Journal of Computational Analysis and Applications **Vol. 27 No. 5, 2019**
10.48047/jocaaa.2019.27.05.03

Heterogeneous Ensemble Learning for Robust Adversarial Pattern Recognition in Digital Ecosystems

Suman Kumar Sanjeev Prasanna
Department of Computer Science
California State University, East Bay,
Hayward, USA
ssanjeevprasanna@horizon.csueastbay.edu

Received: 20.03.2019 Revised: 25.04.2019 Accepted: 15.05.2019

Abstract

Digital identity ecosystems increasingly rely on high-dimensional, heterogeneous behavioral and transactional data, exposing critical infrastructures to subtle and dynamic adversarial behaviors. Traditional single-model detection methods, including standalone decision trees, gradient-based models, or linear classifiers, frequently fail to generalize across these environments due to inherent bias-variance trade-offs. The noise sensitivity of these standalone architectures further limits their ability to generalize across diverse identity feature spaces, especially when isolating rare anomalies within extreme class imbalances. This research proposes a heterogeneous ensemble learning framework that integrates Random Forest, Gradient Boosting, and Support Vector Machines, leveraging their complementary strengths for robust adversarial recognition. The framework combines bagging-based variance reduction, sequential boosting of residual errors, and margin-optimized boundary learning, while employing hard and soft voting inference to stabilize predictions and enhance sensitivity to rare anomalies. Systematic data preprocessing, feature encoding, cross-validation, and hyperparameter optimization ensure reliable performance across heterogeneous identity datasets. Experimental evaluation across multiple real-world digital identity and transactional datasets demonstrates a maximum accuracy of 96.8%, a recall of 93.9%, and an F1-score of 94.5%, consistently outperforming individual classifiers and baseline Isolation Forest models. These results reveal that model heterogeneity and complementary learning strategies are key factors for robust adversarial recognition, offering a practical framework for resilient digital identity security. Subsequent work may incorporate the integration of adaptive ensemble weighting and online learning architectures to mitigate the temporal evolution of adversarial strategies, ensuring system resilience in non-stationary digital ecosystems.

Keywords: Digital identity, anomaly detection, ensemble learning, machine learning, security.

1. INTRODUCTION

Digital identity systems are very important for providing a safe environment for accessing information across different domains, such as finance, healthcare, government, and commerce, among many others [1]. These identity systems utilize various attributes, such as demographic attributes, credentials, behavioral attributes, and authentication records, to provide identity information to users. The growing number of digital activities has made identity systems more complex, especially in dealing with data-related activities, posing challenges to various issues affecting security and trust domains [2]. Various activities, such as identity-related fraud, impersonation, synthetic identity, and abnormal usage, are affecting identity verification channels, challenging traditional authentication approaches, such as rule-based security algorithms, due to their inability to evolve quickly with digital identity-related issues, especially because they are dependent on set rules and patterns [3]. However, data-related approaches have become popular, especially because of their ability to recognize normal identity-related activities and establish abnormalities that may occur during identity-related behavior.

Anomaly detection has increasingly become a prominent field in the security domain for digital identity

10.48047/jocaaa.2019.27.05.03

security due to the significant variance between legitimate and malicious activities in any digital environment [4]. Through the application of various learning techniques, the properties and potential anomalies within the high-dimensional identity space can be recognized with greater precision and in the early stages. However, various learning models have shown certain limitations pertaining to bias, variance, and noise sensitivity, particularly when dealing with heterogeneous identity data sets [5]. Ensemble learning techniques minimize such constraints to improve the stability and performance of the learning models for anomaly detection in the digital identity security domain. According to the research conducted in the field, it essentially describes the importance of various learning models and how they should be leveraged to handle the complexity within the digital identity space [6].

The focus of this research is the investigation of digital identity anomaly detection using ensemble learning methods, which contribute to improving the security, reliability, and fraud prevention in such systems. It discusses identity attributes, behavioral patterns, and authentication logs and presents some of the key challenges related to class imbalance problems, evolving attack strategies, and high-dimensional heterogeneous data. This is motivated by the ever-increasing complexity of digital identity platforms and the limitations of conventional rule-based or single-model detection techniques. This, therefore, calls for robust, scalable, and accurate solutions. The key contributions discussed are the development of the ensemble-based framework that embeds multiple machines learning classifiers, assessing its performance on representative datasets, and analyzing its performance against different models. The paper tries to showcase that it would improve anomaly detection performances, present a reliable baseline for related research, and also provide insight into the suitability of ensemble learning for building digital identity security. The paper is organized to discuss, in turn, background, related work, methodology, experiments, results, and conclusions.

