Artificial Intelligence-Based Private Protection

Anusha Katha
Dept. of Electrical and
Computer Engineering
Queen's University
Kingston, Canada
Email: 22ak3@queensu.ca

Abstract—**Artificial Intelligence (AI) has emerged as a promising solution to address the critical challenge of privacy protection in the digital age. With its ability to analyze and process vast amounts of data, AI can be harnessed to develop robust and efficient privacy protection mechanisms. This research paper explores the potential of AI-based privacy protection methods, examining their current applications, limitations, and challenges, focusing on techniques such as deep learning, federated learning, privacy-preserving machine learning, and user authentication.**

1. Introduction

Privacy has become a top concern for both individuals and corporations in today's linked society. Effective privacy protection is now more important than ever as the digital age continues to speed up information creation and sharing. Conventional techniques of protecting personal data are finding it difficult to keep up with the rapidly changing digital environment, prompting the creation of novel, cutting-edge solutions. With its unmatched capacity to collect and analyze enormous volumes of data, artificial intelligence (AI) presents a promising path for strengthening privacy protection in the digital sphere.

Data has proliferated in the digital age as a result of individuals and organizations producing and disseminating data at an unparalleled rate. The exponential growth of data is a double-edged sword that simultaneously creates new prospects for innovation and economic development while arousing worries about security and privacy. In light of high-profile data breaches and the widespread use of surveillance technology, it is more important than ever to ensure that individual privacy is protected. Advanced privacy protection techniques are therefore urgently needed to adequately protect users' personal data and digital footprints.

In order to imitate, extend, and enhance human intelligence, artificial intelligence (AI), a rapidly expanding field of computer science, studies and develops theories, methodologies, techniques, and application systems. AI has advanced significantly in recent years, in part because of the growth of deep learning (DL) and the development of ultra-performance computing technologies. Particularly, DL technology has made it possible for people to gain access to more data, get better outcomes, and expand their potential[5]. It has fundamentally altered human life and transformed conventional AI technologies. Although AI has many different applications, including robotics, speech recognition, and facial identification, its breadth of use spans much beyond the three domains of image, voice, and behaviour.

Artificial intelligence (AI) techniques have been adopted in a variety of applications as a result of recent technological advancements and increases in computer power[1]. In the healthcare, gaming, and financial industries, machine learning models are being used to spur innovation, and self-driving car makers rely on deep learning models to build self-driving car pipelines. Today's AI systems use machine learning (ML) models and, more recently, deep learning (DL) algorithms to automate jobs and processes, enabling the introduction of new capabilities and functions that were not before conceivable. For instance, in the computer game StarCraft II in 2019, DeepMind's AlphaStar, an AI system built on deep reinforcement learning, attained the Grandmaster level by defeating multiple expert human players[2].

The increasing integration of artificial intelligence (AI) into various aspects of society has brought about numerous benefits and advancements. Machine learning, a subset of AI, has demonstrated its potential in solving complex problems across various domains. However, as the use of machine learning models expands, so too does the risk of adversarial attacks, which aim to compromise the confidentiality, integrity, or availability of these models by exploiting their vulnerabilities [3, 4].

In order to assure the safe and responsible use of AI in this context, this research paper will examine the existing uses, constraints, and challenges of AI-based privacy protection solutions. The paper will explore the many AI privacy-protection methods, including user authentication, federated learning, deep learning, and privacy-preserving machine learning. We will explore the future prospects for AI-based privacy protection, examining the ongoing research and development in the field. This paper aims to inform researchers about the opportunities and risks associated with the implementation of these technologies and to contribute to the ongoing discussion on how to best balance innovation and privacy in the digital age by providing a thorough overview of the current state of AI-based privacy protection.

2. Background

2.1. DeepLearning

Deep artificial neural networks are the main focus of the specialized machine learning field known as ”deep learning.” Conventional machine learning techniques, also known as shallow learning techniques, call for a feature engineer to extract pertinent traits from the input data [6]. The design of these shallow learning techniques is often quite straightforward and frequently consists of a single layer that transforms input data into a problem-specific feature space [7]. To accomplish complex learning tasks and feature selection, deep learning approaches, in contrast, rely on multi-layered representation and abstraction of input data [6, 8].

Shallow learning techniques have been successful in many well-constrained problems, but they frequently fail in fields like computer vision and natural language processing, which demand the extraction of well-represented features from data [9]. In order to represent complicated concepts, deep learning builds several layers of small characteristics [10]. Deep learning architectures have greatly benefited from the recent explosion in data availability and improvements in chip processing power [11]. Deep learning has made significant advances in computer vision, natural language processing, and other fields by enabling machines to learn from and extract useful information from massive volumes of unstructured data. Deep learning techniques have the ability to open up new doors and applications in a variety of industries as they develop and get better.

2.2. Federated Learning

Federated learning is a learning strategy that enables the training of a centralized model on unevenly dispersed data across a network of nodes [14, 15]. It was first developed by Google in 2016 [12, 13]. The requirement to train models with data from users’ mobile devices, which cannot be centrally stored in data centres due to privacy concerns, is the driving force behind federated learning [16]. Federated learning offers considerable privacy advantages over other machine learning models since it transmits just the bare minimum updates required for model improvement. The training aims to determine whether these updates are made, preserving the raw data on users’ devices. Federated learning performs better, creating better models, as there are more nodes available for training. Organizations can create and implement machine learning models that respect user privacy, reduce the amount of data that must be transmitted, and adjust to the particular limitations of distributed data sources by utilizing federated learning. This novel technique has the potential to transform machine-learning applications across a range of sectors, particularly those involving sensitive user data.

2.3. Privacy in Machine Learning

Deep learning is not now the core focus of machine learning privacy security; instead, conventional techniques are. The privacy of the data needed to train a model or serve as input to an existing model, the model itself, and the model’s output are the three main goals of privacy preservation in machine learning.

