Pak. j. sci. ind. res. Ser. A: phys. sci. 2026 **69A**(2) 73-85 https://doi.org/10.52763/PJSIR.PHYS.SCI.69.2.2026.73.85

GIS Overview of Public School of District Lahore Using Spatial Math and AI

Tayyaba Abbas*, Muhammad Azam and Muazzam Ali
Superior University Lahore, Punjab, Pakistan

(received July 25, 2024; revised March 26, 2025; accepted April 04, 2025)

Abstract. This research paper, therefore, tries to identify the educational gaps that exist within Lahore in Pakistan, integrating geospatial analysis with AI tools to draw the blueprint for an equitable and accessible public education system. The study is based on census 2022-2023 dataset of 1120 public schools of Lahore. To analyze the school's education disparity researcher operationalized twenty-six independent and education disparity is one dependent variable from the census dataset encoded to the geodata frame. Dataset of schools was spatially visualized on map and identified cluster of schools covering district Lahore. OLS regression highlighted, that drinking water, security measures, toilets and dangerous classrooms were confounding variables. Hotspot analysis highlighted male (ED value 9.4) schools face more education disparity than female (ED value 9.232). However, urban settings including males and females face more education disparity as compared to rural areas. Proximity analysis identified the 279 schools with the highest nearest neighbor distance having an average area covered 7,282.57 square feet 8 working classrooms and an average of students per school is 378 showed less accessibility. Predictive analysis using AI highlighted 728 additional teachers were required in Lahore school.

Keywords: artificial intelligence, proximity analysis, mapping, resources allocation, spatial mathematics

Introduction

In Lahore's dynamic urban environment, a city with rapid population growth and socio-economic transformation. The population of Lahore as of 2024 is estimated about 14407074 from which 586505 students are studying in public schools according to census 2022-23 that is only 4% of total population which is less from previous year as in 2021 this population in schools was 610,697. (Haider *et al*., 2024) Lahore, a cosmopolitan city with a growing population, faces an ongoing task of adapting its education systems to meet the evolving aspirations of its people. Geospatial assessment provides powerful tools to cope with this task. (Chen *et al*., 2019) The use of spatial mathematics furthermore, the integration of AI-driven methodologies enhances the precision and predictive abilities of the evaluation. Lahore, a vast city in Pakistan, is located alongside the banks of the Ravi river, Lahore holds a strategic role in the coronary heart of the Punjab area. Its geographical coordinates vary approximately from 74°14′ to 74°37′ east longitude and 31°22′ to 31°34′ north range. Covering a diverse landscape, Lahore serves as a cultural, historical and economic hub within Pakistan. The metro region population of Lahore has skilled a steady growth, attaining 13,542,000 in 2022 (a 3.41% boom from 2021), 13,979,000 in 2023 (a 3.23% increase from 2022) and is projected to be 14,407,000 in 2024, marking a 3.06% increase from the preceding 12 months. According to census 2022-23 there were 1100 public schools in Lahore. Traditional methods of allocating resources to public schools were often based on historical facts and demographic forecasts. Geospatial analysis, a discipline that integrates geographic statistical systems (GIS), spatial data and statistical models, provides a powerful tool for knowledge narrative, spatial relationships and modeling spatial statistics, in this context facilitating in-depth knowledge of the distribution of Lahore public schools. Artificial intelligence (AI) has changed many things and education is no exception. AI-driven processes have gained popularity in recent years to optimize interventions, predict instructional improvements and improve strategic choices The aims of the study is to identify spatial disparities: pinpoint areas with inadequate get entry to public schools and those with surplus ability, highlighting spatial disparities, to evaluate accessibility present schools and predict future demand by utilize gadgets to get to

*Author for correspondence;
E-mail: tayyabaabbas9090@outlook.com

know fashions to predict destiny demand for training primarily based on population growth, demographic changes and different relevant elements. Education researchers and policy makers all over the world feel an increasing interest in investigating the gearing factors of student performance and results across varying educational contexts. (Zartashia *et al.*, 2024) The purpose of this literature review is to bring together recent studies that investigate factors influencing student attainment-of variables starting from school infrastructure to teacher qualifications, parental involvement and spatial distribution of educational institutions. According to (Engin-Damir *et al.*, 2009) he researched public schools in Turkish context where he noted that there was a significant correlation between the quality of the infrastructure and the academic score of students. This goes on to underline one major factor: good facilities as support for effective learning environments. Research views parental involvement as one of the major determinant factors of students' success. (Alothman *et al.*, 2024) used Saudi Arabia to explain how parental involvement contributes to positive academic outcomes and went ahead to demonstrate ways of building closer ties between schools and homes. Back home in Ghana, (Antwi, 2024) looked into some of the determinants that accounted for students' performance in secondary schools, unravelling kinds of factors outside infrastructure and parental involvement that play a critical role in educational outcome. Similarly, research has been done on the relationship between teacher qualifications and student performance. (Ahmad *et al.*, 2023) did his study in Pakistan and proved that the level of competence of the teacher greatly influenced the level of achievement among the students and thus there was the reason for continuous professional development programs. (Nawaz *et al.*, 2022) Researched how school physical infrastructure influences the outcome in students in the urban parts in Pakistan and provided an empirical proof that infrastructure investments correlate with improved educational achievements. (Yaakub *et al.*, 2022) Examined factors affecting undergraduate student performance in Malaysia. This research emphasizes the multifaceted nature of educational success, which is influenced by socio-economic factors and learning environments. (Sele *et al.*, 2024) Conducted a comparative study of urban and rural public secondary schools in Nigeria. It was realized that there were huge disparities in resources and academic outcomes for these; therefore, underpinning the importance of having responsive and equitable educational policies. The work of (Singh *et al.*, 2021) from India, which focused on the role of technology in enhancing student learning outcomes, calls for innovative approaches to learner learning, including the use of technology to improve student engagement in and achievement of learning during COVID-19. On their part, these studies identify different factors that have borne on educational outcomes: from infrastructure and technology through socio-economic factors and regional inequalities to global inequalities. It is the dimensions shown here that future research will go through from time to time, enlightened by evidence-based policies in augmenting equity in education and providing quality education for all. (Stone *et al.*, 2021) worked on predicting student success in general chemistry courses from a large public university, outlining the role of academic preparedness, study habits, and instructor effectiveness in driving academic achievement within specified disciplines. (Adewoye *et al.*, 2022) Searched for the effect of computer-assisted instructions on the higher secondary students' academic achievement in India, thus driving home the dormant potential of technology-aided instructions in bringing out the best in student outcomes. (Kumar *et al.*, 2023) Authors used GIS analysis to find the spatial distribution of secondary schools in Quetta city, Pakistan by monitoring water quality and identified geographical factors that affect access to education and educational outcomes. These authors tried to find out the determinants of educational outcomes at the north eastern region of India and manifested the socio-economic, cultural and infrastructural factors which influence the pupil performance in underserved regions. (Fahd *et al.*, 2022) Applied machine learning techniques in predicting students' performance to predict academic outcomes and tailor interventions through advanced analytics in educational research. (Khuma *et al.*, 2023) Tried to signify the determinants of student academic performance regarding the higher institutions of education. In particular, he showed institutional factors, student characteristics and pedagogy, all of which were critical for educational success. (Dabhade *et al.*, 2021) evaluated the various factors that determine the academic achievement of technical institutes in India, using machine learning tools. (Alam *et al*., 2021) conducted an empirical investigation of the infrastructure of schools and student achievement concerning Bangladesh. His findings proved a positive relationship between the two-a relation between the infrastructure and academic achievement. (Meena *et al*., 2023) studied the spatial distribution of rural and urban India's schools to conclude that the access existed and mars equity in terms of quality in education. (Kapoor *et al*., 2021) studied how the school's physical environment influenced the academic achievement of pupils at various levels of schooling in India. (Peprah *et al*., 2024) conducted a cross-national study regarding the relationship between teacher qualifications and pupil achievement in sub-Saharan Africa; it emerged that improvement in educational outcomes requires quality input in teacher training and continuous professional development. (Sharma *et al*., 2022) Contributed an empirical study of the infrastructure of the school and the performance of students in India. The existing literature shows that adequate facilities were positively correlated with educational outcomes. It is in connection with this that (Sajjad *et al*., 2022) were reviewing the spatial inequality of education in Pakistan and insisting that targeted policies be designed to achieve sustainable development goals as regards regional disparities. The running thread in all this educational research points to how much infrastructure is related to student performance. (Zaman *et al*., 2023) used a machine learning-based approach to model the impact of educational interventions in enhancing school quality, therefore underlining potential use of advanced analytics in improving education practice.

