Supervised Learning Based Approach for Mining ABAC Rules from Existing RBAC Enabled Systems

Devendra Singh Kushwaha[1] Granthali Sandesh Mahajan[2], Jayed Altaf Shaikh[3], Naeem Julfikar Shaikh[4], Shweta Suresh Kidile[5]

Department of Computer Engineering
Sandip Institute of Technology and Research Centre (SITRC) Nashik, Maharashtra, India
[1]devendra2904@gmail.com, [2]granthalimahajan.work@gmail.com, [3]zayedshaikhd9@gmail.com, [4]naeemshaikh09sss@gmail.com, [5]shwetakidile@gmail.com

Abstract: Role-Based Access Control (RBAC) has been the leading access control model for a long time due to its simplicity and ease of administration. However, the creation of contemporary computing environments has revealed its limitations in dealing with dynamic and context-aware authorization requirements. Attribute-Based Access Control (ABAC) offers more flexibility since access decisions are based on multiple attributes. Nonetheless, the transition from RBAC to ABAC remains complex and costly. Some recent works have studied supervised learning approaches to automate ABAC policy mining, to reduce the manual efforts while improving the policy accuracy. This review paper offers an in-depth analysis of the existing supervised learning based approaches for mining ABAC policies from RBAC enabled systems. It reviews the machine learning algorithms, feature extraction methods, datasets, evaluation metrics and policy generation strategies that are commonly used in recent studies. The literature review indicates that supervised learning can potentially enhance the efficiency and scalability of RBAC-to-ABAC migration. However, there are still some open research issues including limited labeled datasets, feature selection, policy interpretability and model generalization. This review aims to review the current state of the research, identify existing gaps, and discuss future research directions for developing intelligent and scalable ABAC policy mining solutions.

Keywords: *Role-Based Access Control (RBAC), Attribute-Based Access Control (ABAC), Supervised Learning, Machine Learning, ABAC Policy Mining, RBAC-to-ABAC Migration, Access Control Policies, Policy Mining*.

I. INTRODUCTION

Access control is an important aspect of information security, since it determines who may access a resource and what actions they are allowed to perform. The increasing use of cloud computing, Internet of Things (IoT) devices, distributed applications and large-scale information systems has made the management of access permission more and more difficult [2], [8], [17]. Organizations require access control mechanisms that can protect resources and adapt to changes in users, resources and working environments [1], [12]. Role-Based Access Control (RBAC) is one of the most widely used approaches because it provides a relatively simple way to manage permissions with predefined roles [14], [17], [18].

In an RBAC system, permissions are attached to roles and users acquire permissions through the roles assigned to them [14], [17], [18]. This helps to manage access more easily, especially for organizations where members have well-defined responsibilities. However, RBAC is hard to manage when access requirements become more complex. You can have many roles to reflect different combinations of permissions. This leads to the so-called role explosion problem. RBAC also struggles to handle situations where access is determined by factors such as the location of the user, the time of access, the device, or other contextual information [12], [17]. Such restrictions have motivated researchers to investigate more flexible access control models.

One such model is Attribute-Based Access Control (ABAC). ABAC makes access decisions based on attributes of the user, resource, requested action and surrounding environment [1], [2] instead of relying mostly on predefined roles. This enables organizations to create more granular and adaptive access policies. For example, an employee could only be allowed to view a certain document if they are in a certain department, using a certain device, and attempting to access that document during working hours. Such conditions can be more naturally modeled by ABAC than by a huge number of RBAC roles [1], [2].

While ABAC provides more flexibility, it is not trivial to migrate an existing RBAC system to ABAC. Existing systems may have a large number of roles and permission assignments, and translating this information into meaningful attributes and ABAC policies can be a lot of manual work. And it requires a good knowledge of the existing access requirements of the organization. As the system grows in size and complexity, manual policy creation can become unwieldy, and can lead to inconsistent or incorrect policies. This is the reason why researchers have begun to investigate automated methods to mine ABAC policies from existing access control information [3], [4], [13].

