Frontiers in Health Informatics
ISSN-Online: 2676-7104
www.healthinformaticsjournal.com
2024; Vol 13: Issue 6
Open Access

Analysis of Schools Education Quality by Elephant Herd & Neural Network Model

DIPAK KADVE[1] 0000-*0002-2571-9973* **DR.BINOD KUMAR**[2] *0000-0002-6172-7938*

JSPM'S RAJARSHI SHAHU COLLEGE OF ENGINEERING PUNE MAHARASHTRA INDIA
DR A V SENTHIL KUMAR[3] HINDUSTHAN COLLEGE OF ARTS & SCIENCE COIMBATORE TAMILNADU INDIA

Cite this paper as: DIPAK KADVE , DR.BINOD KUMAR (2024) Analysis of Schools Education Quality by Elephant Herd & Neural Network Model.*Frontiersin Health Informatics, 13* (6) 53-66

ABSTRACT- In the past few years, governments have introduced different programs to make school work better. Details about these efforts are available in various formats, such as simple statistics and detailed reports. This study will examine data to discover ways to enhance schools. This paper focuses on understanding the characteristics of schools, aiming to identify which ones are crucial for analysis of school grades. Input dataset was pre-processed and privacy preserving steps were involved that increases the security in the model. Elephant herd algorithm identified school related features that improves the work accuracy. Selected features were used for the training of neural network. It was found that use of effective training feature improves the school grade prediction accuracy of proposed SQAEHO (School Quality Analysis by Elephant Herd Optimization) model, as compared to other existing model.

Keywords: Data mining, Genetic Algorithm, Privacy Preserving, Neural Network.

I. INTRODUCTION

In the educational landscape of India, the schooling system operates on a two-tier structure, spanning a total of twelve years. The initial decade is dedicated to imparting general education, succeeded by a subsequent two-year phase termed as senior secondary education. This latter segment, often referred to as Higher Secondary Education, assumes paramount significance owing to its pivotal role in shaping the academic trajectory of individuals [1, 2]. It serves as a critical juncture wherein students make crucial decisions regarding their preferred subjects for advanced studies in tertiary education institutions such as colleges and universities. Consequently, the transition from secondary to higher education hinges significantly upon the outcomes of this phase.
Data mining, as a methodology, is intricately woven around the meticulous identification and analysis of patterns, correlations, and prevailing trends within vast datasets. This multifaceted approach encompasses a sophisticated amalgamation of statistical techniques and advanced machine learning algorithms. Through the systematic application of these methodologies, the primary objective is to unearth hidden knowledge and glean valuable insights from the wealth of information encapsulated within the data [3, 4].

People use Educational Data Mining to try to figure out how students will do in school [2, 5]. But it's hard because lots of things like where students are from, what kind of person they are, how much money their family has, and where they live can all affect how well they do. These things can be different depending on what grade they're in, what classes they take, and where they go to school. So, figuring out how students will do in school means looking at all these different things to get a good guess [6].
Many of researchers has developmed modles of prediction student grades like Statical method [8], clustering method [9], pattern mining [10], etc. Some of researchers implement different application models in education data anylysis such as wireless network [11]. Finaly this area moves towards multiclass grade prediction [12], by using leaning models [13]. This work create a show that will anticipate the understudy based on researcher, established, environment, highlights. A few of major issue of this investigate is to recognize the compelling features/factors. As understudy grades depends on

different variables from zone to zone and family to family. These calculate recognizable proof require to be optimize as per work.

Following targets were accomplished by the proposed model:

(a) Foundation of a comprehensive database comprising prescient factors that can offer bits of knowledge into scholarly execution trends.
(b) Prepare a demonstrate that foresee the school grades as per the watched factors.

Through these targets, the examination looks for to outfit important direction and assets to address the needs of understudies confronting challenges in their interest of scholarly victory at the higher auxiliary level.

Rest of paper was partitioned into few areas. Moment area brief the school/student review expectation models done by past analysts. Thirst area clarify the proposed SQAEHO (School Quality Examination by Elephant Group Optimization) show. Another area appears the test yield values with comparison from existing models. At last paper was conclude in fifth section.

