IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 December 2023, accepted 23 December 2023, date of publication 2 January 2024,
date of current version 11 January 2024.

*Digital Object Identifier 10.1109/ACCESS.2023.3349212*

RESEARCH ARTICLE

# Head Impact Detection Using Machine Learning Algorithms

**MOHAMMAD AL BATAINEH[1,2], DANA I. ABU ABDOUN[1], HUDA ALNUAIMI[1], ZOUHAIR AL-QUDAH[3], ZAID ALBATAINEH[4], (Senior Member, IEEE), AND MAHMOUD AL AHMAD[1], (Senior Member, IEEE)**

[1]Electrical and Communication Engineering Department, United Arab Emirates University, Al Ain, United Arab Emirates
[2]Telecommunications Engineering Department, Yarmouk University, Irbid 21163, Jordan
[3]Department of Communication Engineering, Al-Hussein Bin Talal University, Ma'an 71111, Jordan
[4]Electronics Engineering Department, Yarmouk University, Irbid 21163, Jordan

Corresponding author: Mahmoud Al Ahmad (m.alahmad@uaeu.ac.ae)

This work was supported by the UAE University Startup Grant under Fund Code 12N105.

**ABSTRACT** Numerous research studies have emphasized the significance of accurately detecting head impacts and implementing safety measures. This study addresses this crucial need by utilizing machine learning algorithms applied to data from piezoelectric sensors on a simulated head model. Employing a systematic approach, this work utilizes Random Forest (RF) and Extreme Gradient Boosting (XGBoost) models to process the normalized sensor data, aiming to pinpoint impact locations with high precision. Through rigorous k-fold cross-validation and comprehensive performance analysis, the study reveals that the XGBoost model slightly outperforms the RF model, achieving an RMSE of 0.4764 and an R^2 of 0.9485. Feature importance evaluations suggest an optimal sensor placement strategy, potentially reducing the model complexity while retaining predictive accuracy. The superior performance of the XGBoost model, combined with strategic sensor placement, highlights the study's contribution to enhancing head impact safety measures in sports and industrial settings. The findings pave the way for future research into the deployment of intelligent safety systems, leveraging the synergy between wearable technology and machine learning.

**INDEX TERMS** Head impact detection, wearable technology, machine learning, injury prevention, piezoelectric sensors, random forest, eXtreme gradient boosting (XGBoost), predictive modeling.

## I. INTRODUCTION

The significance of head impact detection and the development of safety precautions have been the focal point of numerous studies over the years. The dynamics of head impacts and the urgency to devise systems that can detect and prevent injuries are paramount. This section highlights the convergence of wearables, machine learning, and technology by synthesizing significant and recent developments in this field.

Numerous research have examined the biomechanics of head impacts in the context of sports. Smith et al. [1] utilized instrumented helmets to study the frequency, size, and location of head hits in ice hockey. Defensemen were found to have suffered the most serious head injuries, highlighting the necessity of position-specific preventative measures. Likewise, Mori et al. [2] and Zhang et al. [3] investigated how machine learning techniques may be used to identify brain injuries in American football. While Mori et al. utilized wearable sensors combined with Random Forest and Gradient Boosting machine algorithms, Zhang et al. employed a blend of deep learning and traditional machine learning techniques, achieving an astounding accuracy of 99.3%. These studies collectively accentuate the potential of machine learning in real-world sports scenarios and the importance of sport-specific research to devise tailored protective strategies.

Another key area of attention has been the use of wearable technology in the measurement of actual head impacts. Gabler et al. [4] highlighted the role of wearable devices in understanding brain injury mechanisms. Their work aimed to standardize validation test methods and data reporting for wearable head kinematic devices. In a related vein,

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

Zu et al. [5] introduced a novel wearable electronic device, which can convert impact forces from various directions into electrical signals. This device's integration with machine learning algorithms enables it to assess injury grades with an accuracy of 98%, emphasizing the potential of such devices in collecting standardized data for research into head impacts and mild concussions.

Helmet compliance and detection have also been a significant area of research. Silva et al. [6] designed a system to detect motorcyclists not wearing helmets, achieving an accuracy rate of 91.37%. In industrial settings, Li et al. [7] and [8] proposed methods and systems to detect the wearing of safety helmets using image processing and machine learning techniques. These studies underscore the importance of helmet compliance across different settings, from road safety to industrial environments.

On a broader scale, the integration of technology in safety gear has been explored. Pangestu et al. [9] presented a comprehensive review of the Internet of Things (IoT) in developing a novel smart helmet for motorcyclists. This helmet, integrated with IoT technology, aims to enhance rider safety by detecting and preventing potential accidents in real-time. Similarly, Dhillo et al. [10] introduced a smart helmet prototype for the Industrial IoT using machine learning algorithms, emphasizing the role of information and communication technologies in enhancing occupational health and safety.

Lastly, in the domain of traffic safety, Rezapour et al. [11] and Ijaz et al. [12] utilized machine learning and deep learning techniques to predict motorcycle crash severity and analyze the injury severity of crashes involving three-wheeled motorized rickshaws, respectively. These studies highlight the evolving nature of predictive modeling in traffic safety and the potential for more accurate analyses.

The remainder of the paper is structured as follows: Section II details the methodology, encompassing dataset preparation and algorithmic specifics of Random Forest and XGBoost. Section III addresses model evaluation, highlighting k-fold cross-validation and performance metrics. Section IV analyzes optimal sensor placement, underscoring its significance in model efficacy. Section V presents the main results, comparing the performances of the two models. The paper concludes in Section VI, summarizing key findings and proposing directions for future research. References cited are listed thereafter.

## II. STUDY CONTEXT

Understanding the response and dynamics of impact on the human head is crucial in the realm of safety and injury prevention. While there are various methods to measure these impacts, the utilization of piezoelectric sensors has emerged as a promising technique due to their sensitivity and rapid response time. In this study, we aimed to create a predictive model that can accurately determine the region or regions of impact on a plastic head model, simulating real-life scenarios.