Material and Methods

AI technology, along with predictive analytics and device mastering, play a pivotal characteristic in optimizing resource allocation and improving academic get access. Predictive models forecast pupil enrollment, understand at-danger populations and advocate interventions to decorate fairness. Geospatial Information Systems (GIS) facilitate spatial assessment, mapping and visualization of instructional infrastructure. GIS tools investigate accessibility, demographic trends and spatial disparities, informing city selection-making choices. The convergence of AI and GIS complements choice-making abilities through manner of integrating predictive insights with spatial facts. This synergy helps a proof-based coverage device and fosters sustainable city development.

Framework

AI in education equity

Integeration of AI and Geospatial analysis

Predictive analysis and AI technologies

Geospatial Analysis

GIS tools for spatial analysis and mapping

Data collection. The primary data of geographic coordinates (longitude and latitude) of all schools in the Lahore district were collected from Google maps by searching their respective street addresses. The district monitoring office in Lahore was of great help for the researcher to obtain secondary data from the school census of 2022-23. The dataset consisted of 1121 records and 149 columns, covering information of schools, from student demographic details and teacher strength to the facilities available in the school at that time. Convenience sampling technique were used while collecting data from Google maps. The raw CSV-form data was imported in Python and was pre-processed further. In the initial phase of cleaning, any extra columns irrelevant of the research were deleted from the data.

Data encoding. The categorical variables in the dataset were encoded using the one-hot encoding method to facilitate analysis and visualization of data. The categorical columns contain data about the ownership and physical condition of the school's building and whether it in rural or urban region. The numerical representation of encoding is '0' for 'No' and '1' for 'Yes' with regard to different features of the schools. With respect to incorporating spatial analysis requirements, geometries were incorporated, which showed the geographical point locations of the schools.

This conversion enable the researcher to visualize and map the data on base map of Lahore. To make the dataset capable for spatial analysis; the type of the DataFrame is converted to GeoDataFrame so that it could be much easier for the researcher to have better level of mapping approach. Education disparity (ED) is the dependent variable applied to analyze the school performance and comprehension of educational inequalities in the district. A number of independent variables related to school resources and facilities were tested for their influence on the ED values. The independent variables were:

f_c=functional classrooms
d_c=dangrous classrooms
h_e=has electricity (binary variable)
h_w=has drinking water (binary variable)
h_t=has toilets(binary variable)
h_b = has boundarywall (binary variable)
h_g=has main gate (binary variable)
h_s=has seawrage (binary variable)
h_p=has playground(binary variable)
h_o=has open classrooms(binary variable)
S_a=has security arrangments(binary variable)
h_{sb}=has smart board(binary variable)
h_h=has head(binary variable)
t_{sc}=total sections
t_s=total students
t_t=total teachers

Education disparity (ED) calculated on the bases of the following conditions:

$$ED_1=\begin{cases}1 \text{ if } f_c\neq t_{sc}\\0 \text{ if } f_c=t_{sc}\end{cases}, \quad ED_2=\begin{cases}1 \text{ if } d_c>0\\0 \text{ if } d_c=0\end{cases},$$

$$ED_3=\begin{cases}1 \text{ if } h_o>0\\0 \text{ if } h_o=0\end{cases} \quad ED_4=\begin{cases}1 \text{ if } h_w=0\\0 \text{ if } h_w=1\end{cases},$$

$$ED_5=\begin{cases}1 \text{ if } h_e=0\\0 \text{ if } h_e=1\end{cases}, \quad ED_6=\begin{cases}1 \text{ if } h_t=0\\0 \text{ if } h_t=1\end{cases},$$

$$ED_7=\begin{cases}1 \text{ if } h_b=0\\0 \text{ if } h_b=1\end{cases}, \quad ED_8=\begin{cases}1 \text{ if } h_g=0\\0 \text{ if } h_g=1\end{cases}$$

