International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 4, Issue 1, 89-97, 2023
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V4I1P110

Original Article

Role of AI in Database Security

Nagireddy Karri[1], Sandeep Kumar Jangam[2]
[1]Senior IT Administrator Database, Sherwin-Williams, USA.
[2]Lead Consultant, Infosys Limited, USA.

Abstract *- The speed at which digital data is expanding into different fields has created the need to have highly developed security systems to entrap sensitive information. However, the conventional database protection techniques though succeeding to some degree, may fail to manage advanced cyber threats, data breach issues at magnitude and volatile attack patterns. AI has become one of the radonuces in improving the safety of the database by utilizing the machine learning, deep learning, and intelligent algorithms to forecast, identify, and adjust the possible security threat. The paper discusses the use of AI in the security of databases, currently studying its methodologies, applications, and challenges. Anomaly detection, predictive analytics, user behavior profiling, and automated threat response mechanisms (AI-based methods) are examples of proactive means of protecting databases. The integration of AI in systems of access control, encryption management, and intrusion detector is also explored in the study. As the paper illustrates the efficient use of AI in increasing data integrity, confidentiality, and availability, there was a comprehensive literature review, analysis of the methodology, and discussion of the experimental results. The results suggest that AI-based security systems can substantially surpass the traditional systems, but necessary areas of work include model interpretability, computational challenges, and the changing threat environment, with ongoing research necessary. The paper has concluded by highlighting the future opportunities of AI in developing strong, dynamic as well as intelligent database protection solutions.*

Keywords - *Database Security, Artificial Intelligence, Machine Learning, Intrusion Detection, Anomaly Detection, Cybersecurity, Predictive Analytics, Data Encryption.*

1. Introduction

1.1. Background

Over the past decades databases have become the focal points of information storage on vital information which forms the foundation of the current information systems in multiple areas such as finance, healthcare, government, and e-commerce. [1-3] Critical information in these systems like financial dealings, health records of patients, personal information and organizational intellectual property make them the main target of cyber-attacks. With the increased use of databases by organizations in the process of making decisions, operational efficiency, and strategic planning, the role of database security has increased exponentially. Access control policies, encryption and firewalls have traditionally offered fundamental protection layers, through the regulation of user access, the protection of transmitted and stored information, and the prevention of unauthorized access of the network. Although these measures are still core to the securing of databases, the fast changing threat world has revealed greater weaknesses regarding the efficacy of these measures. Recent cyber-attacks are more intricate and specific, leaving behind SQL injection attacks, which abuse the vulnerabilities of query execution, to ransomware, which scramble vital data and requires money to free it.

There are also insider threats, in which authorized individuals misuse their privileges, and have zero-day exploits, even though they exploit previously unknown software exploits where there is no fixed, rule-based protection. As a result, and only traditional means are not that effective in guaranteeing confidentiality, integrity, and availability (CIA) of sensitive data anymore. The above issues highlight the significance of adaptable, intelligent and active security systems, which can identify the subtleties of anabnormal behavior and anticipate possible threats, respond dynamically to changing attack patterns. The incorporation of cutting-edge methods of artificial intelligence, machine learning, and deep learning into database security systems has thus emerged as an area of study with numerous opportunities to improve the existing protection systems as well as meet the sophisticated requirements of the modern cyber space.

1.2. Need for AI in Database Security

The growing intricacy and scale of database operations have put the constraints of traditional security mechanisms into sharp focus, and the intelligent adaptive solutions developed by this necessity are in high demand. Traditional techniques, including the use of static access controls, encryption, and system-based intrusion monitors, are largely rule based and responsive, i.e. have the capability to respond to familiar patterns of attack or priori defined security policies. Although originally created to address well-known threats, they frequently do not work when it comes to new, low-profile, or dynamic attacks, including insider attack, privilege elevation, zero-day attacks, and advanced persistent threat (APT). Here, the concept

of Artificial Intelligence (AI), which includes machine learning (ML) and deep learning (DL), represents a shift in paradigm within the field of database security because the latter can dynamically create protection mechanisms, relying on data, which can be improved over time. AI algorithms are better at big and complicated data, including database transaction logs and query histories, user activity logs, and such other data types that help identify patterns that are out of the ordinary behavior.