Given the shape and curvature of the head, it is important to note that an impact on one region can influence readings on sensors placed in other regions. Therefore, we strategically placed sensors on different regions of the head and recorded readings after simulating impacts. By understanding these readings and their interplay, we can potentially pinpoint the location of impact with greater precision, contributing to the overall safety and protective measures in real-world applications.

### A. DATASET DESCRIPTION

Data was carefully collected using twenty-four piezoelectric sensors, strategically positioned on various regions of a plastic head model. The regions under study encompass the forehead, left side, right side, back, top, and crown. In each designated region, sensors were oriented in an approximate diamond shape, capturing readings from the top (up), right, bottom (down), and left positions. Impacts were simulated on specific regions, and readings were recorded. These readings, captured within specific time intervals, are visualized in Figures (2–7), offering insights into force dynamics and impact characteristics. Figure 1 shows a plastic head model with piezoelectric sensors distributed in several locations.

**FIGURE 1.** Plastic head model with piezoelectric sensors distributed in several locations.

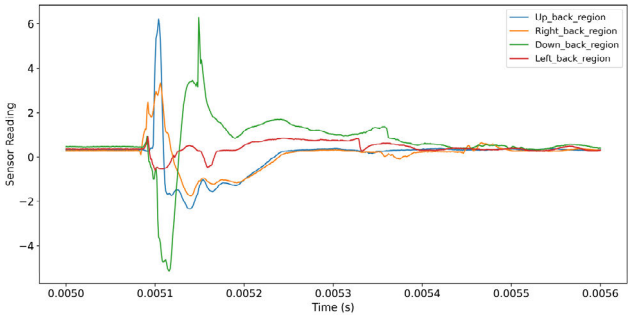


**FIGURE 2.** Back region sensor readings.

## III. METHODOLOGY

### A. MODEL PREPARATION AND PREPROCESSING

The dataset provided was subjected to preprocessing to ensure the accuracy and efficiency of the subsequent machine learning models. This involved data normalization, where

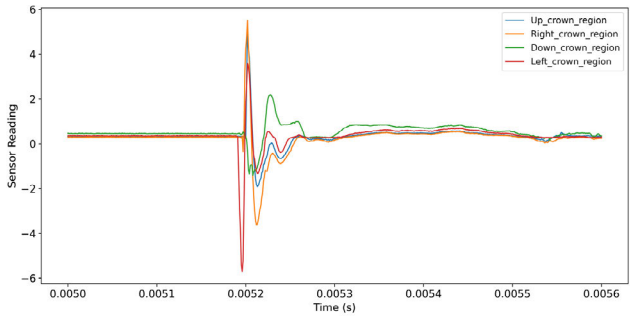


**FIGURE 3.** Crown region sensor readings.

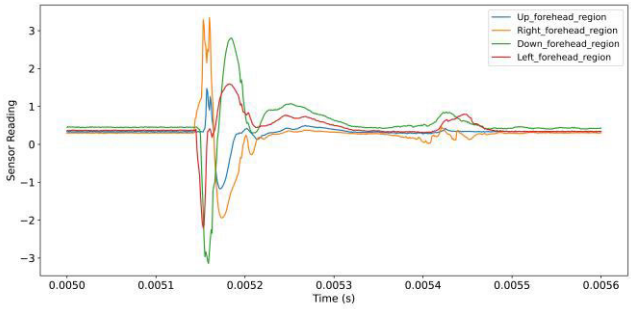


**FIGURE 4.** Forehead region sensor readings.

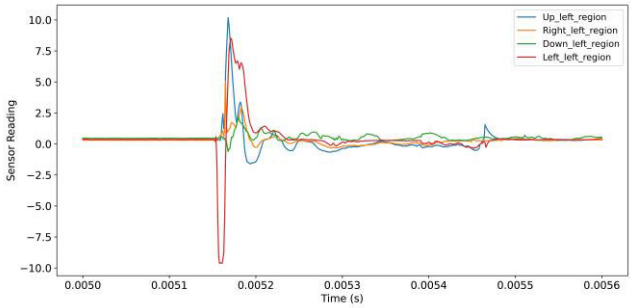


**FIGURE 5.** Left region sensor readings.

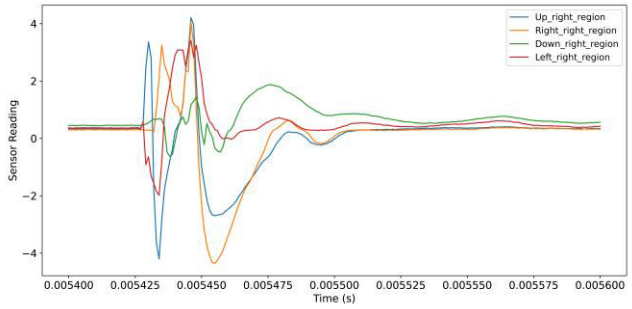


**FIGURE 6.** Right region sensor readings.

the feature variables were transformed into a standard scale, ensuring that they have a mean of 0 and a standard deviation of 1. This preprocessing step is critical for many machine learning algorithms to converge faster and produce more accurate models.

