International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Head Impact Detection Using Machine Learning Algorithms

[1]Ashritha Arelli, [2]N.Kavya Sree, [3]Macha Vardhan ,[4]Jaladi.N.Chandra Sekhar

[1,2,3] UG Scholars, [4]Assistant Professor

[1,2,3,4] Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning)

[1,2,3,4] Guru Nanak Institutions Technical Campus, Ibrahimpatnam,Telangana,India

---***---

Abstract - The need of precisely identifying head impacts and putting in place efficient safety measures has been underlined by numerous research investigations.. This study addresses this critical need by leveraging Logistic Regression, a simpler yet highly effective data collected from piezoelectric sensors installed on a model of a simulated cranium using a machine learning technique. We process normalized sensor data using Logistic Regression in a methodical manner with the goal of precisely identifying impact sites. The model achieves an impressive accuracy of 90%, demonstrating its capability to perform well in real-time applications that require quick, interpretable, and computationally efficient predictions. Rigorous evaluation the model's performance is highlighted by employing k-fold cross-validation, while feature importance analysis identifies an optimal sensor placement strategy. This strategy may reduce model complexity, potentially leading to more efficient implementations without compromising predictive accuracy. The strong performance of Logistic Regression, coupled with its simplicity and interpretability, underscores its potential as an ideal solution for head impact detection in safety-critical environments .The results aid in the creation of intelligent safety systems, which combine machine learning and wearable technologies to boost safety and decision-making in both industrial and sporting settings.

***Key Words*:** Logistic Regression, piezoelectric sensor, k-fold cross-validation, Machine Learning, wearable technology

1. INTRODUCTION

Head impact detection plays a critical role in enhancing safety protocols, especially in environments such as sports, industrial settings, and military operations, where individuals are at risk of experiencing head injuries. Accurate detection of head impacts is essential for initiating timely interventions that can mitigate the severity of injuries, improve safety measures, and guide medical responses [1]. This paper aims to address this challenge by applying machine learning techniques to information gathered from piezoelectric sensors installed on a model of a head. The sensors capture dynamic force measurements during impact events, which are then processed to pinpoint the precise position and severity of the hit Traditional methods of impact detection often rely on simplistic thresholding techniques or mechanical sensors, which can be limited in accuracy and adaptability. In contrast, models for machine learning like logistic regression ,offer a effective method for deciphering intricate patterns from the sensor and make more precise predictions. This study aims to achieve high classification accuracy and real-time decision-making through the application of Logistic Regression, while simultaneously preserving interpretability and computing economy [2]. The study explores the potential for combining wearable technology and machine learning techniques to create smart, flexible safety solutions that improve head impact detection and advance safety technologies in general. In the end, our research opens the door to more efficient, data-driven approaches to head injury prevention and mitigation, especially in high-risk settings where safety is crucial.

In this work, a machine learning-based system for precise head impact detection is developed and evaluated using data gathered from piezoelectric sensors installed on a model of a simulated head. The primary focus is on employing Logistic Regression as the machine learning algorithm to analyze normalized sensor data and identify impact locations with high precision. This study looks into the feasibility of using machine learning techniques, specifically Logistic Regression, to improve traditional mechanical or threshold-based systems, which often struggle with accuracy and flexibility in dynamic environments .The system's scope includes data preparation, feature extraction, model training, and performance evaluation utilizing methods like k-fold cross-validation to ensure generalization and robustness. [3]. In order to improve sensor placement techniques and maybe reduce model complexity while preserving or even boosting predictive accuracy, the research also highlights the significance of feature importance analysis [4]. Because the system was created with real-time performance in mind, it can be used in settings with constrained processing power, like wearable technology or portable security systems. By focusing on both the performance and efficiency of Logistic Regression, By providing a workable method for head impact detection in sports, industrial settings, and other high-risk contexts, the article also seeks to advance the field of intelligent safety systems. Essentially, the scope encompasses the creation of the head impact detection system as well as its possible incorporation into practical applications, which will ultimately improve safety protocols and provide faster, more precise decision-making to reduce head injuries [5–6].

