AIRO JOURNALS
Peer Reviewed Multidisciplinary International

ISSN:2320-3714
Volume: 4 Issue: 3
December 2024
Impact Factor: 10.2
Subject: Engineering

CONSTRUCTION OF RISK EARLY WARNING MODEL FOR SPATIAL ECONOMIC NETWORK IN URBAN AGGLOMERATION AREAS UNDER MULTI-SOURCE HETEROGENEOUS DATA FUSION, AND EMPIRICAL RESEARCH BASED ON RANDOM FOREST

Monika Hinduja
Research Scholar

ABSTRACT - This paper discusses the development of a risk early warning model for spatial economic networks in urban agglomeration areas by employing multi-source heterogeneous data fusion techniques. It studies the effects of urbanization, economic linkages, infrastructure, and labor mobility on the resilience and risk factors associated with spatial economic networks. The empirical analysis and prediction of risk patterns within these networks are done by utilizing the Random Forest algorithm. Against various data sources, the model highlights the interdependence that exists between city nodes in making different risk possibilities that challenge those economic systems. The report postulates that robust infrastructure combined with labor mobility and sound economic integration are critical toward ensuring that urban agglomeration stays resilient. This study contributes to the progress of the risk prediction models and gives useful tools for the policymakers and urban planners to address the complexities of urbanization and the spatial economic network management.

Keywords: Risk prediction, policymaker, urbanization, network management

I. INTRODUCTION

Urban agglomeration areas and spatial economic networks represent an important idea in both urban studies and regional economics: the interrelatedness of towns, cities, and regions, in both physical and economic senses [1]. Here is a brief explanation of each:

A. Urban Agglomeration Areas

Urban agglomeration includes regions where several towns and cities, along with their suburbs, are closely linked with various socio-economic, infrastructural, and geographical factors [2]. The population growth in these areas usually happens very rapidly, economic activities are very huge, and infrastructural development is substantial. An urban agglomeration is not just a single city but is a larger region where multiple urban centers are interconnected sometimes through administrative boundaries [3]. Most of these regions have high population densities, and the migration process significantly contributes to changes in demography. More seriousness of concentration poses problems concerning only increased economic productivity but creates problems in housing, transport, and public services [4]. Urban agglomerations act as nodes for economic activities with a focus on different industries, businesses, and service sectors which promote innovations, job creation, and productivity. Infrastructure development, including transportation networks (roads, railways, airports), and utilities (water, energy, telecommunication), connects cities and suburbs and ensures the smooth flow of goods, services, and people [5]. Urbanization, however, may promote urban sprawl where cities sprawl into rural land and complex urban landscapes with mixed land uses that demand

ISSN:2320-3714
Volume: 4 Issue: 3
December 2024
Impact Factor: 10.2
Subject: Engineering

integrated planning and management. The major agglomeration Examples comprise of Delhi NCR-National Capital Region, Yangtze River Delta in China, and the Tokyo Metropolitan Area in Japan.

B. Spatial Economic Networks

Spatial economic networks can be defined as the interconnectedness of geographic locations or regions that are bound together by economic activities, including trade, investment, labor mobility, and the movement of goods and services [6]. Spatial economic networks represent a network that extends beyond one city or one region to link multiple areas together through both economic and functional relationships. In the urban agglomeration, spatial economic networks define the economic landscape through showing how cities and towns within the agglomeration, as well as those outside it, are economically connected. It influences regional growth and development through exchange of resources and services. Some of the core elements of spatial economic networks include economic linkages, which describe the activities of trade flows, supply chains, financial transactions between businesses or industries. The labor markets in urban agglomerations are relatively integrated, where workers can move freely between cities and regions, thus creating efficient human capital utilization [7]. In the transportation network, which consists of roads, railways, ports, and airports, there is a convenient transportation of goods and services that supports the spatial economic networks further. Lastly, there is a strong concentration of research institutions, universities, and industry clusters that exist within the urban agglomeration to foster knowledge and innovations, thus further enhancing economic tie strength and regional competitiveness [8]. These are networks connected to global systems, therefore, to international trade, foreign direct investment, and the proliferation of technologies. Major agglomerations such as New York, London, and Shanghai are key nodes in these global networks, showing how local economies are plugged into global markets. This, in general, allows cities and regions to draw on each other's resources, capital, and labor, providing means for growth, efficiency, and resilience while reflecting the interdependence of cities within an agglomeration and the broader economic dynamics of the region or nation.