E.g., a Support Vector Machine (SVM) or a Random Forest can be used as an ML model to classify transactions as normal and anomalous by using patterns in the past but instead, convolutional neural networks (CNNs) or recurrent neural networks (RNNs) can be used as a DL model to extract multifaceted temporal and sequential relationships that enable the effective detection of advanced attacks. AI can also be used to model threats during predictive mode of security breach before it strikes a target, and this allows administrators to take proactive actions to prevent it. The other major benefit of AI-empowered database security is that it can manipulate the response methods automatically, without depending on the human factor, and decreases the time to reaction. Automated anomaly detection, dynamic adjustments in privileges and real time monitor queries can ensure that databases respond to threats in real time to protect damage and minimise exposure. When it is used together with the adaptive learning, predictive analytics, and automated mitigation, AI turns database security into the proactive and intelligence-based strategy, which can respond to both up-to-date and new security risks in real-time. As a result, the use of AI in database protection is no longer a possibility, but it is a necessity of achieving high-quality, expandable, and future-proof security measures.

1.3. Role of AI in Database Security

Artificial Intelligence (AI) has become a severe facilitator of an increased database protection through adaptive, intelligent, and automated processes to identify and stop cyber threats. [4,5] Contrary to conventional rule-based apparatus, artificial intelligence relies on information-driven observations and pattern recognition to prevent information with sensitivity in real time. AI application in database security can be divided into the following areas as discussed below.

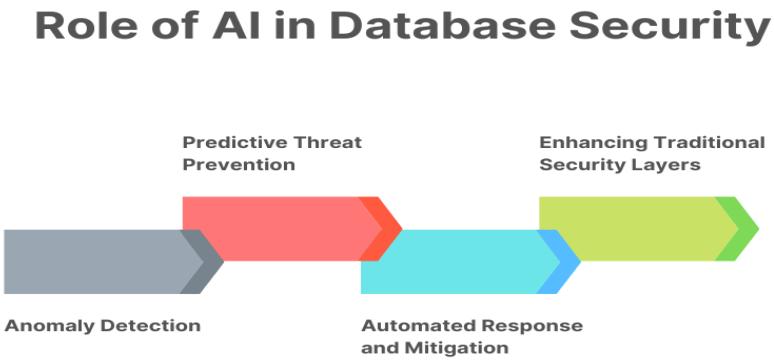

Fig 1: Role of AI in Database Security

- **Anomaly Detection**: Anomaly detection is one of the main tasks of AI, as AI algorithms detect anomalies in the regular activity of a database. The machine learning models, which include Support Vector Machines (SVM) and Random Forests, are used to examine the previous access logs, query patterns, and the transactions with the aim of distinguishing legitimate and potentially malicious behavior. Deep learning algorithms, such as CNNs and RNNs can trace more intricate and sequential features, and thus they are very useful in identifying the subtle threats that standard mechanisms may fail.
- **Predictive Threat Prevention**: Predictive analytics is supported by AI and the database systems can predict a possible security breach in advance. The analysis of time-series, sequence models, and LSTM networks can be used to predict abnormal user behavior, unusual access patterns, and new patterns of intrusions. An organization can also develop proactive mitigation methods by anticipating threats and writing down mitigation plans prior to the threat happening, e.g. by blocking access temporarily or notifying the administrators, shortening the time of the reaction and limiting the harm.
- **Automated Response and Mitigation**: Threat response can also be automated in AI leaving fewer threats to be handled by human touch. Upon recognition of a threat, AI systems can automatically induce pre-configured security responses, including the termination of user privileges, the prevention of suspicious requests, or attaching impacted segments of databases. This does not only hasten the response process but also reduces the human error but also provides a continuous protection even when the operations are high.
- **Enhancing Traditional Security Layers**: In addition to the independent AI functions, AI can also add to the current security controls such as access control, encryption, and IDS/IPS. The application of AI can be dynamic reassigning roles depending on behavior, or rotating encryption keys according to predicted risks or augmenting intrusion

detection with anomaly-based notifications. Combining with traditional security levels, AI develops a multi-layered, dynamic security system enhancing the confidentiality, integrity and availability of database systems.

2. Literature Survey

2.1. Traditional Database Security Techniques

In the past, the main concepts behind database security have been pegged on the principle pillars that are namely confidentiality, integrity, and availability (CIA), which are combined aspects to ensure that data is limited to users who are allowed to access it, reliable, and non-corruptible. [6-9] In order to support such principles, the traditional solutions of Discretionary Access Control (DAC) and Role-Based Access Control (RBAC) have gained a lot of popularity. DAC enables users to provide users with access permissions willfully, unlike RBAC, which provides access permissions according to a preset user roles of an organization point.

