ribeiro Dos Reis[15], Luiza Bastos Campos[16], Luiza Wanzeller Monteiro[17], Maria Eduarda Menezes Almeida[18], Maria Gabriela da Rocha Florêncio[19], Maria Gabriela Perdigão Barros Monteiro[20], Maria Karolina dos Santos Pinto Oliveira Alab[21], Maria Luiza da Silva Oliveira Costa[22], Mateus Augusto Cunha Soares[23], Matheus Silva Sales[24], Sheine Alves de Souza[25]

[1]Centro Universitario Metropolitano da Amazônia, Belém do Pará. Brasil
Email: Amanda.lima.franco@gmail.com

Received: 21 Jan 2024,

Receive in revised form: 05 Mar 2024,

Accepted: 15 Mar 2024,

Available online: 24 Mar 2024

***Keywords*— Extension project. Clown therapy. Health Institution. Teacher. Preceptor. Extension.**

***Abstract*—** *Introduction: Demonstrate in a descriptive way an experience report in participating in the Caçadores da Alegria project from the point of view of a monitor, analyzing how it supported his academic training in a medical course, the contribution to the population served, as well as the performance of clown therapy, humanizing the doctor-patient relationship. Methodology: the procedure in writing the experience report had an observational descriptive process. Longitudinal data were collected orally from the 20 students approved in the selection process, describing their activities and quantifying the actions carried out through attendance and acceptance documents from the educational institution and hospital. Results: there were only positive points with the project, the contributions to building a realistic view of hospitalized patients, especially children, seeking to treat the person and not the disease. The contributions of the activities showed results such as learning humanized care for students, introducing clown therapy to hospitalized patients and experiencing the practice of theoretical activities. Building a smile as an escape from an environment full of sadness. However, the negative points of the research were the difficulties in finding hospitals that accepted the project, bureaucratic documentation required, acceptance time and slow processes in terms of document accuracy. Conclusion: regarding the objectives of the work, it demonstrates the need to apply extension activities in graduation, as a social role considering the demands of the population. The proposal of*

clown therapy is extremely important to escape problems as well as for students to appear more human.

I. INTRODUÇÃO

Quanto a formação acadêmica do ensino superior, o ministério da saúde instituiu a metodologia teórica em três dimensões: ensino, pesquisa e extensão[1]. Tendo como partida o desenvolvimento da extensão, sua abordagem na prática educacional constituiu-se em ideologias voltadas ao assistencialismo e prestação de serviços as principais demandas sociais, buscando um diálogo entre universidade, comunidade e transformação social quanto a participação de discentes as tais demandas, os transformando em produtores de bens e serviços. Tornam-se então atividades que extrapolam o viver da instituição, produzindo atividades que diretamente possam intervir em benefício a sociedade, legalmente escrito pelo Plano Nacional de extensão Universitária aprovado em 2012[2].

As primeiras atividades extensionistas foram realizadas pela universidade do Rio de Janeiro, uma das primeiras fundadas no Brasil, entre os anos de 1911 e 1917. Suas iniciativas continham conferências e semanas abertas ao público, porém com temas abordados não relacionados com a problemática de questões sociais ou econômicas do momento. Apenas em 1930 o ministro da saúde pública e educação Francisco Campos propôs alterações nas diretrizes educacionais de ensino superior em que houvesse um intercâmbio entre instituição e a sociedade com temáticas atuais. Sendo assim, o presidente Getúlio Vargas sancionou o Decreto nº 19.851, de 11 de abril de 1931 que visou estabelecer legalmente a extensão universitária[3].

As ações de extensão atuam como produtoras de novos conhecimentos, seja ampliando seu universo de referências, criação de novas modalidades de pesquisa, ou mesmo uma reflexão sobre assuntos da problemática social. Já para a comunidade, contribui abrindo oportunidades de ação, não desconsiderando a complexidade da realidade entre economia, política e saúde, proporcionando uma melhoria regional. Nesse sentido, a universidade recebe diretamente informações sobre as demandas sociais e de uma forma não ingênua atenda ao mercado, pois quando a extensão é a única responsável em colocar alunos em contado com a sociedade, a mesma se torna alienada se desvinculando as lacunas populares[4].

Todo homem é culpado pelo bem que ele não fez – Voltaire. Frase tema do Projeto Caçadores da Alegria, que é um grupo de atuação filantrópica em todo o brasil, principalmente nas escolas de saúde[5]. O grupo se formou com influência do médico Patch Adams conhecido pela conduta proeminente feliz e apaixonada pelos pacientes, fundador de um instituto que presta atenção totalmente gratuita nos Estados Unidos e admirado por médicos de todo o mundo pela sua filosofia. Como objetivo do projeto, realizam-se visitas a instituições de cuidados como em hospitais e asilos, por voluntários vestidos de palhaços levando alegria e desmistificando a relação médico-paciente. Realiza ações de doação de refeição, roupas, brinquedos e atividades com brincadeiras interativas[6].

Como fator de seleção a ingressar no projeto, os voluntários passam pela primeira etapa em formato de prova teórica, onde os melhores classificados são entrevistados sobre os assuntos abordados em manual disponibilizado antes do processo, assim como suas expectativas quando a participação. Os mesmos passam por um workshop de tema “Como Se Portar nas Ações”, “Oficinas de formação de palhaços”, regidas por preceptores com experiências em circos e atividades em hospitais. Após esta etapa são desenvolvidas aulas quanto a diferenciação na atuação na ala pediátrica e adulta, assim os tornando capacitados para o projeto[7].

A palhaçoterapia teve início em 1986 inspiradas nos ensinamentos de Patch Adams, na sua experiência como paciente em instituições de saúde mental, onde percebeu que o humor promovia o bem esta humano. Atualmente é comum encontrar palhaços vestidos de médicos em hospitais. Assim, para se tornar um médico palhaço com sucesso deve –se aprender a trabalhar com as emoções do paciente, reduzindo a ansiedade relacionada a internação hospitalar, tendo a apresentação ao estudante de medicina que o tratamento é voltado ao paciente e não com a doença[8].

O objetivo central deste estudo é demonstrar de forma descritiva o relato de experiência do autor quanto sua participação no projeto, como a tal corroborou para sua formação acadêmica e como o mesmo contribuiu a população atendida nas ações.

II. METODOLOGIA

Com relação ao objetivo deste trabalho, seu delineamento é descrito observacional. Com procedimentos de relato de experiência, natureza qualitativa e coleta de dados longitudinal. Realizado em uma instituição ensino superior de Belém do Pará, em conjunto com hospitais regionais e locais. O período foi durante todo o ano de 2019, o tempo em que o acadêmico autor desempenhou papel de monitor do projeto.

O universo da amostra foram 20 alunos aprovados no processo seletivo, entre eles 5 eram monitores, auxiliados por um docente coordenador de origem institucional designado para esta atividade. Ao final do período de ensino, não foram aplicados nenhum método de análise da amostra, apena um documento de controle de frequência aos extensionistas comprovando sua presença em cada ação.

A principal contribuição deste estudo é apresentar a comunidade científica a atuação social da atividade, relatando como o Projeto Caçadores da Alegria afetou na formação acadêmica do autor, assim como contribuição para atendimentos humanizados a população abordada.

III. RELATO DE EXPERIÊNCIA

As vivências no projeto de extensão caçadores da alegria mostrou o quanto é necessário a extensão no âmbito universitário, principalmente nos cursos de saúde e em especifico a medicina, pois mostra uma visão real da sociedade sobre suas mazelas, assim como a abordagem da pessoa ao invés da doença, humaniza os futuros formados.

Dentre as contribuições desta, foi possível observar os resultados: amparar o maior número de pessoas em situação de vulnerabilidade; facilitar o processo de aprendizagem humanístico do discente; atendimentos colocando em prática seus conhecimentos teóricos; proporcionou o discente-monitor uma experiência de docente; proporcionou aos extensionistas um cunho de realidade sobre os atendimentos em um ambiente não controlado.

A realidade aponto negatividades quando a escolha dos locais de ações e sua burocracia para aceite do projeto. Os hospitais requeriam documentos de vacinação atualizada, documentos pessoais para cadastro dos discentes e preceptor, e carta de aceite da instituição de origem. A seleção de datas também gerou uma certa dificuldade pois precisavam ser em finais de semana ou dias em que os mesmos não estivessem em período letivo. Quando se obteve o aceite dos ambientes de saúde, os mesmos eram orientados a vestimenta adequada, coleta monetária para os materiais necessários, e conduta correta quanto principalmente os pacientes da pediatria. Ao longo do ano de 2019 foram ofertadas 15 ações, dividindo o grupo em dois e alternando os dias, sempre presentes os monitores e coordenador. Os monitores tiveram como papel monitoras as vivências dos grupos, observar detalhadamente suas capacidades de saber e intervir quando necessário, tirando dúvidas a fim de construir confiança.

IV. CONCLUSÃO

Atendendo aos objetivos propostos neste trabalho, ressalta-se as contribuições e a importância do mesmo quanto seu papel social, tanto para acadêmicos como para os pacientes. As ações de extensão buscam um diálogo entre universidade e comunidade, propondo a palhaçoterapia como tratamento e buscando o sorriso nos aplicados. Desta forma torna-se de suma importância o incentivo do desenvolvimento destas atividades para vivência dos alunos, monitores e preceptores, como também em curso de âmbito da saúde em geral.

Por isso, este trabalho torna-se relevante a sociedade de pesquisa, mediante aos resultados obtidos e podendo servir como cunho teórico a outros trabalhos a se desenvolverem, tendo em vista a metodologia ativa que interfere na graduação acadêmica.

REFERENCES

[1] Santos, J. H. de S., Rocha, B. F., & Passaglio, K. T. (2016). EXTENSÃO UNIVERSITÁRIA E FORMAÇÃO NO ENSINO SUPERIOR. *Revista Brasileira de Extensão Universitária*, *7*(1), 23–28. https://doi.org/10.36661/2358-0399.2016v7i1.3087

[2] Roberto, L., Curi, L., Monteiro, P., & Braga, V. (n.d.). http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=102551-pces608-18&category_slug=novembro-2018-pdf&Itemid=30192

[3] Catapan S de C, Oliveira WF de, Rotta TM. Palhaçoterapia em ambiente hospitalar: uma revisão de literatura. Ciência & Saúde Coletiva [Internet]. 2019 Sep;24(9):3417–29. Available from: https://www.scielo.br/pdf/csc/v24n9/1413-8123-csc-24-09-3417.pdf

[4] *APA PsycNet*. (n.d.). Psycnet.apa.org. https://psycnet.apa.org/record/2013-06406-011

[5] Silva, M. R. da, Marques, M. C. da C., Penha, A. V. X., & Caires, S. (2022). Comportamentos construídos e disseminados no palhaço de hospital. *Ciência & Saúde Coletiva*, *27*(6), 2449–2458. https://doi.org/10.1590/1413-81232022276.13902021

[6] Durán González, A., & José De Almeida, M. (n.d.). *OPINIÃO OPINION*. Retrieved March 21, 2024, from https://www.scielo.br/j/csc/a/3Mz3yMSy6FbZjLwp5Kmq5vM/?format=pdf&lang=pt Philip, E., & Philip, A. (2022). The influence of positive self-affirmation towards Malaysian ESL students at tertiary level of Education. *Journal of Humanities and Education Development*, *4*(4), 09-17. doi:10.22161/jhed.4.4.2

[7] Tavares, J. S., Oliveira, F. R. de, Maia, C. M. A. F. G., & Rodrigues, W. F. G. (2017). Contribuições da monitoria de anatomia humana na formação acadêmica de estudantes de enfermagem: relato de experiência. *Revista de Enfermagem UFPE on Line*, *11*(8), 3176–3179. https://doi.org/10.5205/1981-8963-v11i8a110225p3176-3179-2017

[8] Turci, D. A., Santos, C. A. dos, Aquino, E. R. de J., Souza, R. R. de, Queiroz, L. F. G., & Fragoso, E. M. (2023). Um projeto de extensão em pedagogia hospitalar e o COVID-19. *Revista Brasileira de Extensão Universitária*, *14*(3), 241–250. https://doi.org/10.29327/2303474.14.3-3

International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-11, Issue-3; Mar, 2024
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.113.4

Analysis of the Influence of Systemic Arterial Hypertension and Heart Failure on the Adversement of the Clinical Condition of Patients with Chronic Kidney Disease: A Literature Review

Amanda Lima Franco[1], Rafaela Vieira Saggin[2], Luiza Bastos Campos[3], Wendell Christian Cavalcante Gomes[4], André Angelo Tavares Favacho[5], Renan Willian Costa Silva[6], Brendo Silva Gaia Farias[7], Ana Maria Ferreira Cruz Toledo[8]

[1,5]Centro Universitário Metropolitano da Amazônia. UNIFAMAZ – Belém Pará.
[6 a 7] Universidade Federal do Pará – Belém
[8]Pontifica Universidade Católica do Goiás. PUC – Goiania.

Received: 30 Jan 2024,

Receive in revised form: 11 Mar 2024,

Accepted: 19 Mar 2024,

Available online: 27 Mar 2024

Keywords*— *systemic arterial hypertension, heart failure, literature review, chronic kidney disease, outcomes.

***Abstract*—** *Introduction: systemic arterial hypertension (SAH) and heart failure, epidemiologically, are diseases that model consequences for other systems of the human body, for example chronic kidney disease (CKD). The development of this appears to be a social consequence of lack of knowledge, as its secondary outcomes are controllable and treatable. Countries like Brazil have exorbitant expenses when it comes to financing dialysis and transplant procedures, with an increase in these numbers, especially in young patients decompensated for their underlying diseases. The objective of this work is to observe the incidence in the literature of SAH and heart failure in patients related to the worsening of CKD. Methodology: descriptive study in narrative review, which seeks to answer the PICO acromion "What is the influence of systemic arterial hypertension and heart failure on the worsening of the clinical condition of patients with chronic kidney disease?". Discussion: CKD's pathophysiology is the loss of kidney function, where they lose functionality and destroy their specific cells, resulting in the inability to maintain metabolic balance. It proves to be a problem of public responsibility, where more and more deaths in the population are reported. The main risk factors for CKD are highly prevalent chronic diseases such as hypertension and heart failure, the first being the most described in the literature as a triggering factor. Thus resulting in worsening of renal function laboratory results, resulting in chronic kidney injury (CRF). Results: Analyzing the databases, articles in the last 10 years were observed, where 38.6% had the descriptors systemic arterial hypertension and heart failure, describing them as their main secondary outcome. Conclusion: to the scientific society, it contributes summarized and updated indexes reporting the relationship between these precursor pathologies. To society, it informs the problem and a way to inform the patient about their health condition and better understanding.*

I. INTRODUÇÃO

Em todo mundo, a Doença Renal Crônica (DRC) vem crescendo devido ao aumento da incidência de hipertensão arterial sistêmica (HAS), diabetes, câncer de próstata, entre outras patologias. Assim, há um crescente desenvolvendo de insuficiência renal, por falta de conhecimento sobre as patologias de base e devido à falta de acompanhamento médico adequado, prejudicando a detecção precoce dessas doenças causando a DCR[1]. O desenvolvimento da desta está frequentemente associado à hipertensão e diabetes mellitus, sendo estas as principais causas de insuficiência renal crônica em pacientes no programa de diálise[2].

Nos países subdesenvolvidos e em desenvolvimento, o padrão de morbimortalidade por DRC vem se alterando devido à transição das doenças infecciosas para as doenças crônicas não transmissíveis. Nos EUA, projeções da prevalência desta para os anos de 2020 a 2030 em indivíduos maiores de 30 anos, estimam que a doença passará de 13,2% em 2010 para 14,4% em 2020 e para 16,7% em 2030[3].

A disfunção renal também é comum em pacientes com insuficiência cardíaca (IC), de prevalência superior à da população geral e que aumenta à medida que a taxa de filtração glomerular estimada (TFGe) diminui[4]. Sua prevalência varia entre 20 e 57% em pacientes com IC crônica estável e entre 30 e 67% em grandes registros de pacientes internados com IC aguda ou descompensada. Portanto, danos agudos ou crônicos em um dos órgãos, coração ou rim, podem levar a danos agudos ou crônicos no outro, produzindo o que passou a ser chamado de síndrome cardiorrenal. Além disso, a piora da função renal ocorre em 18-40% dos pacientes durante a hospitalização.

A DRC é uma condição associada a alto risco de doenças cardiovasculares, assim, a associação da HAS e/ou IC com a DRC acelera a progressão da lesão renal para DRC terminal, elevando também a mortalidade cardiovascular. Assim, este projeto tem o objetivo de analisar a influência da hipertensão arterial sistêmica e da insuficiência cardíaca no agravo do quadro clínico de pacientes com doença renal crônica. Salienta-se a importância desta temática devido às altas taxas de incidência e prevalência da DRC no Brasil e no mundo, sendo que o seu conhecimento pode fornecer informações que serão úteis na abordagem precoce e intervenção no tratamento de pacientes renais, podendo incentivar novos estudos sobre o tema.