$$ED_9=\begin{cases}1 \text{ if } h_s=0\\0 \text{ if } h_s=1\end{cases}, \quad ED_{10}=\begin{cases}1 \text{ if } h_s=0\\0 \text{ if } h_s=1\end{cases},$$

$$ED_{11}=\begin{cases}1 \text{ if } h_p=0\\0 \text{ if } h_p=1\end{cases}, \quad ED_{12}=\begin{cases}1 \text{ if } h_l=0\\0 \text{ if } h_l=1\end{cases}$$

$$ED_{13}=\begin{cases}1 \text{ if } s_a=0\\0 \text{ if } s_a=1\end{cases}, ED_{14}=\begin{cases}1 \text{ if } h_{sl}=0\\0 \text{ if } h_{sl}=1\end{cases},$$

$$ED_{15}=\begin{cases}1 \text{ if } h_{cl}=0\\0 \text{ if } h_{cl}=1\end{cases}, \quad ED_{16}=\begin{cases}1 \text{ if } h_i=0\\0 \text{ if } h_i=1\end{cases}$$

$$ED=\sum_{i=1}^{n} ED_i$$

A preview of calculated education disparities on the bases of variables is shown in Fig. 1. All schools geographic data were mapped on Lahore base map to visualize the different schools facing education disparities. The school locations, which were overlaid onto a base map of Lahore to form spatial context and geographical reference upon which the analysis was done. In hot spot analysis the Getis-Ord Gi∗G_i^*Gi∗ statistic were used.

$$G_i^*=\frac{\sum_{j=1}^{n} w_{i,j}-\bar{X}\sum_{j=1}^{n} w_{i,j}}{s\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}^2-\left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}}$$

In density analysis following formulas for area and standard deviation $\mu_A=1/n\sum_{i=1}^{n} nAi \approx 11{,}111$ sqfeet $\sigma_A=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i-\mu_A)^2}\approx 15{,}203$ sqfeet were used and the proximity analysis were done under following calculation:

$$\mu_{NND}=\frac{1}{n}\sum_{i=1}^{n} NND_i \approx 0.0095$$

$$\sigma_{NND}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(NND_i-\mu_{NND})^2} \approx 0.0046$$

Predictive analysis following model were used by the researcher

$$ED=\beta_0+\beta_1\times student+\beta_2\times Teachers+\epsilon$$

Where:
β_0, β_1, β_2 are coefficients of model ϵ is error term
Using the model, it was possible to predict the ED values for each school in the dataset.

Results and Discussions

Geographic visualization of schools. Using the python libraries, some specifically used include Geopandas, Shapely.geometry, Pandas, Contextily and Matplotlib.Pyplot to geographically visualize the cleaned data set for schools on a map of Lahore. It forms the basic exercise before one can actually start making sense of the spatial distribution and clustering of schools across the district. Geopandas and its functionalities were used to handle and analyze geospatial data in such a way that geometries and attributes can be manipulated relatively efficiently.

```
Columns in the DataFrame:
Index(['school id', 'School Names', 'tehsil', 'latitude', 'longitude',
       'official school type', 'Area', 'have building', 'building owner',
       'building condition', 'constructed area sqft', 'functional classrooms',
       'dangerous classrooms', 'dangerous non functional classrooms',
       'total sections', 'has open classes', 'has drinking water',
       'has electricity', 'has toilet', 'has boundary', 'has main gate',
       'has sewerage', 'has playing ground', 'security arrangements',
       'has library', 'has science lab', 'has computer lab', 'has internet',
       'has ece room', 'has care giver', 'has ece kit', 'has multimedia',
       'has smart board', 'has Head', 'Total female students',
       'Total male students', 'Total students', 'Total Teacher', 'PST',
       'ESTs total', 'SSTs total', 'Subject Specialists total',
       'total Teacher', 'geometry', 'ED'],
      dtype='object')
      school id                                       School Names  ED
0         30109      Government girls higher secondary School BARKI   4
1         30167      Government girls high School HAIR LAHORE CANTT   6
2         30193                Government girls high School CHARRAR   8
3         30138   Government girls high School KARBATH (Newly Up...   9
4         30139                 Government girls high School LIDHAR   5
...         ...                                                 ...  ..
1115      30816          Government  high School GHAZI ABAD (EX-MCL)   9
1116      30817   Government  high School GUJAR COLONY JORAY PUL...   9
1117      30841          Government  high School MUSLIM BAGHBAN PURA   8
1118      30861   Government  high School TAJPURA SCHEME LAHORE ...   6
1119      31105   Government  elementry School TALEEM-O-TARBIAT ...   7

[1120 rows x 3 columns]
Total Education Disparity Sum: 10429
Average Education Disparity per School: 9.311607142857143
```

Fig. 1. Review of calculated education disparities of schools

Shapely geometry is used to generate point geometries for the geographical locations of the various schools. Pandas is used for data manipulation in combining the resultant geometries with geospatial operations. Contextily is used to add base maps from various Web tile providers to allow for the visualization of the locationality of schools in the context of Lahore. Matplotlib.pyplot is used in creating visual plots and maps.contextily was employed in extracting the school locations, which were overlaid onto a basemap of Lahore to form spatial context and geographical reference upon which the analysis was done. This kind of integration enhances the interpretability of the patterns of distribution seen among the schools. Besides visualization, the distribution of schools helps to assess the strategies of resource allocation. By spotting clusters or voids in school coverage over urban and rural zones, it would help in planning by the policy-makers for optimal resource allocation to achieve equitable facility distribution for education. With the mapped data in Fig. 2 were the preliminary insights on the geographical distribution of educational resources across Lahore. Future studies can use the techniques of spatial clustering and proximity to make substitute deep optimizations and improvements in the educational outcomes through resource allocations. As in Fig. 3 analysis in the cartography process, two locations of the schools having high and low density were identified, and later on the identified results were plotted on the map. The red spots were the locations where the high concentration of schools is located.

Table 1 shows the summary of high and low density schools and with the mapped data in Fig. 2 were the preliminary insights on the geographical distribution of educational resources across Lahore Future studies can use the techniques of spatial clustering and proximity to make substitute deep optimizations and improvements in the educational outcomes through resource allocations. As in Fig. 3 analysis in the cartography process, two locations of the schools having high and low density were identified and later on the identified results were plotted on the map. The red spots were the locations where the high concentration of schools is located. School areas of low density were denoted by the dots in the color red, which means the number of schools in that region is low. Such spatial representation is provided by means of this illustrative process in order to compare with distribution of schools in the study area and to understand the allocation of educational resources and potential gaps in coverage. To display values of Education Disparity on the district map, a choropleth map was made with the help of Folium. Circle Markers of School Locations and red-dotted regions of high disparities have been indicated. Upon this map, one can easily understand the overall distribution of ED and for which Schools to pinpoint for further targeted intervention to address effectively educational disparities.