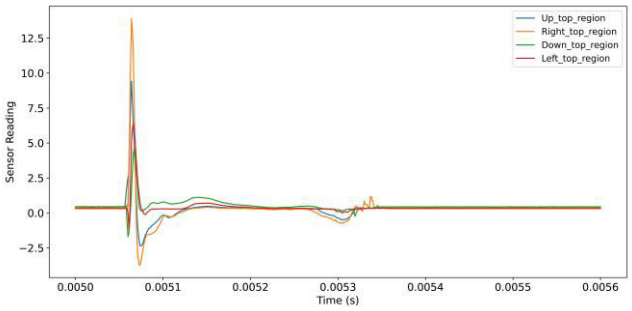


**FIGURE 7.** Top region sensor readings.

### B. RANDOM FOREST

The Random Forest (RF) algorithm is a versatile ensemble learning method grounded in the principles of leveraging multiple decision trees for tasks like classification and regression. The algorithm's genesis traces back to the pioneering work of Ho in 1995 [13], with subsequent expansions by Breiman in 2001 [14]. Recognized for its consistent performance across diverse applications, the RF algorithm has become synonymous with delivering exceptional results for both classification and regression tasks. Central to RF's methodology is the bootstrap resampling technique. This approach involves generating multiple samples from the original dataset, leading to the formation of a "forest" of

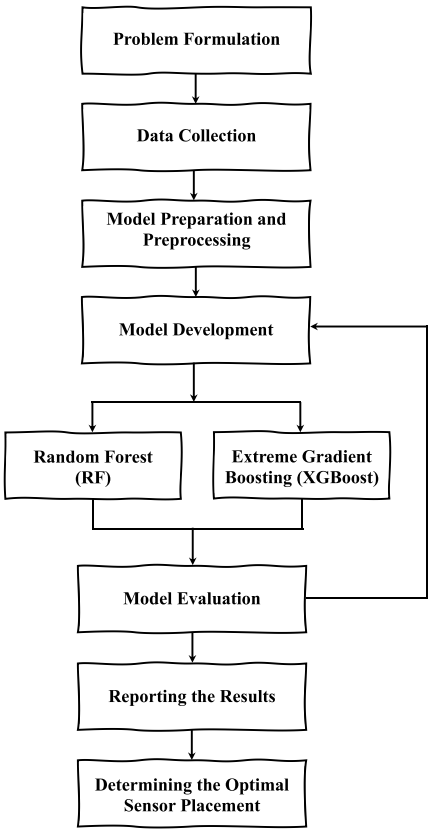


**FIGURE 8.** Proposed predictive model development process.

decision trees [15]. Each tree in this forest is characterized by a series of nodes, where every node represents a decision criterion based on a specific variable [16]. The predictive power of RF stems from aggregating the outcomes of all these individual trees. In practice, each tree offers its prediction (or ''vote''), and the majority consensus dictates the final prediction.

A defining characteristic of the RF algorithm is the utilization of consistent subsampling or bootstrapping for each tree's evolution [14], [16], [17]. A notable portion, approximately one-third, of the data remains unsampled during each tree's bootstrap process. This unsampled subset is labeled as the out-of-bag (OOB) data, providing a unique mechanism for internal validation.

Within the framework of our study, the RFR, a regression-centric variant of RF, becomes especially relevant. Given the intricate and potentially non-linear nature of our sensor data, RFR stands out as an optimal choice. The methodology involved processing a dataset consisting of various sensor readings, segmented into regions such as the forehead, back, crown, left, right, and top. The features for the model comprised these sensor readings, while the target variable was the sensor regions, encoded into numerical values for model compatibility. Our specific RFR model was configured with a series of hyperparameters to cater to the dataset's nuances. The number of regression trees was set to 10, striking a balance between computational efficiency and predictive accuracy. Additionally, a random state was set at 42 to ensure consistency across different runs, making the results reproducible.

$$f_K^{\mathrm{RF}}(x) = \frac{1}{K}\sum_{k=1}^{K} t_i(x), \tag{1}$$

where $f_K^{\mathrm{RF}}(x)$ is the combined regression model's prediction of the Random Forest model based on sensor data, $t_i$ denotes the prediction made by the $i^{th}$ decision tree in the forest for a given sensor reading $x$ derived from the piezoelectric sensors placed on the simulated head model. Each tree independently assesses the sensor data and predicts the impact location. The parameter $K$ represents the total number of trees in the ensemble. In this study, each tree contributes to predicting the impact location, and the ensemble collectively determines the most likely impact region.

### C. EXTREME GRADIENT BOOSTING (XGBOOST) ALGORITHM

Extreme Gradient Boosting (XGBoost) is an advanced and efficient algorithm under the Gradient Boosting framework. This algorithm, initially introduced by Chen and Guestrin [18], has gained immense recognition in various application. This can be attributed to its high efficiency and commendable flexibility. One of the defining features of XGBoost is its unique loss function, which incorporates an added regularization term. This term aids in refining the final learned weights, thereby mitigating the risk of overfitting [19]. The objective function, that XGBoost seeks to minimize, can be mathematically represented as:

$$J(\theta) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \Omega(\theta), \tag{2}$$

where $\sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right)$ is the training loss, where $y_i$ is the actual sensor reading, and $\hat{y}_i$ is the predicted reading. It measures how well the model predictions align with the actual data. The term $\Omega(\theta)$ is a regularization component of the objective function that penalizes the complexity of the model, helping to prevent overfitting [18]. This regularization is further broken down as:

$$\Omega(\theta) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2, \tag{3}$$

where $T$ symbolizes the number of leaves in each tree of the XGBoost model, $w_j$ is the weight associated with the $j^{th}$ leaf, and $\gamma$ and $\lambda$ are hyperparameters that control the degree of regularization [18].

Furthermore, XGBoost harnesses both first and second-order gradient statistics to fine-tune the loss function [19]. In our study, the XGBoost model was configured with the objective parameters set to 'multi:softprob', signifying the use of the softmax objective tailored for multi-class classification. Although explicit regularization parameters were not provided, the algorithm employed its default values, ensuring a balance between bias and variance. As for the row and column sampling, while XGBoost inherently supports this feature to prevent overfitting, our model relied on the default sampling parameters. This decision was grounded in preliminary testing that revealed optimal performance with these settings. Moreover, our implementation leveraged the built-in parallel computational capabilities of XGBoost, facilitating faster tree construction and, by extension, quicker model exploration.

