10-3-2024

Secure Computing with Hidden Markov Models: Overcoming Barriers to GDPR Compliance

Antonio Goncalves
INESC-ID , CINAV Portugal, agoncalveslx@gmail.com

Anacleto Correia
CINAV Portugal, cortez.correia@marinha.pt

Secure Computing with Hidden Markov Models: Overcoming Barriers to GDPR Compliance

António Gonçalves, INESC-ID , CINAV Portugal, agoncalvesLx@gmail.com
Anacleto Correia, CINAV Portugal, cortez.correia@marinha.pt

Abstract

In recent years, the protection of personal data has become a central concern for governments, businesses, and individuals. The General Data Protection Regulation (GDPR), implemented by the European Union in May 2018, set a new standard for data privacy and security. This regulation aims to ensure that personal data is processed fairly, transparently and securely. In parallel, Hidden Markov Models (HMMs) have emerged as a powerful statistical tool for modelling stochastic processes in various areas. However, the application of HMMs in contexts involving sensitive personal data raises serious privacy and security concerns. GDPR compliance poses additional challenges to the secure implementation of these models, requiring organizations to adopt appropriate technical and organizational measures. This study explores the challenges and solutions to implementing secure computing of HMMs, addressing anonymization, encryption, protection against cyberattacks, and regulatory compliance.

Keywords: Secure computing; Data Protection; GDPR; HMMs; Anonymization

1. INTRODUCTION

In recent years, the protection of personal data has become a central concern for governments, businesses, and individuals. The General Data Protection Regulation (GDPR), implemented by the European Union in May 2018, set a new standard for data privacy and security by imposing stringent requirements on organizations that process personal information. This regulation aims to ensure that personal data is handled fairly, transparently, and securely, protecting the rights of individuals and imposing severe penalties for non-compliance (der Vlies & Hesselink, 2017; Townend & others, 2018) .

In parallel, Hidden Markov Models (HMMs) have emerged as a powerful statistical tool for modelling stochastic processes in various areas, including speech recognition, bioinformatics, and finance. HMMs' ability to capture temporal dependencies and hidden patterns in sequential data makes them especially valuable for a wide range of applications. However, the application of HMMs in contexts involving sensitive personal data raises serious privacy and security concerns. GDPR compliance poses additional challenges to the secure implementation of these models, requiring

ISSN 2183-489X

organizations to take steps to protect personal data at every stage of processing (X. Wang et al., 2019; Yu et al., 2019).

In this context, the concept of secure computing emerges as the application of techniques and practices that ensure the protection of sensitive data against unauthorized access, misuse or disclosure during processing, storage or transmission. This includes the use of techniques and measures to ensure the integrity and confidentiality of the data (Pashchenko et al., 2020; S. Wang et al., 2019).

The GDPR sets out fundamental principles for the processing of personal data, including data minimization, purpose limitation, integrity, and confidentiality. To comply with these principles when using HMMs, organizations must ensure that personal data is appropriately anonymized or pseudonymised, that only necessary data is processed, and that robust security measures are in place to protect against unauthorized access and data breaches. In addition, data subjects should be informed about how their data is used and have the right to access, correct, and delete their personal information.

Given the growing importance of data privacy and the effectiveness of HMMs, the central question of this study arises:

- **What are the challenges and solutions for implementing secure computing of Hidden Markov Models in systems that must comply with GDPR regulations?**

This article aims to explore the key challenges and solutions associated with implementing secure computing using GDPR-compliant HMMs. Initially, the technical and regulatory challenges that arise in the integration of HMMs into systems that handle personal data will be discussed. These challenges include the need to protect data privacy while maintaining the functionality and accuracy of templates, implementing appropriate security measures to prevent unauthorized access, and ensuring that all data processing practices are aligned with GDPR requirements.

Subsequently, the article will address the technical solutions that can be used to overcome these challenges. Among the solutions discussed will be the use of anonymization techniques to protect data during processing, the application of homomorphic encryption to allow calculations on encrypted data, and the use of secure multi-party computing to perform joint operations on sensitive data without compromising privacy. In addition, practical examples and case studies illustrating the successful application of some of these techniques in scenarios will be presented. Finally, it will discuss how these solutions can be efficiently integrated into existing systems, highlighting best practices and strategies to ensure ongoing compliance with the GDPR while harnessing the potential of HMMs for complex data analysis.

2. CHALLENGES IN IMPLEMENTING SECURE COMPUTING WITH HMM

Secure computing encompasses the adoption of a series of techniques and practices designed to ensure the effective protection of sensitive data against unauthorized access, misuse, or improper disclosure, strictly complying with the requirements of the GDPR. This concept is crucial in environments where data must be kept secure at every stage of its lifecycle, including processing, storage, and transmission. HMMs represent one of the advanced approaches in data analytics and machine learning, being employed in a wide range of applications, from speech recognition to bioinformatics to the financial sector. Implementing secure computing practices in conjunction with HMMs is crucial to ensure that all analytics and operations are performed within a robust security framework, maintaining data integrity and confidentiality throughout the process. Table 1 presents a summary of the challenges, which will be detailed below.