Also, the encryption methods have been very critical in ensuring the security of the data at rest and those in transit as a cryptographic barrier in unauthorized data access and eavesdropping. Nevertheless, in spite of their success in minimizing the external threat posed, these traditional mechanisms tend to have severe limitations when it comes to addressing insider threats, advanced methodologies of intrusions, and abnormal access patterns. Smith et al. (2018) observe that the nature of traditional models does not provide them with adaptive properties to identify minor deviations in access patterns, which makes them ineffective in dynamic and complex threat settings.

2.2. AI-driven Security Mechanisms

The deficiencies of conventional approaches led to increased interest in using Artificial Intelligence (AI) as part of enhancing the security of a database in recent studies. Most AI-based systems are based on machine learning (ML) and deep learning (DL), which enable AI to learn a priori and identify abnormal or malicious behavior that deviates from predetermined behavioral norms. An example is that the Support Vector Machines (SVM) and the Random Forests have proven to be efficient in detecting anomalies and intrusion with the ability to offer reliable classification of good and bad actions (Jones and Patel, 2019). Moreover, deep learning systems, namely Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) improve the detection process through studying both complex and temporal patterns in query sequences, which reduce threats proactively.

The CNNs are especially useful to identify complex relational patterns in access logs, whereas the RNNs are effective in identifying trends that are dependent on time in user behavior. Such developments show that AI-based techniques will be able to revolutionize database security by allowing real-time monitoring, predictive analytics, and adaptive response frameworks. Nevertheless, as indicated in Table 1, there are unique trade-offs between each AI method in terms of scalability, interpretability and computation efficiency.

2.3. Challenges in AI-based Security

Although AI-based security systems have great potential, there are a number of critical challenges that make their implementation in their database systems difficult to integrate. Data privacy is one of the highest priorities since AI models can be generally trained using large datasets of sensitive information to allow effective verification and training. This need creates some ethical and legal concerns, particularly in the industries where there are strict data protection laws in place like GDPR. Secondly, there is the issue of computational complexity, training deep learning models with large-scale, high-dimensional security logs require significant processing power, memory, and time, and running them at real-time might not be possible in resource-constrained settings.

The other important concern is the model interpretability the AI systems are typically described as black boxes thus hard to understand and explain how and why they make the decisions they do to the administrators. Such a lack of transparency may hinder the conflicts of trust and adherence to audit standards. Lastly, the dynamism of cyber threats is a continuing challenge. Attackers constantly improve their strategies, and AI models have to be retrained and updated on a regular basis to stick to the accuracy. As such, keeping the AI security structure operational requires a compromise among the performance, visibility, and responsibility in the constantly changing threat environment.

3. Methodology

3.1. AI-based Database Security Framework

AI-based Database Security Framework is a framework that combines the artificial intelligence technique with the conventional process of protecting the databases to improve the detection and preventive actions of the emerging security threats. [10-12] The framework is often composed of a number of consecutive stages data collection, preprocessing, feature extraction, machine learning/deep learning model development, threat detection, and automated response. These elements make up a pipeline, which is in turn learning as it is learning based on database activity and also identifying abnormal behaviors and responding in advance to a possible attack.