C. OBJECTIVE OF THE STUDY

To establish early warning models for the spatial economic networks of agglomeration areas, using multi-source heterogeneous data fusion techniques

To use the Random Forest algorithms to conduct empirical analysis and prediction for patterns of risk in spatial economic networks of urban agglomeration

To examine the impacts of urbanization, economic ties, infrastructure, and labor mobility on the resilience and risk factors of the spatial economic networks of urban agglomerations.

II. LITERATURE REVIEW

Zhang et al. (2023) [9] demonstrated the usefulness of integrating information from hydrodynamic models, GIS, and satellite photography to assess the dangers of urban waterlogging. Their work made it possible to anticipate flood risk more precisely by combining these many data sources, which is essential for proactive urban flood management and prevention. The necessity for high-resolution, real-time data that can take into consideration the dynamic and complicated character of urban environments—where infrastructure, land use, and climatic factors interact in intricate ways—was one of the main challenges in urban risk management that this method addressed. This all-inclusive paradigm provided practical insights to reduce waterlogging and its socioeconomic effects, while promoting more resilient urban development.

He et al. (2021) [10] evaluated the Pearl River Delta's polycentric spatial structure using multi-source data fusion methods, with an emphasis on urban agglomeration patterns and their consequences for sustainable development. Big data, such as location-based services and other spatial data, was used in the research to provide a detailed analysis of the interactions and changes that take place in metropolitan centers in this quickly growing region. He et al. were able to determine the distribution and impact of different urban centers by integrating a variety of data sources, providing important insights into how metropolitan areas might more effectively control development, distribute resources, and improve connection between various locations. This research demonstrated how useful data fusion is for analyzing spatial organization and emphasized how it can be used to maximize land usage and encourage balanced urban growth..

Priyashani et al. (2023) [11] aimed at enhancing the accuracy of border demarcation and monitoring urban growth by using multi-source open geospatial big data fusion to define urban agglomeration footprints. To map metropolitan limits and pinpoint sections of the city that are expanding quickly, their research used satellite data, socioeconomic data, and spatial analytics. This strategy was essential for precisely identifying patterns of urban expansion, which are sometimes difficult to characterize using just conventional techniques. The research improved the capacity to detect and measure urban sprawl by integrating several data sources, giving policymakers and urban planners a solid foundation for future zoning and planning choices. This level of accuracy is especially useful for controlling the effects of urbanization and creating sustainable infrastructure that can support future expansion without endangering the welfare of the environment and society.

Hao et al. (2023) [12] demonstrated how to use multi-source data fusion to evaluate the danger of urban fires. They developed a model that offered a thorough mapping of the fire danger zones in Chengdu, China, by combining machine learning algorithms particularly designed for risk prediction with high-resolution spatial data. By examining trends that contribute to fire dangers, such as building density, land use, and environmental conditions, their research demonstrated how data fusion, when paired with machine learning, may significantly improve risk assessments. In highly populated metropolitan areas, this data integration supported attempts to enhance response techniques and implement preventative measures by enabling a more precise evaluation of fire hazards. The study emphasized how crucial it is to combine top-notch data sources with cutting-edge algorithms to produce risk maps that are accurate and flexible enough to adjust to changing urban situations.