Resource impact analysis. By conducting OLS regression described the significant contributors to ED across different schools shown in Fig. 4 and they were taken as aberrant resource.

The quality of drinking water or its availability also tested an influential predictor of ED, therefore, an

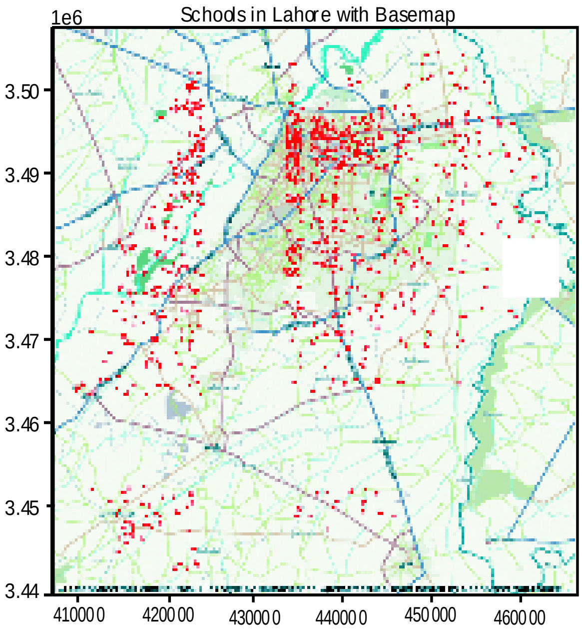

Fig. 2. Visualization of public schools in Lahore with basemap.

important variable. Safety precautions that were in one way or another below the part the school were reflected in higher ED. It means that safety was an influential predictor of disparities in education. Access to the right toilets and sanitation facility is also an influential predictor of ED-impacting variables that seemed to be visibly different between schools with limited or insufficiency of toilet facilities. Dangerous classrooms that were considered unsafe due to damaged infrastructure also affected the ED values significantly from the others, thus highlighting the influence of safety from the perspective of infra-structure on the learning environment. As in Fig. 5.

By identifying urban areas as focal points for high ED values and pinpointing influential resources using the OLS regression as in Fig. 6.

This analysis provides a comprehensive framework for addressing educational disparities. Implementing targeted interventions desired to improve identified resources will work towards equalizing educational opportunities and improving learning outcomes for students in urban and rural settings alike.

Hotspot analysis. Hotspot analysis is the detection of clusters that exhibit low and high ED values. The hot spots exhibited high values and the cold spots exhibited low values of ED as in Fig. 7.

There were some common mathematical formulas have been used to calculate the Getis-Ord G_i^*

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} - \overline{X} \sum_{j=1}^{n} w_{i,j}}{s\sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}}$$

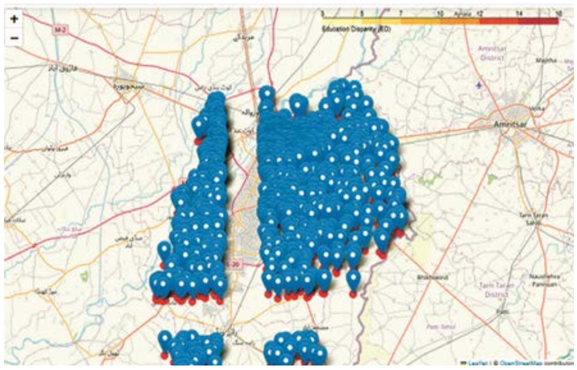

Fig. 3. Schools Locations and High ED Value Region.

For each school here, the numerator represents the weighted sum of the difference of the ED value of school i and the expected sum if no spatial cluster exists. This analysis opens the way for understanding how best interventions can be focused on reducing

Table 1. Summary of high and low density schools

High-density region sources summary					
	Schools identity	Range	Longitude	Have constructing	
Count	1037	1037	1037	1037	-
Mean	31039.26	31.5	74.32	0.999	-
Std Dev	2767.07	0.1	0.12	0.03	-
Min	30100	31.13	74.06	0	-
25%	30419	31.44	74.19	1	-
50%	30769	31.52	74.33	1	-
75%	31047	31.58	74.4	1	-
Max	55030	31.66	74.6	1	-
	Area in kanal	Total area in marla	Constructed area in Sqft	Total teachers	-
Count	1037	1037	1037	1037	-
Mean	5.34	6.71	10474.81	13.71	-
Std Dev	11.31	5.97	37896.54	14.16	-
Min	0	0	0	0	-
25%	1	0	1868	5	-
	1	6	5400	10	-
75%	6	11	10890	16	-
Max	176	19	1105684	103	-
Metric	PST	ESTs total	SSTs total	Subject specialists total	Heads total
Count	1037	1037	1037	1037	1037
Mean	6.95	4.1	2.71	0.31	0.25
Std Dev	5.09	5.67	4.5	1.83	0.43
Min	0	0	0	0	0
25%	3	0	0	0	0
50%	6	1	1	0	0
75%	10	7	4	0	0
Max	51	41	35	17	1
Low-density area resources summary					
Metric	School ID	Latitude	Longitude	Have building	-
Count	83	83	83	83	-
Mean	33169.9	31.31	74.38	0.99	-
Std Dev	7349.47	0.15	0.17	0.11	-
Min	30101	31.11	74.05	0	-
25%	30411.5	31.18	74.2	1	-
50%	30667	31.31	74.4	1	-
75%	31046.5	31.41	74.51	1	-
Max	54196	31.67	74.62	1	-
Metric	Total area in kanal	Total area in marla	Constructed area in Sqft	Functional classrooms	-
Count	83	83	83	83	-
Mean	5.39	7.19	5711.31	6.13	-
Std Dev	7.86	6.55	6723.68	4.31	-
Min	0	0	0	0	-
25%	1	0	1450	4	-
50%	1	8	4500	5	-
75%	7	11	8150	7.5	-
Max	41	19	45000	18	-

