International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Statistical Investigation of Student Behaviour Analysis Models from An Empirical Perspective

Ashwini Raipure[1], Dr. Sarika Khandelwal[2]

Submitted: 20/12/2023 **Revised**: 31/01/2024 **Accepted**: 10/02/2024

Abstract: Student behaviour analysis is a multidisciplinary field which requires exploration of a wide variety of data, including, student's geographical profile, area of behavioural study, temporal responses, situational responses, analytical reasoning, attention profile, etc. Combination of these factors requires design of intelligent machine learning approaches, which work on temporal behavioural responses. For instance, to predict student's inclination towards technical education, models utilize analytical questionnaire, and social media tools to capture student's behaviour. This data is processed using various deep learning architectures to estimate student's inclination probability towards technical education. A wide variety of architectures are proposed for this task, and these architectures vary in terms of performance metrics, area of application, geography of student, etc. This makes it uncertain for researchers to test, validate &select most optimum models for their application, which increases cost & time needed for deployment. In order to reduce the uncertainty of model selection, this paper reviews some of the recently proposed methods for student behaviour analysis, and compares them in terms of performance metrics, area of application, and geographical parameters. The performance metrics include accuracy of analysis, computational complexity, mean squared error (MSE), and speed of analysis. This review will be helpful for researchers & behavioural analysis system designers to select the most optimum models for newer deployments, and will assist in performance upgradation of existing systems. Moreover, this text also recommends various improvements & enhancements in the reviewed models, which assists in upgrading their internal capabilities including scalability, flexibility, and performance analysis.

Keywords: *Behavioural analysis, geography, accuracy, complexity, student, technical, education*

1. Introduction

Analysis of student behaviour requires effective modelling of multiple machine learning components. These components include intelligent data collection, redundancy removal, feature extraction, feature selection, stratification, and post-processing. The accuracy of these models is largely dependent upon quality & quantity of data sources used for information gathering about the student. For instance, due to CoVID-19, students have shifted their mode of learning from offline to online. While the student is learning online, s/he is exposed to distractions like social media, gaming, etc. This causes learning disorders like attention deficiency, lack of retention, etc. Due to these disorders, student's inclination shifts from learning to non-productivity. Thus, it is very important to consider student's behaviour before exposing them to any new kind of learning mechanism. A typical student behaviour analysis model is observed from figure 1, wherein data from multiple sources is aggregated for pre-processing. These sources include, but are not limited to academic data, internet consumption data, lifestyle data, etc. The data is pre-processed using various models, which include but are not limited to, average filters, auto regressive integrated moving average (ARIMA), etc. The pre-processed data is given to feature extraction module, wherein features like frequency of use, temporal patterns, periodic patterns, etc. are extracted.

[1]*Ph. D. Research scholar, G.H.Raisoni Unversity Amaravati,*
[2]*Ph.D. guide Associate Professor in CSE Department, G H Raisoni College of Engineering Nagpur*
Author Emails
[a)] *ashwini.raipure@gmail.com*
[b)] *sarika.khandelwal@raisoni.net*

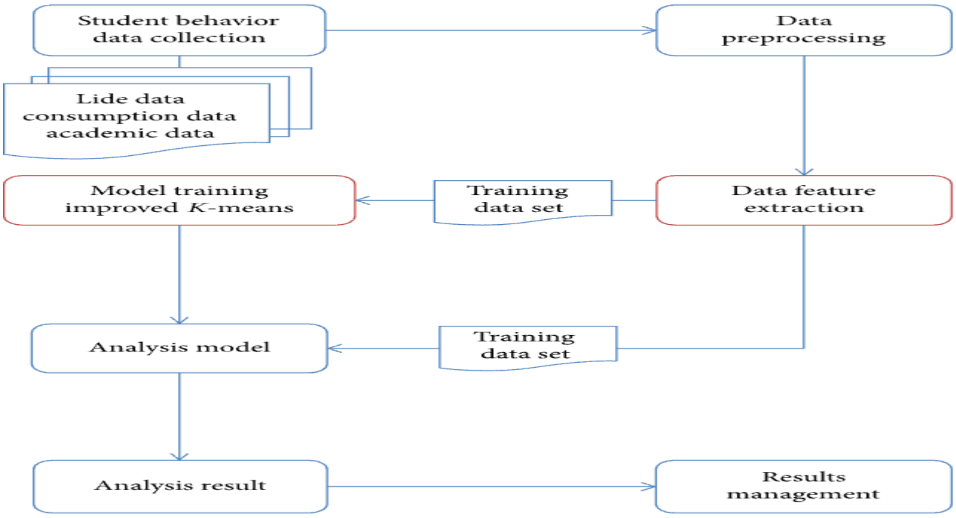

Fig 1. A typical student behaviour analysis model

These extracted features are given to a machine learning model, wherein the features are mapped with student's behaviour. The model assists in estimating final student behaviour w.r.t. the extracted features. A wide variety of models are proposed for this task, which include, convolutional neural network (CNN), recurrent neural network (RNN), long-short-term-memory (LSTM), gated recurrent unit (GRU), etc. Retrograde aggregation of outputs obtained from this trained analysis model results into temporal behavioural patterns. These patterns are processed using a post-processing layer that finally results into student behaviour estimation, and assists in decision making. A wide variety of algorithms are proposed for this task, and each of these algorithms vary in terms of implementation characteristics, input characteristics, behavioural pattern characteristics, etc. This variation causes ambiguity in model selection, which increases time & cost needed for design, testing & validation of student behaviour analysis systems. In order to reduce this ambiguity, the next section reviews these algorithms in terms of their characteristics, nuances, advantages, limitations & future scalability. This section also recommends various hybrid enhancements, which can be applied to these models to improve their internal performance. These reviewed algorithms are compared in terms of performance metrics including accuracy, analysis delay, computational complexity, area of application, and geography of application in section 3 of this text. Finally, this text concludes with some interesting observations about the reviewed models, and recommends various methods to further improve their performance.

2. Literature Review

Analysis of student behavioural patterns requires collection of a wide variety of domain specific data for training and validation purposes. Collection of this data is a primary task, which is facilitated using customized app-based solutions for social media, ecommerce, online learning, and other fields. Due to the recent CoVID pandemic, a large number of students have shifted their engagement to online learning, which has accelerated collection of these data samples. The work in [1] discusses design of such a system model wherein measurement of student engagement levels is performed using single modality, dual modality, and multimodality. The model utilizes facial features, key presses, and mouse movements, in order to estimate parameters like emotional state, click speed, typing speed, mouse speed, etc. These features are given to a Mini Xception Net model to evaluate student's engagement. The model is able to achieve an accuracy of 95.23%, 0.04 mean squared error (MSE) with moderate delay, but high computational complexity. The most optimum classification performance is obtained using Naïve Bayes (NB) model, when measures on the writing task. Other tasks including reading, multimedia viewing, etc. are also observed to have good accuracy using the NB model. This model finds its application in a wide variety of application specific scenarios, and can be used for their performance optimization. For instance, if NB model is used in [2] along with features like relevance between online instruction design & student learning, delivery quality, online support, student participation, and contingency modelling, then the system's overall performance can be improved. Currently, the work in [2] utilizes simpler machine learning models like random forests (RF) to achieve moderate accuracy, with high complexity of deployment. This model is further studied and extended in [3], wherein behavioural engagement (beh), cognitive capabilities (cog), social engagement (soc) and affective engagement (aff) features are used. This model is visualized in figure 2 bi-factor exploratory structure equation modelling framework (BESEM) is defined.