This is an area in which machine learning, and in particular supervised learning, is receiving increasing attention. Supervised learning algorithms learn patterns from labeled data and use those patterns to make predictions or classifications [5], [7], [9], [10]. For the migration from RBAC to ABAC, information such as users, roles, permissions, resources, actions and contextual conditions can be used as input features. ABAC policies can be expressed as patterns discovered by machine learning techniques that analyze these relationships. Different learning algorithms and policy mining approaches have been studied in several works to reduce human effort and improve the efficiency of the migration process [3], [4], [5], [11].

Due to these challenges, there is a need for a comprehensive review of existing supervised learning based approaches for ABAC policy mining. This paper reviews and analyzes

current research on mining ABAC policies from RBAC-enabled systems. The review emphasizes the machine learning techniques, feature extraction techniques, datasets, policy mining strategies, and evaluation metrics adopted in the previous studies. It also compares the reviewed approaches according to accuracy, scalability, interpretability, computational efficiency and practical applicability. Finally, the paper identifies important research gaps and discusses potential directions for developing more efficient and scalable approaches for automated ABAC policy mining.[21]-[23]

II. LITERATURE SURVEY

TABLE I. SUMMARY OF LITERATURE SURVEY

Ref.	Author(s), Year	Method / Technique	Major Contribution	Limitation
[1]	Alohaly et al., 2022	ABAC with cyber deception	Integrates cyber-deception techniques with ABAC to improve insider-threat detection and protection of sensitive resources.	Does not address ABAC policy mining or RBAC-to-ABAC transformation.
[2]	Kim et al., 2022	ABAC-based security model	Develops an ABAC security model for DDS to provide flexible authorization in distributed environments.	Does not consider automatic policy extraction or supervised learning.
[3]	Cotrini et al., 2018	ABAC rule mining from sparse logs	Mines ABAC rules from limited access logs and demonstrates automated policy generation from observed access behavior.	Relies on access logs and does not exploit existing RBAC structures or supervised learning.
[4]	Karimi et al., 2022	Unsupervised policy extraction	Automatically extracts ABAC policies from access logs using learning, rule pruning, and policy refinement.	Does not use labeled RBAC information or supervised learning for policy generation.
[5]	Shan et al., 2020	Support Vector Machine (SVM)	Applies supervised machine learning to develop an intelligent access-control mechanism for online social networks.	Focuses on access decisions rather than ABAC policy mining from RBAC systems.
[6]	Bera et al., 2021	AI and blockchain-based access control	Combines AI, blockchain, and access control for malicious-attack detection and mitigation in IoE environments.	Does not focus on ABAC rule extraction or RBAC-to-ABAC conversion.
[7]	Wang et al., 2024	Fully and semi-supervised learning	Demonstrates the use of supervised and semi-supervised learning for accurate and interpretable classification.	Application is medical image diagnosis and is unrelated to access-control policy mining.
[8]	Luo et al., 2026	Graph-based authorization	Introduces a graph-based approach for efficient and verifiable cloud authorization.	Does not specifically address supervised ABAC rule mining from RBAC data.
[9]	An et al., 2026	Conditional self-supervised learning	Combines supervised and self-supervised learning for improved few-shot learning with limited labeled data.	Not applied to RBAC, ABAC, or authorization-policy mining.
[10]	Wang et al., 2021	Semi-supervised and supervised learning	Proposes a self-trained framework using ensemble transformations to improve learning with limited labeled data.	Designed for image classification rather than access-control policy generation.
[11]	Zhong et al., 2025	eBPF-based automated policy generation	Generates control policies automatically from SSH system behavior and reduces manual policy configuration.	Generates system-call-based policies and does not perform ABAC policy mining.
[12]	Feng et al., 2026	LLM and RAG-based access control	Introduces context-aware authorization by integrating LLMs and RAG with RBAC.	Focuses on runtime authorization rather than mining ABAC rules from RBAC systems.
[13]	Das et al., 2018	Hybrid ABAC policy engineering	Combines top-down policy specification with bottom-up access analysis using entropy and frequency measures.	Uses heuristic techniques and does not employ supervised learning for RBAC-to-ABAC conversion.
[14]	Baralla et al., 2025	Blockchain, SSI and RBAC	Integrates blockchain and Self-Sovereign Identity with RBAC for secure resource management.	Remains RBAC-oriented and does not address ABAC policy mining.
[15]	Rooney and Hinders, 2021	ML-based classification	Uses clustering and supervised learning to recognize MAC protocols in communication networks.	Focuses on communication protocols rather than access-control policies.
[16]	Liu et al., 2024	Self-supervised and supervised learning	Investigates pre-trained encoders to improve secure and privacy-preserving supervised learning.	Does not address RBAC, ABAC, or authorization-policy generation.
[17]	Zhang et al., 2025	Privacy-enhanced RBAC	Enhances RBAC for IoT systems using privacy-preserving cryptographic techniques.	Depends on static roles and does not explore ABAC rule generation.
[18]	Xu et al., 2024	RBAC-based authentication	Proposes one-to-many authentication and key negotiation using RBAC for smart-factory environments.	Focuses on authentication and key management rather than ABAC policy mining.
[19]	Pham et al., 2023	Self-supervised representation learning	Uses residual momentum to improve visual representation learning without extensive labeled data.	Focuses on visual learning and has no direct application to access-control policies.
[20]	Li et al., 2021	Weakly supervised learning	Develops a framework for learning with weak, noisy, or incomplete supervision.	General learning framework and does not address RBAC or ABAC policy mining.