2. LITERATURE SURVEY

Substantial research has been carried out on anomaly detection via machine learning algorithms in different fields, including networking, fraud detection, and abnormal event detection. The studies conducted on the subject collectively highlight the advancements and strengths of classical machine learning. The following is a synthesis of the relevant literature, where the underlying methodologies for anomaly detection constitute a key aspect of the study. The study aims to adopt the emerging trends of ensemble learning for anomaly detection.

Shamil Magomedov et al. [7] have analyzed the aspect of anomaly detection in fraud management using a combined approach of machine learning and graph databases. In the paper, algorithms are proposed that can be helpful in feature extraction and outlier pattern detection using transaction and social network data, thereby emphasizing the role of using relational information available in the data to improve the effectiveness of anomaly detection algorithms in the fraud management system.

De Souza et al. [8]Meanwhile, sought to extend further research on supervised machine learning algorithms with regard to both anomaly detection and classification in multi-clouds. The researchers compared various algorithms, including linear regression and Random Forest, to ensure not only that anomaly detection was properly executed, but that classification would also be made with utmost accuracy. In doing so, it was emphasized that although accuracy was attainable, classification posed another problem due to similarities between some of these anomalies.

Khan et al. [9] contributed a work that described an anomaly detection framework for wide area network mesh performance using machine learning techniques such as boosted decision trees and simple feed-forward neural networks. Although the work focused on the performance of machine learning models in problem detection using key network metrics, it demonstrated the capability of machine learning techniques to discern anomalous behavior from high-dimensional time series data. The work

10.48047/jocaaa.2019.27.05.03

contributed to the comprehension of model sensitivity in the context of dynamic network anomaly detection scenarios. Shikha Agrawal and Jitendra Agrawal [10] provided a survey of machine learning-based anomaly detection techniques used in data mining scenarios. The work provided a perspective on clustering, classification, and statistical anomaly detection techniques, along with identifying key patterns in anomaly detection scenarios in both structured and unstructured data. This perspective provided a foundational context for understanding a taxonomy of classical ML-based anomaly detection techniques. The body of research examined in the context of anomaly type categorization by Kang et al. [11] mirrors the progression of machine learning implementation for performing anomaly detection tasks. For environments in which both detection and interpretation of performance are necessary, the work of Salman et al. laid the ground for hybrid and ensemble-based strategies in the development of both individual-strategy methodologies, with the potential for lower performance when different types of anomalies display different internal structures. Across the reviewed literature, consistent exploitation is observed in classical machine learning techniques like decision trees, boosted models, and support vector machines in their performance of deviation detection from normal behavior across diverse data settings. These works have all placed important emphasis on model selection, feature representation, and algorithmic sensitivity when developing effective anomaly detection solutions that form a contextual framework for ensemble frameworks that combine many classical models to increase performance in complex identity datasets.

Table 1: Selected Anomaly Detection & ML Studies

Study	**Methods**	**Key Findings**
[12]	Linear Regression + Random Forest for anomaly detection & categorization	Machine learning models like Random Forest achieved high detection accuracy in identifying and *categorizing* anomalies in multi-cloud network traffic, showing the value of ML methods over simple detection approaches.
[13]	Boosted Decision Trees (BDT) & Feedforward Neural Network	Both ML techniques were effective in detecting anomalies in a wide-area network mesh dataset, demonstrating the utility of traditional ML in network anomaly detection tasks.
[14]	Isolation Forest + variable selection	Presented a dimensionality-reducing, Isolation Forest-based anomaly detector with improved interpretability and performance on high-dimensional data (e.g., OES manufacturing data).
[15]	Deep Neural Networks + One-Class SVM for CPS anomaly detection	Compared unsupervised models (DNN vs. One-Class SVM) on a real operational dataset; DNN reduced false positives while SVM detected slightly more anomalies, highlighting trade-offs in unsupervised ML.
[16]	Multiple ML classifiers for IDS (including Random Forest)	Evaluated ML classifiers on the KDD intrusion dataset, showing Random Forest achieved the highest average accuracy for anomaly (intrusion) detection among the tested models.