III. RESEARCH METHODOLOGY

The descriptive research methodology used for this study takes a quantitative approach in assessing the relationship between urbanization, economic linkages, infrastructure, labor mobility, and resilience. Data is gathered by surveys, government reports, and public databases through a stratified random sample of 110 observations. Descriptive and inferential statistical techniques are conducted to verify the said relationships between variables through chi-square tests, Pearson's correlation analysis, and multiple regression analysis. Then a Random Forest Classifier is used in predictive modelling of resilience. In terms of ethical considerations, there are consent and confidentiality among others. Aims include bringing out significant insights into what determines resilience within the sphere of urban areas.

A. RESEARCH DESIGN

This research uses a descriptive and correlational design. It mainly employs a quantitative method to analyze data regarding urbanization, economic linkages, infrastructure, labor mobility, and resilience. The study will make use of a stratified random sample with 110 observations, where the sources for data will be primary-gathered from data surveyed, reports submitted by the government, and public databases. Descriptive and inferential statistics such as chi-square tests, Pearson's correlation, as well as multiple regression analysis will be applied on the relationships between variables while utilising predictive modeling using a Random Forest Classifier for forecasting resilience. The study encompasses ethical issues such as informed consent and confidentiality. Generally, the research will consider what factors influence resilience in cities.

B. DATA COLLECTION

Relevant data needed for this study were collated using personal surveys, reports, as well as government public databases. The most significant core areas of collecting data based on the specified variables will range from urbanization, the economic linkages, to infrastructure, labour mobility as well as the resilience elements measured in different scales as;

- Urbanization has different percentage ranges categorized as in 50-59, 60-69%, 70-79%, 80-89%, as well as 90-99% and so to reflect the variability of various scale levels.
- Economic Linkages (Trade Volume): These can be categorized into five according to the ranges (0-1999, 2000-3999, 4000-5999, 6000-7999, and 8000+).
- •Infrastructure: Rating based on 0-10 point scale with ranges: 4-5, 6-7, and 8-9.
- •Labor Mobility (Number of Migrants): Rated within the following three ranges (Less than 999, 1000-1999, More than 2000).
- •Resilience (Risk Factor): This also includes rating on a 1-10 point scale into four categories: 1-3, 4-5, 5-7, and 8-9.

The number of observations needed was 110 to have full representation of the various categories of each variable.

C. DATA ANALYSIS

AIRO JOURNALS
Peer Reviewed Multidisciplinary International
ISSN:2320-3714
Volume: 4 Issue: 3
December 2024
Impact Factor: 10.2
Subject: Engineering

The descriptive statistics of the study indicate that some key information is gathered about the sample population. Urbanization is uniformly distributed in five percentage ranges, each of which contributes 20% of the total sample, thereby implying a balanced representation of different levels of urbanization. The linkages of the economy are more concentrated at the mid-range trade volume categories: 4000-5999 and 6000-7999, representing 30% of the observations for each of the above categories, whereas the lower end of 0-1999 and the higher-end categories of 8000+have fewer occurrences and, thus, only moderate trade activity. Infrastructure scores appear in the middle range at 40% of the sample, while scores are split between the lower and higher ranges at 30% each to suggest generally average to highly developed infrastructure. Labor mobility shows important concentrations in the 0-1999 and 1000-1999 migration ranges at 40% each, indicating middling levels of migration. This is tempered by only 20% of the population having higher levels of migration (2000+). Resilience is generally middle level. That is, about 70% of the recorded observations show that the level of resilience in this population, in general, tends to fall within a moderate to high margin.

IV. DATA ANALYSIS

The data in Table 1 describe the ranges of percentages by which the rates of urbanization vary, along with their frequency and percentage. A peculiar feature of this distribution is that it is even throughout the different percentage ranges that exist-from 50-59% to 60-69%, 70-79%, 80-89%, and finally 90-99%-each containing 22 occurrences, which represents 20% within each category. This equal range implies that there is no dominating range of urbanization since each possesses an equal portion of the total number of observations, which denotes an equal spread of levels of urbanization over the five categories.

