Acta Scientific COMPUTER SCIENCES

Volume 6 Issue 7 July 2024

Research Article

Impact of Biological Factors on the Academic Performance of Students: A Comprehensive Analytical Study

Buoye PA* and Adegboye OJ

Department of Computer Science, Federal Polytechnic, Ilaro, Nigeria

***Corresponding Author:** Buoye PA, Department of Computer Science, Federal Polytechnic, Ilaro, Nigeria.

Received: April 11, 2024

Published: June 14, 2024

Abstract

The research looks into how biological characteristics, such as gender, blood type, and group, affect students' academic achievement as determined by their cumulative grade point average (CGPA). The study looks at these variables' ability to predict CGPA using multiple regression analysis and decision tree regression. The model's weak explanatory ability is indicated by the findings, which show a non-significant F-statistic and an R-squared value of 0.040. Nonetheless, a few factors, including blood type, gender, and specific birth months, are significant in CGPA prediction. A Mean Squared Error (MSE) of 0.2825 for the decision tree regression model indicates a satisfactory fit of the model to the data. To fully comprehend academic accomplishment, the study's conclusion emphasizes the significance of taking biological elements into account in addition to other socioeconomic and psychological aspects. It emphasizes the necessity of using multidisciplinary research methods to disentangle the intricate interactions between variables influencing students' academic paths.

Keywords: Biological Characteristics; Blood-type; Blood Group; Gender; Multiple Regression

Abbreviations

CGPA: Cumulative Grade Point Average; MSE: Mean Squared Error

Introduction

Numerous factors, including biological elements that go beyond conventional educational paradigms, affect pupils' academic achievement. Blood type, blood genotype, and gender are among these biological characteristics that have become interesting research areas because they provide information about the complex interactions between biology and educational results.

According to educational psychology, a variety of factors, from instructional strategies to socioeconomic background, affect students' academic success. The purpose of this study is to investigate this little-studied area in an effort to identify any possible relationships and clarify the complex interactions that exist between these biological variables and academic achievement.

Blood type, a genetic characteristic based on particular antigens found on the surface of red blood cells, has long been linked to a number of health consequences. According to [6], blood genotype is an individual's genetic composition of blood. It is the protein that is found in red blood cells' hemoglobin. From person to person, this protein differs. According to some authors, this protein is the biological or genetic code. [6] states that there are five different varieties of this biological code (genotype) that indicate an individual's uniqueness: AA, AS, AC, SC, and SS.

Recent studies have hinted at potential connections between blood group and cognitive function, raising intriguing questions about its possible influence on academic performance. Furthermore, the genetic variations inherent in blood genotypes may also play a role in shaping individuals' cognitive abilities and learning styles, thus warranting investigation in the academic context.

Numerous studies have examined the differences between genders in academic achievement and learning preferences, with gender serving as a recurring theme in educational research. The relationship between gender and biological characteristics, such as blood type and genotypes, has, however, received less study. This interaction may provide complex insights into academic discrepancies and instructional tactics. Furthermore, new research indicates that birth month—which is frequently disregarded in studies on education—may have a small but important impact on educational attainment and cognitive development.

By performing a thorough analysis of the effects of blood group, genotypes, gender, and birth month on students' academic performance, this research aims to close these discrepancies. Using a mixed-methods approach that includes both quantitative and qualitative investigations, this study seeks to clarify any possible relationship between these biological characteristics and students' academic achievement.

Literature Review

While the direct influence of blood group on academic performance remains relatively understudied [12], found that individuals with blood type AB demonstrated better memory performance compared to those with other blood types. The paper however suggested a further research to elucidate the mechanisms underlying these associations and their implications for educational outcomes.

In [3], it was concluded that blood group and blood genotype have no influence on the general performance of the students after carrying out a statistical analysis on the students' performance and the variations in blood group and blood genotype.

[2] Carried out comparative analysis of male and female students in programming ability, concluded that there is no statistically significant difference in the performance of male and female student. This means that gender might not be a preference to students performance academically.

[4] Compared genotype polymorphism with academic performance of student. In his paper, the blood group of the individual compliments the genotype as well as the achievement of students. He concluded that there is no statistically significant mean difference in the academic achievement of students and their genotypes.

[5] Researched on Blood group and gender impact on personality among medical students. They concluded that there is no considerable difference among personality scores of individuals of different blood groups.