Research Gap

The literature review indicates that existing ABAC policy mining techniques are predominantly based on access logs, unsupervised learning or heuristic approaches [3], [4], [13]. Supervised learning has also shown promising results for classification and pattern discovery in many applications [5], [9], [10}. But the use of supervised learning to learn ABAC rules directly from the structured information of an existing RBAC-enabled system has been insufficiently explored.

Therefore, this research proposes a supervised learning-based approach to mine ABAC rules from an existing RBAC-enabled system by learning and generating more flexible attribute-based authorization rules based on the existing information of roles and permissions.[21-[23]

III. PROBLEM STATEMENT AND OBJECTIVES

As the number of users, roles, and permissions grows, managing existing RBAC systems can be challenging. The goal of this research is to develop a supervised learning-based approach to automatically mine flexible ABAC rules from existing RBAC data, thus decreasing manual policy-conversion effort and improving access-control management. The goals are to examine RBAC data, select important attributes, build a supervised learning model, produce ABAC rules and assess their correctness and consistency.

IV. PROPOSED SYSTEM ARCHITECTURE

We propose a system that provides a supervised learning based framework to mine ABAC (Attribute Based Access Control) rules from an existing RBAC (Role Based Access Control) system. The architecture comprises of ten stages beginning with gathering the existing RBAC information and concluding with the deployment of the generated ABAC policies for decision making.

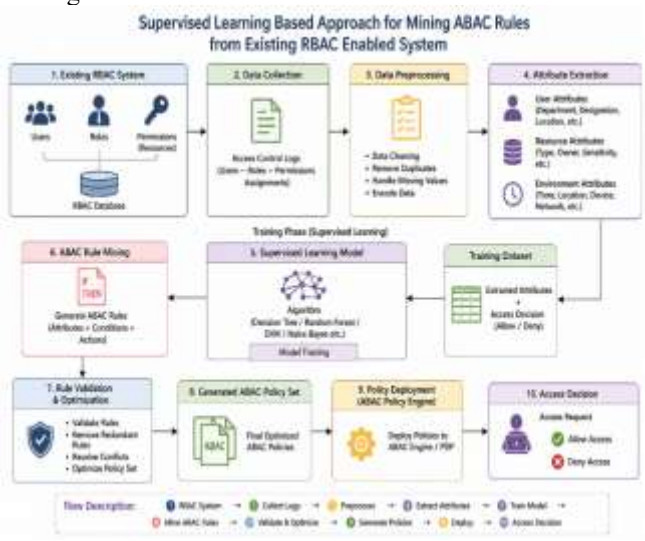


Fig. 1. Proposed System Architecture for Supervised Learning-Based Mining of ABAC Rules from an Existing RBAC-Enabled System.

1. Active RBAC System

The process starts with an existing RBAC-enabled system with information about users, roles and permissions. Roles are assigned to users, and each role has permissions for access to specific resources. This existing RBAC information is the primary source of data required by the proposed system.

2. Data Gathering

The relevant authorization information is collected from RBAC system in the form of access-control records or logs. The information gathered includes relationships between users, roles, permissions and access assignments. The data serves as the basis for learning existing authorization patterns.