II. Writing SURVEY

D. Sun et. al. in [14], proposed a demonstrate centers on analyzing chronicled scholarly grades from different measurements among college understudies to extricate highlights that uncover connections between courses and understudies, among distinctive understudies themselves, or indeed between courses themselves. Also, an consideration component is presented to investigate the relationship between diverse dimensional highlights. This consider collected a triplet set of related courses and students' genuine verifiable grades, demonstrated the relationship between courses through information investigation, and confirmed the viability of distinctive dimensional features.

Chaka(15) explored and anatomized 32 papers on instructional information mining procedures and computations for anticipating understudy educational prosecution. Bracket, clustering, cooperation rules, and fall developed as the most habitually employed procedures to figure understudy educational palm. The Credulous Bayes approach was the fewest employed. The choice tree was the most regularly employed of the seven regularly employed computations honored by the 26 check ponders for anticipating understudy scholarly prosecution.

Chen et al.(16) delved the extent to which classifiers grounded on reading behaviours might prognosticate academic achievement for university scholars. also, he looked into which features taken from the reading logs affected the prognostications. He claimed that grounded on the delicacy, perfection, and recall criteria , logistic retrogression, Gaussian naive Bayes, supporting vector bracket, decision trees, arbitrary timbers, and neural networks each produced relatively accurate prognostications. Turning runners, going back and forth between runners, adding and removing marks, and editing and deleting memos were other pupil online reading behaviours that impacted the vaticination models.

D. A. Bujang et. al. in(17), presents a comprehensive analysis of machine literacy ways to prognosticate the final pupil grades in the first semester courses by perfecting the performance of prophetic delicacy. Two modules will be stressed in this paper. First, we compare the delicacy performance of six well- known machine literacy ways. Second, proposed a multiclass vaticination model to reduce the overfitting and misclassification results caused by imbalancedmulti-classification grounded on oversampling Synthetic nonage Oversampling fashion(SMOTE) with two features selection styles.

Yiwen Zhang and associates(18) proposed a new grade vaticination approach that integrates an educational sphere knowledge graph with cooperative filtering. This system involves collecting course semantic data and erecting a course knowledge graph to support grade vaticination. Experimental results have shown that combining the educational knowledge graph with cooperative filtering reveals fresh semantic connections between courses, leading to more accurate grade prognostications.

III. Proposed Methodology

In this section proposed SQAEHO(School Quality Analysis by Elephant Herd Optimization) was detailed. Fig. 1 shows the inflow of the SQAEHO model where different memos were list in table 1. Dataset was taken from the pupil of primary classes. Data collection was done from different position where colorful features were collect on the base of academy structure and conditioning. This model identifies the features by applying giant herd optimization algorithm. farther optimized data was used for the training of Error back propagation neural network. List of academy features were Library, Electricity, Toilets, Drinking Water, Counseling Room, Girls Restroom, Computer Lab, Dispensary, Science Lab, Smart Classes, Covered Playground, Open Playground, Auditorium, Cycle Stand.

Table 1. Notation used in SQAEHO.

SRD	School Raw Dataset
SPPD	School Pre-Processed Dataset
Hp	Elephant Herd population
Sf	School features in dataset
E	Elephant Herd
e	Number of elephant in population
Sg	School Grades
Hf	Herd Fitness
EHf	Elephant Herd Feature
NN	Initial Neural Network
SGPM	School Grade Prediction Model

Dataset Pre-processing As input raw dataset have limitation of matrix operation hence transformation of data into tabular was done in this step [19]. Further data was processed to remove the noise column data [20, 21]. Many of school data having the features in Yes/No were update into 1/0.

