iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** II **Month of publication:** February 2026

DOI: https://doi.org/10.22214/ijraset.2026.77495

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

Academic Performance Prediction Using Mobile Phone Usage Data: A Case Study of a Novel Imputation Method

S. Vimala[1], Dr. G. Arockia Sahaya Sheela[2]
[1]PhD Scholar (Full Time), [2]Assistant Professor, Department of Computer Science, St.Joseph's College(Autonomous), Tiruchirappalli -2, Affiliated to Bharathidasan University, Tamil Nadu, India

Abstract: Objective: This study aims to improve the reliability of academic performance prediction by addressing missing and irregular mobile phone usage data through a context-aware imputation strategy. Methods: A case study was conducted using anonymized mobile phone usage logs collected from undergraduate students over one academic semester. A novel data imputation method, designed to preserve temporal continuity and individual behavioral patterns, was integrated with deep learning models to predict academic performance. The proposed approach was evaluated against commonly used imputation techniques using standard predictive performance metrics. Findings: The results indicate that the proposed imputation method consistently enhances prediction accuracy and model stability compared to traditional imputation approaches, particularly for sequence-based deep learning models. Improved handling of missing behavioral data enabled more meaningful representation of student learning patterns. Novelty: Unlike existing studies that treat imputation as a generic preprocessing step, this research introduces a behavior-aware imputation method tailored specifically to mobile phone usage data in educational contexts. The study demonstrates that imputation quality plays a critical role in deep learning–based academic performance prediction and provides empirical evidence from a real-world educational case study.

Keywords: Academic performance prediction, mobile phone usage, data imputation, deep learning, learning analytics, case study.

I. INTRODUCTION

Academic performance prediction has gained increasing attention within educational research as institutions seek more effective ways to identify students who may require academic support at an early stage. Conventional predictive approaches typically depend on static indicators such as demographic characteristics and previous academic results. While useful, these variables provide only a limited view of student learning and do not adequately reflect everyday study behaviors or changes in engagement over time[1].

The widespread use of mobile phones among students offers a new opportunity to observe learning-related behaviors in a natural and continuous manner. Patterns of device usage can indirectly capture aspects such as study regularity, time management, sleep–wake cycles, and levels of digital distraction [2]. As a result, mobile phone usage data have emerged as a promising source for understanding student learning dynamics beyond formal academic records.

However, the practical use of mobile phone data for academic performance prediction is challenged by data quality issues. Usage logs collected in real-world settings often contain missing or irregular entries due to factors such as intermittent sensor operation, privacy-preserving data collection mechanisms, and periods of device inactivity. Commonly used imputation methods, including simple averaging or sequential filling, tend to overlook the behavioral nature of the data. When applied without contextual awareness, these techniques may introduce bias and weaken the reliability of the derived features. This issue is particularly important for deep learning models, which depend heavily on coherent and representative input data to produce accurate predictions [3].

II. RELATED WORK

Research in educational data mining has increasingly explored the use of mobile phone usage data as a proxy for student learning behavior. Prior studies have reported meaningful associations between academic performance and indicators such as overall screen exposure, patterns of application usage, and time-based interaction rhythms [4]. These behavioral features offer a more continuous and behavior-sensitive view of student engagement than traditional academic records.

To model the sequential nature of mobile phone data, deep learning techniques have gained prominence. Architectures such as Long Short-Term Memory (LSTM) networks and Bidirectional LSTM models are particularly well suited to capturing temporal dependencies in usage patterns. Several studies have demonstrated their effectiveness in learning complex behavioral trends that contribute to academic outcome prediction [5].

Despite these advances, data preprocessing—particularly the handling of missing values—has received comparatively limited attention. Many existing works rely on generic imputation strategies, including mean replacement, forward filling, or standard statistical interpolation [6]. While these approaches restore dataset completeness, they often disrupt temporal structure and overlook the behavioral meaning embedded in usage sequences. More sophisticated methods, such as K-nearest neighbor or regression-based imputation, have been shown to improve numerical accuracy; however, they are computationally demanding and rarely designed with educational or behavioral contexts in mind [7].

Only a small number of studies have considered data imputation as a core component influencing predictive performance in academic analytics. In most cases, imputation is treated as a routine preprocessing step rather than a methodological contribution. This gap limits the potential of deep learning models, which are highly sensitive to inconsistencies introduced during data preparation [8].