II. METODOLOGIA

Os dados qualitativos coletados dos estudos selecionados foram avaliados, organizados em resultados. Este estudo trata-se de uma revisão integrativa da literatura, não havendo necessidade de apreciação do comitê de ética (CEP) pela sua conformação. Foi-se elaborado a partir da questão do acrônimo PICO (Paciente, intervenção, comparação e desfechos), orientando a construção da pergunta de pesquisa e a busca dos artigos. Assim, a questão de pesquisa PICO formulada para este estudo foi: "Qual é a influência da hipertensão arterial sistêmica e da insuficiência cardíaca no agravo do quadro clínico de pacientes com doença renal crônica?".

A estratégia de busca foi conduzida nas seguintes bases de dados: National Library of Medicine (PUBMED), Scientific Electronic Library Online (SciELO). Serão aplicados os Descritores em Ciências da Saúde (DeCS) em inglês: "Renal Insufficiency, Chronic", "Kidney Failure, Chronic", "Hypertension", "Heart Failure". Estes DeCS foram pesquisados com o operador booleano "AND". Para seleção da amostra, realizando-se um recorte temporal de 2013 a 2023, capturando as produções mais recentes nos últimos 10 anos.

Dentre os critérios de inclusão, estão: espaço temporal de publicação de 2013 a 2023; publicações que contenham informações sobre a abordagem geral da DRC e a sua relação a HAS e a IC; artigos disponíveis na íntegra publicados nos idiomas português, inglês e espanhol; com adultos na faixa etária de 19 anos ou mais. Em relação aos critérios de exclusão: os artigos repetidos serão excluídos do estudo, assim como aqueles que não responderem à questão norteadora do estudo.

III. DISCUSSÃO

A Sociedade Brasileira de Nefrologia define a DRC como uma perda lenta, progressiva e irreversível das funções renais, condição na qual os rins não apresentam mais funcionalidade e destroem os néfrons, resultando na incapacidade do organismo em manter o equilíbrio metabólico e hidroeletrolítico renal[5]. Nesta, tem-se a inclusão de pacientes com marcadores de lesão renal (albuminúria, razão albumina e creatinina maior que 3 mg/mmol, sedimentos urinários anormais, distúrbios eletrólitos e outras anormalidades quanto a lesão tubular detectada por histologia) e aqueles com TFGe menor que 60ml/min/1.73m2, em duas medições em um período de 90 dias, com ou sem presença de marcadores de lesão renal[6].

Na DRC, os estágios 1 a 5, apresenta-se em 14,3% na população geral e 36,1% em grupos de risco. No Brasil, os estágios 3 a 5 mostra-se prevalente em adultos, triplicando-se em indivíduos acima de 60 anos. Nesse contexto, em 2017 obteve-se 1,2 milhões de óbitos, assumindo a 12º posição de morte no mundo. Estima-se estatisticamente que a maior parte dos pacientes que

evoluíram a óbito prematuro não obtiveram acesso à terapia renal substitutiva (TRS), com maiores taxas de óbitos em países de baixa e média renda[7]. Desta forma, esta é um problema de saúde mundial, tendo no Brasil um custo alto no tratamento e reduzindo seu prognóstico. Na atualidade, estima-se 120000 pessoas estão em tratamento dialítico, resultando em valores de 1,4 bilhões de reais ao estado[8].

Os principais fatores de risco para a DRC são doenças crônicas de alta prevalência na atenção primaria do SUS como: hipertensão, diabetes e doenças cardiovasculares. A Hipertensão Arterial Sistêmica (HAS) foi identificada como um dos mais importantes fatores de risco para DRC, pois a idade avançada, aumento da rigidez vascular, atividade do sistema nervoso simpático e não adesão ao tratamento medicamentoso, pioram os resultados laboratoriais de função renal, necessitando um rastreio de proteinúria e aumentar as estratégias de identificação precoce da injúria renal crônica (IRC). Entretanto, há uma escassez de estudos que comprovem a meta ideal de pressão arterial para prevenir o decaimento da função renal, além disso, existem divergências sobre o tratamento intensivo da pressão[9].

Já a insuficiência cardíaca (IC) é uma síndrome clínica complexa de alterações estruturais ou funcionais no coração e caracterizada por sinais e sintomas típicos decorrentes da redução do débito cardíaco e altas pressões de enchimento do ventrículo esquerdo em repouso ou com esforço físico. O comprometimento renal é comum em pacientes com IC, com diminuição da filtração glomerular estimada a IC aumenta[10]. O tratamento objetiva melhorar os sintomas e a qualidade de vida dos pacientes, reduzindo as hospitalizações e a mortalidade. A grande maioria dos pacientes com DRC apresenta doença cardíaca e é comum que o no seu avanço desencadeamento da diálise, apresentando incidência de IC duas vezes maior que a de pacientes sem insuficiência renal[10].

IV. RESULTADOS

Analisando as bases de dados, SCIELO e PUBMED, com os descritores escolhidos em inglês: “Renal Insufficiency, Chronic”, “Kidney Failure, Chronic”, “Hypertension”, “Heart Failure”, encontraram-se 350 artigos. Dentre eles 289 foram na Scielo e 61 na Pubmed. Muitos deles, principalmente no primeiro site citados, eram duplicados, sendo em torno 24%. No total, 63,4 % destes não se enquadravam diretamente na metodologia do trabalho, ultrapassando a delimitação de conteúdo como datas selecionadas e temática de doenças base da DRC, sendo excluídas das referências.

Assim, 12,5% dos trabalhos representaram 44 artigos analisados e escolhidos como base. Os mesmos atendiam a todas as demandas de tempo e espaço e descreviam as doenças de base como fatores de risco a DRC. A partir destes artigos, observou-se que 38,6% tinham os descritores hipertensão Arterial sistêmica e insuficiência cardíaca. 11,3% apresentavam apenas HAS investigada e 38,8% diziam insuficiência cardíaca ou eventos cardiovasculares presentes na DRC. Por outro lado, a utilização do procedimento de diálise apresentava-se em 83,4% da pesquisa, assim como principais medicações de aumento de sobrevida, transplante de rim e coração, interação com outros sistemas do corpo humano, condutas paliativas, e outras doenças como fisiopatologia da DRC.

Como observado nesta revisão, os principais desfechos em pacientes com DRC são as suas complicações, como: acidose metabólica, anemia, desnutrição e alteração do metabolismo mineral, decorrentes da perda funcional renal, óbito, por causas cardiovasculares. Principalmente a descrição da necessidade da diálise e suas diferentes abordagens e quais os parâmetros clínicos de um paciente se tornar elegível a transplante de rins.

V. CONCLUSÃO

Os riscos que envolvem uma revisão de literatura integrativa, incluem-se: artigos com taxas para pagar que não são de acesso livre, artigos duplicados disponíveis em mais de uma base de dado indexada, o que pode reduzir o número total final de artigos selecionados para os resultados. Como benefícios da pesquisa, espera-se contribuir a sociedade científica sobre a temática da hipertensão arterial, insuficiência cardíaca e doença renal crônica, atualizando o paciente sobre a sua condição de saúde e melhor compreensão das patologias envolvidas.

REFERENCES

[1] LOPES, Isabella Katarina Pinto; FIGUEIREDO, Samuel Soares; NUNES, Ronaldo Lima. DOENÇA RENAL CRÔNICA E O PROCESSO DE HEMODIÁLISE. Revista IberoAmericana de Humanidades, Ciências e Educação, v. 8, n. 8, p. 706-717, 2022.

[2] BESSA, J.W.L.; BRILHANTE, F.D.F.; BORGES, G. de O.; BESSA, J.L.; KOSTAKIS, M.E.G.; SOUSA, P.D. de O. de; SILVA, P.A.E; SOUZA, T.S. de; FONTENELLE, V.T. de M.; TAVARES, R. de O.M. Abordagem geral da doença renal crônica e sua relação com a hipertensão arterial sistêmica: uma revisão integrativa. Revista Eletrônica Acervo Médico, v. 1, n. 1, p. e8904, 28 set. 2021.

[3] AMARAL, Thatiana Lameira Maciel et al. Prevalência e fatores associados à doença renal crônica em idosos. Revista de Saúde Pública, v. 53, 2019.

[4] GÓRRIZ, Jose Luis; RICO, Miguel González; NUÑEZ, Julio. Tratamiento de la insuficiencia cardiaca en el paciente con insuficiencia renal avanzada. Revista Española de Cardiología Suplementos, v.18, s.B, p. 31-39, 2019.
[5] SILVA, Saulo Freitas da et al. Fisioterapia durante a hemodiálise de pacientes com doença renal crônica. Brazilian Journal of Nephrology, v. 35, p. 170-176, 2013.
[6] LEITE, Larissa Parada et al. Hipertensão na doença renal crônica em tratamento conservador. Rev. Bras. Hipertens., v. 27, n. 4, p. 115-21, 2020.
[7] SILVA, Saulo Freitas da et al. Fisioterapia durante a hemodiálise de pacientes com doença renal crônica. Brazilian Journal of Nephrology, v. 35, p. 170-176, 2013.
[8] RIBEIRO, WA; JORGE, BO; QUEIROZ, RS. Repercussões da hemodiálise no paciente com doença renal crônica: uma revisão da literatura. Revista Pró-UniverSUS. v.11, n.1, p.88-97, 2020.
[9] FERREIRA, Fernando José Gomes et al. O uso dos inibidores SGLT2 na melhora do desfecho clínico em pacientes com insuficência cardíaca e doença renal crônica: uma revisão sistemática. Revista Eletrônica Acervo Saúde, v. 23, n. 2, p. e11571-e11571, 2023.
[10] HEERSPINKHJL, et al. Dapagliflozin in patients withchronic kidney disease. New England Journal of Medicine, v.383, n.15, p.1436-1446, 2020.
[11] MARQUES, J. et al. The burden of coronary heart disease in simultaneous pancreas-kidney transplantation: coronary angiography as a diagnostic method for all? – a retrospective study. J Bras Nefrol., v. 44, n. 4, p. 522–526, 28 fev. 2022.
[12] LESSA DA COSTA, R. et al. Lesão renal aguda em pacientes com Covid-19 de uma UTI no Brasil: incidência, preditores e mortalidade hospitalar Acute kidney injury in patients with Covid-19 in a Brazilian ICU: incidence, predictors and in-hospital mortality Autores. [s.d.].
[13] LUIS, G. et al. ComuniCação Breve | Brief CommuniCation. [s.d.].
[14] NAKHOUL, G. N. et al. Serum potassium, end stage renal disease and mortality in chronic kidney disease. American journal of nephrology, v. 41, n. 6, p. 456–463, 2015.
[15] SATA, Y.; SCHLAICH, M. P. The Potential Role of Catheter-Based Renal Sympathetic Denervation in Chronic and End-Stage Kidney Disease. Journal of Cardiovascular Pharmacology and Therapeutics, v. 21, n. 4, p. 344–352, 6 jan. 2016.
[16] LAFFIN, L. J.; BAKRIS, G. L. Intersection Between Chronic Kidney Disease and Cardiovascular Disease. Current Cardiology Reports, v. 23, n. 9, 16 jul. 2021.
[17] OUP accepted manuscript. Stem Cells Translational Medicine, 2022.
[18] PECO-ANTIĆ, A.; PARIPOVIĆ, D. Renal hypertension and cardiovascular disorder in children with chronic kidney disease. Srpski arhiv za celokupno lekarstvo, v. 142, n. 1-2, p. 113–117, 2024.
[19] WHITTAKER, C. F. et al. Medication Safety Principles and Practice in CKD. Clinical Journal of the American Society of Nephrology, v. 13, n. 11, p. 1738–1746, 18 jun. 2018.
[20] LESSA DA COSTA, R. et al. Lesão renal aguda em pacientes com Covid-19 de uma UTI no Brasil: incidência, preditores e mortalidade hospitalar Acute kidney injury in patients with Covid-19 in a Brazilian ICU: incidence, predictors and in-hospital mortality Autores. [s.d.].
[21] LESSA DA COSTA, R. et al. Lesão renal aguda em pacientes com Covid-19 de uma UTI no Brasil: incidência, preditores e mortalidade hospitalar Acute kidney injury in patients with Covid-19 in a Brazilian ICU: incidence, predictors and in-hospital mortality Autores. [s.d.].
[22] LUIS, G. et al. ComuniCação Breve | Brief CommuniCation. [s.d.].
[23] AUGUSTO, M. et al. Breve ComuniCação | Brief CommuniCation Razão albumina/proteína em amostras isoladas de urina para análise da seletividade de proteinúria na doença renal crônica Albumin-to-protein ratio in spot urine samples for analysis of proteinuria selectivity in chronic kidney disease Autores. [s.d.].
[24] HALIM, A. et al. FGF23 and Cardiovascular Structure and Function in Advanced Chronic Kidney Disease. Kidney360, v. 3, n. 9, p. 1529–1541, 29 set. 2022.
[25] DALRYMPLE, L. S. et al. Chronic Kidney Disease and the Risk of End-Stage Renal Disease versus Death. Journal of General Internal Medicine, v. 26, n. 4, p. 379–385, 19 set. 2010.
[26] LASH, J. P. et al. Chronic Renal Insufficiency Cohort (CRIC) Study: Baseline Characteristics and Associations with Kidney Function. Clinical Journal of the American Society of Nephrology : CJASN, v. 4, n. 8, p. 1302–1311, 1 ago. 2009.
[27] BORG, R. et al. Chronic kidney disease in primary care: risk of cardiovascular events, end stage kidney disease and death. BMC primary care, v. 24, n. 1, p. 128, 21 jun. 2023.
[28] RAVERA, M. et al. Impaired Left Ventricular Global Longitudinal Strain among Patients with Chronic Kidney Disease and End-Stage Renal Disease and Renal Transplant Recipients. Cardiorenal Medicine, v. 9, n. 1, p. 61–68, 28 nov. 2018.
[29] BARROWS, I. R.; RAJ, D. S. Janus Face of Coronary Artery Disease and Chronic Kidney Disease. Journal of the American Heart Association, v. 5, n. 4, 1 abr. 2016.
[30] IMAI, E. et al. Chronic Kidney Disease Japan Cohort study: baseline characteristics and factors associated with causative diseases and renal function. Clinical and Experimental Nephrology, v. 14, n. 6, p. 558–570, 11 ago. 2010.

International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-11, Issue-3; Mar, 2024
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.113.5

"Determinants of Digital Financial Inclusion and its Impact on Micro Enterprises" Ease of doing Business, A Comprehensive Review

Shabeena[1], Iqra Ashiq[2], Chaudhary Saud Ur Rehman[3]

[1]Department of Commerce, The University of Haripur, Pakistan
[2]Department of Business and Economy, Chongqing University
[3]Department of Civil Engineering, Chongqing University

Received: 03 Feb 2024,

Receive in revised form: 14 Mar 2024,

Accepted: 22 Mar 2024,

Available online: 31 Mar 2024

***Keywords*— Digital Financial Inclusion, Micro-Enterprises, Ease of Doing Business, Logistic Regression, World Bank Enterprises Survey, India**

***Abstract*—** *This study investigates the determinants of digital financial inclusion (DFI) and its influence on the ease of doing business for micro-enterprises in India. Utilizing data from the World Bank's Enterprises Survey of Micro Firms (ESM) 2022, encompassing 998 micro-enterprises, this research examines variables related to access and usage of digital finance. Two categories of independent variables—digital resource capability and firms' and owners' characteristics—are analyzed as explanatory factors for DFI. Analysis of variance (ANOVA) is employed to assess disparities in perceived business obstacles between micro-enterprises with and without access to digital finance. Additionally, a logistic regression model is constructed to identify the determinants of DFI. The findings indicate that DFI plays a significant role in mitigating business obstacles related to regulation and market externalizes for micro-enterprises. Moreover, results from both the logistic regression model and marginal effects estimation underscore the importance of factors such as internet access, educational attainment, and owner experience in facilitating DFI among micro-enterprises. This study offers insights valuable to diverse stakeholders including government agencies, entrepreneurship promoters, financial institutions, and international organizations involved in promoting digital financial inclusion. A noteworthy contribution of this study lies in its examination of the determinants of DFI specifically among micro-enterprises in India, leveraging extensive and specific data collected by the World Bank. By shedding light on this under explored area, the study contributes to a deeper understanding of the factors influencing digital financial inclusion in the context of India's micro-enterprise sector.*

I. INTRODUCTION

In the contemporary global economy, digital financial inclusion (DFI) stands as a crucial instrument for fostering economic empowerment and fostering inclusive growth, particularly for micro-enterprises. Micro-enterprises, often operating in resource-constrained environments, play a significant role in driving economic activity, job creation, and innovation. However, they frequently face numerous challenges, including limited access to formal financial services, which can impede their growth and sustainability.

