International Journal of Current Research and Applied Studies, (IJCRAS)

ISSN: 2583-6781

available at https://ijcras.com/

Volume 3 Issue 5 Sep-Oct 2024, Page 65-74

To cite this article: Aiyegbusi Akinyemi Edward, Ebisin Aderonke Foluso, Abimbola Olayemi Grace, & Adedokun Akeem Ademola (2024). GMDH-SHELL-DS AS A TOOL FOR EDUCATIONAL IMPROVEMENT: AN ANALYTICAL APPROACH TO STUDENT SUCCESS, International Journal of Current Research and Applied Studies (IJCRAS) 3 (5): Article No. 93, Sub Id 147

GMDH-SHELL-DS AS A TOOL FOR EDUCATIONAL IMPROVEMENT: AN ANALYTICAL APPROACH TO STUDENT SUCCESS

Aiyegbusi Akinyemi Edward, Ebisin Aderonke Foluso, Abimbola Olayemi Grace, & Adedokun Akeem Ademola

Department of Computer Science, College of Information and Technology Education, Lagos State University of Education, Lagos Nigeria
Ogun State Institute of Technology, Igbesa Nigeria

DOI: https://doi.org/10.61646/IJCRAS.vol.3.issue5.93

ABSTRACT

In the rapidly evolving landscape of education, data-driven strategies have become paramount in fostering student success. Group Method of Data Handling (GMDH) Shell-DS, an advanced analytical tool, presents a novel approach to addressing educational challenges by leveraging predictive modelling and optimization techniques. This paper explores the potential of GMDH-Shell-DS in enhancing educational outcomes, focusing on its application in identifying key factors influencing student success, optimizing learning strategies, and improving overall educational quality. Through a comprehensive analysis, the paper demonstrates how GMDH-Shell-DS can be a transformative tool for educators, administrators, and policymakers aiming to elevate educational standards.

Keywords: GMDH-Shell-DS, Data Handler, Data-Driven Strategies, Student Success, Educational Improvement

INTRODUCTION

The integration of advanced analytical tools into the educational sector has gained significant attention in recent years. With the increasing availability of educational data, there is a growing need for sophisticated methods to analyze and interpret this data to improve student outcomes. GMDH-Shell-DS, a tool based on the Group Method of Data Handling, offers a robust framework for predictive modeling and data-driven decision-making. This paper aims to explore the application of GMDH-Shell-DS in the educational context, focusing on how it can be used to enhance student success through better understanding and optimization of learning processes. Artificial intelligence (AI) has the power to revolutionize the field of education by enhancing teaching and learning experiences, improving student outcomes, and streamlining administrative tasks. The group method of data handling (GMDH) was first introduced by Ivakhnenko as a proper approach for detecting nonlinear systems (Ivakhnenko, 1968).

According to Omar Boubker (2023) Group Method of Data Handling was applied in a great variety of areas for deep learning and knowledge discovery, forecasting and data mining, optimization and pattern recognition. Inductive GMDH algorithms give possibility to find automatically interrelations in data, to select an optimal structure of model or network and to increase the accuracy of existing algorithms.

Group Method of Data Handling (GMDH)-type neural network algorithms are the heuristic self-organization method for the modelling of complex systems. GMDH algorithms are utilized for a variety of purposes, examples include identification of physical laws, the extrapolation of physical fields, pattern recognition, clustering, the approximation of multidimensional processes, forecasting without models, etc (Dag & Yozgatligil 2016). In the study of Dag & Yozgatligil (2016), the R package GMDH was presented to make short term forecasting through GMDH-type neural network algorithms. The GMDH package had options to use different transfer functions (sigmoid, radial basis, polynomial, and tangent functions) simultaneously or separately.

The group method of data handling (GMDH) employs a family of inductive algorithms for computer-based mathematical modeling grounded on a combination of quadratic and higher neurons in a certain number of variable layers. Therefore, the study answers the research question “Will GMDH-Shell-DS software be used to predict students’ academic success?”