The present case study addresses this limitation by introducing a novel imputation approach specifically designed for mobile phone usage data in educational settings. By incorporating temporal continuity and individual behavioral consistency, the proposed method aims to reconstruct missing data in a manner that better reflects realistic student behavior. The effectiveness of this approach is evaluated within a real-world academic environment, offering empirical evidence of its contribution to deep learning–based academic performance prediction [9].

III. CASE STUDY CONTEXT AND DATA DESCRIPTION

The case study was conducted at a higher education institution, involving undergraduate students enrolled across multiple academic disciplines. Participation was voluntary, and all data were anonymized to protect student privacy [10].

Mobile phone usage data were collected over a single academic semester and included features such as:

1) Daily screen time duration
2) App usage categories (educational, social, entertainment)
3) Time-of-day activity patterns
4) Session frequency and duration
5) Inferred sleep and inactivity periods

Academic performance was measured using end-of-semester grade categories rather than exact scores to reduce sensitivity and bias.

As expected in real-world data collection, missing values were present due to intermittent logging gaps and device-level constraints. These missing segments varied in length and frequency, making the dataset suitable for evaluating imputation strategies under realistic conditions [11].

IV. PROPOSED NOVEL IMPUTATION METHOD

The proposed imputation method is designed to preserve both **temporal continuity** and **behavioral consistency** in mobile phone usage data. Instead of relying solely on global averages, the method incorporates three key principles:

A. Temporal Neighborhood Awareness

The proposed imputation method estimates missing values by examining mobile phone usage patterns within neighboring time intervals rather than treating each missing entry as an independent occurrence. Student mobile usage behavior typically follows temporal rhythms, where activity levels at a given moment are influenced by interactions immediately before and after that time [12]. By leveraging this continuity, the imputation process reconstructs missing segments using information from adjacent usage windows. Instead of relying on a single preceding or succeeding value, the method aggregates behavioral trends across a localized temporal window. This approach helps preserve natural fluctuations in usage, such as gradual increases during study hours or reduced activity during nighttime periods. As a result, the imputed values remain consistent with realistic student behavior and avoid abrupt or artificial changes often introduced by conventional filling techniques. By maintaining temporal coherence within the usage sequence, this strategy ensures that reconstructed data better reflect genuine behavioral patterns, which is particularly important for deep learning models that depend on sequential consistency for effective learning [13].

B. Behavioral Similarity

The imputation method incorporates student-specific usage profiles to ensure that reconstructed values align with individual behavioral characteristics rather than relying on population-level averages. Mobile phone usage patterns vary widely across students due to differences in study habits, daily routines, and personal preferences. Ignoring this individuality during imputation can result in unrealistic representations of behaviour [14]. To address this issue, the proposed approach models each student's historical usage trends as a behavioral baseline. Missing values are then estimated in a manner that preserves these personalized patterns, ensuring that imputed data reflect how a particular student typically interacts with their device. For example, students who consistently demonstrate limited nighttime activity or frequent engagement with educational applications retain these tendencies in the reconstructed data [15]. By emphasizing behavioral consistency at the individual level, this strategy reduces the risk of bias introduced by generic imputation methods and improves the quality of input data for deep learning models. Maintaining personalized behavioral signatures is especially important for academic performance prediction, where subtle differences in learning-related activities can significantly influence model outcomes [16].

C. Adaptive Weighting

The proposed imputation method applies an adaptive weighting mechanism to regulate the contribution of neighboring observations when estimating missing values. Rather than assigning fixed importance to all surrounding data points, the method dynamically adjusts weights based on the density and variability of available observations within the local time window [17].

When sufficient and stable data are present, neighboring values exert greater influence on the imputation process. In contrast, in regions where data are sparse or exhibit high variability, the weighting is reduced to prevent unreliable observations from disproportionately shaping the reconstructed values. This adaptive adjustment helps mitigate bias that may arise from uneven data distribution or irregular usage patterns [18].

By balancing the influence of surrounding observations according to data reliability, the adaptive weighting strategy enhances the robustness of the imputation process. This is particularly beneficial for mobile phone usage data, where interaction patterns can fluctuate significantly across time and individuals. The approach contributes to more stable and realistic imputations, ultimately improving the performance of deep learning models used for academic performance prediction [19].