The goal is to use Logistic Regression to data from piezoelectric sensors put on a simulated head model in order to create a machine learning-based system for precise head impact detection.The primary goal is to accurately determine impact locations and intensities by using sensor data processing to consistently detect head impacts.The goal of the study is to provide a dependable solution for real-time head impact detection by using Logistic Regression to achieve high classification accuracy (targeting 90% or above). By fulfilling these goals, the study hopes to aid in the creation of intelligent safety systems that use machine learning and wearable technologies to improve general safety and decision-making.

2. LITERATURE SURVEY

In order to understand detector data, similar as that attained from piezoelectric detectors deposited on a simulated head model, machine literacy algorithms like Random Forest(RF) and Extreme Gradient Boosting(XG Boost) are constantly used in head impact discovery. In order to determine the class mode for bracket tasks, Random Forest, an ensemble learning fashion, builds a large number of decision trees during training[7]. It's extensively used because it can handle high- dimensional data and mimic complex connections without overfitting. In a analogous tone, XG Boost is another well- liked machine literacy system that has a strong character for delicacy. By combining the prognostications of several weak learners and perfecting them over duplications, it builds a model using a grade boosting frame. With great delicacy in relating impact events, both of these models have been used to pinpoint the position and inflexibility of head impacts. These models can manage the nonlinear correlations between detector readings and impact spots in the environment of brain injury discovery, adding the system's delicacy. The being systems also profit from ways similar ask-foldcross-validation to assess the models' performance and insure their robustness across different datasets [8]. also, point significance analysis is extensively employed to determine the optimal detector point, reducing system complexity while maintaining vaticination accuracy It has been demonstrated that these machine literacy models perform better than conventional threshold- grounded ways, offering real- time impact discovery systems that are more adaptable, scalable, and secure. These systems are generally employed in sports, where timely and accurate head impact discovery is pivotal for safety and in sectors where workers may be exposed to dangerous impact circumstances.

3. PROPOSED SYSTEM

For head impact detection, the suggested method makes use of Logistic Regression, a straightforward yet powerful machine learning approach. Based on input sensor data, the linear classification model known as logistic regression calculates the likelihood of a binary outcome in this case, whether or not a head hit has occurred. This model is a great option for deployment in resource-constrained situations, like wearable technology, because it is computationally efficient and appropriate for real-time applications. The interpretability of logistic regression is one of its main benefits; it offers distinct insights into how the input features affect the prediction, which is essential in applications involving safety. Despite being a simpler model compared to more complex algorithms like Random Forest or XGBoost, Logistic Regression can still achieve high accuracy, especially when the data is well-preprocessed and the relationships between features are relatively linear[18].The suggested system seeks to offer a dependable, interpretable, and real-time solution for head impact detection that guarantees effective deployment without sacrificing performance.

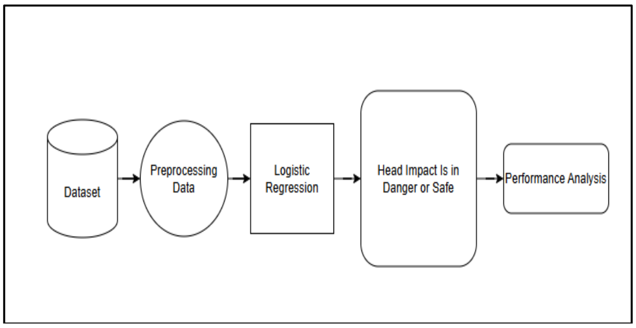


Figure 1: System Architecture

The following modules are used to apply regression:
1) Data Gathering
2) Data Preparation
3) The process of feature engineering
4) Model Creation
5) Model Assessment and Performance Indicators
6) Forecast

- Data accession: The first stage of any machine learning composition is data accession. In order to train and assess the model, it entails collecting raw data. Data for a bracket challenge similar as head impact discovery may appear from a number of sources, similar as tests, simulations, orpre-existing datasets.
- Data preprocessing: To guarantee that the raw data is in a state that can be used for model training, data medication is an essential step. The data needs to be gutted, reused, and arranged before being fed into the machine literacy model. First, as it could dispose the model's performance, managing missing or deficient data is essential. Missing variables can be imputed using styles like mean insinuation or more complex strategies like KNN insinuation. Outliers, or extreme values that diverge significantly from other compliances, should also be controlled because they can have an impact on the model.
- Feature Engineering: point engineering involves creating fresh variables from the raw data in order to ameliorate the model's performance. In order to prize features that more directly reflect the underpinning patterns in the data, sphere moxie is constantly employed in this step. For case, in an impact discovery script, deduced features like the mean, friction, or peak values during particular time frames may be reckoned using raw time- series data from detectors.
- Model Development: The machine literacy model is created once the data has been preprocessed and material characteristics have been recaptured. A popular bracket approach that performs well in situations where categorical result vaticination is the end is logistic retrogression. It finds the best- fitting direct decision border between the classes by calculating the parameters (portions) that maximize the liability of the observed data. During training, the model is fed the pre-processed data, and an optimization approach(similar grade descent) is used