Against this backdrop, understanding the determinants of digital financial inclusion and its impact on micro-enterprises' ease of doing business is of paramount importance, especially in the context of emerging

economies like India. Digital financial inclusion refers to the provision of low-cost digital access to financial services, aiming to reach economically underserved populations and businesses. It encompasses the utilization of digital platforms, such as mobile banking apps, internet banking, and digital wallets, to facilitate financial transactions and access to credit, savings, insurance, and other financial services.

India, with its vast and diverse micro-enterprise sector, presents a compelling case study for examining the dynamics of digital financial inclusion. The country has witnessed significant strides in digital transformation, spurred by government initiatives and technological advancements. Initiatives like the Unified Payments Interface (UPI) and Aadhaar-linked banking have revolutionized the financial landscape, making digital financial services more accessible and inclusive.

Despite these advancements, challenges persist, particularly concerning the effective integration of micro-enterprises into the digital financial ecosystem. Identifying the determinants that influence digital financial inclusion among micro-enterprises and understanding its impact on their ease of doing business is crucial for devising targeted interventions and policy frameworks to address existing barriers and promote inclusive growth.

This study aims to fill this gap by investigating the determinants of digital financial inclusion among micro-enterprises in India and assessing its impact on their ease of doing business. Drawing upon data from the World Bank's Enterprises Survey of Micro Firms, this research will analyze factors such as access to digital finance, digital resource capability, and firm and owner characteristics. Furthermore, it will explore how digital financial inclusion influences micro-enterprises' ability to overcome business obstacles and enhance their operational efficiency.

By providing empirical evidence and insights into the relationship between digital financial inclusion and micro-enterprises' ease of doing business, this study seeks to inform policymakers, financial institutions, and other stakeholders about the importance of fostering an inclusive digital financial ecosystem. Ultimately, the findings of this research can guide the design and implementation of strategies to promote digital financial inclusion, unlock the potential of micro-enterprises, and drive sustainable economic development in India and beyond.

II. LITERATURE REVIEW

Access to financing has been linked to business performance, according to a number of earlier studies (Rokhim et al. 2021; Osano and Languitone, 2016). Anthanasius Fomum and Opperman (2023) assert that financial inclusion raises the likelihood that microenterprises will be categorized as emerging and developed businesses. According to Yang and Zhang (2020), the macroeconomy as well as small and micro businesses can benefit from the restructuring of the financial sector and the promotion of digital financial inclusion. According to Yang and Zhang (2020), the macroeconomy as well as small and micro businesses can benefit from the restructuring of the financial sector and the promotion of digital financial inclusion.

According to Singh et al. (2014), digital financial inclusion is thought to be a workable way to alleviate the barriers that the impoverished have in obtaining financial services. They state that the adoption of digital financial inclusion facilitated the flow of capital, encouraged entrepreneurship, and removed barriers encountered by enterprises. The main cause of the scarcity of profitable opportunities is inadequate financial accessibility, or the inability to quickly obtain capital. Most previous studies have examined how DFI affects the degree of income inequality (Neaime and Gaysset, 2018; Zhang et al., 2020). According to Yang et al. (2022), women are empowered by digital financial inclusion because it encourages entrepreneurship among underprivileged women, especially those with lower levels of education or financial independence.

Digital financial inclusion does not increase listed companies' total factor productivity (TFP), according to Chen et al. (2022b). However, digital financial inclusion can significantly increase the productivity of listed companies in large cities with concentrated financial resources. SMEs believe that the biggest barrier to their expansion is getting access to financing (Wang, 2016). In China, digital finance has expanded quickly, promoting inclusive finance but also posing operational, policy, default, and fraud issues (Song and Dong, 2020). With different incentives for businesses with different property rights and geographic locations, digital finance lowers financing constraints and boosts SME innovation (Yao and Yang, 2022).

According to Beck (2013), financial deepening facilitates company entry into the market, encourages entrepreneurship, and lowers funding barriers for SMEs. It also enhances resource allocation. DFI has improved equity and efficiency by expanding financial inclusion, which has positive externalities and benefits the economy and society. It improves the efficiency and accessibility of financial services. Massive amounts of e-commerce data can be leveraged by digital finance institutions to lower the cost of financial transactions, enhance financial inclusion (Frost et al. 2019), and stimulate economic growth and

entrepreneurship (Xie et al. 2018). Oz-Yalaman (2019) came to the conclusion that business regulation and financial inclusion boost tax receipts. According to Bai et al. (2021), MSEs are using digital finance for business transactions like purchasing and paying bills and taxes.

Nonetheless, there is more research done on the subject of how digital financial inclusion affects business accessibility. The literature mentioned above has led to the formulation of the following research hypotheses:
H1: The micro firms with and without access to digital finance perceive the same business obstacles.
H2: The perceived business barriers by microfirms using and not using digital finance are the same.

Digital financial inclusion and the ability to use digital resources

Scholarly discussions have focused on the impact of information and communication technologies (ICT) (Aziz and Naima, 2021; Karakara and Osabuohien, 2020; Niebel, 2018; Hong, 2017). According to existing research, ICT, particularly mobile phones and the internet, have become important innovations in the financial sector (Suryono et al. 2020; Pradhan et al. 2017). Diniz et al. (2012) state that the use of ICT has made it easier for banking services to be widely distributed and for banks to operate close to one another. The introduction of information technology into the banking sector has fundamentally changed how financial services are provided to clients around the globe (Chavan, 2013).

Technology innovations like automated teller machines (ATMs), internet banking, mobile banking, digital banking kiosks, and the Unified Payments Interface (UPI) have revolutionized the banking industry, according to Gupta and Arya (2019) and Sarkar (2016). These developments have brought about new mechanisms that have the potential to greatly improve banks' capacity to provide more effective and efficient customer service. The current third technological revolution, which is centered on the internet, is having a major impact on the fairness and efficiency of banking, according to Liu et al. (2020). China has benefited greatly from the internet revolution since it has made it easier for the nation's digital economy and digital finance to flourish.

Over the past ten years, China's digital economy—particularly in the field of digital finance—has grown significantly as a result of the country's adoption of cutting-edge technologies like big data and cloud computing (Yin et al. 2019; Gabor and Brooks, 2017). More than 80 countries offer digital banking services that can be accessed through mobile phones, according to the World Bank (The World Bank, 2020; Chu, 2018). Many people use mobile phones and other digital tools, such as artificial intelligence (AI), which increases the number of people who can access banking services (Salampasis and Mention, 2018). People can obtain financial services at a cost and in a manner that suits them best with digital financial inclusion (Gomber et al. 2017).

Indian citizens' standard of living has increased as a result of the use of digital technologies in financial transactions (Malladi et al. 2021). The research hypothesis on digital resources and financial inclusion is as follows, based on the literature mentioned above:
H3: The digital financial inclusion of microfirms is not considerably impacted by digital resource capability.

Features of businesses, owners, and digital financial inclusion

The way a firm responds to financial challenges is likely to depend on its characteristics (Tuffour et al., 2022; Wakaisuka-Isingoma et al., 2016; Yildirim et al., 2013). According to Nguli and Odunga's (2019) empirical research, the size and age of a firm have a positive and negative impact on financial inclusion. Several scholarly investigations have determined that a person's socioeconomic status is a crucial determinant of their ability to obtain financial services (Barcellos and Zamarro, 2021; Kulkarni and Ghosh, 2021; Nandru et al. 2015; Izquierdo and Tuesta, 2015). According to research by Frempong (2009) and Izquierdo and Tuesta (2015), education is a key factor in promoting financial inclusion in businesses. According to Smallbone and Wyer (2000), education enhances exploration, communication, and foresight and inspires entrepreneurs.

According to Miao et al. (2017), seasoned business owners apply their knowledge and skills to help others complete challenging assignments. Dittmar and Duchin (2016) pointed out that seasoned managers make good use of financial resources. The financial affairs of the company are significantly impacted by the experience of the CEO (Matemilola et al. 2018). Experience in business enhances the links between the acquisition of intellectual capital and resources and improves resource acquisition. The effectiveness of financial literacy can also be influenced by individual characteristics, such as the age, education, and experience of managers (Barcellos and Zamarro, 2021; Garg and Singh, 2018). This has important implications for financial inclusion (Goyal and Kumar, 2021; Schuetz and Venkatesh, 2020). One of the main reasons women don't launch or formalize their businesses is a lack of access to capital and information networks (Xheneti et al. 2019).

One of the main reasons women don't launch or formalize their businesses is a lack of access to capital and information networks (Xheneti et al. 2019). According to certain studies (Brixiová and Kangoye, 2016; Demirgüç-Kunt et al., 2013;

and Brush and Cooper, 2012), women have difficulty obtaining loans from outside sources. The literature reviewed above has led to the formulation of the following research hypothesis:

H4: The digital financial inclusion of micro firms is not significantly impacted by the characteristics of the firms or the owners.

III. RESEARCH METHODOLOGY

Source of data and survey instrument

The World Bank's Enterprises Survey of Micro Firms (ESM) 2022 served as the study's foundation. Between December 2021 and March 2022, data was gathered. Nielsen (India) Pvt. Ltd. conducted a survey on behalf of the World Bank to examine the microenterprise business environment and demographics. In accordance with the Factory Act and the Shop and Establishment Act, information was gathered from businesses with fewer than five employees who were registered with a government agency. The universe table of micro-enterprises was derived from the sixth economic census of India. The Enterprises Survey of Micro Firms (ESM) 2022 covers the following major sectors: manufacturing, construction, retail and wholesale trade, transportation and storage, lodging, and food services. Data from the survey were gathered through stratified random sampling, with 998 micro-enterprises participating in the survey. Two steps were involved in the data collection process. First, to establish eligibility and arrange interviews, phone calls were placed to the micro-enterprises. The owners, managers, and directors of the companies were then interviewed in person. See the Enterprise Survey, Micro, India 2022 implementation report (ESM, 2022) for more details. A structured questionnaire has been used for data collection. The questionnaire covers a wide range of factors related to the firm's business operations, perceived business barriers, and demographic profile.

The regression model contains two sets of explanatory variables: firms' and owners' characteristics and digital resource capability. The characteristics of firms and owners have been gathered in relation to the age of the firm, the owner's gender, the size of the household, and the level of education. The ability to access the internet and use computers or tablets is considered a capability of digital resources. Table 1 provides a thorough explanation of the variables, measurement scales, and descriptive statistics.

Table 1

Variables	**Definition**	**Measurement scale**	***N***	**Min**	**Max**	**Mean**	**S.D**
Access to digital financial system	Uses Digital Payments	Yes = 1, No = 0	934	0	1	0.66	0.48
Usage of digital financial system	Uses Digital Payments To Pay Utiility Bills	Yes = 1, No = 0	614	0	1	0.85	0.36
Explanatory Variables							
Computers or Tablet	Establishment Presently Uses: Computers or Tablet	Yes = 1, No = 0	935	0	1	0.43	0.50
Internet	Establishment Presently Uses: Internet	Yes = 1, No = 0	936	0	1	0.75	0.43
Firms Age	Age of the firms	Years in numbers	898	2	68	14.11	9.66
Gender of the owner	Amongst The Owners of The Firm, Are	Yes = 1, No = 0	934	0	1	0.09	0.28

	There Any Females?						
Education level of Owner	Owner Highest Level of Completed Education						
Primary school or less	Primary school or less		922	0	1	0.18	0.38
Secondary	Secondary	Yes = 1, otherwise = 0	922	0	1	0.33	0.47
Diploma	Diploma	Yes = 1, otherwise = 0	922	0	1	0.34	0.48
Bachelors or above	Bachelors or above	Yes = 1, otherwise = 0	922	0	1	0.14	0.35
Households Size	Num. of Household Members of The Owner	Number	864	0	32	4.91	1.92
Managers Experience	How Many Years of Experience Working In This Sector Does The Top Manager Have?	Years in numbers	896	1	50	12.34	9.23

Analytical Approach

SPSS 22 has been used to digitize the ESM 2022 data. The data have been analyzed using basic statistical methods like logistic regression, chi-square testing, and descriptive statistics. To comprehend the nature of the variables extracted for the study, descriptive statistics are utilized. The difference between perceived business barriers by microenterprises with and without access to digital finance, as well as between users and non-users of digital finance, has been measured using analysis of variance (ANOVA). To investigate the factors influencing microenterprises' adoption of digital finance, a binary logistic regression model has been employed. Since the dependent variables are measured on a binary scale, a binary logistic regression model has been applied. Two dependent variables were used in this study: (1) the use of digital payments to gauge how accessible digital finance is. (2) Due to the apparent benefits of digital finance, utilizes digital payments to settle utility bills. where 0 denotes the absence of digital financial inclusion and 1 denotes digital financial inclusion.

IV. RESULTS AND DISCUSSIONS

Profile of sample micro enterprises

Of the total number of micro-enterprises, 66% have reported having access to digital financial systems; of the 998 enterprises, only 614 have reported using these systems, and 85% of micro-enterprises make payments via digital platforms. Seventy-five percent of micro enterprises have internet access, and forty-three percent of micro firms own computers or tablets. Firms have an average age of 14.11 years. Most of the companies are led by men. A diploma and a bachelor's degree or above are held by 14% and 34% of the owners, respectively. Managers in microenterprises have an average of 12.34 percent experience.

Digital financial inclusion and perceived business obstacles

The business environment in any economy is reflected by a number of factors, including market externalities, resource accessibility, and business regulation (Singh et al. 2023b; Khan et al. 2023; Boateng and Poku, 2019; Njiraini et al. 2018; and Ali, 2016). The ease of doing business by micro-enterprises has been measured in the current study using

perceived business obstacles related to access to resources (land, electricity, and other resources for the establishment's operations), market externalities (competitors' informal sector practices, corruption, theft, and disorder), and business regulation (tax rates, tax administrations, business licensing and permits, and labor regulations). The study examined the perceived barriers to business for microenterprises that have access to digital finance and those that do not. The difference in perceived business barriers with and without access to and use of digital finance has been measured using analysis of variance. Table 2 presents the findings. The results of the analysis of variance show that microenterprises with and without access to digital finance differ significantly in the way taxes are administered (F = 4.349, P = 0.037). The means indicate that the administration of taxes is a comparatively larger barrier for businesses without access to digital financing. Similarly, when it comes to business licenses and permits as a barrier, F-statistics show a difference between micro-enterprises with and without access to digital finance (F = 13.153, P = 0.000). Small businesses using digital finance reported fewer difficulties obtaining business licenses and permits. It suggests that obtaining business licenses and permits is facilitated by digital financial inclusion. It might be because, in comparison to other ways, paying fees online is now simpler. An analysis of variance (ANOVA) reveals that businesses with digital finance access perceive a significant reduction in labor regulation barriers (F = 10.303, P = 0.001). The analysis of variance reveals an intriguing result regarding the practice of competitors as an obstacle: micro-enterprises that have access to digital finance perceived it as a significantly greater obstacle (F = 4.393, P = 0.036) in comparison to firms that do not have access to digital finance. The F-test indicates a significant difference in perceptions of crime, theft, and disorder between firms with and without access to digital finance (F = 16.368, P = 0.000). This suggests that microenterprises with access to digital finance see significantly fewer obstacles in relation to crime, theft, and disorder than do firms without such access. By lowering the danger of carrying cash, digital finance lowers the possibility of money-related crimes. One significant finding that has been documented is that the accessibility of digital finance is impeded by land access. Microbusinesses with digital finance access see a lot more barriers to land access (F = 3.056, P = 0.081). As a result, part of hypothesis H1 is rejected. When compared to micro enterprises without access to digital finance, those with access to it reported facing fewer business challenges. Ease of doing business has also been investigated among and without users of digital finance. Analysis of variance conducted under business regulations reveals that there are significantly fewer barriers for micro users of digital finance when it comes to tax rate (F = 3.091, P = 0.079), tax administration (F = 4.902, P = 0.027), and business licensing and permits (F = 11.573, P = 0.001). Additionally, there are no issues with tax administration. Since the entire tax system is now online, microbusinesses that use digital finance ANOVA does not show statistically significant differences between digital finance users and non-users under the market externalities category. Digital finance users perceived significantly more barriers to land access for business purposes when it came to resource groups. The above result indicates a partial rejection of hypothesis H2.