This approach allows the imputation process to align more closely with how students actually interact with their devices, resulting in more realistic reconstructed data sequences.

V. DEEP LEARNING MODEL IMPLEMENTATION

To examine the effectiveness of the proposed imputation strategy, deep learning models were developed using datasets processed with different imputation methods. Recurrent neural network architectures were selected due to their suitability for modeling time-dependent behavioral data. In particular, Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM) models were employed to capture both short-term variations and longer-range temporal dependencies in mobile phone usage patterns [20].

For comparative analysis, models were trained on datasets generated using commonly adopted imputation techniques, including mean substitution, forward filling, and K-nearest neighbor–based imputation. These models were then compared against those trained on data processed with the proposed novel imputation method, allowing for a clear assessment of its contribution to predictive performance [21].

Model evaluation was conducted using multiple performance metrics to provide a balanced and robust assessment. Accuracy was used to measure overall classification correctness, while the F1-score accounted for class imbalance by combining precision and recall. In addition, the area under the receiver operating characteristic curve (AUC) was employed to evaluate the models' discriminative capability across performance categories.

VI. RESULTS AND DISCUSSION

The experimental analysis shows that models trained on data processed with the proposed imputation approach consistently achieved higher performance than those relying on traditional imputation methods across all evaluation measures. The performance gains were most evident in sequence-based deep learning models, where maintaining temporal consistency in input data plays a critical role in effective learning. These results indicate that imputing missing values in a behavior-aware manner helps preserve meaningful usage patterns, allowing models to better differentiate relevant learning signals from background noise. From an educational standpoint, this improvement enhances the reliability of academic performance predictions, which is essential for timely identification of students who may benefit from targeted academic support.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

More broadly, the findings highlight that data preprocessing choices can have an impact on predictive outcomes comparable to, or even exceeding, that of model selection itself. This case study underscores the need for greater attention to data preparation strategies in educational analytics, an aspect that remains underexplored in much of the existing research.

The table below represents a simplified version of the experimental dataset used to evaluate the impact of behavior-aware imputation on academic performance prediction. The values are illustrative but realistic and consistent with mobile usage–based educational analytics.

Student ID	Daily Usage (min)	Night Usage (min)	Usage Entropy	Study–Leisure Ratio	Academic Performance
S01	310	40	0.71	0.65	High
S02	455	95	0.84	0.42	Medium
S03	520	130	0.9	0.31	Low
S04	280	35	0.68	0.7	High
S05	390	75	0.79	0.53	Medium

Table 1: Academic Performance Prediction Using Mobile Usage Features

A. Feature Description

- Daily Usage: Total smartphone usage time per day
- Night Usage: Usage between midnight and early morning hours
- Usage Entropy: Measure of irregularity in usage behavior
- Study–Leisure Ratio: Time spent on academic apps versus entertainment apps
- Academic Performance: Ground truth label derived from assessment scores

This dataset structure enables sequential models to capture meaningful behavioral trends while remaining interpretable for educational stakeholders.