A notable aspect of our implementation is the specification of hyperparameters. The model was set with $n_{estimators} = 100$, dictating the construction of 100 trees. Additionally, a seed value of 42 was designated to ensure reproducibility across multiple runs. Other parameters, such as the maximum depth of the trees, utilized the default values provided by the XGBoost library. Our approach predominantly focused on the practical application of XGBoost. The process entailed encoding labels, training the classifier on resampled data, and subsequently evaluating its predictions. Not only did the model showcase its ability on the test dataset, but its capability was also tested on synthesis data, further emphasizing its robustness.

## IV. MODELS EVALUATION

### A. K-FOLD CROSS-VALIDATION

The imperative of model evaluation cannot be overstressed in predictive modelling tasks. The choice of evaluation strategy can significantly influence the perceived performance of a model. In this study, the application of a k-fold cross validation was deemed appropriate. The dataset was meticulously divided into three distinct subsets: (i) training, (ii) validation, and (iii) testing. The training subset was specifically set aside

for the purpose of model fitting. The validation subset, on the other hand, played a pivotal role in hyperparameter tuning and model selection. The testing subset was reserved for the final evaluation, aimed at gauging the model's capacity for generalization. Given the voluminous nature of the dataset, a 5-fold cross validation was adopted, drawing inspiration from seminal works by Zhou and Wong [20], [21]. This approach not only minimizes bias but also offers a holistic assessment of the model's predictive capabilities.

### B. MODEL EVALUATION AND PERFORMANCE MEASURES

Evaluating the performance of predictive models is a pivotal step in determining their efficacy and accuracy. The root mean square error (RMSE), coefficient of determination (R2), mean square error (MSE), and mean absolute error (MAE) were employed to assess the predictive capabilities of the RFR and XGBoost models. These metrices, as delineated in Equations (4)-(7), offer a comprehensive perspective on how close the predictions are to the actual values. RMSE measures the average magnitude of the errors between predicted and observed values. A lower RMSE, closer to zero, is indicative of a model that can make predictions with greater precision [22]. Whereas, $R^2$ represents the proportion of the variance in the dependent variables that is explained by independent variables in a regression model. An $R^2$ closer to 1 suggests that the model accounts for a significant proportion of the variance in the target variable [23], [25]. On the other hand, MSE calculated the average of the squares of the errors or deviations. A lower MSE value indicates a model with better predictive accuracy [22], [25]. Lastly, MAE computes the average of the absolute difference between predicted and actual values. A smaller MAE indicates that the model can predict closer to the true values [24], [25].

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}, \tag{4}$$

$$R^2 = 1 - \frac{\sum\left(y_i - \hat{y}_i\right)^2}{\sum\left(y_i - \bar{y}\right)^2}, \tag{5}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2, \tag{6}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|, \tag{7}$$

For a detailed understanding, the results of RMSE, $R^2$, MAE, MSE for both the RFR and XGBoost models are presented in Table 1. The table presents the performance measures, wherein the RFR model exhibited an RMSE of 0.4855 and an $R^2$ of 0.9466, while the XGBoost model rendered and RMSE of 0.4764 and an $R^2$ of 0.9485. These figures bear testimony to the robustness and precision of both models. The XGBoost model, however, displayed a marginally superior performance across all metrics, underscoring its adaptability and potency in handling complex datasets.

Furthermore, the scatterplots provide a visual representation of the model's performance, comparing actual versus predicted values. The proximity of the data points to the line of unity indicates the accuracy of predictions. An examination of the scatterplots (Figures 9 and 10) for both RFR and XGBoost models reveals that most of the predicted values closely align with the actual ones. This graphical representation not only validates the performance metrics discussed but also offers an intuitive understanding of the model's predictive capabilities.

**TABLE 1. Prediction models accuracy evaluation.**

| Performance measure | Random Forest model | XGBoost model |
|---|---|---|
| MSE | 0.2357 | 0.2270 |
| RMSE | 0.4855 | 0.4764 |
| MAE | 0.0343 | 0.0169 |
| R2 | 0.9466 | 0.9485 |

In contrast, the XGBoost model demonstrates a more balanced performance across all sensor regions, as evident from its confusion matrix. The model's predictions align more closely with the actual sensor readings, suggesting that it has a better grasp of the underlying patterns in the data. Further evaluation on the XGBoost model also indicates a more consistent performance between training and testing datasets. This consistency is a positive sign, indicating that the model is less prone to overfitting and is more likely to provide reliable predictions on new, unseen data. Nonetheless, further evaluation was performed on the models to measure their performances, Table 2 provides a comprehensive summary of the two models based on the key performance metrics. It is evident from the table that the XGBoost model outperforms the Random Forest model in terms of accuracy. However, both models showcase commendable precision, recall, and F1-score values, indicating their reliability in predicting the sensor regions.

**TABLE 2. Units for magnetic properties.**

| Performance measure | Random Forest model | XGBoost model |
|---|---|---|
| Accuracy | 99.41% | 99.76% |
| Precision | 99.42% | 99.76% |
| Recall | 99.41% | 99.76% |
| F1-score | 99.41% | 99.76% |

## V. OPTIMAL SENSOR PLACEMENT

In this section, our primary focus is on the pivotal aspect of sensor placement and its profound implications on model accuracy and efficiency. The strategic positioning of sensors is fundamental to optimizing the performance of any predictive model. By analyzing feature importances yielded by two prominent machine learning algorithms, RFR and XGBoost, we aim to infer insights for optimal sensor placement.

The computational underpinning of feature importance in both RFR and XGBoost are inherently different.