#	CHALLENGE	DESCRIPTION
D01	Model Complexity and Data Security	Ensure the anonymization and pseudonymization of sensitive data used in the training and validation of HMMs.
D02	Data Exchange and Storage	Implement efficient encryption to protect data at rest and in transit without sacrificing processing power.
D03	Cryptographic Algorithms	Balancing security and efficiency with cryptographic algorithms, such as homomorphic encryption, that still face performance challenges.
D04	Protection against Intrusions and Attacks	Protect against cyber-attacks by applying intrusion detection techniques and secure software development practices.
D05	Sensitive Data Management	Ensure access only to authorized personnel through robust authentication and authorization policies and monitor access.
D06	Compliance with the lawfulness of the processing	Ensure the processing of data in a lawful, transparent manner and with the explicit consent of individuals.
D07	Differential Privacy	Implement differential privacy to protect sensitive information without compromising the accuracy of HMMs.

Table 1 – Challenges in Implementing Secure Computing with HMM.

Model Complexity and Data Security. HMMs are statistical models that describe stochastic systems with hidden states. The complexity of these models means that their implementation requires a significant amount of data for training and validation. Protecting this sensitive data throughout the HMM lifecycle is crucial to ensuring GDPR compliance. The first challenge is to ensure the **anonymization** and **pseudonymization** of data, avoiding the direct or indirect identification of individuals. Anonymization consists of transforming data in such a way that individuals cannot be directly or indirectly identified. This is essential to ensure that the data collected and utilized in HMMs cannot be traced back to specific individuals. In cases where complete anonymization is not feasible, **pseudonymization** can be utilized to replace direct identifiers (such as names and identification numbers) with pseudonyms, which allow for some level

of traceability without exposing the identity of individuals. Ensuring data anonymization and pseudonymization in secure computing with HMMs presents several complex challenges. Technically, complete anonymization is difficult because anonymized data can be re-identified through combination with other data sources, and the preservation of data usefulness is compromised when anonymization techniques distort information. In addition, re-identification techniques are constantly evolving, making continuous updates of protection methods necessary. Legally, complying with the stringent requirements of the GDPR is complex, requiring transparency in anonymization and pseudonymization practices. Operationally, it is challenging to integrate these practices into all business processes in a consistent manner, in addition to requiring ongoing training and awareness of employees. In terms of security, protection against internal and external threats is crucial, along with proper cryptographic key management to prevent re-identification. Finally, striking a balance between protecting privacy and maintaining the efficiency and accuracy of HMMs is a significant technical challenge (El Emam et al., 2020; Li et al., 2021).

Data Exchange and Storage. The secure storage of data used to train HMMs is a primary concern. Data encryption techniques, both at rest[1] and in transit[2], are necessary to ensure that data is not compromised. Implementing such techniques without sacrificing the efficiency and processing power of HMMs is a considerable challenge. Cryptographic key management and access control are fundamental elements to ensure the security of information systems and are crucial to prevent unauthorized access. Cryptographic keys, which allow data to be encrypted and deciphered, must be strictly managed to prevent it from being exposed or compromised, thus ensuring the confidentiality and integrity of the data. On the other hand, effective access control ensures that only authorized users can access sensitive information by applying principles such as least privilege and conducting ongoing audits and monitoring. Negligence in these areas can facilitate security breaches, resulting in significant data loss, reputational damage, and legal penalties, especially under strict data protection regulations like the GDPR (Gupta et al., 2019; Y. Zhang et al., 2020).

Cryptographic Algorithms. The application of cryptographic algorithms in HMMs must balance when trying to balance security and efficiency, due to the computationally intensive nature of HMMs and the heavy computational load imposed by encryption algorithms. In cryptography, cryptographic key management and access control are fundamental elements to ensure the security of information systems, being crucial to prevent unauthorized access. Cryptographic keys, which allow for the encryption and decryption of data, must be strictly managed to prevent their exposure or compromise, thus ensuring the confidentiality and integrity of the data. Advanced algorithms such

[1] Data at rest refers to data that is stored on a physical storage device and is not actively being moved from one location to another.

[2] Data in transit is data that is being transferred from one location to another, either within a local network or over the internet.

as homomorphic encryption, which allows operations to be performed on encrypted data without the need to decrypt it, have great potential in this context, although they still face significant challenges in terms of performance. On the other hand, effective access control ensures that only authorized users can access sensitive information by applying principles such as least privilege and conducting ongoing audits and monitoring. Negligence in these areas can facilitate security breaches, resulting in significant data loss, reputational damage, and legal penalties, especially under strict data protection regulations like the GDPR. Secure computing with HMMs requires the implement of algorithms that ensure data privacy without significantly compromising the speed and accuracy of processing (Johnson et al., 2019; Liu et al., 2019).

Protection against Intrusions and Attacks. Protection against intrusions and cyberattacks poses an ongoing challenge in implementing secure computing with HMMs. Attacks such as a process by which sensitive information is deduced from observable patterns in seemingly anodyne data, and the exploitation of software vulnerabilities, can significantly compromise data integrity and confidentiality. Therefore, the application of robust intrusion detection techniques, along with the implementation of secure software development practices, are essential measures to mitigate these risks. These practices help to identify and prevent attacks that aim to exploit weaknesses in both the processes and technologies used, thus ensuring greater security of data and systems (Shokri et al., 2019; Yeom et al., 2018).