[9] Also researched on relationship between Personality Traits and Blood Groups among the students of Palestinian University. He however concluded that there is no relationship that exists between them.

[7] Explained the difference between blood group and blood genotype. According to them, while blood group is the presence or absent of antigens on the surface of red blood cell, a genotype is the two letters combination that determines which traits or gene are inherited.

[10] in his article written on byjus.com said blood group is a blood classification based on presence or absence of antibodies and inherited antigenic particles on the surface of red blood cells and he explained genotype as genotype form genetic constituency of cells in human. The results from the study of [11] showed that there are gender differences in academic ability and that females generally outperform males in verbal abilities and GPA. They however stated that no gender differences were found in terms of quantitative skills. According to [1] there was no significant association between blood groups and academic performance; they however suggested further studies to be carried out in various other colleges and schools to investigate the association between blood groups and academic performance

Methodology

This study employs a quantitative analysis of academic data with qualitative exploration of students' biological profiles and educational experiences.

Data collection

The Academic Performance of the students in form of Cumulative Grade Point Average (CGPA) used in this research was gotten from the Federal Polytechnic of Ilaro, Computer Science Department. Through the records, biological factors such as blood type, genetic information, and gender identification characteristics were collected which serves as the dataset for the research.

Exploratory data analysis

Using statistical graphics and other data visualization techniques, the key features of the dataset were summarized. The distribution of academic achievement is characterized by statistical descriptive analysis, which calculates the means, standard deviations, and frequencies. The summary of the descriptive analysis is displayed in Table 2. The distribution of students' CGPAs within the dataset is displayed as a histogram in figure 1.

Bar plot was carried out on the blood group, blood genotype, gender and month of birth since these factors are categorical variables. Table 3 shows the graph.

Inferential analysis

Multiple regression and decision tree regression modeling were employed to examine the association between blood group, genotype, gender, month of birth and academic performance.

Data preprocessing

Table 1 shows the first 5 and last 5 rows of the dataset. There are 6 columns and 533 rows in the dataset. The dataset consists of one continuous variable (the CGPA) and four categorical independent variables.

Descriptive statistical analysis

Descriptive statistics analysis involves method and techniques used to summarize and describe the main features of a dataset.

S/N	Month	BLOOD_GNE	BLOOD_GRP	Gender	CGPA	Grade
1	December	AA	A+	M	2.04	PASS
2	January	AA	A+	M	3.18	UPPER
3	January	AA	A+	M	2.21	PASS
4	January	AA	A+	F	2.19	PASS
5	March	AA	O+	M	2.23	PASS
-----	------	-----	----	----	----	-----
-----	------	------	----	----	-----	------
-----	-------	-------	----	----	-----	-------
530	January	AA	A+	M	2.99	LOWER
531	March	AA	A+	M	2.93	LOWER
532	January	AA	B-	F	2.64	LOWER
533	January	AS	O+	F	2.66	LOWER
534	January	AA	A+	M	2.82	LOWER

Table 1: Dataset showing the biological factors (Gender and CGPA).

[535 rows x 6 columns]

The maximum, minimum, standard deviation values for continuous variable. In the case of the dataset used, descriptive statistical analysis is only possible for the CGPA of student because it is the only continuous variable in the dataset the result of the descriptive statistical analysis is shown on the table 2.

Table 2: Summary Statistics: CGPA.

Mean	2.709884
std	0.497881
Min	0.740000
25%	2.380000
50%	2.690000
75%	3.045000
max	3.940000