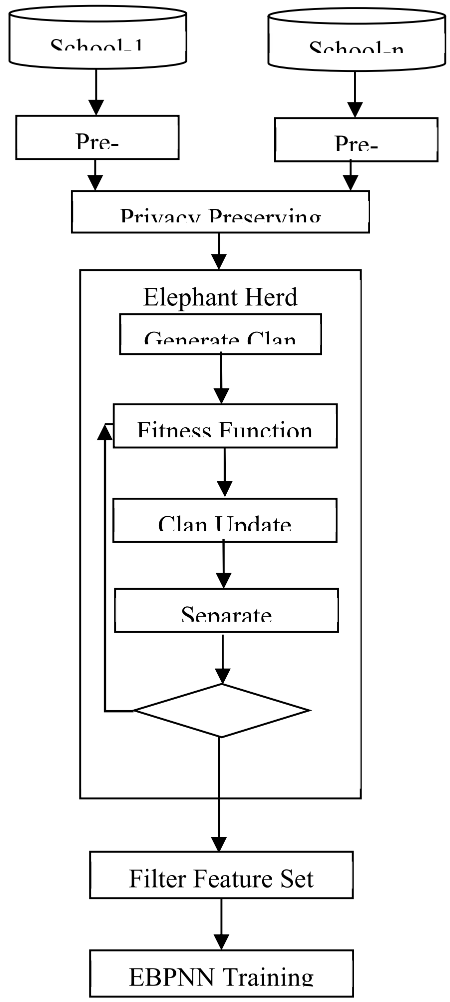

Fig. 1 Block diagram of proposed SQAEHO model.

SPPD←Cleaning_school_Data(SRD)----Eq. 1

Desired output of the school grade class is the average of student marks.

$$Am \leftarrow \frac{\sum_{sm=1}^{st} St\_marks_{SM}}{st}$$

$$Sg=\begin{cases} 3 & Am < c3 \\ 2 & C3 < Am < C2 \\ 1 & C1 < Am \end{cases}$$

Preprocessed dataset need to be optimize for the training of mathematical model. This paper has optimize model by the Elephant herd optimization algorithm.

Privacy Preserving Operations

As dataset have sensitive information like school name, address, student roll number, class, city, etc. Hence few of information were removed other will transformed. This transformation was done on the basis that information of data was not get compromised. School name, student name were of no use in the model hence can be removed [22]. While student roll number, school id will continue. Further student gender were used by transforming information format from male/female to 1/0. In this work in order to provide the school address privacy work has superclass it to the city pincode [22]. Finally the individual student marks were replace by the average marks as this help to understand that what set of parameter like Library, Electricity, Toilets, Drinking Water, Counseling Room, Girls Restroom, Computer Lab, Dispensary, Science Lab, Smart Classes, Covered Playground, Open Playground, Auditorium, Cycle Stand increase or decreases the school performance.

Elephant Clan Population

Random set of features were select from the available dataset. This selection was done on the basis of Gaussian distribution function. Elephant is term as feature set in the work. Each elephant vector is collection of 1 and 0 having Sf number of elements. Each elephant in population is different from other as per the vector, where each position in the vector is representing a feature in SPPD 1 means feature is present and 0 means feature is absent. Eq. 2 shows that herd population Hp has e number of elephants.

Hp←Generate_Herd(SPPD, e, Sf)

Fitness Value

In order to evaluate the fitness of each elephant in C fitness is estimate [23]. Selected elephant features were used for the training of neural network. These features were in binary format where 1 means feature value is pass in the learning model while 0 means feature value is not use for training. After training same set of features were used for the testing and correct school grade class detection accuracy is termed as fitness value.

Input: Hp, SPPD, Sg
Output: Hf // Her Fitness

1. Loop 1:e
2. Loop 1:Sf
3. if Hp[e] is equal to 1
4. Tf←SPPD[Sf]
5. Sg← SPPD[class] // brief in Feature selection
6. Endif
7. EndLoop
8. TNN←Train_Neural_Network(Tf, Sg)
9. Hf←Test_Neural_Network(TNN, Tf, Sg)
10. EndLoop

Mf←max(Hf)

Clan Update

A best solution matriarch, M is derived based on the fitness values Hf, of each elephant in the population [24]. A number of the statuses were randomly changed based on the best matriarch, M feature set. The cloning was done by placing the best elephant set page in other elephant of clan.

Hp←Population_update(M_b, Hf, hp)

Separating

Low fitness elephant were removed from the clan in form of male elephant [25]. This is done after estimating the new clan fitness value.