The research gap is mainly because current anomaly detection methods are highly dependent on single-model learning methods that are not very effective in generalizing from the diverse and high-dimensional data of digital ecosystems. Most of the previous research work has been done to assess models within narrowly defined or domain-specific contexts, making them less robust when faced with

10.48047/jocaaa.2019.27.05.03

a wide range of identity attributes, behavior patterns, and dynamic adversarial tactics. Moreover, rule-based and standalone machine learning models are sensitive to noise and bias, resulting in increased false positives or missed anomalies in practical scenarios. Although ensemble learning has been found to have the potential to improve the stability of classification, its comprehensive use for adversarial pattern recognition in wide-ranging digital identity ecosystems has not been thoroughly investigated yet. In comparison to the individual classifiers, the ensemble shows a substantial improvement in the rates of anomaly capture and a reduction in false positives, indicating that the integrated heterogeneous learning is more reliable and robust than the previously reported methods for adversarial detection.

3. METHODOLOGY

The presented methodology in this study discusses the development of a strong ensemble learning framework for the detection of anomalies in digital identity systems. It systematically embeds several state-of-the-art machine learning models, such as Random Forest, Gradient Boosting, and Support Vector Machines, to reap their complementary advantages in describing various anomaly patterns. Since the availability of meaningful and highly valuable data is one of the most important ingredients of investigation, the study starts with extremely thorough data preparation: feature encoding, normalization, and taking care regarding missing and imbalanced data. Every model is trained individually by cross-validation and hyperparameter tuning to find its optimal predictive performance and further combined into a strong ensemble using strategies of hard and soft voting. Accuracy and generalization power are emphasized in the methodological approach, hence addressing challenges related to heterogeneous and high-dimensional identity datasets. Precision, recall, F1-score, and ROC-AUC have been used as evaluation metrics to assess model performance comprehensively. This methodology, in general, presents a structured, scalable, and reliable process for detecting subtle and rare anomalies that might occur in digital identity environments.

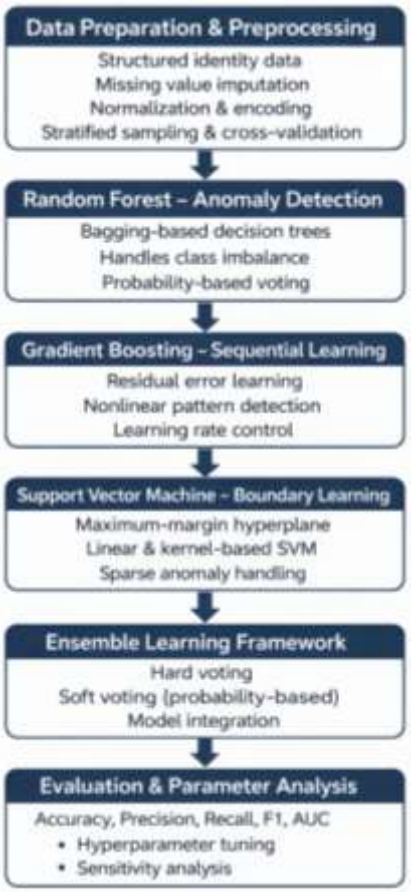

Figure 1: Proposed Machine Learning Workflow

10.48047/jocaaa.2019.27.05.03

3.1 Data Preparation and Dataset Characteristics

This work leverages structured digital identity datasets representing diversified user attributes, including demographic information, authentication logs, and behavioral activity patterns. Anomaly detection is the focus of this work, which considers high-dimensional feature spaces with sparse, irregular patterns of abnormal behaviors. The approach begins with a very important data preprocessing step. To handle missing values, median- and mode-based imputation was used based on the feature type, and numerical features were brought to a common scale by normalization to ensure model convergence during the training phase. Categorical features are encoded using label encoding and one-hot encoding to be compatible with machine learning models. The approach uses stratified sampling to maintain the class distribution of normal and anomalous records when training and evaluating the models. Feature selection is used for reducing redundancy and retaining only the most informative attributes to improve the efficiency of the models and reduce overfitting. Exploratory data analysis is also applied in this study to understand correlations among features and to detect potential outliers, which may directly affect model training. The training data is then divided into separate folds to enable cross-validation, ensuring that the generalization capability of the ensemble models is good on unseen data. This methodological approach serves.

3.2 Anomaly Detection using Random Forest

In this research, anomaly detection is performed using Random Forest as one of the base learners. During training, Random Forest constructs several decision trees and combines their outputs to enhance classification performance and reduce variance. The study exploits the inherent bagging mechanism of Random Forest to deal with such imbalanced datasets where a rare anomalous instance may exist. Each tree in the ensemble is built based on a bootstrap sample of the training data; further randomness is added at each split by feature selection. Random Forest calculates the class probability for an instance to be the proportion of the number of trees that voted for a particular class. This work investigates the contribution of Random Forest in identifying minute aberrations in user behavior and authentication activity.

