

Novel Sonification Methods for Big Data Pattern Recognition: Example from Detecting “Aberrancy” in Standardized Tests

Asiye Şengül Avşar
Dept. of Educational Sciences
Recep Tayyip Erdoğan University
Rize, Türkiye
0000-0001-5522-2514

Bruce N. Walker
School of Psychology
Georgia Institute of Technology
Atlanta, USA
0000-0001-8503-5621

Abstract— As the volume and density of data grow exponentially with the advent of IoT, the prevalence of noise, as well as incomplete or anomalous records within these data streams, also rises. Fraudulent or aberrant data, which falls outside the scope of normal data, can lead to flawed decision-making processes. Various machine learning (ML) algorithms successfully carry out fraud detection processes. In this paper we demonstrate how we can integrate sophisticated and creative sonification approaches into ML pipelines for fraud or aberrancy detection, to improve overall performance. We focus on detecting “aberrant” scores in standardized tests. This study demonstrates that sonification, in conjunction with traditional ML methods, is effective in identifying individuals with aberrant item scores. We recommend the combined use of modern sonification approaches and traditional fraud detection methods when dealing with big data to facilitate more accurate decision-making.

Keywords— fraud, anomaly detection, machine learning models, data sonification

I. INTRODUCTION

The Internet of Things (IoT) is a global network system in which billions of physical devices worldwide connect to the internet to collect and share data. This concept represents a vision where objects via embedded devices sense their surroundings, communicate, and behave like intelligent entities [1]. IoT is defined as a global physical network where objects seamlessly connect to internet infrastructure to communicate or interact with both their internal and external environments [2]. IoT technology enhances operational efficiency, enables real-time monitoring, and facilitates daily life by supporting human activities [3, 4]. Achieving this requires devices to provide uninterrupted, real-time data [4]. A natural consequence of this is the generation of massive data streams. These streams are transmitted to cloud computing platforms, where the data is analyzed using advanced algorithms and artificial intelligence (AI), thereby enabling systems to make autonomous decisions or provide valuable insights to users [4, 5]. As the volume and density of data grow exponentially with the advent of IoT, the prevalence of noise, as well as incomplete or anomalous records

within these data streams, also rises. Whether dealing with energy consumption data from smart meters, patient data from health sensors, or financial transaction records, data quality in big data systems is not merely a technical issue; it is a critical factor that threatens the system’s reliability and validity.

In numerous studies, the massive and complex data streams generated by IoT devices are viewed as both an opportunity and a challenge for fraud detection; particular attention has been paid to financial fraud. Those studies highlight that machine learning (ML)-enabled AI systems can successfully identify suspicious patterns by utilizing continuous data streams to detect credit card and financial fraud [3, 6]. For instance, Choi and Lee (2018) [7] proposed a comprehensive model incorporating artificial neural networks and ML processes to process vast amounts of financial data and detect fraud, while addressing the class imbalance problem through techniques such as sampling and feature selection. Mishra and Pandey (2021) [8] developed a fraud prediction model using logistic regression and k-fold techniques, demonstrating the effectiveness of their methodology on large-scale daily datasets obtained from European banks. The integration of IoT and big data extends beyond the financial sector. Focusing on the Internet of Medical Things (IoMT), Lakhan et al. (2022) [9] presented a framework based on blockchain and federated learning designed to protect the privacy of health data and prevent fraud (such as fraudulent insurance claims). Their study aimed to reduce energy consumption and latency by optimizing data validation and fraud analysis across distributed fog and cloud nodes. Paquet-Clouston et al. (2017) [10] investigated how an IoT botnet (Linux/Moose) engaged in social media fraud specifically the sale of likes and followers on platforms like Instagram. That study illustrates how IoT devices (such as routers and smart TVs) can be utilized as tools for large-scale network manipulation.

Gradient Boosting (GB) models such as specifically the GBM, GBC, and XGBoost variants are frequently employed for fraud detection in IoT environments due to their high accuracy and ability to handle complex data structures [6]. GB models are

also noted as effective algorithms within ensemble learning frameworks for fraud detection [11]. Furthermore, the Light GB algorithm is reported to offer improved time complexity and successfully identify fraud in real-world credit card transactions [12].