Management of Sensitive Data. The management of sensitive data, especially in the context of the operation of HMMs, requires the implementation of[3] strict and well-structured[4] security policies to ensure that only authorized personnel have access to the necessary information. These policies should include strong authentication mechanisms, such as multi-factor authentication, which provides an additional layer of security by requiring multiple forms of identity verification. Additionally, role-based authorization is crucial, as it allows for granular control over data access, ensuring that each user only has the privileges that are strictly necessary to perform their specific roles. To complement these measures, it is imperative to implement robust access monitoring and auditing systems, which allow not only to detect suspicious or anomalous activities in real time, but also to facilitate a rapid response to potential security incidents. These practices are critical to mitigate the risk of unauthorized access and to protect the integrity and confidentiality of the data handled by HMMs (Jang et al., 2021; Khraisat et al., 2019; J. Zhang et al., 2019).

[3] Security policies are guidelines and procedures established by an organization to protect the integrity, confidentiality, and availability of an organization's data and resources.

[4] A well-structured security policy outlines detailed procedures for data protection, specifying responsibilities and ensuring regulatory compliance.

Compliance with the lawfulness of the processing. GDPR compliance[5] presents a significant challenge for organizations, given their requirement for rigor in the protection of personal data. To ensure that the data used in HMMs is handled in a lawful, transparent manner and with clearly defined purposes, organizations must implement meticulous procedures. This includes obtaining explicit consent from individuals, which must be documented and verifiable, as well as ensuring the right to erasure when requested, as stipulated by the right to be forgotten. In addition, the implementation of secure computing solutions must be accompanied by a constant focus on regulatory compliance, which requires frequent reviews and updates of security processes to adapt to legislative changes and new interpretive guidance from data protection authorities. These measures are essential not only to avoid legal penalties, but also to strengthen the confidence of users and business partners in the integrity of the organization's data management practices (Tikkinen-Piri et al., 2018; Voigt & von dem Bussche, 2017).

Differential Privacy. Differential privacy is a mathematical technique used to ensure that HMM results do not reveal sensitive information about specific individuals by introducing controlled random perturbations[6] into the data or results. Implementing differential privacy in HMMs requires a complex balance between data utility and privacy protection, as adding enough noise to protect privacy can compromise the accuracy and effectiveness of the models. This technical challenge is an active area of research, focused on developing efficient algorithms that minimize performance impact, determining optimal privacy values, and creating test methodologies to ensure compliance with differential privacy criteria, without sacrificing the functionality of HMMs in practical applications (Balle et al., 2020; Erlingsson et al., 2019).

Secure computing encompasses techniques to protect sensitive data from unauthorized access, misuse, or disclosure, in compliance with the GDPR, and is essential at every stage of the data lifecycle. HMMs, used in areas such as speech recognition and bioinformatics, must be implemented with robust security practices to maintain data integrity and confidentiality. This involves challenges such as ensuring anonymization and pseudonymization of data, implementing efficient encryption, balancing security with the efficiency of cryptographic algorithms, protecting against cyberattacks, managing sensitive data with restricted access and strong authentication, ensuring strict legal compliance, and enforcing differential privacy without compromising the accuracy of HMMs.

[5] Compliance with the GDPR requires the implementation of strict measures to protect personal data and guarantee the rights of data subjects.

[6] Controlled random disturbances in the data add intentional noise to protect privacy while maintaining the overall usefulness of the results.

3. SOLUTIONS TO OVERCOME BARRIERS IN GDPR COMPLIANCE

In the context of secure computing with HMMs, several barriers need to be overcome to ensure that data is adequately protected without compromising the efficiency and accuracy of the models. Below, we present detailed solutions to each of the key challenges identified, covering everything from data anonymization to implementing differential privacy, providing a comprehensive framework for achieving GDPR compliance and ensuring data security.

D01: Model Complexity and Data Security. To ensure the anonymization and pseudonymization of sensitive data used in HMM training and validation, it is essential to adopt robust data transformation techniques. Anonymization involves modifying data so that individuals cannot be directly or indirectly identified, which is crucial for GDPR compliance. This can be achieved through techniques such as data masking, generalization, and k-anonymity, which obscure identifiable details while preserving the utility of the data. When complete anonymization is not feasible, pseudonymization can be employed. This technique replaces identifiable information with pseudonyms, allowing for some level of traceability without exposing actual identities. Effective pseudonymization requires careful management of pseudonym mappings and ensuring that the pseudonymized data cannot be easily re-identified through data linkage or inference attacks. Both anonymization and pseudonymization present technical challenges, including maintaining data utility, preventing re-identification, and complying with evolving legal standards. Table 2 shows the technical principles used.

CONTEXT	TECHNIQUE	DESCRIPTION
Anonymization	Removal of Direct Identifiers	Eliminates information that can directly identify an individual, such as names, ID numbers, email addresses, phone numbers, etc.
	Generalization	Converts specific data into broader categories, reducing the granularity of the information.
	Suppression	Completely removes values from certain columns or records that could lead to an individual's identification.
	Data Disruption	It introduces small, random changes to data values, making it difficult to re-identify.
	Shuffling	Rearranges the values within a column randomly to preserve the distribution but remove the link between the values and individuals.
Pseudonymization	Replacing Direct Identifiers with Pseudonyms	It swaps direct personal identifiers for aliases, which are surrogate values that can be reversed with the use of a separate key.
	Cryptography	It applies cryptographic algorithms to transform identifiable data into an encrypted form that can only be reversed with the correct key.
	Tokenization	It replaces sensitive data with tokens that have no extrinsic value or that can be mapped back to the original data only through a secure system.
	Lookup Tables	Uses mapping tables to replace actual values with pseudonyms, allowing rollback through table queries.