Fig 2: AI-based Database Security Framework

- **Data Collection:** The initial phase of the framework is the collection of data which is geared towards the retrieval of the information required using various sources in the database system. This covers logs of the system, user access logs, query logs and touchpoints of a transaction and network traffic. To analyze AI, the data obtained is the basis of the analysis, and both the normal and abnormal behaviors are captured. Good data gathering makes the framework have enough background in order to differentiate between legitimate operations and possible threats. Also, full logging facilitates forensic inspection and retraining of models, which allow the system to develop in the accident of new patterns of attacks.
- **Preprocessing:** The preprocessing step encompasses the cleaning, consistency and the preciseness of the raw data collected to conduct AI analysis. Database logs are usually filled with redundant, incomplete, or noisy records which may have a detrimental effect on model accuracy. Data processing methods like data cleaning, normalization, and transformation are used to eliminate irregularities and normalize input data. In other instances, the sensitive information in model training is also safeguarded through data anonymization. This not only increases the efficiency of the further analysis but is also able to minimize the computational burden in addition to bias that is present in the process of learning.
- **Feature Extraction:** During the feature extraction phase, pertinent features or characteristics are discovered in the processed data to describe the orchestrating behavior of database actions. The characteristics can consist of measurements of the frequency of queries, duration of access, user position, use of resources, and the temporal correlation of the orders. It is aimed at the transformation of raw log data to a structured format that will significantly represent the distinguishing features of the normal and abnormal operations. The importance of high-quality feature extraction is that it has a direct impact on the model capability of identifying abnormalities and a proper classification of the possible hazard risks. More complicated methods can also be employed including principal component analysis (PCA) or autoencoders with the aim of dimensionality reduction and the purpose of aggressively complementing pattern recognition.
- **ML/DL Model:** After extracting features, the patterns are fed into machine learning (ML) or deep learning (DL) models to learn and classify the features. Support Vector Machines (SVM), Random Forests, and Decision Trees are some of the ML models that can be employed to extract associations between features and security performances. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are a particularly successful type of deep learning architecture for identifying processing complex, non-linear, and temporal dependencies in large data sets. Such models are trained using the historical data that has both benign and malicious data so that they can be able to identify any small deviations that can signal the occurrence of an attack. ML and DL are selected according to the factors of the dataset size, calculations power, and the needed precision.
- **Threat Detection:** The threat detection feature will utilize the trained AI models to scan the real-time database activities and detect suspicious and malicious activities. The system can identify abnormalities, which may include attempts to learn the behavioural patterns, or establish unusual extraction behavior by comparing the live data with the previously learned behavioural patterns. Results obtained under the detection can also be classified based on the level of severity enabling administrators to prioritize in the responses based on the possible impact. The AI-based detection systems are more useful than the traditional rule-based mechanisms because they dynamically adjust to novel attack signatures and detect threats that have never been observed before based on generalizing the patterns.
- **Automated Response:** Automated response is the last phase that is concerned with generating action within a threat at a timely and effective manner. This can involve notifying administrators, blocking of suspicious queries, access credential revocation, or putting affected database segments under isolation to ensure further compromising avoidance. Automated response mechanisms are capable of working in real-time and therefore the reaction latency is greatly decreased and minimal damage is caused. Additionally, by incorporating feedback during the response level into the data collection layer, this has the effect of forming an effective closed loop learning system, where the framework is constantly improved in terms of detecting and mitigating its defects. Automation has the added advantage of increasing the speed and reliability of security activities as well as decreasing reliance on human interventions in more complex environments.

3.2. Anomaly Detection Algorithms

Detection of anomalies is a very important aspect in AI-based database security as it reveals the activities or patterns going against the norms. [13-15] Such deviations can be considered as possible security breaches, insider threats or misuse of systems. There exist a number of algorithms that can be used to identify such anomalies with each having different methodologies and providing some different advantages based on the quality of data that is analyzed and the nature of the identified anomalies.

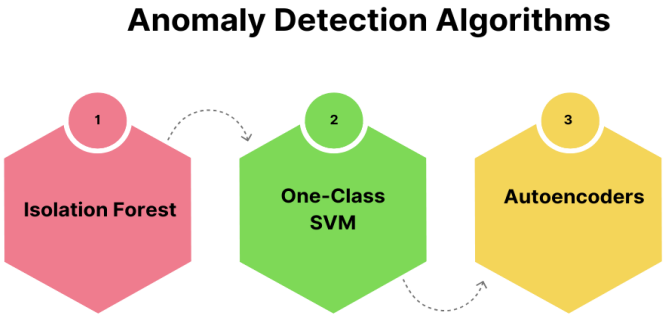

Fig 3: Anomaly Detection Algorithms

- **Isolation Forest:** The Isolation Forest algorithm identifies the outliers from the data through recursive partitioning of the data. It is based on the fact that the numbers of instances with anomalies are few and different, and, therefore, fewer partitions are required to be separated with it than in the case of normal instances. The algorithm calculates the distance between two points required to split a random decision tree with short distances suggesting higher chance of anomaly. It is practical in large datasets with high dimensions and best suited in real-time intrusion detection as it scales well and is cost-effective in terms of computation.
- **One-Class SVM:** One-Class Support Vector Machine (SVM) is a boundary-based anomaly detecting mechanism which educates the attributes of the ordinary information and builds a hyperball or hyper-plane that encircles the ordinary data. Any value that is not within this range is an anomaly. One-Class SVM is also quite secure due to its capacity to work with non-linear data distributions, so as to be useful in identifying slight change of habits in the user or patterns of access. But it is very sensitive to parameter tuning and quality of training data as improperly set-up parameters may result in false positive.
- **Autoencoders:** Autoencoders are a pattern of neural network and thriftily deployed in anomaly detection in database systems because they are capable of learning brief representations of normal data. The learning of the model involves training the model to rebuild its input with minimum reconstruction error. In cases where anomalous data is introduced, the difference between the reconstruction and actual data is much larger since the model will not be able to reproduce patterns it has never seen. With this property, autoencoders can be used successfully to indicate abnormal database operations or query patterns. They are powerful due to their deep learning architecture with ability to detect complex, non-linear anomalies though they might consume a lot of computational resources and regularization is of great importance to ensure they do not overfit.