Hf←Fitness_Value(Hp,Sg,SPPD)

Rank←Sort_decending(Hf)
Hp←Separating_update(Hp, Rank)

Feature Selection
Elephant having highest fitness value is the final feature set for the traing of model. Hence as per the Elephant herd optimization algorithm feature were filtered for the training. As per the school desired Sg is pass for the training. Same set of features were used I the testing of the model.

Training of Neural Network
Error back propogation neural network was used in the work for the learning of selected features [26]. Two hidden layer achetecture was used in the model where first hidden layer have 10 nodes and second has 7 nodes. Further output layer has one node that predict either of grades.

Proposed SQAEHO **Algorithm:**
Input: SRD // School Raw Dataset
Output: SGPM // School Grade Prediction model

1. SPPD←Cleaning_school_Data(SRD)
2. Sg←Desired_output(SPPD)
3. **Hp←Generate_Herd(SPPD, e, Sf)**
4. Loop 1:itr
5. Hf←Fitness_Value(Hp,Sg,SPPD)
6. Hp←Population_update(M_b, Hf, hp)
7. Hf←Fitness_Value(Hp,Sg,SPPD)
8. Rank←Sort_decending(Hf)
9. Hp←Separating_update(Hp, Rank)
10. EndLoop
11. Hf←Fitness_Value(Hp,Sg,SPPD)
12. **Mf←max(Hf)**
13. Ehf←Feature_Selection(Mf, SPPD)
14. SGPM←Train_Neural_Network(Ehf,Sg)

Above algorithm gives a trained SGPM model that predict the school performance grade based on the features selected by the elephant herd optimization.

V. Experiment and Results

Implementation of this model is done on MATLAB 2016a software. Experiment was done on machine having i3 6th generation processor and 4 GB RAM. Details about the dataset was done in table 2. Proposed model SQAEHO (School Quality Analysis by Elephant Herd Optimization) was compared with exsiting model k-PPD-ERT [27].

Table 2 Description of dataset.

Parameters	**Values**
Studnets	4272
School	
Features	29

School Grades	A, B, C
Towns	16

Evaluation Parameters

In order to compared proposed model with other existing models following set of parameters were evaluated.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F\_Measure = \frac{2 * Prcision * Recall}{Precision + Recall}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Results

Table 3. Accuracy value of school grades prediction models for all classes of different testing dataset.

Testing Dataset	SQAEHO	k-PPD-ERT
20	94.12	47.0588
40	94.74	47.368
60	95	40
80	95.35	39.5349
100	96.23	41.6667

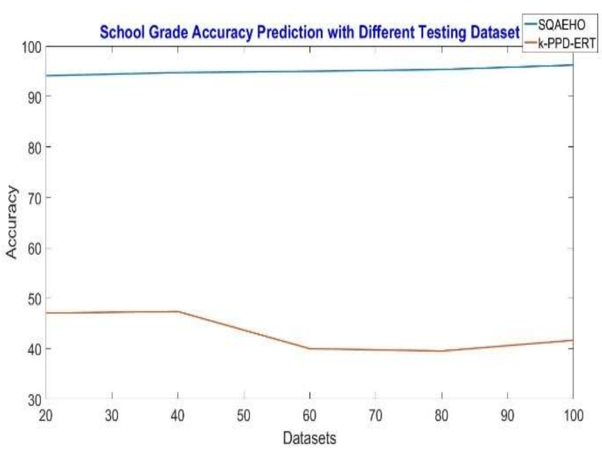

Fig. 2 Accuracy of school grades by varying testing percentage.

Table 3 shows the accuracy percentage of the school grade prediction model. It was found that use of selected features from the elephant herd optimization algorithm has increases the learning rate. It was found that average accuracy was enhanced by more than 50%, this is just because previous work develop tree on the basis of privacy preserving feature values.

Table 4. Precision value of school grades prediction models for all classes of different testing dataset.

Testing Dataset	SQAEHO	k-PPD-ERT
20	1	1
40	1	1
60	1	1
80	1	1
100	1	1

Table 4 shows that precision value of the both comparing model is same in all set of testing dataset. As models find the best class in all sets.

