10TH URV DOCTORAL WORKSHOP IN COMPUTER SCIENCE AND MATHEMATICS

Edited by
Jordi Pascual and Oriol Farràs

UNIVERSITAT ROVIRA i VIRGILI

Title: 10th URV Doctoral Workshop in Computer Science and Mathematics
Editors: Jordi Pascual and Oriol Farràs
March 2025

Universitat Rovira i Virgili
C/ de l'Escorxador, s/n
43003 – Tarragona
Catalunya (Spain)

ISBN: 978-84-1365-208-5
DOI: 10.17345/9788413652085

Preface

This book of proceedings gathers the contributions presented in the 10th Doctoral Workshop in Computer Science and Mathematics - DCSM 2025. It was celebrated in Universitat Rovira i Virgili (URV), Campus Sescelades, Tarragona, on April 3rd, 2025. The aim of this workshop is to promote the dissemination of ideas, methods, and results developed by the students of the PhD program in Computer Science and Mathematics from URV. It has been jointly organized by the research group ITAKA (Intelligent Technologies for Advanced Knowledge Acquisition) and the Doctoral Program on Computer Science and Mathematics of Security of URV. The workshop had two invited talks and sixteen oral presentations. The first invited talk was given by Dr. Jordina Torrents, who is the head of the AI and ML division at HP and course instructor at UOC. The talk discusses HP's AI strategy, focusing on AI-powered personal systems. The second talk was given by Dr. Joan Borràs, who is a data scientist at Inditex since 2022 and former data scientist at Eurecat. This talk discussed that a PhD is not only about academic research, it can also be a gateway to impactful work in industry. In this book, the reader will find the contributions of sixteen PhD students that presented their works in the Workshop. Each chapter presents their current research work, the goals and some preliminary results. All contributions present innovative proposals, methods or applications, with the aim of opening new and strategic research lines. The editors and organizers invite you to contact the authors for more detailed explanations and encourage you to send them your suggestions and comments, which will certainly help them in their PhD theses. The members of the organizing committee were Dr. Jordi Pascual, Dr. Oriol Farràs (Coordinator of the PhD program), Dr. Aïda Valls, Dr. Antonio Moreno, Mr. Gerard Pascual and Mrs. Olga Segú. We would like to thank the invited speakers for such intersting talks. Second, we thank all the participants and, especially, the students that presented their work in this DCSM workshop. Finally, we also want to thank Universitat Rovira i Virgili (URV), the Departament d'Enginyeria Informàtica i Matemàtiques (DEIM) and the Escola Tècnica Superior d'Enginyeria (ETSE) for their support.

Dr. Jordi Pascual and Dr. Oriol Farràs (Editors)

Contents

MemberShield: Federated learning with membership privacy

Faisal Ahmed *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
faisal.ahmed@estudiants.urv.cat

1 Introduction

Federated learning (FL) is a distributed machine-learning paradigm where multiple parties can jointly train deep-learning models without outsourcing their (private) data [1]. Although FL is supposed to provide privacy-by-design, recent work has shown that it is still vulnerable to privacy attacks, because deep neural networks are prone to memorizing (sensitive) information from training data [2]. In particular, *dishonest* or *honest-but-curious* server or clients can orchestrate privacy attacks to infer sensitive information from the updates sent by clients during the federated training rounds. The membership inference attack (MIA) is such a privacy attack, where an adversary with white-box or black-box access to the model tries to identify whether a data sample is part of the training dataset.

Defenses have been proposed to prevent MIAs based on various principles, including *regularization*, *adversarial training*, *knowledge distillation*, and *differential privacy*. However, existing defenses have one or more of the following limitations: (i) they are explicitly designed for centralized training and fail to maintain their performance in FL settings, (ii) they require additional public datasets that are not always available in sensitive domains like healthcare, (iii) they offer privacy protection against a specific category of MIAs, and/or (iv) they require training multiple models on non-overlapping subsets of training data, which is not applicable in the FL setting due to the (usually) small size of client data. On top of that, most defenses offer privacy at the expense of utility, and they add significant training overhead. To the best of our knowledge, no defense can offer meaningful privacy protection against all forms of MIAs in FL without compromising model utility and incurring additional training or inference overheads. In fact, achieving membership privacy while maintaining the model accuracy and without imposing training overhead remains an open challenge for FL due to contradicting requirements [3].

* PhD advisor: Josep Domingo-Ferrer and Zouhair Haddi

To tackle this challenge, in this paper we propose *MemberShield*, an FL framework that offers membership privacy by generalizing the local model on the client end. Specifically, our contributions are:

- We propose a defense mechanism to prevent MIAs by generalizing the local model in two steps: (1) regulating prediction confidence by training the model with soft-encoded ground truth labels, and (2) mitigating training sample memorization by implementing early stopping, considering both global and local model perspectives.
- We report empirical analyses to evaluate the privacy and accuracy that our defense offers. In addition, we compare our defense with several state-of-the-art defenses to show its superiority at delivering both practical membership privacy and high model utility with reduced training time.

2 Threat model

In this work, we consider both the server and the clients as being *honest but curious* adversaries who strictly follow the FL protocol but perform MIAs on local (the server is the global adversary) and global models (clients are the local adversaries) at each training round. We consider a black-box attack scenario where the attacker only queries the model and obtains corresponding prediction vectors or labels.

3 Method

This section presents our solution –called MemberShield– for preventing MIAs in FL. Previous studies have shown that the overfitting exploited by MIAs primarily arises from the overconfidence of the model in predicting and memorizing training samples. Whereas overconfidence is a by-product of training a model with one-hot encoded ground truth labels, memorization happens due to excessive training. Our defense overcomes these issues by generalizing the local model in two steps: *(1) regulating the prediction confidence by training the model with soft-encoded ground truth labels*, and *(2) mitigating training sample memorization by early stopping considering global and local model perspectives.*

Regulating the prediction confidence. This step encourages the model to show less confidence in the training sample prediction, which contributes to making confidence, entropy, and loss distributions for member and non-member samples less distinguishable. This step also helps the model to generalize better, by allowing the model to assign some probabilities to incorrect classes [4].

Consider a uniform prior distribution u over levels $c \in \{1, \ldots, C\}$ where $u_c = 1/C$ for each c, independent of the training example x. We replace the one-

hot encoded ground truth distribution y with a new probability distribution y' given by

$$y'_c = (1 - \theta) \cdot y_c + \theta \cdot u_c \tag{1}$$

for each label $c \in \{1, \ldots, C\}$. For the new distribution, $\sum_{c=1}^{C} y'_c = 1$ and $1 > y'_a > y'_c > 0$ for the ground truth label a and all $c \neq a$. The new ground truth distribution replaces the hard 0 and 1 class labels with θ/C and $1 - \frac{C-1}{C} \cdot \theta$, respectively, thereby mirroring the prediction probability distributions for unseen data.

Mitigating training sample memorization. Due to the usual heterogeneous data distribution across the peers in FL settings, some local models may converge earlier than the expected number of federated training rounds. Specifically, a local model stops enhancing its generalization capability when its validation loss fails to exhibit improvement w.r.t. the validation loss of the global model from the preceding local training round. To prevent such unnecessary overtraining, we include an early stopping regularization into the local model training. Our early stopping criterion differs from traditional early stopping, where training stops when a model's training or validation loss does not change significantly for a predefined number of consecutive local epochs. With that approach, the local model does not consider the global perspective for stopping. In our approach, the peer evaluates the model's loss on its internal validation data after receiving the aggregated global model from the server, and starts local training by tracing the validation loss for successive epochs. consecutive epochs.

4 Experiment and Results

We simulated an FL setting for the CIFAR10 classification task with a custom CNN architecture for ten training rounds with five clients under non-iid data distributions. Defenses were evaluated under three perspectives: *Utility*, *Defense effectiveness* and *training overhead*.

Table 1 report the global adversary's maximum advantage against each client local model, as well as the global model's training accuracy, test accuracy, and training time for vanilla FL (without any defense mechanism), MemberShield, and six defenses in the literature.

5 Conclusions

We have proposed a privacy-preserving FL framework, *MemberShield*, which safeguards MIAs by generalizing the local models. Extensive performance analysis shows that *MemberShield offers much better practical membership privacy than previous defenses except DP, and better or similar global model accuracy than all previous defenses (similar to vanilla FL) and much better*

Table 1: For vanilla FL, six baseline defenses, and MemberShield : the global adversary's auR and maximum advantage (Adv) against each client local model; the global model's training accuracy (Tr Ac), test accuracy (Te Ac), and training time.

Method	Parameters		Client-1		Client-2		Client-3		Client-4		Client-5		Global Model		
	Dr/Ep	Reg	auR ↓	Adv ↓	auR ↓	Adv ↓	auR ↓	Adv ↓	auR ↓	Adv ↓	auR ↓	Adv ↓	Tr Ac ↑	Te Ac ↑	Time ↓
Vanilla FL (No Defense)	-	-	0.84	0.71	0.81	0.58	0.83	0.63	0.84	0.64	0.86	0.69	0.89	0.64	1973.62
Early Stopping	-	-	0.61	0.19	0.60	0.17	0.58	0.13	0.58	0.14	0.59	0.18	0.69	0.61	1215.48
AoF	0.25	-	0.88	0.74	0.82	0.56	0.86	0.62	0.87	0.67	0.91	0.82	0.89	0.65	2117.64
	0.50	-	0.88	0.73	0.83	0.54	0.86	0.62	0.86	0.63	0.92	0.78	0.89	0.65	2080.81
	0.75	-	0.89	0.68	0.82	0.50	0.84	0.54	0.85	0.59	0.92	0.75	0.87	0.65	2015.21
	-	L2	0.87	0.66	0.71	0.33	0.81	0.52	0.83	0.59	0.90	0.77	0.66	0.61	2021.10
	0.25	L2	0.85	0.60	0.67	0.26	0.78	0.43	0.80	0.49	0.88	0.71	0.72	0.66	2058.35
	0.50	L2	0.83	0.56	0.67	0.25	0.75	0.41	0.78	0.45	0.87	0.65	0.75	0.68	2083.54
	0.75	L2	0.81	0.52	0.60	0.16	0.70	0.32	0.73	0.38	0.86	0.64	0.70	0.65	2058.80
DP	0.1	-	0.52	0.09	0.51	0.02	0.52	0.04	0.52	0.04	0.54	0.14	0.09	0.09	3068.95
	0.5	-	0.53	0.06	0.51	0.03	0.52	0.04	0.52	0.04	0.54	0.10	0.12	0.12	3066.95
	1	-	0.53	0.08	0.51	0.02	0.52	0.03	0.52	0.04	0.57	0.14	0.11	0.11	3044.48
	2	-	0.53	0.07	0.52	0.03	0.52	0.04	0.52	0.06	0.56	0.15	0.15	0.15	3022.81
	4	-	0.52	0.06	0.51	0.03	0.52	0.04	0.53	0.05	0.56	0.14	0.19	0.19	3028.14
	8	-	0.52	0.10	0.51	0.03	0.52	0.05	0.51	0.03	0.55	0.11	0.19	0.19	3019.60
	16	-	0.53	0.07	0.51	0.03	0.52	0.05	0.52	0.06	0.53	0.16	0.20	0.20	3014.77
	100	-	0.53	0.07	0.51	0.03	0.51	0.05	0.53	0.05	0.53	0.11	0.23	0.23	3029.29
	1,000	-	0.53	0.07	0.51	0.02	0.52	0.05	0.52	0.07	0.52	0.10	0.26	0.26	3024.98
KD ($\alpha = 0.5, T = 10$)	-	-	0.85	0.59	0.75	0.39	0.81	0.50	0.86	0.62	0.84	0.55	0.48	0.38	4757.52
RelaxLoss ($\alpha = 0.4$)	-	-	0.80	0.53	0.79	0.49	0.80	0.50	0.81	0.54	0.81	0.50	0.85	0.56	1918.79
Adv-Reg ($\lambda = 3$)	-	-	0.70	0.33	0.62	0.18	0.65	0.23	0.62	0.20	0.74	0.39	0.18	0.17	3595.63
MemberShield (Our Defense)	-	-	0.56	0.11	0.60	0.16	0.58	0.13	0.56	0.11	0.57	0.14	0.72	0.66	1780.29

than DP, while reducing the computational cost of model training with respect to all previous defenses.

Acknowledgement. This study was funded by the Marie Skłodowska-Curie Doctoral Network Actions (Horizon MSCA-2021-DN-01-01) under Grant No. 101073222.

References

[1] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282, 2017. PMLR.

[2] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, pages 1322–1333, 2015.

[3] Najeeb Moharram Jebreel, Josep Domingo-Ferrer, Alberto Blanco-Justicia, and David Sánchez. Enhanced security and privacy via fragmented federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. IEEE.