3. Data preparation

The collected data is processed by machine learning. This stage also includes data cleaning, removing duplicate records, handling missing values and encoding categorical information in a suitable format. Proper preprocessing helps to maintain consistency and improves the quality of the training data.

4. Extracting Attributes

After preprocessing, the relevant attributes are selected from the available authorization information. These attributes fall into three main categories:
• User information: department, designation, location, etc.
• Resource properties: type of resource, owner, sensitivity, etc.
• Environmental attributes: time, place, device, network, etc.

These attributes set the conditions that could be later used to build ABAC rules.

5. Training Data Base

The extracted attributes are combined with corresponding access decisions (Allow/Deny) to create the training dataset. The dataset provides labeled examples that allow the supervised learning model to learn the mapping from attributes to authorization decisions.

6. Model for supervised learning

The dataset is then fed into a supervised learning algorithm. Decision Tree, Random Forest, Support Vector Machine (SVM) or Naive Bayes can be considered depending on the characteristics of the dataset. In the training phase the model learns the patterns between the available attributes and the corresponding access decisions.

7. ABAC Rules Mining

After the model has learned the authorization patterns, the learned relationships are used for generating ABAC rules. The generated rules are expressed in terms of attributes, conditions and actions. For example, a rule might state that a user from a particular department may perform a particular action on a particular kind of resource under particular conditions.

8. Rules Validation and Optimization

The generated rules are evaluated and then applied in the authorization system. It eliminates redundant rules, identifies and resolves conflicts between rules, and does overall optimization of the policy set. This yields a consistent and manageable set of ABAC policies.

9. ABAC Policy Set Creation and Deployment

The generated rules are the final ABAC policy set after validation and optimization. These policies are then deployed to an ABAC policy engine or PDP (Policy Decision Point). The PDP assesses incoming access requests against the policies that it has generated.

10. Decision on Access

Finally, when a user submits an access request, the ABAC policy engine evaluates the request according to the generated policies and the relevant attributes. The system then makes an Allow Access or Deny Access decision based on learned rules.

Overall Workflow

The complete process can be represented as:

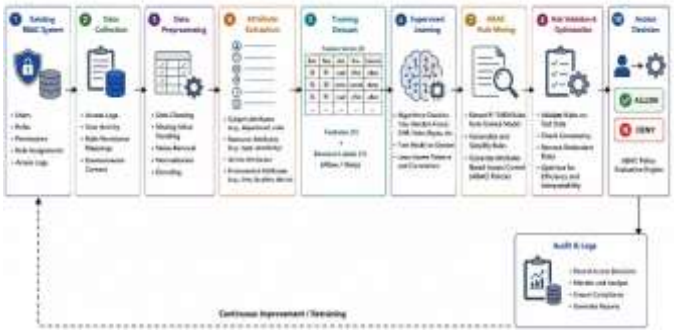


Fig. 2. Proposed Al-Driven RBAC-to-ABAC Policy Mining and Deployment Architecture.

This architecture enables the existing RBAC information to be reused for generating more flexible ABAC policies, while supervised learning reduces the dependency on completely manual policy creation.

V. METHODOLOGY

The methodology proposed here applies supervised learning to infer ABAC rules from an existing RBAC-enabled system. The process includes data collection, preprocessing, attributes extraction, model training, rules generation, and policies validation.

The existing RBAC system is first analyzed to collect users, roles, permissions, and access decisions. The data collected is then cleaned by removing duplicate records, handling missing values and converting categorical information to an appropriate format. Extract relevant user, resource and environmental attributes from processed data

The extracted attributes are combined with the respective Allow/Deny access decisions to form the training dataset. Then you can train a supervised learning algorithm (Decision Tree, Random Forest, SVM, Naive Bayes) to learn the patterns between attributes and access decisions.

The learned patterns are converted into ABAC rules which consist of attributes, conditions and actions. The generated rules are then validated to identify redundant or conflicting policies. The optimized ABAC policy set is deployed to an ABAC policy engine or Policy Decision Point (PDP).

The relevant attributes are then evaluated against the policies generated for the new access request. The final Allow or Deny decision is then made by the system. This way the methodology offers a systematic way of reusing the existing RBAC information and generating more flexible ABAC policies with less manual work.