A method known as Secure Multi-Party Computation (SMC) is utilized to secure the computations performed in the middle of collaborative machine learning on confidential inputs [37]. The use of SMC approaches for privacy-preserving deep learning, however, is still an open issue because they typically carry non-trivial performance overheads. Decision trees [37], linear regression functions [38], association rules [39], Naive Bayes classifiers [40], and k-means clustering [41] are a few examples of SMC approaches used with machine learning algorithms.

The model uses privacy-preserving probabilistic inference [42], speaker identification [43], and computing on encrypted data [44] as privacy-preserving strategies. Contrarily, differential privacy [45] is a well-liked method for privacy-preserving machine learning and has been used with many other algorithms, including boosting [46] and principal component analysis [47].

Despite the existence of these methods, previous research has not addressed the issue of distributed stochastic gradient descent for collaborative deep learning with numerous participants. The system allows participants to control the learning objective, protect the privacy of their training data, and apply the jointly learned model to their inputs without disclosing the inputs or outputs. It also performs better than cryptographic techniques like secure multi-party computation or homomorphic encryption. A system that satisfies all three privacy objectives in the context of collaborative neural network training.

3. AI-Powered Anomaly Detection and Analysis

Finding patterns in data that do not match expected behaviour is known as anomaly detection. In many application fields, these nonconforming patterns are frequently referred to as anomalies, outliers, discordant findings, exceptions, aberrations, surprises, oddities, or contaminants. The two terms most frequently used in the context of anomaly identification are anomalies and outliers, which are occasionally used interchangeably. Anomaly detection is widely used in a range of applications, including military surveillance for enemy activity, intrusion detection for cyber-security, malfunction detection in safety-critical systems, and fraud detection for credit cards, insurance, or healthcare. Anomalies in data can be translated into substantial, and frequently crucial, actionable information in a wide range of application fields, which highlights the significance of anomaly detection[22]. For instance, an unusual pattern of traffic in a computer network may indicate that a compromised computer is transmitting private information to an unauthorized location. Malignant tumours may be present in an abnormal MRI image [23]. Anomalies in credit card transaction data could be a symptom of identity or credit card theft [24], while unusual sensor readings from a spacecraft could signal a problem with one of its components [25]. In the statistics field, research on finding outliers or abnormalities in data dates back to the 19th century[26]. Several anomaly detection methods have been created over time in various research circles. While some of these approaches are more general in nature, many of them have been created expressly for particular application fields.

Anomaly detection is a significant topic that has been researched in numerous disciplines and application domains. Certain anomaly detection techniques are broader in scope, whereas others have been specifically created for a given application domain[22].

A novel anomaly detection model for OpenStack cloud settings is provided in the study[17], and it is represented in Figure 1 by combining stacked and bidirectional long short-term memory (LSTM) neural networks[20,21]. Data from an OpenStack environment[19,20] was used to collect the information, which produced a dataset with 10 features and a class label for each data point. The LSTM-based model received these features as input. For a training set and a test set, the model's performance was assessed using binary cross-entropy as the loss function. The results demonstrated the efficiency of the suggested approach in detecting anomalies within an OpenStack cloud environment, with a high detection accuracy of 94.61 percent on the training set and 93.98 percent on the test set. In an OpenStack setup with three nodes—a controller, storage, and compute node—the experiment was carried out utilizing common library packages. Three components make up the suggested methodology: failure prediction, data preprocessing, and data collecting. Data was gathered using collected, and Iperf and Stress-ng were used to inject artificial load into the virtual computers. When compared to the mean squared error loss function, the Stacked LSTM model performed better when utilizing the binary cross-entropy loss function. This study demonstrates the potential of deep learning techniques for improving security and reliability in cloud computing systems.

Several industries, including network security, finance, and healthcare, have used autoencoders, recurrent neural networks (RNNs), and long short-term memory (LSTM) networks for anomaly identification. These models are proficient at handling high-dimensional data and complicated pattern recognition. Anomaly detection frequently makes use of algorithms like dimensionality reduction (e.g., PCA[28]) and clustering (e.g., K-means[27]). Even without a prior understanding of labels or categories, these approaches can identify patterns and group comparable data pieces. When labelled data is available, methods such as support vector machines (SVMs), decision trees, and ensemble approaches (such as random forests) are used. These techniques work well for finding known anomaly kinds, but they may have trouble with anomalies that haven't been seen before. To better capture, the underlying structure of data, more sophisticated unsupervised and self-supervised learning algorithms should be developed. This would enable more accurate anomaly detection even in extremely complex datasets. Anomalies are often rare events relative to normal data points, hence methods must be improved to solve class imbalance, which is prevalent in anomaly detection activities. Creating more interpretable models and methods would help human specialists validate and comprehend the results by illuminating the causes of abnormalities that are found. improving models' defence against adversarial attacks, which may trigger erroneous alarms or fail to detect anomalies, particularly in security-critical applications. The creation of techniques for transferring knowledge from one domain to another enables models to adapt to new surroundings even in the absence of a large amount of labelled data. By incorporating active learning approaches, the demand for large labelled datasets can be decreased while detection performance is improved with little to no human involvement.