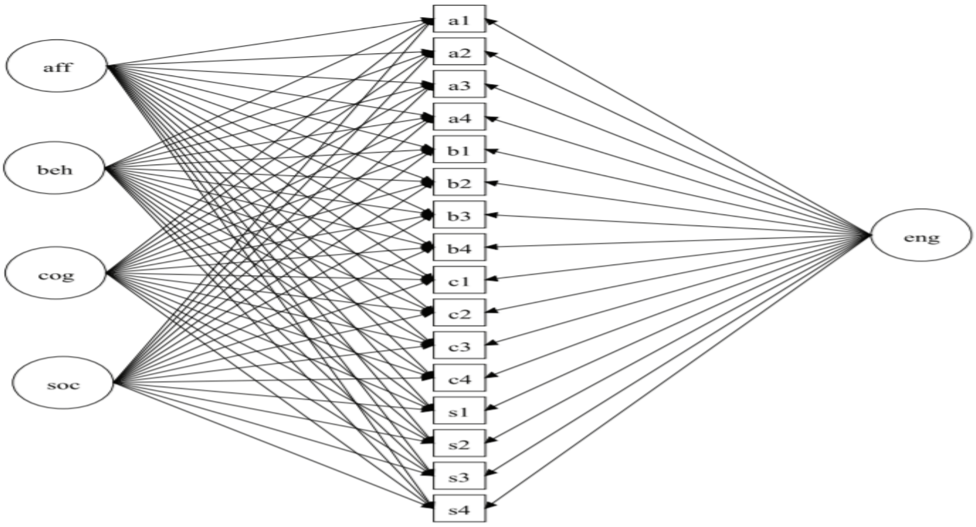

Fig 2. The BESEM framework [3]

This framework utilizes correlation between different features in order to estimate final engagement value. The model showcases 98.2% accuracy, with 0.05 mean squared error (MSE), and moderate delay. This performance is better than ICMCFA which showcases 96.8% accuracy, with 0.07 MSE with high delay, BCFA which showcases 96.4% accuracy, with 0.08 MSE with high delay, and ESEM which showcases 98.3% accuracy, with 0.09 MSE with very high delay. This performance is very high, and thus can be used for real time school-and-college applications. Another such model is described in [4], wherein features like general behaviour, social behaviour, and psychological behaviour are used as visualized in figure 3, along with k Means clustering to find patterns.

The model utilizes adaptive k Means (AKM) for clustering monthly usage, dining habits, work, rest, internet habits, exercise habits, class attentivity, and book borrowing habits. All this data assists in dividing students into 3 types, wherein schedules & eating habits are quantized. The model is observed to have an accuracy of 94%, with moderate error rates, and high delay of computation. This efficiency can be further quantized and explored in order to improve validation performance using extensive analysis as done in [5], wherein over 550 students are evaluated. The research proposes use of intent-to-treat (ITT), two-stage least squares (2SLS), and other models for analysis of student behaviour. ITT model is observed to have low accuracy, with moderate delay, while 2SLS model is observed to have moderate accuracy with high delay of analysis.

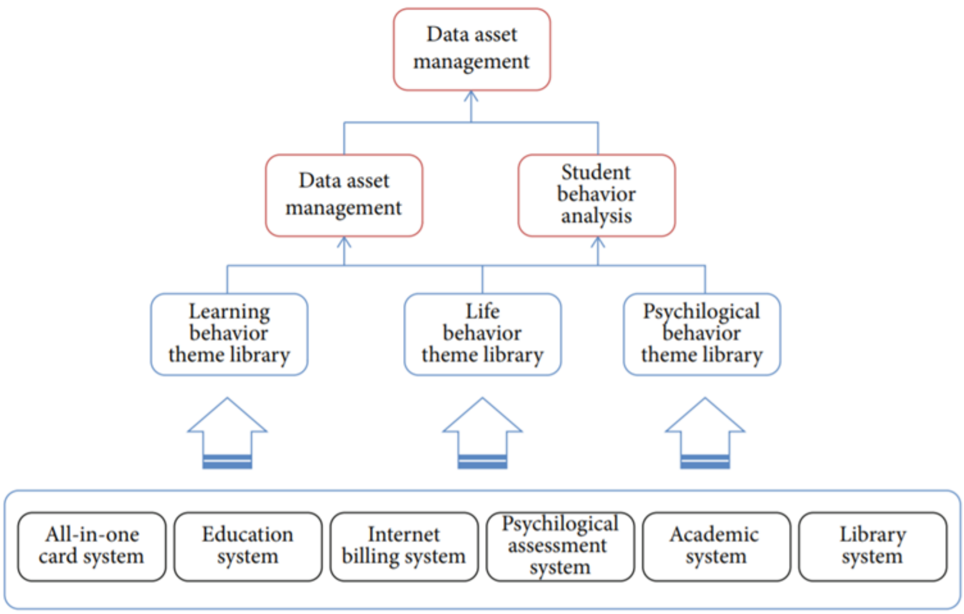

Fig 3. Learning, life & psychological behaviour analysis [4]

A list of different factors that affect student learning behaviour are evaluated in [6], along with the push–pullMooring Model. The model utilizes learning convenience, perceived security risk, service quality, ease of use, usefulness, task technology fit, teacher's teaching attitude, habits and switching cost for analysis of student behaviour. It can be visualized from figure 4, wherein student switching intention is related with different parameters.