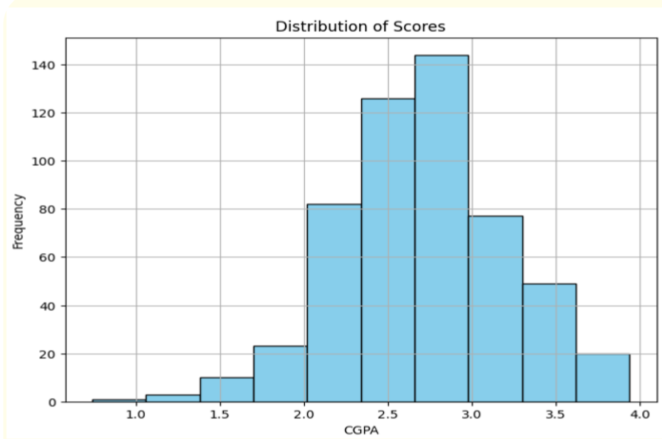
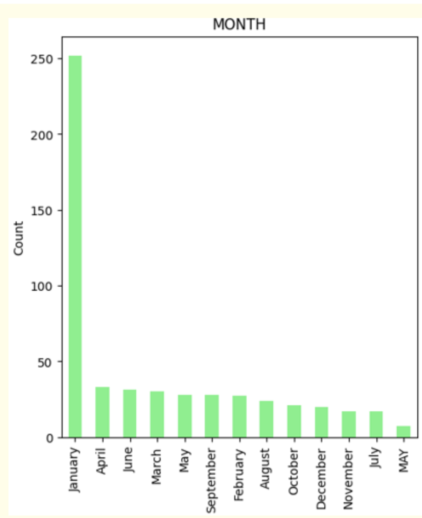


Figure 1: CGPA distribution graph.



From the table 2, the maximum value of CGPA is 3.940 and the minimum value is 0.740, the mean value is 2.709. From figure 1, about 140 students have CGPA that is between 2.5 to 3.0 while above 120 students having CGPA that is between 2.0 and 2.5. Less than 20 students have CGPA that is less than 2.0. No distinction student from this result as the none of the students has a CGPA that is 4.0 or above.

Regression analysis

A statistical method for examining the relationship between a dependent variable and one or more independent variables is a regression model. Determining how the values of the independent variables affect the dependent variable's value is the aim of regression analysis.

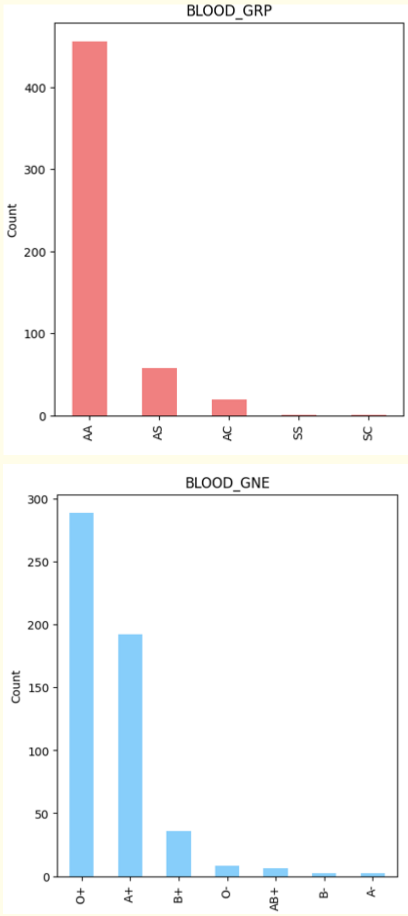


Figure 2: Bar chart of the categorical variables.

A statistical method called multiple linear regression is used to model the connection between one continuous dependent variable (the target) and two or more independent variables (features). By adding more predictors, it expands the idea of simple linear regression, which only considers one independent variable.

The relationship between the independent variables (represented as X1, X2,..., Xp) and the dependent variable (Y) is represented as a linear function in multiple linear regression:

$Y = \beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta pXp + \varepsilon ..\ i$

In this case, the dependent variable (target) is Y.

The independent variables (features) are X1, X2,..., Xp.

When all independent variables are zero, the value of Y is represented by the intercept term, β0. The coefficients, which are often referred to as weights or slopes, are β1, β2,..., βp for each independent variable. They represent the change in Y for a one-unit change in the corresponding independent variable, while keeping all other variables constant.

The error term, ε, denotes the discrepancy between the actual and anticipated values of Y.

Finding the coefficients β0,β1,...,βp that minimize the sum of squared differences between the dependent variable's predicted and actual values is the aim of multiple linear regression.

In Multiple Linear Regression, the coefficients are usually estimated using the least squares method, where the coefficients are selected to minimize the mean squared error (MSE) or the residual sum of squares (RSS). Finding the coefficient values that produce the least overall difference between the observed and predicted values of the dependent variable over all data points is the process at hand.

The use of numerous independent variables in a single linear regression expands its scope. When you have multiple category independent variables, it makes sense.