Methodology Flow:



Fig. 3. Methodology Flow for AI-Driven RBAC-to-ABAC Policy Transformation.

VI. RESULTS AND DISCUSSION

The proposed system provides a systematic way of deriving ABAC rules from the existing RBAC enabled system by supervised learning. First, the RBAC data is analyzed to find proper relationship between users, roles, permissions and access decisions. Then, these relations are used to train the selected machine learning model and the attribute based authorization rules are generated.

The generated ABAC rules are more flexible than the original role-based permissions as access decisions can take into account user, resource and environmental attributes. The rule validation phase also helps in spotting the rules which are redundant or conflicting leading to a more consistent set of policies.

The proposed approach also reduces the manual effort for ABAC policy creation since the existing RBAC information is reused as the basis for rule generation. Instead of using only manually defined policies, the system can learn the way authorizations are made from existing data, by means of supervised learning. The results indicate that the proposed methodology can be a practical approach to migrate from a traditional RBAC structure to a more flexible ABAC based authorization model. Further evaluation of the approach effectiveness can be done using metrics like accuracy, precision, recall, F1-score, rule consistency, and policy generation time.

VII. CONCLUSION AND FUTURE SCOPE

The proposed work proposes a supervised learning based approach to mine ABAC rules from an existing RBAC-enabled system. The approach leverages existing user, role, permission and access-decision information to reduce manual policy creation and to provide more flexible and fine-grained access control. Future work can be done to improve the system by trying different machine-learning algorithms, larger real world datasets, deep learning and explainable AI, real-time policy update and automatic conflict resolution.

REFERENCES

[1] M. Alohaly, O. Balogun and D. Takabi, "Integrating Cyber Deception Into Attribute-Based Access Control (ABAC) for Insider Threat Detection," *IEEE Access*, vol. 10, pp. 108965–108978, 2022, doi: 10.1109/ACCESS.2022.3213645.

[2] H. Kim, D.-K. Kim and A. Alaerjan, "ABAC-Based Security Model for DDS," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 5, pp. 3113–3124, Sept.–Oct. 2022, doi: 10.1109/TDSC.2021.3085475.

[3] C. Cotrini, T. Weghorn and D. Basin, "Mining ABAC Rules from Sparse Logs," *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, London, UK, 2018, pp. 31–46, doi: 10.1109/EuroSP.2018.00011.

[4] 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*, vol. 19, no. 4, pp. 2304–2317, July–Aug. 2022, doi: 10.1109/TDSC.2021.3054331.

[5] 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. 11096–11103, 2020, doi: 10.1109/ACCESS.2020.2963932.

[6] B. Bera, A. K. Das, M. S. Obaidat, P. Vijayakumar, K.-F. Hsiao and Y. Park, "AI-Enabled Blockchain-Based Access Control for Malicious Attacks Detection and Mitigation in IoE," *IEEE Consumer Electronics Magazine*, vol. 10, no. 5, pp. 82–92, Sept. 2021, doi: 10.1109/MCE.2020.3040541.

[7] C. Wang *et al.*, "An Interpretable and Accurate Deep-Learning Diagnosis Framework Modeled With Fully and Semi-Supervised Reciprocal Learning," *IEEE Transactions on Medical Imaging*, vol. 43, no. 1, pp. 392–404, Jan. 2024, doi: 10.1109/TMI.2023.3306781.

[8] Y. Luo, Q. Shen and Z. Wu, "Casbin: A Graph-Based Approach for Efficient and Verifiable Cloud Authorization," *IEEE Transactions on Dependable and Secure Computing*, preprint, 2026, doi: 10.1109/TDSC.2026.3709803.

[9] Y. An, H. Xue and X. Zhao, "Conditional Self-Supervised Learning for Few-Shot Learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 36, no. 7, pp. 9794–9807, July 2026, doi: 10.1109/TCSVT.2026.3670944.

[10] X. Wang, D. Kihara, J. Luo and G.-J. Qi, "EnAET: A Self-Trained Framework for Semi-Supervised and Supervised Learning With Ensemble Transformations," *IEEE Transactions on Image Processing*, vol. 30, pp. 1639–1647, 2021, doi: 10.1109/TIP.2020.3044220.