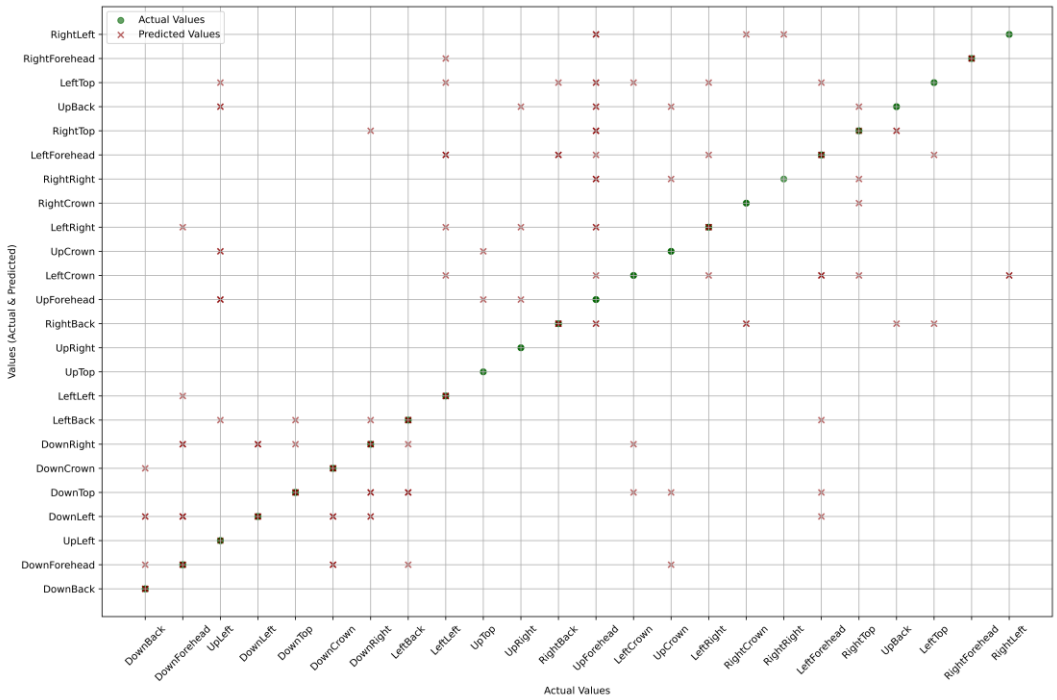


**FIGURE 9.** Scatterplot showcasing the actual vs. predicted values for the RFR model.

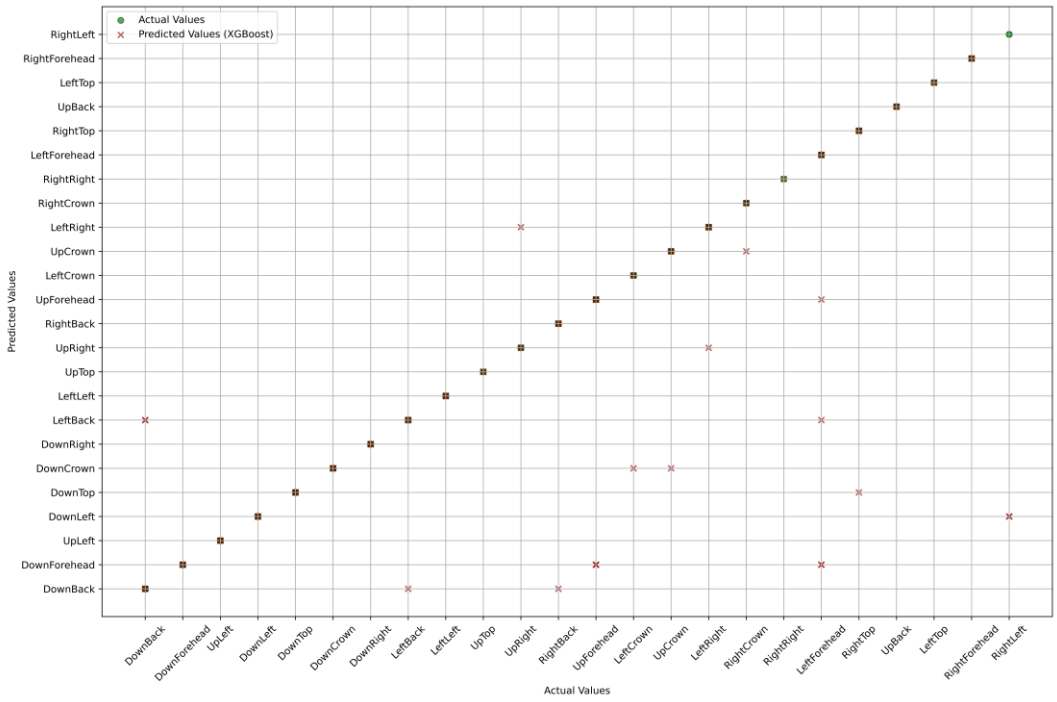


**FIGURE 10.** Scatterplot illustrating the actual vs. predicted values for the XGBoost model.

RFR computes importance based on the average reduction in impurity caused by splits on a particular feature across all trees. The mathematical representation of this process can be articulated using (8), in which $N$ represents the total number of trees in the Random Forest, $t$ is a specific tree, and $\Delta G(t, f)$ is the improvement in the Gini impurity attributed to feature $f$ in tree $t$.