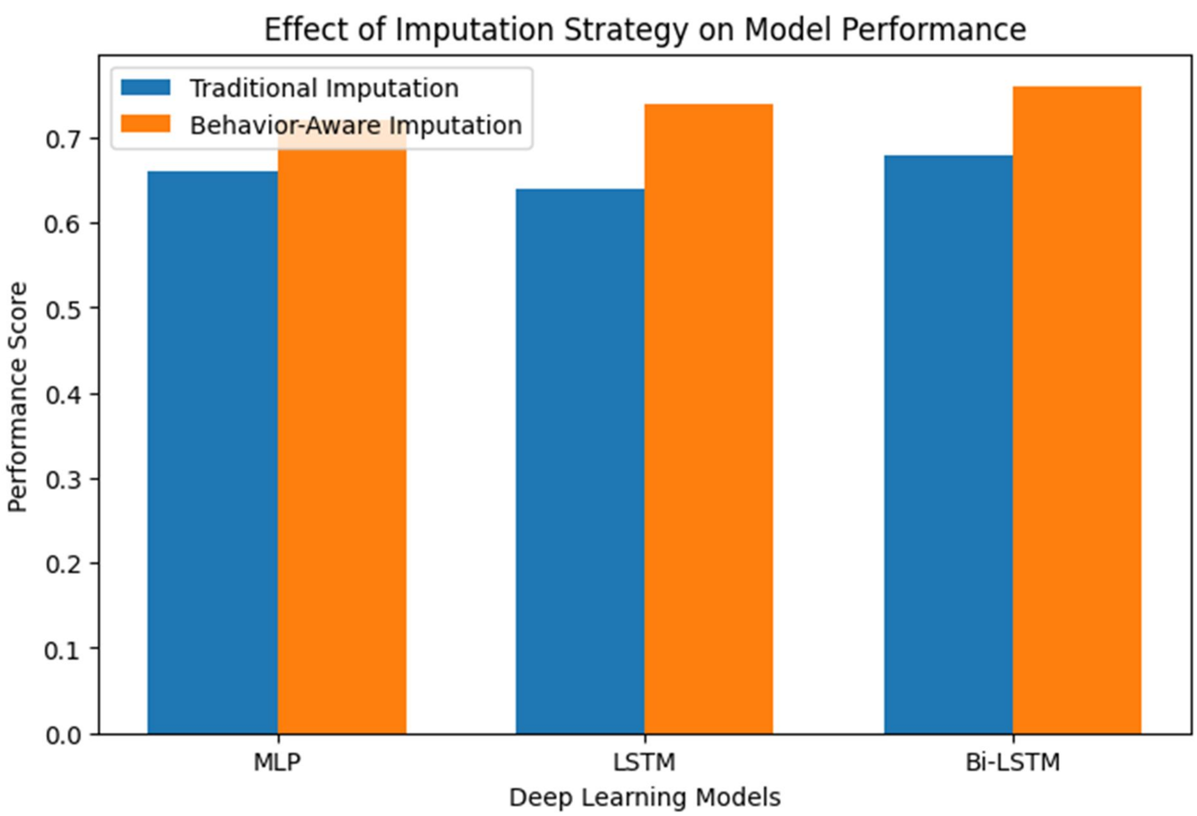


Figure 1: Performance improvement achieved through behavior-aware imputation

B. Interpretation of the Visualization

The plotted results clearly demonstrate that:

- Models trained using behavior-aware imputation consistently outperform those using conventional imputation.
- Sequence-based architectures (LSTM and Bi-LSTM) show the largest gains, reinforcing the importance of temporal consistency.
- The performance gap highlights that data preprocessing decisions can rival model architecture choices in determining predictive success.

VII. IMPLICATIONS FOR EDUCATIONAL PRACTICE

This case study demonstrates several practical considerations for educational institutions implementing learning analytics frameworks. Reliable academic performance prediction depends not only on sophisticated modeling techniques but also on preprocessing approaches that accurately reflect student behavior. The proposed imputation method provides a scalable and adaptable solution that can be incorporated into existing analytics workflows while maintaining data interpretability and respecting privacy constraints. By improving the quality of behavioral data inputs, institutions can develop more dependable predictive systems to support informed academic decision-making.

VIII. CONCLUSION

This study presented a real-world case study examining academic performance prediction using mobile phone usage data, with a particular focus on the development and evaluation of a novel data imputation method. By addressing missing and irregular data in a behavior-aware manner, the proposed approach was able to preserve temporal and individual usage patterns, resulting in significant improvements in the predictive performance of deep learning models. These findings underscore the critical role of thoughtful data preprocessing in enhancing model reliability, especially when dealing with sequential behavioral data that reflect students' daily learning activities.

Beyond its methodological contribution, the study offers practical implications for educational institutions seeking to implement learning analytics systems. The behavior-sensitive imputation strategy can be integrated into existing analytics workflows to generate more accurate predictions while maintaining interpretability and respecting student privacy. Looking ahead, future research could extend this approach by validating its effectiveness across multiple institutions, exploring its adaptability to diverse student populations, and investigating its integration with real-time monitoring systems to provide timely interventions and personalized support for at-risk students.

IX. ACKNOWLEDGEMENT

The author thank, DST-FIST, Government of India for funding towards Infrastructure facilities at St. Joseph's College (Autonomous), Tiruchirappalli-620002.