[11] Y. Zhong, P. Chen and H. Zhang, "ESX: A Self-Generated Control Policy for Remote Access With SSH Based on eBPF," *IEEE Access*, vol. 13, pp. 6487–6506, 2025, doi: 10.1109/ACCESS.2024.3450496.

[12] D. Feng, W. Cui, Y. Jiang, W. Yu and D. Li, "From Static Roles to Context-Aware Decisions: Integrating LLMs and RAG Into Access Control Frameworks for Power Systems," *IEEE Access*, vol. 14, pp. 55864–55876, 2026, doi: 10.1109/ACCESS.2026.3667843.

[13] S. Das, S. Sural, J. Vaidya and V. Atluri, "HyPE: A Hybrid Approach toward Policy Engineering in Attribute-Based Access Control," *IEEE Letters of the Computer Society*, vol. 1, no. 2, pp. 25–29, July–Dec. 2018, doi: 10.1109/LOCS.2018.2889980.

[14] G. Baralla, L. Cocco, M. Di Francesco and R. Tonelli, "Integrating Blockchain, SSI, and RBAC for the Secure Management of Defense Heritage Buildings," *IEEE Access*, vol. 13, pp. 86290–86307, 2025, doi: 10.1109/ACCESS.2025.3570809.

[15] M. M. Rooney and M. K. Hinders, "Machine Learning for Medium Access Control Protocol Recognition in Communications Networks," *IEEE Access*, vol. 9, pp. 110762–110771, 2021, doi: 10.1109/ACCESS.2021.3102859.

[16] H. Liu, W. Qu, J. Jia and N. Z. Gong, "Pre-trained Encoders in Self-Supervised Learning Improve Secure and Privacy-preserving Supervised Learning," *2024 IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, USA, 2024, pp. 144–156, doi: 10.1109/SPW63631.2024.00019.

[17] Y. Zhang, L. Chen, Y. Zhu and X. Kong, "Privacy-Enhanced Role-Based Access Control for IoT Systems," *IEEE Internet of Things Journal*, vol. 12, no. 14, pp. 27269–27279, July 2025, doi: 10.1109/JIOT.2025.3562155.

[18] Z. Xu, G. Dong and R. Yang, "RBAC-Based One-to-Many Authentication and Key Negotiation Scheme in Smart Factory," *IEEE Access*, vol. 12, pp. 189202–189218, 2024, doi: 10.1109/ACCESS.2024.3435550.

[19] T. X. Pham, A. Niu, K. Zhang, T. J. T. Jin, J. W. Hong and C. D. Yoo, "Self-Supervised Visual Representation Learning via Residual Momentum," *IEEE Access*, vol. 11, pp. 116706–116720, 2023, doi: 10.1109/ACCESS.2023.3325842.

[20] Y.-F. Li, L.-Z. Guo and Z.-H. Zhou, "Towards Safe Weakly Supervised Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 334–346, Jan. 2021, doi: 10.1109/TPAMI.2019.2922396.

[21] Pankaj Kawadkar, Deepak Kumar Rathore, Devendra Singh Kushwaha, B. Rebecca, G Priyanka Jeeva Karunya, P. M. Yohan; Performance analysis of image compression techniques for MRI image using machine learning techniques. AIP Conf.

Proc. 4 November 2024; 3214 (1): 020038. https://doi.org/10.1063/5.0 2391 27

[22] Ganesh R. Kadam, Narendra U. Jadhav, Md Abdul Wassay, Yash R. Mohabe, Kaushal B. Loharkar, Devendra Singh Kushwaha" Facial Recognition-Based Classroom Attendance System with Real-Time Group Photo Processing Using Machine Learning Approach" Proceedings of International Conference on Computer Science and Communication Engineering (ICCSCE 2025),Atlantis Press, p:1166-1177, https://doi.org/10. 2991/978-94-64 63-8 58-5_97

[23] P. K. Mannepalli, D. Singh Kushwaha, S. Kalamdhar, V. Nagrale and V. Rajpoot, "Face Recognition Based on Cascade Classifier Using Deep Learning," 2023 1st International Conference on Innovations in High Speed Communication and Signal Processing (IHCSP), BHOPAL, India, 2023, pp. 63-68, doi: 10.1109/IHCSP56702.20 23.101 27172.