Fraud detection is a classification problem; it can be defined as the task of determining whether a given data point is normal or aberrant. Consequently, the accuracy of this decision is crucial. Given the potential issues arising from misclassifying an aberrant financial transaction as normal, the importance of employing a ML model with high accuracy is evident. GB algorithms and Random Forest stand out as leading methods for handling the massive data streams generated by IoT devices, excelling in both general accuracy and the detection of complex, nuanced fraud cases [6, 11, 12].

GB is a powerful ensemble ML technique that aims to create a single strong model by sequentially combining weak prediction models (typically simple decision trees). The fundamental operating principle of this model is based on a stagewise approach [13, 14]. Each new model is trained to correct the errors (residuals) made by the ensemble up to that point [13]. GB models are effectively used in fields such as medical diagnosis (diabetes risk prediction, breast cancer tumor classification, and heartbeat/ECG analysis) [15, 16, 17], robotics and control (robotic arm control and motion recognition using EMG signals) [13], finance and risk management (credit risk assessment) [13, 18], image and text analysis (land cover classification via satellite imagery and automatic categorization of text documents) [13, 19], and Education (student performance prediction and identification of at-risk students) [20, 21, 22].

GB models stand out significantly in precision-oriented fields where accuracy is paramount, such as fraud detection [23]. GB models offer high accuracy and precision. In applications requiring a minimal margin of error like fraud detection, they can achieve extremely high accuracy rates through tens of thousands of iterations [13]. They are considered among the most effective algorithms for assessing risks and the potential for fraud in credit applications [23]. In summary, due to their ability to capture complex non-linear relationships, handle missing data, and deliver high predictive power, GB models are currently among the most preferred choices for critical tasks such as fraud detection, intrusion detection, and risk analysis [6, 12, 13, 23].

Another method that can be used to analyze big data is sonification. Sonification is a method that generally aids in analyzing trends within big data. In its broadest definition, sonification is the use of non-speech sounds to convey information [24]. This process is also defined as the practice of generating sound from datasets [25]. Specifically, data sonification is described as “the transformation of data into acoustic signals to facilitate communication or interpretation” [26]. Defined as the systematic conversion of data into sound, data sonification is a practice aimed at conducting scientific analysis through the use of sound to represent data [27, 28] and making the information derived from data clearly understandable to users [29]. However, basic sonification methods such as parameter mapping may not be sufficient to distinguish patterns within massive datasets. Furthermore, data

analysts might consider modern data analysis approaches (like GB) to be “good enough” and fail to see any value in sonification. Nevertheless, if its effectiveness is demonstrated, there may be opportunities to use sonification either as a substitute for, or as a supplement to, other data analysis methods.

Sonification is employed as an effective method for classifying complex data. Examples of effective sonification applications can be found in areas such as cyber-threats on internet networks [30, 31], medicine [25, 32], biology and genetics [25], and financial markets [33]. Summaries of these studies indicate that sonification significantly enhances classification accuracy and sensitivity. In data mining scenarios where visualization techniques fall short such as instances of “cluster overlapping” sound can facilitate the differentiation of distinct data groups [34]. The human brain is more adept at perceiving temporal changes via the auditory channel, making sound a more effective tool for time-dependent classification tasks (e.g., process monitoring) [35]. Auditory representations enable the rapid “screening” of large datasets and the easy detection of outliers [36].

The process of converting images into sound, transforming visual data into audible signals defines the practice of sonification. Deep learning models (such as ResNet) are employed in this conversion process [37, 38]. In summary, the process of converting images to sound relies on the principle of filtering or mapping visual data based on its attributes and subsequently reconstructing it into audible sound waves through mathematical transformations [37, 38, 39]. Data analysis can be performed either by converting non-visual data into visual data or by directly sonifying visual data.

II. PROBLEM STATEMENT

Both GB and sonification are methods that can be used to detect anomalies in big data. In this paper, we show how advanced and creative sonification approaches for fraud or anomaly detection can be integrated into ML workflows (pipelines) to improve overall performance. To do this, we focus on detecting “aberrant” scores in standardized tests. Student response data from large-scale assessments, like data streams in IoT systems, may contain behavioral anomalies such as inattention, cheating, or random clicking.