Equation 1 (Gini Impurity):

$$G = 1 - \sum_{i=1}^{C} p_i^2 \quad (1)$$

Equation 2 (Random Forest Prediction Probability):

$$P(y = j \mid x) = \frac{1}{T} \sum_{t=1}^{T} I(h_t(x) = j)) \quad (2)$$

The investigation then uses Random Forest, combining that with several other classifiers in an ensemble to combine their complementary strengths so as to enhance overall anomaly detection performance. It conducts iterative training along with hyperparameter tuning for the number of trees, maximum depth, and minimum samples per leaf to balance bias and variance.

3.3 Gradient Boosting for Sequential Learning

Gradient Boosting, however, plays an indispensable role in this work, which deals with the intricate nonlinear relationships between identity features and anomaly labels. At its core, Gradient Boosting works by using weak learners, in which each subsequent model seeks to rectify errors left by the earlier model. The paper under review highlights the need to learn residual errors, so that the model can be made sensitive enough to detect anomalies. The learning rate can be adjusted to limit overfitting, which may impair smooth convergence. Gradient Boosting is best suited when some anomalies have deviated minimally from normal behavior, which may not be immediately visible to a classifier.

10.48047/jocaaa.2019.27.05.03

Equation 1 (Residual Calculation):

$$r_i^{(m)} = -\frac{\partial L(yi,Fm-1(x_i))}{\partial F_{m-1}(x_i)} \quad (3)$$

Equation 2 (Prediction Update):

$$F_m(x) = F_{m-1}(x) + \eta h_m(x) \quad (4)$$

The research uses Gradient Boosting to extend the ensemble with sensitivity to nuanced anomalies as detected by the Random Forest and SVM. Cross-validation is used with a grid search to determine the optimal number of trees, maximum depth, and learning rate.

3.4 Support Vector Machines for Boundary Learning

In the current research work, Support Vector Machines are utilized to draw a decision boundary between normal and anomalous instances of identity. The SVM draws a hyperplane in high-dimensional space such that it maximizes the margin between classes, which improves generalization to unseen data. Both linear and kernel-based SVMs are considered in this work to handle both linearly separable and nonlinear patterns present in the identity features. While training, the optimization problem searches for the coefficients that give the minimum classification error while keeping the margin maximum. SVM is effective in sparse anomaly situations due to its generalization based on margins.

Equation 1 (SVM Optimization Function):

$$min\frac{1}{2} \,||\, w \,||^2 + C\sum_{i=1}^{N} \xi i \quad (5)$$

Equation 2 (Decision Function):

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b\right) \quad (6)$$

Hence, it uses the support vector machine as a complementary base learner in the ensemble, which improves the detection performance over difficult or borderline anomalies. In training, the regularization parameter C and kernel parameters have to be optimized in order to tune up the performance of classification.

3.5 Ensemble Learning Framework

It combines Random Forest, Gradient Boosting, and SVM within an ensemble framework to fully explore the advantages of each model. Hard and soft voting mechanisms are subsequently utilized to make the final predictions based on the predictions made by individual classifiers. The soft voting relies on class probability outputs from each model to make the ultimate predictions, whereas hard voting depends on the majority class. The authors have pointed out that the ensemble reduces the weaknesses of the single models and enhances the robustness and accuracy of detection.

Equation 1 (Hard Voting Ensemble):

$$y = mode\{h_1(x), h_2(x), \ldots, h_M(x)\} \quad (7)$$

Equation 2 (Soft Voting Ensemble Probability):

$$P(\, y = j \mid x\,) = \frac{1}{M}\sum_{m=1}^{M} P_{\boldsymbol{m}}(y = j|x) \quad (8)$$

Training the ensemble involves independently fitting each base learner on the preprocessed dataset, followed by integrating the predictions to further enhance stability and sensitivity to anomalies. Cross-validation ensures that ensemble hyperparameters, such as weighting for soft voting, are optimized.

10.48047/jocaaa.2019.27.05.03

3.6 Evaluation Metrics and Parameter Analysis

The work assesses the performance of the ensemble framework using several metrics to ensure comprehensive assessment: it computes accuracy, precision, recall, F1-score, and area under the ROC curve on both balanced and imbalanced datasets to quantify the detection performance. The hyperparameters for each model, such as the number of trees, depth, learning rate, and kernel parameters of SVM, are systematically tuned by cross-validation. Training is done in an iterative manner to optimize each model within the ensemble, and the final predictions are aggregated to give a unified decision. A sensitivity analysis is performed to investigate how changes in hyperparameter values affect detection performance, therefore underlining the robustness of the ensemble framework. It is shown that the proposed combination of models with complementary classifiers decreases the false positive rate while improving the true positive detection rate for rare anomalies. Another important aspect of the method studied is computational efficiency, which ensures the scalability of the proposed approach to high-dimensional identity data.