Table 2

	Access to digital finance				**Usage of digital finance**			
	Yes	**No**	***F***	**Sig.**	**Yes**	**No**	***F***	**Sig.**
Access to resources								
Access to land	2.04	1.91	3.056***	0.081	2.07	1.85	3.023***	0.083
Access to electricity	2.05	2.06	0.025	0.876	2.02	2.2	2.661	0.103
Business regulation								
Tax rates	2.3	2.24	0.451	0.502	2.26	2.51	3.091***	0.079
Tax administrations	1.99	2.16	4.349**	0.037	1.95	2.24	4.902**	0.027
Business licensing and permits	1.92	2.19	13.153*	0.000	1.86	2.27	11.573*	0.001

Labor regulations	1.8	2.02	10.303*	0.001	1.79	1.84	0.231	0.631
Market externalities								
Practices of Competitors	2.21	2.06	4.393**	0.036	2.21	2.21	0.001	0.979
Corruption	2.09	2.17	0.952	0.329	2.09	2.09	0.003	0.957
Crime, theft and disorder	1.75	2.03	16.368*	0.000	1.73	1.87	1.696	0.193

Factors influencing digital financial inclusion in Indian microbusinesses

A logistic regression model and the marginal effect, which was determined using Stata software, have been used to calculate the factors that influence digital financial inclusion in terms of access and utilization. Table 3 provides estimates for the pseudo-R square, level of log-likelihood, marginal effect, regression coefficient, and significant level. As dependent variables, two aspects of digital financial inclusion—access to and usage of digital finance—have been identified. As a result, two regression models have been created to determine the factors that influence microenterprises' adoption of digital finance. Digital financial inclusion has been the subject of numerous studies conducted under various conditions (Lu et al., 2022; Tay et al., 2022; Yang and Zhang, 2020). Logistic regression estimates indicate that six variables—computers or tablets, internet, age of the firm, owner gender, owner education level, and household size—significantly influence access to digital finance out of a total of seven explanatory variables. Access to computers is less likely to affect access to digital finance, according to an unexpected finding indicated by the regression coefficient of computers or tablets, which is negative and significant ($\beta = -0.404$, $P = 0.029$). This could be because smartphones are typically used to access digital banking. Access to digital finance is positively and significantly impacted by the estimated regression coefficient of internet accessibility ($\beta = -1.539$, $P = 0.000$); this effect may be attributed to the requirement for digital resource capability in order to access the digital finance interface. It is not feasible to use and complete the digital payment process without internet access. As a result, part of the hypothesis H3 is rejected. It can be observed from the results that older firms are more likely to adopt digital financial inclusion, with a positive regression coefficient and marginal effects of a firm's age on access to digital finance ($\beta = 0.032$, $P = 0.004$). It might be because more established businesses have better infrastructure and resource capacities to manage the digital platform.

Gender significantly influences access to digital finance, according to the estimated regression coefficient ($\beta = -0.756$, $P = 0.066$); the value of marginal effects indicates that female-headed micro-enterprises are 14.8 percent more likely to access digital finance than male-headed micro-enterprises. Regarding the education categories, the results of the logistic regression analysis showed that education has a positive and significant impact on digital financial inclusion for both secondary education ($\beta = -0.551$, $P = 0.026$) and bachelor's degree or above ($\beta = -0.916$, $P = 0.005$). This indicates that micro-entrepreneurs' access to digital financial services is significantly influenced by their level of education. It might be as a result of education's ability to empower people and inform them of current developments. Thus, more access to digital finance is revealed by educated micro-entrepreneurs. Even though using or making payments online has gotten too easy, less educated people still don't trust digital transactions. Micro-entrepreneurs' access to digital finance was significantly and inversely impacted by the size of their household ($\beta = -0.129$, $P = 0.008$). According to the model summary, the regression model fits the data fairly well. The regression model's explanatory variables collectively and individually significantly explain the determinants of digital financial inclusion, as indicated by the chi-square value of 93.540. A good fit for a regression model is indicated by the negative value and the high log-likelihood value (−473.306).

V. CONCLUSIONS

The study delves into the determinants of digital financial inclusion (DFI) among micro-enterprises in India and its impact on ease of doing business, leveraging the comprehensive dataset from the World Bank's Enterprises Survey of Micro-Firms. Through analysis of variance, it is revealed that micro-enterprises equipped with digital finance experience fewer business obstacles compared to those without such access. Logistic regression analysis further elucidates that factors such as internet accessibility, firm age, owner gender and education level, and household

size significantly influence DFI. These findings underscore the importance of digital resource capability, owner characteristics, and internet accessibility in driving digital financial inclusion among micro-enterprises.

The study holds theoretical and managerial implications for various stakeholders, including promoters of micro-entrepreneurship, the banking industry, industrial economists, and the government. It emphasizes the necessity of encouraging micro-enterprises to adopt digital financial inclusion to overcome business barriers. Recommendations include promoting the benefits of digital finance through live demonstrations, implementing comprehensive policies for digital financial inclusion, and providing subsidized digital resources to micro-enterprises. Furthermore, the study underscores the importance of financial literacy programs, affordable digital financial services, regulatory measures, and integration with supply chains to enhance accessibility and transparency for micro-enterprises.

While the study offers valuable insights, it acknowledges limitations inherent in secondary data analysis and suggests avenues for future research. Future studies could incorporate theoretical models, utilize appropriate indicators, and explore behavioral aspects of micro-enterprise owners to enrich understanding of DFI determinants. Additionally, integrating in-depth interviews to explore characteristic variables of respondents could further enhance the comprehensiveness of research models. Overall, the study provides critical insights for creating an enabling environment for digital financial inclusion among micro-enterprises in India and beyond.

REFERENCES

[1] Ali J (2016) Performance of small and medium-sized food and agribusiness enterprises: evidence from Indian firms. Int Food Agribus Manag Rev 19(4):53–64

[2] Abor JY, Amidu M, Issahaku H (2018) Mobile telephony, financial inclusion and inclusive growth. J Afr Bus 19(3):430–453

[3] Anthanasius Fomum T, Opperman P (2023) Financial inclusion and performance of MSMEs in Eswatini. Int J Soc Econ 50(11):1551–1567

[4] Aziz A, Naima U (2021) Rethinking digital financial inclusion: Evidence from Bangladesh. Technol Soc 64:101509

[5] Bai C, Quayson M, Sarkis J (2021) COVID-19 pandemic digitization lessons for sustainable development of micro-and small-enterprises. Sustain Prod Consum 27:1989–2001

[6] Barcellos SH, Zamarro G (2021) Unbanked status and use of alternative financial services among minority populations. J Pension- Econ Financ 20(4):468–481

[7] Beck T (2013) Bank financing for SMEs–lessons from the literature. Natl Inst Econ Rev 225:R23–R38

[8] Boateng S, Poku KO (2019) Accessing finance among women-owned small businesses: evidence from lower Manya Krobo municipality, Ghana. J Glob Entrepreneurship Res 9:1–17

[9] Brixiová Z, Kangoye T (2016) Gender and constraints to entrepreneurship in Africa: New evidence from Swaziland. J Bus Venturing Insights 5:1–8

[10] Brush CG, Cooper SY (2012) Female entrepreneurship and economic development: An international perspective. Entrepreneurship Regional Dev 24(1-2):1–6

International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-11, Issue-3; Mar, 2024
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.114.1

Navigating the Dark Web of Hate: Supervised Machine Learning Paradigm and NLP for Detecting Online Hate Speeches

Njideka Nkemdilim Mbeledogu and Mishael Somtochukwu Ike-Okonkwo

Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria

Received: 03 Feb 2024,

Receive in revised form: 08 Mar 2024,

Accepted: 19 Mar 2024,

Available online: 30 Mar 2024

***Keywords*— Natural Language Processing, Tokenization, Logistic Regression, Hyperparameter**

***Abstract*—** *Many online platform's participants are worried about hate speeches that usually trigger cyberbully attitudes that dissuades users' interest in their platforms. The study investigates hate speech in online platforms using Natural Language Processing (NLP) techniques and supervised machine learning paradigm. It specifically focused on developing a robust model capable of classifying text as 'hateful' or 'non-hateful' accurately. The approaches applied included compiling a large dataset from multiple online textual sources; preprocessing the dataset through normalization, tokenization, stop-word removal, and lemmatization; advanced feature extraction techniques such as negation handling, n-gram analysis, and Term Frequency-Inverse Document Frequency (TF-IDF) to capture the intricacies of the textual material and the model implementation phase using Logistic Regression for its efficiency in binary classification problems. The model's performance was evaluated using metrics such as accuracy, precision, recall, F1-score and confusion matrix. The baseline performance of the model with default hyperparameters achieved a test accuracy of 93%. When optimized with hyperparameter tuning and cross-validation procedures to guarantee more generalizable performance, the model achieved an accuracy of 95%. The study concluded that NLP and logistic regression technique can effectively identify hate speeches.*

I. INTRODUCTION

The growing frequency of hate speech on online communication platforms poses a major danger to the digital age's guiding principles of inclusivity, tolerance, and respectful conversation. The internet's obscurity has given people the confidence to indulge in abusive language, thereby, establishing harm as normal way of life. Popular initiatives aimed at reducing hate speech frequently rely on manual content moderation. This approach is cumbersome, time-consuming, resource-intensive and biased. Yet more, the dynamic and ever-changing character of online conversation makes it difficult to effectively recognize and respond to hate speech in real time.

Addressing the issues of hate speech needs provident approach and adequate technologies that can quickly detect it. Based on this, the research aims to use Natural Language Processing (NLP) and machine learning techniques to create a hate speech detection and sentiment analysis system that automates the detection of hate speeches based on the emotional tones to improve the safety and civility of digital communication spaces.

II. LITERATURE REVIEW

Hate speech can be said to be expressions that belittle, extricate, or support both physical and emotional violence based on any socio-attributes such as religion and ethnicity.

The UN Strategy and Plan of Action on hate speech defined it as any kind of communication in speech, writing or behavior that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, on their religion, ethnicity, nationality, race, color, descent, gender or other identity factor" (United Nations, n.d). It is characterized by expressions that demean, discriminate, or incite violence and poses significant threats to the well-being of a man and the world at large.

Hate speeches have been on increase not only among peers but also political and religious leaders. The high rate of social media and online comments have provided users with eccentric avenues to voice their opinions without any regards. This democratization of expression has necessitated this also. From targeted harassment campaigns by political elites to the least of common man, its impact on individuals and societies cannot be overemphasized.

An agreement was reached that online platforms have the responsibility to mitigate the exigencies of hate speech while upholding principles of free speech and open dialogue. In the light of this, many actions have been taken to address the occurrences of hate speeches by online platforms, pressure groups and governments, thus, creating the need for the use of supervised machine learning paradigm and natural language processing to mitigate this.

Supervised Machine Learning Paradigm

This is the learning approach of machines when under supervision whereby labeled data are used in the form of input-output pairs. The major tasks of this type of learning are regression, classification and forecasting (Kotsiantis, 2007).

Natural Language Processing (NLP)

The field of NLP is a branch of Artificial Intelligence that focuses on the interaction between humans and computers using natural language (Johnson, 2023). It leverages on computational linguistics and machine learning techniques to analyze and understand human language. By developing sophisticated algorithms and models, researchers and practitioners in NLP can automate machine translation, speech recognition, information retrieval, spam detection, text summarization, intelligent web searching, intelligent spell checking and human-computer communication.

Review of Related Work

Zhang *et al*. (2018) worked on "Detecting hate speech on Twitter using a convolution-GRU based deep neural network". The paper introduced a new method based on a deep neural network combining convolutional and gated recurrent networks. The authors conducted an extensive evaluation of the method against several baselines and state of the art on the largest collection of publicly available Twitter datasets to date. The researchers' proposed method captured both word sequence and order information in short texts.

Khanday *et al*. (2022) delved into detecting twitter hate speech in COVID-19 era using machine learning and ensemble learning techniques. The authors carried out hate speech detection using machine learning and ensemble learning techniques during COVID-19. The twitter data used were extracted using the publicly available twitter API with the help of trending hashtags during the COVID-19 pandemic. The tweets were manually annotated into two categories based on different factors. Feature extraction was performed using Term Frequency/Inverse Document Frequency (TF/IDF), Bag of Words and Tweet Length. The study found the Decision Tree Classifier to be effective when compared to other typical Machine Learning (ML) classifiers. It had 98% precision, 97% recall, 97% F1-Score, and 97% accuracy.

Rodriguez *et al*. (2022) developed a framework for detection and integration of unstructured data of hate speech on Facebook using sentiment and emotion analysis. The aim of the research was to locate and analyze the unstructured data of selected social media posts that intend to spread hate in the comment sections. To address this issue, they proposed a novel framework called FADOHS, which combines data analysis and natural language processing strategies to sensitize all social media providers to the pervasiveness of hate on social media. Specifically, they used sentiment and emotion analysis algorithms to analyze recent posts and comments on these pages. Posts suspected of containing dehumanizing words will be processed before fed to the clustering algorithm for further evaluation. According to the experimental results, the proposed FADOHS framework surpassed the state-of-the-art approach in terms of precision, recall, and F1 scores by approximately 10%.

Pamungkas *et al*. (2020) on "Do you really want to hurt me? Predicting abusive swearing in social media". They explored the phenomenon of swearing in Twitter conversations, taking the possibility of predicting the abusiveness of a swear word in a tweet context as the main investigation perspective. They developed the Twitter English corpus SWAD (Swear Words Abusiveness Dataset), where abusive swearing was manually annotated at the word level. Their collection consists of 1,511 unique swear words from 1,320 tweets. They developed models to automatically predict abusive swearing to provide an intrinsic evaluation of SWAD and confirm the robustness of the resource. They also presented the results of a glass box ablation study in order to investigate which lexical, syntactic and effective features that are more informative

towards the automatic prediction of the function of swearing.

Zimmerman *et al*. (2019) researched on improving hate speech detection with deep learning ensembles. They utilized a publicly available embedding model and tested against a hate speech corpus from Twitter. To confirm the robustness of their results, they additionally tested against a popular sentiment dataset. Their method had a nearly 5 point improvement in F-measure when compared to original work on a publicly available hate speech evaluation dataset. The major difficulties they encountered was reproducibility of deep learning methods and comparison of findings from other work.

Yun *et al*. (2023) worked on BERT-Based logits ensemble model for gender bias and hate speech detection. They aimed to solve the problem on gender bias and hate speech detection, and to detect malicious comments in a Korean hate speech dataset constructed in 2020. They explored bidirectional encoder representations from transformers (BERT)-based deep learning models utilizing hyperparameter tuning, data sampling, and logits ensembles with a label distribution. They evaluated the model in Kaggle competitions for gender bias, general bias, and hate speech detection. For gender bias detection, an F1-score of 0.7711 was achieved using an ensemble of the Soongsil-BERT and KcELECTRA models. The general bias task included the gender bias task, and the ensemble model achieved the best F1-score of 0.7166.

Siino *et al*. (2021) analyzed the detection of hate speech spreaders using convolutional neural network. The authors developed a deep learning model based on a convolutional neural network (CNN) for the profiling hate speech spreaders (HSSs). Their classification (HSS or not HSS) takes advantage of the CNN based on a single convolutional layer. In this binary classification task, they performed tests using a 5-fold cross validation, in which the proposed model reached a maximum accuracy of 0.80 on the multilingual (i.e., English and Spanish) training set, and a minimum loss value of 0.51 on the same set. The trained model presented was able to reach an overall accuracy of 0.79 on the full test set.

Mozafari *et al*. (2019) worked on a BERT-Based transfer learning approach for hate speech detection in online social media. The study introduced a novel transfer learning approach based on an existing pre-trained language model called Bidirectional Encoder Representations from Transformers (BERT). The transfer learning-based fine-tuning techniques to explore BERT's capacity to detect hateful context in social media content. To evaluate the proposed approach, they made use of two publicly available datasets that have been annotated for racism, sexism, hate, or offensive content on Twitter. The results showed that their solution could obtain considerable performance on these datasets in terms of precision and recall in comparison to existing approaches. Also, their model captured some biases in data annotation and collection process and can potentially lead to a more accurate model.

III. RESEARCH METHODOLOGY

This section outlines the methodology adopted for sentiment analysis on hate speeches using a supervised learning approach. The chosen approach involves training models on labeled datasets, leveraging the rich body of research and techniques in supervised learning for sentiment classification.

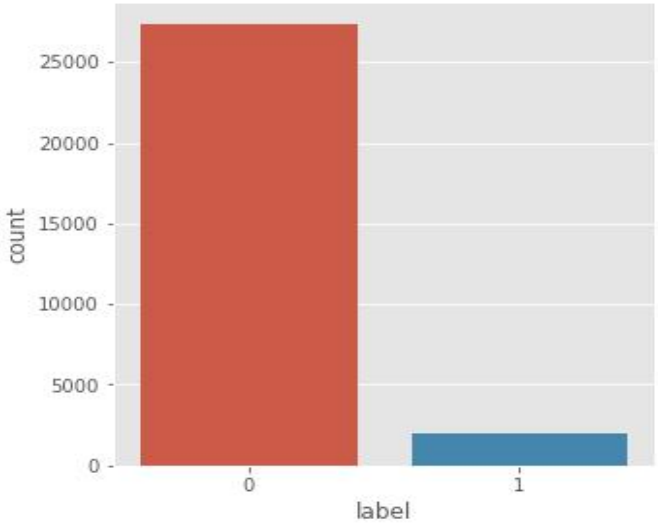


Fig.1: Graph of the data used for analysis (0-Non Hate Speech1- Hate Speech)

Data Collection

The dataset was collected from Twitter and contains a diverse set of tweets from various sources and user backgrounds, spanning over a year of data collection. The `hateDetection_train.csv` dataset utilized consists of 31964 tweets in total, with 93.2% labeled as hateful and 6.8% as non-hateful, making it an imbalanced dataset as seen in Figure 1.