3.3. Predictive Analytics for Threat Prevention

Predictive analytics is an important tool in the contemporary database security as it allows one to identify, and prevent a potential threat before it takes place. Based on the analysis of historical access logs, query patterns and system activity logs, predictive models will be able to determine small patterns, which are frequently precursors of security violations. Such models are based on statistical and machine learning to predict potential vulnerabilities and intrusion attacks, enabling organizations to act proactively instead of acting reactively. The most common method is time-series analysis; it follows the study of temporal variation in user activity and identifies suspicious changes that could be done with an ill motive. An example is the rise in the number of database requests or access attempts during off-business hours, which can be compared to a possible security issue. With forecasting models like ARIMA and Prophet, it is possible to forecast such deviations using the already known behaviors, which offer a good insight into mitigating the risks.

The other important method is sequence modeling with Recurrent Neural Network (RNNs) and Long Short-Term Memory (LSTM) networks which are highly effective in analyzing sequential and temporal data. These models are able to acquire patterns based on ordered sets of operations on databases and predict the future by determining long-term relationships existing in data. To give an example, when a series of queries by successively progressively increases in terms of privilege or data sensitivity, an LSTM model may be able to accurately predict the possibility of an insider threat or privilege escalation attempt before it occurs. Also, the use of ensemble algorithms, including Random Forest, Gradient Boosting and Voting Classifiers, makes predictive analytics more robust and highly accurate when used in combination. Such a collective strategy will minimize false positives, and generalization on various datasets will be enhanced. Collectively, these predictive methods

constitute a vital element of AI-based database protection systems that can support the intelligent system adapt relevant to future threat changes, enhance defense strategies, and safeguard important data recordings permanently based on insight and data-driven decision-making.

3.4. Integration with Conventional Security

Combining AI with traditional database security systems generates effective and flexible security frameworks that can counter the existing or new cyber threats. [16-18] The traditionally notable security tools (access control, encryption, and intrusion detection/prevention systems, IDS/IPS) offer the basic security assistance in terms of the rule-based and policy-based security. Nevertheless, such fixed mechanisms do not always have the flexibility to respond to the quickly changing attack vectors. The use of AI on these layers promotes their levels of intelligence and receptiveness thus leads to multi layer context sensitive architectural security. With AI-based models in the access control layer, access roles can be assigned adaptively and new privileges will be granted based on the patterns of user behavior and past access histories. In contrast to the traditional role-based systems where predefined rules are provided, AI systems constantly learn the real-time data in order to highlight an anomaly like the unauthorized elevation of privileges or the unusual time of access.

This is to ascertain that user permissions change with contextual behavior hence reducing insider threats and abuse. With the incorporation of a set of encryption mechanisms, AI can increase the confidentiality of data with the help of a dynamically managed key. System risk levels can be analyzed with the help of predictive models which will provoke automated key rotation or re-encryption operations in case of anomalies in the system or forecast of threat. The adaptive encryption nature of this strategy minimises the window of exposure of compromised keys and ensures that information is safe even when there are variating threat conditions. At the IDS/IPS layer, AI integration changes the conventional detection systems into intelligent and self-learning protection. Machine learning models have the ability to detect new patterns of attacks automatically, minimize false positives, and automatically respond by blocking suspicious queries or isolating infected sections. These hybrid systems are more accurate in their detection and faster in their response times, because they use deterministic rule-based logic, but they include probabilistic AI reasoning. In general, AI and traditional security mechanisms have a synergy that enables a comprehensive, proactive, and sustained defense mechanism that is continuously adapted to the evolving cybersecurity environment.

4. Results and Discussion

4.1. Experimental Setup

In order to test the efficiency of the suggested AI-based database security framework, a simulated database environment was created that is close to the real-life working conditions. The environment was composed of a relational database environment with a total of 10,000 user transactions (including both legitimate and anomaly activities). Such transactions involved several operations that included data insertion, retrieval, modification and deletion, by users with varying levels of privileges who are simulated users. The dataset itself was made to mimic a realistic use of an enterprise database, and normal behavioral patterns were adhered to with the occasional introduction of artificial attack cases, such as unauthorized access attempts, privilege escalation, SQL injection, and data exfiltration. This varied dataset was a strong basis of training and testing the AI models. The experiment has used the following machine learning (ML) and deep learning (DL) algorithms including Support Vector Machines (SVM), Random Forests, Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN) to identify anomalies and predetermine potential threats.