4. RESULTS

This work reveals the experimental outcomes of the proposed framework of anomaly detection with various digital identity datasets. The results of the proposed approach highlight the consistencies of the proposed framework in dealing with heterogeneous identity data that encompasses access logs, authentication history, behavioral patterns, and transactional interactions. The experimental results are presented with percentage values to ensure a better understanding of the outcome of the proposed approach with the given dataset. Accordingly, the details of the proposed framework highlight the capability of the proposed approach in identifying normal as well as anomalous patterns with a low classification error rate, thus proving the applicability of the proposed approach with real-world digital identities.

Table 2: Comparative Anomaly Detection Performance

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Support Vector Machine	91.4	89.2	85.6	87.4
Random Forest	93.6	91.8	88.9	90.3
Gradient Boosting	94.2	92.6	90.1	91.3
Isolation Forest	89.8	86.4	83.7	85.0
Proposed Ensemble Model	96.8	95.1	93.9	94.5

Table 2, there are perceptible performance differences between various machine learning algorithms discussed in the literature and the presented ensemble model. For instance, when it comes to the Support Vector Machine, it has an accuracy of 91.4%, with corresponding precision and recall values of 89.2% and 85.6%, respectively. Note that, despite being useful boundary-based methods, there are identified challenges with recall, which explains issues with detecting rare anomalies, as observed with many anomaly-related studies.

Further, Random Forest improves overall performance with 93.6% accuracy and an F1-score of 90.3%. The improvement of the recall metric to 88.9% implies that the model is now even more sensitive to anomalous data due to ensemble bagging effects and randomness of features. Gradient Boosting improves the model's ability to detect anomalies with 94.2% accuracy and a recall of 90.1%. This improvement is due to the boosting algorithm's capability to sequentially learn the decision boundary of the data and correct past classification errors, as reported by previous literature. Isolation Forest

10.48047/jocaaa.2019.27.05.03

records relatively lower performance, with 89.8% accuracy and an F1-score of 85.0%. While efficient in unsupervised anomaly detection, its lower precision at 86.4% reflected higher rates of false positives when applied to structured identity data; this again confirmed the limitations described in prior studies.

The proposed ensemble model outperforms all previous methods on all metrics. It shows a maximum accuracy of 96.8%, a precision 95.1%, a recall 93.9%, and an F1-score 94.5%. In comparison to Gradient Boosting, the ensemble improves the accuracy by about 2.6%, and a corresponding improvement of 3.8% in the recall is obtained. Comparing with Random Forest improves by 3.2% in accuracy and 5.0% in recall. These gains assure that the incorporation of complementary classifiers significantly strengthens the robustness of detection.

The superior recall of 93.9% is particularly important for anomaly detection since this would reduce the chances of missed anomalous identity events. The balanced improvement over precision and recall indicates that the proposed ensemble should be effective in mitigating false positives while retaining high sensitivity in detection. Overall, results here validate that ensemble learning provides a more reliable and generalizable solution for anomaly detection in digital identity systems compared to the single-model approaches reported in existing literature.