TABLE 1: URBANIZATION

Urbanization (%) Range	Frequency	Percentage (%)
50-59	22	20%
60-69	22	20%
70-79	22	20%
80-89	22	20%
90-99	22	20%
Total	**110**	**100%**

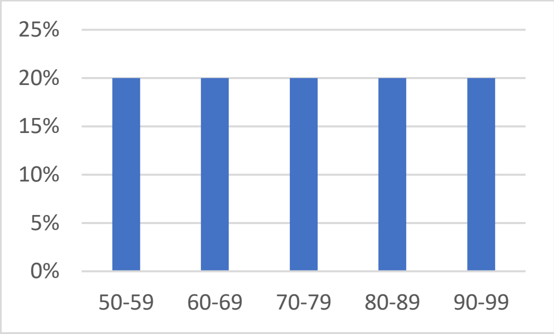

FIGURE 1: GRAPHICAL REPRESENTATION ON THE PERCENTAGE OF URBANIZATION

A balance across these ranges would imply that the population sampled is experiencing a spread of levels of urbanization from moderate (50-59%) through to high (90-99%) without concentrations in individual categories. The total sample of 110 observations, with equal numbers at each level of urbanization, might reflect a design choice in balancing sampling across levels or indicate real patterns of urbanization in the population under study. In short, the dataset for this group indicates near uniformity among different levels of urbanization, and thus disperses broad representation across all stages.

TABLE 2: ECONOMIC LINKAGES (TRADE VOLUME) RANGE

Economic Linkages (Trade Volume) Range	Frequency	Percentage (%)
0-1999	22	20%
2000-3999	11	10%
4000-5999	33	30%
6000-7999	33	30%
8000+	11	10%
Total	**110**	**100%**

AIRO JOURNALS
Peer Reviewed Multidisciplinary International
ISSN:2320-3714
Volume: 4 Issue: 3
December 2024
Impact Factor: 10.2
Subject: Engineering

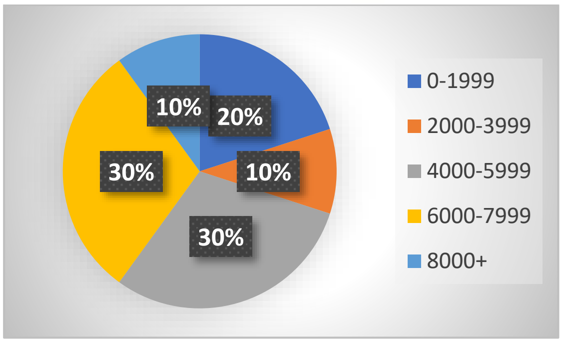

FIGURE 2: GRAPHICAL REPRESENTATION ON THE PERCENTAGE OF ECONOMIC LINKAGES

Table 2, Trade volume ranges for economic linkages, clearly shows that it is not uniform on its frequency cutting across the categories but dominantly highly dominated in some. The 4000-5999 and 6000-7999 both dominate with a 30% frequency in each, which indicates that this makes up a huge percentage of the volume of trade. The trade volume range of 2000-3999 and 8000+ each takes up 10% as a concern with 11 observations each, indicating that such large volumes do not occur as frequently. At the other end, the smallest trade volume range of 0-1999 takes up 20% of all observations; therefore, relatively balanced trade activity in this range is observed. The total number of observations is 110; most of the economic linkages are concentrated in the mid-range volumes, whereas the lower and higher ranges have relatively fewer observations. Thus, the data suggests that the modal group of this population concentrates its majority of economic linkages involving moderate to higher trade volumes and few linkages in the extremes of the trade volume spectrum.