to minimize the logistic loss function. In a double bracket task, this will mean figuring out the probability that an case will belong to one of the two classes.

- Model Evaluation and Performance Metrics: To make sure the Logistic Retrogression model can effectively generalize to new data, it's pivotal to assess its performance after training. Dividing the dataset into training and test sets is one of the original evaluation stages. The test set is also used to assess how well the model performs on data that has not been seen ahead. The most abecedarian performance statistic is delicacy, which shows the proportion of accurate prognostications the model makes.
- Vaticination: Using the learned model to read fresh, untested data is the last phase. At this point, the model can be used in practical operations. The model should be prepared to make prognostications on incoming data, whether it be batch or real- time data, after it has been trained and assessed. For illustration, in head impact discovery, the model would read the probability of a head impact and, if one were linked, pinpoint the impact's position.

4. EXPERIMENTAL PROCESS

- **RANDOM FOREST with XGBOOST**

Both Random Forest and XG Boost(Extreme Gradient Boosting) are extremely effective ensemble knowledge ways that are constantly used for type and regression problems. still, their working propositions differ, which may have an impact on how well they perform depending on the specifics of the problem. Using bootstrap slice, sometimes appertained to as arbitrary slice with relief, Random Forest is a collection of decision trees that are trained on various arbitrary subsets of the data. likewise, each tree split avoids overfitting and decreases correlation between trees by only taking into account a arbitrary sample of features. Following training, Random Forest aggregates each tree's affair to give prognostications using either regression averaging or maturity voting for type. Random Forest's strength, ease of use, and capacity to manage a range of data types including continuous and categorical variables are its main advantages. It also provides a point connection metric that makes it easier to understand which attributes are most vital to the projections. still, as the number of trees rises, Random Forest can come computationally ferocious and may not be suitable to handle datasets with largely imbalanced or complex point relations(19). On the other hand, XG Boost uses grade boosting, which is a fashion that fixes the crimes of former models by training models one after the other. The procedure uses a grade descent fashion to minimize the loss function for each new model(constantly a decision tree) that focuses on the residuals(crimes) from the former model. In XG Boost," extreme" refers to a range of sophisticated styles, analogous as regularization, which reduces overfitting, and a sparsity-alive strategy, which is better at managing missing variables.When compared to Random Forest, XG Boost's boosting fashion produces hastily convergence and generally advanced predictive delicacy, especially for big or largely dimensional datasets.

- **LOGISTIC REGRESSION**

The suggested approach, logistic regression, is a straightforward but effective machine knowledge model that is constantly applied to challenges involving double and multi-class categorization. It describes the probability of a categorical dependent variable using one or further independent factors. A direct equation's outgrowth is colluded to a probability value between 0 and 1 by the logistic function, constantly known as the sigmoid function. This value can also be threshold to give double vaticinations. The recommended system is logistic regression, a simple yet important machine knowledge model that is considerably used for double and multi- class type problems. It describes the probability of a categorical dependent variable using one or further independent factors. The logistic function, also called the sigmoid function, maps the result of a direct equation to a probability value between 0 and 1. The performing model is a great option for issues demanding precise, computationally effective prognostications since it's simple to understand and performs well in real- time operations.(21).