[4] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

The Role of Lesion Counting in Optimizing Diabetic Retinopathy Diagnosis

Amal Esmail Alzubairi *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
amalesmailqasem.al-zubairi@urv.cat

1 Abstract

Diabetic Retinopathy (DR) is a major cause of vision loss in individuals with Diabetes. Detecting the disease early and understanding its stages of development are crucial for preventing serious complications. This research builds upon the ITAKA group's prior work by combining two existing software tools, Retiprogram and LezioSeg, to enhance DR diagnosis and prognosis. LezioSeg is a deep learning-based computer vision model that segments DR lesions and identifies their locations in different regions of the retina. The current study aims to incorporate the number of lesions per region into Retiprogram, a tool that classifies DR using fuzzy random forests. Specifically, this work focuses on two primary types of lesions: microaneurysms and hemorrhages. We collected data from Sant Joan Hospital in Reus and used it for training and testing. This research aims to improve DR diagnostic systems, making it easier to provide personalized risk assessments, timely treatments, and better outcomes for patients.

2 Introduction

Diabetic Retinopathy (DR) is a progressive eye disease caused by prolonged high blood sugar, affecting the retinal blood vessels and potentially leading to blindness. It produces lesions such as cotton wool spots, exudates, microaneurysms (MA), and hemorrhages (HM). MA is an early indicator of DR, while HM results from blood vessel leakage.
DR progresses from mild nonproliferative DR (NPDR), marked by MA, to moderate and severe NPDR with HM and venous abnormalities. The most severe stage, Proliferative DR (PDR), involves abnormal blood vessel growth, risking retinal detachment and vision loss. Our study classifies Diabetic Retinopathy (DR) severity, a multiclass problem with stages from no

* PhD advisor: Aida Valls

DR to severe, using key indicators MA and HM. This classification is crucial for timely treatment. Machine learning algorithms, like Decision Trees(DT), aid in this by analyzing retinal features to identify patterns of DR severity. Our research group developed Retiprogram [1], a Fuzzy Random Forest (FRF) model for Diabetic Retinopathy (DR) classification. Retiprogram currently uses clinical and analytical risk factors like age, gender, and hemoglobin, achieving approximately 80% accuracy. This thesis aims to improve Retiprogram by incorporating MA and HM lesion counts, obtained using LezioSeg [2]. LezioSeg is a deep learning-based computer vision tool that segments, localizes, and counts these lesions in four regions of the eye fundus.

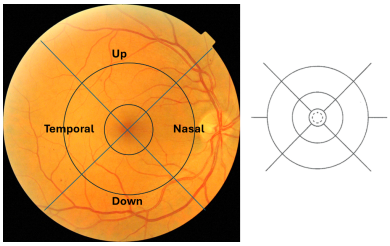

Fig. 1: The 4 regions of the eye fundus image defined by ETDRS.

The FRF uses fuzzy logic to handle uncertainty in the data. Instead of making strict yes-or-no decisions, it uses flexible boundaries, which helps it work better with complicated or unclear data. The inputs are fuzzy linguistic variables, which consist of a set of labels:Ł= {l_1, l_2, . . . , l_r}, each one mapped into a numerical reference scale, S, by means of a fuzzy membership function:$\mu\_l\_i(x) \in [0, 1], \quad x \in S$. Fuzzy linguistic variables effectively convert numerical values into descriptive terms like 'low," moderate,' or 'high,' simplifying information for clinicians and researchers. The goal of this doctoral thesis is to develop a methodology for automatically constructing fuzzy linguistic variables for lesion counts in each region based on data distributions.

3 Methodology

3.1 Dataset Collection

We collected eye fundus images from Sant Joan Hospital in Reus to study and measure retinal lesions for diagnosing and analyzing DR. Using LezioSeg,counts were recorded in four major directions around the center of the retina:up, down, temporal, and nasal to capture how lesions are distributed a cross different regions of the retina. 1 shows the number of images per class in the training and testing datasets. For each image in the training and testing sets, we used LezioSeg to obtain the following 10 variables:

Class	Train	Test
0 (No DR)	1397	598
1 (Mild)	500	214
2 (Moderate)	903	387
3 (Severe)	1016	435

Table 1: Number of images per class in the training and testing datasets.

- Total_MA: total no. of MA in the eye.
- Total_HM: total no. of HM in the eye.
- temporal_MA, nasal_MA, up_MA, down_MA: no. of MA in each region.
- temporal_HM, nasal_HM, up_HM, down_HM: no. of HM in each region.

3.2 Handling Imbalanced Data:

We used SMOTE (Synthetic Minority Oversampling Technique), which generates synthetic samples for minority classes to balance the dataset.

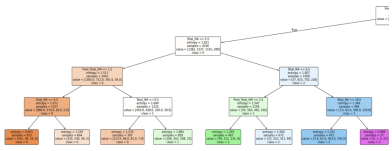

Fig. 2: Decision tree model for classifying DR based on extracted features.

3.3 Decision Tree-Based Classification:

The DT algorithm is applied to analyze lesion counts and identify the best values for classifying the DR degrees. This algorithm selects the best attributes at each node and identifies the optimal cut points for classification, Fig. 2.

3.4 Defining Linguistic Variables:

We propose to use these cut points as reference thresholds that divide the range of each attribute into meaningful categories to distinguish the different levels of lesion severity, Fig 2. These points define linguistic variables by applying fuzzy membership functions, with input values as integers, not real numbers. To represent these intervals, we build triangular fuzzy sets defined by the tuple (a, b, c). We consider the center of the intervals defined by the cut points either as the value with maximal membership, b, or as one of the extremes a/c depending on the length of the interval.For example, looking at the tree obtained in Fig 2, we can see some cut points of the Total_MA variable. Considering a range of $[0, 10]$ and cut points: $[0, 0.5, 2.5, 3.5, 6.5]$, we generate the following labels:

Name	Type	Parameters
none	Triangular	[0, 0, 1.5]
scarce	Triangular	[0, 1.5, 3]
few	Triangular	[1.5, 3, 5]
mid	Triangular	[3, 5, 8]
high	Trapezoidal	[5, 8, 10, 10]

Table 2: Fuzzy Membership Functions

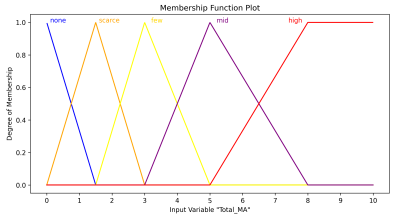

Fig. 3: Membership function plot for the input variable Total_MA .

4 Conclusion and Future Work

Our methodology constructs fuzzy linguistic variables for DR lesion counts using cut points from the DT model. These cut points are transformed into descriptive categories, improving data interpretability and providing meaningful inputs for the RF model. Using fuzzy linguistic variables helps capture the uncertainty in lesion counts, crucial for accurate DR severity assessment. We plan to combine these fuzzy variables with other clinical risk factors in the Retiprogram software to evaluate the overall impact on DR classification performance.

Acknowledgement. This work was supported by the Programa Martí Franquès under the grant 2023PMF-PIPF-19

References

[1] Jordi Pascual-Fontanilles. Fuzzy-based machine learning methods for continuous diagnosis and prognosis of Diabetic Retinopathy. PhD thesis, University of Rovira i Virgili, Spain, 2024. http://hdl.handle.net/10803/690588.

[2] Mohammed Yousef Salem Ali. Deep Learning-based Methods for Extracting Fundus Image Landmarks and Signs of Eye Diseases. PhD thesis, University of Rovira i Virgili, Spain, 2023. http://hdl.handle.net/10803/687502.

Capture The Flag applied to security research

Miquel Calonge *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
miquel.calonge@estudiants.urv.cat

1 Motivation

In addition to being a researcher in security topics, I am also an active participant in Capture The Flag (CTF) competitions. A CTF is a cybersecurity challenge designed to test and enhance technical skills in different areas such as cryptography, reverse engineering, and web security. These challenges simulate real-world security vulnerabilities and require analytical thinking, problem solving, and creativity to identify and exploit systems weaknesses [1].
Since the workshop for which this abstract is intended is designed to be accessible to students with no prior expertise in the field, I see it as an excellent opportunity to introduce those unfamiliar with CTFs to this concept, sparking their curiosity, just as it did for me. Engaging with the CTF community not only strengthened my technical skills, but also boosted my confidence, ultimately encouraging me to take the leap into pursuing a PhD.
In this abstract, I have chosen to focus on how CTF challenges have unexpectedly contributed to my research. To illustrate this, I have organized it into three sections: encoding standards consideration, custom wordlists creation, and web archiving services use, each covering a specific instance where CTFs have influenced my daily work as a researcher and enhanced my methodologies.

2 Encoding standards consideration

One of the most popular categories in CTFs is OSINT, which stands for Open Source Intelligence. This category involves gathering publicly available information from various sources on the internet, such as social media, websites, and public databases. While solving an OSINT challenge, the system did not accept the flag, although I was confident that my answer was correct. It was

* PhD advisors: Agusti Solanas & Julio Hernández-Castro

the name of a city in Iceland, which contained characters not present in the Spanish alphabet.
The issue was that the hashing algorithm used, MD5 in this case, could produce different hashes for the same string depending on the encoding specified for the input. This difference is not noticeable with strings containing only common characters, as they tend to be hashed the same way across all encoding standards 2.

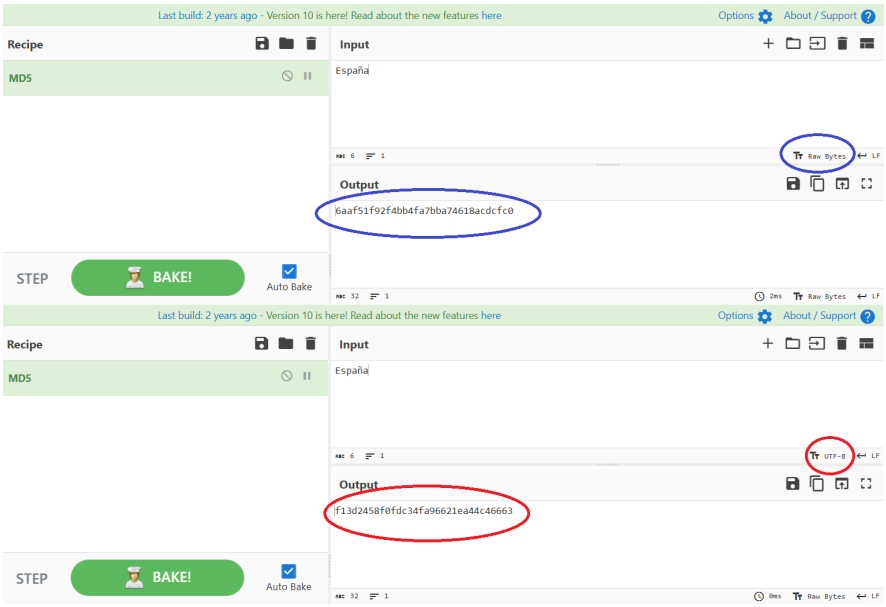

Fig. 1: Comparison of MD5 hash results using different encodings in CyberChef.

This realization led me pay closer attention to an aspect I had previously overlooked, which ultimately had a significant impact on my research. As part of my thesis, I conduct cracking attempts to access Bitcoin wallets someone used in the past. One type of them are *Brainwallets*, where the owner had to choose a password from which the wallet is generated. However, just like in the CTF challenge, the encoding of the chosen password can result in the generation of different wallets.
By applying this concept, I started processing candidate passwords using multiple encodings, which led to new matches that I had previously missed. This approach provided a considerable enhancement to my research, adding an extra layer of depth and effectiveness to my findings.

3 Custom wordlists creation

I actively participate in CTFs on the **Atenea** platform, which is managed by the Spanish **Centro Criptológico Nacional (CCN)**. One of the most intriguing challenge categories is called *Acknowledgments*, reserved for users who achieve something unexpected: solve challenges in unconventional ways or discover vulnerabilities on the platform. Once I had already started my PhD and was immersed in cracking techniques daily, I decided to approach this challenge category using what was becoming my area of expertise.
I conducted a study on how flags were structured to predict the format of new flags. I identified interesting patterns, with the most striking one being that organisers often set the flags using data from Wikipedia articles related to the theme of the challenge. By narrowing down my list of possible candidates using this and other patterns, I successfully solved a challenge through a targeted flag attack. This approach proved successful and as a result, the competition organizers awarded me a flag in this category.
I have also applied this strategy to my research. I generate wordlists tailored specifically for cracking wallets, incorporating elements relevant to the cryptocurrency ecosystem. Using easily reproducible private keys to generate new private keys has led me to some remarkable results, including gaining access to wallets that contained Bitcoin worth several hundred thousand euros (valued at the time the wallets held the Bitcoin, not based on Bitcoin's later price increases).