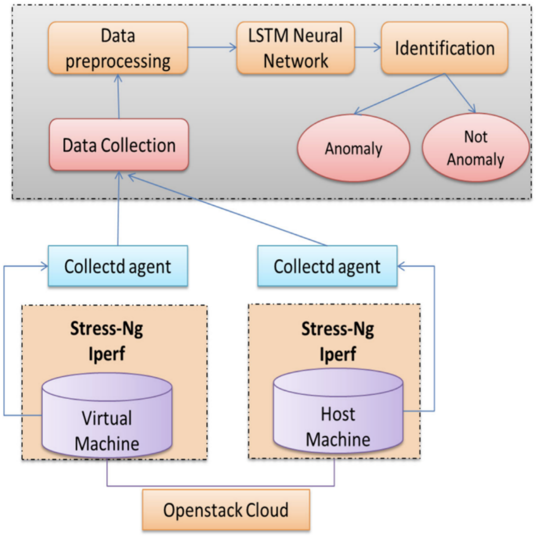
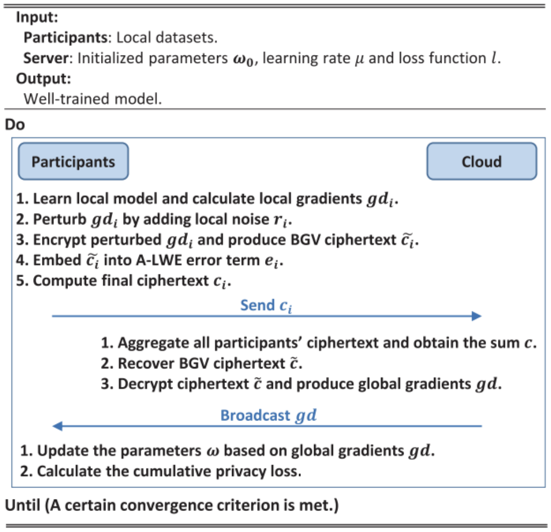


Figure 1: Overall view of the proposed methodology

4. Federated Learning and AI-Driven Privacy Enhancement

Due to its capacity to train machine learning models on decentralized data while protecting privacy, federated learning is drawing more attention from academia and industry. Using techniques like homomorphic encryption[29], secure multi-party computation[30], or federated learning, AI models can be trained to operate on encrypted data. In order to safeguard privacy, this enables businesses to do calculations on sensitive data without disclosing the actual information. Many industries, including healthcare, finance, telecommunications, smart cities, and autonomous cars, are using federated learning. To address privacy and security problems, researchers are creating new federated learning algorithms and methodologies, including secure aggregation, differential privacy, and asynchronous communication.

Deep learning and industrial artificial intelligence (IAI) technologies have been used to address a variety of difficulties across numerous industries with the introduction of Industry 4.0. Yet, traditional centralized training is no longer appropriate in sensitive data-driven contexts like healthcare and driverless vehicles due to privacy concerns. Federated learning has become a well-liked substitute because it allows users to study a common model collectively without disclosing their personal information. The paper[31] suggests a Federated Learning Method for Industrial AI that is Efficient and Privacy-Enhanced. Federated learning has drawbacks despite the potential it holds. Shared parameters can be used by adversaries to compromise industrial applications, including autonomous vehicle navigation systems, wearable medical gadgets, and industrial robot decision-making. The authors provide an effective and privacy-enhanced federated learning (PEFL) strategy for IAI to allay these worries. Figure 2 depicts the proposed scheme's high-level view.



```
Input:
  Participants: Local datasets.
  Server: Initialized parameters $\omega_0$, learning rate $\mu$ and loss function $l$.
Output:
  Well-trained model.
Do
  Participants
    1. Learn local model and calculate local gradients $gd_i$.
    2. Perturb $gd_i$ by adding local noise $r_i$.
    3. Encrypt perturbed $gd_i$ and produce BGV ciphertext $\widetilde{c_i}$.
    4. Embed $\widetilde{c_i}$ into A-LWE error term $e_i$.
    5. Compute final ciphertext $c_i$.
  Send $c_i$
  Cloud
    1. Aggregate all participants' ciphertext and obtain the sum $c$.
    2. Recover BGV ciphertext $\tilde{c}$.
    3. Decrypt ciphertext $\tilde{c}$ and produce global gradients $gd$.
  Broadcast $gd$
    1. Update the parameters $\omega$ based on global gradients $gd$.
    2. Calculate the cumulative privacy loss.
Until (A certain convergence criterion is met.)
```

Figure 2: High-level view of the proposed scheme.

PEFL is intended to be non-interactive, avoiding the disclosure of personal information even when numerous parties conspire. The authors show that PEFL beats existing methods in terms of accuracy and efficiency through comprehensive trials using real-world data. For IAI applications, especially those involving sensitive data or crucial activities, this strategy offers a more secure and resilient federated learning framework.

It's difficult to provide reliable security and privacy in federated learning. It's important to handle potential weaknesses like model inversion and membership inference attacks. Enhancing privacy may be accomplished by creating new cryptographic methods and incorporating them with federated learning. Heterogeneous hardware and data distributions are frequently present in real-world federated learning scenarios. Future success will depend on algorithms that can handle data heterogeneity, device capability fluctuations, and changing network conditions. The majority of federated learning strategies concentrate on training global models, however, personalized models can deliver greater performance for particular people or devices. Future development will focus on ways for individualized federated learning, where local models are customized for each student. The right incentive mechanisms are needed to promote participation in federated learning networks. Federated learning systems can encourage greater participation and enhance overall performance by designing reward systems, accounting for resource contributions, and assuring fairness.

5. Privacy-preserving machine learning

The authors of the paper Privacy-preserving Machine Learning in Cloud[32] provide novel methods for enabling the deployment and training of deep neural networks on encrypted data. They hope to allay security and privacy worries raised by the use of Machine Learning as a Service (MLaaS) platforms by accomplishing this. Finding a balance between the degree of polynomial approximation utilized for activation functions and the performance of the final model is the key problem with this approach. The authors concentrate on using Chebyshev polynomials, which are orthogonal systems of polynomials, to approximate activation functions like Sigmoid and ReLU. They find the best approximation polynomial for these activation functions by considering the mean of the feature values in the dataset and approximating the function over the interval [-mean, mean]. Chebyshev polynomials are used to approximate the function over a symmetric interval for Sigmoid. The ReLU function's non-continuous nature and the makeup of its derivative, however, make approximating it difficult. The authors suggest that ReLU can be approximated by a continuous and indefinitely differentiable function, similar in structure to Sigmoid. The authors show that it is feasible and practical to train neural networks using encrypted data and create encrypted predictions by incorporating these polynomial approximations into neural networks and carrying out operations over encrypted data. Their empirical findings demonstrate that the suggested strategies offer precise training and classification that protects privacy while keeping reasonable computation speeds.