	coef	std err	t	P>|t|	[0.025	0.975]
functional classrooms	-0.0060	0.003	-1.751	0.080	-0.013	0.001
total sections	0.0122	0.004	3.359	0.001	0.005	0.019
dangerous classrooms	0.1019	0.010	10.286	0.000	0.082	0.121
has open classes	-0.2187	0.017	-12.958	0.000	-0.252	-0.186
has drinking water	4.0430	0.180	22.504	0.000	3.690	4.396
has electricity	-1.2037	0.475	-2.537	0.011	-2.135	-0.273
has toilet	4.0430	0.180	22.504	0.000	3.690	4.396
has boundary	-1.0603	0.240	-4.420	0.000	-1.531	-0.590
has main gate	4.0430	0.180	22.504	0.000	3.690	4.396
has sewerage	-0.9584	0.089	-10.795	0.000	-1.133	-0.784
has playing ground	-0.9975	0.033	-30.068	0.000	-1.063	-0.932
security arrangements	4.0430	0.180	22.504	0.000	3.690	4.396
has library	-1.0437	0.044	-23.858	0.000	-1.130	-0.958
has science lab	-0.9396	0.064	-14.684	0.000	-1.065	-0.814
has computer lab	-0.9984	0.065	-15.251	0.000	-1.127	-0.870
has internet	-1.0155	0.038	-26.878	0.000	-1.090	-0.941
has multimedia	-1.0528	0.059	-17.699	0.000	-1.169	-0.936
has smart board	-0.9441	0.070	-13.407	0.000	-1.082	-0.806
has Head	-1.0689	0.048	-22.180	0.000	-1.163	-0.974
have building	-1.1326	0.475	-2.382	0.017	-2.065	-0.200
has ece room	0.1928	0.050	3.870	0.000	0.095	0.291
has care giver	-0.9629	0.037	-25.942	0.000	-1.036	-0.890
has ece kit	-0.2729	0.049	-5.606	0.000	-0.368	-0.177
Total students	1.376e-05	6.37e-05	0.216	0.829	-0.000	0.000
Total Teacher	-0.0025	0.003	-0.983	0.326	-0.008	0.003

Omnibus:	55.174	Durbin-Watson:	2.011
Prob(Omnibus):	0.000	Jarque-Bera (JB):	62.839
Skew:	-0.591	Prob(JB):	2.26e-14
Kurtosis:	3.296	Cond. No.	4.55e+19

Fig. 4. OLS regression analysis std err results.

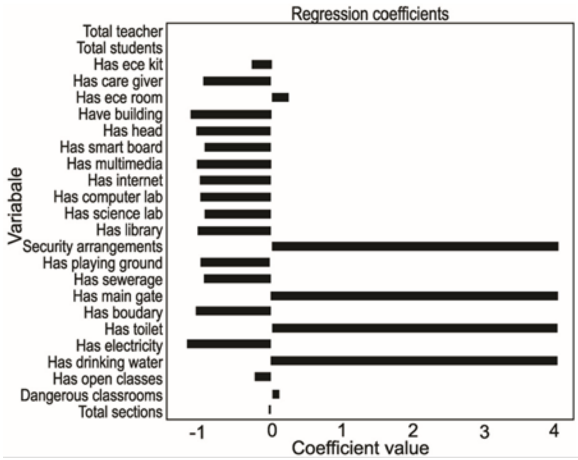

Fig. 5. Regression coefficients (identified confounding variables).

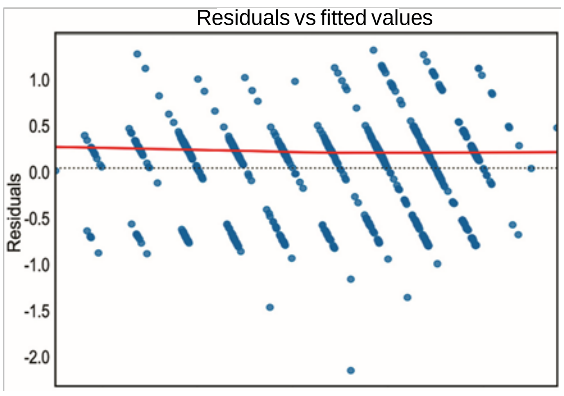

Fig. 6. OLS regression line.

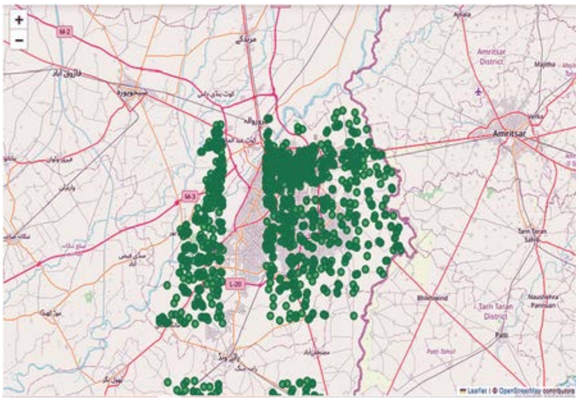

Fig. 7. Visualization of schools with high and low ED value depending upon confounding variables.

disparities and for optimal designing and allotment of resources in Fig. 7 to ensure the even distribution of facilities in the education sector. Therefore, the study includes the assessment of correlations between variables to identify important factors affecting the ED value. Further, the Getis-Ord Gi∗G_i^*Gi∗ statistic was employed to contrast ED values in the male and female schools. Pearson's correlations were populated to identify in which the factors affect the ED values through the most significant degree and thus which would account for the most variation about differences in the education system among schools. It identified the most significant factors in ED by the exploration, as such ranking about which the most attention needs to be given to effectively reduce ED.

Gender differences in education. A contrast to the Getis-Ord Gi∗G_i^*Gi∗ statistic revealed as in Fig. 8 that a male school had an average ED value of 9.4, whereas a female school had a slightly reduced average ED value of 9.232. This further tends to suggest that male schools, on average, also suffer from slightly more educational disparities than those experienced in female schools as in Fig. 9. Educational disparities, especially with male schools, warrant critical evidence-based interventions through policy changes. By considering those influential factors identified by the correlation analysis, policymakers and educators might take real and material steps in working that there must be more justice and equality to increase the effectiveness of the outcomes of education in all schools. By considering those influential factors identified by the correlation analysis, policymakers and educators might take real and material steps in working that there must be more justice and equality to increase the effectiveness of the outcomes of education in all schools.