4 Web archiving services use

In OSINT challenges, it is common to use **Wayback Machine** or similar web archiving services, which I was previously completely unaware of. These tools allow viewing content that is no longer available on the original hosting websites.
This has proven to be extremely useful in my research when trying to reproduce vulnerabilities related to the creation of Bitcoin private keys through web-based JavaScript implementations. No matter how well you understand the theoretical process, without access to the original implementation, it is highly likely that you will not achieve the same results.

Acknowledgments

This work has been developed within the "Plan de Recuperación, Transformación y Resiliencia", project C084/23 Ada Byron INCIBE-UAH, funded by the European Union (Next Generation).

References

[1] L. McDaniel, E. Talvi, and B. Hay. Capture the Flag as Cyber Security Introduction. *Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 5479–5486, Koloa, HI, USA, 2016.

Resource Allocation in Serverless Data Analytics: The Insidious Exchange

Germán T. Eizaguirre *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
germantelmo.eizaguirre@urv.cat

1 Serverless data analytics

Serverless services are a practical approach to making the Cloud more accessible, as much of the management burden is abstracted away by the cloud provider. Serverless functions —or Function as a Service (*FaaS*)— such as AWS Lambda are the leading embodiment of the serverless paradigm: stateless, event-driven, ephemeral computations with low start-up latency and broad concurrency. FaaS provides a convenient pay-as-you-go billing model and intuitive, easy-to-use abstractions over the underlying resources. Developers can deploy their code to the Cloud in seconds, running in a secure, isolated and managed environment that auto-scales during peak demand and to zero when there is none.

The potential of serverless services has not gone unnoticed in data analytics; particularly, in elastic multi-stage workloads where the number of parallel tasks changes drastically from one stage to another. Traditional cluster technologies on virtual machines (VMs) often fall short in such scenarios due to expecting a fixed pool of resources or slow (if any) auto-scaling capabilities. These conditions lead to a mismatch between cluster size and resource requirements in elastic multi-stage workloads, resulting in overprovisioning (where more resources are provisioned than needed, leading to inflated costs), or underprovisioning (where fewer resources are provisioned than required, leading to degraded performance). Serverless functions address this issue more efficiently, quickly scaling out and in based on the needs of the workload, and fitting the exact resource requirements of each stage.

However, serverless functions face a major limitation: they are ephemeral, making them non-addressable and stateless, thus dependent on external disaggregated storage for communication. Functions are not suitable for complex communication patterns like sorting, grouping, or re-partitioning, which require extensive data sharing between workflow stages. These *stateful opera-*

* PhD advisor: Marc Sánchez-Artigas

tions are a significant performance challenge for serverless data analytics due to the load placed on the storage. Overloading the storage slows down I/O throughput, reduces CPU utilisation, and hurts cost-efficiency. As the number of functions per stage increases, so does the number of storage requests, gradually degrading I/O performance and exacerbating resource waste.

2 Stateful backends for data exchanges

With sufficient abstraction, serverful (VMs) and serverless (functions) resources can be seamlessly combined within a complex workflow. By deploying servers in a *serverless* flavour, i.e., launching select workflow stages on a short-lived, right-sized VM, we mitigate the limitations of functions to achieve fast state sharing in stateful phases with minimal developer intervention [1]. The overhead of creating VMs can be acceptable to launch stateful stages, as the cost of (indirect) communication in serverless functions is frequently prohibitive.

We bridge the gap between serverless and serverful technologies by transparently integrating VMs into serverless analytics. We do this by extending Lithops, a serverless programming library, with the ability to leverage hybrid architectures. With minimal in-code changes, a developer can choose to seamlessly run their parallel code on either serverless functions or VMs. Embarrassingly parallel stages, such as data processing, benefit most from *FaaS*, while serverful services are better suited to stateful stages with strong dependencies. Exploiting Lithops's unified programming framework, we develop a hybrid serverless architecture selectively choosing the right service for each stage. We use AWS Lambda for embarrassingly parallel stages, while hosting stateful operations in properly scaled EC2 instances.

The end-to-end execution times of a cloud function-based implementation, a static Spark deployment and the proposed hybrid architecture are listed in Table 1. Results correspond to a preprocessing stage in a metabolomics pipeline. When running the annotation pipeline in a hybrid architecture, we achieve a speedup of 3.64 between and 2.21 compared to Spark. Overall, by coherently alternating cloud functions and VMs, we achieve up to 75% better performance at a similar cost to a cluster implementation through a smarter use of resources.

3 Optimizing serverless exchanges

Even if using cloud VMs may result practical, provisioning VMs is a costly operation and can take minutes depending on the instance type. Moreover, VM scaling tends to act at a coarse granularity, which may lead to either system degradation or resource underutilization in the absence of prior knowledge

Table 1: Execution time of each metabolomics job executed in the studied architectures.

Job size	**Cloud functions**	**Hybrid**	**Spark**
small	152.20s	105.49s	54.83s
medium	351.57s	398.70s	889.54s
large	488.86s	709.14s	2582.66s

We argue that a fully serverless implementation of data analytics pipelines may be feasible, avoiding the use of serverful components. We propose tuning shuffles at runtime in order to maximize the I/O efficiency of object storage. For that, we take profit of fine-grained scaling of compute resources, finding a way to determine at runtime the optimal number of functions (degree of parallelism) that better utilizes remote object storage.

Table 2: Comparison of Seer to state-of-the-art data analytics systems for 100GB TeraSort.

System	**# workers**	**Storage layer**	**Exec. time**
Qubole [2]	400	AWS S3	597.7s
Locus [3]	dynamic	AWS S3/Redis	80s to 140s
Primula [4]	200	IBM COS	192.3s
Caerus [5]	100	Jiffy [6] (VMs)	105s
Serverful Spark	100	HDFS	600.12s
	160	HDFS	493.94s
Seer	100	IBM COS	95.06s
	225	IBM COS	89.96s
	256	IBM COS	90.12s

We propose a smart shuffle auto-tuner called Seer [7]. More precisely, Seer chooses "on the fly" the optimal number of parallel workers in a shuffle to deliver improved efficiency, based on analytical models. The only input these models need from a job is the volume of data to be shuffled at the current stage, which can be dynamically inferred as the sum of the individual data partitions. Simply put, Seer does not require jobs to know (even an estimate of) intermediate data sizes a priori. When a shuffle operation is encountered, Seer dynamically determines the right shuffle implementation, along with the optimal degree of parallelism that minimizes shuffle time. The performance models require the characterization of the remote object storage service via a series of basic measurements.

Table 2 lists the execution time of state-of-the-art serverless data exchange approaches in a classic TeraSort benchmark. In all cases, Seer outperforms the existing alternatives, with the exception of Locus for only 9 seconds, which is surprising, since Locus has been "bolstered up" with in-memory Redis instances. Seer is able to improve execution time by 1.1 - 6.3X for the same number of vCPUs, which contributes to confirm that object storage is a practical solution to serverless shuffling.

Acknowledgement. Germán T. Eizaguirre is recipient of a pre-doctoral FPU grant from the Spanish Ministry of Universities (ref. FPU21/00630).

References

[1] Eizaguirre, Germán T., Barcelona-Pons, Daniel, Arjona, Aitor, Vernik, Gil, García-López, Pedro, and Alexandrov, Theodore. Serverful Functions: Leveraging Servers in Complex Serverless Workflows (industry track). In *Proceedings of the 25th International Middleware Conference Industrial Track*, pages 15–21, Hong Kong, Hong Kong, 2024. Association for Computing Machinery. ISBN 9798400713194. DOI: `https://doi.org/10.1145/3700824.3701095`.

[2] Qubole. *Spark on Lambda*. Available at: `https://www.qubole.com/blog/spark-on-aws-lambda`, accessed on February 10, 2025, 2017.

[3] Qifan Pu, Shivaram Venkataraman, and Ion Stoica. Shuffling, fast and slow: scalable analytics on serverless infrastructure. In *Proceedings of the 16th USENIX Conference on Networked Systems Design and Implementation (NSDI'19)*, pages 193–206, Boston, MA, USA, 2019. USENIX Association.

[4] Marc Sánchez-Artigas, Germán T. Eizaguirre, Gil Vernik, Lachlan Stuart, and Pedro García-López. Primula: a practical shuffle/sort operator for serverless computing. In *Proceedings of the 21st International Middleware Conference Industrial Track (Middleware '20)*, pages 31–37, Delft, Netherlands, 2020. Association for Computing Machinery. https://doi.org/10.1145/3429357.3430522.

[5] Hong Zhang, Yupeng Tang, Anurag Khandelwal, Jingrong Chen, and Ion Stoica. Caerus: NIMBLE Task Scheduling for Serverless Analytics. In *18th USENIX Symposium on Networked Systems Design and Implementation (NSDI 21)*, pages 653–669, Virtual Event, USA, 2021. USENIX Association.

[6] Anurag Khandelwal, Yupeng Tang, Rachit Agarwal, Aditya Akella, and Ion Stoica. Jiffy: elastic far-memory for stateful serverless analytics. In *Proceedings of the Seventeenth European Conference on Computer Systems (EuroSys '22)*, pages 697–713, Rennes, France, 2022. Association for Computing Machinery. https://doi.org/10.1145/3492321.3527539.

[7] Marc Sánchez-Artigas and Germán T. Eizaguirre. A seer knows best: optimized object storage shuffling for serverless analytics. In *Proceedings of the 23rd ACM/IFIP International Middleware Conference (Middleware '22)*, pages 148–160, Quebec, QC, Canada, 2022. Association for Computing Machinery. https://doi.org/10.1145/3528535.3565241.

Smart Contracts for a Secure and Privacy-Preserving Smart Grid

Joan Ferré-Queralt *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
joan.ferreq@urv.cat

Introduction

To guarantee a nation's energy needs, establishing an appropriate balance between supply and demand is essential. Historically, power generation has depended on building large-scale plants that run on fossil fuels (such as coal or natural gas) or nuclear energy. These conventional methods, however, bring with them significant drawbacks: high construction and maintenance costs, greenhouse gas emissions, pollution, and issues with nuclear waste management. For these reasons, shifting toward alternative primary energy sources appears promising.

In recent years, renewable energy sources—like wind and solar—have become much more cost-effective. This cost reduction has increased their competitiveness and has helped steer regions such as the European Union toward a cleaner, decarbonized power sector. Still, the rate of renewable adoption remains slower than expected[2]. One contributing factor is that renewables, in some cases, do not yet match the reliability, availability, or efficiency of their conventional counterparts.

Another promising strategy is to locate energy production closer to the end user. Reducing the distance between production and consumption minimizes energy losses and transmission costs. In addition, using distributed renewable technologies—such as solar panels—creates the opportunity for localized, decentralized production and consumption of electricity [1]. Many households already incorporate distributed energy resource (DER) systems to generate and self-consume electricity, and when these systems produce a surplus, homeowners can feed the extra power back into the main grid. Unfortunately, the prices offered for this surplus are often unattractive due to imposed tariffs or taxes.

* PhD advisor: Dr. Jordi Castellà-Rova, Dr. Alexandre Viejo

[2] IEA: Wholesale energy costs made simple, https://www.edfenergy.com

Related Work

A growing body of research has examined various smart grid implementations, with each study focusing on different aspects of the new energy paradigm. In this work, we concentrate on those solutions that enable the trading of electricity among prosumers.

The very first formal description of a smart grid appeared in [2], where a set of initial requirements for this innovative energy model was proposed. Later, in [3], several communication standards for smart grids were discussed—most of which assumed a traditional centralized network structure. Unfortunately, centralized systems are prone to security issues, such as Denial of Service attacks [4], and they suffer from the well-known single point of failure problem. These concerns can be alleviated by moving to a distributed framework [5].

An early distributed solution was offered in [6]. In that work, the authors introduced an energy-currency concept that builds on the blockchain technology originally described in [7]. This proposal laid the groundwork for integrating blockchain and smart contracts into the design of future smart grids. Following this idea, a number of studies [8,9,10,11,6] have explored diverse blockchain architectures to support smart grid applications.

It is important to note that most of these approaches rely on blockchains using a Proof-of-Work consensus mechanism. Because Proof-of-Work requires extensive computational power to validate transactions, it is highly energy-intensive [12]—a serious drawback when considering a real-world smart grid. For this reason, there is a pressing need to develop alternative consensus methods that reduce energy consumption while still ensuring the system's integrity and security.