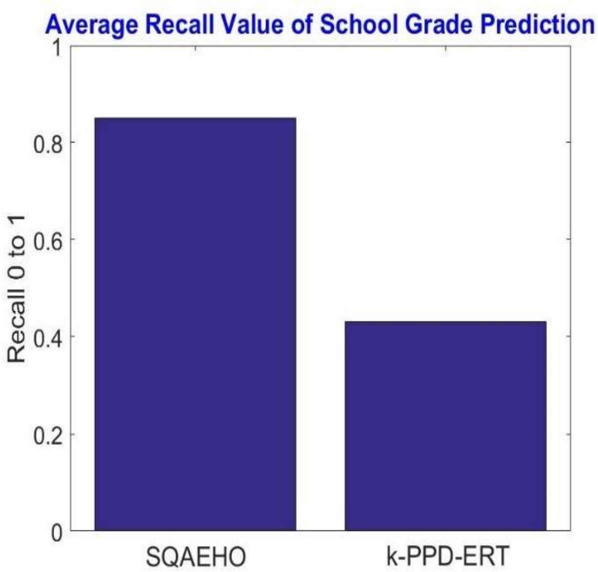

Fig. 3 Average recall value of school grades prediction.

Table 5. Recall value of school grades prediction models for all classes of different testing dataset.

Testing Dataset	SQAEHO	k-PPD-ERT
20	0.8	0.4706
40	0.833	0.4737
60	0.8667	0.4
80	0.8667	0.3953
100	0.8824	0.4167

Recall values of the school grades prediction was shown in table 5 and it was found that use of privacy preserving operation has not much effect the training dataset. Recall values were enhanced by the average 0.49 value, as compared to k-PPD-ERT.

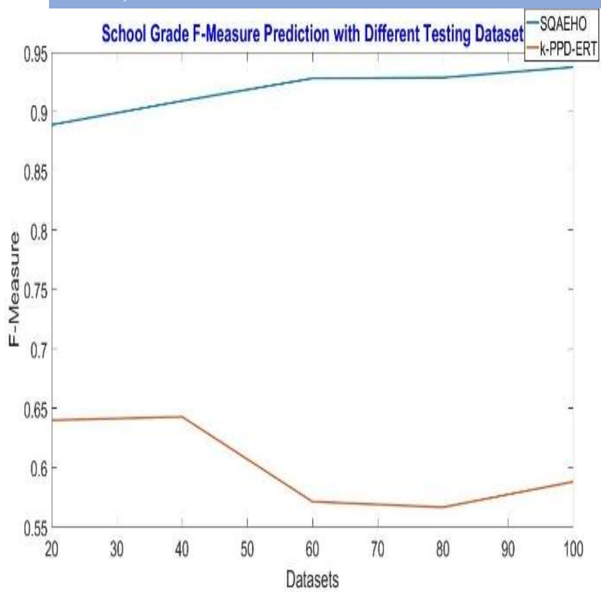

Fig.4 F-measure value of school grades prediction at different testing dataset percentage.

F-measure values shown in table 6, and it was found that use of elephant herd optimization algorithm has enhanced the work efficiency by removing unwanted features that reduces the learning.

Table 6. F-measure value of school grades prediction models for all classes of different testing dataset.

Testing Dataset	SQAEHO	k-PPD-ERT
20	0.8889	0.64
40	0.909	0.6429
60	0.928	0.5714
80	0.9286	0.5667
100	0.9375	0.5882