LITERATURE REVIEW

Overview of GMDH-Shell-DS

GMDH-Shell-DS is a powerful software tool designed for predictive analytics, data mining, and optimization. It is based on the Group Method of Data Handling (GMDH), a self-organizing algorithm that builds predictive models from data sets. GMDH-Shell-DS automates the process of model creation, selecting the most relevant variables, and determining the optimal structure for the model. This makes it particularly useful in environments where large volumes of data need to be analyzed quickly and accurately. Accurate forecasting model for under-five mortality rate (U5MR) is essential for policy actions and planning. While studies have used traditional time series modeling techniques (e.g., autoregressive integrated moving average (ARIMA) and Holt-Winters smoothing exponential methods), their appropriateness to predict noisy and non-linear data (such as childhood mortality) has been debated. The objective of this study was to model long-term U5MR with group method of data handling (GMDH)-type artificial neural network (ANN), and compare the forecasts with the commonly used conventional statistical methods (Adeyinka & Muhajarine, 2020). Sarycheva, (2003) presented an investigation into the efficacy of two validation strategies in predicting the net power output from the plant using GMDH Shell software. Using the combinatorial algorithm, the k-fold cross-validation strategy and the training/testing validation technique were applied to empirical data of a power plant in Nigeria.

Adeyinka et al., (2021) used an artificial intelligence time series (GMDH- type ANN) to forecast age-specific childhood mortality rates (neonatal and under- five) and sex- specific U5MR from 2018 to 2030. The data sets were the yearly historical mortality rates between 1960s and 2017, obtained from the World Bank website. Two scenarios of mortality trajectories were simulated: (1) status quo scenarios—assuming the current trend continues; and (2) acceleration scenarios—consistent with the SDG targets. Yefimenko, (2018) analyzed approaches to prediction of economic processes in business intelligence systems. Contemporary tools of predictive analytics, used for effective making of business decisions, are considered. The concept of advanced GMDH-based predictive analytics tool was proposed.

Babatunde et al., (2020) presented an investigation into the efficacy of two validation strategies in predicting the net power output from the plant using GMDH Shell software. Using the combinatorial algorithm, the k-fold cross-validation strategy and the training/testing validation technique were applied to empirical data of a power plant in Nigeria. The performance of the models returned from the two validation strategies was evaluated using maximum negative error, maximum positive error, mean absolute percentage error (MAPE), root mean square percentage error (RMSPE), residual sum, the standard deviation of residuals, coefficient of determination (R^2) and correlation.

Alfaifi & Bonakdari, (2023) postulated a new approach for predicting the geometrical characteristics of the mixing behaviour of an inclined dense jet for angles ranging from 15° to 85° proposed in the study. The approach is called the group method of data handling (GMDH) which was based on the artificial neural network (ANN) technique. The proposed model was trained and tested using existing experimental data reported in the literature. The results of the coefficient of determination (R2) indicated a high accuracy of the proposed model with values of 0.9719 and 0.9513 for training and testing for the dimensionless of the distance from the nozzle to the return point xr/D, and 0.9454 and 0.9565 for training and testing for the dimensionless of the terminal rise height yt/D. Moreover, four previous analytical models were used to evaluate the GMDH model. The results showed the superiority of the proposed model in predicting the geometrical characteristics of the inclined dense jet for all tested angles. Finally, the standard error of estimate (SEE) was applied to demonstrate which model performed the best in terms of getting closer to the actual data. The results illustrated that all fitting lines of the GMDH model performed very well for all geometrical parameter predictions and was the best model with approximately 10% error, which was the lowest value of error among the models. Therefore, this study confirms that the GMDH model can be used to predict the geometrical properties of the inclined negatively buoyant jet with high performance and accuracy.

Mithiya; Lakshmikanta & Mandal, (2019) looked into the need to address the deficit motivated a systematic study of the oilseeds economy to formulate appropriate strategies to bridge the demand-supply gap. In their study, an effort was made to forecast oilseed production by using Autoregressive Integrated Moving Average (ARIMA) model, which was the most widely used model for forecasting time series. One of the main drawbacks of this model is the presumption of linearity. The Group Method of Data Handling (GMDH) model has also been applied for forecasting oilseed production because it contains nonlinear patterns. Both ARIMA and GMDH are mathematical models well-known for time series forecasting. The results obtained by the GMDH are compared with the results of the ARIMA model. The comparison of modelling results shows that the GMDH model performed better than the ARIMA model in terms of mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE). The experimental results of both models indicated that the GMDH model was a powerful tool to handle the time series data and it provided a promising technique in time series forecasting methods. Дикан et al., (2020) use the Group Method of Data Handling (GMDH)-GMDH Shell DS software to deals with the construction of the norm-pathology states classifiers according to statistical features of the ultrasound images texture in diffuse liver diseases. The work was performed on data, had provided by the Nuclear Medicine and Radiation Diagnostics Institute of the National Academy of Medical Sciences of Ukraine.