Other works [13,10,18] have integrated blockchain technology into energy systems, yet they often pay little attention to protecting user identities and consumption patterns. This oversight jeopardizes user privacy, leaving sensitive data vulnerable to unauthorized access or profiling [14]. More recent proposals [15,16,17] have attempted to address privacy concerns. Although some of these approaches enhance anonymity, they often do not verify whether the energy production or consumption reported by users is genuine. As a consequence, dishonest users might submit falsified records, thereby misleading the system. Even when a reputation system is introduced—as in [17]—the lack of linkage between smart grid accounts and real-world identities prevents effective enforcement against misbehavior.

Contributions

This paper presents a novel energy trading system that builds on a smart grid infrastructure to empower users to both consume and produce electricity through distributed energy resources. Our proposed system improves upon earlier models by emphasizing user privacy, system availability, and security—all while incorporating measures to detect and penalize misbehaving participants. These advances are achieved through the integration of blockchain technology and smart contracts.
Our smart grid framework continuously monitors the energy usage of each participant, which in turn allows for accurate calculations of individual prosumers' financial positions. In addition, the system supports smooth electricity trading among nearby prosumers and offers the flexibility to either borrow from or contribute surplus energy to the main grid as needed. In summary, by combining modern blockchain solutions with smart grid technology, our approach not only fosters decentralized energy production but also enhances overall system resilience and user protection against fraudulent behavior.
In conclusion, the integration of these technologies paves the way for a more sustainable, efficient, and secure energy future.

Acknowledgement. This research is supported by Project PID2021-125962OB-C32 "SECURING/DATA' funded by MCIN/AEI/ 10.13039/501100011033 and by "ERDF A way of making Europe"; by Project HERMES funded by the European Union NextGenerationEU/PRTR via INCIBE; and by Grant SGR2021-00115 funded by AGAUR, Generalitat de Catalunya.

References

[1] Madison K. Hoffacker and Rebecca R. Hernandez. Local Energy: Spatial Proximity of Energy Providers to Their Power Resources. *Frontiers in Sustainability*, 1, 2020. ISSN 2673-4524.

[2] US Congress. Energy independence and security act of 2007. *Public law*, 2:110–140, 2007.

[3] K.S. Reddy, Madhusudan Kumar, T.K. Mallick, H. Sharon, and S. Lokeswaran. A review of Integration, Control, Communication and Metering (ICCM) of renewable energy based smart grid. *Renewable and Sustainable Energy Reviews*, 38:180–192, 2014.

[4] Kishan Bhat, Vikram Sundarraj, Shravani Sinha, and Ankur Kaul. IEEE Cyber Security for the Smart Grid. *IEEE Cyber Security for the Smart Grid*, 2013, pages 1–122.

[5] Muhammad Baqer Mollah, Jun Zhao, Dusit Niyato, Kwok-Yan Lam, Xin Zhang, Amer M. Y. M. Ghias, Leong Hai Koh, and Lei Yang. Blockchain for Future Smart Grid: A Comprehensive Survey. *IEEE Internet of Things Journal*, 8(1):18–43, 2021.

[6] Mihail Mihaylov, Sergio Jurado, Narcís Avellana, Kristof Van Moffaert, Ildefons Magrans de Abril, and Ann Nowé. NRGcoin: Virtual currency for trading of renewable energy in smart grids. In *Proceedings of the 11th International Conference on the European Energy Market (EEM14)*, pages 1–6, 2014.

[7] Satoshi Nakamoto. Bitcoin whitepaper. *URL: `https://bitcoin.org/bitcoin.pdf` (Accessed: 17.07.2019)*, 2008.

[8] Anak Agung Gde Agung and Rini Handayani. Blockchain for smart grid. *Journal of King Saud University - Computer and Information Sciences*, 34(3):666–675, 2022. ISSN 1319-1578.

[9] Nurzhan Zhumabekuly Aitzhan and Davor Svetinovic. Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams. *IEEE Transactions on Dependable and Secure Computing*, 15(5):840–852, 2018.

[10] Jiawen Kang, Rong Yu, Xumin Huang, Sabita Maharjan, Yan Zhang, and Ekram Hossain. Enabling localized peer-to-peer electricity trading among plug-in hybrid electric vehicles using consortium blockchains. *IEEE Transactions on Industrial Informatics*, 13(6):3154–3164, 2017.

[11] Esther Mengelkamp, Benedikt Notheisen, Carolin Beer, David Dauer, and Christof Weinhardt. A blockchain-based smart grid: towards sustainable local energy markets. *Computer Science-Research and Development*, 33:207–214, 2018. Springer.

[12] Alex de Vries. Bitcoin's energy consumption is underestimated: A market dynamics approach. *Energy Research & Social Science*, 70:101721, 2020.

[13] Jianchao Hou, Haicheng Wang, and Pingkuo Liu. Applying the blockchain technology to promote the development of distributed photovoltaic in China. *International Journal of Energy Research*, 42(6):2050–2069, 2018.

[14] Peng Zhuang, Talha Zamir, and Hao Liang. Blockchain for cybersecurity in smart grid: A comprehensive survey. *IEEE Transactions on Industrial Informatics*, 17(1):3–19, 2021.

[15] Rabiya Khalid, Nadeem Javaid, Ahmad Almogren, Muhammad Umar Javed, Sakeena Javaid, and Mansour Zuair. A blockchain-based load balancing in decentralized hybrid P2P energy trading market in smart grid. *IEEE Access*, 8:47047–47062, 2020.

[16] Zhitao Guan, Guanlin Si, Xiaosong Zhang, Longfei Wu, Nadra Guizani, Xiaojiang Du, and Yinglong Ma. Privacy-preserving and efficient aggregation based on blockchain for power grid communications in smart communities. *IEEE Communications Magazine*, 56(7):82–88, 2018.

[17] Omaji Samuel and Nadeem Javaid. GarliChain: A privacy preserving system for smart grid consumers using blockchain. *International Journal of Energy Research*, 46(15):21643–21659, 2022. Wiley Online Library.

[18] Omaji Samuel and Nadeem Javaid. A secure blockchain-based demurrage mechanism for energy trading in smart communities. *International Journal of Energy Research*, 45(1):297–315, 2021.

Towards a Surgical Semantic Video Search

Gerard Finol Peñalver *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
gerard.finol@urv.cat

1 Introduction, motivation and challenges

Event streaming systems, such as Apache Kafka [1], AWS Kinesis [2], or Apache Pulsar [4], are enabling organizations of all kinds to ingest, store, and process data in real-time with high performance and scalability. While streaming systems are often associated with managing event-like data types (*e.g.*, logs, sensors), there is an increasing interest in using these systems for managing multimedia. For instance, AWS Kinesis can ingest video streams and serve them in real-time to analytics applications [3].
In collaboration with the National Center for Tumor Diseases (NCT) in Germany, we are exploring a computer-assisted surgery use case. The NCT is an institution that combines data scientists and surgeons to apply AI techniques to surgery-related multimedia. Specifically, NCT requires the ability to search for specific video fragments, or even individual frames, within a collection of surgical video streams. This is essential for various reasons, such as creating specialized AI training datasets and aiding surgeons or medical students in locating specific video fragments. Metadata alone is insufficient for content-based queries in large video collections. Additionally, data scientists may need to find videos using unstructured data, like images. Therefore, a content-based approach for indexing, querying, and retrieving relevant video stream data is necessary.
Our main goal is to devise a flexible semantic video search solution for streaming systems. Achieving this goal could provide added value to data streaming platforms supporting video stream ingestion and analytics [2]. For example, users could index video streams based on their own models and get accurate query results in the form of video fragments. Even more, data loaders in AI inference frameworks could exploit such a mechanism by ingesting only relevant video fragments for training a model, discarding the rest.
This goals entails some challenges:

* PhD advisors: Pedro García-López and Raúl Gracia-Tinedo

(C1) Flexible content-based video stream indexing: Although a video stream is immutable, its contents can be indexed in multiple ways. For example, one user may index a video for liver segmentation, while another may index the same video for surgical instrument detection. Our solution should allow reindexing the same video stream based on different AI models.
(C2) Scalable semantic search: Dealing with video streams for long retention periods requires managing large amounts of indexed data. In use cases like health video analytics, we may require indexing each video stream with fine granularity (*e.g.*, embedding per key video frame). Our solution must be able to handle large datasets.
(C3) Programmatic search interface: We see potential in exposing a semantic search mechanism via APIs to external programs. For instance, AI inference frameworks could reduce data transfers when loading data that is related to the specific model to train, instead of bulk loading a whole collection of video streams.
We built our prototype, named **StreamSense**, on top of a tiered streaming storage system (Pravega [5]) and validated it on a computer-assisted surgery use case from the National Center for Tumor Diseases.

2 StreamSense Usage

In Fig. 1, we provide a high-level overview of the use of StreamSense. First, we have the video ingestion phase. As visible in step ①, video frames from surgery cameras are ingested in Pravega as *video streams*. Users start an index request for a video stream (step ②). The indexer controller spawns a new *indexing instance* (*e.g.*, container) with the requested embedding model running inside and consuming data from the stream (step ③). The indexing container image ingests the video frames, passes them through the model, and writes the index data in the vector DB (step ④). Once a video has been indexed, users (or programs) can perform video searches using an image as input (step ⑤). The indexer controller performs a similarity search in the vector DB, which returns the most similar video frames (step ⑥). Finally, results are retrieved from Pravega (step ⑦) and then displayed to the user (step ⑧).

3 Evaluation

Datasets and Embedding Model. We utilize three publicly available datasets of surgical videos: CATARACTS [6], which includes videos of cataract surgeries; CHOLEC80 [7], which comprises videos of cholecystectomy surgeries; and AUTOLAPARO [8], which consists of videos of laparoscopic hysterectomy surgeries. To extract the embeddings, we make use of ResNet50 [9] trained with the `IMAGENET1K_V1` dataset.

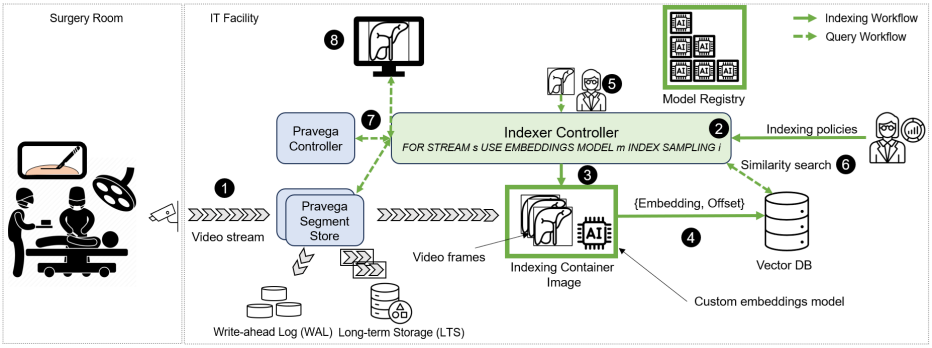

Fig. 1: Overview of StreamSense design and main components in the NCT use case.

Video Indexing Performance. Fig. 2 shows the total indexing latency for key video frames and the latency breakdown. The total key frame indexing latency ranges between 63ms and 360ms in average, depending on the decoding configuration. This is the waiting time for a data scientist to perform semantic search on ingested video frames in StreamSense. Regarding the latency breakdown, the Pravega IO latency shows a p99 latency under 7ms. For comparison, the latency of writing and reading 10KB events in AWS Kinesis at a rate of 25 events/s (same rate as our datasets) is 10.7x higher at p95. The AI inference latency to generate embeddings takes around 12-14ms per key frame and inserts to Milvus take 3-6ms. This latency meets the requirements for streaming workloads. The video decoding phase dominates latency (30-330ms).

Data Transfer Savings in AI Data Loading. In this experiment, we address the problem of data scientists downloading relevant video data for training their AI models. We evaluate the preliminary integration of our PyTorch data loader for StreamSense. We allow the data loader to use an embedding as input and call the StreamSense service for downloading similar video fragments. As can be observed in Table 1, StreamSense achieves a transfer reduction from 83.79% up to 99.79% compared to bulk transferring the whole dataset.

4 Conclusions

In this work we introduced StreamSense, which enables data scientists to index video frames using embeddings from AI models. StreamSense abstracts the complexities of vector database interactions, allowing users to perform semantic searches using images as input and visualize the results. We have shown that it can index video streams with low latency and reduce data transfers by over 80%.

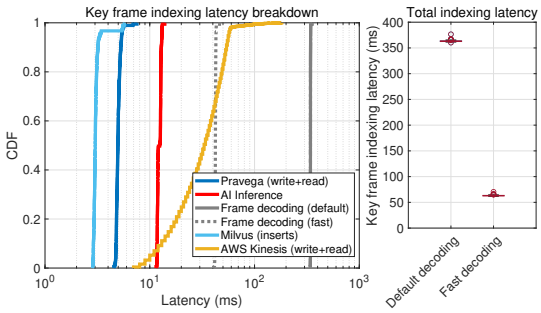

Fig. 2: Key frame streaming indexing latency with `h264` encoding.

