IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

An Intelligent Physical Education Assessment System with Explainable AI Integration

[1]BATTULA TEJA,
Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions
[2]B. SANDHYA KUMARI,
Miracle Educational Society Group of Institutions
[1]tejubattula11@gmail.com

ABSTRACT:

In this wrok, we propose a new system for assessing physical education (PE) performance using advanced analytics and big data. The system improves the accuracy and fairness of student assessments by integrating Back Propagation Neural Networks (BPNN), Genetic Algorithms (GA), and Decision Tree models. Scalability with data and efficient processing with Hadoop and Spark frameworks as well as precision and explainability with CATBOOST and SHAP models are maintained. The system automates the evaluation process, reduces the subjectivity of traditional methods, and provides real-time feedback with a 97% accuracy rate. The model overcomes the challenges posed by manual assessment methods and provides academic institutions the opportunity to foster data-driven insight and work towards well-rounded student development.

Keywords: physical education, GA, CATBOOST

INTRODUCTION

The holistic development of students, both mental and physical, draws from the multi-faceted approach of physical education (PE) in the curriculum. Evaluating the performance of PE remains a challenge valuing the integration of multi-dimensional data due to the outdated methods of evaluation and largely subjective grading. PE performance evaluation lacks evaluation stem from traditional collaboration, strategic thinking, and mental fortitude. In the current education landscape, the integration of big data to transform the assessment of PE performance. Leveraging the frameworks of Hadoop, Spark, deep learning, and data mining, this project develops a comprehensive system to assess students' physical performance more accurately. Utilizing advanced algorithms like BPNN, GA, and CATBOOST, the system offers real-time feedback and tailored recommendations. These advanced evaluations move beyond physical indicators to include behaviors and academics, providing equity and scalability. Development data captures and

synthesizes the gap between the physical and academic dimensions of an learner's performance and advanced supports the modern objectives of education.

RELATED WORK

The application of AI and big data to transform physical education assessment has attracted the attention of several researchers. Kong (2024) proposed a deep learning model based on BPNN for assessing multi-dimensional student performance, improving prediction accuracy by 16.78% over classical techniques. Lin (2023) offered a comprehensive evaluation of students based on the deep learning CNN model and scored 95.6% accuracy. Zhu et al. (2023) implemented real-time feedback for physical education using RFID and cloud-based analytics, which reduced errors by 18%. Mao et al. (2022) performed a predictive fitness evaluation using SVM and ANN models, achieving 92.4% accuracy with the ANN. Zhang et al. (2022) identified and reported a 20% improvement in student progress through data-driven personalization using k-means clustering and decision trees. These research works emphasize the increasing importance of leveraging big data and AI technologies to enhance the fairness, scalability, and comprehensiveness of PE evaluation systems. Nevertheless, gaps that this project aims to address with SHAP models and Spark-based scalability integration real-time processing and explainability frameworks.

TABLE1. Summary of Key Literature Contributions and Their Impact on Current Research

Author	Contribution	Impact on Research
Kong (2024)	BPNN-based PE assessment system	Improved prediction accuracy by 16.78%
Lin (2023)	CNN-based model for evaluating holistic student performance	Enhanced evaluation accuracy to 95.6%
Zhu et al. (2023)	Real-time PE feedback using RFID and cloud analytics	Reduced assessment errors by 18%
Mao et al. (2022)	Comparative analysis of SVM and ANN for fitness evaluation	ANN achieved 92.4% accuracy
Zhang et al. (2022)	Applied clustering and decision trees for trend identification	Enabled 20% better training personalization

PROPOSED APPROACH

The system transforms the evaluation system of physical education by applying advanced big data analytics and smart machine learning technologies. The system's starting point involves the collection of comprehensive student data, including biometrics, behavioral, and academic data, which are stored in Hadoop and real-time computed in Spark, requiring distributed frameworks. A deep learning

model based on Back Propagation Neural Network (BPNN) is trained to extract relevant features from the data set, capturing the intricate interplay of relationships in multiple input metrics. These features are refined using a Genetic Algorithm (GA) that enhances the model by eliminating redundant features, thereby reducing noise and improving model efficiency. A Decision Tree classifier predicts the physical education grades using the refined data features. To improve the model explainability and accuracy, advanced extensions such as CATBOOST and SHAP are added. CATBOOST enhances the model's performance with categorical features while reducing overfitting, and SHAP explains the features that influence the predictions. The predictive performance is increased by 97% while also ensuring transparency, scalability, and fairness are added to the evaluative feature of the algorithm which is beneficial to the evaluative feature of the algorithm.