They used HELib [33], a homomorphic encryption toolkit, to create neural networks over encrypted data. They performed tests on a virtual system with 8GB RAM, a Core i5, and Ubuntu 14.04, calculating the training process's running time as well as the precision of the models it produced.

Based on the trade-off between running time and algorithm complexity, they selected a security level of 80 (k=80) and a parameter L=20 to build encryption schemes in HELib. They employed the Crab, Fertility, and Climate Model datasets from the UC Irvine Machine Learning Repository [34], each of which had a distinct number of cases and attributes.

They developed neural networks using different batch sizes and hidden layer counts (1, 2, 3, 4, and 5). They discovered that when the batch size rose, the training time remained constant, even when the batch size tripled. This proves that batch learning is effective for training over encrypted data. They were only able to run the feed-forward function and carry out concurrent computations for a batch of unseen instances during the classification phase. According to the findings, classification with a batch size of 576 instances had a running time of 0.04 seconds for one hidden layer and 0.1 seconds for two hidden layers. The running time was 0.014 seconds for one hidden layer and 0.036 seconds for two hidden layers when the batch size was increased to 6144. The study shows that neural networks can be used over encrypted data with respectable performance and economy. The authors note that they did not use any parallelization techniques in their implementation, suggesting that further improvements in performance could be achieved with such optimizations.

Modern privacy-preserving neural networks based on homomorphic encryption (HE) and secure multi-party computing (SMC) methods were compared to the authors' findings. Contemporary HE-based methods: The authors made a comparison between their work and CryptoNets [35], another project that seeks to use HE to build neural network classification. The accuracy and throughput of their method outperform CryptoNets. When categorizing ciphertext with a batch size of 8,192, they outperformed CryptoNets, making 163,840 predictions per hour as opposed to 51,739 predictions per hour, regardless of whether they used Sigmoid or ReLU activation functions. Contemporary SMC-based methods: The authors contrasted their research with that of Mohassel and Zhang [16], a cutting-edge SMC-based strategy. Their approach has a number of benefits over SMC-based ones, including a reduction in client-server connections. In contrast to SMC protocols, which demand contacts for each operation, the authors' solution only requires communication when the noise hits a threshold. Their solution has a substantially lower communication count than [36]. Their method, which uses the ReLU activation function over the MNIST dataset, yields an accuracy of 95.15 percent, compared to the procedure in [36]'s 93.4 percent. Overall, the suggested method outperforms cutting-edge HE and SMC-based strategies in terms of training and classification accuracy while maintaining privacy. The authors intend to investigate the approximation of non-continuous functions employed in deep neural network algorithms in future work by running experiments on larger datasets and more intricate neural networks.

The document[48] Privacy-Preserving With the use of a collaborative deep learning system, Deep Learning demonstrates how several parties can cooperatively develop an accurate neural network model without disclosing their input datasets. A typical neural network with two hidden layers is depicted in Figure 3. This system's primary elements and protocols consist of: 1. N participants, each of whom has a local, private dataset for training. 2. A learning aim and network architecture that is shared by all participants. 3. A parameter server that keeps the most recent parameter values accessible to all parties.

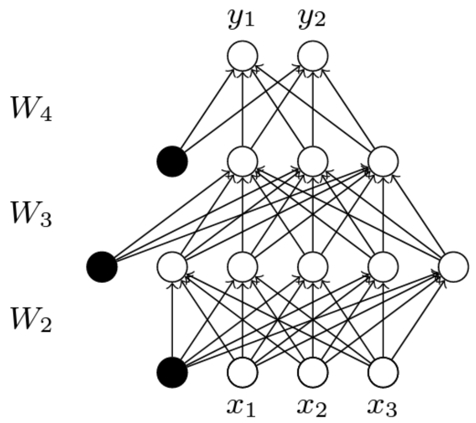


Figure 3: A neural network with two hidden layers. Black circles represent the bias nodes.

The system features a parameter exchange protocol that enables users to get the most recent parameter values at each local SGD epoch and post gradients of chosen neural-network parameters to the parameter server. This prevents overfitting to a single participant's local training dataset and enables participants to independently converge to a set of parameters. Each person can independently and privately evaluate the network on fresh data after it has been trained, without interacting with other participants. Each participant individually executes the distributed selective SGD (DSSGD) method, which has five phases for each learning epoch:

1. Download a θd fraction of parameters from the server and overwrite local parameters with the downloaded values.
2. Run one epoch of SGD training on the local dataset.
3. Compute Δw(i), the vector of changes in all parameters in step 2.
4. Select which gradients to share at the end of each local epoch.
5. Update Δw(i) with bound(Δw(i),γ) and add random noise before uploading it.

The parameter server manages participant upload and download requests as well as initializes the parameter vector w(global). Distributed selected SGD functions because it makes the learning process more stochastic, preventing local SGD from overfitting to a limited local dataset. Round-robin, random order, and asynchronous updates are just a few of the situations that the parameter exchange protocol supports. These scenarios help the distributed SSGD work well by boosting stochasticity. The primary elements and procedures of the cooperative deep learning system are shown in Figure 4.