Dataset	CATARCT	CHOLEC	LAPARO
15 *videos*	83.79%	87.97%	99.77%
45 *videos*	93.94%	92.82%	99.79%

Table 1: Data transfer reduction in AI data loading.

Acknowledgement. This work has received funding from the European Union Horizon Europe research and innovation programme under grant agreements 101092644 (NEARDATA) and 101092646 (CLOUDSKIN). We also thank the Cloud Native Computing Foundation (CNCF) for sponsoring the Pravega project. Gerard Finol is a Martí i Franquès PhD fellow at the Universitat Rovira i Virgili.

References

[1] J. Kreps et al. Kafka: A distributed messaging system for log processing. In *NetDB'11*, vol. 11, pp. 1–7, 2011.

[2] Amazon Web Services. Amazon Kinesis. `https://aws.amazon.com/es/kinesis`, 2024.

[3] Amazon Web Services. Kinesis Video Streams. `https://docs.aws.amazon.com/kinesisvideostreams/latest/dg/examples.html`, 2024.

[4] Apache Software Foundation. Apache Pulsar. `https://pulsar.apache.org`, 2024.

[5] R. Gracia-Tinedo et al. Pravega: A Tiered Storage System for Data Streams. In *Middleware'23*, pp. 165–177, 2023.

[6] H. Alhajj et al. CATARACTS. IEEE Dataport, 2021. DOI: `https://dx.doi.org/10.21227/ac97-8m18`.

[7] C. I. Nwoye et al. Rendezvous: Attention mechanisms for the recognition of surgical action triplets in endoscopic videos. *Med. Image Anal.*, 78:102433, 2022. DOI: `https://doi.org/10.1016/j.media.2022.102433`.

[8] Z. Wang et al. AutoLaparo: Multi-task dataset for laparoscopic hysterectomy automation. *arXiv*, 2022. `https://arxiv.org/abs/2208.02049`.

[9] K. He et al. Deep Residual Learning for Image Recognition. *arXiv*, 2015. `https://arxiv.org/abs/1512.03385`.

Towards the Use of Keypoint Matching Techniques for Region of Interest Registration in Whole Slide Images

Alessio Fiorin *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
Pathology Department, Oncological Pathology and Bioinformatics Research Group, Hospital de Tortosa Verge de la Cinta (HTVC), ICS, Institut d'Investigació Sanitària Pere Virgili (IISPV), 43500 Tortosa, Tarragona, Spain
alessio.fiorin@estudiants.urv.cat

1 Introduction

Digital pathology enhances efficiency, collaboration, and accessibility by converting tissue examinations into a digital format, allowing pathologists to analyse whole slide images (WSI) instead of traditional glass slides [1]. This technology is pivotal in breast cancer (BC) research, the most prevalent cancer among women worldwide, integrating digital tools with AI-driven analysis to improve diagnostic accuracy. A key focus is the evaluation of tumour-infiltrating lymphocytes (TILs), critical indicators of prognosis and treatment response, particularly in HER2-positive and triple-negative BC subtypes [2]. TIL evaluation is typically performed on specific regions of interest (ROI) identified on haematoxylin and eosin (H&E)-stained tissues according to established criteria [2]. These regions often contain diverse immune cells that influence tumour behaviour, necessitating detailed analysis through subsequent immunohistochemical (IHC) staining to identify distinct biomarkers. Automating ROI registration across sequential tissue sections enhances efficiency by reducing the repetitive manual selection of ROIs for different immune markers, facilitating the process for future diagnostic assessments.
Consequently, the primary objective of this study is to assess the efficacy of keypoint matching techniques in automating the registration of ROIs across sequential WSIs without the need for extensive preprocessing to correct for staining variations and tissue artefacts. By leveraging advanced keypoint detection and matching algorithms, this work aims to streamline the pathological analysis of BC tissues, focusing on the accurate alignment of H&E and IHC-

* PhD advisors: Prof. Marylène Lejeune, Prof. Carlos López Pablo, prof. Domenec Puig, and prof. Hatem A. Rashwan

stained slides. This could significantly reduce the manual effort required in current practices and improve the consistency of TIL evaluations, potentially impacting diagnostic and therapeutic decisions in BC treatment [3].

2 Materials

This study was conducted using tissue samples from twenty patients diagnosed with non-specific invasive BC between 2007 and 2015, sourced from the Tumor Banks of Hospital de Tortosa Verge de la Cinta (HTVC) and Hospital Joan XXIII (HUJ23), Tarragona, Spain, and the Biobank of the Pere Virgili Research Institute (IISPV). Ethics approval was granted by the IISPV Ethics Committee, with reference numbers 128/2022 and 218/2022. For each case, three consecutive WSIs were analysed: one WSI stained with H&E for initial ROI identification, followed by WSIs stained with CD8 and CD163 to highlight specific immune cell biomarkers. CD8 is a marker for cytotoxic T cells, which play a crucial role in anti-tumour immunity, whereas CD163 is a marker for macrophages associated with tumour progression and immune suppression. These stains highlight specific immune cell biomarkers within axillary lymph node tissues, crucial for evaluating tumour-infiltrating lymphocytes (TILs) following Salgado's criteria [2].

3 Methods

WSI registration involves aligning the H&E-stained tissue, containing the pathologist-identified ROI, termed the reference WSI, with the IHC-stained slide, called the moving WSI. The registration process uses a keypoint matching approach, where keypoints identified on the reference WSI are matched to those found in the moving WSI. After the initial keypoint matching, a filtering step is required to refine the keypoint pairs using the Random Sample Consensus (RANSAC) method. This step is crucial for ensuring the accuracy of the matches. Subsequently, the homogeneous transformation matrix is calculated to apply an Euclidean transformation to the moving WSI to the reference WSI, aligning the IHC-stained BC tissue with the H&E-stained BC tissue. This transformation also facilitates the warping of the ROI from the reference WSI to the moving WSI, adjusting the four vertices of the ROI to maintain its rectangular shape. This study evaluates traditional and advanced keypoint matching techniques on WSI registration, specifically the Scale-Invariant Feature Transform (SIFT) and a deep learning (DL) approach known as SuperPoint + SuperGlue (SP + SG) [4]. These techniques have been tested in the ACROBAT challenge for BC WSI registration. However, due to staining variance issues, both methods require a preprocessing step for the reference and moving WSIs. Recent DL-based methods, Dense Keypoint Matching (DKM),

and Robust Dense Feature Matching (RoMa) have been explored to address these challenges. These methods use a dense keypoint matching strategy and accept RGB images as input. Notably, they are available in two versions based on their training datasets: indoor (I), trained on the Scannet dataset, and outdoor (O), trained on the Megadepth dataset. In this study, these two models have been refined and adapted to enhance their applicability for WSI registration, leveraging advanced DL architectures to address the unique challenges of the high variability and complex morphological features inherent in medical imaging datasets. The number of registration failures is used to evaluate the strategies quantitatively. Failures are determined by testing a threshold (t) for the relative target registration error of the centroids ($rTREc$) [3]. The WSI registration is considered failed if the $rTREc$ exceeds or equals t. As noted, the maximum number of failures is 40 since there are two pairs for each case study (H&E-CD8 and H&E-CD163). The evaluation is conducted using a t of 0.025.

4 Results and Discussion

The evaluation of keypoint matching techniques for WSI registration showed varying performance, as depicted in Figure 1. This figure compares the number of registration failures among traditional SIFT, DL-based SP+SG, and the advanced DKM and RoMa techniques over 40 points. Figure 1 highlights that RoMa (O version) had the fewest failures, with only 3 out of 40 cases exceeding the relative target registration error threshold ($rTREc$) of 0.025. Conversely, SIFT had the highest failure rate, reflecting its inability to handle staining variations and the presence of extra tissues in WSIs. The $SP + SG$ method showed intermediate performance but was outperformed by RoMa, underscoring RoMa's efficacy due to its training on diverse datasets, which likely improves its adaptability without preprocessing. These findings suggest significant potential for RoMa in enhancing digital pathology by reducing manual effort and increasing the speed and accuracy of diagnostic workflows, aligning with the trend towards AI integration in medical imaging.

5 Conclusion and Future Work

This preliminary study demonstrates that the DL approaches, mainly the RoMa technique, especially its O version, achieve the lowest failure rates in WSI registration without requiring preprocessing image pairs, unlike methods such as SIFT and SP+SG. RoMa effectively handles WSIs with additional tissues, eliminating the need for their preliminary removal. This capability is advantageous for the pre-alignment phase in non-rigid registration processes. Future efforts will aim to expand data collection and explore additional techniques to refine WSI analysis further.

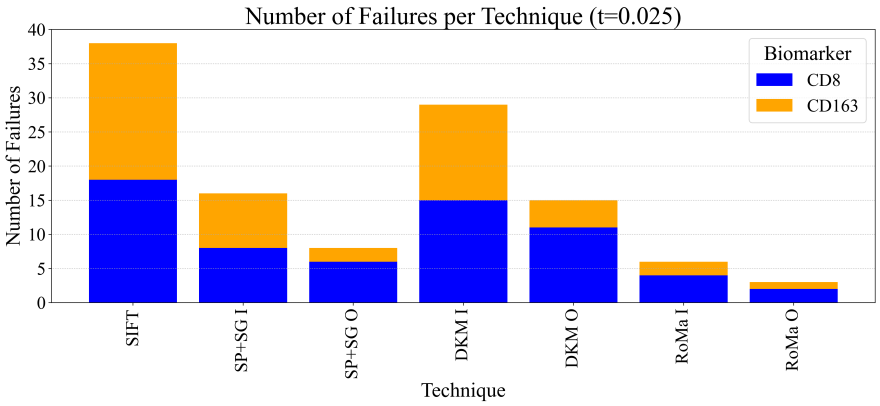

Fig. 1: The number of failures with a threshold of 0.025 is shown for each technique. The number of failures for the CD8 biomarker is shown in blue, while those for the CD163 biomarker are shown in orange.

Acknowledgement. This work was supported by the BosomShield project, funded by the MSCA-DN Actions (HORIZON-MSCA-2021-DN-01-01) under grant agreement 101073222, and SCARLET, funded by the Ministerio de Ciencia e Innovación under grant number TED2021-130081B-C22 via Proyectos Estratégicos Orientados a la Transición Ecológica y a la Transición Digital from NextGeneration EU. Special thanks to the pathology department at the HTVC and the Xarxa de Bancs de Tumors de Catalunya (XBTC), sponsored by the Pla Director d'Oncologia de Catalunya, for their support in data collection and provision.

References

[1] Kiran, N., Sapna, F., Kiran, F. et al. Digital pathology: transforming diagnosis in the digital age. *Cureus.* **15** (2023)

[2] Salgado, R., Denkert, C., Demaria, S. et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in BC: recommendations by an International TILs Working Group 2014. *Annals Of Oncology.* **26**, 259-271 (2015)

[3] Fiorin, A., Adalid Llansa, L., Goyda, E. et al. Optimising Region of Interest Registration for Multiple-Tissue Whole Slide Images. *International Workshop On Biomedical Image Registration.* pp. 333-345 (2024)

[4] Wodzinski, M., Marini, N., Atzori, M. & Müller, H. RegWSI: Whole slide image registration using combined deep keypoint-and intensity-based methods: Winner of the ACROBAT 2023 challenge. *Computer Methods And Programs In Biomedicine.* **250** pp. 108187 (2024)

Weighted Threshold Secret Sharing Schemes and Chow Parameters Approximation

Miquel Guiot *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
miquel.guiot@urv.cat

1 Introduction

A secret sharing scheme is a cryptographic primitive allowing a dealer to share a secret among parties such that only certain subsets, called authorized, can recover the secret. These subsets form the access structure of the scheme. Introduced by Shamir [1] and Blakley [2] in 1979, secret sharing schemes have broad applications in cryptographic protocols, particularly in secure multiparty computation and threshold encryption.
While schemes for threshold access structures are common, some applications require more general access structures. For example, in proof-of-stake models, validators have stakes proportional to their coin holdings. In such cases, there is a need of secret sharing schemes with *weighted threshold access structures* (WTASs). In these access structures, each party is assigned a weight according to its importance and the authorized subsets are those in which the sum of their weights is at least the threshold value.
The best-known secret sharing schemes for WTASs have share sizes that are either linear in the weights [1] or quasipolynomial in the number of parties [3], resulting in long shares. Recent approaches [4,5,6] address this by approximating weights with smaller ones, leading to WTAS approximations. This work [7] addresses the open problem of achieving a better tradeoff between the efficiency and the accuracy of the approximation.