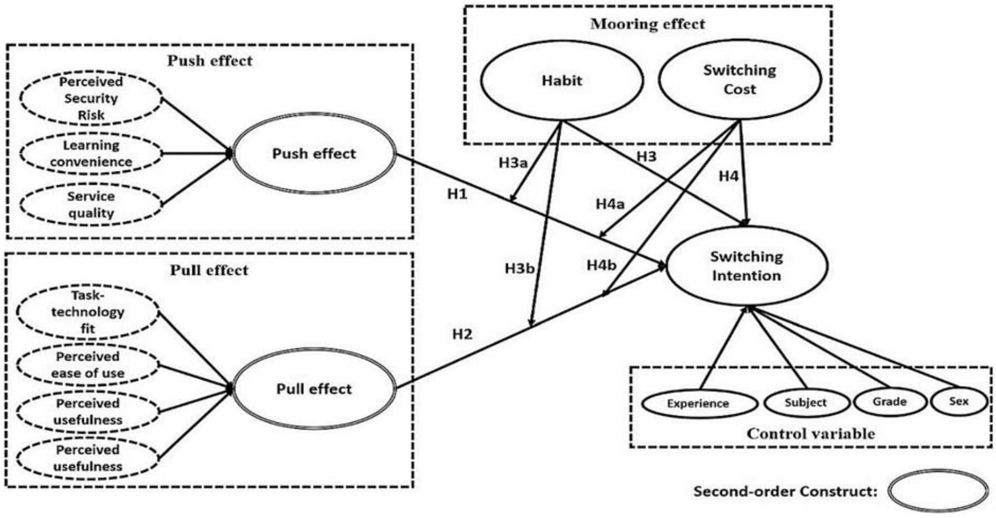

Fig 4.oore Model for student behaviour analysis [6]

The model is observed to have an accuracy of over 93%, with moderate delay and low error rate. An improvement to this model is proposed in [7], wherein student's intention towards usage of online learning platforms is observed using Technology Acceptance Model (TAM).

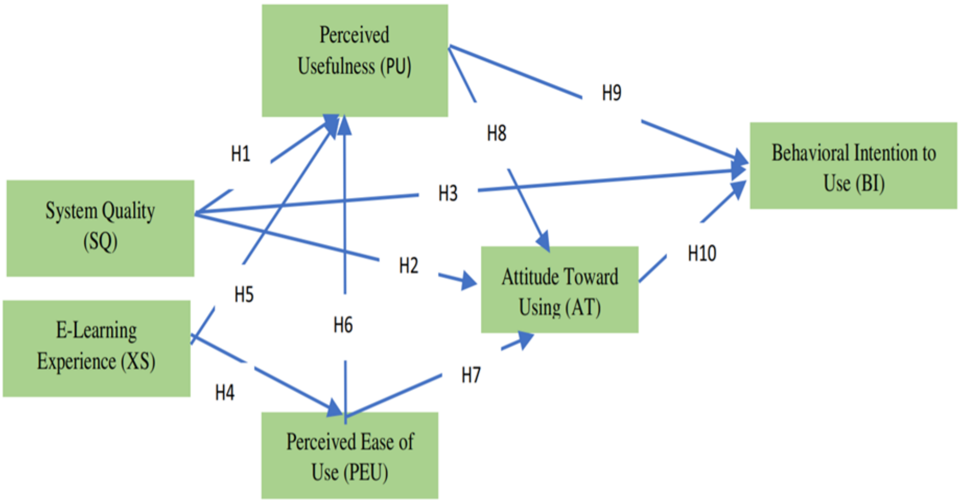

Fig 5. Method used for generation of TAM [7]

The model utilizes perceived usefulness, perceived ease of use, attitude towards utilization, behavioural intention to use, and actual use parameters for evaluation of student behaviour. The model utilizes 10 different relationship weights as observed from figure 5 in order to estimate student's inclination towards online education. The model is observed to have an accuracy of 91.5% for estimation of user behaviour, but requires a large amount of data for efficient analysis. But collection of this data requires low complexity, thereby extending usability of the model for a wide variety of scenarios including analysis of students with unique abilities. Such a model is proposed in [8], wherein future of subjects like Science, Engineering, Mathematics and Technology for Gifted & Talented students in Australia is discussed. The model is developed for studying behaviour of students in rural areas, and uses machine learning with local knowledge (LK). The model is able to achieve an accuracy of 90% with low MSE and low delay performance, thereby making it useful for a multiple scenario.

The models in [7], and [8] can be combined in order to design a highly efficient technique that covers majority of student categories for behaviour analysis. Such hybrid models and their performance are observed in [9], wherein students from different farming communities are studied, and their behaviour is analyzed. The study evaluates opportunities, aspirations, challenges, and barriers in order to estimate aspiration achievement gaps for rural students. Data from web of science (WoS), IFPRI library, MDPI, CAB abstracts, and other sources is analyzed in order to perform this task. It is observed that majority of rural student population wants to shift into towns, and opt for higher studies and better education opportunities. To provide such opportunities mobile learning platforms are needed, which can be achieved through smart phone-based technologies as suggested in [10], wherein recordings from rural & urban schools are combined in order to develop a highly efficient learning model. The proposed BISM model suggests use of focus group discussions, pre & post-test scenarios, fundamental analysis, etc. for estimation of student behaviour. The model is able to achieve an accuracy of 89% for different student categories, but has high delay and high MSE performance, which limits it usability. Study of parameters like geographical background, various social capital sources, gender, race, immigration status, meal availability, parental education, and high school rank are also some of the most useful parameters for social behavioural analysis as suggested in [11]. The model proposed in [11] uses these parameters to train a machine learning model (MLM) which is capable of achieving high accuracy, and low error rate, but requires large delay when compared with simpler models.

Similar models are proposed in [12, 13, 14], wherein multiple person behaviour analysis, classroom behaviour analysis, and learning pattern analysis are discussed. These models utilize data usage pattern, in-class behaviour patterns, and online behaviour analysis to develop different models that achieve moderate accuracy. The MUFIC (Multiuser fitness coach) model in [12] is observed to achieve an accuracy of 85%, with moderate error and high delay, while the Online Hard Example Mining model with recurrent convolutional neural network (OHEM RCNN) is observed to have an accuracy of 94% with very high delay, & low error, and the Felder and Silverman learning style model (FSLSM) in [14] is observed to have an accuracy of 85.71% when using decision tree classifier (DT), and 85.95% when using gradient boosted tree (GBT) classifier. Both these models have moderate error, and high delay due to larger training data requirement. These models must be combined in order to cover larger range of scenarios, and obtain lower MSE with faster response time. Thereby improving their applicability to a large number of scenarios. An application of these models is proposed in [15], wherein group presentations are used for analysis of student behaviour. The model utilizes body movements, body pose, expression of face, pauses taken during presentation, eye contact, pace of speech, post of the body, etc. in order to evaluate student behaviour. Due to these parameters, the model is observed to have an accuracy of over 83%, with moderate delay, and high MSE due to large variations in pose & other body parameters.

The work in [16] proposes design of an operating system (OS) for behavioural analysis (BAOS), the OS model utilizes a wide variety of input parameters including login duration, log sizes, file open counts, etc. in order to perform this task. It was tested on over 850 students, and was observed to achieve an accuracy of 75%, which is mainly due to the variety of data used for analysis. This efficiency can be improved using customized models, like the one proposed in [17], wherein online learning sets were used for clustering. The model uses self-organizing maps (SOMs) along with neural networks (SOMNNs) for performing this task. Over 1.7 million records were analyzed for this purpose, and parameters like grades, examination performance, continuous evaluation, etc. were used. Metrics including number of homework submitted, resources created, posts created, pages read, etc. were evaluated to obtain an accuracy of 93.61% which makes the system usable for real time scenarios. The model is highly complex, thus requires large processing delays, but has low MSE due to multiple parametric selection.