Table 2 – Data Anonymization and Pseudonymization Techniques.

D02: Data Exchange and Storage. Implementing efficient encryption to protect data at rest and data in transit is critical to ensure security without sacrificing processing power. Data at rest refers to information stored in databases, servers, or other storage devices, which must be protected through encryption to prevent unauthorized access and breaches. Techniques such as full-disk encryption, file-level encryption, and database encryption are commonly used to secure data at rest. Data in transit, on the other hand, refers to information being transferred across networks, which requires encryption methods like TLS (Transport Layer Security) and VPNs (Virtual Private Networks) to prevent interception and eavesdropping. Efficient encryption techniques must balance strong security with minimal performance impact, ensuring that data remains secure without significantly slowing down system operations. Table 3 shows the technical principles used.

CATEGORY	TECHNIQUE	DESCRIPTION	APPLICATION
Encryption at Rest	Advanced Encryption Standard (AES)	AES is one of the most widely used and secure symmetric encryption algorithms. It uses 128-, 192-, or 256-bit keys to encrypt and decrypt data.	It can be implemented to protect files, databases, and any data stored on hard drives or SSDs.
	Full Disk Encryption (FDE)	FDE encrypts all data stored on a hard drive or SSD, ensuring that any data on the disk is protected.	Used in entire storage devices such as laptops, desktops, and servers.
	File or Folder Encryption	Encrypts specific files or folders instead of the entire disk.	Used to protect individual files and folders that contain sensitive data.
Encryption in Transit	Transport Layer Security (TLS)	TLS is an encryption protocol that protects data during transmission between systems, ensuring the security of communication.	Used to secure web communications (HTTPS), emails (STARTTLS), and other data transmissions.
	Secure Sockets Layer (SSL)	SSL is the predecessor of TLS, used to encrypt communication between a client and a server.	Although it is being replaced by TLS, it is still used in some legacy applications.
	Internet Protocol Security (IPsec)	IPsec is a set of protocols that encrypts and authenticates IP packets, ensuring the security of communication between networks.	Used to set up VPNs (Virtual Private Networks) and secure communication between devices on private networks.
	Secure Shell (SSH)	SSH is a network protocol that provides a secure way to access and transfer data between computers.	Used for secure login to remote systems and secure file transfer.

Table 3 – Data Exchange and Storage Techniques.

D03: Cryptographic Algorithms. Balancing security and efficiency with cryptographic algorithms, such as homomorphic encryption, is challenging due to the impact on performance. Homomorphic encryption allows computations on encrypted data without decrypting it, providing a high level of security. However, this process is computationally intensive, often leading to performance bottlenecks. To overcome this, it is necessary to develop algorithms that offer robust security without compromising efficiency. Techniques such as optimizing encryption schemes, leveraging hardware accelerators, and using hybrid cryptographic methods can enhance performance. Hybrid approaches combine different encryption techniques to balance security and speed, utilizing the strengths of each method to mitigate their individual weaknesses. Continuous research and development in cryptographic algorithms focus on reducing computational overhead while maintaining strong security guarantees.

SOLUTION	DESCRIPTION	APPLICATION	BENEFITS
Symmetric Encryption (AES)	It uses the same key to encrypt and decrypt data, offering high speed and efficiency.	Can be used to encrypt training data and HMMs model parameters.	High security and performance, widely supported and efficient for large volumes of data.
Asymmetric Encryption (RSA)	It uses a key pair (public and private) for encryption and decryption, ensuring that only the intended recipient can decrypt the message.	It can be used to protect the exchange of symmetric keys and other sensitive data during model setup.	High security for key exchange and authentication, although less efficient for encryption of large volumes of data.
Homomorphic Encryption	Allows operations to be performed directly on encrypted data without the need to decrypt it.	Ideal for performing complex calculations on HMMs without compromising the privacy of the underlying data.	Maintains data privacy during processing, which is essential for environments where security and privacy are critical.
Public Key Cryptography and Public Key Infrastructure (PKI)	It uses public keys to encrypt data and public key infrastructures to manage and distribute those keys.	It can be used to authenticate and protect the integrity of data exchanged during communication between different parts of the HMM system.	It provides a secure and scalable method for managing cryptographic keys and ensuring the authenticity and integrity of data.
Digital Signatures	It uses asymmetric cryptography to ensure the authenticity and integrity of the data by digitally signing the data or models.	It can be used to validate the origin and integrity of data and HMM models exchanged between different entities.	Ensures that the data has not been altered and that it comes from a trusted source, which is essential to the integrity of the models.
Secure Multi-Party Computation (SMPC) Protocol	It allows multiple parties to jointly perform calculations on data without revealing their private information.	Useful in scenarios where multiple entities collaborate to train HMMs without exposing their sensitive data.	Protects data privacy while enabling collaboration, which is essential for applications in consortia and partnerships.

Table 4 – Cryptographic Algorithm Techniques.