2 Our Results

Given a WTAS Γ, our goal is to construct a similar WTAS Γ' with smaller weights. To achieve this, we translate the problem of approximating WTASs into approximating monotone Boolean functions. Two functions are considered ε-close if they differ on at most an ε fraction of inputs.

* PhD advisor: Oriol Farràs

In the case of a WTAS defined by a threshold T and a vector of positive weights $\boldsymbol{w} = (w_1, \ldots, w_n)$ assigned to the parties, the access structure is determined by a monotone Boolean function of the form

$$f(\boldsymbol{x}) = \operatorname{sign}(\boldsymbol{w} \cdot \boldsymbol{x} - T),$$

which are known as monotone LTFs in the context of complexity theory. Therefore, WTASs can be studied via complexity theory simply by considering the monotone LTFs assigned to them. In this context, our proposal reduces the weights of the associated LTF using the five-step process summarized in Figure 1.

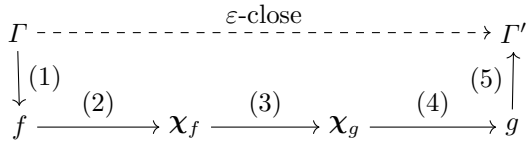

Fig. 1: Procedure for approximating any WTAS.

1. Consider the monotone LTF f associated to Γ.
2. Compute the Chow parameters $\boldsymbol{\chi}_f$ of f.
3. Find the Chow parameters $\boldsymbol{\chi}_g$ of a monotone LTF g that is ε-close to f and has small weights.
4. Construct the monotone LTF g from $\boldsymbol{\chi}_g$.
5. Consider the WTAS Γ' associated to g.

Steps (1), (2), and (5) of Figure 1 are immediate, so all the effort remains in deriving steps (3) and (4). To do so, we adapt the results of De et al. [8], in which they construct an approximate LTF with smaller weights by solving a problem related to the Chow parameters, described next.

2.1 The Chow Parameters Problem

Any Boolean function can be uniquely expressed as a real multilinear polynomial whose degree-0 and degree-1 coefficients are known as the Chow parameters. In this context, the Chow parameters problem consists in efficiently reconstructing a LTF from its Chow parameters.
De et al. [8] provide an algorithm to solve the approximate Chow parameters problem, constructing a close LTF with weights that scale sublinearly with input length and quasipolynomially with error. They achieve this result starting with the Chow parameters of the given LTF and modifying them step by step until the desired LTF is obtained.

In this work [7], we adapt the construction of De et al. [8] to the monotone setup to guarantee that the resulting LTF is monotone. Our main technical result is the following.

Theorem 1. *Let f be a monotone LTF. For any $0 < \varepsilon$ there exists an ε-close monotone LTF g represented by an integer vector with norm $\sqrt{n} \cdot \text{quasipoly}\left(\frac{1}{\varepsilon}\right)$.*

2.2 Secret Sharing Schemes Construction

Once we have derived a low-weight approximator for the original access structure, we still need to construct the information-theoretic and the computational schemes. In both cases, we apply known constructions that allow us to maximize the benefits of having an approximate WTAS with small weights. In the case of the information-theoretic scheme, we use Shamir's virtualization technique [1]. The total share size of the resulting scheme is $W \log W$, where W is the total weight, which we know that is small due to the approximation procedure. This leads to our main result.

Theorem 2 (Informal). *For any weighted threshold access structure Γ on n parties, there exists a secret sharing scheme with share size $n^{1+o(1)}$ whose access structure is $o(1)$-close to Γ.*

In the computational setting, Applebaum et al. [9] introduce the notion of Projective Pseudorandom Generator (pPRG), which yields secret sharing schemes with polylogarithmic share sizes for monotone circuits of polynomial size. Combined with the existence of such circuits for any WTAS [3], this results in highly efficient schemes. This is stated in the following theorem.

Theorem 3. *Under the subexponential RSA assumption, any weighted threshold access structure over n parties admits a computational secret sharing scheme where the size of the shares is* $\text{polylog}(n)$ *and the size of the public information is* $\text{poly}(n)$.

Figure 2 compares our results with the state-of-the-art solutions.

Acknowledgement. The author is supported by grant 2021 SGR 00115 from the Government of Catalonia, by the project ACITHEC PID2021-124928NB-I00 funded by MCIN/AEI/10.13039/501100011033/FEDER, EU, and by the project HERMES, funded by the European Union NextGenerationEU/PRTR via INCIBE.

References

[1] A. Shamir. How to Share a Secret. *Commun. ACM*, 22:612–613, 1979.

[2] G. R. Blakley. Safeguarding Cryptographic Keys. In *Proc. 1979 AFIPS Nat. Comput. Conf.*, pp. 313–317, AFIPS Press, 1979. *R. E. Merwin, J. T. Zanca, and M. Smith, editors.*

	Total Share Size	**Access structure**	**Error**	**Privacy**
[1]	$W \log W = 2^{O(n \log n)}$	WTAS	0	Perfect
[3]	$n^{O(\log n)}$	WTAS	0	Perfect
[5]	$W = 2^{O(n \log n)}$	$(t, t + \Omega(\lambda))$-ramp WTAS	-	$2^{-\lambda}$-Stat.
[4] Rounding, [6]	$O\left(\frac{n}{\beta - \alpha}\right)$	$(\alpha W, \beta W)-$ramp WTAS	-	Perfect
[4] BS Channels	$n \cdot \max\left\{\lambda^2, \text{poly}(\frac{1}{\beta - \alpha})\right\}$	$(\alpha W, \beta W)-$ramp WTAS	-	$2^{-\lambda}$-Stat.
Theorem 2	$n^{2+o(1)}$	WTAS	$o(1)$	Perfect
[3]	$\text{poly}(n)$	WTAS	0	Comp.
[9,3] Theorem 3	$n \cdot \text{polylog}(n)$	WTAS	0	Comp.

Fig. 2: Summary of secret sharing schemes for WTAS assuming 1-bit secrets.

[3] A. Beimel and E. Weinreb. Monotone Circuits for Monotone Weighted Threshold Functions. *IPL*, 97(1):12–18, 2006. Conference version: *Proc. 20th IEEE Conf. Comput. Complexity*, pp. 67–75, 2005.

[4] F. Benhamouda, S. Halevi, and L. Stambler. Weighted Secret Sharing from Wiretap Channels. In *4th Conf. Inf.-Theoretic Cryptography, ITC 2023*, vol. 267 of *LIPIcs*, pp. 8:1–8:19, 2023.

[5] S. Garg et al. Cryptography with Weights: MPC, Encryption, and Signatures. In *Adv. Cryptology – CRYPTO 2023*, pp. 295–327, 2023.

[6] A. Tonkikh and L. Freitas. Swiper: A new paradigm for efficient weighted distributed protocols. In *Proc. 43rd ACM Symp. Principles of Distr. Comput.*, pp. 283–294, PODC '24, 2024.

[7] O. Farràs and M. Guiot. Reducing the Share Size of Weighted Threshold Secret Sharing Schemes via Chow Parameters Approximation. In *TCC*, vol. 15367 of *LNCS*, pp. 517–547, Springer, 2024.

[8] A. De et al. Nearly Optimal Solutions for the Chow Parameters Problem and Low-Weight Approximation of Halfspaces. *J. ACM*, 61(2):11, April 2014. ACM, New York, NY, USA.

[9] B. Applebaum, A. Beimel, Y. Ishai et al. Succinct Computational Secret Sharing. In *STOC 2023*, pp. 1553–1566, ACM, 2023.

Hybrid Deep Learning Architectures for Robust End-to-End Camera Localization

Hussein Hasan Hameed *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
husseinhasanhameed.al-sinayyid@estudiants.urv.cat

Abstract
Camera localization, also known as camera pose estimation, involves determining the position and orientation of a camera in a 3D environment using visual data. Traditional methods, such as structure-based and image retrieval-based approaches, are effective but often encounter challenges in dynamic or visually ambiguous settings. Deep learning-based methods have emerged as alternatives, offering end-to-end learning and improved feature extraction. However, these methods often fail to balance fine-grained local feature extraction with global contextual understanding. To address these limitations, hybrid methods have been introduced, combining convolutional and attention-based mechanisms.The combination of convolutional and attention-based blocks ensures that the models dynamically adapt to the scene's requirements. While the convolutional layers focus on extracting detailed local features, the attention layers refine these features by incorporating global context. By strategically integrating these components, hybrid approaches provide robust and scalable solutions, achieving state-of-the-art results across both indoor and outdoor datasets.

1 Introduction

Accurate camera localization is essential in robotics, augmented reality, and autonomous navigation. It involves estimating the camera pose $\mathbf{p} = (\mathbf{t}, \mathbf{R})$, where $\mathbf{t}$ is the translation vector and $\mathbf{R} \in SO(3)$ is the rotation matrix. Given an image I, the task is to find $\mathbf{p}$ such that it aligns the camera with its environment, typically formulated as a regression problem $\mathbf{p} = f(I; \theta)$.
Traditional methods rely on structure correspondences or image retrieval. structure based approaches estimate the pose by solving 2D-3D correspondences using the Perspective-n-Point (PnP) algorithm, often with RANSAC

* PhD advisor: Miguel Angel Garcia , Domenec Puig

for robustness. Image retrieval methods compare the query image to a database of images with known poses and derive the camera pose from the closest matches. While effective in controlled settings, these methods struggle with scalability, texture-less surfaces, and repetitive patterns [1].
Deep learning approaches address some of these challenges by directly regressing the pose from input images. Convolutional neural networks (CNNs) extract fine-grained local features but lack the ability to capture global relationships. Transformers overcome this limitation using attention mechanisms to model long-range dependencies, but they often miss fine spatial details [1]. Hybrid architectures integrate convolutional and attention-based mechanisms to effectively capture both fine-grained local details and long-range global relationships. By combining these components, they overcome the limitations of traditional and standalone deep learning methods. This synergy enables hybrid models to deliver robust and scalable performance, making them well-suited for camera localization in complex, dynamic, and diverse environments [2].

2 Methodology

Hybrid architectures are designed to combine the strengths of convolutional and attention-based mechanisms. These models typically stack convolutional layers in the initial stages to capture local spatial features like textures and edges, followed by attention-based layers in later stages to model global relationships across the entire input [3].

2.1 Sequential Hybrid Architecture

As shown in figuer 1, in sequential hybrid models, the pipeline begins with a series of convolutional layers that progressively extract localized features while downsampling the spatial dimensions of the input. These layers often include specialized blocks designed for efficient feature extraction, such as depthwise separable convolutions or residual connections. After the initial feature maps are generated, the model transitions to attention-based layers. These layers use mechanisms like self-attention to capture long-range dependencies and spatial relationships across the image. By stacking these layers in sequence, the architecture balances computational efficiency with feature richness, enabling precise localization in both small-scale and large-scale environments[4].

2.2 Dual-Stream Hybrid Architecture

Dual-stream architectures take a parallel approach, processing multiple modalities, such as RGB and depth images, through separate but interconnected pipelines. Figure 2 demonstrates the dual stream architecture which Each

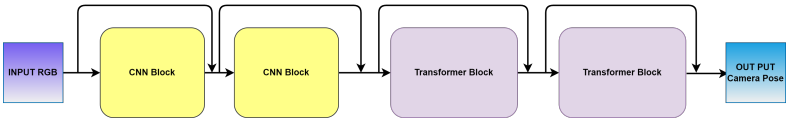

Fig. 1: Sequential Hybrid Model Architecture

stream begins with convolutional layers that extract features specific to its input modality—such as texture and color information from RGB data or geometric depth cues from depth images. The streams are then intertwined using attention mechanisms, such as cross-attention, which allow information flow between the two modalities. This integration enhances the model's ability to learn complementary features, leading to improved localization performance in scenarios where a single modality might be insufficient.

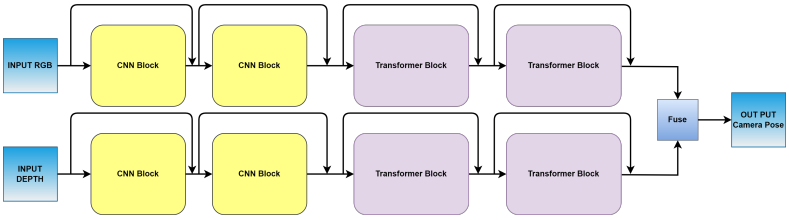

Fig. 2: Dual-Stream Hybrid Model Architecture

2.3 Results

The performance of hybrid architectures has been evaluated on widely used indoor and outdoor datasets. These datasets include a variety of challenges, such as cluttered scenes, repetitive patterns, and texture-less areas [1].
Hybrid architectures have demonstrated significant improvements in camera localization performance across both small-scale indoor and large-scale outdoor environments. In indoor scenes characterized by repetitive features, limited spatial extent, clutter, or complex lighting, these models effectively combine local and global features, achieving lower error rates compared to traditional methods and standalone deep learning models. Similarly, in outdoor environments with wider spatial extents and varying viewpoints, hybrid architectures excel at modeling relationships between distant features while preserving essential local information. As figure 3 shows, the estimated trajectories exhibit alignment with ground truth, providing insights into the models' robust estimating process.