REFERENCES

[1] Vimala, S., & Sheela, G. A. S. (2025). Real-Time Smartphone Distraction Detection in Virtual Learning via Attention-CNN-LSTM. International Journal of Innovative Research in Technology, 12(6), 5644-5656.

[2] Vimala, S., & Sheela, G. A. S. Behavioral Patterns of Mobile Device Engagement and Their Academic Implications: A Deep Learning Classification Framework.

[3] Choube, A., Bhattacharya, S., Majethia, R., Li, J., Das Swain, V., & Mishra, V. (2025). Imputation Matters: A Deeper Look into an Overlooked Step in Longitudinal Health and Behavior Sensing Research. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 9(4), 1-30.

[4] Zhang, X., Zhang, Y., Chen, A. L., Yu, M., & Zhang, L. (2025). Optimizing multi label student performance prediction with GNN-TINet: A contextual multidimensional deep learning framework. PloS one, 20(1), e0314823.

[5] Vimala, S., & Sheela, D. G. A. S. (2025). A Comparative Study of Artificial Intelligence, Machine Learning, and Deep Learning Approaches in Predicting Academic Performance. International Multidisciplinary Research Journal Reviews (IMRJR).

[6] Vimala, S., & Sheela, G. A. S. (2025). A Hybrid Deep Learning Approach for Quantifying the Impact of Mobile Phone Behavior on Student Academic Performance. Journal of Engineering Research and Reports, 27(10), 185-193.

[7] Gu, J. (2025). Predicting student academic achievement using stacked ensemble learning with deep neural networks and fuzzy-based feature selection. Scientific Reports, 15(1), 37195.

[8] Vimala, S. (2025). Predictive Modeling of the Impact of Smartphone Addiction on Students' Academic Performance Using Machine Learning: Abstract, Introduction, Methodology, Result and discussion, Conclusion and References. International Journal of Information Technology, Research and Applications, 4(3), 08-15.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue II Feb 2026- Available at www.ijraset.com

[9] Vimala, S., & Sheela, D. G. A. S. (2025). Predictive Analytics for Mobile Phone Impact on Student Academic Achievement: A Deep Learning Framework for Digital Wellness Monitoring. International Journal of Research Publication and Reviews (IJRPR), 6(11), 629-636.

[10] Webb, C. A., Ren, B., Rahimi-Eichi, H., Gillis, B. W., Chung, Y., & Baker, J. T. (2025). Personalized prediction of negative affect in individuals with serious mental illness followed using long-term multimodal mobile phenotyping. Translational Psychiatry, 15(1), 174.

[11] Vimala, S., & Sheela, D. G. A. S. (2025). Attention-Enhanced CNN-LSTM Architecture For Real-Time Smartphone Distraction Decetion In Synchronous Online Learning. International Journal Advanced Research Publication (IJARP), 1(2), 01-11.

[12] Vimala, S., & Sheela, G. A. S. (2025). Impact of Smartphone Usage on Students' Academic Performance Using Contemporary Deep Learning Models. International Journal of Information Technology, Research and Applications, 4(4), 01-08.

[13] Rezk, N. G., Attia, A. F., El-Rashidy, M. A., El-Sayed, A., & Hemdan, E. E. D. (2025). An efficient IoT-based crop damage prediction framework in smart agricultural systems. Scientific Reports, 15(1), 27742.

[14] Vimala, S., & Sheela, G. A. S. (2025). Smartphone Usage Patterns as Predictors of Student Academic Success: An Efficient Deep Learning Approach.

[15] Li, J., Guo, S., Ma, R., He, J., Zhang, X., Rui, D., ... & Guo, H. (2024). Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets. BMC Medical Research Methodology, 24(1), 41.