Applications of GMDH-Shell-DS in Education

In the educational domain, GMDH-Shell-DS can be applied in various ways, including:

Predicting Student Performance: By analyzing historical data on student performance, GMDH-Shell-DS can identify patterns and predict future outcomes, allowing educators to intervene early and provide targeted support to students at risk.

Optimizing Learning Strategies: GMDH-Shell-DS can help in determining the most effective teaching methods and learning strategies for different groups of students, based on their unique characteristics and learning styles.

Resource Allocation: The tool can be used to optimize the allocation of educational resources, ensuring that they are used where they are most needed to maximize student success.

According to Mir, (2023) to improve the country's educational system and provide a blueprint for the future, the government has come up with a National Education Policy. National Education Policy emphasizes the use and integration of technology in school and higher education that will enable people of the population to meet the requirements of quality education in line with the SDGs in the 21st century. In the year 2013, Najafzadeh & Azamathulla, (2013) used the neuro-fuzzy-based group method of data handling (NF-GMDH) as an adaptive learning network was used to predict the scour process at pile groups due to waves. The NF-GMDH network was developed using the particle swarm optimization (PSO) algorithm and gravitational search algorithm (GSA). Effective parameters on the scour depth include sediment size, geometric property, pile spacing, arrangement of pile group, and wave characteristics upstream of group piles. Seven dimensionless parameters were obtained to define a functional relationship between input and output variables. Published data were compiled from the literature for the scour depth modelling due to waves. The efficiency of training stages for both NF-GMDH-PSO and NF-GMDH-GSA models was investigated. The results indicated that NF-GMDH models could provide more accurate predictions than those obtained using model trees and traditional equations.

METHODOLOGY

To utilize GMDH-Shell-DS effectively, it is essential to have access to high-quality data. In this study, data was collected from various educational institutions, including student demographics, and academic performance records. GMDH Shell DS (General Modeling for Data Handling Shell for Data Science) 3.8.9 was used to predict students learning outcomes using Mock examination. The dataset ratio splits used was based on 70% training set and 30% testing set. The data was cleaned and preprocessed to ensure that it was suitable for analysis. The GMDH-Shell-DS software was used to develop predictive models for student success. The data was divided into training and testing sets, with the training set used to build the model and the testing set used to evaluate its accuracy. The software's automated model-building capabilities were leveraged to select the most relevant variables and determine the optimal model structure. Once the model was developed, it was validated using the testing data set. The model's predictions were compared to actual outcomes to assess its accuracy. The model was then optimized to improve its predictive power, ensuring that it could provide reliable insights for decision-making.

ANALYSIS AND RESULTS

Using GMDH-Shell-DS Software Application for Prediction

GMDH Shell DS (General Modeling for Data Handling Shell for Data Science) 3.8.9 was used to predict students learning outcomes using Mock examination. The dataset ratio splits used was based on 70% training set and 30% testing set. The result from the table classified students learning outcomes into three: Evidence to pass WASSCE with good grade, Evidence to pass WASSCE with average grade, and Evidence to pass WASSCE with poor grade based on their performances in Mock examination. Table 1 shows the predicted results of first 20 results while table 4.20 shows the last 20 results respectively.

Table 1: The first 20 results

ID	Actual	Prediction	Hit/Miss
1	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
2	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
3	Evidence to pass with average grade	Evidence to pass with average grade	Hit
4	Evidence to pass with average grade	Evidence to pass with average grade	Hit
5	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
6	Evidence to pass with poor	Evidence to pass with poor	Hit

	grade	grade	
7	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
8	Evidence to pass with average grade	Evidence to pass with average grade	Hit
9	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
10	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
11	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
12	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
13	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
14	Evidence to pass with average grade	Evidence to pass with average grade	Hit
15	Evidence to pass with average grade	Evidence to pass with average grade	Hit
16	Evidence to pass with average grade	Evidence to pass with average grade	Hit
17	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
18	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
19	Evidence to pass with poor grade	Evidence to pass with average grade	Miss
20	Evidence to pass with average grade	Evidence to pass with average grade	Hit