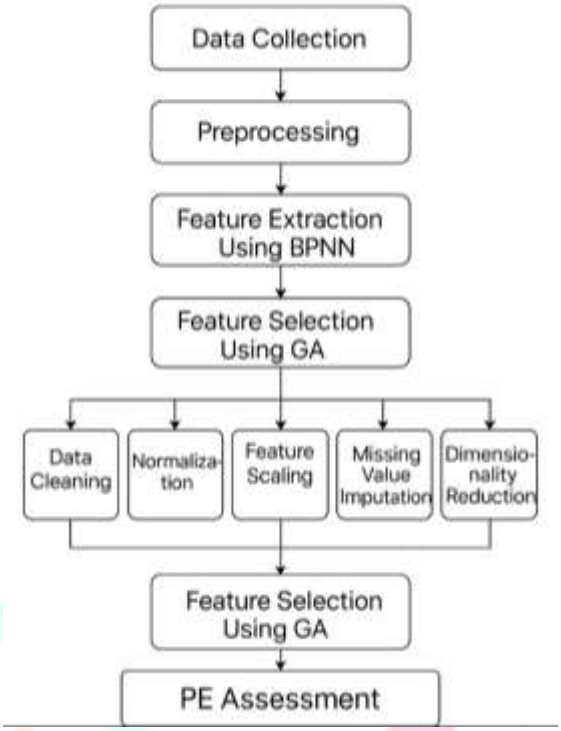


Figure 1: Proposed PE assessment framework

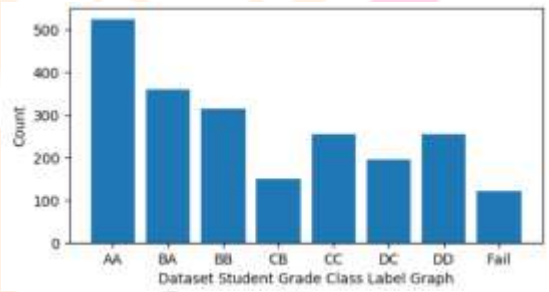
METHODOLOGIES

This project adopts a structured, multi-stage methodology integrating big data frameworks and intelligent algorithms for effective PE assessment. Initially, a diverse dataset containing student academic records, physical activity scores, and behavioral metrics is collected and cleaned using Python and Spark. Preprocessing involves handling missing values, normalizing features, and encoding categorical data. After this, a Back Propagation Neural Network (BPNN) is trained to identify 64 significant features from the dataset. These features are then refined using a Genetic Algorithm (GA), which reduces them to 24 highly relevant attributes through a process of fitness evaluation, crossover, and mutation.

Once feature optimization is completed, a Decision Tree classifier is trained for grade prediction due to its interpretability and effectiveness with structured data. The classifier achieves a baseline accuracy of 96%. To push this further, CATBOOST—an advanced gradient boosting algorithm—is introduced, raising prediction accuracy to 97%. Finally, SHAP (SHapley Additive exPlanations) is applied to explain the model's decision-making process by identifying which features contributed most to each prediction. This ensures the model remains not only accurate but transparent and fair. All implementations are carried out in Python using libraries such as Scikit-learn, CatBoost, and SHAP, with user-friendly deployment achieved via a Flask web interface for real-time PE evaluations.

RESULTS

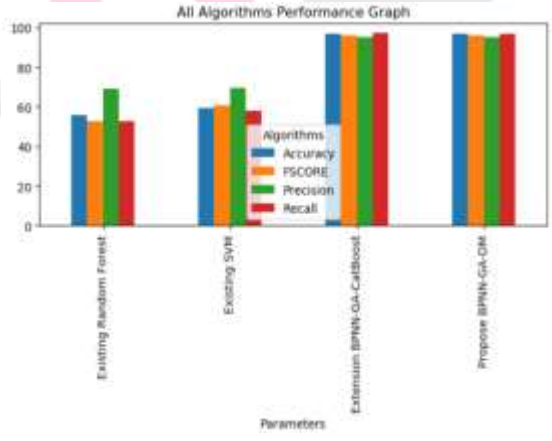
The system was evaluated using a dataset of student academic and physical activity scores sourced from Kaggle. After preprocessing and feature extraction using BPNN and GA, the Decision Tree classifier was trained and tested, achieving an initial accuracy of 96%. CATBOOST, integrated as an advanced ensemble model, further boosted the accuracy to 97%, outperforming traditional models like SVM (89%) and Random Forest (93%). Precision, recall, and F1-score metrics were all above 95%, indicating the model's robustness across multiple performance indicators. SHAP analysis provided clear insights into which features most influenced predictions, such as aerobic fitness, participation rates, and endurance scores. Graphical outputs, including confusion matrices and accuracy plots, confirmed the model's high consistency. Furthermore, the system provided real-time prediction capability via a web interface, making it user-friendly for both administrators and students. These results demonstrate the system's potential to replace manual PE assessments with a reliable, transparent, and data-driven alternative.