D04: Protection against Intrusions and Attacks. Protecting against cyberattacks is crucial to ensure the integrity, confidentiality, and availability of data processed by HMMs. Intrusion detection techniques and secure software development practices are essential to prevent unauthorized access, malicious manipulations, and system disruptions. These methods help maintain the accuracy of models, protect sensitive data, and comply with regulations, increasing user and stakeholder confidence in the system. A multi-layered security approach is often employed, combining firewalls, intrusion detection systems (IDS), and intrusion prevention systems (IPS) to monitor and defend against potential threats in real time. Regular security audits and penetration testing are critical for identifying vulnerabilities and enhancing security measures. Additionally, implementing robust encryption protocols and secure coding practices reduces the risk of data breaches. Employee training on cybersecurity best practices and establishing a culture of security awareness are also vital components. These comprehensive strategies not only safeguard the data and systems but also ensure regulatory compliance and build trust among users and stakeholders. Table 5 shows the technical principles used.

SOLUTION	DESCRIPTION	APPLICATION	BENEFITS
Intrusion Detection Techniques	Monitor and analyse suspicious activity in real-time to identify and block threats.	Applied in network systems to monitor suspicious traffic and activity.	Identifies and blocks threats in real-time, increasing system security.
Strong Encryption	Use strong encryption to protect sensitive data during storage and transmission.	Used to protect data stored and transmitted across networks and devices.	Ensures the protection of data against unauthorized access and interception.
Secure Software Development	Implement practices such as code reviews and security testing to prevent vulnerabilities.	Applied during the software development cycle to prevent security breaches.	Prevents the introduction of vulnerabilities during software development.
Multi-Factor Authentication	Require multiple forms of identity verification for access to systems and data.	Used to access critical systems and data, adding an extra layer of security.	Reduces the risk of unauthorized access to critical systems and data.
Continuous Monitoring	Maintain constant vigilance over the IT environment to quickly detect and respond to incidents.	Implemented in IT operations for continuous surveillance and threat response.	It enables early detection of threats and rapid response to incidents.
Regular Updates and Patches	Keep all systems and software up to date to fix known vulnerabilities.	Applied to all systems and software to keep security up to date.	Keeps systems safe from known threats.
Data Backup & Recovery	Implement backup and recovery strategies to ensure data availability.	Used to create regular backups and data recovery plans.	Ensures data availability and integrity, even after security incidents.
Network Segmentation	Divide the network into segments to limit lateral movement of attackers within the environment.	Implemented in corporate networks to control and limit access to different segments.	Limits lateral movement of attackers, reducing the impact of potential intrusions.

Table 5 – Protection Techniques against Intrusions and Attacks.

D05: Management of Sensitive Data. Ensuring that only authorized personnel have access to sensitive data requires the implementation of robust authentication and authorization policies. This involves multiple layers of security and advanced techniques to ensure that data is protected from unauthorized access. Table 6 shows the technical principles used.

SOLUTION	DESCRIPTION	APPLICATION	BENEFITS
Multi-Factor Authentication	Requires multiple forms of identity verification to ensure that only authorized users access the systems.	Used to access critical systems and data, increasing security.	Significantly reduces the risk of unauthorized access.
Role-Based Authorization (RBAC)	Controls access to resources based on users' roles within the organization.	Implemented in IT systems to limit access based on job roles.	Ensures that users only have access to the data they need for their roles.
Continuous Monitoring and Auditing	Continuously monitors access to systems and data, logging and analysing all suspicious activity.	Applied in IT operations to detect and respond to unauthorized access.	Enables early detection and rapid response to suspicious activity.
Strict Security Policies	Define and enforce clear security policies to manage access to and use of sensitive data.	Used at every layer of the organization to maintain compliance and security.	Helps maintain compliance with GDPR and other privacy regulations.

Table 6 – Sensitive Data Management Techniques.

D06: Compliance with the lawfulness of the processing. To ensure compliance with the GDPR, it is necessary to implement rigorous processes that ensure the protection of personal data. Table 7 shows the technical principles used.

SOLUTION	DESCRIPTION	APPLICATION	BENEFITS
Informed Consent	Obtain explicit consent from individuals before collecting and processing their data.	Implemented through clear and detailed consent forms.	Ensures that individuals are aware of and agree to the use of their data.
Data Transparency	Keeping individuals informed about how their data is being used, stored, and protected.	Enforced through accessible privacy policies and regular communications to individuals about the use of their data.	It promotes trust and transparency by ensuring that individuals know how their data is handled.
Legal Compliance	Ensure that all data collection and processing practices comply with the GDPR and other applicable regulations.	Integrated into all data collection and processing processes to ensure they are in line with legal requirements.	Avoids fines and penalties associated with non-compliance with GDPR and other regulations.
Consent Management	Utilize tools and systems to manage and record individuals' consent, allowing for easy access and auditing.	Used to maintain up-to-date records of consent and ensure that individuals can easily modify them	It facilitates consent management and ongoing compliance with privacy and data protection requirements.

Table 7 – Techniques to maintain compliance with the lawfulness of the processing.

D07: Differential Privacy. Implementing differential privacy to protect sensitive information without compromising the accuracy of HMMs involves adding statistical noise to data or analysis results. Using Laplacian or Gaussian mechanisms can prevent the re-identification of individuals while ensuring data privacy. Table 8 shows the technical principles used.