Table 2: The last 20 results

1180	Evidence to pass with average grade	Evidence to pass with average grade	Hit
1181	Evidence to pass with average grade	Evidence to pass with average grade	Hit
1182	Evidence to pass with average grade	Evidence to pass with average grade	Hit
1183	Evidence to pass with average grade	Evidence to pass with average grade	Hit
1184	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1185	Evidence to pass with good grade	Evidence to pass with good grade	Hit
1186	Evidence to pass with average grade	Evidence to pass with average grade	Hit
1187	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1188	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1189	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1190	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1191	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1192	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1193	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1194	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1195	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1196	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1197	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1198	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1199	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit
1200	Evidence to pass with poor grade	Evidence to pass with poor grade	Hit

The results show that most of actual and predictions hit – means they tallied with each other

Table 3: Summary of prediction strength

	Evidence to pass with average grade	**Evidence to pass with good grade**	**Evidence to pass with poor grade**
Accuracy	0.989	0.998	0.991
Precision	0.972	1.00	1.00
Recall	1.00	0.971	0.984
F-Measure	0.986	0.985	0.992
RMSE	0.195		

All the above evidences prove that the classification model used was accurate with high precision, f-measure, RMSE and recall to predict students learning outcomes (see fig.1)

Correctly classified instances 1187 98.9% RMSE 0.195 x-measure 0.980
Incorrectly classified instances 13 1.1% Weighted F-measure 0.989

		Predicted class				
		Evidence to pass with average grade	**Evidence to pass with good grade**	**Evidence to pass with poor grade**	**Total**	**Recall**
Actual class	**Evidence to pass with average grade**	454	0	0	454	1.000
	Evidence to pass with good grade	2	67	0	69	0.971
	Evidence to pass with poor grade	11	0	666	677	0.984
	Total	467	67	666	1200	
	Precision	0.972	1.000	1.000		
	F-measure	0.986	0.985	0.992		
	Baseline	0.622	0.943	0.564	0.564	
	Accuracy	0.989	0.998	0.991	0.989	

Figure 1: Classification performance of the algorithm

DISCUSSION

The GMDH-Shell-DS models demonstrated a high level of predictive accuracy, successfully identifying students at risk of underperforming and suggesting effective interventions. The results showed that the tool could accurately predict student performance based on a variety of factors, including previous academic achievements, attendance records, and engagement in class activities. The insights gained from the GMDH-Shell-DS analysis were used to inform the development of new educational strategies. For example, the tool identified specific teaching methods that were particularly effective for certain student groups, leading to the adoption of more personalized learning approaches. Additionally, the analysis highlighted areas where resources could be better allocated to improve student outcomes. GMDH approach can be useful because: Optimal complexity of the model structure is found, adequate to the level of noise in data sample. For real problems, with noised or short data, simplified optimal models are more accurate. The number of layers and neurons in hidden layers, model structure and other optimal hyperparameters are determined automatically. It guarantees that the most accurate or unbiased models will be found - method doesn't miss the best solution during sorting of all variants (in the given class of functions) As input variables can be used non-linear functions or features, that may have influence on output variable. It automatically finds interpretable relationships in data and selects effective input variables. GMDH sorting algorithms are rather simple for software development. Twice-multilayered neural nets can be used to increase the accuracy of another modelling algorithms. Method gets information directly from data sample and minimizes influence of apriori author assumptions about results of modeling. Approach gives possibility to find the objective physical model of the object (law or segmentation) - the same on future samples.

GMDH-Shell-DS proved to be a valuable tool, there were some challenges and limitations. The quality of the input data significantly impacted the accuracy of the predictions, underscoring the importance of robust data collection and preparation processes. Additionally, while the tool is powerful, it requires a certain level of expertise to interpret the results and implement the recommended strategies effectively.

CONCLUSION

GMDH-Shell-DS has the potential to significantly enhance educational outcomes by providing data-driven insights that can be used to optimize learning strategies and improve resource allocation. As educational institutions continue to collect more data on student performance, tools like GMDH-Shell-DS will become increasingly important in helping educators make informed decisions that support student success. However, it is essential to address the challenges associated with data quality and to ensure that educators are adequately trained in using these tools to fully realize their potential. GMDH-type neural network performed better in predicting and forecasting of under-five mortality rates for Nigeria, compared to the ARIMA and Holt-Winters models. Therefore, GMDH-type ANN might be more suitable for data with non-linear or unknown distribution, such as childhood mortality. GMDH-type ANN increases forecasting accuracy of childhood mortalities in order to inform policy actions in Nigeria.