5. RESULTS AND DISCUSSION

TABLE I: Accuracy Evaluation of Prediction Models

Metric of performance	Random forest	XG Boost
MSE	0.2368	0.2390
RMSE	0.4965	0.4875
MAE	0.0354	0.0159
R2	0.9577	0.9596

Vaticination delicacy is indicated by how near the data points are to the line of concinnity. The maturity of the prognosticated values nearly match the factual bones , according to an analysis of the scatterplots(numbers 9 and 10) for the RFR and XG Boost models. This graphical representation not only gives an intuitive sense of the prophetic power of the model, but it also confirms the preliminarily indicated performance pointers. The confusion matrix of the XG Boost model, on the other hand, shows performance that's further slightly spread throughout all detector regions. The model appears to have a better understanding of the underpinning patterns in the data since its prognostications more nearly match the factual detector readings. A more invariant performance across training and testing datasets is also suggested by fresh examination of the XG Boost model. This thickness is cheering since it shows that the model is less likely to overfit and more likely to produce accurate prognostications when applied to fresh, untested data.Of particular interest is the RFR, a retrogression- grounded variant of RF that forms the base of our study. Given the complexity and implicitnon-linearity of our detector data, RFR is plainly the stylish choice. The methodology included recycling a dataset of colorful detector readings that were separated into regions similar as the forepart, back, crown, left, right, and top state was placed at 42 to insure uniformity across trials and make the results replicable. The detector readings made up the model's characteristics, and the target variable was the detector areas, which were converted into numerical values for model comity. To take into consideration the craft of the dataset, our particular RFR model was set up with a number of hyper parameters. To balance prognosticated delicacy with computational effectiveness, ten retrogression trees were used.

$$f^{SE}(x) = {}^{1/k} \sum_{k=1}^{K} t\,(x), \qquad (1)$$

where fkSE (x) is the forecast generated by the ith decision tree in the test for a specific sensor reading x acquired from the piezoelectric sensors attached on the simulated head model, and fkSE (x) is the combined regression model's prediction of the Random Forest model based on sensor data. Every tree evaluates the sensor data on its own and makes an impact location prediction. The ensemble's total number of trees is denoted by the parameter K. Each tree helps forecast the impact site in this study, and the ensemble chooses the impact explanation that seems the most likely.
Our main attention is on the crucial aspect of sensor orientation and how it significantly affects the accuracy and effectiveness of the model. To maximize the effectiveness of any prediction model, sensors must be positioned strategically. We intend to learn more about the ideal sensor site by examining feature importance produced by two well-known machine learning approaches, RFR and XGBoost. RFR and XGBoost have fundamentally distinct computational foundations for feature importance. The average impurity reduction brought about by splits on a certain feature across all trees is utilized by RFR to calculate significance.

TABLE II: FOR RFR and XGBOOST MODELS, COMPARITIVE OVERVIEW OF SENSOR IMPORTANCE STATISTICS.

Metric	RFR	XGBoost
Mean	0.041857	0.041857
Standard deviation	0.134540	0.134540
25% quartile	0.012416	0.001840
Count	26.000000	26.000000
Median	0.032124	0.002540
75% quartile	0.031526	0.031435
Maximum	0.523732	0.272515
Minimum	0.002725	0.002430