We present a novel approach for identifying aberrant item scores in the context of Trends in International Mathematics and Science Study (TIMSS). In this approach, we enrich the GB model based on traditional “person-fit statistics (PFS)” features with a ResNet-18 model trained on sonified representations of the same features.

III. METHOD

This study focuses on detecting aberrant response patterns in large-scale educational assessment data. To this end, traditional PFS scores were calculated based on students’ responses. The study’s target variable was defined by classifying individuals as “aberrant” if they were flagged according to at least two PFS measures, and as “normal” otherwise. The effectiveness of GB and sonification methods in classifying individuals based on this target variable was investigated.

A. Data and Preprocessing

This study utilizes responses provided by eighth-grade students who participated in the TIMSS 2023 assessment and completed the student questionnaire to specific items selected from that questionnaire. The *Izpoly*, *Gnormedpoly*, and *U3poly* statistics were calculated based on the scores derived from these polytomous questionnaire items. The model inputs consist of 21 variables, comprising these statistics (PFS scores), their standard deviation values, and the students' ability scores.

The z-score standardization was applied to prepare 21 variables for ML models. This process transforms each variable so that it has a mean of 0 and a standard deviation of 1. The transformation was performed using the mean and standard deviation of the training data, and the same parameters were applied to the validation and test sets. This approach aimed to ensure that variables on different scales are weighted equally by the models, to maintain data consistency prior to Borderline-SMOTE, and to establish a standard basis for transforming variables into the 0–1 range during the sonification phase. Although tree-based algorithms do not theoretically require scaling, this standardization step was chosen to ensure the performance of methods such as Logistic Regression used during the ensemble learning stage and to maintain overall system integrity. A distinct scaling strategy was employed for the sonification phase; each variable was normalized to the [0, 1] range using the minimum and maximum values from the training set. This approach ensures that all variables are represented within a consistent range prior to sound synthesis.

The dataset was divided into three subsets: train, validation, and test. During the preprocessing stage, observations containing missing values were removed from the dataset. Due to the natural imbalance between the normal and aberrant classes, the Borderline-SMOTE method was applied to ensure the model learned the decision boundary without bias. This method aims to enhance the model's discriminative power by generating synthetic data, specifically targeting minority class samples located near the decision boundary. In this study, the oversampling ratio was set to 0.30 to prevent excessive synthetic data generation. Additionally, class weights were employed during the training phase to increase the model's sensitivity to the minority class.

B. Classification Models and Sonification-Based Approach

The HistGradientBoosting algorithm was selected as the classification model. Compared to traditional GB methods, this algorithm offers faster training times and better scalability with large datasets. HistGradientBoosting reduces computational load by binning continuous features using a histogram-based approach and yields effective results, particularly with high-dimensional data [40].

One of the key innovations of this study is the implementation of a sonification-based method alongside traditional ML approaches. In this research, feature scores for ML were first converted into audio signals (sound waves), and subsequently, log-mel spectrogram images were generated from these audio signals. These resulting images were classified using the ResNet-18 architecture. This methods follows from [32].

The probability outputs obtained from both models were combined using two different ensemble learning approaches. First, the Blend Ensemble method was employed. In this method, a weighted average of the probability outputs from the HistGradientBoosting and ResNet-18 models was calculated using the validation set, and the weight combination yielding the highest F1-score was determined [41]. Second, the Stacking Ensemble approach was applied. Stacking is an ensemble method in which the outputs of base learners serve as inputs to a higher-level model known as a meta-learner which then makes the final decision [42]. In this study, Logistic Regression was used as the meta-learner. By combining the strengths of different models, the stacking method can produce results that are more generalizable and stable compared to those of a single model [43]. The overall pipeline of the proposed approach is illustrated in Figure 1.