In conclusion, the suggested method enables collaborative deep learning while protecting user privacy, allowing users to benefit from each other's models without disclosing

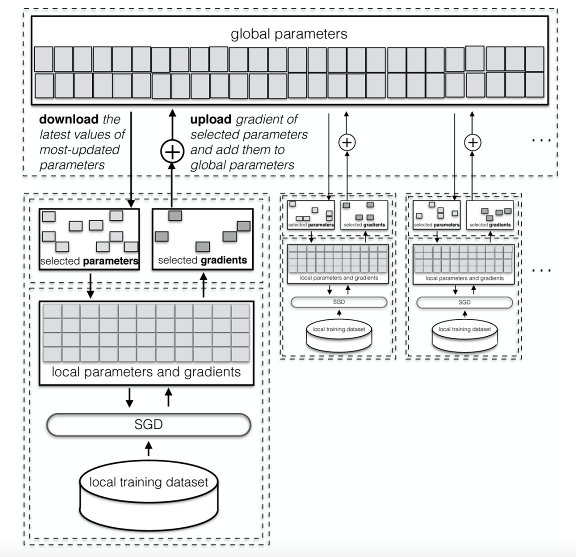


Figure 4: High-level architecture of the deep learning system. An abstract model of the parameter server, which maintains global values for the parameters, is depicted at the top.

their input data. On the MNIST dataset and the SVHN dataset, two significant datasets, the researchers examined their systems. The SVHN dataset has 600,000 training photos and 10,000 test instances of house numbers taken from Google Street View images, compared to the MNIST dataset's 60,000 training examples and 10,000 test examples of handwritten digits.

Multi-layer perceptron (MLP) and convolutional neural network architectures were employed in the study (CNN). Round-robin, random order, and asynchronous parameter exchange protocols were used to create distributed SSGD. Because random order performed nearly identically to round robin, it was skipped. Centralized SGD and Standalone SGD served as the baseline scenarios. For choosing gradients to submit to the parameter server, they used two criteria: biggest values and random with the threshold. The findings demonstrated that SSGD can achieve almost the same accuracy as SGD by simply sharing a tiny portion of gradients. The most accurate method was distributed SSGD with round-robin parameter exchange, which was almost as accurate as centralized SGD. The fact that the percentage of shared parameters had a greater influence on accuracy than the number of participants suggests that distributed SSGD does not need a large number of participants to increase accuracy. The number of participants, the frequency of parameter changes, and the timing of parameter exchange all have an impact on the communication cost of distributed SSGD. The conclusions were drawn with the assumption that every participant would share the biggest gradients with the other participants. The alternative approach involves randomly selecting gradients whose values are higher than a predetermined threshold. Analysis was also done on the learning rates, participant count, and convergence of DSSGD for various datasets. The findings demonstrate that, regardless of the number of participants, higher learning rates lead to faster convergence to maximum accuracy, demonstrating that the distributed and selective nature of DSSGD does not alter the gradient descent algorithm's general behaviour. It is conceivable to create a parameter server that is unaware of the identity of uploaders in order to avoid a nosy server from associating the changes of each participant. Participants can upload their gradients and authenticate themselves in an anonymous manner. To conceal participants' identities, scalable anonymous communication methods with verifiable security can be utilized. A fully distributed implementation of the parameter storage system, in which each participant is in charge of a randomly chosen subset of the parameters, is also made possible by the independence of parameters from one another in distributed SSGD. This scheme's intricate design is reserved for later work.

The researchers concluded by proposing a novel distributed training method based on selective stochastic gradient descent that can be used to any kind of neural network and maintains participant privacy while maintaining model accuracy. This strategy can assist in bringing the advantages of deep learning to fields where data owners are prohibited from disclosing their information owing to privacy concerns.

The applicability of privacy-preserving approaches for large-scale machine learning applications may be constrained by the fact that they frequently have higher computing costs. More scalable and effective algorithms will enable PPML to be used in more applications. Several privacy-preserving methods are intricate and challenging to use in real-world situations. Adoption of these techniques will be aided by the development of user-friendly tools and frameworks that streamline their deployment. It's essential to provide interoperable solutions that can be integrated to fulfil diverse privacy requirements in various contexts, especially now that there are so many privacy-preserving strategies available. Techniques for protecting privacy should stand up well to a variety of threats and hostile actions. The future of PPML depends on creating more robust algorithms that can withstand various forms of attacks. In conclusion, even though machine learning for privacy protection has advanced significantly, there is still much to be done. The use of PPML techniques will be facilitated by improvements in scalability, usability, interoperability, robustness, and benchmarking, which will also help to protect data privacy in AI applications.

6. AI-driven User Authentication

In order to protect the security and privacy of systems and data, user authentication is essential. Researchers and businesses are concentrating on creating advanced techniques to confirm and identify user identities as a result. Figure 5 illustrates three general categories for authentication systems: solutions that are knowledge-based, physiologically biometric, and behaviorally biometric [49, 50]. Understanding-based authentication The user must explicitly submit credentials for this sort of authentication, such as a password, personal identification number (PIN), or graphical PIN. These documents serve as identification proof for the person. Biological characteristics, such as fingerprints, iris patterns, and face photographs, are used in this method of user verification. Based on these physiological traits, user identities are differentiated using machine learning techniques.

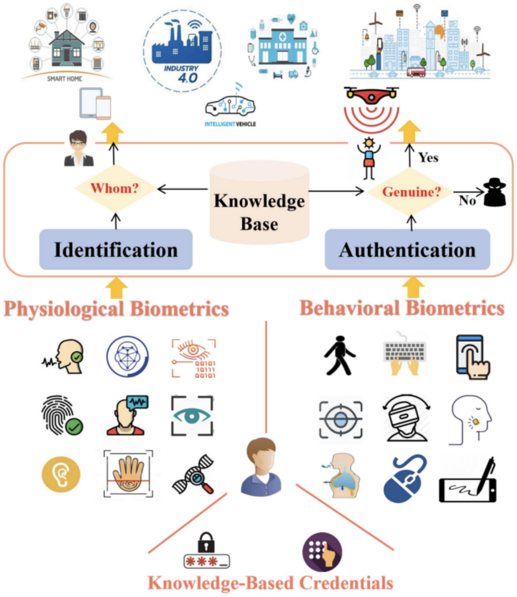


Figure 5: Overview of credentials for user authentication and identification and their applications.

The category of behavioural biometric-based authentication covers authentication techniques that take into account a person's distinctive characteristics, such as their walking style, typing patterns, and touchscreen dynamics. Users can be authenticated using these behavioural patterns. There are additionally two subcategories of authentication systems:

Identifying whether a user is an authorized person or an unauthorized visitor is the goal of user authentication. It checks to see if the system access request comes from a legitimate user.