To ensure transparency and reproducibility, it is crucial to provide a detailed account of the dataset's origin, size, and composition. The first step in understanding the dataset was loading it into a Pandas DataFrame. Figure 2 shows the first 5 tweets visualized from the dataset after loading it into the Pandas DataFrame.

	id	label	tweet
0	1	0	@user when a father is dysfunctional and is s...
1	2	0	@user @user thanks for #lyft credit i can't us...
2	3	0	bihday your majesty
3	4	0	#model i love u take with u all the time in ...
4	5	0	factsguide: society now #motivation

Fig.2: Raw testing data (Top 5 tweets)

Subsequently, a comprehensive exploration of the dataset was essential. This included calculating descriptive statistics, such as the mean tweet length, character distribution, and class distribution (i.e., the number of hateful and non-hateful tweets). Visualizations, such as word clouds, can also provide valuable insights into the most common words used in each category. Figure 3 shows the processes involved in hate speech detection.

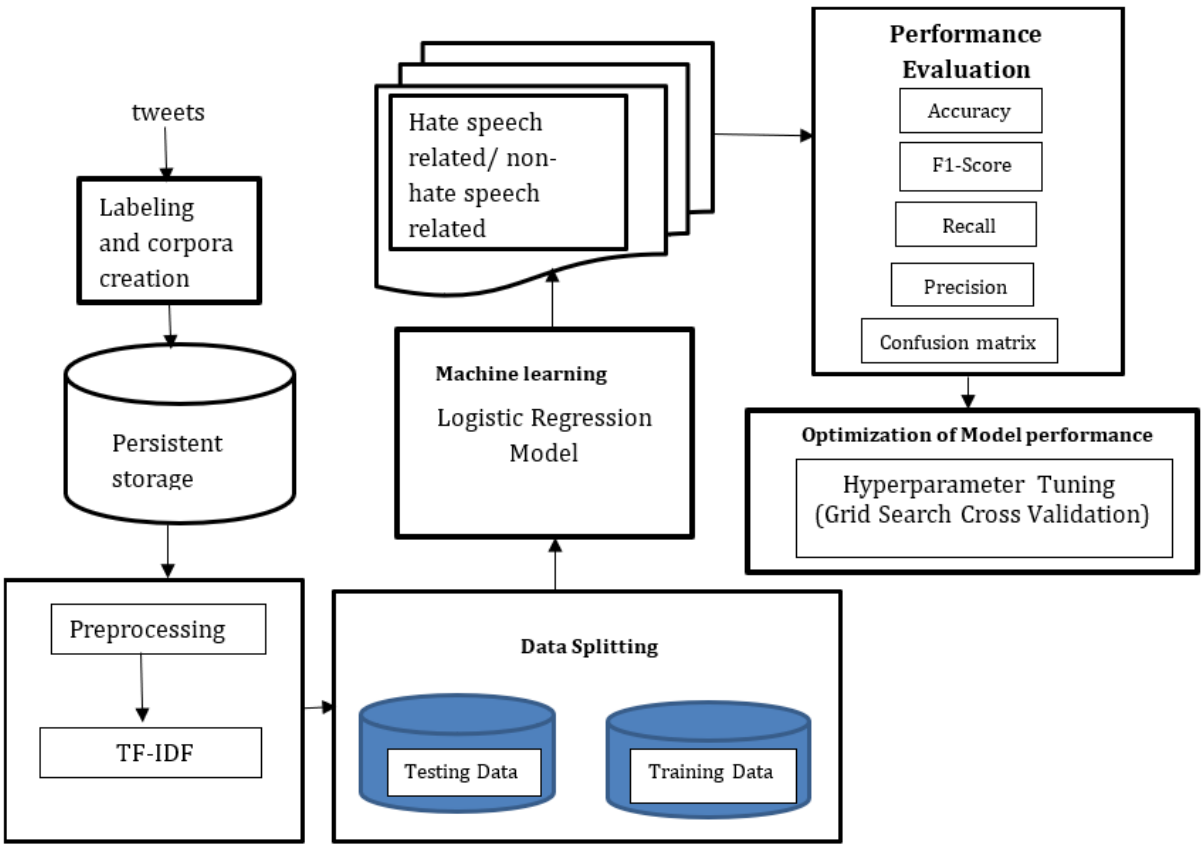


Fig.3: Hate Speech Detection System

To prepare the text data for modeling, the following preprocessing steps were applied. Figure 4 also shows the code block that carried out these preprocessing steps:

a. Removal of URLs, mentions, and hashtags: These elements do not carry significant semantic meaning and can be safely removed.

b. Conversion to lowercase: To ensure consistency in word representation and avoid treating the same word differently due to case variations.

c. Handling special characters and emojis: Special characters and emojis are retained as they may convey sentiment or context.

d. Stop word removal: Common words like "the," "and," and "in" are removed as they carry little informative value.

e. Lemmatization: Reducing words to their root forms helps in capturing the core meaning of words.

f. Duplicate tweet removal: Duplicate tweets are removed to prevent bias in the training process.

```
#creating a function to process the data
def data_processing(tweet):
    tweet = tweet.lower()
    tweet = re.sub(r"https\S+|www\S+http\S+", '', tweet, flags = re.MULTILINE)
    tweet = re.sub(r'\@w+|\#','', tweet)
    tweet = re.sub(r'[^\w\s]','',tweet)
    tweet = re.sub(r'ð','',tweet)
    tweet_tokens = word_tokenize(tweet)
    filtered_tweets = [w for w in tweet_tokens if not w in stop_words]
    return " ".join(filtered_tweets)
```

```
tweet_df.tweet = tweet_df['tweet'].apply(data_processing)
```

```
tweet_df = tweet_df.drop_duplicates('tweet')
```

```
lemmatizer = WordNetLemmatizer()
def lemmatizing(data):
    tweet = [lemmatizer.lemmatize(word) for word in data]
    return data
```

```
tweet_df['tweet'] = tweet_df['tweet'].apply(lambda x: lemmatizing(x))
```

Fig.4: Code block for Data preprocessing

Data Splitting: To evaluate model performance effectively, the dataset was split into training (80%) and testing (20%) sets using a random split with a fixed random state. This ensures reproducibility and allows me to assess the model's generalization ability on unseen data. Figure 5 shows the code block used in splitting the dataset.

```
X = tweet_df['tweet']
Y = tweet_df['label']
X = vect.transform(X)
```

```
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.2, random_state=42)
```

```
print("Size of x_train:", (x_train.shape))
print("Size of y_train:", (y_train.shape))
print("Size of x_test: ", (x_test.shape))
print("Size of y_test: ", (y_test.shape))
```

```
Size of x_train: (23476, 380305)
Size of y_train: (23476,)
Size of x_test:  (5869, 380305)
Size of y_test:  (5869,)
```

Fig.5: Data Splitting Code Block Module

Feature Extraction

Feature extraction is a crucial aspect of natural language processing tasks. In the research, the textual data was represented as numerical features using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization. It measures the importance of words in a document relative to the entire corpus. The experiment was carried out with different n-gram ranges (1-2 and 1-3) to investigate the impact of capturing word sequences on model performance as seen in Figure 7. This experiment is vital as it helped to identify which textual features are most informative for hate speech detection. Figures 6 and 7 show the word cloud of the most frequent words in hate speech tweets and the feature extraction code block respectively.

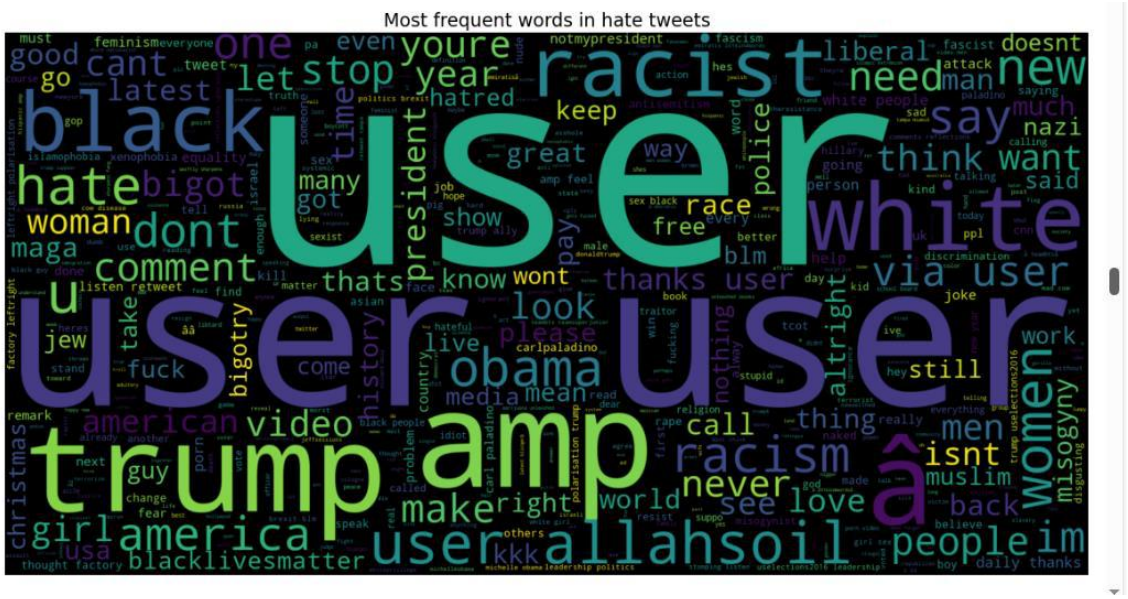


Fig.6: Word Cloud of the hate speech detected tweets

```
vect = TfidfVectorizer(ngram_range=(1,2)).fit(tweet_df['tweet'])
```

```
feature_names = vect.get_feature_names_out()
print("Number of features: {}\n".format(len(feature_names)))
print("First 20 features: \n{}".format(feature_names[:20]))
```

```
Number of features: 208579

First 20 features:
['0000001' '0000001 polluting' '00027' '00027 photooftheday' '001' '0035'
 '00h30' '01' '01 4995' '01 7900' '01 blog' '01 croatia' '01 may'
 '01 shopalyssas' '0115' '0115 8599968' '0161' '0161 manny' '019'
 '019 previous']
```

```
vect = TfidfVectorizer(ngram_range=(1,3)).fit(tweet_df['tweet'])
```

```
feature_names = vect.get_feature_names_out()
print("Number of features: {}\n".format(len(feature_names)))
print("First 20 features: \n{}".format(feature_names[:20]))
```

```
Number of features: 380305

First 20 features:
['0000001' '0000001 polluting' '0000001 polluting niger' '00027'
 '00027 photooftheday' '00027 photooftheday music' '001' '0035' '00h30'
 '01' '01 4995' '01 4995 rustic' '01 7900' '01 7900 shopalyssas' '01 blog'
 '01 blog silver' '01 croatia' '01 croatia happy' '01 may' '01 may actual']
```

Fig.7: Feature Extraction Code Block

Model Selection and Training

After extensive experimentation with various machine learning algorithms, Logistic Regression was selected as the most suitable model for the binary classification task of hate speech detection. Logistic Regression is well-suited for this task due to its simplicity, interpretability and effectiveness in handling textual data as seen in Figure 8.

```
logreg = LogisticRegression()
logreg.fit(x_train, y_train)
logreg_predict = logreg.predict(x_test)
logreg_acc = accuracy_score(logreg_predict, y_test)
print("Test accuarcy: {:.2f}%".format(logreg_acc*100))
```

```
Test accuarcy: 93.15%
```

Fig.8: Code block of Logistic Regression

The performance of the Logistic Regression model heavily depends on its hyperparameters. To optimize these hyperparameters, Grid Search Cross-Validation was

employed. The grid search explored different combinations of hyperparameters (C and solver) and selected the ones that can maximize the model's performance on the validation set. This process enhances the model's ability to discriminate between hateful and non-hateful tweets.

Performance Evaluation

Five performance metrics were used – Accuracy, Confusion matrix, Precision, Recall and F1-Score.

Accuracy: This is the primary evaluation metric used in the research. It measures the proportion of correctly classified tweets. While accuracy provides an overall assessment of model performance, it may not be sufficient for imbalanced datasets. Figure 9 depicts the code block for determining accuracy. It is calculated as:

$$Accuracy = \frac{Number\ of\ correctly\ classified\ tweets}{Total\ number\ of\ tweets} \quad (1)$$

```
logreg = LogisticRegression()
logreg.fit(x_train, y_train)
logreg_predict = logreg.predict(x_test)
logreg_acc = accuracy_score(logreg_predict, y_test)
print("Test accuarcy: {:.2f}%".format(logreg_acc*100))
```

```
Test accuarcy: 93.15%
```

Fig.9: Code block for calculating the accuracy of the model

Confusion Matrix: To gain deeper insight into model performance, a confusion matrix that visualizes the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) was employed as shown in Figure 10. This information helped to identify specific patterns of errors made by the model, such as whether it tends to have more false positives or false negatives.

```
style.use('classic')
cm = confusion_matrix(y_test, logreg_predict, labels=logreg.classes_)
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=logreg.classes_)
disp.plot()
```

```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x27795b87f90>
```

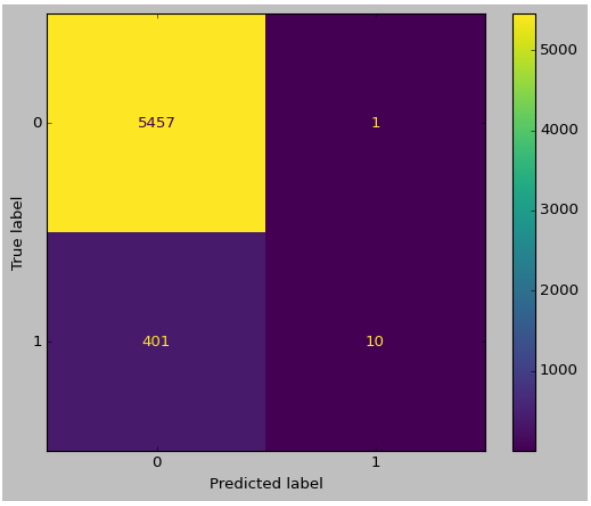


Fig.10: Confusion Matrix Display

Precision: This is used to determine the exactness of the measurement. It measures the accuracy of positive predictions. It is calculated as:

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)} \quad (2)$$

Recall: This measures the completeness of positive predictions, that is, measure of how well a model correctly identifies True Positives. Equ. (3) shows its calculation:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (3)$$

F1-Score: A measure of a model's accuracy on a dataset. It is a harmonic means of both precision and recall of the model. It is determined as:

$$F1_{Score} = 2\ X\ \frac{(precision\ x\ recall)}{(precision + recall)} \quad (4)$$

IV. RESULTS AND DISCUSSIONS

The performance metrics analysis provides insights into the model's effectiveness in distinguishing 'hateful' from 'non-hateful' texts and discusses potential factors influencing misclassifications.

The baseline performance of the model with default hyperparameters, achieved a test accuracy of 93% as shown in Figure 11. This baseline performance serves as a reference point for evaluating the effectiveness of subsequent improvements.

```
Test accuarcy: 93.15%
```

```
print(confusion_matrix(y_test, logreg_predict))
print("\n")
print(classification_report(y_test, logreg_predict))
```

```
[[5457    1]
 [ 401   10]]


              precision    recall  f1-score   support

           0       0.93      1.00      0.96      5458
           1       0.91      0.02      0.05       411

    accuracy                           0.93      5869
   macro avg       0.92      0.51      0.51      5869
weighted avg       0.93      0.93      0.90      5869
```

Fig.11: Baseline performance without hyperparameter tuning

Tuning hyperparameter enhances the performance of a model. Through the grid search cross-validation, the hyperparameters of the logistic regression model was optimized, resulting in an improved accuracy of 95% as seen in Figure 12. The optimal hyperparameters are determined to be C = 0.1 and solver = newton-cg.

```
[[5450    8]
 [ 292  119]]


              precision    recall  f1-score   support

           0       0.95      1.00      0.97      5458
           1       0.94      0.29      0.44       411

    accuracy                           0.95      5869
   macro avg       0.94      0.64      0.71      5869
weighted avg       0.95      0.95      0.94      5869
```

Fig. 12: Performance Evaluation after hyperparameter tuning

V. CONCLUSION

The research focused on creating a machine learning model for detecting hate speech in online textual content using NLP techniques. Based on the performance evaluation, Logistic Regression model showed reliable results in classifying text as either a hate speech or non-hate speech.

REFERENCES

[1] United Nations (n.d). What is hate speech? Retrieved from https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech

[2] Johnson, .A. (2023). NLP Vs Computational Linguistics: Understanding the Differences. Retrieved from https://medium.com/@andrew_johnson_4/nlp-vs-computational-linguistics-understanding-the-differences-57044aa41ad2.