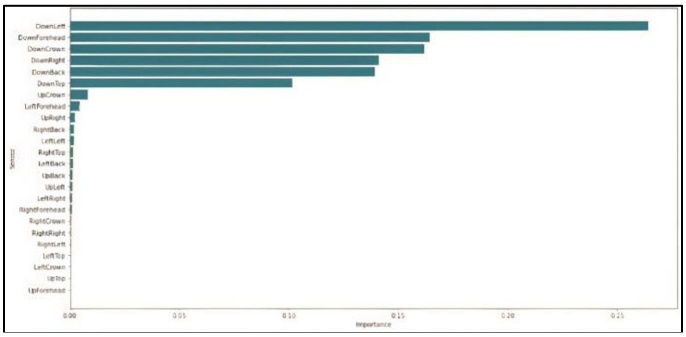


Figure 2.Characteristic Value from XGBoost framework

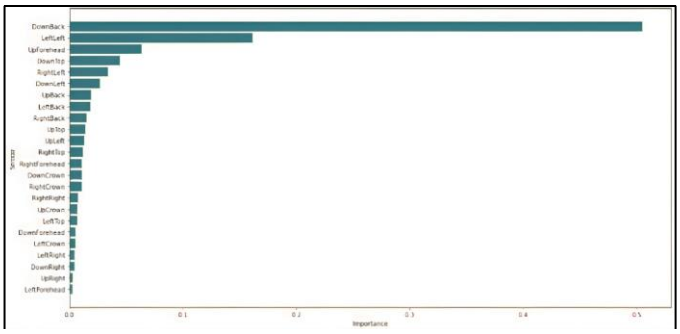
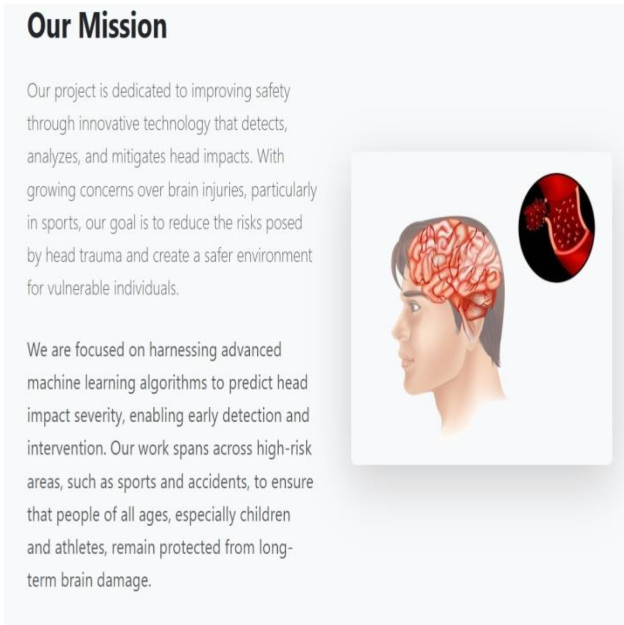
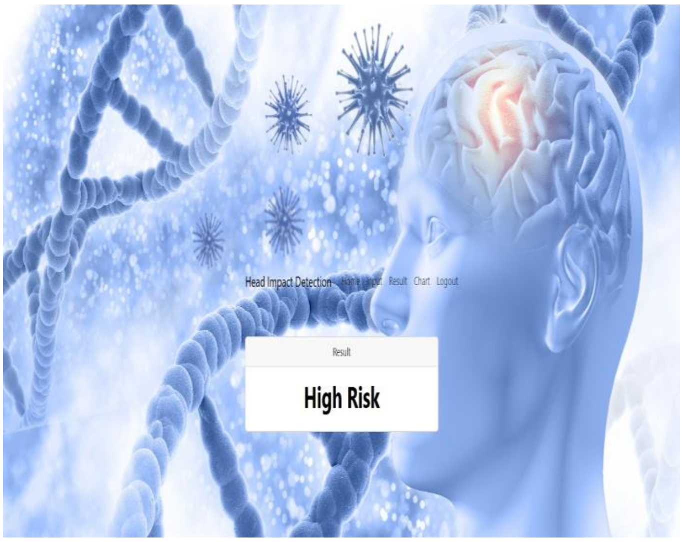


Figure3. Characteristic Value from RFR framework

The output screen as the application is used for the head impact are as follows.



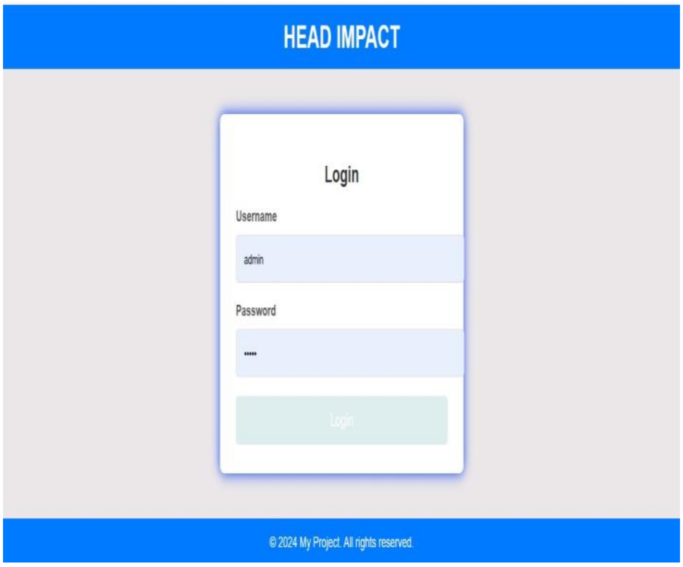
4.(a)