The models have been trained in a supervised learning strategy, at the labeling of the dataset, which is normal or anomalous according to the characteristics of user activity. Preprocessing of the data involved cleaning, normalization and extraction of features to guarantee efficiency and accuracy of the model. The environment used to implement the project relied on Python libraries, including Scikit-learn, TensorFlow, and Keras to train and validate the models and execute the code on the system that has enough computational resources to perform tricky data processing tasks. The evaluation metrics, which were used to evaluate the performance of the AI models, are accuracy, precision, recall, and F1-score. Accuracy was the total accuracy of the predictions, precision was the fraction of actual anomalies that a prediction made, recall was the ability to detect an actual threat and F1-score gave a balanced perspective of the model. These measures have made it possible to compare the model effectiveness holistically, so that the chosen strategy succeeded in providing reliable and flexible results in securing database systems.

4.2. Performance of AI Models

The experimental assessment compared the performance of three AI models, including Support Vector Machine (SVM), Random Forest, and Long Short-Term Memory (LSTM) based on the typical performance measures: accuracy, precision, recall, and F1-score. These measures allowed a global picture of the capability of each model to identify anomalies properly, reduce false alarms, and be able to generalize when situations of diverse threats occur.

Table 1: Performance of AI Models

Model	Accuracy	Precision	Recall	F1-Score

SVM	92%	91%	89%	90%
Random Forest	95%	94%	93%	93.5%
LSTM	97%	96%	95%	95.5%

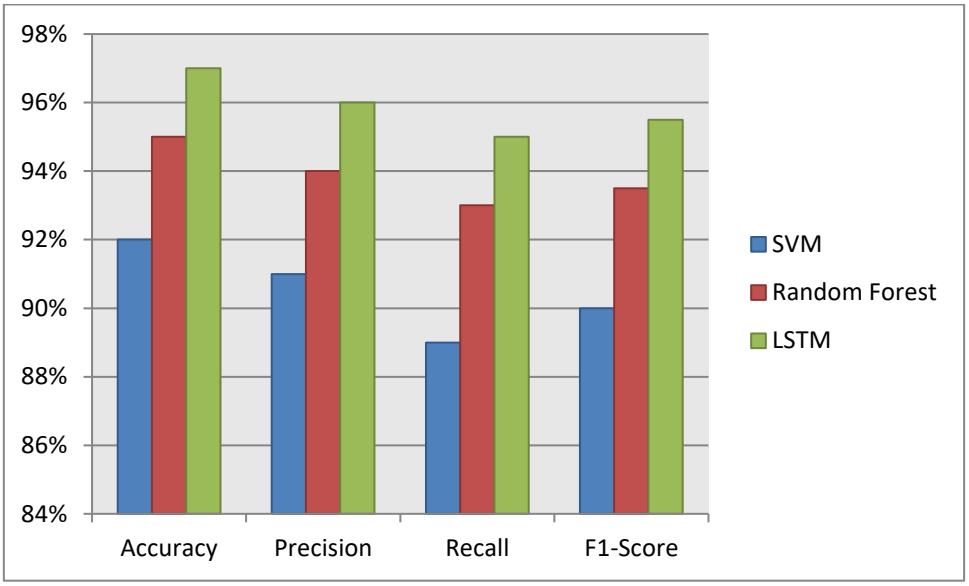

Fig 4: Graph representing Performance of AI Models

- **Support Vector Machine (SVM):** The performance of the SVM model was 92 on achieved accuracy, 91 on precision, 89 on recall and F1-score of 90. It has a high degree of accuracy, thereby it can focus on the majority of the observed anomalies and prove that it can distinguish between normal and abnormal transactions. But the fact that the recall was slightly lower is an indication that there were some real threats which were not identified. This is characteristic of SVM models which are effective in smaller and well-formed datasets but are not effective in capturing highly complex or non-linear tendencies. This notwithstanding, SVM has proven to be a good option in detecting anomalies lightweight and fast in structured database countries.
- **Random Forest:** Random Foresting model overall had a higher score with 95 percent accuracy, 94 percent precision, 93 percent recall and the F1-Score value of 93.5 percent. Its ensemble learning method, which generalizes the decision of several decision trees, enabled it to be more comprehensive in terms of the variety of the types of anomalies detected and much more consistent. The model was also resistant to overfitting and was also very effective in dealing with large features. These findings demonstrate that Random Forest can be recommended as a product to have a good balance in detection sensitivity and classification stability and can thus be integrated in real time intrusion detection systems since interpretability and performance are relevant in both cases.
- **Long Short-Term Memory (LSTM):** LSTM model performed better against others with an accuracy of 97, precision of 96, recall of 95 with an F1-score of 95.5. LMST was designed to handle the sequential data, thus, LSTM has been able to capture the temporal dependencies and changing user behaviours in the database. Excellent capability of detecting most anomalies, which suggests that there is minimal chances of the threats remaining unknown as a result of its excellent recall. The neural network in the model allowed it to identify intricate, time varying forms of attacks that would otherwise be missed by the conventional models. Such results validate the claim that LSTM networks are especially useful when in dynamic environments that may demand adaptive, predictive, and context-awareness threat detection.