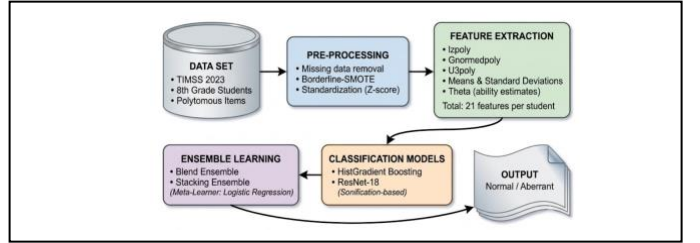


Fig. 1. Overview of the proposed methodology.

C. Evaluation Metrics

The performance of the models was analyzed using multiple metrics to evaluate classification success from various perspectives. Since accuracy alone is not a sufficient indicator due to the imbalanced class distribution, precision, recall, F1-score, Area Under the ROC Curve (AUC), Area Under the Precision-Recall Curve (AUPR), and Matthews Correlation Coefficient (MCC) were calculated. Additionally, true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values were reported based on the confusion matrix to enable a detailed examination of the model's classification performance.

IV. RESULTS

This section presents findings regarding the classification performance of four different model architectures: HistGradientBoosting, ResNet-18, Blend Ensemble, and Stacking Ensemble. The models were compared using F1-score, AUC, AUPR, MCC, sensitivity, specificity, and the confusion matrix. Additionally, the F1-score variation during the ResNet-18 training process and spectrogram examples of normal and aberrant classes are illustrated.

A. Model Performance Comparison

The classification performance of four different model architectures was compared using F1-score, AUC, AUPR, MCC, sensitivity, and specificity metrics. The obtained performance metrics are presented in Table 1.

TABLE I. CLASSIFICATION PERFORMANCE METRICS

Model	GB Baseline	ResNet- 18	Blend Ensemble	Stacking Ensemble
F1-score	0.825	0.824	0.830	0.830
AUC	0.989	0.988	0.989	0.989
AUPR	0.924	0.927	0.932	0.932
MCC	0.804	0.801	0.809	0.809
Precision	0.857	0.818	0.849	0.849
Recall	0.795	0.830	0.813	0.812
Specificity	0.983	0.976	0.981	0.982

Table 1 shows that the AUC values for all models range from 0.988 to 0.989, indicating a high level of discriminative power in classification. Regarding the F1-score, the ensemble models achieved the highest performance. Blend Ensemble model also yielded the highest AUPR and MCC values. The Stacking Ensemble model demonstrated performance comparable to the Blend Ensemble and achieved the highest precision. The ResNet-18 model exhibited the highest recall compared to the other models, suggesting it is more successful at identifying the “aberrant” class. Specificity values for all models ranged between 0.976 and 0.983, reflecting high accuracy in correctly identifying the “normal” class. Overall, ensemble learning methods (Blend and Stacking Ensemble) delivered superior and more balanced performance compared to individual models (GB and ResNet-18). The high sensitivity of ResNet-18 demonstrates the effectiveness of the sonification-based approach in identifying aberrant samples. It can be concluded that the Blend Ensemble model achieved the best overall performance.

B. ROC and Precision-Recall Curves

The classification performance of the models is visualized using ROC (Receiver Operating Characteristic) and Precision-Recall (PR) curves. ROC curves illustrate the true positive and false positive rates of the models at various threshold values, while PR curves are employed to evaluate model performance more precisely, particularly in cases of imbalanced class distributions.

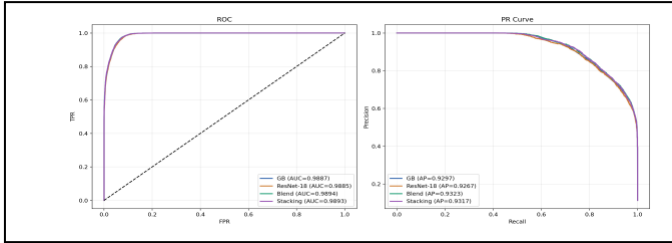


Fig. 2. ROC and Precision-Recall curves.