$$Importance(f) = \frac{1}{N} \sum_{t=1}^{N} \Delta G(t, f), \tag{8}$$

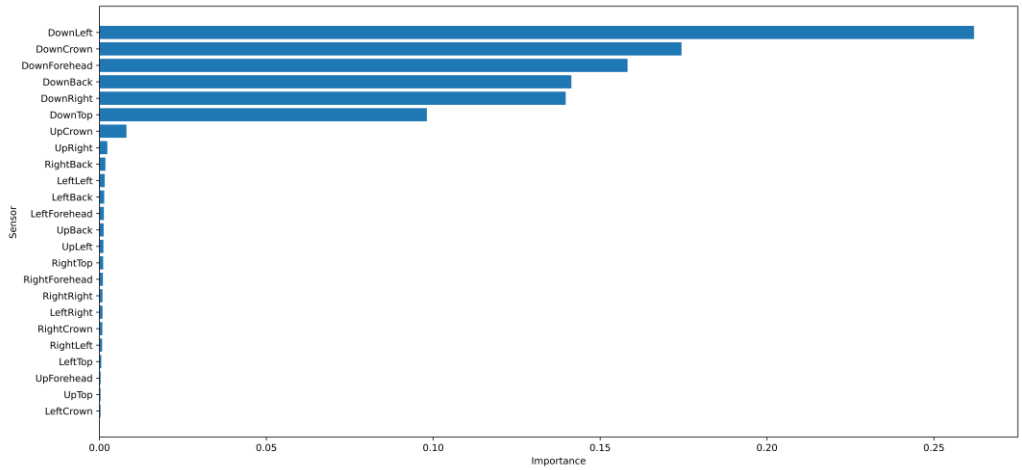


**FIGURE 11.** Feature Importance from XGBoost model.

Contrariwise, XGBoost computes feature importance predicted on the frequency of a feature's appearance across all trees. The mathematical delineation of this procedure can be expressed using (9). The $\Delta G(t, f)$ in (9) represents the gain associated with feature $f$ in tree $t$, reflecting its contribution to the model's performance, and Count($f$) is the number of times feature $f$ appears in the model.

$$Importance\,(f) = \frac{\sum_{t=1}^{N} \Delta G\,(t, f)}{\text{Count}\,(f)}. \tag{9}$$

### A. DESCRIPTIVE STATISTICS OF SENSOR IMPORTANCE

Table 3 delineates the comparative descriptive statistics of sensor importance as determined by the RFR and XGBoost models. Although both models analyzed an identical set of 24 sensors, discernible differences emerge in their perceived importance distributions. Notably, the RFR model's maximum importance value markedly surpasses that of XGBoost, highlighting the distinct weighting methodologies inherent to each model.

**TABLE 3.** Comparative descriptive statistics of sensor importance for RFR and XGBoost models.

| Metric | RFR | XGBoost |
|---|---|---|
| Count | 24.000000 | 24.000000 |
| Mean | 0.041667 | 0.041667 |
| Standard deviation | 0.104040 | 0.075394 |
| Minimum | 0.002625 | 0.000340 |
| 25% quartile | 0.006116 | 0.000740 |
| Median | 0.011224 | 0.001540 |
| 75% quartile | 0.020626 | 0.031424 |
| Maximum | 0.504732 | 0.264015 |

### B. FEATURE IMPORTANCE ANALYSIS

As we delved deeper into the sensor feature importance, a richer grasp of their nuanced methodologies came to the fore. Figure 11, elucidating the feature importances as determined by the XGBoost algorithm showcases a notably narrow band of importance scores for a vast majority of sensors. Particularly discernible are sensors such as down sensor in left region, down sensor in forehead region, down sensor in crown region, down sensor in right region, and down sensor in back region, which exhibit markedly elevated importance values. This underscores their pivotal roles in XGBoost's analytical framework. On the other end of the spectrum, numerous sensors gravitate towards the lower echelons of importance, hinting at their more marginal roles in the model's predictive endeavors. The narrow dispersion of scores underscores the equitable distribution of feature contributions inherent to the XGBoost mechanism.

On the other hand, Figure 12, associated with the RFR model offers a richer tapestry of feature importance. Sensor, namely left sensors in back region, left sensor in left region, and up sensor in forehead region, stand out with prominently high importance values. Of particular note is the down sensor in back region, whose value approaches the 0.5 threshold, accentuating its paramountcy in RFR's inferential logic. Contrasted with its XGBoost counterpart, RFR presents a broader spectrum of importance values, signifying its adeptness at harnessing and interpreting intricate data patterns. The pronounced interquartile range observed further amplifies this assertion, bearing testament to the algorithm's sophisticated handling of feature nuances.

### C. CUMULATIVE FEATURE IMPORTANCE: OPTIMAL NUMBER OF SENSORS

In Figure 13, the cumulative feature importance for both RFR and XGBoost is plotted against the number of sensors used. For the XGBoost algorithm, the curve shows a steep increase at the start, indicating that significant predictive power is obtained with the inclusion of only a few sensors. Specifically, the model rapidly approaches a high level of cumulative importance with the first 10 sensors, suggesting that these sensors capture the majority of the predictive power. In contrast, the curve for the RFR model is more gradual, suggesting