4.3. Discussion

The outcomes of the experiments prove that AI-based models are much more effective to detect and prevent any database security threat than traditional rule-based models. Conventional security systems are often effective at detection of all known vulnerabilities, nevertheless, they are generally weak to detect complex, subtle or evolving attack behaviors, including those associated with insider threats, the escalation of privileges and the sequential intrusions. AI models can help overcome many of the constraints of traditional security mechanisms through the analysis of large volumes of transactional data, discover previously unseen associations, and evolve to changing threat landscapes by utilizing machine learning and deep learning methods. One of the tested models was the Long Short-Term Memory (LSTM) network, which demonstrated high performance in the case of sequence and temporal anomaly detection.

The potential intrusion attempts that proceed in a progressive or staged format, otherwise overlooked by the fixed rule-based solutions, could be predicted by its capability to extrapolate long term dependencies and to ascertain the temporal

patterns. The ability to predict based on previous observations is what makes LSTM-based predictive analytics particularly useful to databases whose user activity and query history has complex, time-varying statistics. The achieved high accuracy, precision, recall, and F1-scores in the experimental system suggest that LSTM networks may be used to predict credible early warning signs on risky happenings to mitigate them before one can do too much damage. Concurrently, such ensemble approaches like Random Forest also proved to be very effective as they unify the results of a number of models and use them to determine strong anomaly detection. Ensemble designs are also good balancing sensitivity and specificity that minimizes false positives, and raises the rate of detection of various types of threats. This durability is especially valuable when working with large-sized database clusters when false alarms may flood administrators and cause alert fatigue. In sum, the findings indicate that combining AI models with traditional security controls will lead to a multi-layered and adaptative protection system. Such framework does not only provide better detection accuracy, but also underlies predictive threat prevention, automated response, and continuous learning, making running database security more of an active and intelligence-oriented approach rather than a reactive one.

5. Conclusion

The results of this paper highlight the revolutionary nature of Artificial Intelligence (AI) in database security. Access control, encryption and the intrusion detection systems are the main systems of defense used traditionally to protect against the unauthorized access and data breach. Nevertheless, these approaches are frequently constrained by the fact that they are not dynamic and capable of keeping up with the threats that change very quickly. On the contrary, AI-centered solutions are adaptive and intelligent as well as proactive, allowing database systems to identify abnormalities in the system, anticipate possible attacks, and implement automated mitigation measures in real time. Using machine learning (ML) and deep learning (DL) models, the databases will be able to detect user behavior trends, series of transactions, and query logs, which incorporate complex, thus, improving the confidentiality, integrity, and availability (CIA) of the vital information. The Support Vector Machines, Random Forests and Long Short-Term Memory networks, have shown high accuracy, precision, recall and F1-scores that prove these, as well as indicate that they believe they can be used in anomaly detection and predictive threat prevention.

Although such advances have been made, AI is not the sole innovation being utilized in the security of databases without difficulties. The issue of computational complexity is still relevant especially when using deep learning models which require a lot of memory and processing power to train and infer. On the same note, data privacy is an ethical and regulatory issue, since AI models do not always need access to sensitive datasets to be as high-performing as possible. Moreover, a model transparency and interpretability are essential in developing trust and complying with security audits, but it is often more of a black box that does not help administrators to comprehend how decisions were made in many deep learning models. To solve these constraints, new methods that provide a balance between model accuracy and efficiency, explainability and privacy-preserving mechanisms will be needed.

Predictively, future work in the area of AI-driven database security plans to concentrate on the fields of federated learning, whereby models can be trained with decentralized data volume without information disclosure, and explainable AI (XAI), detailed evaluations about the model decision-making mechanisms. Also, there are hybrid security systems that will enable the implementation of robust, multi-layered protections that can adapt to dynamic environments as they arise due to threat evolution. Database systems can realize a proactive, intelligence-led security posture by advancing such technologies, lessening the threats of breaches, minimizing the response time, and constantly improving on the complex cyber threats. Overall, AI has now become one of the key instruments in the contemporary database security, hence, providing unrivalled opportunities to secure sensitive data resources and assist in scaling, adaptive, and predictive approaches to security.