TABLE 3: INFRASTRUCTURE SCORE

Infrastructure Score (0-10)	Frequency	Percentage (%)
4-5	33	30%
6-7	44	40%
8-9	33	30%
Total	**110**	**100%**

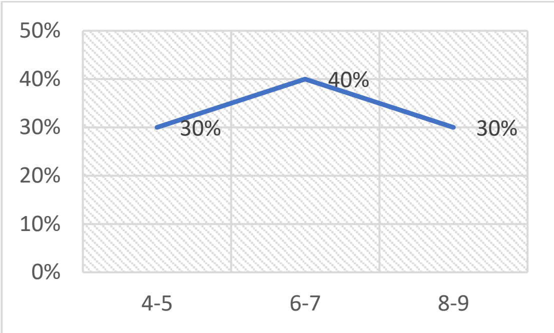

FIGURE 3: GRAPHICAL REPRESENTATION ON THE PERCENTAGE OF INFRASTRUCTURE SCORES

Table 3 Infrastructure Scores: Distribution of scores, divided into three segments - 4-5, 6-7, and 8-9. This distribution of scores shows that the observations primarily lie in the middle range, which is 6-7. This translates to roughly 40% in terms of frequency. This in turn causes a conclusion to be drawn that the type of infrastructure in the population studied tends to be of the moderately developed kind. Both the 4-5 and the 8-9 range equally contribute to 30% of the total observations, which reflects relatively balanced distribution between lower and higher infrastructure scores. The scores in the range of 4-5 reflect a cut from the population who face below average infrastructural conditions, but those in the 8-9 range reflect more developed infrastructural conditions. It is a total of 110 observations with a balanced mix of lower, middle, and higher ranges of infrastructure scores. Overall, this data can adequately justify the opinion that infrastructure is evenly spread across these three categories with the slight concentration in the moderate range (6-7), meaning an average to good level of infrastructure within the population under consideration.

TABLE 4: LABOR MOBILITY (NUMBER OF MIGRANTS) RANGE

Labor Mobility (Number of Migrants) Range	Frequency	Percentage (%)
Less than 999	44	40%
1000-1999	44	40%
More than 2000	22	20%
Total	**110**	**100%**

AIRO JOURNALS
Peer Reviewed Multidisciplinary International
ISSN:2320-3714
Volume: 4 Issue: 3
December 2024
Impact Factor: 10.2
Subject: Engineering

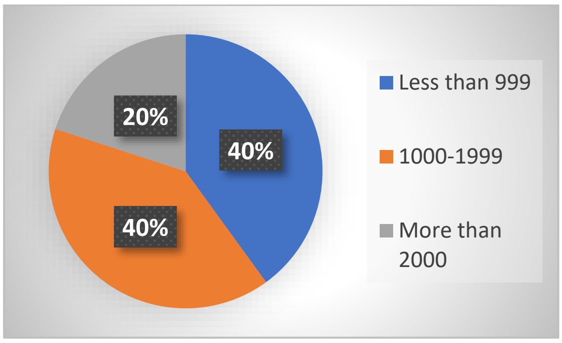

FIGURE 4: GRAPHICAL REPRESENTATION ON THE PERCENTAGE OF LABOR MOBILITY

Table 4 provides information on labour mobility, aggregated by number of migrants. The distribution is highly concentrated in the lower to middle ranges; both 0-999 and 1000-1999 ranges represent 40% of all observations. This implies that most labor mobility falls into these two ranges: moderate to high in the surveyed population. By the 2000+ range, which makes up 20% of the total, it shows less people fall into this category; thereby indicating that there is a proportion of people where migrants take up less space. There are a total of 110 observations and the general trend is that it is quite well balanced between the lower and middle ranges. The higher migration portion has relatively few people. This distribution shows that the overall migration of the populations in the region is mostly moderate in nature with an even smaller portion of the population migrating over a wider area.