User identification entails identifying the particular user who is currently logging in to the system. The goal of user identification is to ascertain the real identity of the system user.

We will now talk about AI-based techniques used in Continuous Authentication (CA) for IoT devices to identify user identities based on behavioural fingerprints. Figure 6 illustrates the pipeline for AI-based CA approaches, which consists of three main parts: data preparation, feature extraction, and classification algorithms.

A. Data Preprocessing: The ability to separate high-quality data from unprocessed, noisy, inconsistent, and redundant data is crucial. For continuous person identification based on behavioural patterns, data segmentation and filtering are required [51].

B. Feature Extraction: From the training dataset, useful features reflecting user behaviour are extracted via feature engineering. Behaviours are frequently described by statistical features and frequency-domain patterns. Handcrafted feature engineering is time-consuming and strongly dependent on the knowledge of domain specialists [52]. Researchers have used deep learning-based techniques to learn behaviour representations as a result of these limitations [53].

C. Classification Algorithms: When there are few negative samples (imposters), supervised classification algorithms for user authentication may struggle [54]. For unbalanced datasets with skewed class distributions, one-class classification algorithms, such as one-class support vector machines (SVMs) [55] and isolated forests (iForest) [56], are frequently used as anomaly detection techniques.

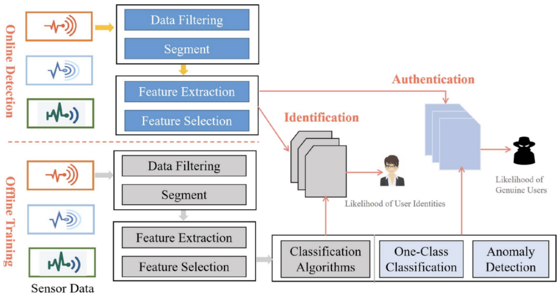


Figure 6: General workflow of machine learning-based user identification and authentication.

The one-class SVM algorithm, which finds a hyperplane encircling the majority of positive samples from the origin with the greatest margin, is a semi-supervised classification technique [57]. Anomaly identification using the unsupervised algorithm iForest is based on the finding that anomalies are uncommon and considerably distinct from the rest of the data [56]. Anomaly sites are isolated closer to the tree's root in an ensemble of binary search trees called iTrees. Due to its capacity for handling big data quantities and high-dimensional problems, linear time complexity, and suitability for resource-constrained IoT devices, iForest is a good choice for user authentication [58].

The goal of user identification is to identify the present user and ascertain their eligibility to utilize IoT applications or devices. Deep learning-based solutions and traditional classification algorithms can both be used to identify users based on their behaviour.

Conventional Classification: These methods, including SVM, random forest, naive Bayes, and artificial neural networks, rely on specially created characteristics that have been approved by professionals. These techniques take a lot of time and effort, though, and they might not be appropriate for fields that are changing quickly.

Deep Learning-Based Classification: Due to their enhanced performance, these algorithms are becoming more and more popular. Based on the kinds of neural networks used, they can be categorized into three groups:

a) Convolutional Neural Networks (CNN): CNNs are widely adopted in user authentication and identification systems to detect personal patterns from fingerprints, eyes, and more. They are also used to detect the liveness of biometrics against presentation attacks.

b) Recurrent Neural Networks (RNN) and their variants, such as Long Short-Term Memory (LSTM): RNNs are useful for processing sequential behaviour data with outstanding performance. They have been used in various user authentication and identification studies based on behavioural biometrics.

c) Generative Adversarial Networks (GAN): GANs consist of two submodels, a generator model to generate new examples, and a discriminator model to classify whether generated examples are real data or generated examples. GANs have been widely adopted to generate high-fidelity human biometrics that can bypass authentication systems.

User identification systems' performance has been greatly improved by deep learning-based approaches like CNN, RNN, and GAN. When compared to standard classification algorithms, they can adjust to changes in user behaviour and offer more reliable and accurate responses.

AI has been used to authenticate users using a variety of biometric modalities, including voice, iris, facial, and fingerprint recognition. Convolutional neural networks (CNNs), one type of deep learning technology, have been particularly effective in enhancing the precision and effectiveness of these systems. The analysis of distinctive user behaviour patterns, such as keyboard dynamics, mouse movement, gait, or touch-screen interaction patterns, is increasingly done using artificial intelligence (AI). For processing data on sequential activity, methods like recurrent neural networks and long short-term memory networks have proven successful. Systems powered by AI have been created to continuously monitor and authenticate individuals as they interact with hardware or software. This strategy contributes to security maintenance without sacrificing user experience.

AI systems must be built to withstand adversarial and presentational attacks (spoofing). To create methods for detecting and thwarting these dangers, more study is required. Systems for authentication must be scalable and effective to support large-scale deployments as the number of users and devices increases. Lightweight model development and existing algorithm optimization should be the main areas of research. Systems for AI-based authentication should be developed to safeguard users' privacy. To protect user data throughout the authentication process, strategies including federated learning, secure multi-party computing, and differential privacy can be used. Systems powered by AI ought to be able to adjust to evolving user behaviour patterns and new device or communication modalities. Transfer learning, meta-learning, and unsupervised learning techniques can be explored to improve the adaptability of these systems.

Considering the tremendous improvements in behavioural and biometric systems, the current state of AI in user authentication is encouraging. In terms of robustness, scalability, privacy protection, adaptability, and usability, there is still potential for improvement. Further advancements in these fields will further improve the functionality and efficacy of AI-driven user authentication systems.

7. Conclusion

The paper has shown the substantial potential of artificial intelligence-based privacy protection techniques in tackling the escalating difficulties associated with preserving user privacy in the digital age. Deep learning, federated learning, privacy-preserving machine learning, and user authentication are examples of AI-driven approaches that have become effective tools for strengthening the security and privacy of personal data and digital footprints.