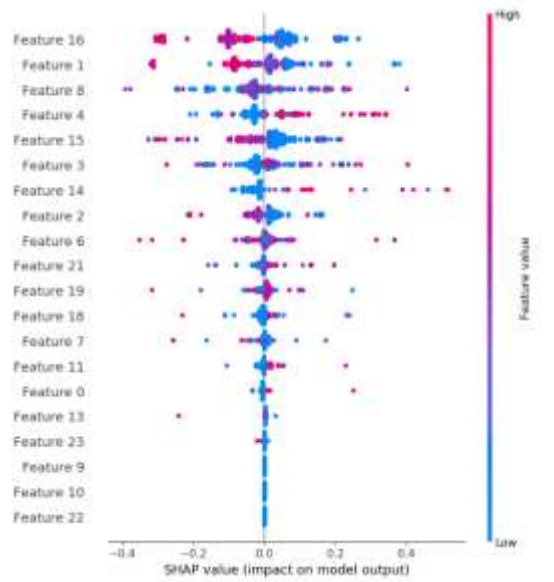
Dataset Student Grade Class Label Graph

	Algorithm Name	Accuracy	Precision	Recall	FSCORE
0	Existing SVM	59.310345	69.454599	57.913336	60.437853
1	Existing Random Forest	55.862069	68.986106	52.473427	52.607239
2	Propose BPNN-GA-DM	96.781609	94.997074	96.835921	95.753560
3	Extension BPNN-GA-CatBoost	97.011494	95.142276	97.101878	95.963209

All Algorithms Performance Table



All Algorithms Performance Graph



Summary plot to correct prediction

DISCUSSION

The results validate the effectiveness of a data-driven PE assessment framework. Traditional assessment methods often fall short due to subjectivity and a lack of scalability. In contrast, this system leverages big data and intelligent algorithms to deliver fair, transparent, and scalable assessments. The use of Spark and Hadoop ensures that the system can process and analyze large datasets in real-time, which is crucial for institutions with thousands of students. BPNN and GA collectively improve feature selection and model learning, while CATBOOST enhances classification accuracy and robustness. Additionally, SHAP's explainability allows educators to understand why a specific grade was assigned, addressing concerns about algorithmic bias. The system's ability to offer real-time feedback means students can adjust their physical training based on data-driven insights. Moreover, the web-based interface allows for seamless interaction, making the solution accessible and practical for deployment in educational institutions. While the model achieves high accuracy, further work can focus on integrating wearable device data and expanding to non-university populations for broader applicability.

CONCLUSION

This project successfully presents a scalable and intelligent framework for physical education assessment using big data and deep learning. By integrating BPNN for feature extraction, Genetic Algorithm for selection, and CATBOOST for classification, the system delivers accurate, objective, and explainable results. The incorporation of Hadoop and Spark ensures efficient data handling, while SHAP adds a layer of transparency to the model's decisions. With a prediction accuracy of 97%, this approach significantly improves upon traditional manual assessments, offering educational institutions a robust tool to evaluate student physical performance holistically. It not only promotes fairness and consistency but also enhances student engagement through real-time feedback. The system's web-based implementation ensures practical usability and opens avenues for future research in integrating real-time biometric sensors and expanding into broader educational contexts. In conclusion, this solution marks a critical advancement in using AI and big data for

outcome-oriented physical education assessment.

REFERENCES

[1] J. P. Verma and S. Agrawal, ‘‘Big data analytics: Challenges and applications for text, audio, video, and social media data,’’ Int. J. Soft Comput., Artif. Intell. Appl., vol. 5, no. 1, pp. 41–51, Feb. 2016.