TABLE 5: RESILIENCE

Risk Factor (1-10)	Frequency	Percentage (%)
1-3	11	10%
4-5	44	40%
5-7	33	30%
8-9	22	20%
Total	**110**	**100%**

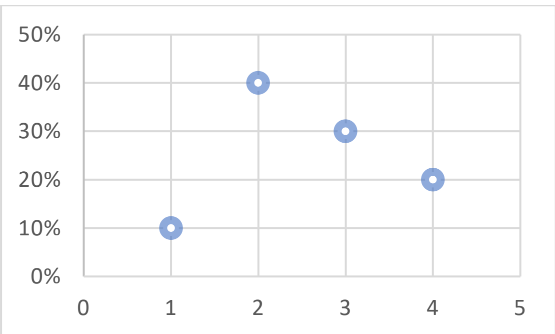

FIGURE 5: GRAPHICAL REPRESENTATION ON THE PERCENTAGE OF RESILIENCE

Table 5 reports data on resilience, grouped according to scores by risk factor. The distribution is such that most scores are concentrated in middle ranges, with the 4-5 range accounting for 40% of total observations; this means that the great majority of people in the population possess a moderate level of resilience. The 5-7 range accounts for 30% of the total sample, indicating that a large majority of the population has a level of resilience that is relatively high. The range of 8-9 accounts for 20% of the total sample, representing a very small group with an even greater degree of resilience. Conversely, a range of 1-3 accounts for a paltry 10% of the total sample, which means only a small proportion of the population, is at a relatively low level of resilience. There were 110 total observations for the analysis, and the data strongly indicate that resilience usually is concentrated at the moderate to high end of the spectrum, with less being distributed at the lower end. This extreme distribution of this population relative to resilience suggests that this is a very resilient population in which most demonstrate at least moderate degrees of resilience.

V. CONCLUSION

The overall purpose of the study is the provision of proper emphasis to the critical role played by multi-source data fusion techniques, especially the application of Random Forest algorithms, in assessing the patterns of resilience and risk within spatial economic networks of urban agglomerations. These analyses of multifaceted interconnected factors like urbanization, the linkages between the agents of economic transactions, the growth of infrastructure, and labor mobility thus depict the shapes of collective resilience and the vulnerability of urban areas. It has facilitated a better reflection of how changes in one area, be it economic fluctuations, infrastructural expansion, or population shift, could affect the larger urban network. The outcomes show the necessity of an integrated approach to urban planning that forms a process of risk assessment and management incorporating data-driven insights. The research also highlights the demand for advanced analytics and predictive models in the improvement of the quality of risk forecasting regarding challenges related to fast-growing urban agglomerations. Finally, the research presents a richer framework for enhancing

ISSN:2320-3714
Volume: 4 Issue: 3
December 2024
Impact Factor: 10.2
Subject: Engineering

urban resilience and guiding policymakers and city planners toward more sustainable and adaptable ways of handling urban growth and development under pressure from the environment, economics, and society.