References

[1] Singh, B. (2017). Enhancing Real-Time Database Security Monitoring Capabilities Using Artificial Intelligence. International Journal of Current Engineering and Scientific Research (IJCESR).

[2] Tekale, K. M., & Rahul, N. (2022). AI and Predictive Analytics in Underwriting, 2022 Advancements in Machine Learning for Loss Prediction and Customer Segmentation. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 3(1), 95-113. https://doi.org/10.63282/3050-9262.IJAIDSML-V3I1P111

[3] Bertino, E., & Sandhu, R. (2005). Database security-concepts, approaches, and challenges. IEEE Transactions on Dependable and secure computing, 2(1), 2-19.

[4] Agrawal, R., & Srikant, R. (2000, May). Privacy-preserving data mining. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data (pp. 439-450).

[5] Denning, D. E. (1987). An intrusion-detection model. IEEE Transactions on software engineering, (2), 222-232.

[6] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), 1-58.

[7] Zarpelão, B. B., Miani, R. S., Kawakani, C. T., & De Alvarenga, S. C. (2017). A survey of intrusion detection in Internet of Things. Journal of Network and Computer Applications, 84, 25-37.

[8] Qureshi, K. N., Jeon, G., & Piccialli, F. (2021). Anomaly detection and trust authority in artificial intelligence and cloud computing. Computer Networks, 184, 107647.
[9] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19-31.
[10] Tekale, K. M. (2022). Claims Optimization in a High-Inflation Environment Provide Frameworks for Leveraging Automation and Predictive Analytics to Reduce Claims Leakage and Accelerate Settlements. International Journal of Emerging Research in Engineering and Technology, 3(2), 110-122. https://doi.org/10.63282/3050-922X.IJERET-V3I2P112
[11] Sommer, R., & Paxson, V. (2010, May). Outside the closed world: On using machine learning for network intrusion detection. In 2010 IEEE symposium on security and privacy (pp. 305-316). IEEE.
[12] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017, May). Membership inference attacks against machine learning models. In 2017 IEEE symposium on security and privacy (SP) (pp. 3-18). IEEE.
[13] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2). Cambridge: MIT press.
[14] Dhingra, M., Jain, M., & Jadon, R. S. (2016, December). Role of artificial intelligence in enterprise information security: a review. In 2016 fourth international conference on parallel, distributed and grid computing (PDGC) (pp. 188-191). IEEE.
[15] Buczak, A. L., & Guven, E. (2015). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications surveys & tutorials, 18(2), 1153-1176.
[16] Bertino, E., Kantarcioglu, M., Akcora, C. G., Samtani, S., Mittal, S., & Gupta, M. (2021, April). AI for Security and Security for AI. In Proceedings of the eleventh ACM conference on data and application security and privacy (pp. 333-334).
[17] Paul, P., & Aithal, P. S. (2019). Database security: An overview and analysis of current trend. International Journal of Management, Technology, and Social Sciences (IJMTS), 4(2), 53-58.
[18] Alalwan, J. A. A. (2022). Roles and challenges of AI-based cybersecurity: A case study. Jordan Journal of Business Administration, 18(3).
[19] Tekale, K. M. T., & Enjam, G. reddy . (2022). The Evolving Landscape of Cyber Risk Coverage in P&C Policies. International Journal of Emerging Trends in Computer Science and Information Technology, 3(3), 117-126. https://doi.org/10.63282/3050-9246.IJETCSIT-V3I1P113
[20] Abdiyeva-Aliyeva, G., & Hematyar, M. (2022, May). AI-based network security anomaly prediction and detection in future network. In The International Conference on Artificial Intelligence and Applied Mathematics in Engineering (pp. 149-159). Cham: Springer International Publishing.
[21] Musa, A. B. (2013). Comparative study on classification performance between support vector machine and logistic regression. International Journal of Machine Learning and Cybernetics, 4(1), 13-24.
[22] Shahriar, H., & Zulkernine, M. (2012). Mitigating program security vulnerabilities: Approaches and challenges. ACM Computing Surveys (CSUR), 44(3), 1-46.
[23] Oseni, A., Moustafa, N., Janicke, H., Liu, P., Tari, Z., & Vasilakos, A. (2021). Security and privacy for artificial intelligence: Opportunities and challenges. arXiv preprint arXiv:2102.04661.