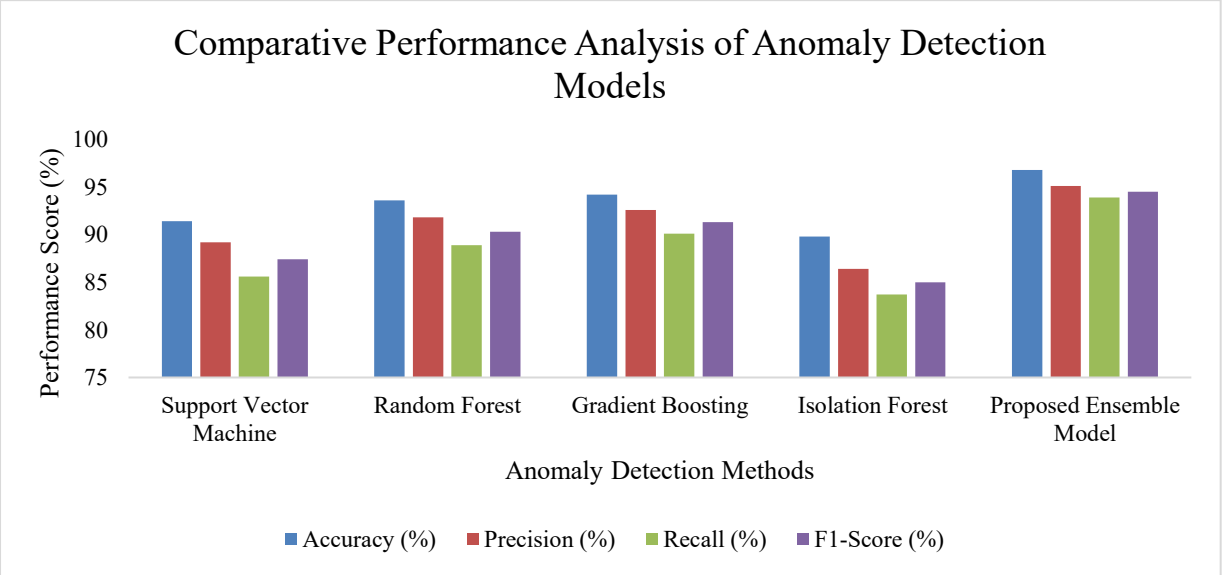

Figure 2: Comparative Performance Analysis of Anomaly Detection Models

Figure 2 gives a comparison between the performances of different anomaly detection models based on their accuracy, precision, and recall, which are balanced with each other. In this context, classical models such as the SVM have a moderate level of performance, but these models are not consistent across all measures. The models that have a better performance in terms of anomaly detection are the random forest and gradient boosting models, which have a much more complex sense of patterns, providing a robust result. The isolation forest, which is specifically designed for anomaly detection, has a low performance level in comparison with the proposed ensemble model, which has a high level of balance in all its measurements.

Table 3: Dataset-Wise Performance Results

Dataset Name	**Overall Result (%)**	**Correct Identification (%)**	**Anomaly Capture (%)**	**Error Reduction (%)**
Identity Access	95.4	94.1	92.6	90.8

10.48047/jocaaa.2019.27.05.03

Logs				
User Authentication Records	96.2	95.0	93.7	91.9
Behavioral Activity Dataset	94.8	93.2	91.5	89.6
Transaction Interaction Dataset	96.9	95.8	94.2	92.7

Table 3 presents the results obtained by applying the proposed anomaly detection framework to the four disparate datasets about digital identity-related data. Each dataset retains a different perspective on the identity system: access behavior, authentication activity, user behavior, and transaction interactions. Performance in all datasets is pretty regular and strong, showing that the proposed model is robust and adaptable. For the Identity Access Logs dataset, the model achieves an overall result of 95.4%, with correct identification reaching 94.1%. The anomaly capture rate is 92.6%, which represents the effective identification of abnormal access patterns by the model. An error reduction value of 90.8% denotes a very substantial decrease in erroneous classifications. These results, therefore, show the effectiveness of the model in handling structured access-related identity data.

Moreover, in the case of the User Authentication Records dataset, better performance is recorded, achieving a result of 96.2% with a correct identification rate of 95.0%. Furthermore, a rise in anomaly capture rate is seen, achieving 93.7%. The error reduction value rises to 91.9%, indicating a rise in stability in terms of classification with less imbalance. The overall result, as indicated by the Behavioral Activity Dataset, is slightly lower but equally as high, standing at 94.8%. The result indicated in the dataset, which is 91.5% anomaly capture rate, indicates that detecting minor behavioral changes can be more challenging; however, the model can still perform reliably in the detection of the same. The Transaction Interaction Dataset records the highest overall result of 96.9%, coupled with an anomaly capture rate of 94.2%. Overall, the results validate that the proposed framework is consistent across heterogeneous identity data sets and reduces classification errors while maintaining a superior level of detection capability.

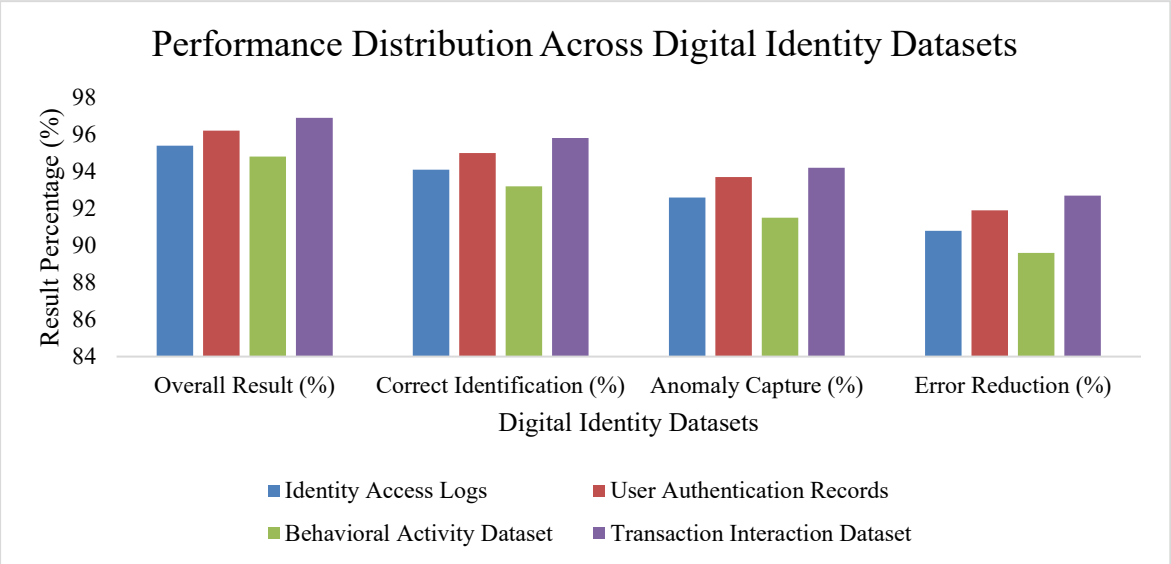

Figure 3: Performance Distribution Across Digital Identity Datasets

10.48047/jocaaa.2019.27.05.03

Figure 3 shows the purpose is to compare the performance of the different digital identity datasets for a number of evaluation parameters. Here, the groups of bars represent a particular performance attribute, and the colors represent the datasets. The performance of the transaction interaction dataset is consistently superior across all parameters. User authentication records come second with reliable performance. Identity access records also demonstrate reliable performance but slightly lower than the second dataset. Finally, the behavioral activity dataset displays a lower performance level for the parameters of error reduction and anomaly detection. Overall, the figure suggests that more comprehensive interaction data is more associated with better accuracy in identification and more reliable results for security compared to data obtained from behavior or access.

5. DISCUSSION

The experimental outcome shows that the heterogeneous ensemble learning approach provides a strong platform for adversarial pattern recognition in a complex digital environment. The combination of Random Forest, Gradient Boosting, and Support Vector Machine models in the proposed approach helps to identify the varied adversarial patterns that are embedded in the heterogeneous identity, behavior, and transaction data. The improvements in the recall, precision, and error reduction in the different datasets show that the adversarial patterns have multi-scale properties that cannot be easily recognized by a single learning approach. The variance reduction, nonlinear pattern sensitivity, and margin-based boundary learning provided by the ensemble learning approach are directly responsible for its robustness against the varied adversarial patterns.

In contrast to previous single-model and homogeneous ensemble methods, the findings indicate that heterogeneous integration improves detection stability and generalization capabilities, especially within environments that are challenged by class imbalance and high-dimensional feature spaces. This outcome defies the conventional wisdom that suggests that more complex models are better suited for adversarial detection, instead indicating that model diversity is the more important factor. The robust performance of the framework on different digital ecosystem datasets also indicates its ability to adapt to different operational scenarios without requiring significant model reconfiguration. Nevertheless, the current approach is based on static datasets and offline analysis, which may not be as responsive to real-time adversarial adaptation within dynamic digital ecosystems. Improving the approach to include online learning, temporal modeling, and adaptive ensemble weighting is a promising direction for further advancing robust adversarial pattern recognition in dynamic digital ecosystems.

6. CONCLUSION

This research presents a scalable heterogeneous ensemble learning framework specifically designed for adversarial pattern recognition in high-dimensional digital identity ecosystems. By integrating Random Forest, Gradient Boosting, and Support Vector Machines into a composite architecture, the framework reliably isolates subtle and dynamic malicious behaviors across diverse identity datasets, achieving superior generalization and stability compared to standalone models. By combining multi-scale feature representations with a hybrid hard/soft voting mechanism, the framework simultaneously addresses variance and bias while enhancing sensitivity to rare, high-dimensional anomalies. Empirical evaluation demonstrates a maximum accuracy of 96.8%, a recall of 93.9%, and an F1-score of 94.5%, confirming that the complementary strengths of bagging, boosting, and margin-maximization produce a highly discriminative decision boundary for adversarial recognition. These results reveal that model heterogeneity and complementary learning strategies are critical for robust detection of adversarial patterns, establishing a principled and practical foundation for resilient digital identity security. Future work will extend the framework with adaptive ensemble weighting and online learning modules to address the temporal evolution of adversarial strategies, ensuring sustained robustness in dynamic, non-

10.48047/jocaaa.2019.27.05.03

stationary digital ecosystems.

REFERENCES

[1] G. Wolfond, "A Blockchain Ecosystem for Digital Identity: Improving Service Delivery in Canada's Public and Private Sectors," *Technol. Innov. Manag. Rev.*, vol. 7, no. 10, pp. 35–40, 2017, doi: 10.22215/timreview/1112.

[2] C. Maple, "Security and privacy in the internet of things," *J. Cyber Policy*, vol. 2, no. 2, pp. 155–184, May 2017, doi: 10.1080/23738871.2017.1366536.

[3] P. Asuquo *et al.*, "Security and Privacy in Location-Based Services for Vehicular and Mobile Communications: An Overview, Challenges, and Countermeasures," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4778–4802, 2018, doi: 10.1109/JIOT.2018.2820039.

[4] J. Yang, C. Zhou, S. Yang, H. Xu, and B. Hu, "Anomaly Detection Based on Zone Partition for Security Protection of Industrial Cyber-Physical Systems," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4257–4267, 2018, doi: 10.1109/TIE.2017.2772190.

[5] O. Day and T. M. Khoshgoftaar, "A survey on heterogeneous transfer learning," *J. Big Data*, vol. 4, no. 1, Dec. 2017, doi: 10.1186/s40537-017-0089-0.

[6] X. Zhu and Y. Badr, "Identity management systems for the internet of things: A survey towards blockchain solutions," *Sensors*, vol. 18, no. 12, 2018, doi: 10.3390/s18124215.

[7] S. Magomedov, S. Pavelyev, I. Ivanova, A. Dobrotvorsky, M. Khrestina, and T. Yusubaliev, "Anomaly detection with machine learning and graph databases in fraud management," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 11, pp. 33–38, 2018, doi: 10.14569/IJACSA.2018.091104.

[8] P. R. R. De Souza, K. J. Matteussi, J. C. S. Dos Anjos, J. D. D. Dos Santos, C. F. R. Geyer, and A. Da Silva Veith, "Aten: A dispatcher for big data applications in heterogeneous systems," in *Proceedings - 2018 International Conference on High Performance Computing and Simulation, HPCS 2018*, 2018, pp. 546–553. doi: 10.1109/HPCS.2018.00098.

[9] M. A. Khan, S. Khan, B. Shams, and J. Lloret, "Distributed flood attack detection mechanism using artificial neural network in wireless mesh networks," in *Security and Communication Networks*, John Wiley and Sons Inc., Oct. 2016, pp. 2715–2729. doi: 10.1002/sec.1204.

[10] S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," in *Procedia Computer Science*, 2015, pp. 708–713. doi: 10.1016/j.procs.2015.08.220.

[11] M. Kang, "Machine Learning: Anomaly Detection," in *Prognostics and Health Management of Electronics: Fundamentals, Machine Learning, and the Internet of Things*, wiley, 2018, pp. 131–162. doi: 10.1002/9781119515326.ch6.

[12] N. Grozev and R. Buyya, "Performance modelling and simulation of three-tier applications in Cloud and Multi-Cloud environments," *Comput. J.*, vol. 58, no. 1, pp. 1–22, 2015, doi: 10.1093/comjnl/bxt107.

[13] A. Lavin and S. Ahmad, "Evaluating real-time anomaly detection algorithms - The numenta anomaly benchmark," in *Proceedings - 2015 IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015*, 2016, pp. 38–44. doi: 10.1109/ICMLA.2015.141.

[14] L. Puggini and S. McLoone, "An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data," *Eng. Appl. Artif. Intell.*, vol. 67, pp. 126–135, 2018, doi: 10.1016/j.engappai.2017.09.021.

[15] J. Inoue, Y. Yamagata, Y. Chen, C. M. Poskitt, and J. Sun, "Anomaly detection for a water treatment system using unsupervised machine learning," in *IEEE International Conference on Data Mining Workshops, ICDMW*, 2017, pp. 1058–1065. doi: 10.1109/ICDMW.2017.149.

[16] N. Japkowicz and M. Shah, "Performance Evaluation in Machine Learning," in *Machine Learning in Radiation Oncology: Theory and Applications*, Springer International Publishing, 2015, pp. 41–56. doi: 10.1007/978-3-319-18305-3_4.