The multiple regression analysis was carried out on the provided dataset using stats models in Python. The summary output will include coefficients, standard errors, t-values, p-values, and other relevant statistics for each variable in the model.

```
# Defining independent and dependent variables
X = dfresult.drop(columns=['CGPA'])
y = dfresult['CGPA']
# Fitting the regression model
model = sm.OLS(y, X)
results = model.fit()
# Printing the regression summary
print(results.summary())
```

This OLS (Ordinary Least Squares) regression analysis provides insights into the relationship between the dependent variable (CGPA) and the independent variables included in the model.

Features importance analysis

In machine learning and statistical modeling, feature importance analysis is a technique used to evaluate the importance or contribution of various features (also called variables, predictors, or independent variables) to the prediction of the target variable (also called the dependent variable or outcome). This study sheds light on the underlying links between predictors and the target variable and aids in determining which traits are most important in making predictions.

The relevance score assigned to each feature indicates how much it contributes to the model's predictions. A larger positive

Dep. Variable:	**CGPA**	**R-squared:**	**0.040**
Model:	OLS	Adj. R-squared:	-0.005
Method:	Least Squares	F-statistic:	0.8988
Date:	Wed, 20 Mar 2024	Prob (F-statistic):	0.601
Time:	13:34:54	Log-Likelihood:	-371.02
No. Observations:	518	Df Model:	23
Df Residuals:	494		

SN		**coef**	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
1	MONTH_April	0.755	0.089	0.848	0.397	-0.099	0.250
2	MONTH_AUG	0.0840	0.101	0.831	0.406	-0.114	0.282
3	MONTH_DEC	0.1592	0.111	1.438	0.151	-0.058	0.377
4	MONTH_FEB	0.0920	0.097	0.953	0.341	-0.098	0.282
5	MONTH_JAN	0.1506	0.043	3.498	0.001	0.066	0.235
6	MONTH_JUL	0.2643	0.120	2.211	0.027	0.029	0.499
7	MONTH_JUN	0.0736	0.092	0.797	0.426	-0.108	0.255
8	MONTH_MAY	0.3572	0.181	1.979	0.048	0.003	0.712
9	MONTH_MAC	0.2292	0.095	2.409	0.016	0.042	0.416
10	MONTH_NOV	0.0554	0.125	0.442	0.659	-0.191	0.302
11	MONTH_OCT	0.0242	0.108	-0.224	0.823	-0.236	0.188
12	MONTH_SEP	0.0534	0.095	0.564	0.573	-0.133	0.240
13	BLOOD_GRP_AA	0.0345	0.134	0.257	0.797	-0.229	0.298
14	BLOOD_GRP_AC	0.2700	0.166	1.623	0.105	-0.057	0.597
15	BLOOD_GRP_AS	0.0336	0.143	0.235	0.815	-0.248	0.315
16	BLOOD_GRP_SC	0.5063	0.435	1.163	0.245	-0349	1.362
17	BLOOD_GRP_SS	0.7670	0.428	1.790	0.074	-0.075	1.609
18	BLOOD_GNE_A+	0.1222	0.086	1.418	0.157	-0.047	0.292
19	BLOOD_GNE_A-	0.6598	0.323	2.044	0.042	0.025	1.294
20	BLOOD_GNE_AB+	0.2436	0.198	1.229	0.220	-0.146	0.633
21	BLOOD_GNE_B+	0.2138	0.109	1.962	0.050	-0.000	0.428
22	BLOOD_GNE_B-	0.0579	0.319	0.182	0.856	-0.563	0.684
23	BLOOD_GNE_O+	0.1212	0.083	1.454	0.147	-0.043	0.285
24	BLOOD_GNE_O-	0.1928	0.175	1.101	0.272	-0.151	0.537
25	GENDER_F	0.8315	0.053	15.616	0.000	0.727	0.936
26	GENDER_M	0.7799	0.049	15.856	0.000	0.683	0.877
27	Intercept	1.6114	0.089	18.044	0.000	1.436	1.787

Omnibus:	2.375	Durbin-Watson:	1.386
Prob(Omnibus):	0.305	Jarque-Bera (JB):	2.333
Skew:	-0.071	Prob(Omnibus):	0.311
Kurtosis:	3.297	Cond. No.	1.34e+16

Table 3: OLS Regression Analysis.

value denotes a stronger positive connection, whilst a lower positive value denotes a negative correlation, with the target variable (CGPA).

Table 4 illustrates that features with higher absolute values are thought to be more significant in predicting the target variable.