REFERENCES

1) X. He, X. Yuan, D. Zhang, R. Zhang, M. Li, and C. Zhou, "Delineation of urban agglomeration boundary based on multisource big data fusion—A case study of Guangdong–Hong Kong–Macao Greater Bay Area (GBA)," Remote Sensing, vol. 13, no. 9, p. 1801, 2021.
2) Wang, J. Zou, X. Fang, S. Chen, and H. Wang, "Using Social Media and Multi-Source Geospatial Data for Quantifying and Understanding Visitor's Preferences in Rural Forest Scenes: A Case Study from Nanjing," Forests, vol. 14, no. 10, p. 1932, 2023.
3) L. Chen, X. Ge, L. Yang, W. Li, and L. Peng, "An Improved Multi-Source Data-Driven Landslide Prediction Method Based on Spatio-Temporal Knowledge Graph," Remote Sensing, vol. 15, no. 8, p. 2126, 2023.
4) X. Guan, J. Li, C. Yang, and W. Xing, "Development Process, Quantitative Models, and Future Directions in Driving Analysis of Urban Expansion," ISPRS Int. J. Geo-Inf., vol. 12, no. 4, p. 174, 2023.
5) M. Ren, Z. Zhang, J. Zhang, and L. Mora, "Understanding the use of heterogeneous data in tackling urban flooding: an integrative literature review," Water, vol. 14, no. 14, p. 2160, 2022.
6) K. Yao, S. Yang, S. Wu, and B. Tong, "Landslide susceptibility assessment considering spatial agglomeration and dispersion characteristics: a case study of Bijie City in Guizhou Province, China," ISPRS Int. J. Geo-Inf., vol. 11, no. 5, p. 269, 2022.
7) Z. Bu, J. Fu, D. Jiang, and G. Lin, "Production–Living–Ecological Spatial Function Identification and Pattern Analysis Based on Multi-Source Geographic Data and Machine Learning," Land, vol. 12, no. 11, p. 2029, 2023.
8) Chen, Y. Wang, Z. Shen, J. Liao, J. Chen, and S. Sun, "Long time-series mapping and change detection of coastal zone land use based on Google Earth Engine and multi-source data fusion," Remote Sensing, vol. 14, no. 1, p. 1, 2021
9) Z. Zhang, Y. Zeng, Z. Huang, J. Liu, and L. Yang, "Multi-source data fusion and hydrodynamics for urban waterlogging risk identification," Int. J. Environ. Res. Public Health, vol. 20, no. 3, p. 2528, 2023.
10) X. He, Y. Cao, and C. Zhou, "Evaluation of polycentric spatial structure in the urban agglomeration of the pearl river delta (PRD) based on multi-source big data fusion," Remote Sensing, vol. 13, no. 18, p. 3639, 2021.
11) N. Priyashani, N. Kankanamge, and T. Yigitcanlar, "Multisource open geospatial big data fusion: application of the method to demarcate urban agglomeration footprints," Land, vol. 12, no. 2, p. 407, 2023.
12) Y. Hao, M. Li, J. Wang, X. Li, and J. Chen, "A High-Resolution Spatial Distribution-Based Integration Machine Learning Algorithm for Urban Fire Risk Assessment: A Case Study in Chengdu, China," ISPRS Int. J. Geo-Inf., vol. 12, no. 10, p. 404, 2023.

AIRO JOURNALS
Peer Reviewed Multidisciplinary International
ISSN:2320-3714
Volume: 4 Issue: 3
December 2024
Impact Factor: 10.2
Subject: Engineering

Author's Declaration

I as an author of the above research paper/article, here by, declare that the content of this paper is prepared by me and if any person having copyright issue or patent or anything otherwise related to the content, I shall always be legally responsible for any issue. For the reason of invisibility of my research paper on the website /amendments /updates, I have resubmitted my paper for publication on the same date. If any data or information given by me is not correct, I shall always be legally responsible. With my whole responsibility legally and formally have intimated the publisher (Publisher) that my paper has been checked by my guide (if any) or expert to make it sure that paper is technically right and there is no unaccepted plagiarism and hentriacontane is genuinely mine. If any issue arises related to Plagiarism/ Guide Name/ Educational Qualification /Designation /Address of my university/ college/institution/ Structure or Formatting/ Resubmission /Submission /Copyright /Patent /Submission for any higher degree or Job/Primary Data/Secondary Data Issues. I will be solely/entirely responsible for any legal issues. I have been informed that the most of the data from the website is invisible or shuffled or vanished from the database due to some technical fault or hacking and therefore the process of resubmission is there for the scholars/students who finds trouble in getting their paper on the website. At the time of resubmission of my paper I take all the legal and formal responsibilities, If I hide or do not submit the copy of my original documents (Andhra/Driving License/Any Identity Proof and Photo) in spite of demand from the publisher then my paper maybe rejected or removed from the website anytime and may not be consider for verification. I accept the fact that as the content of this paper and the resubmission legal responsibilities and reasons are only mine then the Publisher (Airo International Journal/Airo National Research Journal) is never responsible. I also declare that if publisher finds Any complication or error or anything hidden or implemented otherwise, my paper maybe removed from the website or the watermark of remark/actuality maybe mentioned on my paper. Even if anything is found illegal publisher may also take legal action against me.

Monika Hinduja

*****