[3] Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., and Malik, S. H. (2022). Detecting twitter hate speech in COVID-19 era using machine learning and ensemble learning techniques. International Journal of Information Management Data Insights, *2*(2), Pgs. 100-120.

[4] Kotsiantis, S.B. (2007). Supervised Machine Learning: A Review of Classification Techniques, Informatica 31, Pgs. 249-268.

[5] Mozafari, M., Farahbakhsh, R. and Crespi, N. (2020). A BERT-based transfer learning approach for hate speech detection in online social media. In Complex Networks and Their Applications VIII: Volume 1 Proceedings of the 8th International Conference on Complex Networks and Their Applications, COMPLEX NETWORKS 2019 Vol.8, Pgs. 928-940. Springer International Publishing.

[6] Pamungkas, E. W., Basile, V. and Patti, V. (2020, May). Do you really want to hurt me? Predicting abusive swearing in social media. In Proceedings of the 12th Language Resources and Evaluation Conference, Pgs. 6237-6246.

[7] Rodriguez, A., Chen, Y. L. and Argueta, .C. (2022). FADOHS: framework for detection and integration of unstructured data of hate speech on facebook using sentiment and emotion analysis, IEEE Access, 10, Pgs. 22400-22419.

[8] Siino, M., Di Nuovo, E., Tinnirello, I. and La Cascia, M. (2021). Detection of hate speech spreaders using convolutional neural networks. In *CLEF (*Working Notes*)*, Pgs. 2126-2136.

[9] Wang, Z. and Cha, Y. J. (2021). Unsupervised deep learning approach using a deep auto-encoder with a one-class support vector machine to detect damage. Structural Health Monitoring, 20 (1), Pgs. 406-425.

[10] Yun, S., Kang, S. and Kim, H. (2023). BERT-Based Logits Ensemble Model for Gender Bias and Hate Speech Detection. Journal of Information Processing Systems, 19 (5).

[11] Zhang, D., Mao, R., Song, X., Wang, D., Zhang, H., Xia, H., and Gao, Y. (2023). Humidity sensing properties and respiratory behavior detection based on chitosan halloysite nanotubes film coated QCM sensor combined with support vector machine. Sensors and Actuators B: Chemical, 374, 132824.

[12] Zhang, Z., Robinson, D. and Tepper, J. (2018). Detecting hate speech on twitter using a convolution-gru based deep neural network. In The Semantic Web: 15th International Conference Proceedings, ESWC 2018, Heraklion, Crete, Greece, June 3–7, Springer International Publishing, Pgs.745-760.

[13] Zimmerman, S., Kruschwitz, U. and Fox, C. (2018). Improving hate speech detection with deep learning ensembles. In Proceedings of the 11th international conference on language resources and evaluation (LREC 2018).

International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-11, Issue-3; Mar, 2024
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.114.2

Machine Learning Model for Attenuating Outliers in Stock Data

[1]Njideka Nkemdilim Mbeledogu and [2]Kaodilichukwu Chidi Mbeledogu

[1]Department of Computer Science, Nnamdi Azikiwe University, Awka, Nigeria
[2]Department of Mechanical Engineering, University of Ottawa, Canada

Received: 07 Feb 2024,

Receive in revised form: 12 Mar 2024,

Accepted: 20 Mar 2024,

Available online: 30 Mar 2024

***Keywords*— Outliers, Uncertainty, Artificial Neural Network, Fuzzy Artificial Neural Network, Root Mean Square Error**

***Abstract*—** *The presence of outliers has deleterious effects on the stock value because the unreliable information may discourage investors from investing in the stock. This associated problem paved way for the importance of intelligent prediction paradigm. However, existing stock forecasting models like Artificial neural network (ANN) performed better than traditional statistical models in handling the problem of non-linearity and complexity in forecasting stock price, it still lacked the capacity to handle outliers which are inherent in the stock market. Based on this, the researchers were motivated to forecast stock market based on the observation from literature that researchers do not report checking, prediction and proper management of outliers of any sort in stock market price forecasting. This paper addressed outliers' deleterious effects on the stock value by proposing a hybrid Fuzzy Artificial Neural Network (FANN) Model that attenuates outliers in stock forecasting accurately. The proposed model was simulated using MATLAB. The historical Nigerian stock quantitative datasets from Nigeria Stock Exchange (NSE) for 2008-2011 were used to test the simulated model. The Gaussian Membership function, due to its ability to handle minimum uncertainty principle, was used for the fuzzification of the extracted stock features to capture the stock dynamism. The proposed model's predictive performance was calculated using Root Mean Square Error (RMSE). The outlier detection analysis of the actual historical stock data, ANN and the proposed FANN Model predictions was calculated using Z-score. The proposed model had a RMSE value of about 3.83 which shows that it is a reliable stock forecasting model. The Z-score value of the proposed model was calculated to be about 0.78 which shows that it significantly attenuates outliers in stock forecasting. In overall, the results proved that the proposed FANN model can handle outliers in stock forecasting.*

I. INTRODUCTION

When dealing with real-world issues, one can rarely avoid uncertainty. At the experiential echelon, it cannot be avoided from any calculation, resulting from a combination of unavoidable experimental errors and limits of the calibrating instruments. At the mental rank, it is presented as a result from the vagueness and ambiguity that are fundamental in natural language. At the social stratum, it is originated and sustained by individuals for various needs (privacy, secrecy, propriety) while at economic level, taking uncertainty in stock can be caused by financial data characterized by erroneous data, outliers (Jatinder, 2012), geopolitical events, market sentiment and news. Whatever

the level is, uncertainty affects the nation's economy by affecting individual companies and the economy as a whole respectively.

Uncertainty is an unclear multi-faceted characterization about data or predictions made from data that may include error, accuracy, validity, quality, noise and confidence and readability (Dungan et al., 2002). In any analyses, it can feature due to insufficient data, indistinct, conflicting, incoherent, not fully reliable, deficient in some way, erroneous, outliers , transnational events and information deficiency in market views and thus, results in different types of uncertainty.

Mendel (2001) stated that three types of uncertainty exists - *fuzziness* (or vagueness), which results from the imprecise boundaries of fuzzy sets; *nonspecificity* (or imprecision), which is connected with sizes (cardinalities) of relevant sets of alternatives; and *strife* (discord, outlier), which expresses conflicts among the various sets of alternatives. The researcher went further to group these into two major classes. They are *fuzziness* and *ambiguity*, where ambiguity (one-to-many relationships) encompasses *nonspecificity* and *strif*e.

II. LITERATURE REVIEW

Outliers are observations in the dataset that appear to be unusual and discordant (Yanfang, 2014). In Statistics, it is an observation that is significantly distant from the rest of the data. Hawkins (1980) defined it as an observation which deviates so much from the other observations as to spur suspicions that it was generated by a different mechanism. Aggarwal (2005) referred to them as abnormalities, discordant, deviants or anomalies in the data mining and statistics.

In many analyses, a relatively small size of outliers can disrupt even simple analysis because they can shoot up the error variance, thereby, lessening the power of statistical tests, normality in a non-randomly distributed situation by altering the propability of producing Type I and Type II errors and can greatly affect estimates that may be of valuable interest (Osborne and Overbay, 2004).

Based on research, many researchers gave different views on how best to handle outliers observed in a data. Barnett and Lewis (1994) proposed its removal when not in accordance with the other valid data. This notion was generally accepted by many researchers even in the situations when they are licit or has vague purpose. However, some researchers such as Orr *et al*. (1991) and Osborne and Overbay (2004) felt otherwise but suggested it as a triggering factor for investigation amidst its erroneous value that might contain valuable information in a more global sense (Osborne and Overbay, 2004). Based on this, Osborne (2002) opted that it is needful to use a transformation technique to keep the individual observation in the dataset as well as reduce its disruption in the statistical inference.

Forecasting and Circuit Breakers in Stock Market

A forecast is a prediction of some future event or events. It is a needful aspect of life that spreads across many fields such as business and industry, government, economics, environmental sciences, medicine, social science, politics, and finance. It is often categorized as short-term, medium-term, and long-term. Historical data are used in forecasting because they display inactiveness to change fast. Securities are traded in the Stock Market. The last price at which it is traded in a day gives the most up-to-date valuation of that security until the next trading day begins (Investopedia, 2013). This last price is termed Closing Price. Close Price is needful in stock market because it shows a valuable benchmark for investors to check differences in stock prices over time such that measuring the market opinion for a given security over a trading day, the closing price of one day can be compared to the preceding day. For investors to predict the closing stock price, daily historical stock data are collected.

Companies normally forecast sales and trends in production for the investors to follow with the believe that the market conditions are normal but the observed variables (predicting/attributes) usually contain outliers. If a stock technical indicator contains outliers, the incorrect evaluation from contaminated observations may be highly misrepresented, thus leading to unreliable results (Yanfang, 2014). Based on this, there is need to identify the outliers using the circuit breakers that is within $\pm 10\%$ by the Security and Exchange Commission (SEC) (Ohuche and Ikoku, 2015) and minimize their discordance. Therefore, the key part of this stock data analysis is the detection and proper handling of outliers.

Artificial Neural Network (ANN)

ANNs provide a way to emulate biological neurons to solve complex problems in the same manner as the human brain. It is an enormously parallel distributed processor that has a natural capacity for storing experiential knowledge and making it available for use (Haykin, 1998). It is capable of learning because it is modeled after the human brain (Haag *et al*., 1998). Though it is a mathematical model of information processing, ANN is relatively different from turing machines with stored programs (Fausett, 1996). Its information processing system (Figure 1) is developed based on the idea of mathematical models of the brain's cognition on the belief that the capability of processing information occurs at its simple elements called neurons.

The signals are transmitted amidst these neurons through the connection links. Every connection link has an associated weight that multiplies the transmitted signal. The output signal is determined by an activation function (usually nonlinear) of each neuron upon the sum of the weighted input signals (net input).

ANN offers so many benefits. Some of such benefits are nonlinearity, input-output Mapping, adaptation, evidential response, contextual information, fault tolerance, VLSI Implementation, uniformity of analysis and design, and neurobiological analogy.

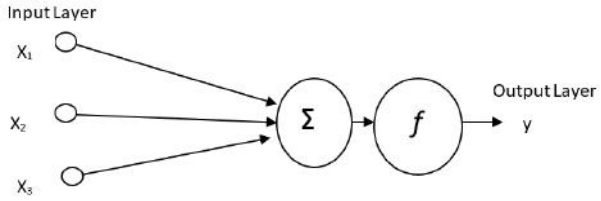


Fig.1: Single layer feed forward network

The mapping of single unit perceptron can be expressed as:

$$y = f(\sum_{i=1}^{n} w_i x_i + b) \quad \text{Eq. 1}$$

Where w_i are the individual weights, x_i are the inputs and b is the bias.

REVIEW OF RELATED WORKS

Moghar and Hamiche (2020) predicted the future stock market values using Long-Short Term Memory (LSTM) and Recurrent Neural Network. Also, they determined which precision a machine learning algorithm can predict and how much epochs can improve their model.

Noel (2023) employed dynamic neural networks to predict the future closing stock price using two prevailing theories that sought to qualify market behavior. The two theories were the Efficient Market Hypothesis (EMH) and Chaos.

Thakkar and Chaudhari (2021) delved into a comprehensive survey on deep neural networks for stock market. The authors presented the need, challenges, future directions and the applicability of deep neural network variations to the temporal stock market data and NN meta-heuristics approaches were discussed.

III. RESEARCH METHODOLOGY

Data and Sources of data

The daily stock data of three randomly selected companies from the 27 blue chip companies in Nigeria were collected from the Nigeria Stock Exchange from 2008-2011. The companies are Dangote Sugar Refinery, GlaxoSmith Kline Consumer Plc and Julius Berger Construction Company.

The inputs into the system were four stock variables (Open price, High price, Low price and Close price) that are independent in nature.

Outlier Detection

When the sample data is drawn from a non-normally distributed population and the sample size is large enough (that is, ≥ 30), a standard normal test can be employed. This was empowered by the Central Limit Theorem (CLT) which states that if random variable S_n is defined as the sum of n independent and identically distributed (i.i.d) random variables, $X_1, X_2, \ldots, X_n$; with mean,μ and standard deviation, σ. Then, for large enough n (typically $n \geq 30$), S_n is approximately normally distributed with parameters : $\mu_{sn} = n\mu$ and $\sigma_{sn} = \sqrt{n\sigma}$. This result holds regardless of the shape of the X distribution (that is, X do not have to be normally distributed (Eze *et al*., 2005; Filmus, 2010).

Therefore, so long as n is large and are independent variables, by CLT, the Z-score (Z-test) is distributed as standard normal and was adopted.

Z-score

The basic idea of this rule is that if X follows a normal distribution, N (μ, σ^2), then Z follows a standard normal distribution, N (0, 1), and Z-scores that exceed 3 in absolute value are generally considered as outliers. This method is simple and it is the same formula as the 3 standard deviation method when the criterion of an outlier is an absolute value of a Z-score of at least 3.

The actual close stock price data was used because the scope of the prediction was the next day's closing price.

The Z-score can be computed as

$$Z_i = \frac{ABS\,(X_i - \mu)}{\sigma} \quad \text{Eq. 2}$$

Where X_i is the value, μ is the mean and σ is the standard deviation.

Outlier Analysis of the Actual Stock Dataset

Three different stock datasets: Dangote Sugar [2008-2011 (938 data values)], GlaxoSmith Kline Consumer Plc [2012 (240 data values)] and Julius Berger [2012 (237 data values)] were used. Table 1 shows sample of the stock dataset and the detected outliers using Z-score (Z_i) based on the parameter analysis stated in Table 2. Figure 2 depicts the polar plot of Dangote Sugar.

Table 1: Outlier detection of the sample stock dataset

Dangote Sugar	Z_i **DS**	**GlaxoSmith Kline Cons. Plc**	Z_i **GS**	**Julius Berger**	Z_i **JB**
15	0.4237941	23	0.563775983	30.99	0.558068152
15.11	0.406774133	23	0.563775983	31.06	0.577889533
15.25	0.385112356	23	0.563775983	31.06	0.577889533
15.2	0.392848705	23	0.563775983	31.06	0.577889533
15.2	0.392848705	23	0.563775983	31.06	0.577889533
15.01	0.422246831	21.85	0.6901164	31.06	0.577889533
115.3	**15.09532188**	21.85	0.6901164	32.61	1.016791541
15.55	0.338694262	22.9	0.574762106	34.24	1.478346557
15.49	0.347977881	23	0.563775983	**0**	**8.217140398**
15.46	0.35261969	23	0.563775983	34	1.410387536
15.99	0.270614391	23	0.563775983	34	1.410387536
16	0.269067122	23	0.563775983	34	1.410387536
16	0.269067122	**0**	**3.090584321**	33.01	1.130056576
16.31	0.221101758	22.7	0.596734353	30.61	0.450466369

Where DS is Dangote Sugar, GS is GlaxoSmith and JB is Julius Berger.

Table 2: Outlier Parameter Analysis

Stock	μ	σ
Dangote Sugar	17.73898	6.462997
GlaxoSmith Kline Consumer Plc	28.13171	9.102392
Julius Berger	29.01916	3.53154

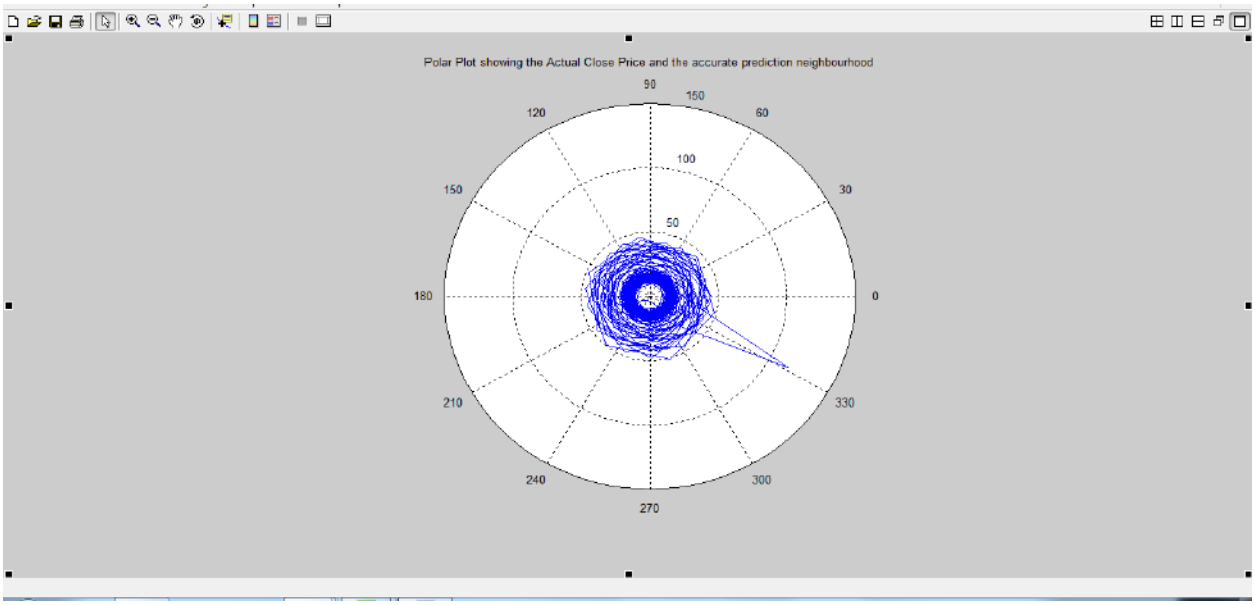


Fig.2: Polar plot showing the Actual close price for Dangote Sugar and accurate neighbourhood of prediction

From Table 1, it will be observed that an outlier was detected from each of the three data sets. For Dangote Sugar, the stock data value 115.30 with Z_i of 15.09532188, stock data 0 of GlaxoSmith Kline Consumer Plc with Z_i of 3.090584321 and stock data 0 of Julius Berger with Z_i of 8.217140398. Figure 1 depicts the Polar plot showing the Actual Close Price for Dangote Sugar. Polar plot gives the position of a point in a 2D surface (Tsishchanks, 2010). This shows how the data are clustered thereby showing the accurate neighbourhood of prediction and the outlier.