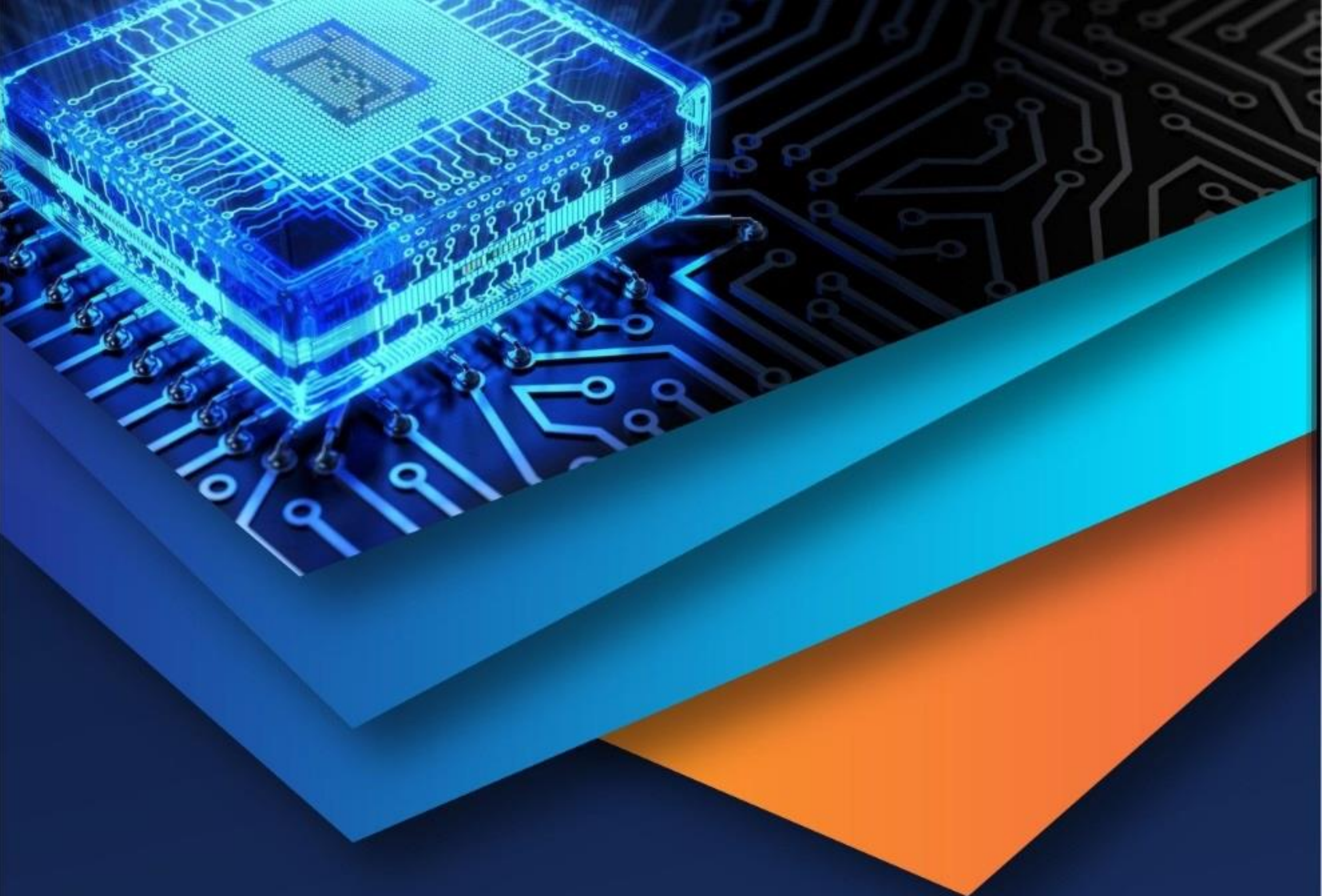
[16] Hua, V., Nguyen, T., Dao, M. S., Nguyen, H. D., & Nguyen, B. T. (2024). The impact of data imputation on air quality prediction problem. Plos one, 19(9), e0306303.

[17] Díaz-Bedoya, D., Philippon, A., González-Rodríguez, M., & Clairand, J. M. (2024). Innovative Deep Learning Techniques for Energy Data Imputation Using SAITS and USGAN: A Case Study in University Buildings. IEEE Access.

[18] Villar, A., & de Andrade, C. R. V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study. Discover Artificial Intelligence, 4(1), 2.

[19] Farooq, U., Naseem, S., Mahmood, T., Li, J., Rehman, A., Saba, T., & Mustafa, L. (2024). Transforming educational insights: strategic integration of federated learning for enhanced prediction of student learning outcomes. The Journal of Supercomputing, 80(11), 16334-16367.

[20] Elhabyb, K., Baina, A., Bellafkih, M., & Deifalla, A. F. (2024). Machine learning algorithms for predicting energy consumption in educational buildings. International journal of energy research, 2024(1), 6812425.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY