International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 14, No.1 (Oct-23)

http://dx.doi.org/10.12785/ijcds/1401107

# Exploring the Role of Machine Learning in Enhancing Access Control Systems: A Comprehensive Review

**Maryem Ait el hadj**

*Laboratory for Sustainable Innovation and Applied Research (L.I.D.R.A), Universiapolis, Agadir, Morocco*

*Received 23 Jul. 2023, Revised 24 Sep. 2023, Accepted 27 Sep. 2023, Published 1 Oct. 2023*

**Abstract:** Access control (AC) systems are crucial for safeguarding sensitive data and resources, yet the increasing complexity of dynamic environments has underscored the need for more accurate and efficient solutions. Therefore, there is a growing need to enhance the accuracy, efficiency, and decision-making capabilities of AC systems. Machine learning (ML) techniques offer promising solutions to address these challenges and automate access control processes. This paper conducts a systematic review of ML techniques in AC systems, involving a comprehensive analysis covering the identification and classification of ML models and their specific applications. Through a meticulous examination of 62 relevant studies published between 2000 and 2023, the review reveals a predominant focus on innovative solutions and adaptations to enhance access control decision-making. Comparative studies play a significant role in assessing different ML approaches, with a focus on identifying the most effective methods for addressing access control challenges. Notably, Attribute-Based Access Control (ABAC) and Role-Based Access Control (RBAC) emerge as the predominant models, while Support Vector Machine (SVM) tops the list as the most commonly employed ML technique, followed by Random Forest (RF) and Decision Tree (DT). The review evaluates the performance and effectiveness of ML models in AC systems, highlighting their strengths and limitations. Additionally, it addresses significant challenges in the field and identifies potential directions for future research. This systematic review provides valuable insights into the current state of ML-based access control systems, fostering further advancements in this vital domain.

**Keywords:** Access Control, Machine Learning, Systematic Review, Empirical Studies, Dataset, Evaluation metrics

## 1. INTRODUCTION

Machine Learning (ML) has emerged as a powerful and successful approach for tackling complex problems across diverse domains [1]. Unlike manually designed processes, ML has the ability to automatically capture intricate properties and extract relevant information. It excels in handling large datasets, identifying patterns, and making accurate predictions or decisions based on learned patterns. By leveraging computational algorithms, ML surpasses the capabilities of human-designed systems and reveals hidden insights within data. ML has revolutionized fields such as image recognition [2], [3], natural language processing [4], recommendation systems [5], [6], and medical diagnosis [7], among others. Similar to other domains. ML techniques have revolutionized access control research, leading to the development of innovative solutions that outperform traditional approaches in terms of robustness and effectiveness [8], [9], [10]. ML has found application in various areas within access control, such as attribute extraction, policy mining [11], [12], and automating decision-making [13], [14]. These advancements have enabled access control systems to tackle complex challenges and improve their overall performance.

To support researchers and practitioners in selecting suitable ML algorithms for access control and identifying areas for improvement in the development of ML-based access control systems, it is crucial to have an exhaustive overview of the application of ML in this domain. However, the current lack of such comprehensive information makes it challenging to gain in-depth insights and plan accordingly. Moreover, the limited research efforts in determining the most appropriate ML method for specific access control problems further complicate the selection of the best model for novel problems. In light of these challenges, the decision was made to conduct a systematic review to thoroughly investigate the current state of ML-based access control.

This paper presents a systematic review and analysis of ML approaches in the field of access control systems. By considering specific requirements and characteristics of access control systems, researchers can identify the ML models that align with their objectives and achieve optimal

*E-mail: maryem.aitelhadj@e-polytechnique.ma*

TABLE I. Research questions

| ID | Research questions | Motivation |
|---|---|---|
| RQ1 | How has the frequency of using *ML* techniques in *AC* systems changed over time? | Identify the publication trend of ML application in AC systems studies over time. |
| RQ2 | What are the main publication sources? | Supply researchers with a list of relevant studies on the application of ML in AC system . |
| RQ3 | What research types were applied? | Identify the different research types used in the studies on the application of ML in AC system. |
| RQ4 | What empirical approaches were applied? | Identify the empirical approaches that have been used to validate the application of ML in AC system. |
| RQ5 | What access control models were used | Identify and classify the models used in the selected studies to express security policies |
| RQ6 | What ML technique were implemented? | Identify and classify the common ML techniques widely used in access control-related research |
| RQ7 | What datasets were used? | Identify the datasets employed including their types. |
| RQ8 | What performance metrics were used? | identify the most used measurement metrics to evaluate the accuracy of the selected studies |

results. Furthermore, the identification of relevant databases and evaluation metrics helps researchers in designing robust experiments and accurately assessing the performance of ML-based access control systems. Overall, this study offers practical guidance to researchers, facilitating the advancement of ML applications in the field of access control. This study contributes to a deeper understanding of the intersection between machine learning and access control, highlighting the benefits, limitations, and potential research opportunities.

The remainder of this paper is organized as follows, Section 2 provides the methodology employed to conduct the study, including research questions, selection criteria and process of data extraction. In section 3, the findings and results of the analysis, categorization, and evaluation of the ML-based access control systems are presented. Section 4 presents an overall discussion and open challenges. Finally, in section 5 conclusions and directions for future work are presented.

## 2. RESEARCH METHODOLOGY

In this paper, we used the recommended guidelines proposed by Petersen et al. [15] to carry out a systematic mapping study. The purpose is to provide a broad overview of a research area, determine the presence of research evidence on a specific topic, and assess the quantity of available evidence. The results of a mapping study can identify suitable areas for conducting systematic literature reviews and also areas where primary studies are more appropriate.

The methodology employed in this study encompassed the following steps: a) formulation of research questions to address the review's issues, b) identification of search terms, selection of sources for the search process, c) application of inclusion and exclusion criteria to select the set of primary studies, d) mapping of publications by extracting relevant data from each selected study, and e) synthesis of data by grouping overall results to facilitate analysis and provide answers to the research questions. A detailed description of each of these steps is presented in the following subsections, providing an in-depth explanation of the methodology employed in this study.

### A. Research Questions

The main objective of this paper is to present a comprehensive summary, analysis, and evaluation of the application of machine learning (ML) in access control (AC) systems. To accomplish this goal, the paper addresses eight specific research questions (*RQs*) as outlined in Table I. These *RQs* serve as a framework for organizing the research area based on the properties and categories specified and defined in the next sections. By formulating the *RQs* in a structured manner, the paper aims to systematically explore the research area and offer valuable insights into the application of ML in access control systems.

### B. Search Strategy

In order to identify pertinent studies concerning the implementation of machine learning in the domain of access control and to address our research inquiries, we conducted a systematic search comprising three distinct steps. Initially, we meticulously formulated a search string and identified pertinent search terms that encompassed the fundamental concepts of machine learning and access control. Subsequently, we employed these search terms to target specific electronic databases, with the aim of retrieving potential studies for further examination. Finally, we executed a comprehensive search procedure designed to guarantee the inclusion of all applicable studies. Elaborate explanations of these steps are provided below, delineating our approach to acquiring a comprehensive set of studies for our analysis.

*1) Search Terms*

The search terms utilized in this study were carefully selected based on the research questions and included keywords and their synonyms. The main search terms employed were: "access control," "Management," "Machine learning," "empirical," and "technique". Table II presents these main search terms along with their corresponding synonyms and alternatives. It is important to note that this systematic study did not consider any synonyms for "machine learning" The search terms were combined using boolean operators, with "OR" used to connect synonymous terms and retrieve records containing any of the terms, and "AND" used to connect the main terms and retrieve records containing all of them [16]. The complete set of search terms was formulated as follows:

("access control" OR "security policy") AND (management* OR administration* OR regulation* OR operation* OR enforcement*) AND "machine learning" AND (empirical* OR evaluation* OR validation* OR experiment* OR "case study" OR survey) AND (method* OR technique* OR model* OR tool* OR approach* OR algorithm*).

TABLE II. Search terms

| Main terms | Alternative terms |
|---|---|
| Access Control | Security Policy |
| Management | Administration, regulation, operation, enforcement |
| Machine Learning | – |
| Empirical | Evaluation, validation, experiment, case study, survey |
| Technique | Method, technique, model, tool, approach, algorithm |

*2) Literature Resources*

To address our research questions, we conducted an automated search utilizing specified search terms in various digital libraries, including IEEE Digital Library, Science Direct, Springer Link, ACM Digital Library, Google Scholar, Scopus, and DBLP. The selected databases were chosen due to their comprehensive coverage, relevance to the research field, and established reputation for hosting high-quality academic publications, ensuring a well-rounded and credible dataset for the study. The search was restricted to articles published between 2000 and the first quarter of 2023. In each digital library, the searches were performed separately, targeting the title, abstract, and keywords of the articles. This approach optimizes the retrieval of highly relevant literature while minimizing the inclusion of less pertinent or tangential works. It is important to note that the selection of search terms was tailored to the search engine capabilities of each electronic database. For instance, in some cases, a straightforward query combining terms such as "access control" and "machine learning" was employed.

*3) Search Process*

In order to ensure the integrity and reliability of the review, a meticulous and rigorous selection process was conducted on the candidate papers identified during the initial search phase. Each paper underwent careful evaluation based on a predefined set of inclusion and exclusion criteria, which served as objective guidelines for determining its eligibility for inclusion or rejection. The criteria were carefully designed to select only relevant papers that align with the research objectives, thus upholding the quality and validity of the review.

*C. Study Selection*

To ensure the reliability of the review, a rigorous selection process was conducted on the candidate papers identified during the initial search phase. Each paper was evaluated based on a set of predefined inclusion and exclusion criteria, which served as guidelines for deciding whether to retain or reject a paper.

- **Inclusion Criteria (IC)**: The paper should be an original study that focuses on the utilization of machine learning techniques to address various access control problems. The publication should provide adequate explanation of the research findings and results. The publication year should fall within the range of 2000 to 2023.
- **Exclusion criteria (EC)**: Papers presenting only conceptual ideas, magazine publications, interviews, and discussion papers were excluded. Secondary research, review papers, and other non-relevant publications were excluded. Papers not written in English were excluded. In the case of multiple publications covering the same study, only the most comprehensive and recent publication was included in the review.

A paper was retained if it met all of the inclusion criteria and none of the exclusion criteria. Conversely, a paper was rejected if it met at least one of the exclusion criteria. This strict filtering process aimed to ensure that only eligible and relevant papers meeting the research objectives were included in the review.

While conducting our systematic review, we recognized that there are certain limitations inherent in this methodology [17], [18], [19]. For instance, our choice of search terms and databases, while carefully considered, may have introduced some bias and possibly omitted relevant studies . To mitigate this, we employed a comprehensive search strategy and conducted thorough sensitivity analyses. Additionally, we accounted for potential publication bias by using established methods to assess it [20]. It is important to acknowledge these limitations and take steps to address them in order to ensure the validity and reliability of our findings [17].

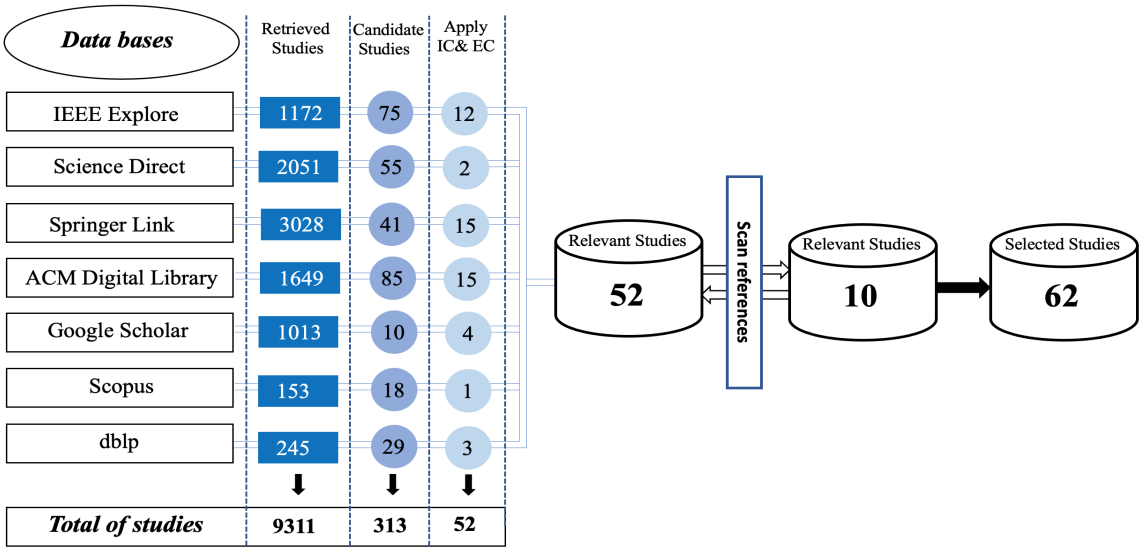


Figure 1. Search process steps and results

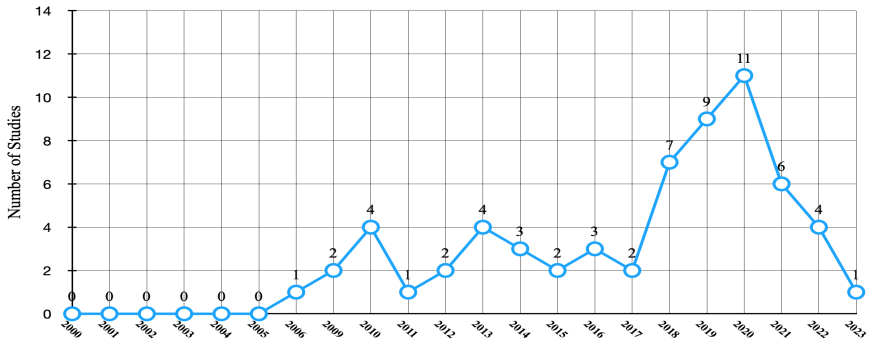


Figure 2. Distribution of selected studies per year

## 3. MAPPING RESULTS

To carry out the study, we followed the planned process. This involved retrieving relevant data, selecting studies, extracting data, and synthesizing information as per the predefined review protocol. Firstly, the protocol was utilized to search for studies focusing on the application of machine learning in access control models. After removing any duplicates, the selected studies underwent a comprehensive evaluation by reading their full texts to determine their eligibility for inclusion in the primary study list.

Figure 1 presents the search steps undertaken and their corresponding outcomes: (1) Conducting the search across seven online databases yielded a total of 9311 studies. (2) After eliminating duplicate studies, the number of candidate studies reduced to 313. (3) Applying the predefined inclusion and exclusion criteria led to the identification of 52 relevant studies. (4) A comprehensive review of the references and citations resulted in the inclusion of 10 additional studies, resulting in a total of 62 relevant studies.

In this section, the results obtained from the review of the 62 primary studies are presented and analyzed, offering insights and answers to the research questions $RQ1 - 8$.

### A. Publication Years (RQ1)

Figure 2 displays the distribution of the selected studies per year. Before 2006, we did not find any ML-based access control solutions. Interest in using ML increased slowly from 2006 to 2017, reached a peak in 2020 (11 studies). The results show that the scientific community pursues ML-based access control solutions. Therefore, with the increasing interest in using machine learning benefits in access control systems, a review that provides an extensive analysis of previous research becomes even more essential. The reason only one study is shown for 2023 in Figure

2 is because, at the time of conducting this review, most of the published studies for that year were not yet available online. Therefore, the data presented in the figure represents a limited snapshot of the studies that were accessible at the time of the review. It is possible that more studies on ML-based access control were conducted in 2023 but were not included in the figure due to their unavailability online during the review period.

### B. Publication Sources (RQ2)

Figure 3 graphically presents the distribution of the selected studies across various publication sources. Out of the total of 62 selected studies, 20 (32%) were published in journals, 26 (42%) were presented at conferences, 14 (23%) were published in symposiums, and one study (2%) was published as both a workshop paper and a book chapter. The distribution of the selected studies over publication sources is further detailed in Table III. Notably, only three journals (MTA, IEEE Access, and IEEE TDSC), three conferences (DBSec, ACSAC, and TrustCom), and two symposiums (SACMAT and ASIACCS) had more than one selected study. The remaining 35 publication sources each had only one study and were grouped under the category of "Others". SACMAT stands out as the publication source with the highest number of selected studies, totaling 9 studies. This represents approximately 14.5% of the papers included in the research study.

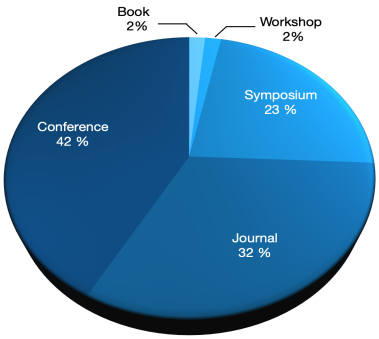


Figure 3. Distribution of selected studies by publication source

### C. Research Types (RQ3)

Based on the selected studies, we identified two main types of research conducted as presented in Table IV. The first type is solution proposal (SP), which involves the introduction of new techniques or the adaptation of existing ones. The second type is evaluation research (ER), which focuses on the assessment and comparison of existing techniques to enhance access control decisions [21]. Figure 4 illustrates the distribution of these research types, indicating that SP was the most frequently employed, accounting for 54 studies or 88% of the selected studies. ER, on the other hand, comprised 8 studies or 12% of the total.

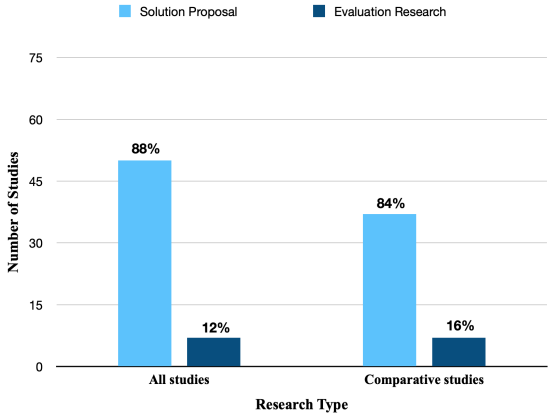


Figure 4. Research types of the selected studies

These findings suggest that the primary objective of researchers in this domain was to propose innovative solutions or modifications to existing techniques (SP). Additionally, they conducted evaluations and comparisons of different ML approaches to improve access control decisions (ER). It is noteworthy that out of the 62 selected studies, 48 (77%) performed comparative studies to identify the most relevant and accurate ML approaches for addressing various access control challenges. Among these, 40 studies (84%) belonged to the SP category, while 8 studies (16%) fell under the ER category. These results highlight the significance of both proposing novel techniques and evaluating their effectiveness in addressing access control issues using machine learning.

### D. Empirical Approaches for Validating ML Application in the Access Control Systems (RQ4)

Based on the selected studies, we identified two main types of empirical approaches as presented in Table V. Figure 5 illustrates the two primary empirical approaches used to validate the application of machine learning models in the access control domain: experiments (Ex) and case studies (CS). It can be observed that the experiment approach was more commonly employed, with 56 studies (90%) validating their ML models under controlled conditions. On the other hand, 6 studies (10%) relied on case studies for validation. Furthermore, Table VI provides a chronological perspective, indicating that the usage of both experiment and case study approaches has demonstrated an increasing trend over time. This trend suggests a growing interest in conducting empirical research to validate ML models in the access control domain, with a specific emphasis on controlled experiments to assess performance and effectiveness.

### E. Access Control Models (RQ5)

The selected studies encompassed a variety of access control models used to express security policies. These

TABLE III. Publication sources

| Publication source | Type | # of studies | Proportion |
|---|---|---|---|
| IEEE Access Journal | Journal | 3 | 4.8% |
| Multimedia Tools and Applications Journal | Journal | 2 | 3.2% |
| IEEE Transactions on Dependable and Secure Computing | Journal | 2 | 3.2% |
| International Symposium on Access Control Models and Technologies (SACMAT) | Symposium | 9 | 14.5% |
| International Symposium on Information, Computer and Communications Security (ASIACCS) | Symposium | 2 | 3.2% |
| International Conference on Data and Applications Security and Privacy(DBSec) | Conference | 5 | 8% |
| Annual Computer Security Applications Conference (ACSAC) | Conference | 2 | 3.2% |
| International Conference on Trust, Security, and Privacy in Computing and Communications (TrustCom) | Conference | 2 | 3.2% |
| Others (conference, symposium, journal, workshop, book) | | 1 each source | 56.7 % |

TABLE IV. Research types

| Research types | Description [22] |
|---|---|
| SP | Empirical studies in which a ML-based solution to design access control policies is proposed, either as a new technique or as a significant adaptation of an existing one, or propose a solution to a defined problem. |
| ER | Empirical studies that evaluate and/or compare existing techniques. |

TABLE V. Empirical approaches

| Empirical approaches | Description [22] |
|---|---|
| Ex | An empirical process applied under controlled conditions to evaluate the application of ML technique in access control systems. |
| CS | An empirical study that investigates the application of ML in access control in a real-life context. |

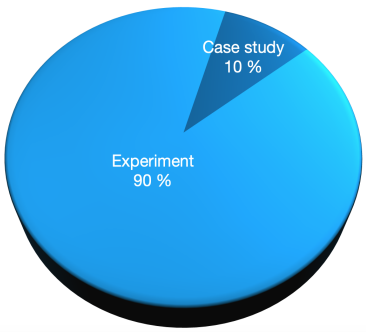


Figure 5. Empirical approaches for validating ML-based Access Control models

TABLE VI. Temporal Distribution of Empirical Approaches for Validating ML Application in the Access Control Domain

| Empirical approach | 2000-2006 | 2007-2013 | 2014-2023 | Total |
|---|---|---|---|---|
| Experiment (Ex) | 1 | 11 | 44 | 56 |
| Case Study (CS) | 0 | 2 | 4 | 6 |

models include ABAC (Attribute-Based Access Control), ATBAC (Attribute Tree Based Access Control), BBAC (Behavior Based Access Control), CAAC (Context Aware Access Control), DTAC (Decision Trees Access Control), FirewallAC, IDAC (Intent-Driven Access Control), NLAC (Natural Language Access Control), PAAC (Differential Privacy Based Access Control), PBAC (Provenance Based Access Control), RBAC (Role Based Access Control), ReBAC (Relationship Based Access Control), SBAC (Situation Based Access Control), SVM-BSAC (SVM Based Smart Access Control), TBAC (Trust Based Access Control), TVRB-BSP (Time Varying Risk Budget Based Security Policy), DLBAC (Deep Learning Based Access Control), and MLBAC (Machine Learning Based Access Control).

Table VII provides insights into the distribution of studies across these access control models. ABAC emerged as the most frequently utilized model, with 18 studies (29%), followed by RBAC with 14 studies (22%). NLAC and ReBAC were employed in three studies each (5%), while BBAC, RiskBAC, TBAC, and SBAC were utilized in two studies each (3%). ATBAC, CAAC, DTAC, FirewallAC, IDAC, PAAC, PBAC, SVM-BSAC, TVRB-BSP, DLBAC,

TABLE VII. Distribution of access control models

| AC Model | Used in | # of studies | Percentage |
|---|---|---|---|
| ABAC | [23], [24], [25], [26], [12], [27], [28], [29], [10], [30], [31], [32], [33], [34], [13], [35], [36], [37] | 18 | 29 % |
| ATBAC | [38] | 1 | 2 % |
| BBAC | [39], [40] | 2 | 3 % |
| CAAC | [41] | 1 | 2 % |
| DTAC | [42] | 1 | 2 % |
| FirewallAC | [43] | 1 | 2 % |
| IDAC | [44] | 1 | 2 % |
| NLAC | [45], [46], [47] | 3 | 5 % |
| PAAC | [48] | 1 | 2 % |
| PBAC | [49] | 1 | 2 % |
| RBAC | [50], [51], [52], [53], [54], [55], [56], [57], [58], [59], [60], [61], [8], [62] | 14 | 22 % |
| ReBAC | [9], [11] , [25] | 3 | 5 % |
| RiskBAC | [63], S75 | 2 | 3 % |
| SBAC | [56], [64] | 2 | 3 % |
| SVM-BSAC | [65] | 1 | 2 % |
| TBAC | [66], [67] | 2 | 3 % |
| TVRB-BSP | [68] | 1 | 2 % |
| DLBAC | [14] | 1 | 2 % |
| MLBAC | [69] | 1 | 2 % |
| Not Identified | [70], [71], [72], [73], [74], [75], [76] | 7 | 11 % |

and MLBAC were each used in one study (2%). Some studies may have employed multiple access control models, such as [25], which utilized both ABAC and ReBAC. It is important to note that the "Not Identified" group of studies did not specify the type of access control model employed. The high percentage of ABAC usage can be attributed to its generalized and scalable nature, making it a widely adopted access control model in the domain.

Figure 6 illustrates the distribution of research types (RQ3), empirical approaches (RQ4), and access control models (RQ5). Several notable observations can be made from the figure:

- The most frequently conducted experimental studies in the selected research employed the ABAC model for solution proposal. This was followed by the RBAC model, which was utilized in 11 studies for solution proposal and three studies for evaluation research. These findings indicate a strong focus on exploring the application of ML in access control through experimental approaches, with particular attention given to the ABAC and RBAC models.
- Furthermore, it can be observed that a majority of access control models employed experiment as the empirical approach. In contrast, case study approaches were less commonly utilized, with only one study employing it for evaluation research and five studies for solution proposal (as depicted in Figure 7).

These findings highlight the dominance of experimental studies in investigating the application of ML in the access control domain. The ABAC model emerged as a popular choice for proposing solutions, while RBAC was also widely used for both solution proposal and evaluation research. Additionally, the preference for experimental approaches indicates a focus on controlled settings to assess the effectiveness and performance of access control models.

### *F. Machine Learning Techniques (RQ6)*

From the 62 selected primary studies, we have identified a range of machine learning methods utilized in the research. These methods include AdaBoost (Adaptive Boosting), AHC (Agglomerative Hierarchical Clustering), AL (Adaptive Learning), ARM (Association Rule Mining), ASSISTANT'86, ATM (Author-Topic Model), BNMF (Bayesian nonnegative matrix factorization), C4.5, CNN (Convolutional neural network), CNN-LSTM (CNN Long Short Term Memory), DBSCAN (density-based spatial clustering of applications with noise), DT (Decision Tree), EA (evolutionary algorithm), ET (Extra Trees), FIS (Fuzzy inference system), GA (Genetic Algorithm), GB (Gradient Boosting), GMM (Gaussian mixture model), GNB (Gaussian Naive Bayes), K-Means, K-modes, KNN (K Nearest Neighbors), LDA (Linear Discriminant Analysis), LPCA (Logistic Principal Component Analysis), LR (Logistic Regression), LSIA (Limited Search Induction Algorithm), LSTM (Long Short Term Memory), LSVC (Linear Support Vector Classification), MLP (Multilayer Perceptron), NC (Nearest Centroid), NMF (non-negative matrix factorization), NN (Neural network), RBMs (Restricted

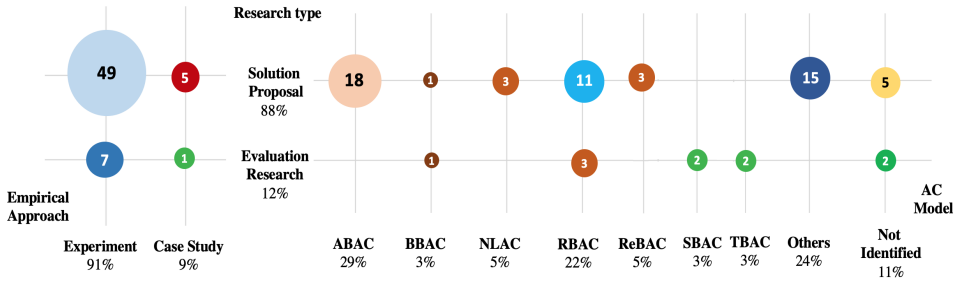


Figure 6. Frequency of research types, empirical approaches and access control models

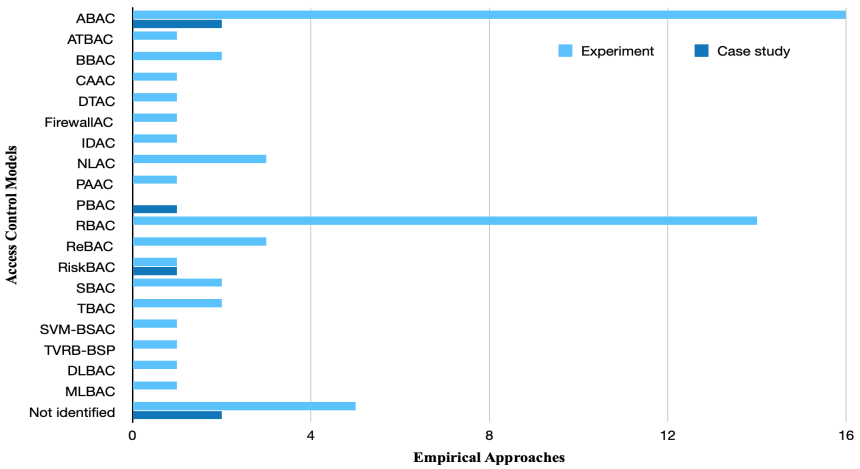


Figure 7. Frequency of access control models per empirical approach

Boltzmann Machines), RF (Random Forest), RNN (Recurrent Neural Network), RR (Ridge Regression), SVD (Singular Value Decomposition), SVM (Support Vector Machine),XGB (XGBoost), ResNet (Residual Network), and DenseNet (Densely Connected Convolutional Network).

Figure 8 presents a comprehensive overview of the various machine machine learning techniques identified in the selected studies.This categorization, based on the functionality of the methods, facilitates the identification of the predominant approaches utilized in the field of access control.

Figure 9 and Table VIII provide a comprehensive overview of the frequency of usage for these machine learning methods. Among the identified methods, SVM stands out as the most frequently employed, appearing in 16 studies (25% of the total). RF was utilized in 14 studies (22%), DT in nine studies (14%), KNN in seven studies (11%), and CNN, GNB, and LR each appeared in five studies (8%). Additionally, C4.5 and NN were utilized in four studies each (7%), while K-Means was used in three studies (4%). DBSCAN, LDA, FIS, and ResNet were each employed in two studies (3%), while the remaining methods were reported only once. It is worth noting that some studies may have utilized multiple machine learning methods to address their research objectives. This analysis offers a comprehensive investigation into the utilization of machine learning methods within the domain of access control. The results reveal that SVM, RF, DT have emerged as the predominant choices in the selected studies. The prominence of these methods underscores their efficacy and applicability in addressing access control challenges. These findings provide valuable contributions to the existing body of knowledge, offering insights for researchers and practitioners in the field.

Based on the findings derived from the examination of comparative studies conducted in response to RQ3, Table IX provides an inclusive overview of these studies, highlighting the compared methods and the corresponding comparison results, with a specific focus on identifying the most efficient methods. The analysis of Table IX reveals that RF consistently outperformed other methods in eight studies, demonstrating its superiority. Likewise, KNN was identified as the most efficient method in three studies, while the SVM approach exhibited its effectiveness in two studies. It is important to note that the remaining methods were reported as the most effective in only one study each, as denoted in the table. These results contribute valuable insights into

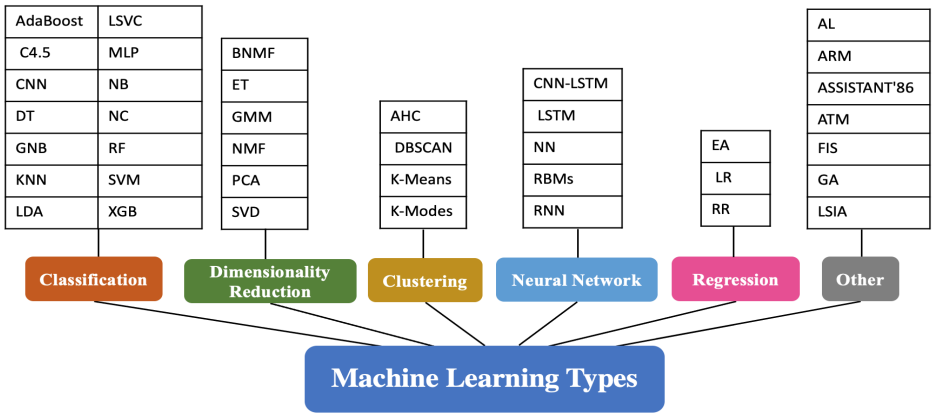


Figure 8. Machine Learning Types

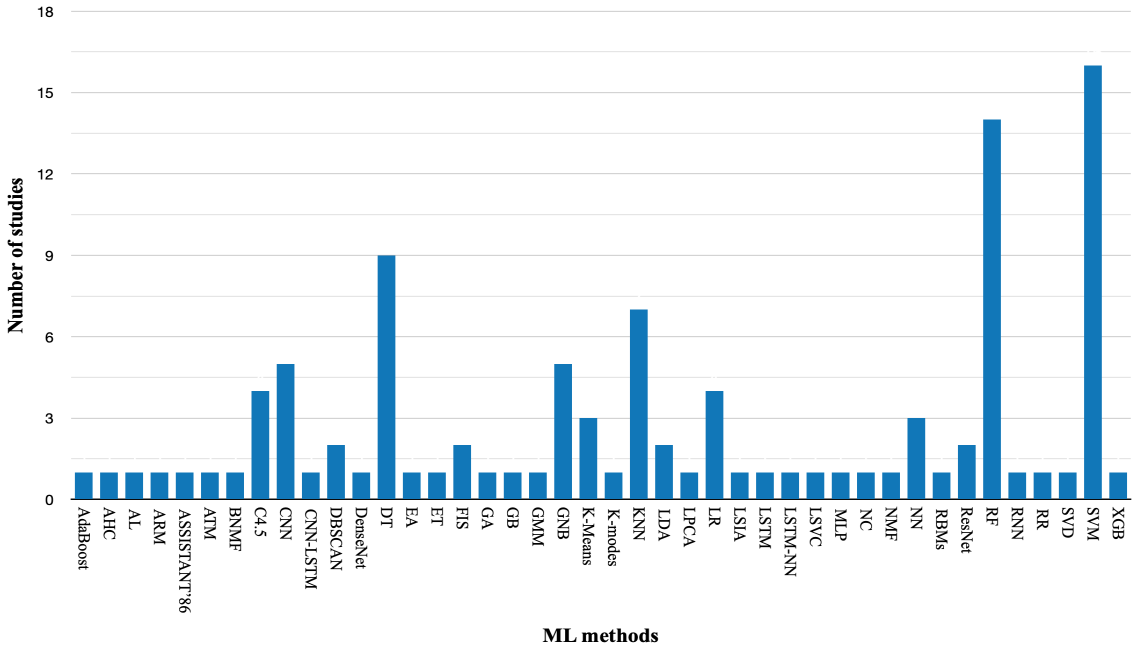


Figure 9. Number of studies per machine learning methods

the comparative performance of various machine learning methods within the scope of the conducted research.

### G. *Datasets (RQ7)*

The selected studies encompassed a diverse array of datasets, which can be categorized into two principal groups:

- Real dataset (RD): These datasets originate from actual organizations, policy documents, or applications. They may either be publicly accessible or obtained from private sources that are not publicly available.
- Synthetic dataset (SynD): The creation of these datasets involves diverse methodologies, including data augmentation techniques, merging data from multiple sources, and the synthetic construction of datasets from scratch.

The utilization of both real and synthetic datasets enables researchers to conduct a comprehensive evaluation of machine learning algorithms in the access control domain. This approach allows for a holistic assessment of algorithm performance and efficacy. Real datasets provide insights into system behavior in real-world scenarios, allowing researchers to measure the practical applicability of the algorithms. On the other hand, synthetic datasets offer a controlled environment for exploring and examining specific access control scenarios. By combining both types

TABLE VIII. Distribution of machine learning methods

| ML method | Used in | # of studies | Percentage |
|---|---|---|---|
| SVM | [51], [39], [56], [65], [46], [64], [40], [59], [38], [74], [8], [75], [62], [76], [77],[37] | 16 | 25 % |
| RF | [50], [41], [44], [26], [64], [29], [59], [60], [10], [33], [34], [13], [37], [69] | 14 | 22 % |
| DT | [9], [42], [25], [71], [72], [49], [66], [33], [77] | 9 | 14 % |
| KNN | [41], [45] , [46], [66], [59], [10], [34] | 7 | 11 % |
| CNN | [23], [57], [28], [59], [76] | 5 | 8 % |
| GNB | [56], [66], [74], [34], [37] | 5 | 8 % |
| C4.5 | [55], [27], [73], [48] | 4 | 6 % |
| LR | [56], [66], [74], [34], [37] | 5 | 8 % |
| K-Means | [39], [12], [58] | 3 | 4 % |
| NN | [11] , [47], [33], [37] | 4 | 6 % |
| DBSCAN | [31], [77] | 2 | 3 % |
| LDA | [54], [74] | 2 | 3 % |
| FIS | [63], [67] | 2 | 3 % |
| ResNet | [14], [69] | 2 | 3 % |
| AdaBoost, AHC, AL, ARM, ASSISTANT'86, ATM, BNMF, CNN-LSTM, EA, ET, GA, GB, GMM, K-modes, LPCA, LSIA, LSTM, LSVC, MLP, NC, NMF, RBMs, RNN, RR, SVD, XGB, DenseNet, LSTM-NN | [8], [77], [43], [70], [73], S15, [53], [61], [68], [33], [52], [33], [77], [35], [53], [73], [26], [24], [66], [66], [66], [53], [32], [66], [53], [33], [14], [36] (respectively) | 1 each method | 45 % |

TABLE IX. Summary of ML methods reported to be superior

| Ref | Compared techniques | Techniques reported superior |
|---|---|---|
| [41] | KNN, GNB, RF | KNN, RF |
| [53] | BNMF, LPCA, NMF, SVD | LPCA |
| [56] | LR, SVM | SVM |
| [46] | KNN, GNB, SVM | KNN |
| [26] | LSTM, RF | LSTM |
| [64] | GNB, RF, SVM | RF |
| [58] | NNMF, Convex NNMF, K-Means, C-Means | K-Means |
| [73] | ASSISTANT'86, C4.5, LSIA | C4.5 |
| [66] | DT, GNB, KNN, LR, LSVC, MLP, NC, RR | KNN, RF+KNN |
| [59] | CNN, KNN, RF, SVM, GNB | SVM, RF |
| [74] | LDA, LR, SVM | LR |
| [76] | CNN, SVM | CNN |
| [34] | KNN, LR, RF | RF |
| [37] | GNB, LR, NN, RF, SVM | RF |
| [14] | ResNet, DenseNet, SVM, RF, MLP | ResNet, DenseNet |
| [69] | ResNet, RF | RF |
| [67] | FIS, KNN, NC, GNB, DT, LSVC, LR, RC | FIS |

of datasets, researchers gain a deeper understanding of algorithm performance under various conditions, leading to advancements in the field of access control through machine learning.

The majority of studies (61%) relied on real datasets, highlighting their prominence in the domain. Synthetic datasets were employed in 21% of the studies, enabling researchers to address limitations or augment the available data. Interestingly, a hybrid approach combining both real and synthetic datasets was adopted in 14.5% of the studies, emphasizing the importance of leveraging multiple data sources. It is worth noting that two studies (i.e. [55] and [77]) did not specify the datasets used.

Researchers often face challenges when using either real-world datasets or synthetic datasets alone in the context of access control studies. Real-world datasets may possess

TABLE X. Number of studies per dataset source

| Dataset | Used in | Percentage |
|---|---|---|
| Domino, Firewall; Healthcare | [60], [53] | 3 % |
| Amazon (AZ) | [10], [13], [37], [14] | 6 % |
| Collected ACP Documents | [23], [45], [24], [30], [31],[47] | 9 % |
| CyberChair | [23], [45], [24], [30], [31], [47] | 9 % |
| electronic medical records (EMR) | [9], [11] , [25] | 4 % |
| IBM Course Management App | [23], [45], [24], [30], [31], [47] | 9 % |
| iTrust | [23], [50], [45], [24], [46], [30], [31], [47] | 12 % |
| The workforce management case study | [9], [11] , [25] | 4 % |
| Health Care Sample Policy | [9], [11] , [25] | 4 % |
| Project Management Sample Policy | [9], [11] , [25], [10] | 4 % |
| University Sample Policy | [9], [11] , [53], [25], [33] | 8 % |

limitations, such as data quality issues or incomplete access control information, that hinder their suitability for comprehensive analysis. Similarly, synthetic datasets may not capture the full range of access control scenarios present in real-world environments. To address these limitations, researchers frequently adopt a hybrid approach, combining both real-world and synthetic datasets. This approach allows for a more comprehensive exploration of access control systems, facilitating a deeper understanding of their performance and effectiveness. Furthermore, researchers may modify sample policies to align them with the specific access control model under investigation, ensuring the relevance and applicability of the datasets used in the study.

For in-depth analysis and to further identify the relevant datasets used in empirical studies, Table X presents the number and percentage of studies per dataset source. The datasets listed in the table are those that have been used in more than one study. The most frequently used dataset is iTrust [78], accounting for 12% of the studies. This is followed by IBM Course Management App [79], CyberChair [80], and Collected ACP Documents [81], each representing 9% of the studies. The University Sample Policy dataset [82] is used in 8% of the studies. It should be noted that some studies may have used more than one dataset. For example, [9] utilized several datasets to support its proposed approach.

### *H. Performance Metrics (RQ8)*

In the selected primary studies, a range of performance and evaluation metrics were employed, which can be categorized into three main groups based on their application. The distribution of these metrics per category is presented in Table XI.

- Access control system metrics: These metrics assess the overall security properties, quality, and support of access control models [83], [84]. Examples include anonymization time (AnonyT), average permission risk prediction (APRP), authorization time (AuthT), completeness, coverage, error in role permission (ERP), error in user role (EUR), generality, inferred AC rules rate (InACRR), incompleteness, number of requests denied/permitted (NRD/P), percentage of controlled requests (PCR), risk-adjusted utility (RAUti), syntactic similarity, policy semantic similarity (SS/PSS), scalability, stability, and weighted structural complexity (WSC).

- Machine learning metrics: These metrics are specifically designed to evaluate machine learning models [85]. They include accuracy, area under the ROC curve (AUC), F-score, false negative rate (FNR), false positive rate (FPR), mean absolute error (MAE), median error rate (MER), precision, recall, root mean square error (RMSE), receiver operating characteristic (ROC), role prediction accuracy (RPA), and weighted average of precision (WAP).

- Miscellaneous metrics: This category encompasses various metrics that do not fit into either the access control system or machine learning metrics. Examples of such metrics are complexity, fitness, and processing time (PT).

Despite the smaller number of machine learning (ML) metrics compared to access control system (ACS) metrics, the overall number of reported metrics in the ML category across all studies is remarkably high, as depicted in Figure 10. This observation can be attributed to several factors, which are further explained below.

Figure 11 and Table XII provide a breakdown of the frequently used ML metrics in the selected studies. The analysis reveals that recall was the most commonly employed metric, appearing in 28 studies (45%). Accuracy was the second most prevalent metric, reported in 27 studies (43.5%). F-score and precision were each utilized in 21 studies (34%), while FPR appeared in 10 studies (16%). Other metrics such as WSC, PT, SS/PSS, complexity, FNR, and RMSE were used in varying frequencies, ranging from three to five studies. Metrics like AUC, ROC, and FNR were employed in two studies each, and the remaining metrics were reported only once. It is important to note that multiple

TABLE XI. Distribution of evaluation metrics per category

| Metrics category | metrics | # of metrics | Percentage |
|---|---|---|---|
| ACS metrics | AnonyT, APRP, AuthT, completeness, coverage, ERP, EUR, generality, InACRR, incompleteness, NRD/P, PCR, RAUti, SS/PSS, scalability, stability, WSC | 17 | 51.6 % |
| ML metrics | Accuracy, AUC, F-score, FNR, FPR, MAE, MER, precision, recall, RMSE, ROC, RPA, WAP | 13 | 39.4 % |
| Miscellaneous metrics | complexity, fitness, PT | 3 | 9 % |

TABLE XII. Number of studies per dataset performance metrics

| Performance metrics | Used in | # of studies | Percentage |
|---|---|---|---|
| Recall | [23], [41], [44], [53], [45], [24], [42], [70], [39], [56], [46], [26], [71], [64], [40], [28], [43], [66], [10], [30], [31], [47], [62], [33], [76], [13], [37], [67] | 28 | 45 % |
| Accuracy | [52], [53], [70], [39], [65], [12], [71], [72], [64], [40], [57], [28], [43], [66], [59], [10], [61], [38], [48], [8], [62], [33], [34], [13], [77],[37], [69] | 27 | 43 % |
| F-score | [23],[44], [45], [24], [42], [46], [64], [40], [28], [66], [10], [30], [31], [47], [33], [76], [13], [35], [37], [14], [67] | 21 | 34 % |
| Precision | [23], [45], [24], [42], [46], [71], [64], [40], [28], [43], [66], [10], [30], [31], [47], [33], [76], [34], [37], [14], [67] | 21 | 34 % |
| FPR | [41], [53], [70], [39], [12], [71], [64], [74], [62], [14] | 10 | 16 % |
| WSC | [9], [11], [25], [35], [32] | 5 | 8 % |
| PT | [52], [58], [66] | 3 | 5 % |
| SS/PSS | [9], [11] , [25] | 3 | 5 % |
| Complexity | [27], [49], [77] | 3 | 5 % |
| FNR | [53], [74], [14] | 3 | 5 % |
| RMSE | [72], [34], [67] | 3 | 5 % |
| ROC | [26], [29] | 2 | 3 % |
| MAE | [66], [67] | 2 | 3 % |
| AUC | [26], [71] | 2 | 3 % |
| AnonyT, APRP, AuthT, completeness, coverage, ERP, EUR, fitness, generality, InACRR, incompleteness, MER, NRD/P, PCR, RAUti, RPA, scalability, stability, WAP | [48], [60], [55], [49], [54], [58], [58], [61], [54], [50], [73], [68], [55], [10], [75], [51], [63], [54], [44] (respectively) | 1 each metric | 30 % |

evaluation metrics were often used within a single study, supporting the results depicted in Figure 10.

The prominence of recall, accuracy, F-score, and precision as the most commonly employed metrics can be attributed to their widespread recognition and relevance within the machine learning field. These metrics hold well-defined interpretations and serve as reliable indicators for evaluating the performance of machine learning models. Researchers frequently rely on these metrics to assess the effectiveness of their models across different tasks, including classification and prediction.

It is important to mention that various performance metrics within the three groups (access control system, machine learning, and miscellaneous) interact and influence each other as a whole. This comprehensive perspective helps in understanding the broader implications of performance optimization. Certainly, it is undeniable that alterations or enhancements in one metric can have a ripple effect on others, contingent upon the specific objectives for the study. For example, when considering user-permission assignments or authorized requests [53], [74], [14], it is essential to note that emphasizing a decrease in the False Positive Rate (FPR) could lead to a reduction in legitimate transactions being mistakenly identified as fraudulent. However, this approach may inadvertently elevate the False Negative Rate (FNR) and lead to quicker processing times due to the use of a simpler model. Conversely, if the aim shifts towards maxi-

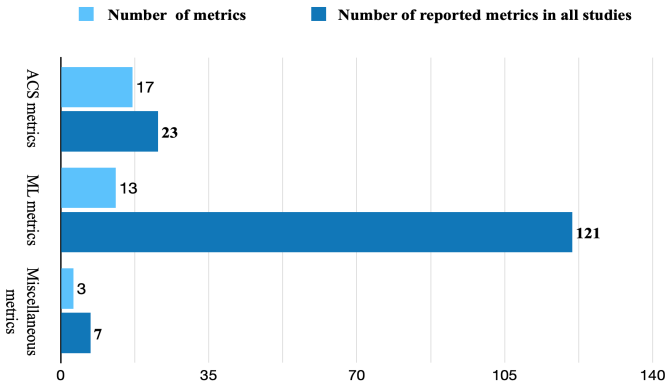


Figure 10. Frequency of evaluation metrics per category

mizing FNR to capture almost all fraudulent transactions, it might elevate FPR and processing times due to the adoption of a more cautious model. Achieving a balance between minimizing FPR and FNR requires meticulous fine-tuning, influencing both metrics while considering processing time. Ultimately, the study objectives steer the optimization process, shaping the interplay of performance metrics in attaining the desired outcomes. In a nutshell, considering how metrics interact with each other is instrumental in achieving established objectives for optimizing access control systems enhanced by machine learning.

## 4. Discussion

The analysis of the selected studies investigating the utilization of machine learning in the access control domain reveals several key findings. Researchers predominantly concentrated on presenting inventive solutions or adaptations to existing techniques in order to enhance access control decision-making. Comparative studies were performed to assess and compare various machine learning approaches, with a considerable portion of the studies focusing on determining the most pertinent and precise methods for tackling access control challenges. Among several access control models, ABAC emerged as the most commonly employed, followed by RBAC. These models offer adaptable and scalable frameworks for the management of access control.

Regarding machine learning methods, Support Vector Machine (SVM) emerged as the predominant choice among the selected studies, followed by Random Forest (RF) and Decision Tree (DT). Additionally, other methods such as K-Nearest Neighbors (KNN), Convolutional Neural Network (CNN), Gaussian Naive Bayes (GNB), and Logistic Regression (LR) were also observed. SVM's prevalence in the access control domain can be attributed to its ability to perform binary classification makes it well-suited for determining access authorization. Moreover, SVM's capability to handle non-linear decision boundaries enables it to capture intricate relationships in access control scenarios. The well-established nature and familiarity of SVM among researchers make it a natural choice for addressing access control challenges. However, the choice of the machine learning algorithm should depend on specific requirements, the nature of the access control problem, and the available data. Alternative algorithms such as Decision Trees, Random Forests, or Neural Networks may also be suitable depending on the context and objectives of the access control system.

Research studies have provided compelling evidence showcasing numerous advantages of employing machine learning models to enhance the accuracy of access control decision-making. To provide valuable insights, we combine RQ5 (Access Control Models) and RQ6 (ML Techniques) and provide a consolidated summary of how ML techniques are applied to various access control models. From the studies we have reviewed, it is apparent that researchers have been applying machine learning methods in an ad hoc manner, addressing specific cases individually, without the presence of a standardized strategy for the utilization of machine learning within the access control domain. Nonetheless, it is worth noting that numerous researchers have employed machine learning algorithms for the purpose of mining access control policies. In addition, there has been the development of a probabilistic model to tackle the role mining problem, and this model is primarily driven by machine learning techniques [32], [14], [9], [11], [53]. The integration of user behavior into access control systems, particularly ABAC model, has been explored in various studies [34], [39]. These studies employ machine learning to analyze user behavior, enabling the dynamic adaptation of access control policies based on observed actions [54], [32], [35]. By analyzing historical access logs, geolocation data, and time of access, ML-driven access control conducts a comprehensive assessment, allowing for finely-grained decisions like granting or denying access based on dynamic risk assessments [63], [60], [75].

In order to illustrate these concepts, consider the following example: traditional access control systems, based on predefined roles and static permissions, are superseded by dynamic, adaptive models [86], [87]. Machine learning continually scrutinizes user behavior, discerning nuanced patterns in access, resource utilization, and data retrieval. For instance, an employee's typical activities are characterized and learned over time, enabling the system to identify deviations or unusual behavior that might signify a security threat. Access permissions are no longer static; they evolve in real-time, adapting to the changing roles, responsibilities, and requirements of individual users. This adaptive approach enhances security by proactively identifying and mitigating potential threats, all while ensuring operational efficiency through streamlined access management. The integration of machine learning into access control in this corporate network setting represents a leap forward in cybersecurity and user access management.

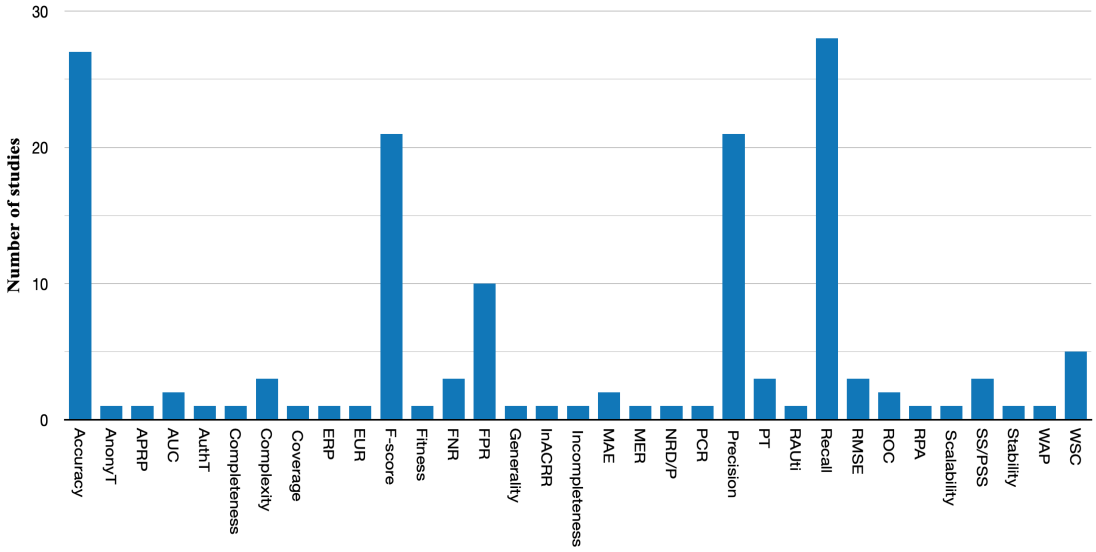


Figure 11. Number of studies per evaluation metrics

In the context of our review, where Role-Based Access Control (RBAC) is a commonly employed model, the integration of machine learning brings transformative changes to access control in several domains [61], [50], [51]. Traditional RBAC structures are replaced as ML algorithms continuously analyze user behavior and access patterns. For example, healthcare professionals like doctors, nurses, and administrative staff are no longer confined to static roles but are dynamically assigned permissions based on their real-time behavior and evolving contextual factors. ML plays a pivotal role in detecting anomalies, responding to unusual access patterns, and conducting dynamic risk assessments, including recognizing non-standard access hours for additional authentication. Additionally, the system operates proactively, identifying potential security threats and autonomously restricting access while notifying security personnel when abnormal activities are detected. This adaptive and data-driven approach not only bolsters security but also streamlines operational efficiency by reducing the administrative burden of manual policy adjustments, ensuring that access control remains agile and responsive.

The advantages of integrating machine learning into access control models can be summarized as follows:

- Adaptability: Machine learning enables access control systems to adapt in real-time to evolving user behavior and emerging security threats.
- Enhanced Security: ML-driven systems can detect and respond to anomalies, reducing the risk of unauthorized access and potential security breaches.
- Dynamic Policies: Machine learning facilitates the creation of adaptive security policies, automatically adjusting permissions based on user behavior and contextual factors.
- Efficiency: Automation and proactive threat detection reduce the administrative burden of manual policy adjustments, enhancing operational efficiency.
- Data-Driven Decisions: ML models make access decisions based on data-driven insights, ensuring a more informed approach to access control.
- Risk Assessment: Access control systems can perform dynamic risk assessments, considering multiple factors for access requests, improving decision accuracy.
- Scalability: ML-based solutions can scale to handle large and complex access control environments effectively.

In essence, machine learning algorithms possess the capability to analyze massive volumes of data and uncover hidden patterns and correlations that may not be easily discernible to humans. By leveraging historical access data and user behavior, machine learning models can acquire valuable insights, enabling them to make informed and intelligent access control decisions. As a result, machine learning significantly enhances the accuracy of determining whether access should be granted or denied [11], [12], [54]. The complexity of access control systems, characterized by dynamic environments and a multitude of factors including user attributes, resource attributes, and contextual information, can be effectively managed by machine learning techniques. Classification, regression, and clustering are examples of machine learning methods that are capable of handling this complexity and capturing intricate relationships between various access control attributes. These

techniques provide the means to analyze and model the complex interdependencies within access control systems, enabling more accurate and effective decision-making [68], [43].

Despite the positive impact of machine learning on access control, there are still several challenges and promising areas for future research in this field. These challenges and research directions encompass:

- Effective administration is vital for managing access authorization changes, policy configurations, and access control-related attributes (creation, adjustment and updates) [88], [89]. Administering access control systems is necessary to accommodate changes and ensure their continued operation. However, when it comes to ML-based access control systems, administration poses unique challenges. It involves updating the corresponding ML model to incorporate policy and configuration changes, as well as adjusting access control-related information like user attributes and resource attributes to meet evolving system requirements. Unfortunately, there is currently a lack of comprehensive methodologies and frameworks that address administration issues from both ML and access control perspectives [90].

- High-quality datasets are essential for effective research and development in the domain of machine learning-based access control. Nevertheless, the current availability of publicly accessible access control datasets is constrained, and the existing ones often exhibit imbalanced class distributions and lack comprehensive access control information. To overcome these challenges, researchers rely on synthetic dataset [12], [61]. In some cases, a hybrid approach that combines real-world datasets with synthetic ones is adopted to address the limitations of the available data [9], [25]. However, many of the existing datasets from real-world organizations suffer from high anonymization or incompleteness in terms of access control information. To advance in ML-based access control, it is imperative to acquire high-quality datasets that accurately represent the semantic and granularity aspects of permissions, while also capturing the complete access control state of a system.

- The integration of machine learning in access control presents new challenges, including adversarial attacks [91], [92]. Adversarial attacks aim to manipulate input data in a manner that deceives the machine learning model, resulting in incorrect or undesirable outcomes. In the context of access control, these attacks exploit vulnerabilities in the decision-making process of the machine learning model to gain unauthorized access. Additionally, attribute-hiding attacks can further undermine the access control system by concealing or removing crucial information that should be considered during the decision-making process. Consequently, it is imperative to thoroughly investigate and comprehend these attacks from an access control perspective in order to enhance the design of machine learning models.

Addressing these open challenges and pursuing these future research directions will contribute to the advancement and practical implementation of ML-based access control systems, enhancing their effectiveness, security, and usability in various domains and applications.

## 5. Conclusion and Future Guidelines

This paper emphasizes the importance of machine learning (ML) in the field of access control systems. ML has proven to be a powerful approach for addressing complex decision-making processes and considering multiple attributes. By leveraging ML techniques, access control systems can efficiently analyze large datasets, detect patterns, and make intelligent decisions based on learned patterns. The study investigates current ML research trends in access control systems and provides a novel taxonomy for categorizing and evaluating ML-based access control models. This taxonomy enables researchers to select the most suitable ML model for their access control systems, choose appropriate databases, and consider relevant evaluation metrics. The findings of this research contribute to a deeper understanding of the intersection between machine learning and access control, highlighting the benefits and limitations of ML-based approaches. Moreover, it offers valuable insights for researchers to make informed decisions and advance the application of ML in the field of access control.

Directions for future work in the application of ML in access control systems should focus on several key aspects. First, there is a need to develop more comprehensive and diverse datasets that accurately represent real-world access control scenarios. Additionally, the advancement of ML techniques specifically tailored for access control, along with improved interpretability and explainability of ML models, is crucial. The establishment of standardized evaluation metrics and performance benchmarks for ML-based access control systems is necessary to facilitate fair comparisons and objective assessments. Furthermore, research should address the robustness and security of ML models against adversarial attacks and vulnerabilities. The integration of domain knowledge into ML models, the practical deployment considerations, and the development of guidelines for implementation in real-world scenarios are also important areas for future investigation. By addressing these directions, researchers can contribute to the advancement and effectiveness of ML-based access control systems, ensuring the security and integrity of sensitive data and resources.

## References


[1] X. Li, W. Lin, and B. Guan, "The impact of computing and machine learning on complex problem-solving," *Engineering Reports*, vol. 5, no. 6, 2023.

[2] P. Rani, S. Kotwal, J. Manhas, V. Sharma, and S. Sharma, "Machine learning and deep learning based computational approaches in automatic microorganisms image recognition: methodologies, challenges, and developments," *Archives of Computational Methods in Engineering*, vol. 29, no. 3, pp. 1801–1837, 2022.

[3] W. Ma, X. Tu, B. Luo, and G. Wang, "Semantic clustering based deduction learning for image recognition and classification," *Pattern Recognition*, vol. 124, p. 108440, 2022.

[4] I. Lauriola, A. Lavelli, and F. Aiolli, "An introduction to deep learning in natural language processing: Models, techniques, and tools," *Neurocomputing*, vol. 470, pp. 443–456, 2022.

[5] R. Zheng, L. Qu, B. Cui, Y. Shi, and H. Yin, "Automl for deep recommender systems: A survey," *arXiv preprint arXiv:2203.13922*, 2022.

[6] A. H. Khan, J. Siddqui, and S. S. Sohail, "A survey of recommender systems based on semi-supervised learning," in *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 3*. Springer, 2022, pp. 319–327.

[7] M. Shehab, L. Abualigah, Q. Shambour, M. A. Abu-Hashem, M. K. Y. Shambour, A. I. Alsalibi, and A. H. Gandomi, "Machine learning in medical applications: A review of state-of-the-art methods," *Computers in Biology and Medicine*, vol. 145, p. 105458, 2022.

[8] L. Zhou, C. Su, Z. Li, Z. Liu, and G. P. Hancke, "Automatic fine-grained access control in scada by machine learning," *Future Generation Computer Systems*, vol. 93, pp. 548–559, 2019.

[9] T. Bui and S. D. Stoller, "A decision tree learning approach for mining relationship-based access control policies," in *Proceedings of the 25th ACM Symposium on Access Control Models and Technologies*, 2020, pp. 167–178.

[10] A. A. Jabal, E. Bertino, J. Lobo, M. Law, A. Russo, S. Calo, and D. Verma, "Polisma-a framework for learning attribute-based access control policies," in *European Symposium on Research in Computer Security*. Springer, 2020, pp. 523–544.

[11] T. Bui, S. D. Stoller, and H. Le, "Efficient and extensible policy mining for relationship-based access control," in *Proceedings of the 24th ACM Symposium on Access Control Models and Technologies*, 2019, pp. 161–172.

[12] L. Karimi and J. Joshi, "An unsupervised learning based approach for mining attribute based access control policies," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 1427–1436.

[13] A. Liu, X. Du, and N. Wang, "Efficient access control permission decision engine based on machine learning," *Security and Communication Networks*, vol. 2021, 2021.

[14] M. N. Nobi, R. Krishnan, Y. Huang, M. Shakarami, and R. Sandhu, "Toward deep learning based access control," in *Proceedings of the Twelfth ACM Conference on Data and Application Security and Privacy*, 2022, pp. 143–154.

[15] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and software technology*, vol. 64, pp. 1–18, 2015.

[16] A. Idri, F. azzahra Amazal, and A. Abran, "Analogy-based software development effort estimation: A systematic mapping and review," *Information and Software Technology*, vol. 58, pp. 206–230, 2015.

[17] J. K.Owens, "Systematic reviews:brief overview of methods,limitations,and resources," *Nurse Author EdItion. 31:69–72. doi: 10.1111/nae2.28*, 2021.

[18] L. Uttley, D. S. Quintana, P. Montgomery, C. Carroll, M. J. Page, L. Falzon, A. Sutton, and D. Moher, "The problems with systematic reviews: a living systematic review," *Journal of Clinical Epidemiology*, 2023.

[19] M. Gusenbauer and N. R. Haddaway, "Which academic search systems are suitable for systematic reviews or meta-analyses? evaluating retrieval qualities of google scholar, pubmed, and 26 other resources," *Research synthesis methods*, vol. 11, no. 2, pp. 181–217, 2020.

[20] A. Grewal, H. Kataria, and I. Dhawan, "Literature search for research planning and identification of research problem," *Indian journal of anaesthesia*, vol. 60, no. 9, p. 635, 2016.

[21] R. Wieringa, N. Maiden, N. Mead, and C. Rolland, "Requirements engineering paper classification and evaluation criteria: a proposal and a discussion," *Requirements engineering*, vol. 11, pp. 102–107, 2006.

[22] S. Elmidaoui, L. Cheikhi, A. Idri, and A. Abran, "Empirical studies on software product maintainability prediction: a systematic mapping and review," *E-Informatica Software Engineering Journal*, vol. 13, no. 1, 2019.

[23] M. Alohaly, H. Takabi, and E. Blanco, "A deep learning approach for extracting attributes of abac policies," in *Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies*, 2018, pp. 137–148.

[24] M. Narouei, H. Khanpour, H. Takabi, N. Parde, and R. Nielsen, "Towards a top-down policy engineering framework for attribute-based access control," in *Proceedings of the 22nd ACM on Symposium on Access Control Models and Technologies*, 2017, pp. 103–114.

[25] T. Bui and S. D. Stoller, "Learning attribute-based and relationship-based access control policies with unknown values," in *International Conference on Information Systems Security*. Springer, 2020, pp. 23–44.

[26] A. Alkhresheh, K. Elgazzar, and H. S. Hassanein, "Adaptive access control policies for iot deployments," in *2020 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 2020, pp. 377–383.

[27] R. A. Shaikh, K. Adi, L. Logrippo, and S. Mankovski, "Inconsistency detection method for access control policies," in *2010 Sixth International Conference on Information Assurance and Security*. IEEE, 2010, pp. 204–209.

[28] A. Liu, X. Du, and N. Wang, "Unstructured text resource access control attribute mining technology based on convolutional neural network," *IEEE Access*, vol. 7, pp. 43 031–43 041, 2019.

[29] L. Argento, A. Margheri, F. Paci, V. Sassone, and N. Zannone, "Towards adaptive access control," in *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, 2018, pp. 99–109.

[30] M. Narouei, H. Khanpour, and H. Takabi, "Identification of access control policy sentences from natural language policy documents," in *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, 2017, pp. 82–100.

[31] M. Alohaly, H. Takabi, and E. Blanco, "Automated extraction of attributes from natural language attribute-based access control (abac) policies," *Cybersecurity*, vol. 2, no. 1, pp. 1–25, 2019.

[32] D. Mocanu, F. Turkmen, A. Liotta *et al.*, "Towards abac policy mining from logs with deep learning," in *Proceedings of the 18th International Multiconference, ser. Intelligent Systems*, 2015.

[33] V. Gumma, B. Mitra, S. Dey, P. S. Patel, S. Suman, and S. Das, "Pammela: Policy administration methodology using machine learning," *arXiv preprint arXiv:2111.07060*, 2021.

[34] M. Afshar, S. Samet, and H. Usefi, "Incorporating behavior in attribute based access control model using machine learning," in *2021 IEEE International Systems Conference (SysCon)*. IEEE, 2021, pp. 1–8.

[35] L. Karimi, M. Aldairi, J. Joshi, and M. Abdelhakim, "An automatic attribute based access control policy extraction from access logs," *IEEE Transactions on Dependable and Secure Computing*, 2021.

[36] E. Psarra, D. Apostolou, Y. Verginadis, I. Patiniotakis, and G. Mentzas, "Context-based, predictive access control to electronic health records," *Electronics*, vol. 11, no. 19, p. 3040, 2022.

[37] M. You, J. Yin, H. Wang, J. Cao, K. Wang, Y. Miao, and E. Bertino, "A knowledge graph empowered online learning framework for access control decision-making," *World Wide Web*, vol. 26, no. 2, pp. 827–848, 2023.

[38] D. Ye, Y. Mei, Y. Shang, J. Zhu, and K. Ouyang, "Mobile crowdsensing context aware based fine-grained access control mode," *Multimedia Tools and Applications*, vol. 75, no. 21, pp. 13 977–13 993, 2016.

[39] A. Adler, M. J. Mayhew, J. Cleveland, M. Atighetchi, and R. Greenstadt, "Using machine learning for behavior-based access control: Scalable anomaly detection on tcp connections and http requests," in *MILCOM 2013-2013 IEEE Military Communications Conference*. IEEE, 2013, pp. 1880–1887.

[40] B. Tay and A. Mourad, "Intelligent performance-aware adaptation of control policies for optimizing banking teller process using machine learning," *IEEE Access*, vol. 8, pp. 153 403–153 412, 2020.

[41] M. Miettinen, S. Heuser, W. Kronz, A.-R. Sadeghi, and N. Asokan, "Conxsense: automated context classification for context-aware access control," in *Proceedings of the 9th ACM symposium on Information, computer and communications security*, 2014, pp. 293–304.

[42] C. Xiang, Y. Wu, B. Shen, M. Shen, H. Huang, T. Xu, Y. Zhou, C. Moore, X. Jin, and T. Sheng, "Towards continuous access control validation and forensics," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 2019, pp. 113–129.

[43] M. Noferesti and R. Jalili, "Acope: An adaptive semi-supervised learning approach for complex-policy enforcement in high-bandwidth networks," *Computer Networks*, vol. 166, p. 106943, 2020.

[44] M. L. Rahman, A. Neupane, and C. Song, "Iac: On the feasibility of utilizing neural signals for access control," in *Proceedings of the 34th Annual Computer Security Applications Conference*, 2018, pp. 641–652.

[45] J. Slankas, X. Xiao, L. Williams, and T. Xie, "Relation extraction for inferring access control rules from natural language artifacts," in *Proceedings of the 30th Annual Computer Security Applications Conference*, 2014, pp. 366–375.

[46] J. Slankas and L. Williams, "Access control policy extraction from unconstrained natural language text," in *2013 International Conference on Social Computing*. IEEE, 2013, pp. 435–440.

[47] M. Narouei, H. Takabi, and R. Nielsen, "Automatic extraction of access control policies from natural language documents," *IEEE Transactions on Dependable and Secure Computing*, vol. 17, no. 3, pp. 506–517, 2018.

[48] N. Metoui and M. Bezzi, "Differential privacy based access control," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 2016, pp. 962–974.

[49] J. Pei and X. Ye, "Towards policy retrieval for provenance based access control model," in *2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE, 2014, pp. 769–776.

[50] H. T. Le, C. D. Nguyen, L. Briand, and B. Hourte, "Automated inference of access control policies for web applications," in *Proceedings of the 20th ACM Symposium on Access Control Models and Technologies*, 2015, pp. 27–37.

[51] W. Zhang, Y. Chen, C. Gunter, D. Liebovitz, and B. Malin, "Evolving role definitions through permission invocation patterns," in *Proceedings of the 18th ACM symposium on Access control models and technologies*, 2013, pp. 37–48.

[52] I. Saenko and I. Kotenko, "Genetic algorithms for solving problems of access control design and reconfiguration in computer networks," *ACM Transactions on Internet Technology (TOIT)*, vol. 18, no. 3, pp. 1–21, 2018.

[53] I. Molloy, N. Li, Y. Qi, J. Lobo, and L. Dickens, "Mining roles with noisy data," in *Proceedings of the 15th ACM symposium on Access control models and technologies*, 2010, pp. 45–54.

[54] I. Molloy, Y. Park, and S. Chari, "Generative models for access control policies: applications to role mining over logs with attribution," in *Proceedings of the 17th ACM symposium on Access Control Models and Technologies*, 2012, pp. 45–56.

[55] D. Praveena and P. Rangarajan, "A machine learning application for reducing the security risks in hybrid cloud networks," *Multimedia Tools and Applications*, vol. 79, no. 7, pp. 5161–5173, 2020.

[56] A. A. Boxwala, J. Kim, J. M. Grillo, and L. Ohno-Machado, "Using statistical and machine learning to help institutions detect suspicious access to electronic health records," *Journal of the American Medical Informatics Association*, vol. 18, no. 4, pp. 498–505, 2011.

[57] A. Liu, X. Du, and N. Wang, "Recognition of access control role based on convolutional neural network," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*. IEEE, 2018, pp. 2069–2074.

[58] A. Dath and K. Praveen, "Role mining in distributed firewall using matrix factorization methods," in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*. IEEE, 2020, pp. 625–629.

[59] R. Julian, E. Guyot, S. Zhou, G. S. Poh, and S. Bressan, "Role-based access classification: Evaluating the performance of machine learning algorithms," in *Transactions on Large-Scale Data-and Knowledge-Centered Systems XLIII*. Springer, 2020, pp. 1–39.

[60] N. Badar, J. Vaidya, V. Atluri, and B. Shafiq, "Risk based access control using classification," in *Automated Security Management*. Springer, 2013, pp. 79–95.

[61] T.-Y. Kim and S.-B. Cho, "Particle swarm optimization-based cnn-lstm networks for anomalous query access control in rbac-administered model," in *International Conference on Hybrid Artificial Intelligence Systems*. Springer, 2019, pp. 123–132.

[62] Q. Ni, J. Lobo, S. Calo, P. Rohatgi, and E. Bertino, "Automating role-based provisioning by learning from examples," in *Proceedings of the 14th ACM symposium on Access control models and technologies*, 2009, pp. 75–84.

[63] Q. Ni, E. Bertino, and J. Lobo, "Risk-based access control systems built on fuzzy inferences," in *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, 2010, pp. 250–260.

[64] G. Misra, J. M. Such, and H. Balogun, "Improve-identifying minimal profile vectors for similarity based access control," in *2016 IEEE Trustcom/BigDataSE/ISPA*. IEEE, 2016, pp. 868–875.

[65] F. Shan, J. Liu, X. Wang, W. Liu, and B. Zhou, "A smart access control method for online social networks based on support vector machine," *IEEE Access*, vol. 8, pp. 11 096–11 103, 2020.

[66] P. M. Khilar, V. Chaudhari, and R. R. Swain, "Trust-based access control in cloud computing using machine learning," in *Cloud Computing for Geospatial Big Data Analytics*. Springer, 2019, pp. 55–79.

[67] A. Kesarwani and P. M. Khilar, "Development of trust based access control models using fuzzy logic in cloud computing," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 5, pp. 1958–1967, 2022.

[68] Y. T. Lim, P.-C. Cheng, P. Rohatgi, and J. A. Clark, "Dynamic security policy learning," in *Proceedings of the first ACM workshop on Information security governance*, 2009, pp. 39–48.

[69] M. N. Nobi, R. Krishnan, Y. Huang, and R. Sandhu, "Administration of machine learning based access control," in *Computer Security–ESORICS 2022: 27th European Symposium on Research in Computer Security, Copenhagen, Denmark, September 26–30, 2022, Proceedings, Part II*. Springer, 2022, pp. 189–210.

[70] T. Kalbarczyk, C. Liu, J. Hua, and C. Julien, "Lad: Learning access control polices and detecting access anomalies in smart environments," in *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. IEEE, 2019, pp. 485–493.

[71] D. Sun, Z. Wu, Y. Wang, Q. Lv, and B. Hu, "Cyber profiles based risk prediction of application systems for effective access control," in *2019 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 2019, pp. 1–7.

[72] R. Singru, P. Bhandari, K. Patel, P. Mane, and C. Gulhane, "Efficient electronic document access control management using natural language processing," in *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*. IEEE, 2020, pp. 714–719.

[73] R. A. Shaikh, K. Adi, L. Logrippo, and S. Mankovski, "Detecting incompleteness in access control policies using data classification schemes," in *2010 Fifth International Conference on Digital Information Management (ICDIM)*. IEEE, 2010, pp. 417–422.

[74] Y.-W. Seo and K. Sycara, "Cost-sensitive access control for illegitimate confidential access by insiders," in *International Conference on Intelligence and Security Informatics*. Springer, 2006, pp. 117–128.

[75] I. Molloy, L. Dickens, C. Morisset, P.-C. Cheng, J. Lobo, and A. Russo, "Risk-based security decisions under uncertainty," in *Proceedings of the second ACM conference on Data and Application Security and Privacy*, 2012, pp. 157–168.

[76] J. Heaps, R. Krishnan, Y. Huang, J. Niu, and R. Sandhu, "Access control policy generation from user stories using machine learning," in *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, 2021, pp. 171–188.

[77] J. Liu, M. Simsek, B. Kantarci, M. Erol-Kantarci, A. Malton, and A. Walenstein, "Risk-aware fine-grained access control in cyber-physical contexts," *Digital Threats: Research and Practice*, 2021.

[78] A. Meneely, B. Smith, and L. Williams, "Appendix b: itrust electronic health care system case study," *Software and Systems Traceability*, p. 425, 2012.

[79] IBM, "Course registration requirements," *https://khanhn.files.wordpress.com/2016/08/vidu-ibm.pdf*, 2004.

[80] R. Van De Stadt, "Cyberchair: A web-based groupware application to facilitate the paper reviewing process," *arXiv preprint arXiv:1206.1833*, 2012.

[81] X. Xiao, A. Paradkar, S. Thummalapenta, and T. Xie, "Automated extraction of security policies from natural-language software documents," in *Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering*, 2012, pp. 1–11.

[82] Z. Xu and S. D. Stoller, "Mining attribute-based access control policies," *IEEE Transactions on Dependable and Secure Computing*, vol. 12, no. 5, pp. 533–545, 2014.

[83] A. Li, Q. Li, V. C. Hu, and J. Di, "Evaluating the capability and performance of access control policy verification tools," in *MILCOM 2015-2015 IEEE Military Communications Conference*. IEEE, 2015, pp. 366–371.

[84] A. A. Jabal, M. Davari, E. Bertino, C. Makaya, S. Calo, D. Verma, A. Russo, and C. Williams, "Methods and tools for policy analysis," *ACM Computing Surveys (CSUR)*, vol. 51, no. 6, pp. 1–35, 2019.

[85] G. S. Handelman, H. K. Kok, R. V. Chandra, A. H. Razavi, S. Huang, M. Brooks, M. J. Lee, and H. Asadi, "Peering into the

black box of artificial intelligence: evaluation metrics of machine learning methods," *American Journal of Roentgenology*, vol. 212, no. 1, pp. 38–43, 2019.

[86] J. Park and R. Sandhu, "Towards usage control models: beyond traditional access control," in *Proceedings of the seventh ACM symposium on Access control models and technologies*, 2002, pp. 57–64.

[87] J. Qiu, Z. Tian, C. Du, Q. Zuo, S. Su, and B. Fang, "A survey on access control in the age of internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 4682–4696, 2020.

[88] S. Jha, S. Sural, V. Atluri, and J. Vaidya, "An administrative model for collaborative management of abac systems and its security analysis," in *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)*. IEEE, 2016, pp. 64–73.

[89] M. P. Singh, S. Sural, J. Vaidya, and V. Atluri, "A role-based administrative model for administration of heterogeneous access control policies and its security analysis," *Information Systems Frontiers*, pp. 1–18, 2021.

[90] D. Servos and S. L. Osborn, "Current research and open problems in attribute-based access control," *ACM Computing Surveys (CSUR)*, vol. 49, no. 4, pp. 1–45, 2017.

[91] V. Ballet, X. Renard, J. Aigrain, T. Laugel, P. Frossard, and M. Detyniecki, "Imperceptible adversarial attacks on tabular data," *arXiv preprint arXiv:1911.03274*, 2019.

[92] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," *arXiv preprint arXiv:1706.06083*, 2017.

**Maryem Ait el hadj** holds a Ph.D. in the Detection and Resolution of Anomalies in Large Scale Attribute-Based Access Control Policies. She obtained her Ph.D. from ENSIAS at Mohammed V University of Rabat. Currently, she serves as an assistant professor at the Polytechnic School within the International University of Agadir (Universiapolis). Her current research and teaching interests extend to various topics in Computer Science, Data Base Management and Cybersecurity.