1	MONTH_April	0.075497
2	MONTH_August	0.083977
3	MONTH_December	0.159195
4	MONTH_February	0.091975
5	MONTH_January	0.150559
6	MONTH_July	0.264275
7	MONTH_June	0.073595
8	MONTH_MAY	0.357240
9	MONTH_March	0.229164
10	MONTH_November	0.055372
11	MONTH_October	-0.024173
12	MONTH_September	0.053403
13	BLOOD_GRP_AA	0.034491
14	BLOOD_GRP_AC	0.269999
15	BLOOD_GRP_AS	0.033609
16	BLOOD_GRP_SC	0.506269
17	BLOOD_GRP_SS	0.766988
18	BLOOD_GNE_A+	0.122213
19	BLOOD_GNE_A-	0.659814
20	BLOOD_GNE_AB+	0.243642
21	BLOOD_GNE_B+	0.213821
22	BLOOD_GNE_B-	0.057915
23	BLOOD_GNE_O+	0.121195
24	BLOOD_GNE_O-	0.192757
25	GENDER_F	0.831455
26	GENDER_M	0.779902
27	intercept	1.611357

Table 4: Feature importance coefficient.

Decision tree regression

Regression analysis that employs a decision tree as the predictive model is known as choice tree regression. Decision trees can be modified for regression tasks, where the objective is to predict a continuous outcome (CGPA), even though they are primarily used for classification tasks, where the goal is to categorize data into predetermined categories.

Mean Squared Error (MSE) is a metric used in decision tree regression that quantifies the average squared difference between the target values predicted by the decision tree model and the actual target values in the dataset.

For each data point (i) in the dataset, the mean squared difference (MSE) between the actual target values (yi) and the anticipated target values (Þi) is computed. Mathematically, it can be represented as:

MSE=n1∑i=1n(yi−y^i)2 ……ii

Where:

- n is the number of data points in the dataset.
- yi is the actual target value for the *h*ith data point.
- y^i is the predicted target value for the *h*ith data point.

The dataset was trained for the decision tree regression model using the python code below.

```
model = DecisionTreeRegressor(random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print('Mean Squared Error:', mse)
Mean Squared Error: 0.28250500491181
```

Result and Discussion

The dependent variable (CGPA) has a variance that can be explained by the independent variables in the model. The coefficient of determination, or R-squared, is 0.040, meaning that the independent variables in the model account for about 4% of the variance in the dependent variable. The R-squared value is adjusted for the number of predictors in the model using this adjusted R-squared value. In this instance, a negative score (-0.005) indicates that the explanatory capacity of the model may be outweighed by its complexity.

The model's overall significance is tested using the F-statistic. In this case, the corresponding probability (p-value) is 0.601, and the F-statistic is 0.8988. The model may not be statistically significant in explaining the variation in CGPA as a whole if the p-value is higher than 0.05, which is the standard significance level.

These findings suggest that the model is not a very good explanation for the variation in CGPA. The non-significant F-statistic and low R-squared imply that the independent factors may not be enough accounting for the variance in the dependent variable. Furthermore, the negative adjusted R-squared value suggests that the complexity of the model may not be warranted.

The features in figure above are classified by gender, blood groups, months, and blood genotypes. Several months, blood types, and gender have been found by the model to be significant predictors of the target variable (CGPA).

With relevance scores of 0.831 and 0.780, respectively, gender (GENDER_F and GENDER_M) appears to be extremely relevant, for instance. This shows that the model's gender component is a powerful predictor.

Blood group features (e.g., BLOOD_GRP_A-, BLOOD_GNE_SS) also appear to have high significance scores, suggesting that they have a significant impact on the predictions made by the model. A few more months with reasonably high relevance scores include May, March, January, and December, indicating that these months are equally important in predicting the target variable.

The squared difference between the actual target values and the predicted values produced by the decision tree regression model is roughly 0.2825 on average, according to the decision tree regression result with an MSE of 0.28250500491181.

In regression tasks, mean square error (MSE) is frequently employed in conjunction with other metrics such as mean absolute error (MAE) and R-squared to evaluate the model. It offers a numerical indicator of how well the model conforms to the facts.