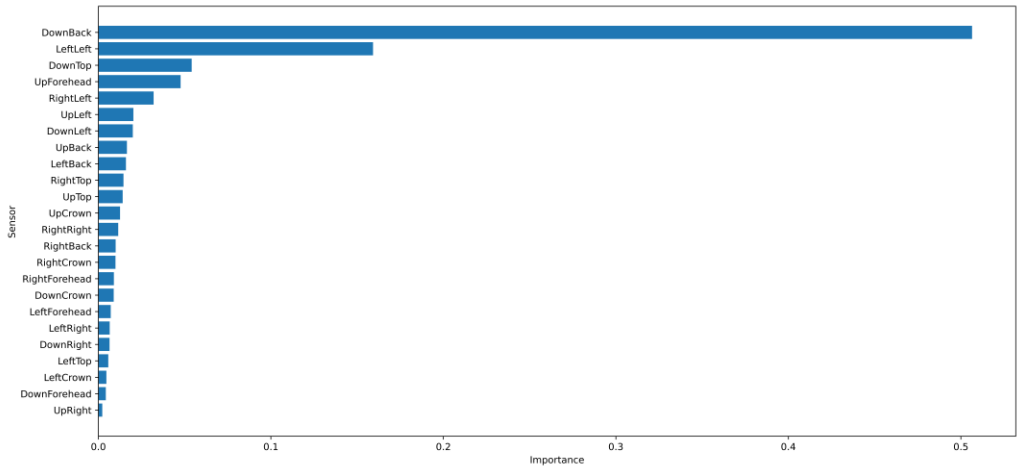


**FIGURE 12.** Feature Importance from RFR model.

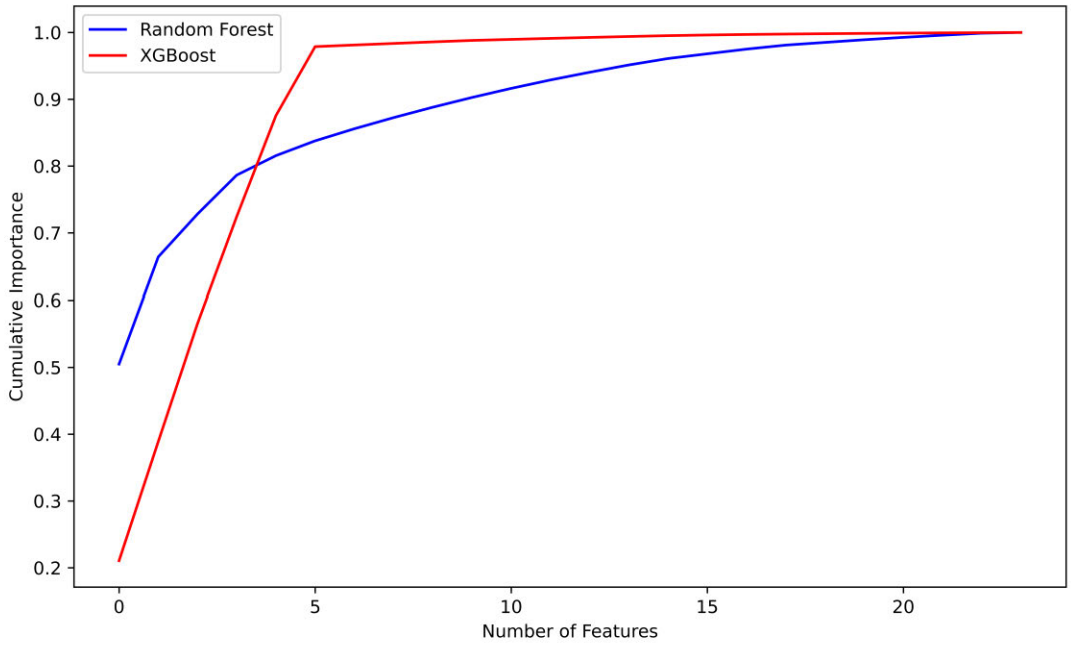


**FIGURE 13.** Cumulative feature importance with respect to the number of sensors.

a more even distribution of feature importance across the sensors. While the RFR model also gains a significant amount of predictive power from the first 10 sensors, it continues to integrate additional sensors more smoothly, without the abrupt plateau observed in the XGBoost curve.

This indicates that while both models benefit from the initial set of sensors, XGBoost concentrates more importance in the first few sensors compared to RFR. Consequently, for deployments focusing on resource efficiency, XGBoost may achieve satisfactory predictive accuracy with fewer sensors, whereas RFR would benefit from a larger sensor set to capture more nuanced patterns in the data.

## VI. MAIN RESULTS

Following the comprehensive evaluation and comparison methodologies delineated in the preceding sections, the results reported in Table 1 provide a clear exposition of the performance dynamics between the RFR and XGBoost models in the context of head impact detection.

The RFR model delivered an RMSE value of 0.4855, signifying that on average, the deviation between the model's predictions and the actual values is approximately 0.4855 units. As it is well-established in machine learning literature, an RMSE value that gravitates closer to zero typically denotes a model with heightened predictive precision [22]. The XGBoost model, registering an RMSE of 0.4764, subtly but distinctly outperforms the RFR in this metric, hinting at its superior predictive acuity.

Diving deeper into the $R^2$, the RFR model's value of 0.9466 suggests that about 94.66% of the variability in the dependent variable is elucidated by the model. In comparison, the XGBoost model, with its R2 of 0.9485, accounts

for approximately 94.85% of this variability. It is a canonical understanding that an $R^2$ value inching closer to 1 is emblematic of a model's adeptness at predicting the dependent variable, given the independent variables.

Further analysis of the MSE and MAE metrics repeats similar insights. The XGBoost model, with its marginally lower MSE and MAE values, exhibits a nuanced edge over the RFR model, reinforcing its aptitude in handling datasets of the nature explored in this research. Figures 9 and 10, the scatterplots, visually accentuate these analytical insights. A model's predictive ability becomes profoundly evident when its scatterplot data points coalesce tightly around the line of unity. Both the RFR and XGBoost models compellingly exhibit this trait, highlighting the consistency and reliability of their predictions.

Upon examining Figures 9 and 10, the scatterplots, we discern a marked difference in the predictive capabilities of the two models. In the scatterplots, the green circles represent the actual values and serve as a visual reference. A model with perfect prediction would result in the dark red 'x' markers (representing the predicted values) aligning perfectly with these green circles. The XGBoost scatterplot displays dark red 'x' markers that cluster closely around the green circles, visually affirming its exceptional predictive ability. This convergence of predicted values to the actual ones is indicative of the model's accuracy and reliability. In contrast, the RFR model's scatterplot exhibits a more dispersed spread of the dark red 'x' markers, hinting at a less consistent prediction paradigm compared to the XGBoost model. This indicates that while the RFR model showcased commendable predictive capabilities, the XGBoost model emerged as the superior contender across all performance metrics.