Another methodology that uses life cycle assessment (LCA) is proposed in [18], wherein Sakai learning management system (LMS) was taken as a use case. The model showcases 93% accuracy, and uses resource utilization, lesson evaluation, tests, polls, assignment, and other data for analysis. These data metrics are given to a Gradient Boosted Decision Tree (GBDT) model for analysis and classification, due to which moderate delay, and low MSE values were obtained. Similar models are proposed in [19, 20, 21], wherein student behaviour during current CoVID pandemic is evaluated using k Means, MFR model (model time, frequency, and recent activity), TeSLA (Adaptive Trust-based e-Assessment System for Learning) are discussed. These models have proven to be incrementally effective in estimation of student behaviour, and thus can be used for application specific scenarios. The k Means is supposed to have an accuracy of 66.5%, MFR 72%, and TeSLA 89.2% on different applications. These models must be combined in order to achieve good accuracy across different applications, and make them more generic in nature. Accuracy of these models would be improved if machine learning techniques like enhanced extended nearest neighbour (EENN) [22], Structured Equation Modelling (SEM) [23], and association rule mining using apriori [24] are integrated into the system. The EENN model is observed to have low complexity, moderate MSE and an accuracy of 91%, while SEM has moderate level of complexity, low MSE, and an accuracy of 92.5%, which makes them suitable for optimization of existing systems. The apriori model is observed to have an accuracy of 84.75%, but is context-sensitive, and can be used depending upon the type of dataset being processed.

Student behaviour can be predicted with better efficiency if student feedback is considered while analysing it. The work in [25] proposes such a model, wherein an adaptive feedback system is designed for collaborative behaviour analysis. The model uses distance learning for performance tracking, behaviour analysis, engagement analysis and suggestive analysis, which enables the model to achieve an accuracy of over 83% on multiple datasets. This accuracy can be further improved using the work in [26, 27, 28], wherein LMS based models (LMSM), deep knowledge tracing with multiple feature fusion attention mechanism (DKTMFAM), and blended learning (BT) methods are described. The LMSM method utilizes online behavioural patterns including connection distribution, average time per lecture, average number of sessions, etc. to achieve an accuracy of 91%, with moderate delay and moderate MSE performance. While the DKTMFAM model uses recurrent neural network (RNN) as observed in figure 6 in order to process parameters like skill, response time, number of practice sets, first action type, etc. The DKTMFAM model achieves an accuracy of 98% for various scenarios, and thus finds its utility in multiple student behaviour analysis scenarios. It has a MSE of 0.2, which is higher than some other models, and requires large delay for training and validation due to use of long-short-term-memory (LSTM), and other RNN components. In contrast, the BT model utilizes study duration, access time, number of posts, etc. to train a GBDT and achieves an accuracy of 97.4%, with moderate delay and moderate MSE performance. Thus, the models in [27] and [28] can be fused in order to design a highly effective student behaviour analysis system.

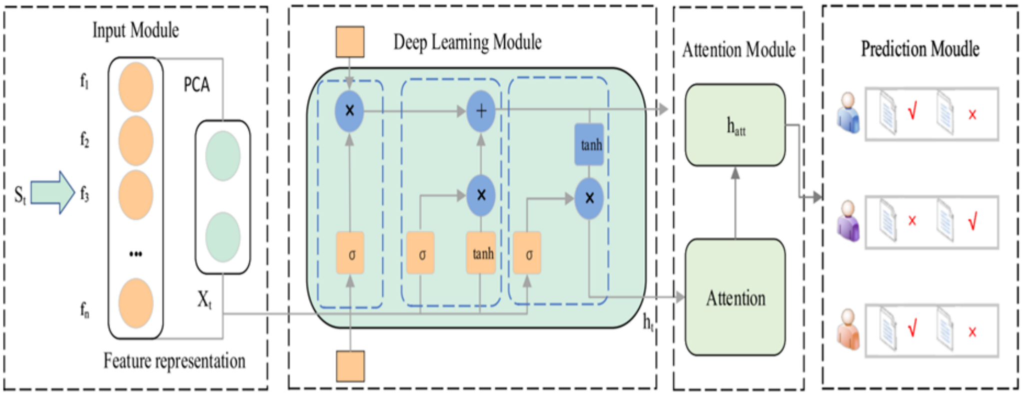

Fig 6. The DKTMFAM model [27]

Other models like preference cognitive diagnosis method (PrefCD) [29], Firefly Grey Wolf-Assisted Nearest Neighbour (FGWANN) [30], self-regulation models (SRM) [31], and DBSCAN (density-based spatial clustering of applications with noise) with k Means (DBkMeans) [32] are also discussed. These models are essentially extensions of previously discussed models, and are capable for obtaining high accuracy, with moderate delay, and moderate MSE performance. The PrefCD model has an accuracy of 76.14%, with MSE of 0.23, and moderate delay, while FGWANN has an accuracy of 96.3%, which MSE of 0.09, and moderate delay. Similarly, SRM has an accuracy of 96%, with MSE of 0.15, and DBkMeans that utilizes food habits along with study patterns, has an accuracy of 91.5%, which makes them useful for real time scenarios. Some models also assist in improving student's behaviour during certain scenarios, for instance, the work in [33] utilizes mobile health, temporal parameters & geographical features to recommend health-based suggestions to students, and improve their real time online learning performance.

Models like cellular automata (CA) [34], cyber engagement (CE) for bullying estimation [35], predictive game theory model (PGTM) for programming students [36], and profile-based cluster evolution analysis (PBCEA) [37] are value adds to behaviour analysis system models because they allow for processing incremental inputs. These incremental inputs include student depression levels [35], programming capabilities [36], and migration patterns [37]. The CA model is capable to achieve an accuracy of 79%, while CE has an accuracy of 85%, PGTM has an accuracy 89%, while PBCEA has an accuracy of 85% on different datasets. These models must be combined, and then applied to deep learning networks for design of a comprehensive behavioural analysis model. Models that can further assist in incrementally improve student behaviours are also discussed in [38, 39, 40], wherein disengaged behaviour, app-specific context-based behaviours, model thinking methods are proposed. These methods are observed to have moderate accuracy, with low MSE and moderate delay when applied to application specific scenarios. This performance can be further enhanced via use of deep learning methods like CNNs, RNNs, and LSTMs. Thus, it is observed that multiple methods are available for analysis of student behaviour, and deep learning-based methods are the most effective ones for this task. The next section compares these models, and recommends their utility for effective behaviour analysis.