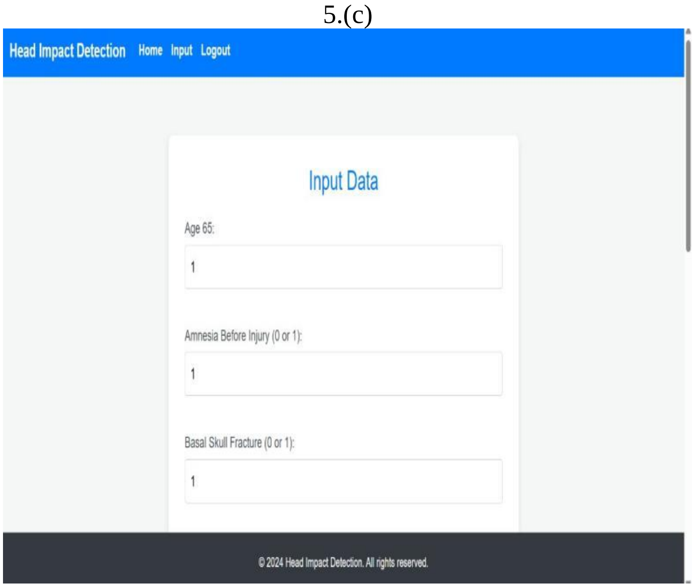
4.(b)

Figure 4.(a) (b) Introduction screens of the application

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 06 | June - 2025 SJIF Rating: 8.586 ISSN: 2582-3930



5.(c)



5.(d)

Figure (c) (d) Login and Input screens of the application respectively

6. CONCLUSIONS

In conclusion, this paper demonstrates the effectiveness of Logistic Regression for accurately detecting head impacts using sensor data, achieving high performance with 90% accuracy. The simplicity, interpretability, and computational efficiency of Logistic Regression make it a promising solution for real-time safety applications, especially in environments where rapid decision-making is critical. The model is well-suited for use in sports and industrial settings where head injuries must be identified quickly to stop additional damage by utilizing a methodical approach that includes data preprocessing, feature engineering, and thorough evaluation. Although the model's performance is already strong, it might be made much more accurate and flexible in the future by adding more sophisticated feature engineering, fine-tuning model parameters, and integrating real-time learning. The results of the study contribute to the broader goal of developing intelligent safety systems that reduce the risks of head impacts and improve overall protective measures in a range of high-risk environments by utilizing wearable technology and machine learning to improve safety and decision-making.