## VII. CONCLUSION

In this work, a comprehensive investigation into head impact detection has been conducted, utilizing the strengths of wearable sensor technology and advanced machine learning algorithms. The integration of piezoelectric sensors on a plastic head model has enabled the simulation of real-life impacts, facilitating the accurate prediction of impact regions. The application of Random Forest and eXtreme Gradient Boosting (XGBoost) models has proven to be instrumental in deciphering the complex patterns associated with head impacts, with XGBoost exhibiting a slight edge in predictive accuracy and reliability.

The study has underscored the critical importance of meticulous data preprocessing, including normalization, to optimize the performance of the machine learning models. Rigorous evaluation using k-fold cross-validation and various performance metrics has established the reliability and robustness of the proposed models, particularly highlighting the superior performance of the XGBoost model.

Additionally, the manuscript has addressed the crucial aspect of sensor placement, employing the XGBoost algorithm to quantitatively assess the importance of each sensor. This analysis not only contributes to the field of impact detection but also provides valuable guidance for future endeavors aimed at optimizing sensor configurations.

In light of the broader implications of this work, the manuscript makes a substantial contribution to the domains of safety and injury prevention, especially in sports and industrial settings. The convergence of wearable technology and machine learning, as demonstrated in this research, opens up new possibilities for more nuanced and accurate head impact detection, potentially leading to the development of more effective protective measures.

Looking forward, the insights derived from this study lay a solid foundation for future research, with potential avenues including the exploration of additional machine learning algorithms, the optimization of sensor placements, and the extension of these models to diverse application scenarios. The ultimate goal remains to enhance our understanding of head impact dynamics, provide robust predictive models, and contribute significantly to the overarching objective of enhancing safety and preventing injuries in various contexts.

## 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 of head impact exposure and brain microstructure response in a season-long 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 learning-assisted 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.

[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 k-fold and leave-one-out cross validation," *Pattern Recognit.*, vol. 48, no. 9, pp. 2839–2846, Sep. 2015.
[22] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. Melbourne, VIC, Australia: OTexts, 2018.
[23] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models*. New York, NY, USA: McGraw-Hill, 2005.
[24] C. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Res.*, vol. 30, no. 1, pp. 79–82, 2005.
[25] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 112. Cham, Switzerland: Springer, 2013.

**MOHAMMAD AL BATAINEH** received the B.S. degree (Hons.) in telecommunications engineering from Yarmouk University, Jordan, in 2003, and the M.S. and Ph.D. degrees in electrical engineering from the Illinois Institute of Technology (IIT), USA, in 2006 and 2010, respectively. Subsequent to his academic pursuits, he held noteworthy positions at institutions, including Yarmouk University, where he was promoted to an Associate Professor, in 2018, and roles with Argonne National Laboratories and MicroSun Technologies. In August 2020, he joined United Arab Emirates University (UAEU), as an Assistant Professor. His research interests include the application of communications, coding theory, and information theory to the interpretation and understanding of information flow in biological systems, particularly gene expression. His additional research avenues encompass machine learning, network information theory, and optimization.

**DANA I. ABU ABDOUN** received the B.Sc. and M.Sc. degrees in industrial engineering and engineering management from the University of Sharjah, United Arab Emirates, in 2018 and 2022, respectively. She is currently a Research Assistant with the College of Engineering (COE), specifically the Electrical and Communications Engineering Department, United Arab Emirates University. Her research interests include data mining, machine learning, additive manufacturing, and supply chain management.

**HUDA ALNUAIMI** received the B.S. degree in electrical engineering from United Arab Emirates University, Al Ain, United Arab Emirates, in 2019, where she is currently pursuing the M.S. degree in electrical engineering. Her research interests include electronic and sensor systems.

**ZOUHAIR AL-QUDAH** received the B.Sc. degree in electrical engineering from Yarmouk University, Jordan, in September 2002, the M.Sc. degree in electrical engineering from Kalmar University College, Sweden, in November 2006, and the Ph.D. degree in electrical engineering from Southern Methodist University, Dallas, TX, USA, in May 2013. Since August 2013, he has been with Al-Hussein Bin Talal University, Ma'an, Jordan, where he is currently a Professor of wireless communications. His research interests include coding and information theory in wireless communication systems, cooperative communications 5G communication, resource allocation problems, interference management, and signal processing.

**ZAID ALBATAINEH** (Senior Member, IEEE) received the B.S. degree in electrical engineering from Yarmouk University, Irbid, Jordan, in 2006, the M.S. degree in communication and electronic engineering from the Jordan University of Science and Technology, Irbid, in 2009, and the Ph.D. degree from the Electrical and Computer Engineering Department, Michigan State University, East Lansing, MI, USA, in 2014. He is currently an Assistant Professor with the Electronics Engineering Department, Yarmouk University. His research interests include signal processing for energy conversion systems, blind source separation, independent component analysis, and RF and analog integrated circuit.

**MAHMOUD AL AHMAD** (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from Birzeit University, Ramallah, Birzeit, in 1999, and the M.Sc. and Ph.D. degrees in microwave engineering from Technische Universität München, Munich, Germany, in 2002 and 2006, respectively. He is currently a Faculty Member with the Department of Electrical Engineering, United Arab Emirates University. He has also conducted research in energy harvesting technologies and frequency agile circuits with Siemens AG/CNRS and the King Abdullah University of Science and Technology (KAUST). He has over ten years of electronic materials and device fabrication research experience in academia, national laboratories, and industry. He has managed several research projects and teams with annual budgets of up to U.S. $1 million. He has published more than 20 journal articles and 35 conference papers in this domain with more under review. He is a principle (lead) author of around 55 published papers in journals and international conferences and has over 40 presentations at international conferences (several of which has been invited). His research interests include the design and fabrication of self-powered, low powered nano-based electronic devices and systems, and applied electromagnetic for biomedical applications.