Urban influences on disparities in education and resource effects analysis. The results showed not only the presence of extremely high ED values in the urban area of between the two groups in Figs. 10 and 11 but also indicated what resources were negatively, ED affecting and how intense. The strikingly high proportion of variance explained by OLS regression for factors that increase ED is superior to what has been reported in OLS literature thus far for any other educational contrast.

Resource influence relative to education inequalities. Urban areas showed high ED values for males and females alike. The inequality that nonetheless emerges from this spatial trend is significant in that the variation in resources and outcomes extends beyond what would be expected and is therefore the focus of redress in urban settings, where such inequality can best be mediated.

Spatial mathematics. ***Density analysis.*** On the subject of high-intensity areas, the dataset of Lahore schools produced several results describing it.

There were 280 schools identified in these high-density areas. The schools showed striking variations and patterns across the following key attributes. Following were the mathematical calculations

Male schools Gi* values:

	School Names	ED	Gi_star
60	Government high School HADYARA	5.0	2.462667e-15
61	Government high School PADHANA LAHORE	7.0	1.964080e-15
62	Government high School LIDHAR	7.0	1.964080e-15
63	Government high School HAIR LAHORE CANTT	6.0	2.458393e-15
64	Government high School KARBATH	5.0	2.462667e-15

Female schools Gi* values:

	School Names	ED	Gi_star
0	Government girls higher secondary School BARKI	4.0	-6.451228e-15
1	Government girls high School HAIR LAHORE CANTT	6.0	-6.432508e-15
2	Government girls high School CHARRAR	8.0	-6.915541e-15
3	Government girls high School KARBATH (Newly Up...	9.0	-6.418991e-15
4	Government girls high School LIDHAR	5.0	-6.440885e-15

Fig. 8. Preview of Gi* value of some male and female schools

constructed.Area(A):
here
n=count=280 schools
Mean=m
stand.deviation=s

$$\mu_A=1/n\sum_{i=1}^{n}nAi \approx 11{,}111 \text{ sqfeet}$$

$$\sigma_A=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i-\mu_A)^2}\approx 15{,}203 \text{ sqfeet}$$

Functional.classroom(C_f):

$$\mu_{C_f}=\frac{1}{n}\sum_{i=1}^{n}C_{f_i}{}^2=13.36 \text{ classrooms}$$

$$\sigma_{C_f}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}((C_f)_i-\mu_{C_f})^2}=9.59 \text{ classrooms}$$

Dangerous.classroom(C_d):

$$\mu_{C_d}=\frac{1}{n}\sum_{i=1}^{n}(C_d)_i \approx 0.09 \text{ classrooms}$$

$$\sigma_{C_d}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}((C_{d_i})-\mu_{C_d})^2}\approx 0.44 \text{ classrooms}$$

Total.Section(S):

$$\mu_S=\frac{1}{n}\sum_{i=1}^{n}(S)_i \approx 10.07 \text{ sections}$$

$$\sigma_S=\sqrt{\frac{1}{n}\sum_{i=1}^{n}((S_i)-\mu_S)^2}\approx 6.62 \text{ sections}$$

Total.Students(N):

$$\mu_N=\frac{1}{n}\sum_{i=1}^{n}(N)_i \approx 378.05 \text{ students}$$

$$\sigma_N=\sqrt{\frac{1}{n}\sum_{i=1}^{n}((N_i)-\mu_N)^2}\approx 434.14 \text{ students}$$

Total.Teachers(T):

$$\mu_T=\frac{1}{n}\sum_{i=1}^{n}(T)_i \approx 9.01 \text{ Teachers}$$

$$\sigma_T=\sqrt{\frac{1}{n}\sum_{i=1}^{n}((T_i)-\mu_T)^2}\approx 10.97 \text{ Teachers}$$

In the high-density schools, the constructed area had a mean of about 11,111 square feet, while the standard

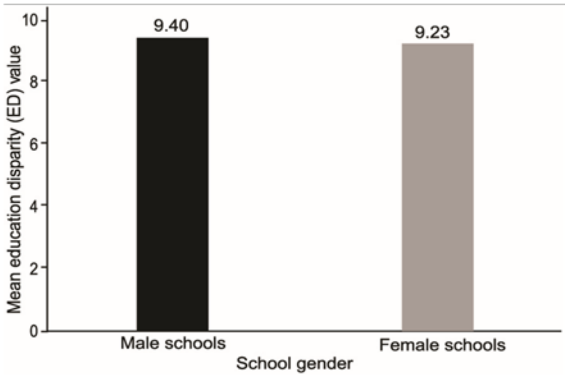

Fig. 9. Comparison of mean ED value of male and female schools.

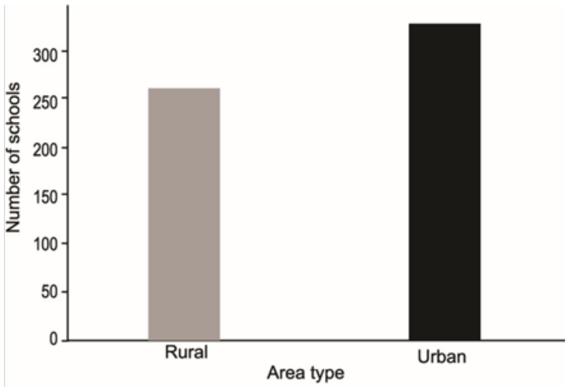

Fig. 10. Comparison of female schools with high ED values in rural *vs* urban areas.

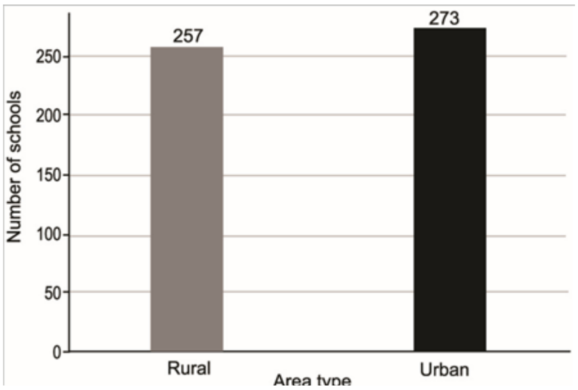

Fig. 11. Comparison of male schools with high ED values in rural *vs* urban areas

deviation was 15,203 square feet. This extreme deviation denotes tremendous dispersion in the number of schools hosted groups. On average, the schools had 13.36 functional classrooms, with a standard deviation of 9.59, pointing toward wide variations in the availability of classes. In terms of safety, the figures indicated that every school had about 0.49 risky classroom sections on average, while the standard deviation of sections was 2.69. This measure showed the need for better infrastructural facilities in some schools. Schools in a high-density area also told that each school had an average of 14.45 total sections, while the standard deviation of this number stood at 10.45; hence, there was a very large difference in the number of sections between one school and another. The average total of student population with regard to these high-density schools settled at 655.64 students, with a standard deviation of 573.44, a clear indication that there is huge variation in the size of strength. Similarly, the distribution of the teacher force was wide, with an average of 17.58 and a standard deviation of 15.40 teachers per school. More statistical analysis went into giving insights on how these attributes were distributed. For instance, the 25th percentile of this constructed area was 3,584 square feet and the median was 518.5 students. The 75th percentile of total sections was 16 sections and the maximum total student that their population in a school could reach is 4,090 students.

$$\mu_{ED} = \frac{1}{n}\sum_{i=1}^{n} ED_i \approx 9.85$$

It shows how the educational facilities were up to school, the mean score of average density (ED) of these high-density schools being 8.29. Similarly, the mean density from other attributes thus obtained was approximately 16.69, referring to better educational infrastructure in high-density areas. These findings provide an alternative perspective on the school infrastructure and its distribution in high-density zones of Lahore to help identify improvement areas and understand resource allocation.

Proximity analysis*. *Nearest neighbor analysis (NNA). In this light, to determine how close schools were to one another, all while identifying potential areas of over attention, or areas where there were gaps in the coverage of education, a nearest neighbor analysis was conducted. This analysis evaluated the spatial distribution of schools, particularly focusing on those with the highest nearest neighbor distances destined to explain their distribution. The location of schools is shown in Fig. 3.

In research investigation, 279 schools containing large nearest neighbor distances were found. Various key attributes were tested to provide an insight into the nature of the schools. The following descriptions paint a very clear picture of these schools. Average constructed area is 7,282.57 square feet (10,746.77 square feet as a standard deviation), 7.91 functional classrooms (standard deviation of 6.26) and 0.09 dangerous classrooms (standard deviation of 0.44). The total number of mean sections was 10.07 per school with the standard deviation 6.62. The population of students was on average 378.05 with a standard deviation of 434.14 and an average of 9.01 teachers with a standard deviation of 10.97.

$$\mu_{NND} = \frac{1}{n}\sum_{i=1}^{n} NND_i \approx 0.0095$$

$$\sigma_{NND} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (NND_i - \mu_{NND})^2} \approx 0.0046$$

The percentile analysis stated that 25% of the schools had an built up area less than 1,700 sq.ft., the median students per school were 185 and 75% of school has up to 11 sections and the maximum population of student observed is around 3,880 with average ED density score rates at 9.85, which implies that a school records slightly more than 9 deficiency points, while the average score for nearest neighbor distance is approximately 0.0095 with a standard error of 0.0046; this provides an explanation for the spatial distribution and has point inclusion of gaps in the coverage limit of the school.

AI modeling. *Predictive analysis with linear regression model.* This study proceeded with predictive analysis of the school-based data to forecast the ED values and subsequently compute the number of additional teachers that would help to reduce this value. The procedure and results were presented. A linear regression model was implemented for the estimation of the ED value based on the numbers of students, teachers in each school as according to school education department policy the ratio of student teacher is 40:1 respectively. That’s why student’s strength is directly related to the number of teachers.

The model implemented was
$ED=\beta_0+\beta_1\times student+\beta_2\times Teacher\times\epsilon$

where:
β_0, β_1, β_2 are coefficients of model
ϵ is error term

using the model, it was possible to predict the ED values for each school in the dataset.

Evaluation of additional needs of teachers. The next step was estimating the number of excess teachers needed specifically to bring down the ED values to the level needed. This was done by assuming that the ED value would reduce by the predicted value from the model and then adjusting the model to find the number of teachers change needed to fulfill the required increase:

$$\text{Additional teachers required}=\frac{\text{Targeted.ED-Predicted.Ed}}{\beta_2}$$

The results showed that 728 teachers among the schools were needed such that the reduction in the value of the ED is achieved.

To have an insight into the spatial query in which extra teachers were required in different schools, locations have been mapped by availing the tools of a geographic information system (GIS). Point shapes of schools, overlaying on a base map of Lahore, cast the tone of feeling needed most. Schools that have needs for more teachers were stored as shapefile format to also be subject to a spatial. The analysis found that a total of 728 teachers required to lower the educational disparity, which can thus bring about positive changes within education. The implication is that resource allocation into given school sites and rational planning were central factors in educational policy and reinforcement in the quest for quality in student learning environments.

Conclusion

This study has demonstrated a micro-level approach to assessing educational disparities and optimal resource allocation in schools in Lahore using both geospatial and statistical techniques. Infrastructure provision, including, at minimum, drinking water, security, toilets and classrooms in good order, is instrumental in ensuring equity in education. Similar spatial bias in educational density is also observed as the male schools and urban schools were more, so they again need to receive the intervention in terms of the infrastructural gap. School coverage is adequate and thus, there were many spatial gaps through which NNA gives no results in showing spatial pattern of result in school coverage; other parts of Lahore were also provided with enough resources. AI data analyses, amongst other AI techniques, have improved enormously in ascertaining major contributors of ED values like access to drinking water, security measures, sanitation facilities, and safety in classrooms for the census 2022-23 dataset. Such an example underlines the potential with which AI techniques can contribute to the enhancement of decision-making processes by capturing areas most critical and sensitive to interventions or improvement. Spatial mathematics has provided a good framework for the analysis of the spatial distribution of the educational facilities and the spatial clustering of high ED values. With techniques like the Getis-Ord Gi* statistic available, identification of urban areas and the gender-specific gaps in the educational outcomes has been possible for targeting the interventions. Insights into education policies were critical for policymakers to optimize resource allocation and get better educational outcomes in Lahore toward sustainable development through informed and strategic planning. Practical glitches may face government policies intending to translate research findings into action, such as bureaucratic hitches, scarcity of resources and most importantly, resistance to change from stakeholders. Although this work indicates areas for optimizing resource allocation, the effectiveness of the implemented interventions to achieve educational results warrants continued monitoring and evaluation. Data collection along with its analysis could also introduce biases inadvertently or raise issues of privacy, especially when it relates to sensitive information regarding students, teachers or educational institutions.

Acknowledgment

Targeted resource allocation based on the findings of the spatial analysis, through which educational outcomes were to be improved, has to be given the highest priority. Infrastructure needs to be improved on, more teachers to deploy strategically and there has to be educational resource accessibility equity for all.

Conflict of Interest. The authors declare that they have no conflict of interest.

References

Adewoye, S.E., Olaseni, A.O. 2022. A systematic review of factors that influences the efficacy of computer based learning. *Journal of Education, Teaching and Learning*, **7:** 166-175.

Ahmad, N., Sumera, R., Ali, Z. 2023. Investigating primary school teachers' perceptions about professional development and its impact on students achievement. *Journal of Social Sciences Review*, **3:** 809-823.

Alam, M.M., Alauddin, M., Yasin, M., Dooty, E.N., Syed, M., Chowdhury, M.M. 2021. Students' satisfaction and university reputation through service quality in private higher educational institutions in Bangladesh. *The Journal of Asian Finance, Economics and Business*, **8:** 91-100.

Alothman, H.F., Bashatah, L., Aldossari, A.S., Alfaifi, M.S., Almutairi, A., Alshuaibi, A.A., Alajlan, H.A., Aseri, S., Aseery, A.Y., Naeema, A.A. 2024. How Saudi parents rationalize the choice of school for their children. *Humanities and Social Sciences Communications*, **11:** 1-11.

Antwi, V.N. 2024. *Global trends in student anxiety: a comparative review with a focus on Ghana.* Link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5021682

Chen, A., Chen, N., Liu, W., Bai, R. 2019. Application of computational intelligence technologies in emergency management. *Artificial Intelligence Review*, **52:** 2131-2168.

Dabhade, P., Agarwal, R., Alammen, K., P., Fathima, A., T., Sridharem, R., Gopakumar, G. 2021. Educational data mining for predicting students' academic performance using machine learning algorithms. *Materials Today: Proceedings*, **47:** 5260-5267.

Engin-Damir, C. 2009. Factors influencing the academic achievement of the Turkish urban poor. *International Journal of Educational Development*, **29:** 17-19.

Fahd, K., Venkatraman, S., Miah, S.J., Ahmed, K.E.U. 2022. Application of machine learning in higher education to assess student academic performance, at-risk and attrition: a meta-analysis of literature. *Education and Information Technologies*, **27:** 3743-3775. DOI: 10.1007/ s10639-021-10741-7

Haider, H. 2024. *Internationalisation of Education: Focus on Central Asia and Eastern Neighbourhood, K4DD Rapid Evidence Review 8*, 29 pp., Institute of Development Studies, Brighton, UK. DOI 10.19088/K4DD.2024.012

Kapoor, N.R., Kumar, A., Alam, T., Kumar, A., Kulkarni, K., Blecich, P. 2021. A review on indoor environment quality of Indian school classrooms. *Sustainability*, **13:** 11855.

Khuma, M., Utete, R. 2023. Factors that influence academic performance of students: an empirical study. *The Seybold Report*, **18:** 1710-1722.

Kumari, R., Kumar, A., Tunio, I.A., Sassanelli, S. 2023. Water quality assessment and monitoring in Pakistan: a comprehensive review. *Sustainability*, **15:** 6246.

Meena, D., Tripathi, R., Agarwal, S. 2023. An evaluation of primary schools and its accessibility using GIS techniques: a case study of Prayagraj district, India. *GeoJournal*, **88:** 1921-1951.

Nawaz, S.M.N., Alvi, S., Rehman, A., Tayyabah, R. 2022. How do beliefs and attitudes of people influence energy conservation behavior in Pakistan? *Heliyon*, **8:** e11054. DOI: http://doi.org/10.1016/j.heliyon.2022.e11054

Peprah, O.M., Nketsia, W., Belbase, S., Side, A.S., Jiya, A.N., Gemeda, F.T. 2024. A cross-national study of teacher retention and job satisfaction in sub-Saharan Africa. *Preventing School Failure: Alternative Education for Children and Youth*, **69:** 137-148. DOI: doi.org/10.1080/1045988x.2024.2361874

Sajjad, M., Munir, H., Kanwal., S., Naqvi, S.A.A. 2022. Spatial inequalities in education status and its determinants in Pakistan: a district-level modelling in the context of sustainable development goal-4. *Applied Geography*, **140:** 102665. DOI: 10.1016/japgeog.2022.102665

Sele, J.P., Wanjiku, C. 2024. The role of education in economic development: a comparative study of Nigeria and Kenya. *Greener Journal of Social Sciences*, **14:** 136-146.

Sharma, G., Patil, G.R. 2022. Spatial and social inequities for educational services accessibility-a case study for schools in greater Mumbai. *Cities*, **122:** 103543.

Singh, M., Adebayo, S.O., Saini, M., Singh, J. 2021. Indian government e-learning initiatives in response to COVID-19 crisis: a case study on online learning in Indian higher education system.

Education and Information Technologies, **26:** 7569-7607.

Stone, D.C. 2021. Student success and the high school-university transition: 100 years of chemistry education research. *Chemistry Education Research and Practice*, **22:** 579-601.

Yaakub, N.M. 2022. GIS-based spatial correlation analysis: sustainable development and two generations of demographic changes. *Sustainability*, **14:** 1490.

Zaman, B., Sharma, A., Ram, C., Kushwah, R., Muradia, R., Warjri, A., Dany, K., Lyngdoh, D.K., Lyngdoh, M.K. 2023. Modeling education impact: a machine learning-based approach for improving the quality of school education. *Journal of Computers in Education*, **11:** 1181-1214.

Zartashia, K.J., Ramazan, M., Ijaz, S. 2024. A systematic review on cognitive and motivational impact on english language learning through artificial intelligence. *International Journal of Literature, Linguistics and Translation Studies*, **4:** 44-71.