REFERENCES

1. G. D. Myer, W. Yuan, K. D. Barber Foss, S. Thomas, D. Smith, J. Leach,A. W. Kiefer, C. Dicesare, J. Adams, P. J. Gubanich, K. Kitchen,D. K. Schneider, D. Braswell, D. Krueger, and M. Altaye, ‘‘Analysis ofhead impact exposure and brain microstructure response in a seasonlong Application of a jugular vein compression collar: A prospective,neuroimaging investigation in American football,’’ Brit. J. Sports Med.,vol. 50, no. 20, pp. 1276–1285, Oct. 2016.
2. L. C. Wu, C. Kuo, J. Loza, M. Kurt, K. Laksari, L. Z. Yanez, D. Senif,S. C. Anderson, L. E. Miller, J. E. Urban, and J. D. Stitzel, ‘‘Detection of American football head impacts using biomechanical features and support vector machine classification,’’ Sci. Rep., vol. 8, no. 1, p. 855, 2017.
3. X. Wu, J. Zhou, M. Zheng, S. Chen, D. Wang, J. Anajemba, G. Zhang,M. Abdelhaq, R. Alsaqour, and M. Uddin, ‘‘Cloud-based deep learningassisted system for diagnosis of sports injuries,’’ J. Cloud Comput., vol.11, no. 1, pp. 1–18, 2022.
4. L. Gabler, D. Patton, M. Begonia, R. Daniel, A. Rezaei, C. Huber,G. Siegmund, T. Rooks, and L. Wu, ‘‘Consensus head acceleration measurement practices (CHAMP): Laboratory validation of wearable head kinematic devices,’’ Ann. Biomed. Eng., vol. 50, no. 11, pp. 1356–1371, Nov. 2022.
5. L. Zu, J. Wen, S. Wang, M. Zhang, W. Sun, B. Chen, and Z. L. Wang,‘‘Multiangle, self-powered sensor array for monitoring head impacts,’’ Sci.Adv., vol. 9, no. 20, May 2023, Art. no. eadg5152.
6. R. R. V. e. Silva, K. R. T. Aires, and R. d. M. S. Veras, ‘‘Helmet detection on motorcyclists using image descriptors and classifiers,’’ in Proc. 27th SIBGRAPI Conf. Graph., Patterns Images, Aug. 2014, pp. 141–148.
7. J. Li, H. Liu, T. Wang, M. Jiang, S. Wang, K. Li, and X. Zhao, ‘‘Safety helmet wearing detection based on image processing and machine learning,’’in Proc. 9th Int. Conf. Adv. Comput. Intell. (ICACI), Feb. 2017, pp. 201–205.
8. K. Li, X. Zhao, J. Bian, and M. Tan, ‘‘Automatic safety helmet wearing detection,’’ 2018, arXiv:1802.00264.
9. A. Pangestu, M. N. Mohammed, S. Al-Zubaidi, S. H. K. Bahrain, and A. Jaenul, ‘‘An Internet of Things toward a novel smart helmet for motorcycle,’’AIP Conf. Proc., vol. 2320, no. 1, Mar. 2021, Art. no. 050026.
10. N. Dhillon, A. Sutandi, M. Vishwanath, M. Lim, H. Cao, and D. Si,‘‘A raspberry pi-based traumatic brain injury detection system for single-channel electroencephalogram,’’ Sensors, vol. 21, no. 8, p. 2779, Apr. 2021.
11. M. Rezapour, S. Nazneen, and K. Ksaibati, ‘‘Application of deep learning techniques in predicting motorcycle crash severity,’’ Eng. Rep., vol. 2, no. 7, Jul. 2020, Art. no. e12175.
12. M. Ijaz, L. Lan, M. Zahid, and A. Jamal, ‘‘A comparative study of machine learning classifiers for injury severity prediction of crashes involving three-wheeled motorized rickshaw,’’ Accident Anal. Prevention, vol. 154, May 2021, Art. no. 106094. [10]. S. Dongre and J. Agrawal, "Drug Recommendat ion and ADR Detect ion Healthcare in IEEE Transact ions on Computat ional Social Systems, doi: 10.1109/TCSS.2022.3231701.
13. T. K. Ho, ‘‘The random subspace method for constructing decision forests,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 8, pp. 832–844, Aug. 1998.
14. L. Breiman, ‘‘Random forests,’’ Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

15. M. Hanko, M. Grendár, P. Snopko, R. Opšenák, J. Šutovský, M. Benčo, J. Soršák, K. Zeleňák, and B. Kolarovszki, ''Random Forest–based prediction of outcome and mortality in patients with traumatic brain injury undergoing primary decompressive craniectomy,'' World Neurosurg., vol. 148, pp. e450–e458, Apr. 2021.
16. M. Haddouchi and A. Berrado, ''A survey of methods and tools used for interpreting random forest,'' in Proc. 1st Int. Conf. Smart Syst. Data Sci. (ICSSD), Oct. 2019, pp. 1–6.
17. A. Chaudhary, S. Kolhe, and R. Kamal, ''An improved random forest classifier for multi-class classification,'' Inf. Process. Agricult., vol. 3, no. 4, pp. 215–222, Dec. 2016.
18. T. Chen and C. Guestrin, ''XGBoost: A scalable tree boosting system,'' in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2016, pp. 785–794.
19. W. Zhang, C.Wu, H. Zhong, Y. Li, and L.Wang, ''Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization,'' Geosci. Frontiers, vol. 12, no. 1, pp. 469–477, Jan. 2021.
20. J. Zhou, E. Li, H. Wei, C. Li, Q. Qiao, and D. J. Armaghani, ''Random forests and cubist algorithms for predicting shear strengths of rockfill materials,'' Appl. Sci., vol. 9, no. 8, p. 1621, Apr. 2019.
21. T.-T. Wong, ''Performance evaluation of classification algorithms by kfold and leave-one-out cross validation,'' Pattern Recognit., vol. 48, no. 9, pp. 2839–2846, Sep. 2015.