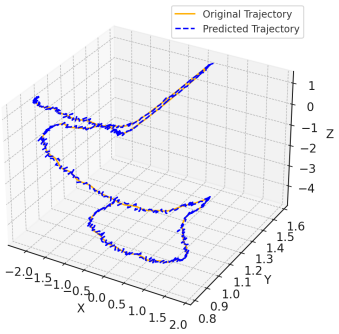

Fig. 3: Estimated trajectories compare to ground truth.The yellow line indicates the orginal trajectory while blue dots represents the estimated camera pose

3 Conclusion

Hybrid deep learning architectures represent a significant step forward in camera localization, addressing the limitations of both traditional methods and standalone deep learning approaches. By strategically stacking convolutional and attention-based layers, these models achieve a balanced representation of local and global features, enabling robust performance in complex and dynamic environments. The adoption of sequential and dual-stream designs further enhances their versatility, making them suitable for a wide range of applications. Future research can explore optimizing these architectures for real-time deployment and integrating additional modalities to further improve localization accuracy.

References

[1] M. Xu, Y. Wang, B. Xu, J. Zhang, J. Ren, Z. Huang, S. Poslad, and P. Xu, "A critical analysis of image-based camera pose estimation techniques," *Neurocomputing*, vol. 570, p. 127125, 2024. Available: https://doi.org/10.1016/j.neucom.2023.127125

[2] K. Bayoudh, "A survey of multimodal hybrid deep learning for computer vision: Architectures, applications, trends, and challenges," *Information Fusion*, vol. 105, p. 102217, 2024. Available: https://doi.org/10.1016/j.inffus.2023.102217

[3] T. Xie, K. Dai, K. Wang, R. Li, J. Wang, X. Tang, and L. Zhao, "A deep feature aggregation network for accurate indoor camera localization," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3687-3694, 2022. Available: https://doi.org/10.1109/LRA.2022.3146946

[4] Z. Dai, H. Liu, Q. V. Le, and M. Tan, "CoAtNet: Marrying Convolution and Attention for All Data Sizes," *arXiv preprint arXiv:2106.04803*, 2021. Available: https://arxiv.org/abs/2106.04803

Non-Invasive Classification of Breast Cancer Molecular Subtypes Using Mammographic Images

Adnan Khalid *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
adnan.khalid@urv.cat

1 Introduction

Breast Cancer (BC) remains an immense global health concern, recognized as a leading cause of cancer-related death among women, requiring accurate and effective diagnostic methods. The widespread influence of this issue requires efficient screening and diagnostic methods to address its high recurrence and death rates. Mammography is considered a fundamental method in the detection of breast cancer. It is widely acknowledged for its effectiveness in reducing mortality rates by enabling the early identification of the illness by 42% since 1989 [3]. BC is a heterogeneous disease that consists of different molecular subtypes. The presence of both intra and inter-tumor heterogeneity plays a significant role in drug resistance and treatment effectiveness. The molecular subtypes are categorized into four classes based on the expression of immunohistochemical markers: Luminal A, Luminal B, HER2, and Triple-Negative [2]. In addition to facilitating tailored treatment methods, having an understanding of the molecular subtype of breast cancer helps with risk assessment and prognostication, which ultimately leads to improved patient outcomes

2 Challenges and Objectives

To identify a molecular subtype of BC, doctors usually need to perform a biopsy. This process involves taking a tissue sample, which is invasive, takes time, and adds extra costs to the treatment.
This work aims to automate the process of classifying the molecular subtypes of BC using artificial intelligence (AI), leveraging advanced Deep Learning algorithms in multi-modal radiological imaging, mainly mammography, to provide a non-invasive, efficient, and cost-effective alternative to biopsies [1].

* PhD advisor: Domenec Puig, Hatem A.Rashwan

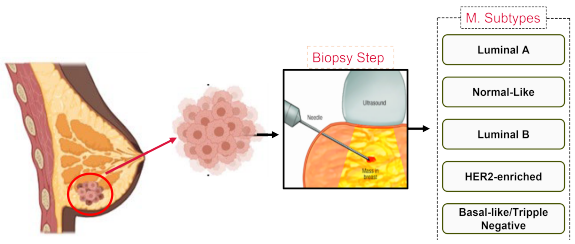

Fig. 1: Overview of an Invasive Biopsy Process.

2.1 Proposed Methodology

This work presents an end-to-end framework for classifying BC molecular subtypes from radiological images (e.g., mammograms), structured into three key stages, as illustrated in Figure 2.

- **Tumor Detection and Segmentation Based on Deep Learning:** In the first stage, a deep learning-based segmentation model is employed to detect and precisely segment the tumor region. The model ensures accurate delineation of tumor boundaries, enhancing the reliability of subsequent feature extraction.
- **Feature Extraction via Image Processing & Radiomics:** The second stage focuses on extracting key morphological and radiomic (texture-based) features from the segmented tumor region. Advanced image processing techniques are applied to quantify tumor shape and margin characteristics, which are crucial for distinguishing molecular subtypes.
- **Molecular Subtype Classification Based on Machine Learning:** In the final stage, the extracted morphological and radiomic features are fed into machine learning models—such as Random Forest (RF), CatBoost, and Support Vector Machine (SVM)—to classify tumors into molecular subtypes (Luminal A, Luminal B, HER2+, and Triple-Negative).

The proposed framework enhances subtype classification accuracy, providing a more precise approach to BC diagnosis from mammography.

3 Results and Future Direction

We employed quantitative and qualitative measures to compare the proposed framework to determine its effectiveness. We conducted an in-depth analysis of the relationship between tumor shape, margin, radiomics features, and BC molecular subtypes. This analysis identified novel correlations that are strongly supported by clinical data, offering new insights into the biological behavior of different subtypes.

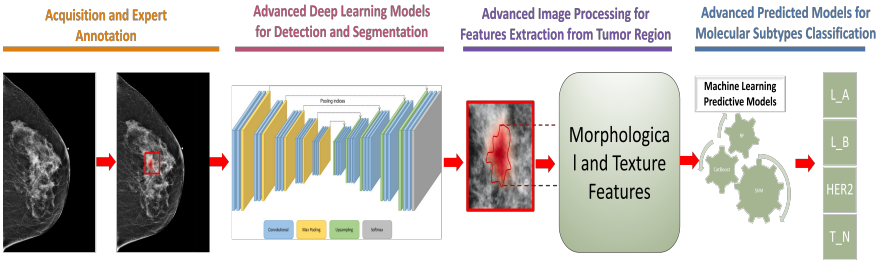

Fig. 2: The Proposed Approach For The Automatic Classification Of Molecular Subtypes Of BC (I.E., Four Classes: Luminal A, Luminal B, Triple-Negative, HER2) From The Analysis Of Mammographic Images.

Table 1: Classification Results of Molecular Subtypes for Dataset A and Dataset B

	Accuracy	**Precision**	**Recall**	**F1 Score**	**ROC/AUC**
Dataset A	**0.9252**	0.9244	0.9246	0.9252	0.9724
Dataset B	**0.9448**	0.9476	0.9447	0.9451	0.9838

Based on the results we got from our private in-house dataset using this framework for subtype classification, we present a more accurate and efficient approach to classifying breast cancer molecular subtypes compared to existing methods. Our framework consistently outperforms other approaches in distinguishing subtypes such as Luminal, HER2-positive, and Triple-negative, highlighting its potential for use in clinical and research settings, contributing to more precise and personalized breast cancer treatment strategies.

Acknowledgement. This work was supported by the Bosomshield Project, a grant from Marie Sklodowaka-Curie Doctoral Networks Actions(HORIZON-MSCA-2021-DN-01-1-101073222)

References

[1] Nariya Cho. *Molecular subtypes and imaging phenotypes of breast cancer*. Ultrasonography, 35(4):281, 2016.

[2] Ana M. Mota, João Mendes, and Nuno Matela. *Breast Cancer Molecular Subtype Prediction: A Mammography-Based AI Approach*. Biomedicines, 12(6):1371, 2024.

[3] J. Ferlay, M. Ervik, F. Lam, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram, and F. Bray. *Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. [Internet]*. 2018.

Exact and Efficient Unlearning for Large Language Models.

Hajar Lachheb *

CRISES: Security and Privacy Research Group
Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
hajar.lachheb@urv.cat

1 Introduction

Large Language Models (LLMs) excel in tasks like text generation and translation, but their training data often includes sensitive or copyrighted content, raising privacy concerns. Regulations like GDPR and CCPA require data removal mechanisms, making efficient unlearning techniques essential [3]. Retraining is costly, and approximate unlearning lacks guarantees.
The Sharded, Isolated, Sliced, and Aggregated (SISA) framework offers exact unlearning but is computationally expensive as represented in figure 1 [4]. We enhance SISA for text-based models, improving data partitioning to boost unlearning efficiency while maintaining performance. Our approach, optimized for textual data, explores multiple distribution strategies to improve unlearning in SISA.

Fig. 1: Standard SISA Distribution

We apply this to BERT models for copyright and privacy protection [1], comparing strategies and evaluating trade-offs between guarantees, cost, and performance. Our results show that smarter partitioning reduces unlearning time, making large-scale LLMs more compliant with privacy regulations.

* PhD advisor: Josep Domingo Ferrer, David Sánchez Ruenes

2 Methodology

This work addresses exact machine unlearning for LLMs in privacy and copyright contexts, enhancing the SISA approach to reduce retraining costs and environmental impact while ensuring legal compliance.
Our methodology consists of five key components:

- **Data Preprocessing**: We preprocess the dataset to remove noise and standardize the format, enabling efficient prioritization of data points for unlearning.
- **Forget Priority Score**: We introduce a forget priority score to evaluate which data samples are most critical to forget. The score $F(i)$ for each data point i is computed based on two factors: recency and sensitivity. It is given by:

$$F(i) = \alpha R(i) + \beta S(i), \tag{1}$$

  where $R(i)$ represents the recency of the data point (e.g., publication year or birth year), $S(i)$ represents the sensitivity (We will be trying different sensitivity features, and observing the results while combining them for both the copyright and privacy requests), and α, β are the weights assigned to each factor.
- **Sharding and Slicing**: We partition the data into S shards and R slices to optimize unlearning and minimize retraining, prioritizing recent and sensitive data for removal. The partitioning ensures that (1) the most recent data is in the final slices of each shard and (2) recent data appears earlier in the shards. This strategy, shown in figure 2, ensures exact unlearning with minimal cost and environmental impact, while complying with privacy and copyright regulations.

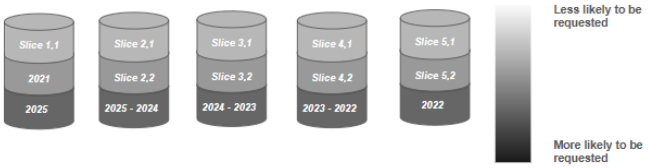

Fig. 2: SISA Distribution Strategy

- **Training Pipeline**: Our training pipeline optimizes the SISA framework [2] by strategically partitioning the dataset into *shards* and *slices*, with each slice added incrementally during training. The model trains sequentially on each slice, updating its parameters, and uses an ensemble method with majority voting for inference, aggregating predictions from models trained on individual shards.

- **Unlearning mechanism**: Upon receiving an unlearning request, our system identifies the relevant shard and slice. Requests are assumed to follow an exponential distribution, favoring newer data. The forgetting probability $P_{\text{forget}}(t)$ is modeled as:

$$P_{\text{forget}}(t) \propto e^{-\lambda(t-t_0)}, \tag{2}$$

where t is the publication year, t_0 is a reference year, and λ controls the decay rate, in line with cognitive psychology's forgetting model [5]. Each request is processed independently by retraining the affected slices, and after all requests, we aggregate the results to update the model while preserving utility and unlearning guarantees similar to the original SISA framework.