The applications of these AI-based techniques currently in use, their inherent limits, and the difficulties that must be addressed to ensure their responsible and successful adoption have all been highlighted throughout the article. I have also examined the numerous privacy protection situations that these strategies can be used for, illustrating their adaptability and potential for advancement.

Yet, there are possible hazards and ethical issues that need to be taken into account, just like with any new technology. It takes continual cooperation between researchers, decision-makers, and industry stakeholders to ensure the appropriate and transparent use of AI in privacy protection. This partnership will be crucial for improving current practices, creating fresh approaches, and creating legal frameworks that support both innovation and consumer privacy.

In the end, artificial intelligence has the ability to completely transform how personal information is protected in the digital era, but it's critical to find a balance between utilizing its powers and upholding people's rights and expectations. We must continue to be cautious in our efforts to guarantee that these technologies are used for the benefit of society while limiting any risks and unexpected consequences. As research develops and AI-based privacy protection approaches grow more advanced.

References

[1] Nicolas Papernot, Patrick McDaniel, Arunesh Sinha, and Michael P Wellman. 2018. SoK: Security and privacy in machine learning. In 2018 IEEE European Symposium on Security and Privacy (EuroSP). IEEE, 399–414.

[2] Sebastian Risi and Mike Preuss. 2020. Behind DeepMind's AlphaStar AI that Reached Grandmaster Level in StarCraft II. KI-Künstliche Intelligenz 34, 1 (2020), 85–86.

[3] Marco Barreno, Blaine Nelson, Russell Sears, Anthony D Joseph, and J Doug Tygar. 2006. Can machine learning be secure?. In Proceedings of the 2006 ACM Symposium on Information, computer and communications security. 16–25.

[4] Luis Munoz-Gonzalez, Javier Carnerero-Cano, Kenneth T. Co, Emil C. Lupu.2019. ChallengesandAdvances in Adversarial Machine Learning. Resilience and Hybrid Threats: Security and Integrity for the Digital World 55 (2019), 102

[5] Li, Jian-hua. "Cyber security meets artificial intelligence: a survey." Frontiers of Information Technology & Electronic Engineering 19.12 (2018): 1462-1474.

[6] Giovanni Apruzzese, Michele Colajanni, Luca Ferretti, Alessandro Guido, and Mirco Marchetti. 2018. On the effectiveness of machine and deep learning for cyber security. In 2018 10th International Conference on Cyber Conflict (CyCon). IEEE, 371–390.

[7] Li Deng. 2014. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Transactions on Signal and Information Processing 3 (2014).

[8] In Lee and Yong Jae Shin. 2020. Machine learning for enterprises: Applications, algorithm selection, and challenges. Business Horizons 63, 2 (2020), 157–170.

[9] LiDeng.2012. Threeclassesofdeeplearningarchitectures andtheir applications: a tutorial survey. APSIPA transactions on signal and information processing (2012).

[10] Xiaoyong Yuan, Pan He, Qile Zhu, and Xiaolin Li. 2019. Adversarial examples: Attacks and defenses for deep learning. IEEE transactions on neural networks and learning systems 30, 9 (2019), 2805–2824.

[11] Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal. 2017. Chapter 10- Deep learning. In Data Mining: Practical Machine Learning Tools and Techniques (fourth ed.), Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal (Eds.). Morgan Kaufmann, 417–466.

[12] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communicationefficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics. PMLR, 1273–1282.

[13] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. 2019. Federated machine learning: Concept and applications. ACMTransactions on Intelligent Systems and Technology (TIST) 10, 2 (2019), 1–19.

[14] K. A. Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloé M Kiddon, Jakub Konečný, Stefano Mazzocchi, Brendan McMahan, Timon Van Overveldt, David Petrou, Daniel Ramage, and Jason Roselander. 2019. Towards Federated Learning at Scale: System Design. In SysML 2019. https://arxiv.org/ abs/1902.01046

[15] Jakub Konečn‘y, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. 2016. Federated Learning: Strategies for Improving Communication Efficiency. In NIPS Workshop on Private Multi-Party Machine Learning. https://arxiv.org/abs/1610.05492

[16] Jakub Konečn‘y, H Brendan McMahan, Daniel Ramage, and Peter Richtárik. 2016. Federated optimization: Distributed machine learning for on-device intelligence. arXiv preprint arXiv:1610.02527 (2016).

[17] Girish, L., and Sridhar KN Rao. ”Anomaly detection in cloud environment using artificial intelligence techniques.” Computing (2021): 1-14.

[18] Radford B, Apolonio L, Trias A, Simpson J (2018) Network traffic anomaly detection using recurrent neural networks. arXiv:1803.10769

[19] Girish L (2019) Efficient monitoring of time series data using dynamic alerting. IndiaRxiv. https://doi. org/10.26634/jcom.6.2.14870

[20] Girish L, Rao SKN (2020) Quantifying sensitivity and performance degradation of virtual machines using machine learning. J Comput Theor Nanosci. https://doi.org/10.1166/jctn.2020.9019

[21] Rashmi TV, Prasanna MK, Girish L (2015) Load balancing as a service in Openstack-Liberty. Int J Sci Technol Res 4(8):70–73

[22] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. ”Anomaly detection: A survey.” ACM computing surveys (CSUR) 41.3 (2009): 1-58.

[23] Spence, Clay, Lucas Parra, and Paul Sajda. ”Detection, synthesis and compression in mammographic image analysis with a hierarchical image probability model.” Proceedings IEEE workshop on mathematical methods in biomedical image analysis (MMBIA 2001). IEEE, 2001.

[24] Aleskerov, Emin, Bernd Freisleben, and Bharat Rao. ”Cardwatch: A neural network based database mining system for credit card fraud detection.” Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr). IEEE, 1997.

[25] Fujimaki, Ryohei, Takehisa Yairi, and Kazuo Machida. ”An approach to spacecraft anomaly detection problem using kernel feature space.” Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. 2005.