SOLUTION	DESCRIPTION	APPLICATION	BENEFITS
Laplacian Mechanism	Adds Laplacian noise to the results of the calculations to ensure that the outputs do not reveal information about specific individuals.	Used in scenarios where it is necessary to provide aggregated results without compromising individual privacy.	It ensures the privacy of individuals by providing accurate and protected results.
Gaussian Mechanism	Applies Gaussian noise to data or results to provide a layer of privacy while balancing data accuracy.	Applied in statistical analysis and machine learning where data accuracy is critical, but privacy must be maintained.	Maintains a balance between data accuracy and privacy, which is essential for in-depth analytics.
Additive Noise	It introduces random noise to the data to obscure individual information while maintaining the usefulness of the data for analysis.	Used in data storage and processing systems to protect sensitive information while maintaining data functionality.	Protects sensitive information without compromising the usefulness of the data for analysis.
Local Differential Privacy	Implements differential privacy directly into individuals' data before it is sent for processing, ensuring privacy at the source.	Implemented in mobile applications and data collection systems to ensure that the privacy of individuals' data is preserved from the time of collection.	Ensures data privacy from the point of collection, increasing user trust and GDPR compliance.

Table 8 – Techniques for implementing Differential Privacy.

In summary, this section comprehensively addresses the techniques and challenges associated with secure computing with HMMs, highlighting the need to protect sensitive data and ensure GDPR compliance. The proposed solutions, ranging from anonymization and pseudonymization to the implementation of differential privacy, offer a robust framework for addressing the issues of security, privacy, and efficiency.

These measures are designed to prevent unauthorized access, protect against data breaches, and ensure that sensitive information is not inadvertently disclosed. By incorporating advanced encryption techniques, secure key management, and rigorous access controls, organizations can create a secure environment for processing and analyzing data using HMMs, thus maintaining the integrity and confidentiality of the data throughout its lifecycle.

Additionally, the integration of secure computing practices with HMMs not only ensures compliance with legal regulations but also enhances the overall trust and reliability of the system.

4. CASE STUDIES

In the context of medical data analytics, patient privacy is of utmost importance. The application of anonymization techniques allows data to be used for modelling and predictions without compromising individual privacy. This case study demonstrates the use of Hidden Markov Models (HMMs) to predict health states with synthetic medical data. The comparison between the results obtained with non-anonymized and anonymized data illustrates the effectiveness of anonymization.

Data anonymization is crucial for several reasons: 1) **Privacy Protection**: Anonymization ensures that personally identifiable information cannot be associated with specific individuals, complying with regulations such as the General Data Protection Regulation (GDPR). 2) **Data Security**: In case of unauthorized access or security breaches, anonymized data minimizes the risk of re-identification and misuse and 3) **Foster Trust**: Patient trust is critical to the ongoing collection and use of medical data, and anonymization helps maintain that trust.

4.1. Study Methodology

In the methodology of the case study presented, meticulous steps were followed to ensure both the protection of data privacy and the effectiveness of the analytical model. The main steps of the methodology include the generation of synthetic data, the anonymization of this data, the training of a Hidden Markov Model (HMM), and the evaluation and comparison of the results. Below, I detail each of these steps:

- **Synthetic Data Generation**. This step involves creating a synthetic dataset that mimics real medical data, making it easier to develop and test the model without compromising real patient information. The variables generated include age, glucose levels, blood pressure, and body mass index (BMI). The selection of these variables was guided by their relevance in real-world medical applications, especially in contexts where continuous monitoring may be critical for diagnosis and treatment.
- **Anonymization of Data.** Anonymization is performed to protect the privacy of simulated individuals by applying specific techniques that reduce the granularity of information and make it difficult to directly identify or infer individual identities. In the specific case:
    - **Age Categorization**: Ages are grouped into categories instead of being used as exact values.
    - **Rounding Values:** Glucose levels, blood pressure, and BMI are rounded to remove accurate accuracy, thus decreasing the possibility of re-identification.
- **Hidden Markov Model (HMM) training.** With the prepared data, both anonymized and non-anonymized, a Hidden Markov Model is trained. HMM is chosen for its ability to

model temporal sequences and hidden states that may represent unobservable health transitions directly in the data. The model is configured to identify potential health states based on observable variables and their transitions over time.

- **Evaluation and Comparison of Results.** Finally, the results are evaluated and compared to understand the impact of anonymization on the model results:
    - **Visualization of Hidden States:** Visualization helps to visually compare the patterns and consistency of predicted hidden states between anonymized and non-anonymized datasets.
    - **Classification Report:** The accuracy, recall, and F1 score of predicted states are calculated to quantify differences in the performances of models trained with anonymized versus non-anonymized data.

The described methodology not only respects data privacy, but also allows us to evaluate how different anonymization techniques can impact the usefulness of data in predictive applications. Comparing the results helps to ensure that anonymization does not significantly degrade the quality of the analyses, ensuring a balance between privacy and usefulness of the data in healthcare settings.

In this case study, synthetic data will be used. The use of synthetic data in case studies is valid mainly because it allows detailed analysis to be carried out without compromising the privacy of real individuals, ensuring compliance with regulations such as the GDPR. Synthetic data can be generated to replicate the characteristics and statistical relationships of the actual data, giving you control and flexibility to test diverse scenarios. However, to ensure validity, it is crucial that synthetic data is faithfully representative, and that rigorous validation and verification methods are carried out. Thus, synthetic data can be a powerful tool for developing and testing AI models, performing simulations and impact analyses, and providing educational materials without risks of sensitive data exposure (Patki et al., 2016; Yoon et al., 2020).

4.2. Case Study Results

Synthetic data were generated for 1000 patients, including: Ages ranging from 30 to 50 years, Blood glucose levels with a mean of 150 mg/dL and standard deviation of 10 mg/dL. Blood pressure with a mean of 130 mmHg and standard deviation of 5 mmHg and Body mass index (BMI) with a mean of 28 and standard deviation of 2.