FANN Modeling

The FANN model was simulated using MATLAB. Figure 3 shows the model diagram of FANN that hybridized Fuzzy and ANN. Four stock inputs (Open price, High price, Low price and Close price) were used.

Fuzzy Modeling in FANN

The Gaussian membership function was used to transform the crisp stock data into fuzzy values (fuzzification). The choice of membership was with the aim of capturing the stock dynamism by the function's smoothness at the edges.

The Gaussian Membership Function is calculated as follows:

$$f(x_i, \sigma, m) = \exp\left\{-\frac{(x_i - m)^2}{2\sigma^2}\right\} \quad \text{Eq. 3}$$

Where m represents the mean value, σ represents the standard deviation for a given membership function and x_i represents the raw stock training data.

$$net_i = x_i \quad \text{Eq. 4}$$

$$out_j = f(net_i, \sigma_{ij}, m_{ij}) \quad \text{Eq. 5}$$

Where out_j represents the output corresponding to the j^{th} membership function that corresponds to the input x_i.

Rule Generation

The Multiple input multiple output (MIMO) form of representing the expert knowledge was used such that:

Fact: μ_1 is A_1^i and μ_2 is A_2^i and … and μ_n is A_n^i

Rule R_{ij}: If μ_1 is A_1^i and μ_2 is A_2^i and … and μ_n is A_n^i then T_1 is C_1^j, T_2 is C_2^j, …, T_m is C_m^j, W_{ij}

Result: T_1 is C_1^j, T_2 is C_2^j, …, T_m is C_m^j

Where $u_1, \dots, u_n$ are the stock input linguistic variables (process state variables) and $T_1, \dots, T_m$ are the stock control Linguistic variables, A_1^i, …, A_n^i and C_1^j, …, C_m^j are the stock linguistic values of the linguistic variables $u_1, \dots, u_n$ and $T_1, \dots, T_m$ in the stock universe of discourse X and Y. The rules are weighted such that the degree of confidence with which the stock input fuzzy set A_1^i, …, A_n^i (which is composed of fuzzy intersection (AND) of several univariate stock fuzzy sets) is related to the stock output fuzzy set C_1^j, …, C_m^j is given by $W_{ij} \in [0,1]$. When W_{ij} is zero, the rule is inactive and does not contribute to the output. Otherwise, it partially fires whenever its antecedent is activated to a degree greater than zero. R_{ij} represents the stock rule number.

Defuzzification

Center of Gravity (CoG) which is also called Center of Area (CoA) defuzzification method was employed to transform the fuzzified data into its equivalent crisp form. The centroid of the area bounded by the controller output MF is determined and its abscissa taken as the crisp controlling value. Its computation is given as:

$$CoA\,[C(Z)] = \frac{\Sigma_{i=1}^{q} Z_i C(Z_i)}{\Sigma_{i=1}^{q} C(Z_i)} \quad \text{Eq. 6}$$

where q is the number of sample values of the stock dataset and Z_i is the value of the control output at the sample value. Because they are four inputs of the stock control output MF, it resulted to four groups of CoA defuzzification groups.

ANN modeling in FANN

A Multi-layer perceptron model (MLP) was used because it can be trained to approximate most functions arbitrarily well while Single-Layer networks cannot. The MLP comprised of four layers – input layer, two hidden layers and the output layer. The output from the fuzzy formed the inputs into the ANN. The financial time series under consideration is highly non-linear and requires a sufficiently non-linear function to represent all the properties of the series. Log-sigmoid and Purelin activation functions were deployed within the layers because they can be trained to approximate most functions arbitrarily well (Hagan *et al.*, 1996).

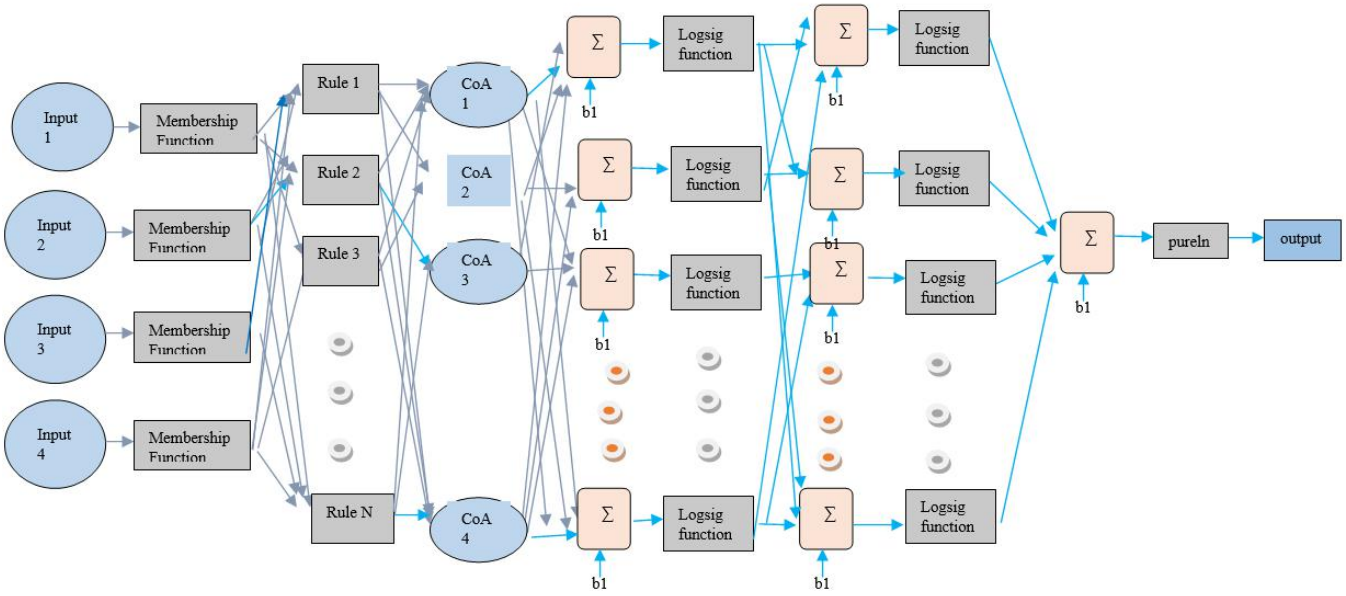


Fig.3: High Level Model of FANN Architecture

Algorithm

Training and learning functions are mathematical procedures used to automatically adjust the network's weights and biases. The training function dictates a global algorithm that affects all the weights and biases of a given network while the learning function can be applied to individual weights and biases within a network (Mathworks, 2014).

The system deployed a supervised paradigm by adopting Levenberg Marquardt (LM) back propagation for training and Gradient Descent (GD) for learning.

Step 1: The defuzzified data are the inputs to the neural network. Propagate the input forward through the network by selecting random weights and biases:

$$a^{m+1} = f^{m+1}(W^{m+1}a^m + b^{m+1}) \quad \text{for} \quad m = 0,1,\dots,M-1 \qquad \text{Eq. 7}$$

Step 2: Calculate the errors

$$e_q = t_q - a_q^M \qquad \text{Eq. 8}$$

Where e_q is the error, t_q is the target value and a_q^M is the output.

Step 3: Compute the sum of squared errors over all inputs, $F(x)$:

$$F(x) = \sum_{q=1}^{Q}(t_q - a_q)^T(t_q - a_q) \qquad \text{Eq. 9}$$

$$= \sum_{q=1}^{Q} e_q^T e_q \qquad \text{Eq. 10}$$

$$= \sum_{q=1}^{Q}\sum_{j=1}^{s^m}(e_{j,q})^2 \qquad \text{Eq. 11}$$

$$= \sum_{i=1}^{N}(v_i)^2 \qquad \text{Eq. 12}$$

Where $e_{j,q}$ is the jth element of the error for the qth input/target pair and v is the error vector.

$$v^T = [v_1\ v_2 \dots\ v_N] \qquad \text{Eq. 13}$$

$$= [e_{1,1}\ e_{2,1} \dots\ e_{s^m,1}\ e_{1,2}\ \dots e_{s^m,Q}] \qquad \text{Eq. 14}$$

Step 4: Compute the Jacobian matrix J(x):

$$J(x) = \begin{bmatrix} \frac{\partial e_{1,1}}{\partial w_{1,1}^1} & \cdots & \frac{\partial e_{1,1}}{\partial b_1^1} \\ \vdots & \ddots & \vdots \\ \frac{\partial e_{s^m,1}}{\partial w_{1,1}^1} & \cdots & \frac{\partial e_{s^m,1}}{\partial b_1^1} \end{bmatrix} \qquad \text{Eq. 15}$$

Step 5: Calculate the sensitivity with the recurrence relations

Initialize the back propagation with

$$S_q^{-m} = -F^m(n_q^m) \qquad \text{Eq. 16}$$

Step 6: Propagate each column of the matrix S_q^{-m}

$$S_q^{-m} = F^m(n_q^m)(W^{m+1})^T S_q^{-m+1} \qquad \text{Eq. 17}$$

Step 7: Augment the individual matrices into Marquardt sensitivities:

$$S^{-m} = \langle S_1^{-m} | \dots | S_Q^{-m} \rangle \qquad \text{Eq. 18}$$

Note: For each input presented to the network, the sensitivity vectors are propagated back. This is because the derivatives of each individual error is computed and not the derivative of the sum of squares of the errors. For every input applied to the network there will be S^m errors (one for each element of the network output). For each error, there will be one row of the Jacobian matrix.

Step 8: Compute the elements of the Jacobian matrix:

$$[J]_{h,l} = \frac{\partial v_h}{\partial x_l} = \frac{\partial e_{k,q}}{\partial w_{i,j}^m} = \frac{\partial e_{k,q}}{\partial n_{i,q}^m} \; x \; \frac{\partial n_{i,q}^m}{\partial w_{i,j}^m} = s_{i,h}^{-m} \; x \; \frac{\partial n_{i,q}^m}{\partial w_{i,j}^m} = s_{i,h}^{-m} \; x \; a_{j,q}^{m-1} \quad \text{Eq. 19}$$

$$[J]_{h,l} = \frac{\partial v_h}{\partial x_l} = \frac{\partial e_{k,q}}{\partial b_i^m} = \frac{\partial e_{k,q}}{\partial n_{i,q}^m} \; x \; \frac{\partial n_{i,q}^m}{\partial b_i^m} = s_{i,h}^{-m} \; x \; \frac{\partial n_{i,q}^m}{\partial b_i^m} = s_{i,h}^{-m} \quad \text{Eq. 20}$$

if x_l is a bias

Step 9: Solve to obtain Δx_k

$$\Delta x_k = -[J^T(x_k)J(x_k) + \mu_k I]^{-1} J^T(x_k)v(x_k) \quad \text{Eq. 21}$$

Step 10: Re-compute the sum of squares errors using $x_k + \Delta x_k$. If this new sum of squares is smaller than the value computed in step 1, then divide μ by v, let

$$x_{k+1} = x_k + \Delta x_k \quad \text{Eq. 22}$$

and go back to step 1. If the sum of squares is not reduced, then multiply μ by v and go back to step 9.

The algorithm is assumed to have converged when the norm of the gradient is less than some predetermined value, or when the sum of squares has been reduced to some error goal.

$$\nabla F(x) = 2J^T(x)v(x) \quad \text{Eq. 23}$$

Learning Algorithm

Gradient descent learning algorithm with learning rate and momentum coefficient were used for the learning. It was implemented using the Learngdm Function.

Training

Functional tests and comparisons presented in Tables 3, 5 and 7 are the actual close price of the stock data, the predictions of ANN and FANN, and the Z-score for the actual stock price, Dangote Sugar, GlaxoSmith Kline and Julius Berger datasets respectively based on their outlier parameter analysis presented on Tables 4, 6 and 8. Figures 4 and 5 show the graphical representation of FANN predictions and the Polar plot showing the Actual close price and FANN Model for Dangote Sugar respectively.

Table 3: Actual close price with ANN and FANN and their z-scores for Dangote Sugar

Days	Actual Close Price	Z_i of Actual Close Price	ANN Predictions	Z_i of ANN	FANN prediction	Z_i of FANN
518	15.20	0.385112356	15.25	0.422214959	15.05	0.544055695
519	15.20	0.392848705	15.20	0.427652245	15.00	0.550129038
520	15.01	0.392848705	15.20	0.427652187	14.99	0.55141744
521	115.30	0.422246831	15.01	0.448312847	**14.85**	0.570113596
522	15.55	**15.09532188**	**115.30**	**10.45726241**	25.00	**0.780386738**
523	15.49	0.338694262	15.55	0.389593024	25.00	0.780386738
524	15.46	0.347977881	15.49	0.396117468	15.29	0.511555605
525	15.99	0.35261969	15.46	0.399379704	15.21	0.522823472
526	16.00	0.270614391	15.99	0.341747185	15.67	0.461188991

Table 4 : Outlier Parameter Analysis of Dangote Sugar

System	μ	σ
Actual Close Price	17.73898	6.462997
ANN	19.13278063	9.196213649
FANN Model	19.13458208	7.515948831

Table 5: Actual close price with ANN and FANN predictions of GlaxoSmith Kline Plc

Days	**Actual Close Price**	**Zi of Actual Close Price**	**ANN Predictions**	**Z_i of ANN**	**FANN Predictions**	**Zi of FANN**
6	21.85	0.6901164	22.99999	0.555166773	22.44521	0.678771729
7	21.85	0.6901164	21.85	0.682216357	22.44519	0.678774061
8	22.9	0.574762106	21.84999	0.682217461	23.37474	0.553814392
9	23	0.563775983	22.9	0.566213554	23.45614	0.542871598
10	23	0.563775983	23	0.555165668	23.45614	0.542871598
11	23	0.563775983	23	0.555165668	23.45614	0.542871598
12	23	0.563775983	23	0.555165668	23.45614	0.542871598
13	0	3.090584321	23	0.555165668	13.35331	1.901003085
14	22.7	0.596734353	0.183169	3.075943133	23.20793	0.576238447
15	22.7	0.596734353	22.70001	0.588308221	23.20793	0.576238447
16	22.7	0.596734353	22.70001	0.588308221	23.20793	0.576238447
17	22.7	0.596734353	22.70001	0.588308221	23.20793	0.576238447

Table 6: Outlier Parameter Analysis for GlaxoSmith Plc

System	μ	σ
Actual Close Price	28.13171	9.102392
ANN	28.02508508	9.051505466
FANN Model	27.49443484	7.438769163

Table 7: Actual close price with ANN and FANN predictions for Julius Berger.

Days	Actual Close Price	Zi of Actual Close Price	ANN Predictions	Z_i of ANN	FANN Predictions	Zi of FANN
87	31.06	0.577889533	31.06012	-1.417360474	30.89801	0.642466247
88	31.06	0.577889533	31.06012	-1.417360474	30.89801	0.642466247
89	31.06	0.577889533	31.06012	-1.417360474	30.89801	0.642466247
90	31.06	0.577889533	31.06012	-1.417360474	30.89801	0.642466247
91	31.06	0.577889533	31.06027	-1.417210474	30.89825	0.642542611
92	32.61	1.016791541	32.61015	0.132669526	32.8368	1.259353294
93	34.24	1.478346557	34.24013	1.762649526	**34.50837**	1.79121589
94	**0**	**8.217140398**	**2.078105**	**23.48781559**	**13.38482**	**4.929906245**
95	34	1.410387536	34.00015	1.522669526	34.32237	1.732034137
96	34	1.410387536	34.00015	1.522669526	34.32237	1.732034137
97	34	1.410387536	33.99997	1.522489526	34.09875	1.660882397
98	33.01	1.130056576	33.01014	0.532659526	33.29709	1.405809042

Table 8: Outlier Parameter Analysis for Julius Berger

System	μ	σ
Actual Close Price	29.01916	3.53154
ANN	29.02170053	3.455779944
FANN Model	28.87882814	3.142860608

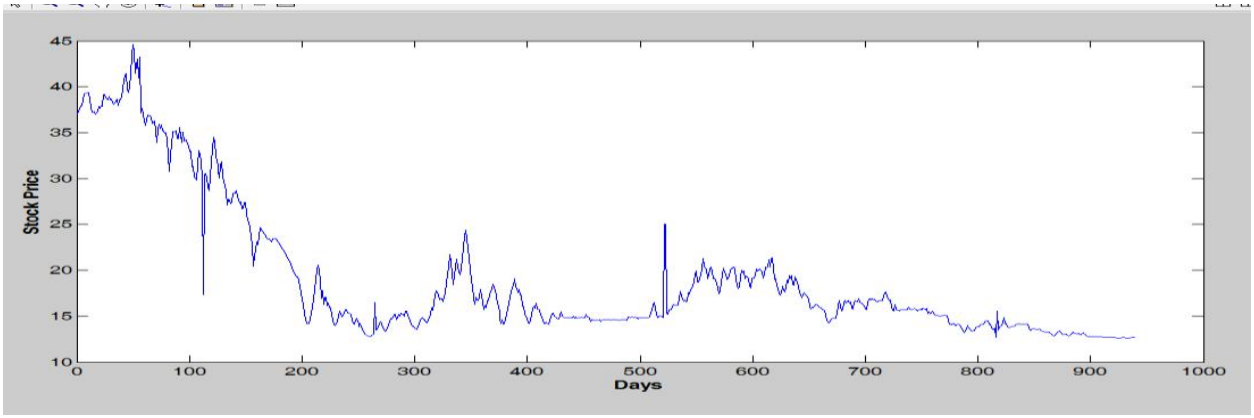


Fig.4: Graphical representation of FANN predictions for Dangote Sugar

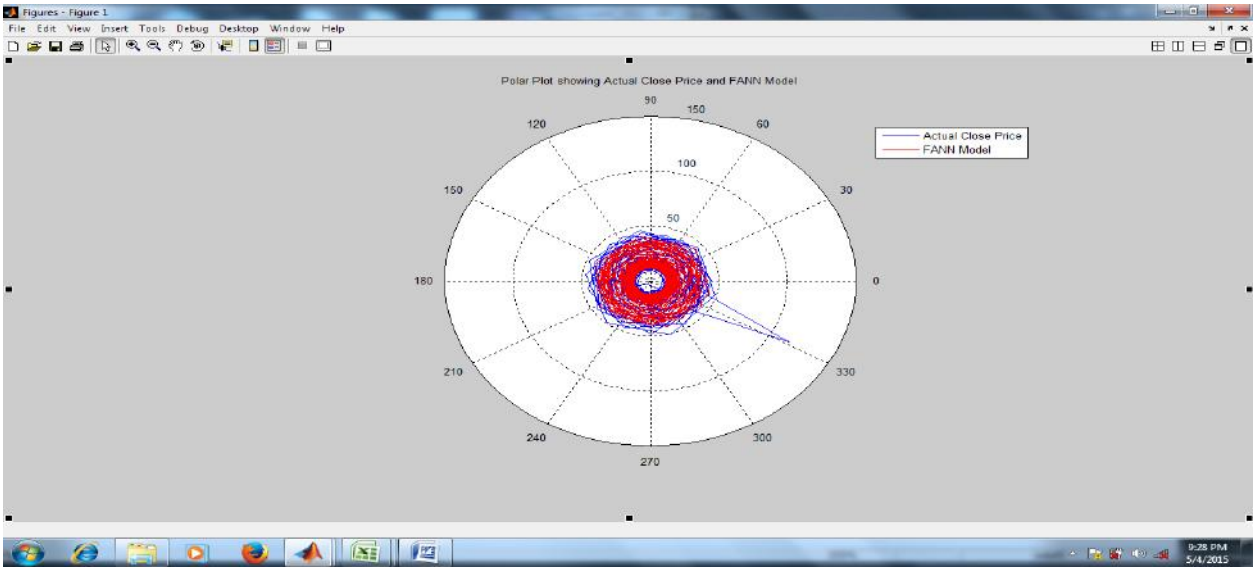


Fig.5: Polar plot showing Actual close price and FANN Model for Dangote Sugar

IV. RESULTS AND DISCUSSION

From Tables 3, 5 and 7, both systems can be used for predictions. Figures 3 and 4 show the graphical predictive capability and mitigating power of FANN respectively. The results analysis was based on the predictive performance using error analysis and its mitigation capability.

Error Analysis

The proposed model's predictive performance was evaluated using the Root Mean Square Error (RMSE) and compared with ANN as shown on Table 9.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i}-X_{model,i})^2}{n}} \quad \text{Eq.24}$$

Where $X_{obs,i}$ is observed values and $X_{model,i}$ is modeled values at time.

Table 9: Performance Evaluation using RMSE for Dangote Sugar

Model	RMSE
ANN	4.7308
FANN	3.8321

The proposed model returned a RMSE value of about 3.83 which shows that it competes favourably with the existing systems and it is a reliable stock forecasting model.

Mitigation Capability

From Tables 3 and 5, the outliers in the actual stock price were significantly reduced to values that returned very low Z-scores of about 0.78 and 1.90 respectively, that is, to the accurate neighborhood of prediction that are not outliers. From Table 7, the Z-score of FANN Model prediction for Julius Berger was significantly reduced from 8.217140398 to 4.929906245. This shows that the proposed model significantly attenuates outliers in stock forecasting. Overall, the results proved that the proposed FANN Model can handle outliers in stock forecasting. Therefore, with this model, stock forecasting will be more accurate and hence, more reliable.

V. CONCLUSION

Stock forecasting has always been an active area of research because successful prediction of a stock's future price could yield significant profit, thereby affecting the entire nation's economy. In order to handle outliers that are inherent in stock data as well as predict stock price on normal condition, a new solution was proposed. The proposed solution combined approaches in soft computing - Fuzzy and ANN to automatically find patterns from trading data. The idea of their hybridization originated from the fact that fuzzy logic is appropriate for uncertainty and ANN with the aim of learning.

The experimental results indicated the combinational strength of fuzzy logic with artificial neural network in attenuating outliers. Therefore, a model now exists that significantly attenuates outliers in stock forecasting.

REFERENCES

[1] Aggarwal, C.C. (2005). On Abnormality Detection in Spuriously Populated Data Streams SIAM Conference on Data Mining, Kluwer Academic Publishers Boston London.

[2] Barnett, V. and Lewis, T. (1994). Outliers in Statistical data. John Wiley & Sons, 3rd edition, Kluwer Academic Publishers Boston London.

[3] Dungan, J.L., Gao, D and Pang, A.T (2002). Definitions of uncertainty. Retrieved from ftp://.cse.ucsc.edu/pub/reinas/papers/white.pdf

[4] Eze, J.I, Obiegbu, M.E and Jude-Eze, E.N. (2005). Statistics and Quantitative Methods for Construction and Business Managers, The Nigerian Institute of Building, pp. 1- 402.

[5] Fausett, L. (1996). Fundamentals of Neural network: Architectures, Algorithms and Applications, Prentice Hall, Upper Saddle River, New Jersey 07458, pp. 1-14.

[6] Filmus, Y. (2010). Two Proofs of the Central Limit Theorem. Retrieved from www.cs.toronto.Edu/~yuvalf/CLT.pdf.

[7] Haag, S., Cummings, M. and Dawkins, J. (1998). Management Information System for the information Age, Mc-Graw-Hill, USA, pp. 526.

[8] Hagan, M.T., Demth, H.B. and Beale, M. (1996). Neural network design, PWS Publishing company.

[9] Hawkins, D. (1980). Identification of Outliers, Chapman and Hall, Kluwer Academic Publishers Boston London.

[10] Haykin, S. (1998). Neural networks: A comprehensive Foundation, Macmillan College Publishing company, Inc. USA, pp. 1-41.

[11] Investopedia (2013). Stock Market. Retrieved from www.investopedia.com/terms/s/stock market.asp#axzz2IjJu5E13

[12] Jatinder, R. P. (2012). Root cause Analysis of IMRT QA outliers. Retrieved from www.aapm.org/meetings/amos2/pdf35-9828-49165-839.pdf

[13] Mathworks (2014). The mathworks. Help Guide. Mathworks [online] mathworks. Retrieved from www.mathworks.com/products/neural.network/features.html

[14] Mendel, J.M (2001). Uncertainty in fuzzy logic systems. Retrieved from www.inFormit.com/articles/article.aspx?=21313

[15] Moghar, .A. and Hamiche, .M. (2020). Stock Market Prediction using LSTM Recurrent Neural Networks, Procedia Computer Science 170, pp. 1168-1173

[16] Noel, .D.(2023). Stock Price Prediction using Dynamic Neural Networks, Computational Engineering, Finance and Science (cs.CE)

[17] Ohuche, F.k. and Ikoku, A.E. (2015). Financial Management Focus on Price Volatility and Circuit Breakers in the Nigerian Equity Market Implications for Monetary Policy. Journal of Financial Management and Analysis Vol. 27 No. 2.

[18] Orr, J.M., Sacket, P.R. and DuBois, C.L. (1991). Outlier detection and treatment in I/OPsychology: A Survey of researcher beliefs and an empirical illustration. Personnel Psychology, No. 44, pp. 73-486.

[19] Osborne, J.W. and Overbay, A. (2004). The Power of Outliers (and why researchers should Always check for them). Practical Assessment Research and Evaluation, Vol. 9, No. 6.

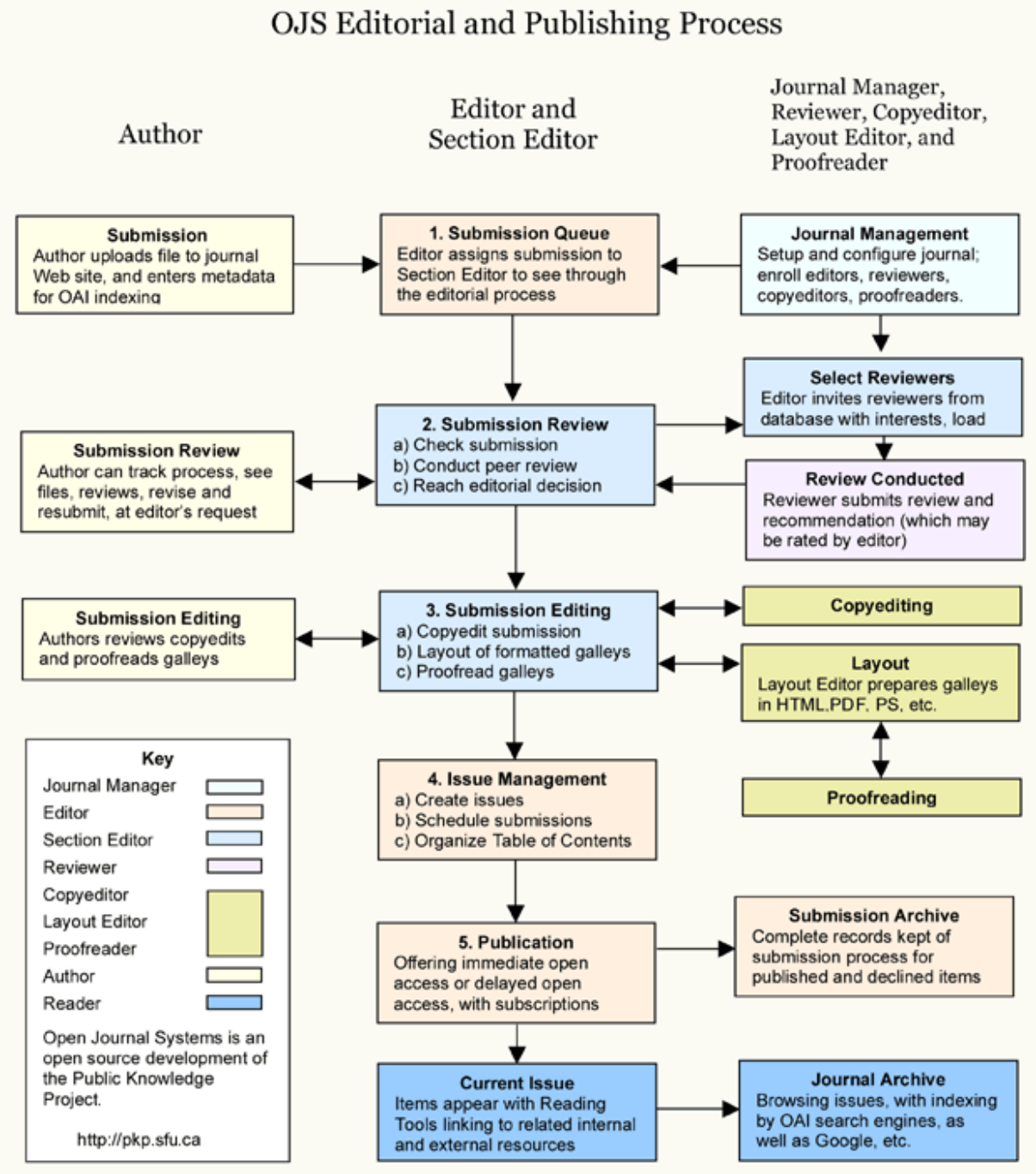
[20] Osborne, J.W. (2002). Notes on the use of data transformations. Practical Assessment Research and Evaluation, Vol. 8(6).

[21] Songwon, S.M.S. (2006). A Review and Comparison of Methods for detecting outliers in Univariate Data Sets. Retrieved from d-scholarship.pitt.edu/7948/1/seo.pdf.

[22] Thakkar, .A. and Chaudhari, .K. (2021). A Comprehensive Survey on Deep Neural Network for Stock Market: The need, Challenges and Future Directions, Expert System with Applications, Vol. 177

[23] Tsishchanks, K. (2010). Polar Coordinates. Retrieved from https://cims.nyu.edu/~kiryl /calculus/section_9.3--polar_coordinates/polar_coordinatinates.pdf

[24] Yanfang, L. (2014). Detection of Outliers in Panel Data of Intervention Effects Model Based on Variance of Remainder Disturbance, Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article 10902602.



~IJAERS Workflow~

Important links:

Paper Submission Link:
https://ijaers.com/submit-paper/
Editorial Team:
https://ijaers.com/editorial-board/
Peer Review Process:
https://ijaers.com/peer-review-process/
Publication Ethics:
https://ijaers.com/publication-ethics-and-publication-malpractice-statement/
Author Guidelines:
https://ijaers.com/instruction-to-author/
Reviewer Guidelines:
https://ijaers.com/review-guidelines/

Journal Indexed and Abstracted in:

- Qualis-CAPES (A2)-Brazil
- Normatiza (Under Review-Ref.020191511)
- NAAS Score: 3.18
- Bielefeld Academic Search Engine(BASE)
- Aalborg University Library (Denmark)
- WorldCat: The World's Largest Library Catalog
- Semantic Scholar
- J-Gate
- Open J-Gate
- CORE-The world’s largest collection of open access research papers
- JURN
- Microsoft Academic Search
- Google Scholar
- Kopernio - powered by Web of Science
- Pol-Index
- PBN(Polish Scholarly Bibliography) Nauka Polaska
- Scilit, MDPI AG (Basel, Switzerland)
- Tyndale University College & Seminary
- Indiana Library WorldCat
- CrossRef DOI-10.22161/ijaers
- Neliti - Indonesia's Research Repository
- Journal TOC
- WIKI-CFP
- Scinapse- Academic Search Engine
- Mendeley-Reference Management Software & Researcher Network
- Dimensions.ai: Re-imagining discovery and access to research
- Index Copernicus Value(ICV): 81.49
- Citeseerx
- Massachusetts Institute of Technology (USA)
- Simpson University (USA)
- University of Louisville (USA)
- Biola University (USA)
- IE Library (Spain)
- Mount Saint Vincent University Library (Halifax, Nova Scotia Canada)
- University Of Arizona (USA)
- INDIANA UNIVERSITY-PURDUE UNIVERSITY INDIANAPOLIS (USA)
- Roderic Bowen Library and Archives (United Kingdom)
- University Library of Skövde (Sweden)
- Indiana University East (campuslibrary (USA))
- Tilburg University (The Netherlands)
- Williams College (USA)
- University of Connecticut (USA)
- Brandeis University (USA)
- Tufts University (USA)
- Boston University (USA)
- McGill University (Canada)
- Northeastern University (USA)
- BibSonomy-The blue social bookmark and publication sharing system
- Slide Share
- Academia
- Archive
- Scribd
- ISRJIF
- Cite Factor
- SJIF-InnoSpace
- ISSUU
- Research Bib
- infobaseindex
- I2OR
- DRJI journal-repository

AI Publication
International Journal of Advanced Engineering Research and Science (IJAERS)
104/108, Sector-10, Pratap Nagar, Jaipur, India