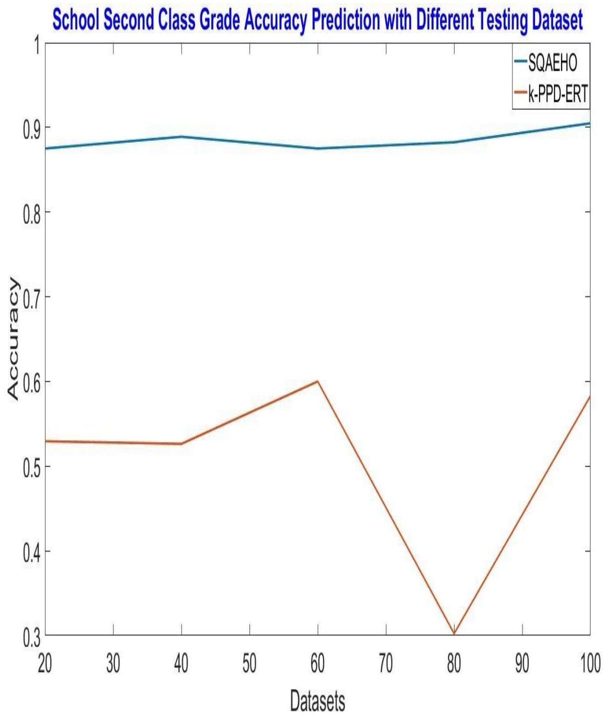

Fig.5 School first class grades prediction at different testing dataset percentage.

Table 7. Accuracy of First class grade prediction models for different testing dataset.

Testing Dataset	SQAEHO	k-PPD-ERT
20	0.8	0.2941
40	0.8	0.2632
60	0.909	0.275
80	0.9231	0.6047
100	0.9412	0.3125

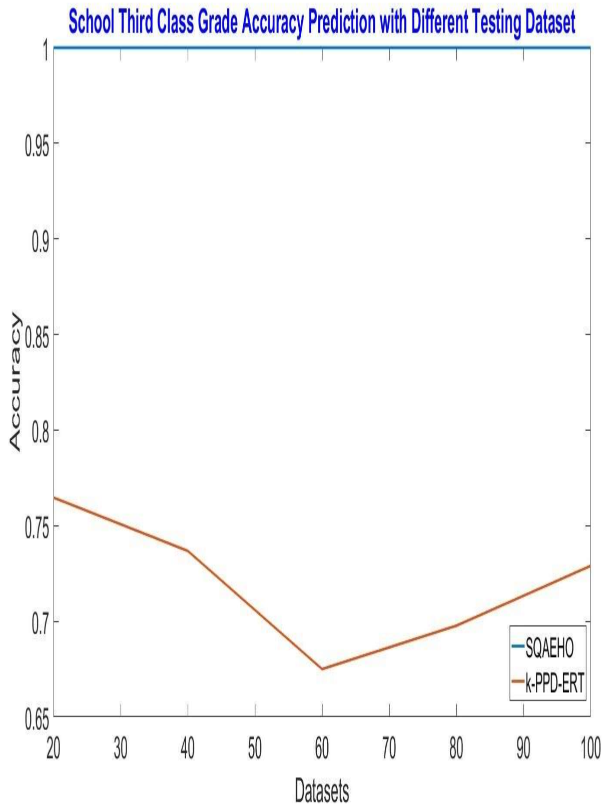

Fig.6 School second class grades prediction at different testing dataset percentage.

Table 8. Accuracy of Second class grade prediction models for different testing dataset.

Testing Dataset	SQAEHO	k-PPD-ERT
20	0.875	0.5294
40	0.8889	0.5263
60	0.875	0.6
80	0.8824	0.3023
100	0.9048	0.583

Table 7 and fig. 5 shows that proposed model has improved the work performance by 52.47%. It was found that by increase in dataset image from 20 to 100 detection accuracy was enhanced. Similarly table 8 and fig. 6 shows that proposed model has improved the second class grade prediction accuracy by 37.7%.

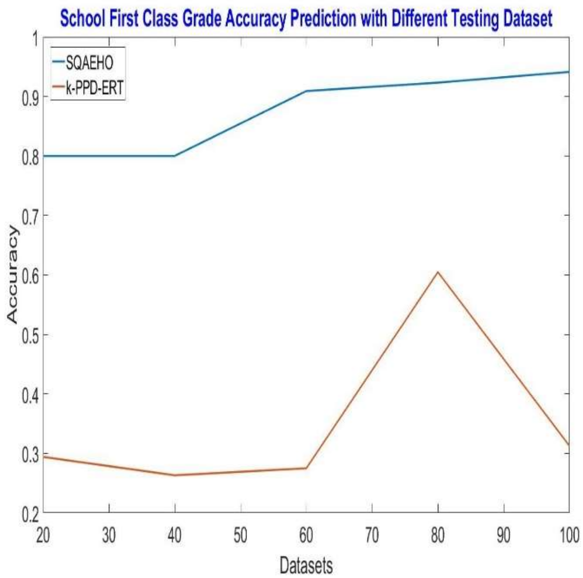

Fig.7 School third class grades prediction at different testing dataset percentage.