The ROC and PR curves presented in Figure 2 demonstrate that all four models exhibited high and comparable performance. The area under the ROC curve (AUC) ranged from 0.988 to 0.989, while the area under the PR curve (AUPR) ranged from 0.924 to 0.932. The highest AUPR value was achieved by the Blend Ensemble model. These findings support the effectiveness of the proposed ensemble methods in detecting aberrant samples, particularly in the context of the imbalanced class problem.

C. Confusion Matrix Analysis

To examine the classification performance of the models in detail, confusion matrices were generated for the four models. By visualizing true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values, these matrices enable a comparative assessment of the models’ error types. The confusion matrices are shown in Figure 3.

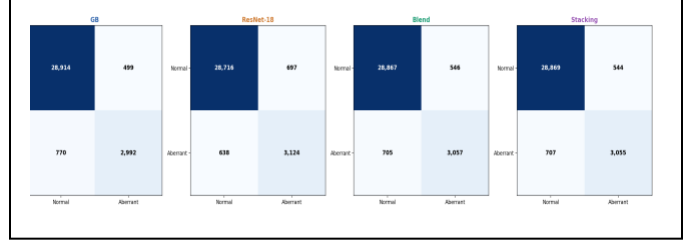


Fig. 3. ROC and Precision-Recall curves.

According to the confusion matrices presented in Figure 3, all models classified the normal class with high accuracy (TN > 28,700). ResNet-18 was the model that best captured the aberrant class (TP = 3,124). Meanwhile, the ensemble methods (Blend and Stacking) demonstrated more balanced performance regarding error distribution.

D. Training Dynamics

The performance of the ResNet-18 model during the training process was monitored using F1 scores calculated on the validation set at the end of each epoch. The model’s learning progress was evaluated by taking the early stopping criterion into account. The resulting training curve is presented in Figure 4.

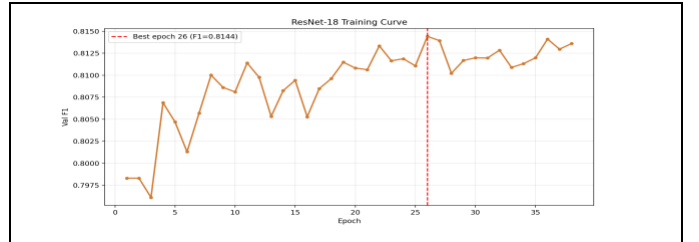


Fig. 4. The training curve of the ResNet-18 model.

An examination of Figure 4 reveals that the ResNet-18 model’s F1-score value increased rapidly from the initial epochs, reaching its peak value (F1-score = 0.814) at the 26th epoch. No significant improvement in the F1-score was observed beyond this point; as the model failed to demonstrate a meaningful performance gain over the subsequent 12 epochs (up to the 38th epoch), early stopping was implemented and training was terminated. These findings indicate that the ResNet-18 model reached its optimum point at the 26th epoch during the learning process and that extended training yielded no additional benefits.

E. Spectrograms

Spectrograms are three-dimensional representations of audio signals composed of log-mel, delta, and delta-delta channels. These representations illustrate the visual input provided by data

obtained through sonification during the classification process. Figure 5 presents three-channel (log-mel, delta, delta-delta) spectrogram images of normal and aberrant samples.

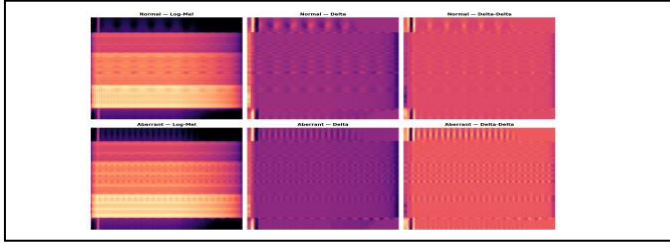


Fig. 5. Spectrograms (Respectively: log-mel, delta, delta-delta for the normal sample; log-mel, delta, delta-delta for the aberrant sample.)

An examination of Figure 5 reveals that, in the Log-Mel channel, the response pattern of a typical student exhibits broad frequency bands with smooth transitions; in contrast, for the aberrant student, these bands transform into a structure characterized by higher frequency and sharper edges, with abrupt energy transitions. In the context of TIMSS, this pattern may indicate behaviors such as “rapid guessing” where the student responds to questionnaire items very quickly without reflection or “high response variability,” where the student displays an inconsistent response rhythm by spending very little time on some questions and excessive time on others. In the Delta channel, while the typical student’s wide-interval patterns reflect a measured and consistent response rhythm, the aberrant student’s patterns become noticeably denser, with abrupt transitions between responses. This finding is consistent with the interpretation of the Log-Mel channel. Finally, regarding the Delta-Delta channel, it can be observed that the typical student exhibits a pattern distributed homogeneously across frequency regions, whereas the aberrant student displays a layered structure in which the lower, middle, and upper bands are distinct from one another.

V. CONCLUSION AND DISCUSSION

This study aimed to classify students participating in the TIMSS 2023 large scale assessment specifically those with and without aberrant item scores using ML and sonification methods based on their responses to questionnaire items. To achieve this objective, student responses were analyzed using three traditional PFS; individuals flagged as aberrant by at least two of these statistics were designated as such and defined as the target variable for the GB model. In this research, data sonification was employed alongside the GB model to classify aberrant students. Spectrogram images of the sonified data were modeled using a deep learning architecture to determine the effectiveness of sonification in classifying aberrant students. The HistGradientBoosting algorithm was used for the GB model. The GB model achieved high classification accuracy. This finding of the study aligns with research findings [13, 15, 16, 23].

Sonification-based classification using ResNet-18 also demonstrated relatively high performance. Numerous studies [25, 30, 32, 33, 36] indicate that sonification is a useful method for both classification and fraud detection, and our research reached a similar conclusion. In our study, the ResNet-18

approach increased sensitivity by identifying many aberrant students that the GB model had overlooked. Ensemble models combining these two approaches via Blending and Stacking techniques struck a balance between precision and sensitivity by merging the GB model’s precision with the sonification approach’s high sensitivity. Our findings demonstrate that sonification can reveal non-linear patterns of anomaly that traditional models fail to detect.

Although ensemble learning combined with sonification generally yields positive results in this research, it must be noted that sonification has certain limitations. Converting numerical data into sound via parameter mapping in a way that the human auditory system can easily distinguish is not a simple task [24, 44]. The generated sounds may lack meaning for listeners. The interpretation of these sounds can vary from user to user, depending on individual hearing ability, experience, and cultural background [25, 44]. Converting massive datasets into sound in real-time requires significant computational power, particularly with model-based approaches [44]. Furthermore, mapping too many data dimensions to sound simultaneously can lead to a lack of clarity [25, 45]. There is no universally accepted method for determining how data should be mapped to specific sound parameters; consequently, design decisions are often subjective or based on trial and error [44].

Despite all these limitations, this study demonstrates that a multi-model ensemble approach can be an effective tool, particularly in the context of training data with imbalanced class distributions, such as the detection of aberrant responses. Furthermore, to facilitate more accurate decision-making processes, we propose the combined use of modern sonification approaches and traditional fraud detection methods when working with big data.

REFERENCES

- [1] T. N. Fitria and N. E. Simbolon, "Internet of things (IoT) in education: opportunities and challenges," in *Prosiding Seminar Nasional & Call for Paper STIE AAS*, vol. 6, no. 1, 2023.
- [2] L. A. Alexei and A. Alexei, "Analysis of IoT security issues used in Higher Education Institutions," *International Journal of Mathematics and Computer Research*, vol. 26, no. 5, pp. 2277–2286, 2021.
- [3] T. Ganesan, "Machine learning-driven AI for financial fraud detection in IoT environments," Available at SSRN 5665670, 2019.
- [4] T. C. Olowu, "ICT, computers, internet and cyber security in education system," *J. Teach. Learn. Afr.*, vol. 5, no. 1, pp. 1–12, 2022.
- [5] A. Dehbi, A. Bakhoui, R. Dehbi, and M. Talea, "Smart evaluation: A new approach improving the assessment management process through cloud and IoT technologies," *Int. J. Inf. Educ. Technol.*, vol. 14, no. 1, pp. 107–118, 2024.
- [6] M. N. Alatawi, "Detection of fraud in IoT based credit card collected dataset using machine learning," *Mach. Learn. Appl.*, vol. 19, p. 100603, 2025.
- [7] D. Choi and K. Lee, "An artificial intelligence approach to financial fraud detection under IoT environment: A survey and implementation," *Security and Communication Networks*, vol. 2018, p. 5483472, 2018.
- [8] K. N. Mishra and S. C. Pandey, "Fraud prediction in smart societies using logistic regression and k-fold machine learning techniques," *Wireless Pers. Commun.*, vol. 119, no. 2, pp. 1341–1367, 2021.
- [9] A. Lakhan et al., "Federated-learning based privacy preservation and fraud-enabled blockchain IoMT system for healthcare," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 2, pp. 664–672, 2022.

- [10] M. Paquet-Clouston, O. Bilodeau, and D. Décary-Héty, "Can we trust social media data? Social network manipulation by an IoT botnet," in *Proc. 8th Int. Conf. Social Media & Society*, 2017, pp. 1–9.
- [11] R. Soleymanzadeh, M. Aljasim, M. W. Qadeer, and R. Kashef, "Cyberattack and fraud detection using ensemble stacking," *AI*, vol. 3, no. 1, pp. 22–36, 2022.
- [12] P. Srikanth, "An efficient approach for clustering and classification for fraud detection using bankruptcy data in IoT environment," *Int. J. Inf. Technol.*, vol. 13, no. 6, pp. 2497–2503, 2021.
- [13] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurobot.*, vol. 7, p. 63623, 2013.
- [14] F. Zhang, B. Du, and L. Zhang, "Scene classification via a gradient boosting random convolutional network framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1793–1802, 2015.
- [15] "Early diabetes prediction using voting based," in *Advances in Computing and Data Sciences*, Springer, 2018, p. 95.
- [16] M. R. Abbasniya, S. A. Shekholeslamzadeh, H. Nasiri, and S. Emami, "Classification of breast tumors based on histopathology images using deep features and ensemble of gradient boosting methods," *Comput. Electr. Eng.*, vol. 103, p. 108382, 2022.
- [17] H. Shi et al., "A hierarchical method based on weighted extreme gradient boosting in ECG heartbeat classification," *Comput. Methods Programs Biomed.*, vol. 171, pp. 1–10, 2019.
- [18] R. Sun, G. Wang, W. Zhang, L. T. Hsu, and W. Y. Ochieng, "A gradient boosting decision tree based GPS signal reception classification algorithm," *Appl. Soft Comput.*, vol. 86, p. 105942, 2020.
- [19] Q. T. Bui et al., "Gradient boosting machine and object-based CNN for land cover classification," *Remote Sens.*, vol. 13, no. 14, p. 2709, 2021.
- [20] R. Guevara-Reyes, I. Ortiz-Garcés, R. Andrade, F. Cox-Riquetti, and W. Villegas-Ch, "Machine learning models for academic performance prediction: interpretability and application in educational decision-making," *Front. Educ.*, vol. 10, p. 1632315, 2025.
- [21] C. Wang, L. Chang, and T. Liu, "Predicting student performance in online learning using a highly efficient gradient boosting decision tree," in *Int. Conf. Intell. Inf. Process.*, Springer, 2022, pp. 508–521.
- [22] W. Ahmed, M. A. Wani, P. Plawiak, S. Meshoul, A. Mahmoud, and M. Hammad, "Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions," *Sci. Rep.*, vol. 15, no. 1, p. 26879, 2025.
- [23] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 1937–1967, 2021.
- [24] B. N. Walker and M. A. Nees, "Theory of sonification," in *The Sonification Handbook*, 2011, pp. 9–39.
- [25] R. Braun, M. Tfirm, and R. M. Ford, "Listening to life: Sonification for enhancing discovery in biological research," *Biotechnol. Bioeng.*, vol. 121, no. 10, pp. 3009–3019, 2024.
- [26] G. Kramer et al., "Sonification report: status of the field and research agenda," 2010.
- [27] J. Kim, Y. Lee, and I. Seo, "Math graphs for the visually impaired: audio presentation of elements of mathematical graphs," in *Ext. Abstr. CHI Conf. Hum. Factors Comput. Syst.*, 2019, pp. 1–6.
- [28] S. Lenzi and P. Ciuccarelli, "Intentionality and design in the data sonification of social issues," *Big Data Soc.*, vol. 7, no. 2, p. 2053951720944603, 2020.
- [29] T. Ziemer and H. Schultheis, "Pampas: A psychoacoustical method for the perceptual analysis of multidimensional sonification," *Front. Neurosci.*, vol. 16, p. 930944, 2022.
- [30] A. Moreno-Fernandez-de-Leceta, S. Lenzi, P. Galán-García, D. de Tiso, and L. E. Insagurbe, "A sonification model for real-time anomaly detection in machine learning supported cybersecurity," (in press).
- [31] S. Lenzi, G. Terenghi, D. Meacci, A. Moreno Fernandez-de-Leceta, and P. Ciuccarelli, "The design of Datascape: toward a design framework for sonification for anomaly detection in AI-supported networked environments," *Front. Comput. Sci.*, vol. 5, p. 1254678, 2024.
- [32] B. N. Walker et al., "Dermoscopy diagnosis of cancerous lesions utilizing dual deep learning algorithms via visual and audio (sonification) outputs: laboratory and prospective observational studies," *EBioMedicine*, vol. 40, pp. 176–183, 2019.
- [33] M. Ballora, R. J. Cole, H. Kruesi, H. Greene, G. M. Iyer, and D. L. Hall, "Use of sonification in the detection of anomalous events," in *Proc. SPIE*, vol. 8407, 2012, pp. 213–219.
- [34] Z. Halim and S. Bashir, "Sonification: A novel approach towards data mining," in *2nd Int. Conf. Emerg. Technol.*, 2006, pp. 13–14.
- [35] J. Cooke, W. Díaz-Merced, G. Foran, J. Hannam, and B. García, "Exploring data sonification to enable, enhance, and accelerate the analysis of big, noisy, and multi-dimensional data," *Proc. Int. Astron. Union*, vol. 14, no. S339, pp. 251–256, 2017.
- [36] T. Hermann and H. Ritter, "Listen to your data: Model-based sonification for data analysis," *Advances in Intelligent Computing and Multimedia Systems*, 1999.
- [37] S. Vinod and M. V. Thomas, "A comparative analysis on deep learning techniques for skin cancer detection and skin lesion segmentation," in *Int. Conf. Commun. Control Inf. Sci.*, 2021, pp. 1–6.
- [38] H. N. Bicer, P. Götz, C. Tuna, and E. A. Habets, "Explainable acoustic scene classification: Making decisions audible," in *Int. Workshop Acoust. Signal Enhancement*, 2022, pp. 1–5.
- [39] S. A. ElGhany, M. R. Ibraheem, M. Alruwaili, and M. Elmogy, "Diagnosis of various skin cancer lesions based on fine-tuned ResNet50 deep network," *Comput. Mater. Contin.*, vol. 68, no. 1, 2021.
- [40] M. A. Agebure, J. K. Wiredu, and S. Akobre, "Gradient boosting revisited: comparative analysis of selected advances on real-world tabular data," *Compos. Mater.*, vol. 11, no. 1, pp. 37–52, 2026.
- [41] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*, Springer, 2000, pp. 1–15.
- [42] D. H. Wolpert, "Stacked generalization," *Complex Syst.*, vol. 5, no. 2, pp. 191–215, 1992.
- [43] L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, vol. 33, no. 1–2, pp. 1–39, 2010.
- [44] T. Hermann, "Sonification for exploratory data analysis," Ph.D. dissertation, Faculty of Technology, Bielefeld University, Bielefeld, Germany, 2002.
- [45] L. Axon, B. A. AlAhmadi, J. R. C. Nurse, M. Goldsmith, and S. Creese, "Data presentation in security operations centres: exploring the potential for sonification to enhance existing practice," *J. Cybersecurity*, vol. 6, no. 1, pp. 1–16, 2020.