3. Statistical Analysis of Various Behaviour Analysis Models

From the review it is observed that various models are surveyed, and their performance in terms of accuracy, delay, and error rates is discussed. In this section, the performance of these models is compared in terms of the discussed parameters, which will allow readers to select the best models for their behaviour analysis application. This performance comparison is tabulated in table 1, wherein various models and their parameters are aggregated. The values of delay and MSE are dependent on the execution environment, thus they are quantized into low (L), medium (M), high (H), and very high (VH) ranges, which assists in comparing them on the same test environment.

Table 1. Performance evaluation of different models

Model	Accuracy (%)	Delay	MSE
Mini Xception Net [1]	95.23	VH	L
NB [2]	95.7	M	L
BESEM [3]	98.2	VH	M
ICMCFA [3]	96.8	VH	L
BCFA [3]	96.4	H	M
ESEM [3]	98.3	H	M
AKM [4]	94	M	H
ITT [5]	75	L	H
2SLS [5]	79	M	M
Push Pull [6]	93	M	L
TAM [7]	91.5	H	L
LK [8]	90	L	L
BISM [10]	89	H	H
MLM [11]	91	H	M
MUFIC [12]	85	H	H
OHEM RCNN [13]	94	H	L
FSLSM DT [14]	85.7	H	M
FSLSM GBT [14]	85.95	H	M
Pose body [15]	83	M	H
BAOS [16]	75	L	M
SOMNN [17]	93.61	H	L
LCA [18]	93	M	L
k Means [19]	66.5	M	M
MFR [20]	72	L	M
TeSLA [21]	89.2	M	L
EENN [22]	91	M	L
SEM [23]	92.5	M	M
Apriori [24]	84.75	H	M
Adaptive feedback [25]	83	H	M
LMSM [26]	91	M	M
DKTMFAM [27]	98	VH	L
BT [28]	97.4	H	L
PrefCD [29]	76.14	M	H
FGWANN [30]	96.3	M	M
SRM [31]	96	H	M
DBkMeans [32]	91.5	M	H
CA [34]	79	M	M
CE [35]	85	H	H
PGTM [36]	89	M	H
PBCEA [37]	85	M	L

Based on this evaluation, the accuracy of these models was compared, and it is observed that ESEM [3], BESEM [3], DKTMFAM [27], BT [28], ICMCFA [3], BCFA [3], FGWANN [30], SRM [31], NB [2], and Mini Xception Net [1] have better accuracy when compared with other models. Thus, they must be used while performing student behaviour analysis for a wide variety of scenarios. This performance can be observed from figure 7, wherein different models and their absolute accuracy levels are compared.

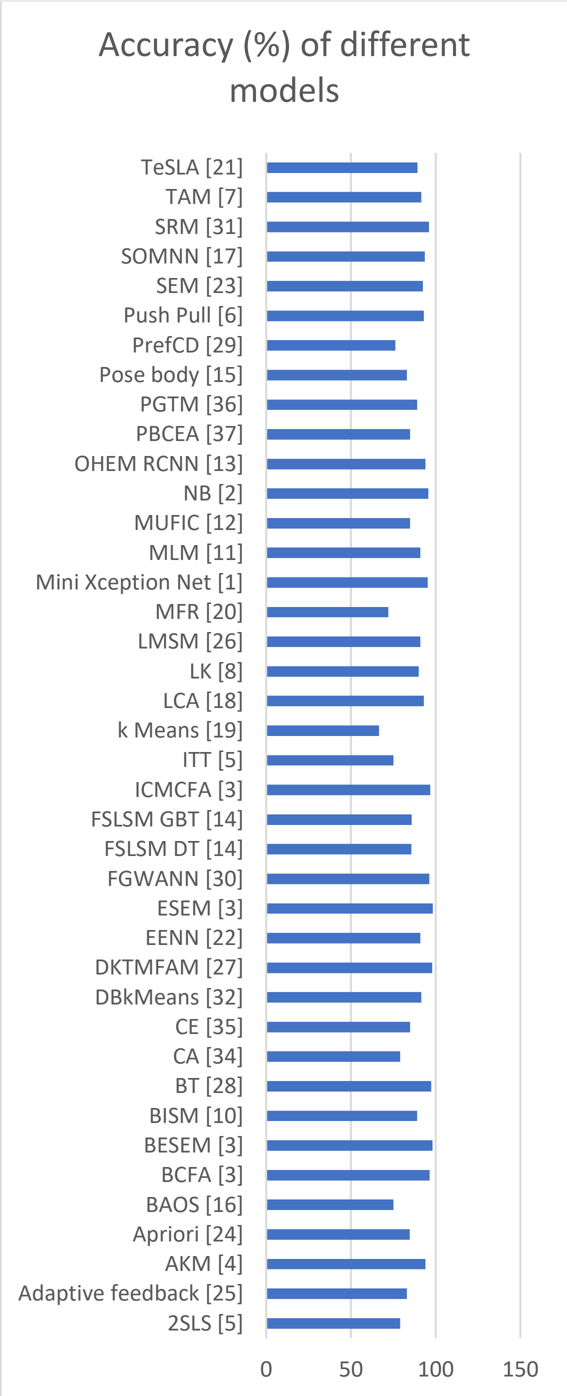

Fig 7. Accuracy of different models

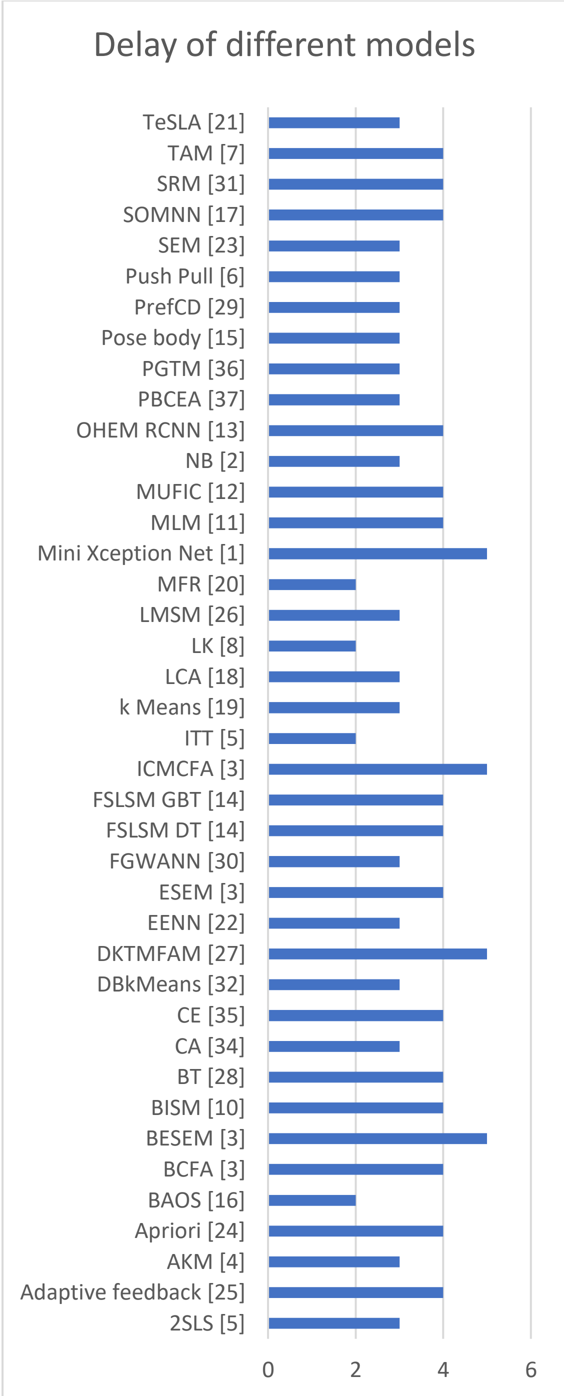

Fig 8. Delay performance of different models

Similarly, the delay performance is visualized from figure 8, wherein it can be observed that LK [8], ITT [5], BAOS [16], MFR [20], FGWANN [30], NB [2], AKM [4], Push Pull [6] and LCA [18] have the lowest delay. These models have faster performance, but might have lower accuracy and higher error.

The MSE performance these models are visualized from figure 9, wherein it is observed thatLK [8], NB [2], Push Pull [6], LCA [18], EENN [22], TeSLA [21], PBCEA [37], BT [28], OHEM RCNN [13], SOMNN [17], TAM [7], DKTMFAM [27], ICMCFA [3], and Mini Xception Net [1] have minimum value of error.

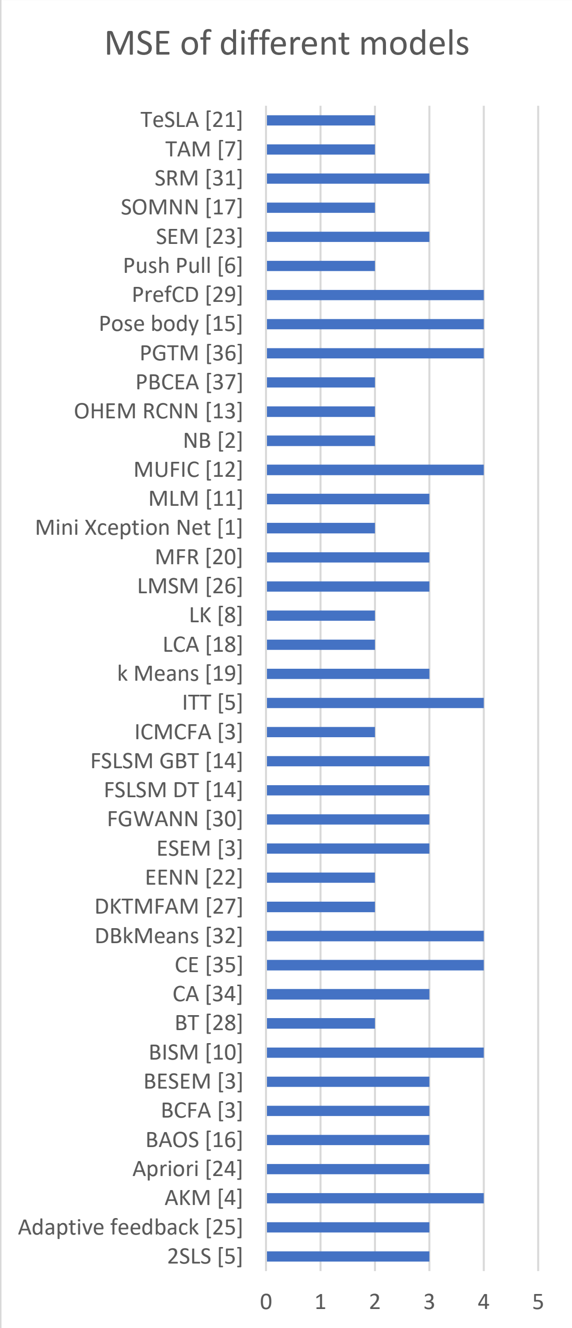

Fig 9. MSE performance of different models

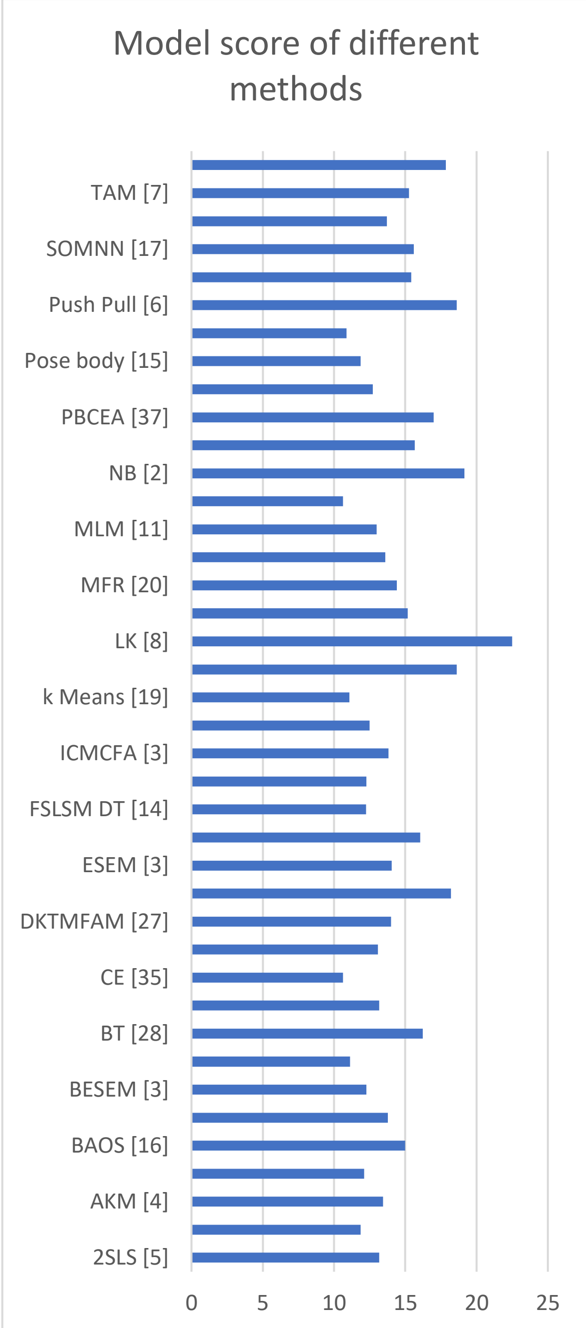

Fig 10. Overall model score for all methods

This error performance must be correlated with delay & accuracy performance in order to estimate the best performing model. To perform this task, equation 1 was used for evaluation of algorithm score, wherein values of accuracy, delay and MSE are combined for a correlative evaluation of the reviewed models. The results of this evaluation can be observed from figure 10, wherein final model score is visualized.

$$A_{Score} = \frac{Accuracy}{MSE + Delay} \ldots (1)$$

As all values of MSE and delay are quantized to the same level, thus this evaluation provides algorithms with low delay, low MSE, and high accuracy of student behaviour prediction.

From this evaluation it can be observed that, LK [8], NB [2], Push Pull [6], LCA [18], EENN [22], TeSLA [21], PBCEA [37], BT [28], FGWANN [30], OHEM RCNN [13], SOMNN [17], SEM [23], TAM [7], LMSM [26], and BAOS [16] have better performance than other models, and thus can be used for real time deployment of student behaviour analysis methods.

4. Conclusion & Future Scope

In this extensive review, various methods are compared based on the accuracy of student behaviour analysis, delay needed to perform that analysis, and error evaluated during the analysis. The value of error is not related with accuracy, because it indicates the total deviation from real world scenario for the given algorithm, when compared with a standard evaluation model. From these observations, it can be concluded that ESEM [3], NB [2], BESEM [3], DKTMFAM [27], BT [28], and ICMCFA [3] have highest accuracy, but LK [8], ITT [5], BAOS [16], MFR [20], FGWANN [30], NB [2], and AKM [4] are the fastest, moreover, LK [8], NB [2], Push Pull [6], LCA [18], EENN [22], and TeSLA [21] have minimum MSE performance. When these metrics are combined to form an algorithmic model fitness score, then it is observed that, LK [8], NB [2], Push Pull [6], LCA [18], EENN [22], TeSLA [21], PBCEA [37], and BT [28] are the most recommended models for any kind of student behaviour analysis system. Thus, these models and their combinations must be used during design of any behaviour analysis system for students or their community at large. In future, it is recommended that these models should be combined, and their performance must be evaluated on larger datasets for real time validation& applicability.

References

[1] Altuwairqi, K., Jarraya, S.K., Allinjawi, A. *et al.* Student behavior analysis to measure engagement levels in online learning environments. *SIViP* **15,** 1387–1395 (2021). https://doi.org/10.1007/s11760-021-01869-7

[2] Bao, W. COVID-19 and online teaching in higher education: A case study of Peking University. *Hum Behav&Emerg Tech*. 2020; 2: 113–115. https://doi.org/10.1002/hbe2.191

[3] Hoi, VN, Le Hang, H. The structure of student engagement in online learning: A bi-factor exploratory structural equation modelling approach. *J Comput Assist Learn*. 2021; 37: 1141– 1153. https://doi.org/10.1111/jcal.12551

[4] Xiaoying Shen, Chao Yuan, "A College Student Behavior Analysis and Management Method Based on Machine Learning Technology", *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 3126347, 10 pages, 2021. https://doi.org/10.1155/2021/3126347

[5] Hellings, J., Haelermans, C. The effect of providing learning analytics on student behaviour and performance in programming: a randomised controlled experiment. *High Educ* (2020). https://doi.org/10.1007/s10734-020-00560-z

[6] Lin, CL., Jin, Y.Q., Zhao, Q. *et al.* Factors Influence Students' Switching Behavior to Online Learning under COVID-19 Pandemic: A Push–Pull–Mooring Model Perspective. *Asia-Pacific Edu Res* **30,** 229–245 (2021). https://doi.org/10.1007/s40299-021-00570-0

[7] Mailizar, M., Burg, D. &Maulina, S. Examining university students' behavioural intention to use e-learning during the COVID-19 pandemic: An extended TAM model. *Educ Inf Technol* (2021). https://doi.org/10.1007/s10639-021-10557-5

[8] Morris, J., Slater, E., Fitzgerald, M.T. *et al.* Using Local Rural Knowledge to Enhance STEM Learning for Gifted and Talented Students in Australia. *Res Sci Educ* **51,** 61–79 (2021). https://doi.org/10.1007/s11165-019-9823-2

[9] Nandi, R., Nedumaran, S. Understanding the Aspirations of Farming Communities in Developing Countries: A Systematic Review of the Literature. *Eur J Dev Res* **33,** 809–832 (2021). https://doi.org/10.1057/s41287-021-00413-0

[10] Khan, Sharifullah& Hwang, Gwo-Jen & Abbas, Muhammad & Rehman, Arshia. (2018). Mitigating the urban–rural educational gap in developing countries through mobile technology-supported learning. British Journal of Educational Technology. 50. 10.1111/bjet.12692.

[11] Nelson, I.A. (2016), Rural Students' Social Capital in the College Search and Application Process. Rural Sociology, 81: 249-281. https://doi.org/10.1111/ruso.12095

[12] L. Wenyuan, W. Siyang and W. Lin, "A MultipersonBehavior Feature Generation Model Based on Noise Reduction Using WiFi," in IEEE Transactions on Industrial Electronics, vol. 67, no. 6, pp. 5179-5186, June 2020, doi: 10.1109/TIE.2019.2927179.

[13] R. Zheng, F. Jiang and R. Shen, "Intelligent Student Behavior Analysis System for Real Classrooms," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 9244-9248, doi: 10.1109/ICASSP40776.2020.9053457.

[14] Y. Ikawati, M. U. H. Al Rasyid and I. Winarno, "Student Behavior Analysis to Detect Learning Styles in Moodle Learning Management System," 2020 International Electronics Symposium (IES), 2020, pp. 501-506, doi: 10.1109/IES50839.2020.9231567.

[15] A. Fekry, G. Dafoulas and M. Ismail, "Automatic detection for students behaviors in a group presentation," 2019 14th International Conference on Computer Engineering and Systems (ICCES), 2019, pp. 11-15, doi: 10.1109/ICCES48960.2019.9068128.

[16] L. Wang, Z. Zhen, T. Wo, B. Jiang, H. Sun and X. Long, "A Scalable Operating System Experiment Platform Supporting Learning Behavior Analysis," in IEEE Transactions on Education, vol. 63, no. 3, pp. 232-239, Aug. 2020, doi: 10.1109/TE.2020.2975556.

[17] S. Delgado, F. Morán, J. C. S. José and D. Burgos, "Analysis of Students' Behavior Through User

Clustering in Online Learning Settings, Based on Self Organizing Maps Neural Networks," in IEEE Access, vol. 9, pp. 132592-132608, 2021, doi: 10.1109/ACCESS.2021.3115024.
[18] H. Wan, Q. Yu, J. Ding and K. Liu, "Students' behavior analysis under the Sakai LMS," 2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE), 2017, pp. 250-255, doi: 10.1109/TALE.2017.8252342.
[19] J. Deng, "College Students’ Online Behavior Analysis During Epidemic Prevention and Control," 2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 2021, pp. 796-799, doi: 10.1109/ICMTMA52658.2021.00183.
[20] P. Sun, Z. Wu, Y. Zhang and Y. Yang, "Analysis and Modeling of Learning Behaviors on Intelligent Tutoring Website," 2014 Tenth International Conference on Computational Intelligence and Security, 2014, pp. 729-731, doi: 10.1109/CIS.2014.170.
[21] M. E. Rodríguez, A. -E. Guerrero-Roldán, D. Baneres and I. Noguera, "Students’ Perceptions of and Behaviors Toward Cheating in Online Education," in IEEE RevistaIberoamericana de Tecnologias del Aprendizaje, vol. 16, no. 2, pp. 134-142, May 2021, doi: 10.1109/RITA.2021.3089925.
[22] S. Lai, B. Sun, F. Wu and R. Xiao, "Automatic Personality Identification Using Students’ Online Learning Behavior," in IEEE Transactions on Learning Technologies, vol. 13, no. 1, pp. 26-37, 1 Jan.-March 2020, doi: 10.1109/TLT.2019.2924223.
[23] F. B. Shaikh, M. Rehman, A. Amin, A. Shamim and M. A. Hashmani, "Cyberbullying Behaviour: A Study of Undergraduate University Students," in IEEE Access, vol. 9, pp. 92715-92734, 2021, doi: 10.1109/ACCESS.2021.3086679.
[24] D. Aviano, B. L. Putro, E. P. Nugroho and H. Siregar, "Behavioral tracking analysis on learning management system with apriori association rules algorithm," 2017 3rd International Conference on Science in Information Technology (ICSITech), 2017, pp. 372-377, doi: 10.1109/ICSITech.2017.8257141.
[25] M. Awais Hassan, U. Habiba, H. Khalid, M. Shoaib and S. Arshad, "An Adaptive Feedback System to Improve Student Performance Based on Collaborative Behavior," in IEEE Access, vol. 7, pp. 107171-107178, 2019, doi: 10.1109/ACCESS.2019.2931565.
[26] A. M. Villa, C. P. Molina and M. Castro, "Students’ Behavior When Connecting to the LMS: A Case Study at UNED," in IEEE RevistaIberoamericana de Tecnologias del Aprendizaje, vol. 14, no. 3, pp. 87-94, Aug. 2019, doi: 10.1109/RITA.2019.2942253.
[27] D. Liu, Y. Zhang, J. Zhang, Q. Li, C. Zhang and Y. Yin, "Multiple Features Fusion Attention Mechanism Enhanced Deep Knowledge Tracing for Student Performance Prediction," in IEEE Access, vol. 8, pp. 194894-194903, 2020, doi: 10.1109/ACCESS.2020.3033200.
[28] Z. Xu, H. Yuan and Q. Liu, "Student Performance Prediction Based on Blended Learning," in IEEE Transactions on Education, vol. 64, no. 1, pp. 66-73, Feb. 2021, doi: 10.1109/TE.2020.3008751.
[29] P. Jiang and X. Wang, "Preference Cognitive Diagnosis for Student Performance Prediction," in IEEE Access, vol. 8, pp. 219775-219787, 2020, doi: 10.1109/ACCESS.2020.3042775.
[30] H. Tang et al., "Predicting Green Consumption Behaviors of Students Using Efficient Firefly Grey Wolf-Assisted K-Nearest Neighbor Classifiers," in IEEE Access, vol. 8, pp. 35546-35562, 2020, doi: 10.1109/ACCESS.2020.2973763.
[31] R. A. Rasheed, A. Kamsin and N. A. Abdullah, "An Approach for Scaffolding Students Peer-Learning Self-Regulation Strategy in the Online Component of Blended Learning," in IEEE Access, vol. 9, pp. 30721-30738, 2021, doi: 10.1109/ACCESS.2021.3059916.
[32] X. Li, Y. Zhang, H. Cheng, F. Zhou and B. Yin, "An Unsupervised Ensemble Clustering Approach for the Analysis of Student Behavioral Patterns," in IEEE Access, vol. 9, pp. 7076-7091, 2021, doi: 10.1109/ACCESS.2021.3049157.
[33] S. Vhaduri, S. V. Dibbo and Y. Kim, "Deriving College Students’ Phone Call Patterns to Improve Student Life," in IEEE Access, vol. 9, pp. 96453-96465, 2021, doi: 10.1109/ACCESS.2021.3093493.
[34] P. Song, Y. Gao, Y. Xue, J. Jia, W. Luo and W. Li, "Human BehaviorModeling for Evacuation From Classroom Using Cellular Automata," in IEEE Access, vol. 7, pp. 98694-98701, 2019, doi: 10.1109/ACCESS.2019.2930251.
[35] W. M. Al-Rahmi, N. Yahaya, M. M. Alamri, N. A. Aljarboa, Y. B. Kamin and F. A. Moafa, "A Model of Factors Affecting Cyber Bullying Behaviors Among University Students," in IEEE Access, vol. 7, pp. 2978-2985, 2019, doi: 10.1109/ACCESS.2018.2881292.
[36] F. D. Pereira et al., "Explaining Individual and Collective Programming Students’ Behavior by Interpreting a Black-Box Predictive Model," in IEEE Access, vol. 9, pp. 117097-117119, 2021, doi: 10.1109/ACCESS.2021.3105956.
[37] S. A. Priyambada, M. Er, B. N. Yahya and T. Usagawa, "Profile-Based Cluster Evolution Analysis: Identification of Migration Patterns for Understanding Student Learning Behavior," in IEEE Access, vol. 9,

pp. 101718-101728, 2021, doi: 10.1109/ACCESS.2021.3095958.

[38] S. Nam, G. Frishkoff and K. Collins-Thompson, "Predicting Students Disengaged Behaviors in an Online Meaning-Generation Task," in IEEE Transactions on Learning Technologies, vol. 11, no. 3, pp. 362-375, 1 July-Sept. 2018, doi: 10.1109/TLT.2017.2720738.

[39] W. -Y. Hwang, I. Q. Utami, S. W. D. Purba and H. S. L. Chen, "Effect of Ubiquitous Fraction App on Mathematics Learning Achievements and Learning Behaviors of Taiwanese Students in Authentic Contexts," in IEEE Transactions on Learning Technologies, vol. 13, no. 3, pp. 530-539, 1 July-Sept. 2020, doi: 10.1109/TLT.2019.2930045.

[40] R. Lavi, Y. J. Dori, N. Wengrowicz and D. Dori, "Model-Based Systems Thinking: Assessing Engineering Student Teams," in IEEE Transactions on Education, vol. 63, no. 1, pp. 39-47, Feb. 2020, doi: 10.1109/TE.2019.2948807.