Table 9. Accuracy of Third class grade prediction models for different testing dataset.

Testing Dataset	SQAEHO	k-PPD-ERT
20	1	0.7647
40	1	0.7368
60	1	0.675
80	1	0.6977
100	1	0.729

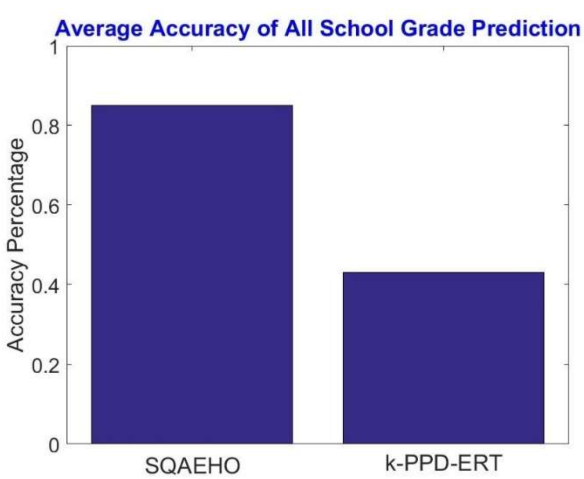

Fig. 8 Average accuracy of all set of grades for comparing school grades prediction models.

Table 7, 8 and 9 shows that proposed model SQAEHO has increased the work performance in all grades and for all set of testing dataset. School privacy operations has not affects the dataset features. Table 9 and fig. 7 shows that proposed model has improved the second class grade prediction accuracy by 27.93%.

VII.CONCLUSION

This paper has identified the features that help to grade a school performance. So selected features will be generalize for estimating the school performance. As school has public reputation and major part of school is students, so privacy of such sensitive information was done in this model applying suppression and super class substitution methods. In order to select features elephant herd optimization algorithm was implemented. Selected features were used for the training of error back propogation neural network. Experiment was perform on real data of 16 cities having 132 school. Result shows that proposed that average accuracy was enhanced by more than 50%. In future scholar can apply same model for college grade prediction.

REFERENCES

1. ABhatnagar et al. Improved governance of Indian school system through school consolidation J Pol Model (2019)
2. Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. Learning analytics (pp. 61–75). Springer.
3. Jose M. Cordero et al. The effect of teaching strategies on student achievement: an analysis using TALIS-PISA-link J Pol Model (2018).
4. E. Faham et al. Technological Forecasting & Social Change Using system dynamics to develop education for sustainable development in higher education with the emphasis on the sustainability competencies of students Technol Forecast Soc Change, 2017.
5. C.O. Henriques et al. Efficiency of secondary schools in Portugal: a novel DEA hybrid approach Soc Econ Plann Sci 2021.
6. Babić, I. D.. Machine learning methods in predicting the student academic motivation. Croatian Operational Research Review, 8(2), 2017, 443–461.
7. R. Paul, S. Gaftandzhieva, S. Kausar, S. Hussain, R. Doneva and A. K. Baruah, "Exploring student academic performance using data mining tools", Int. J. Emerg. Technol. Learn., vol. 15, no. 8, pp. 195-209, Apr. 2020.
8. V. Ramesh, P. Parkavi, K. Ramar, Predicting student performance: a statistical and data mining approach, International Journal of Computer Applications 63 (8) (2013) 35–39.
9. A. Bogar´ın, C. Romero, R. Cerezo, M. Sanchez-Santill ´ an, ´ Clustering for improving educational process mining, in: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, ACM, 2014, pp. 11–15.
10. C. Coffrin, L. Corrin, P. de Barba, G. Kennedy, Visualizing patterns of student engagement and performance in moocs, in: Proceedings of the fourth international conference on learning analytics and knowledge, ACM, 2014, pp. 83–92.
11. K. Bunkar, U. K. Singh, B. Pandya, R. Bunkar, Data mining: Prediction for performance improvement of graduate students using classification, in: Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on, IEEE, 2012, pp. 1–5.
12. S. D. A. Bujang, A. Selamat, R. Ibrahim, O. Krejcar, E. Herrera-Viedma, H. Fujita, et al., "Multiclass prediction model for student grade prediction using machine learning", IEEE Access, vol. 9, pp. 95608-95621, 2021.
13. P. M. Arsad, N. Buniyamin, J.-l. A. Manan, A neural network students' performance prediction model (nnsppm), in: Smart Instrumentation, Measurement and Applications (ICSIMA), 2013 IEEE International Conference on, IEEE, 2013, pp. 1–5
14. D. Sun et al., "A University Student Performance Prediction Model and Experiment Based on Multi-Feature Fusion and Attention Mechanism," in IEEE Access, vol. 11, pp. 112307-112319, 2023.
15. Chaka C. Educational data mining, student academic performance prediction, prediction methods, algorithms, and tools: an overview of reviews. 2023.
16. C.H. Chen, S.J. Yang, J.X. Weng, H. Ogata, C.Y. Su Predicting at-risk university students based on their e-book reading behaviours by using machine learning classifiers Aust. J. Educ. Technol., 37 (4) (2021), pp. 130-144.