RECOMMENDATIONS

- Investment in Data Quality: Educational institutions should invest in improving the quality of the data they collect to ensure that tools like GMDH-Shell-DS can provide accurate and reliable insights.
- Training for Educators: Educators should be trained in using GMDH-Shell-DS and interpreting its results to make informed decisions that benefit students.
- Further Research: Additional research should be conducted to explore the full range of applications of GMDH-Shell-DS in education and to develop best practices for its use.

REFERENCES

Adeyinka DA, Petrucka PM, Isaac EW & Muhajarine Nazeem (2021). Changing patterns of gender inequities in childhood mortalities during the Sustainable Development Goals era in Nigeria: findings from an artificial neural network analysis. BMJ Open 2021;11:e040302. doi:10.1136/ bmjopen-2020-040302

Adeyinka, D. A., & Muhajarine, N. (2020). Time series prediction of under-five mortality rates for Nigeria: comparative analysis of artificial neural networks, Holt-Winters exponential smoothing and autoregressive integrated moving average models. BMC medical research methodology, 20(1), 292. https://doi.org/10.1186/s12874-020-01159-9

Alfaifi, Hassan & Bonakdari, Hossein. (2023). A comprehensive Study for Predicting the Geometrical Characteristics of an Inclined Negatively Buoyant Jet Using Group Method of Data Handling (GMDH) Neural Network. 10.20944/preprints202312.0853.v1.

Babatunde, Damilola & Anozie, Ambrose & Omoleye, James & Oyebode, Oluwaseun & Babatunde, Olubayo & Ayeni, A.O.. (2020). Modelling the power output from a steam power plant in Nigeria. Energy Reports. 6. 822-828. 10.1016/j.egyr.2020.11.126.

Dag, Osman & Yozgatligil, Ceylan (2016). GMDH: An R Package for Short Term Forecasting via GMDH-Type Neural Network Algorithms, The R Journal 8(1), 379 - 386

Damilola E. Babatunde, Ambrose N. Anozie, James A. Omoleye, Oluwaseun Oyebode, Olubayo M. Babatunde & Augustine O. Ayeni (2020). Modelling the power output from a steam power plant in Nigeria, https://doi.org/10.1016/j.egyr.2020.11.126

Ivakhnenko AG (1968). The group method of data of handling; a rival of the method of stochastic approximation. Soviet Auto Contr 13(1):43–55

Mir, Aijaz. (2023). THE SCOPE OF TECHNOLOGY IN NATIONAL EDUCATION POLICY 2020: A STUDY. 83. 1-4.

Mithiya, Debasis; Lakshmikanta, Datta & Mandal, Kumarjit. (2019). Time Series Analysis and Forecasting of Oilseeds Production in India: Using Autoregressive Integrated Moving Average and Group Method of Data Handling – Neural Network. Asian Journal of Agricultural Extension, Economics & Sociology. 1-14. 10.9734/ajaees/2019/v30i230106.

Najafzadeh, Mohammad & Azamathulla, Hazi. (2013). Neuro-Fuzzy GMDH to Predict the Scour Pile Groups Due to Waves. Journal of Computing in Civil Engineering. 29. 04014068. 10.1061/(ASCE)CP.1943-5487.0000376.

Omar Boubker (2023). From chatting to self-educating: Can AI tools boost student learning outcomes? https://doi.org/10.1016/j.eswa.2023.121820Get rights and content

Sarycheva, L. (2003). Using GMDH in ecological and socio-economical monitoring problems. Systems Analysis Modelling Simulation, 43(10), 1409–1414. https://doi.org/10.1080/02329290290024925

Yefimenko, Serhiy (2018). Advances in GMDH-based Predictive Analytics Tools for Business Intelligence Systems, ACIT 2018, June 1-3, Ceske Budejovice, Czech Republic

Дикан, I; Tarasyuk, Boris; Павлов, В; Круглий, В; Диба, Марина; Babenko, Vitalii; Nastenko, Ievgen & Nosovets, Olena. (2020). Diffuse Liver Disease States Classification Based on Statistical Features of the Ultrasound Images Texture and GMDH [in Ukranian]. 11. 54-66.