Data were anonymized using the following techniques:

- **Age Categorization:** Grouping into age groups of 30-40 and 40-50 years.
- **Rounding Values:** Glucose levels and blood pressure were rounded to the nearest 10, and BMI was rounded to the nearest integer.

An HMM was trained using both non-anonymized and anonymized data, with 3 hidden states representing possible health states of the patients. The predicted hidden states using non-anonymized and anonymized data are presented below (Tables 8 and 9):

ACTS	GLUCOSE	BLOOD_PRESSURE	BMI	HIDDEN_STATE
42	154.967142	131.768291	29.0	2
48	148.617357	129.767135	27.0	2
31	151.476885	132.98433	27.0	2
35	165.230299	133.692585	27.0	1
48	145.658466	129.974242	28.0	2

Table 8 – HMM for non-anonymized data.

ACTS	GLUCOSE	BLOOD_PRESSURE	BMI	HIDDEN_STATE
40-50	150.0	130.0	29.0	2
40-50	150.0	130.0	27.0	2
30-40	150.0	130.0	27.0	2
30-40	170.0	130.0	27.0	1
40-50	150.0	130.0	28.0	2

Table 9 – HMM for anonymized data.

Ranking Report

The performance of the hidden states predicted by the HMM was evaluated using a confounding matrix. Comparing the results of the trained models with non-anonymized and anonymized data reveals that anonymization does not significantly compromise the quality of the predictions. The accuracy, recall, and F1-score metrics remain elevated, indicating that the HMM model can effectively identify patterns of health status transitions across both datasets. The graphs of the hidden states show significant consistency, with overall patterns maintained even after anonymization. This demonstrates that anonymization can protect data privacy while preserving its usefulness for predictive analytics.

This case study demonstrated that the application of anonymization techniques to medical data allows to protect the privacy of individuals without compromising the accuracy of predictions made by HMM models. The effectiveness of the model on anonymized data suggests that similar techniques can be applied in real-world scenarios, ensuring the privacy and usefulness of the data for predictive analytics in healthcare.

5. CONCLUSION

The protection of personal data has become a central concern in an increasingly digitized and interconnected world. This study addressed the challenges and solutions associated with implementing Hidden Markov Models (HMMs) in compliance with the General Data Protection Regulation (GDPR). With the increasing use of personal data in a variety of applications, from speech recognition to bioinformatics and finance, it is essential to ensure that this data is processed securely and in compliance with stringent regulations.

The main challenges identified in this study include the complexity of anonymizing and pseudonymizing data, implementing efficient encryption to protect data at rest and in transit, and protecting against intrusions and cyberattacks. Each of these challenges presents significant technical and operational barriers, which need to be overcome to ensure the security and privacy of personal data.

Data anonymization and pseudonymization are key to protecting the privacy of individuals. However, these techniques present complex technical challenges. Complete anonymization is difficult to achieve, as anonymized data can be re-identified through combination with other data sources. In addition, the preservation of the usefulness of the data is compromised when anonymization techniques distort the information. Pseudonymization, which replaces direct identifiers with pseudonyms, offers an intermediate level of protection, allowing for some degree of traceability without exposing the identity of individuals.

Implementing efficient encryption to protect data at rest and in transit is another significant challenge. Encryption at rest ensures that data stored on hard drives or SSDs is protected from unauthorized access. Techniques such as Advanced Encryption Standard (AES) and Full Disk Encryption (FDE) are widely used for this purpose. Encryption in transit, which protects data during transmission, uses protocols such as Transport Layer Security (TLS) and Secure Sockets Layer (SSL).

The application of cryptographic algorithms must balance security and efficiency. Algorithms such as homomorphic encryption allow operations to be performed on encrypted data without the need to decrypt it, offering a high level of security. However, these algorithms still face performance challenges that need to be overcome to be viable in practical applications.

Protection against intrusions and cyberattacks is an ongoing challenge in implementing secure computing with HMMs. Attacks such as data inference and exploiting software vulnerabilities can compromise data integrity and confidentiality. Applying intrusion detection techniques, as well as implementing secure software development practices, are essential to mitigate these risks. Continuous monitoring and regular security updates help protect the system from emerging threats and ensure compliance with data protection regulations.

Managing sensitive data involves ensuring that only authorized personnel have access to the data necessary for the operation of HMMs. This requires the implementation of robust security policies, including the use of strong authentication and role-based authorization. Monitoring and auditing access to data is crucial for detecting and responding to suspicious activity, ensuring that data is handled in a secure and GDPR-compliant manner.

To ensure compliance with the GDPR, it is necessary to implement rigorous processes that ensure the protection of personal data. This includes obtaining explicit consent from individuals, maintaining transparency about how data is used, and ensuring that data is processed lawfully and for specific purposes. Managing consent and enforcing accessible privacy policies are key to fostering trust with individuals and ensuring ongoing compliance with regulations.

Differential privacy is a mathematical technique used to ensure that HMM outputs do not reveal sensitive information about specific individuals. This is achieved by introducing controlled random perturbations into the data or results. Implementing differential privacy in HMMs requires a complex balance between data utility and privacy protection. Adding enough noise to protect privacy can compromise the accuracy and effectiveness of the models, posing a significant technical challenge.