17. S. D. A. Bujang et al., "Multiclass Prediction Model for Student Grade Prediction Using Machine Learning," in IEEE Access, vol. 9, pp. 95608-95621, 2021.
18. Y. Zhang, V. Y. Mariano and R. P. Bringula, "Prediction of Students' Grade by Combining Educational Knowledge Graph and Collaborative Filtering," in IEEE Access, vol. 12, pp. 68382-68392, 2024.
19. S. D. A. Bujang et al., "Multiclass Prediction Model for Student Grade Prediction Using Machine Learning," in IEEE Access, vol. 9, pp. 95608-95621, 2021.
20. S. D. Abdul Bujang et al., "Imbalanced Classification Methods for Student Grade Prediction: A Systematic Literature Review," in IEEE Access, vol. 11, pp. 1970-1989, 2023.
21. N. A. Butt, Z. Mahmood, K. Shakeel, S. Alfarhood, M. Safran and I. Ashraf, "Performance Prediction of Students in Higher Education Using Multi-Model Ensemble Approach," in IEEE Access, vol. 11, pp. 136091-136108, 2023
22. H. C. Tanuwidjaja, R. Choi, S. Baek and K. Kim, "Privacy-Preserving Deep Learning on Machine Learning as a Service—a Comprehensive Survey," in IEEE Access, vol. 8, pp. 167425-167447, 2020.
23. Li, J.; Guo, L.; Li, Y.; Liu, C. Enhancing Elephant Herding Optimization with Novel Individual Updating Strategies for Large-Scale Optimization Problems. Mathematics 2019, 7, 395. https://doi.org/10.3390/math7050395.
24. Syed Muqthadar Ali, N. Kumaran, G. N. Balaji, "A hybrid elephant herding optimization and harmony search algorithm for potential load balancing in cloud environments". 2022, International Journal of Modeling, Simulation, and Scientific Computing, 2250042, 13 05.
25. Wei Li, Gai-Ge Wang, Amir H. Alavi. "Learning-based elephant herding optimization algorithm for solving numerical optimization problems", Knowledge-Based Systems, Volume 195, 2020 105675,
26. Pankaj Kumar, Asst. Prof. Komal Kanojia. "Channel Estimation and BER Reduction Using Artificial Neural Network". IJSRET_V8_issue2, 2022.
27. A. Aminifar, M. Shokri, F. Rabbi, V. K. I. Pun and Y. Lamo, "Extremely Randomized Trees With Privacy Preservation for Distributed Structured Health Data," in IEEE Access, vol. 10, pp. 6010-6027, 2022, doi: 10.1109/ACCESS.2022.3141709