Model representation

$Y=\beta 0+\beta 1X1+\beta 2X2+...+\beta pXp+\varepsilon$ becomes CGPA = β0+ β(MONTH)+ β(BLOOD_GRP)+ β(BLOOD_GNE)+ β(GENDER)+ ε. ….iii This model allows for the reasonably accurate prediction of a student's CGPA based on their blood group, blood genotype, month of birth, and gender. Not only that, but it is also possible to forecast this student's grade.

Consider the case where MONTH January has a coefficient of 0.1506, BLOOD_GRP_A+ has a coefficient of 0.1222, BLOOD_GNE_AA has a coefficient of 0.0345, GENDER_F has a coefficient of 0.8315, and the intercept value is 1.6114 in table 2.

After changing the values of the elements (independent variables) in the model mentioned above (equation iii), we obtain:

1.6114+0.1506+0.1222+0.0345+0.8315 is the CGPA.

2.7502 is the CGPA.

Conclusion

The thorough analytical research on how biological characteristics like gender, blood type, and blood group affect students' academic performance has shed important light on the intricate relationship between biological characteristics and academic success. Several important discoveries and conclusions have been drawn from the thorough study and review of pertinent data, providing insight into the connection between these biological parameters and academic outcomes.

In the first instance, the research has shown links between a few biological characteristics and academic achievement. For example, it has been discovered that there are various degrees of correlation between students' academic achievements and blood group and blood genotype. These results point to possible directions for future research into the underlying processes that could account for these associations, Also, the study has emphasized how crucial it is for academic research to take gender into consideration as a fundamental biological issue. Gender equity in educational results can be promoted by educational policies and initiatives, as evidenced by the gender differences in academic achievement that were noted.

The study has also brought attention to the necessity of a comprehensive strategy for comprehending the factors that influence academic achievement. Although biological elements are important, they work in concert with numerous other socio-economic, environmental, and psychological factors to influence kids' academic paths. It is consequently imperative that future research projects use interdisciplinary approaches in order to fully capture the intricacy of these connections.

Bibliography

1. Barun M., *et al*. "Association between blood groups and academic performance: A systematic review". *Educational Research Review* 48 (2022): 112-125.

2. Buoye A and Ojuawo A. "Gender differences in programming ability among students: A comparative analysis". 1st National Conference of WITED, Ilaro Chapter (2019).

3. The Federal Polytechnic, Ilaro August 13-16 (2019).

4. Buoye PA and Alawode A J. "Analytical Approach to Investigating the Influence of Blood Group and Blood Genotype on the Performance of Students of Federal Polytechnic, Ilaro". *Federal Polytechnic Ilaro Journal of Pure and Applied Sciences* 4.2 (2022): 27-36.

5. Joseph K. "Genotype polymorphism and academic performance: A comparative study". *Educational Psychology Review* 31.4 (2019): 567-580.

6. Kumar A., *et al*. "Blood group-and gender-wise comparison of big five models of personality among medical students". *National Journal of Physiology, Pharmacy and Pharmacology* 11.1 (2021): 1-15.

7. Medbury M. "How blood genotype and blood group increases your quality of life" (2024).

8. Melissa A and Dominic B. "Understanding blood group and blood genotype". *Journal of Genetics Education* 45.3 (2021): 213-225.

9. Nikam L H., *et al*. "Association of ABO blood groups with academic stress and academic performance of the first-year undergraduate medical students in an Indian Medical College". *National Journal of Physiology, Pharmacy and Pharmacology* 13.4 (2023): 698-698.

10. Rihab A. "Relationship between personality traits and blood groups among Palestinian university students: A cross-sectional study". *International Journal of Psychology and Behavioral Sciences* 10.2 (2020): 34-42.

11. Sachin S. "Blood group and genotype: A comprehensive explanation". *Journal of Biological Sciences* 87.2 (2022): 89-102.

12. Tsaousis I and Alghamdi M H. "Examining academic performance across gender differently: Measurement invariance and latent mean differences using bias-corrected bootstrap confidence intervals". F*rontiers in Psychology* 13 (2022): 896638.

13. Wang X., *et al*. "Influence of ABO blood group on memory performance in elderly subjects with subjective memory complaints". *Archives of Gerontology and Geriatrics* 59.3 (2014): 645-649.

14. Wrigley-Asante C., *et al*. "Gender differences in academic performance of students studying Science Technology Engineering and Mathematics (STEM) subjects at the University of Ghana". *SN Social Sciences* 3.1 (2023): 12.