[26] Edgeworth, Francis Ysidro. ”Xli. on discordant observations.” The london, edinburgh, and dublin philosophical magazine and journal of science 23.143 (1887): 364-375.

[27] Hartigan, John A., and Manchek A. Wong. ”Algorithm AS 136: A k-means clustering algorithm.” Journal of the royal statistical society. series c (applied statistics) 28.1 (1979): 100-108.

[28] Reddy, G. Thippa, et al. ”Analysis of dimensionality reduction techniques on big data.” Ieee Access 8 (2020): 54776-54788.

[29] Yi, Xun, et al. Homomorphic encryption. Springer International Publishing, 2014.

[30] Goldreich, Oded. ”Secure multi-party computation.” Manuscript. Preliminary version 78.110 (1998).

[31] Hao, Meng, et al. ”Efficient and privacy-enhanced federated learning for industrial artificial intelligence.” IEEE Transactions on Industrial Informatics 16.10 (2019): 6532-6542.

[32] Hesamifard, Ehsan, et al. ”Privacy-preserving machine learning in cloud.” Proceedings of the 2017 on cloud computing security workshop. 2017.

[33] Shai Halevi and Victor Shoup. 2014. Algorithms in HElib. In Advances in Cryptology- CRYPTO- 34th Annual Cryptology Conference, Santa Barbara, CA, USA, Proceedings, Part I. 554–571.

[34] M. Lichman. 2013. UCI Machine Learning Repository. (2013). http://archive.ics. uci.edu/ml

[35] Nathan Dowlin, Ran Gilad-Bachrach, Kim Laine, Kristin Lauter Michael Naehrig, and John Wernsing. 2016. CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy. Technical Report MSR-TR-2016-3.

[36] P. Mohassel and Y. Zhang. 2017. SecureML: A System for Scalable PrivacyPreserving Machine Learning. In 2017 IEEE Symposium on Security and Privacy (SP). 19–38.

[37] Y. Lindell and B. Pinkas. Privacy preserving data mining. In CRYPTO, 2000.

[38] W. Du, Y. Han, and S. Chen. Privacy-preserving multivariate statistical analysis: Linear regression and classification. In SDM, volume 4, pages 222–233, 2004.

[39] J. Vaidya and C. Clifton. Privacy preserving association rule mining in vertically partitioned data. In KDD, 2002.

[40] J. Vaidya, M. Kantarcıo˘glu, and C. Clifton. Privacy-preserving naive bayes classification. VLDB, 17(4):879–898, 2008.

[41] M. Pathak, S. Rane, W. Sun, and B. Raj. Privacy preserving probabilistic inference with Hidden Markov Models. In ICASSP, 2011.

[42] M. Pathak and B. Raj. Privacy-preserving speaker verification and identification using gaussian mixture models. Trans. Audio, Speech, and Language Processing, 21(2):397–406, 2013.

[43] M. Barni, P. Failla, R. Lazzeretti, A. Sadeghi, and T. Schneider. Privacy-preserving ECG classification with branching programs and neural networks. Trans. Info. Forensics and Security, 6(2):452–468, 2011.

[44] C. Dwork. Differential privacy. In Encyclopedia of Cryptography and Security, pages 338–340. Springer, 2011.

[45] P. Simard, D. Steinkraus, and J. Platt. Best practices for convolutional neural networks applied to visual document analysis. In Document Analysis and Recognition, 2013.

[46] C. Dwork, G. Rothblum, and S. Vadhan. Boosting and differential privacy. In FOCS, 2010.

[47] K. Chaudhuri, A. Sarwate, and K. Sinha. A near-optimal algorithm for differentially-private principal components. JMLR, 14(1):2905–2943, 2013.

[48] Shokri, Reza, and Vitaly Shmatikov. ”Privacy-preserving deep learning.” Proceedings of the 22nd ACM SIGSAC conference on computer and communications security. 2015.

[49] A. Ometov, V. Petrov, S. Bezzateev, S. Andreev, Y. Koucheryavy, and M. Gerla, “Challenges of multi-factor authentication for securing advanced IoT applications,” IEEE Netw., vol. 33, no. 2, pp. 82–88, Mar./Apr. 2019.

[50] A. K. Jain, A. Ross, and S. Pankanti, “Biometrics: A tool for information security,” IEEE Trans. Inf. Forensics Security, vol. 1, no. 2, pp. 125–143, Jun. 2006.

[51] J. Zhang, B. Wei, W. Hu, S. S. Kanhere, and A. Tan, “Human identification using WiFi signal,” in Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops), Mar. 2016, pp. 1–2.

[52] Z. Cai, C. Shen, and X. Guan, “Mitigating behavioral variability for mouse dynamics: A dimensionality-reduction-based approach,” IEEE Trans. Human–Mach. Syst., vol. 44, no. 2, pp. 244–255, Apr. 2014.

[53] G. Batchuluun, R. A. Naqvi, W. Kim, and K. R. Park, “Body-movement-based human identification using convolutional neural network,” Expert Syst. Appl., vol. 101, pp. 56–77, Jul. 2018.

[54] L. M. Manevitz and M. Yousef, “One-class SVMs for document classification,” J. Mach. Learn. Res., vol. 2, pp. 139–154, Mar. 2002.

[55] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, “Support vector method for novelty detection,” in Proc. 12th Int. Conf. Neural Inf. Process. Syst. (NIPS), Cambridge, MA, USA, 1999, pp. 582–588.

[56] F. T. Liu, K. M. Ting, and Z. Zhou, “Isolation forest,” in Proc. 8th IEEE Int. Conf. Data Min., 2008, pp. 413–422.

[57] D. Shin and S. Kim, “Nearest mean classification via one-class SVM,” in Proc. Int. Joint Conf. Comput. Sci. Optim., vol. 1, 2009, pp. 593–596.

[58] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation-based anomaly detection,” ACM Trans. Knowl. Discovery Data, vol. 6, no. 1, pp. 1–39, Mar. 2012.