Case studies illustrated the feasibility of the technical solutions presented in this study. For example, in the analysis of medical data, the application of anonymization techniques has made it possible to use the data for modelling and predictions without compromising individual privacy. The comparison between the results obtained with non-anonymized and anonymized data demonstrated that anonymization could protect data privacy while preserving its usefulness for predictive analytics. The precision, recall, and F1-score metrics remained elevated, indicating that the HMM model was able to identify patterns of health status transitions effectively in both datasets.

In conclusion, the integration of HMMs into systems that handle sensitive personal data can be carried out in a secure and effective manner, provided that appropriate technical and organizational measures are taken to protect data privacy and integrity. This includes implementing robust encryption techniques, ensuring secure data exchange and storage, and employing advanced anonymization and pseudonymization methods to safeguard individuals' identities. Additionally, adopting differential privacy and secure software development practices helps mitigate risks associated with data breaches and cyberattacks. The continuous evolution of data protection techniques and compliance with regulations such as GDPR are key to ensuring the security of personal data and fostering individuals' confidence in the use of advanced technologies for complex analysis.

By prioritizing data privacy and security, organizations can leverage the full potential of HMMs for sophisticated analyses while maintaining public trust and meeting regulatory requirements, thus paving the way for innovative and responsible use of advanced machine learning technologies.

REFERENCES

Balle, B., Bell, J., Gasc'on, A., & Nissim, K. (2020). Privacy amplification by mixing and diffusing data. *ArXiv Preprint ArXiv:2009.03777*.

der Vlies, L., & Hesselink, M. W. (2017). GDPR: A New Era in Data Protection. *European Data Protection Law Review*.

El Emam, K., Arbuckle, L., Jonker, E., & Malin, B. (2020). A systematic review on data anonymization and pseudonymization techniques in the healthcare sector. *Journal of the American Medical Informatics Association*, *27*(3), 447–456.

Erlingsson, 'Ulfar, Feldman, V., Mironov, I., Raghunathan, A., Talwar, K., & Thakurta, A. G. (2019). Amplification by shuffling: From local to central differential privacy via anonymity. *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2468–2479.

Gupta, P., Verma, A., & Bhushan, B. (2019). Privacy-preserving data storage and sharing techniques for cloud computing. *Journal of Information Security and Applications*, *47*, 284–295.

Jang, Y., Lee, S., & Kim, S. (2021). A systematic review on security policy enforcement using blockchain. *IEEE Access*, *9*, 24015–24029.

Johnson, S., Kumar, R., & Wang, W. (2019). Secure HMM-based speech recognition using homomorphic encryption. *IEEE Transactions on Information Forensics and Security*, *14*(2), 213–225.

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). A survey of intrusion detection systems, and intrusion prevention systems. *IEEE Communications Surveys & Tutorials*, *21*(4), 407–438.

Li, N., Li, T., & Venkatasubramanian, S. (2021). Privacy-preserving data publishing: A survey on recent developments. *ACM Computing Surveys (CSUR)*, *53*(4), 1–36.

Liu, W., Chen, X., & Zhou, L. (2019). Secure and scalable HMM-based mobile health monitoring using cryptographic techniques. *IEEE Transactions on Mobile Computing*, *18*(5), 1165–1178.

Pashchenko, I., Gurses, S., & Joosen, W. (2020). Security and privacy in programming languages. *ACM Computing Surveys*, *53*(1), 1–30.

Patki, N., Wedge, R., & Veeramachaneni, K. (2016). The synthetic data vault. *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 399–410.

Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2019). Privacy risks of securing machine learning models against adversarial examples. *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, 241–253.

Tikkinen-Piri, C., Rohunen, A., & Markkula, J. (2018). The EU General Data Protection Regulation: Toward a holistic approach for the protection of privacy in the age of analytics. *Journal of Information Privacy and Security*, *14*(1), 10–31.

Townend, D., & others. (2018). The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era. *Journal of Law, Medicine & Ethics*.

Voigt, P., & von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR): A practical guide. *Springer International Publishing*.

Wang, S., Liu, X., Xu, C., Chang, Y., & Ma, Z. (2019). Automated test generation for deep learning systems. *Proceedings of the 41st International Conference on Software Engineering*, 913–923.

Wang, X., Takaki, S., & Yamagishi, J. (2019). Deep variational HMM-based speech synthesis. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *27*(1), 148–158.

Yeom, S., Fredrikson, M., & Jha, S. (2018). Privacy-preserving data analysis against inference attacks. *Proceedings of the 2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, 375–390.

Yoon, J., Jarrett, D., & Van Durme, B. (2020). Anonymization through data synthesis using generative adversarial networks (ADS-GAN). *IEEE Journal of Biomedical and Health Informatics*, *24*(8), 2378–2388.

Yu, T., Zhang, P., Qin, Z., Zhou, C., & Su, L. (2019). Hidden Markov model-based traffic anomaly detection in SDN. *Journal of Network and Computer Applications*, *136*, 86–92.

Zhang, J., Huang, D., Xiao, S., & Zhang, X. (2019). Secure and efficient policy update framework for software-defined networks. *IEEE Journal on Selected Areas in Communications*, *37*(3), 616–628.

Zhang, Y., Li, X., Sun, W., & Chen, L. (2020). Efficient and secure data exchange protocol for healthcare systems. *IEEE Transactions on Industrial Informatics*, *16*(9), 6105–6114.