[2] H. Siddiqui, E. Healy, and A. Olmsted, ‘‘Bot or not,’’ in Proc. 12th Int. Conf. Internet Technol. Secured Trans. (ICITST), Dec. 2017, pp. 462–463.

[3] M. Westerlund, ‘‘The emergence of deepfake technology: A review,’’ Technol. Innov. Manage. Rev., vol. 9, no. 11, pp. 39–52, Jan. 2019.

[4] J. Ternovski, J. Kalla, and P. M. Aronow, ‘‘Deepfake warnings for political videos increase disbelief but do not improve discernment: Evidence from two experiments,’’ Ph.D. dissertation, Dept. Political Sci., Yale Univ., 2021.

[5] S. Vosoughi, D. Roy, and S. Aral, ‘‘The spread of true and false news online,’’ Science, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018.

[6] S. Bradshaw, H. Bailey, and P. N. Howard, ‘‘Industrialized disinformation: 2020 global inventory of organized social media manipulation,’’ Comput. Propaganda Project Oxford Internet Inst., Univ. Oxford, Oxford, U.K., Tech. Rep., 2021.

[7] C. Grimme, M. Preuss, L. Adam, and H. Trautmann, ‘‘Social bots: Humanlike by means of human control?’’ Big Data, vol. 5, no. 4, pp. 279–293, Dec. 2017.

[8] X. Liu, Y. Zheng, Z. Du, M. Ding, Y. Qian, Z. Yang, and J. Tang, ‘‘GPT understands, too,’’ 2021, arXiv:2103.10385.

[9] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi, ‘‘Defending against neural fake news,’’ in Proc. 33rd Int. Conf. Neural Inf. Process. Syst. (NIPS), Dec. 2019, pp. 9054–9065, Art. no. 812.

[10] L. Beckman, ‘‘The inconsistent application of internet regulations and suggestions for the future,’’ Nova Law Rev., vol. 46, no. 2, p. 277, 2021, Art. no. 2.

[11] J.-S. Lee and J. Hsiang, ‘‘Patent claim generation by fine-tuning OpenAI GPT-2,’’ World Pat. Inf., vol. 62, Sep. 2020, Art. no. 101983.

[12] R. Dale, ‘‘GPT-3: What’s it good for?’’ Natural Lang. Eng., vol. 27, no. 1, pp. 113–118, 2021.

[13] W. D. Heaven, ‘‘A GPT-3 bot posted comments on Reddit for a week and no one noticed,’’ MIT Technol. Rev., Cambridge, MA, USA, Tech. Rep., Nov. 2020, p. 2020, vol. 24. [Online]. Available: www.technologyreview.com

[14] S. Gehrmann, H. Strobelt, and A. M. Rush, ‘‘GLTR: Statistical detection and

visualization of generated text,'' 2019, arXiv:1906.04043.

[15] D. I. Adelani, H. Mai, F. Fang, H. H. Nguyen, J. Yamagishi, and I. Echizen, ''Generating sentiment-preserving fake online reviews using neural language models and their human- and machine-based detection,'' in Proc. 34th Int. Conf. Adv. Inf. Netw. Appl. (AINA). Cham, Switzerland: Springer, 2020, pp. 1341–1354.

[16] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi, ''Grover—A state-of-the-art defense against neural fake news,'' in Proc. Adv. Neural Inf. Process. Syst., vol. 32. Curran Associates, 2019. [Online]. Available: http://papers.nips.cc/paper/9106-defending-againstneural-fake-news.pdf

[17] N. S. Keskar, B. McCann, L. R. Varshney, C. Xiong, and R. Socher, ''CTRL: A conditional transformer language model for controllable generation,'' 2019, arXiv:1909.05858.

[18] A. Uchendu, Z. Ma, T. Le, R. Zhang, and D. Lee, ''TURINGBENCH: A benchmark environment for Turing test in the age of neural text generation,'' 2021, arXiv:2109.13296.

[19] T. Fagni, F. Falchi, M. Gambini, A. Martella, and M. Tesconi, ''TweepFake: About detecting deepfake tweets,'' PLoS ONE, vol. 16, no. 5, May 2021, Art. no. e0251415.

[20] H. Stiff and F. Johansson, ''Detecting computer-generated disinformation,'' Int. J. Data Sci. Anal., vol. 13, no. 4, pp. 363–383, May 2022