3 Experiments and Results

Experiments Setup For evaluation, we used two request-specific datasets: the *BookMIA dataset* (10,000 literary snippets) for copyright and the *Wikipedia Biographies Text Generation dataset* (10,000 biographies) for privacy. Both were processed using a fine-tuned BERT-based architecture (`BertForMaskedLM('bert-base-uncased')`). The datasets were split into 5 shards and 3 slices, with training for one epoch, a batch size of 8, BERT tokenization, and a modified vocabulary for privacy. Sentence transformers were used for representation learning, and Adam optimized the training.

Experiments Results: Copyright Requests The results in Table 1 compare various unlearning methods for copyright-related requests.

Table 1: Copyright Requests - Final Evaluation

Experiments	Training		Unlearning	
	Training Time	Training Loss	Unlearning Time	Unlearning Loss
SISA Standard	93.19s	0.8184	280.37s	0.9941
SISA Standard Random Requests	93.19s	0.8184	305.45s	1.0084
SISA Enhanced				
First Distribution	89.92s	0.7116	120.25s	1.5381
Second Distribution	90.86s	0.9918	115.69s	1.0545
Third Distribution	93.91s	0.8729	110.73s	1.0708
Fourth Distribution	95.58s	0.8408	109.18s	1.0991
Fifth Distribution	91.56s	0.8622	107.97s	1.0275
Sixth Distribution	92.41s	0.8092	96.27s	1.0997
Approximate Unlearning				
Gradient Ascent	339.36s	2.0510	102.80s	1.2337
Gradient Difference	339.36s	2.0510	103.20s	1.3422
KL Minimization	339.36s	2.0510	101.00s	2.3937
Preference Optimization	339.36s	2.0510	105.30s	2.0548
SISA Enhanced Random Requests				
First Distribution	89.92s	0.7116	242.52s	0.9975
Second Distribution	90.86s	0.9918	230.58s	1.0270
Third Distribution	93.91s	0.8729	222.68s	1.0104
Retraining on Retain Set				
One User Request	339.36s	2.0510	337.61s	2.0620

SISA Standard methods show high unlearning times (280.37s–305.45s) with consistent training losses but offer limited efficiency gains compared to **SISA Enhanced** methods, which reduce unlearning times, especially for the Sixth Distribution (96.27s). The **Fifth Distribution** balances low unlearning loss (1.0275) and moderate time (107.97s), making it a strong choice. **Approximate Unlearning** methods are faster (101.00s–105.30s) but incur higher training and unlearning losses, demonstrating the trade-off between efficiency and accuracy. **SISA Enhanced Random Requests** result in longer unlearning times (e.g., 242.52s for the First Distribution) due to random requests, but still outperform **SISA Standard** in efficiency and accuracy. **Retraining on the Retain Set** is the least efficient approach, particularly given that it derives from a single user request, resulting in high unlearning times (337.61s) and significant losses (2.0620). Overall, **SISA Enhanced** improves over **SISA Standard** but varies across distributions, with the Fifth and Sixth Distributions being the most balanced for unlearning tasks.

4 Conclusion

In addition to the promising results observed for copyright unlearning, we also found encouraging outcomes for privacy-related unlearning tasks. These findings suggest that our approach is effective across different types of sensitive data. Moving forward, the next step will be to deepen our understanding of unlearning in large language models (LLMs) by exploring other advanced techniques, such as explainable AI, which can provide greater insight into the decision-making process behind unlearning.

Acknowledgement. This work was supported by the Joan Oró grant for predoctoral researcher recruitment (FI 2024). We also thank the CRISES research group at Universitat Rovira i Virgili for their valuable contributions.

References

[1] A. Vaswani, N. Shazeer, et al. Attention is all you need. *NeurIPS*, 2017.

[2] L. Bourtoule, F. B. Bastani, Y. A. K. Kim, et al. Machine unlearning. *NeurIPS*, 2019.

[3] Regulations such as GDPR and CCPA. Data privacy laws and AI compliance. *Data Privacy and Security Review*, 2023.

[4] R. L. Cook, P. T. Wang, J. K. Thompson. The SISA framework for exact unlearning. *ACM Transactions on AI*, 8(4):301–315, 2021.

[5] P. Wozniak. SuperMemo: The forgetting curve and the exponential decay model. *Journal of Cognitive Psychology*, 34(5):1–10, 2016.

Effective Unlearning in Large Language Models

Tamim Al Mahmud *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
tamimal.mahmud@urv.cat

Abstract Numerous AI applications, including virtual assistants, chatbots, and recommendation systems, use large language models (LLMs). However, there are serious privacy issues, since these models could retain confidential and copyright-protected data. LLMs may unintentionally reveal private information during inference since they are trained on large datasets, endangering individual privacy. The European Union (EU), in order to tackle these issues, introduces the Right-to-Be Forgotten (RTBF) clause under the GDPR (General Data Protection Regulation). People can request the deletion of their data under this provision. Unfortunately, removing unwanted data requires retraining the model from scratch, which can be costly and impractical for LLMs in a continuous manner. This research proposes a novel technique called LLM unlearning using Differential Privacy (DP), which can substantially lower the costs of unlearning by approximately 50% compared to more conventional approaches like retraining from scratch or fine-tuning pre-trained models exclusively on retained data. Our method also prevents models from retaining private and copyright-protected information during training.

1 Introduction

In recent years, due to training on massive text corpora [1] with trillions of tokens, LLMs have brought a revolution in the event of the creation of human-like text, language comprehension in a sophisticated manner, and other problems related to natural language. Despite this advancement, LLMs have the potential of leaking sensitive information [2], ranging from personal information identity, including critical identifiers (e.g., name, SSN, phone number, email address) and quasi-identifiers (e.g., date of birth, sex, zip code, ethnicity).
To protect personal privacy, the European Union (EU) General Data Protection Regulation (GDPR) has established rules giving people the right to have

* PhD advisors: Josep Domingo-Ferrer and David Sánchez

their personal information removed, which is known as the person's right to be forgotten (RTBF). A similar act in the USA, known as the CCPA, allows resident the right to know, access, delete, and opt out of the sale of their personal information. Consequently, LLMs' unlearning is becoming a concerning issue among organizations in compliance with the different data privacy and security institutions' acts.

Unlike deleting a file from a computer, removing data from an AI model is extremely difficult and costly. The only guaranteed method is to retrain the entire model from the beginning, which can take months and cost millions of euros. For example, training ChatGPT-4 reportedly cost over 100 million dollars. If AI companies had to retrain models every time someone requested data removal, it would be financially impractical. In recent times, researchers [3] have been trying to tackle this problem using different model editing-based approximate forgetting such as Gradient Ascent (GA), Gradient Difference (GD), Preference Optimization (PO), and Kullback-Leibler (KL) divergence, etc. However, due to the non-linear parameter relationship, neither approach can guarantee personal data removal, even when offering small-scale forgetting in trade for model utility.

2 Proposed Method

In LLMs, the general goal of unlearning is to remove target knowledge while accurately maintaining model performance for non-targets. Specifically, when a Forget ($\mathcal{D}_f$) request arrives, the model must forget the information associated with the requested entity. At the same time, it can perform at the same level on Retain ($\mathcal{D}_r = \mathcal{D} \setminus \mathcal{D}_f$) data as it did on the whole ($\mathcal{D}$) data.

In this research, we present a method for unlearning that utilizes differential privacy (DP), offering a cost-effective strategy for LLM unlearning. The primary aspect of the DP method is safeguarding personal data confidentiality by introducing calibrated noise during model training, effectively minimizing the risk of data exposure. Despite the inclusion of noise, the approach maintains critical patterns found in the original data, with further fine-tuning on retained data, effectively restoring the model performance and allowing for the effective unlearning of unwanted information.

The workflow architecture of our method is shown in Figure 1. Our method works in two stages.

In stage one, referred to as unlearning-ready training, we create the privacy-protected private base model (BM_D^{DP}) by training (for a new model from scratch) or fine-tuning (for the pre-trained model) on raw data (D), ensuring DP constraints. This private base model is then further fine-tuned on raw data (D) to make it M_D^{DP} ready to deploy for real-world applications.

In stage two, referred to as unlearning request handling, for each forgetting request, we delete the current operational model and then fine-tune the private

base model on D_r to make the unlearned model ($UM_{D_r}^{DP}$), which is ready to deploy for the real-world applications.

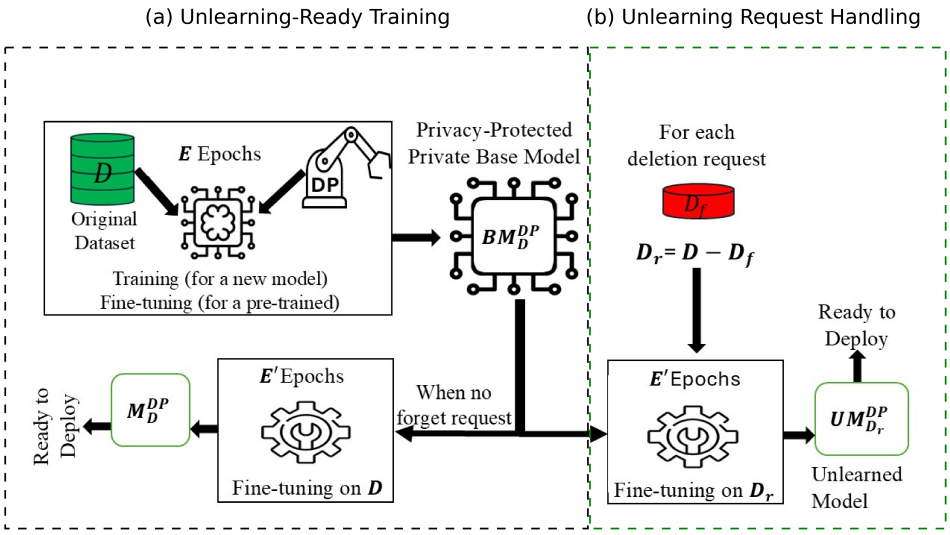

Fig. 1: Our Method Workflow Architecture

3 Dataset and Model

We used the TOFU [3] dataset, which contains 200 fictitious author profiles, with each profile having 20 questions and answers pairs. The dataset is further divided into Forget set and Retain set with 1%-99%, 5%-95%, and 10%-100% pairs, respectively.
We employed the pre-trained Microsoft Phi model for the experiment, which has approximately 1.5 billion parameters.

4 Experimental Results

Our experiment focuses on two primary evaluation metrics: forget quality and model utility. The forget quality and model utility scores at 5% forgetting are illustrated in Figure 2. Our method achieves a forget quality and model utility close to FR-exact (exact forgetting). In contrast, all approximate methods, except GD, fail to achieve the minimum level of forgetting (threshold ≥ 0.05).

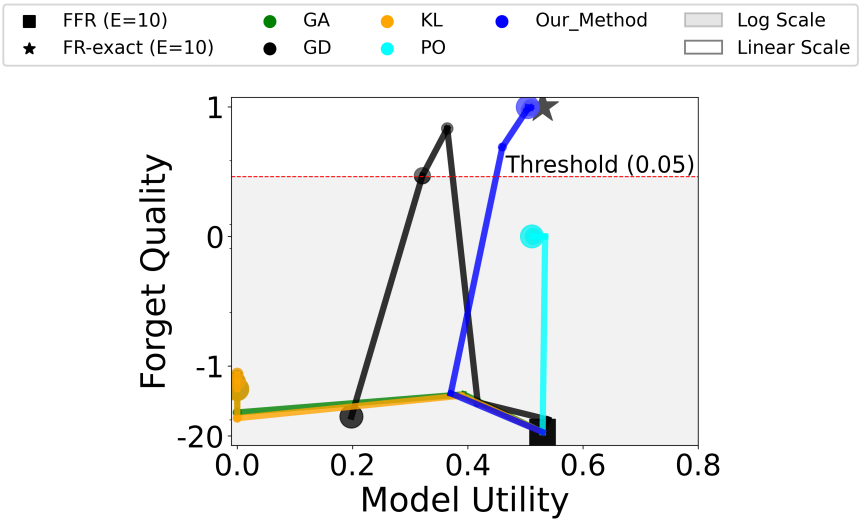

Fig. 2: Forget Quality vs Model Utility. FFR stands for fine-tuning the pre-trained model on the TOFU dataset, both the Forget Set and Retain Set. FR-exact stands for fine-tuning the model on the Retain set only. Epochs are $E = 10$ for FFR and FR-exact and $E' = 5$ for others. The comparative size of the markers represents the number of epochs: the larger the epoch number, the larger the marker size.

5 Conclusion

We proposed an efficient unlearning approach for LLMs, compliance with regulations like GDPR. By integrating differential privacy, our method enables selective data removal with minimal performance impact, offering a scalable, cost-effective solution for responsible AI deployment.

Acknowledgement. My PhD grant is from the Spanish Ministry of Science and Innovation under the FPI scholarship PID2021-123637NB-I00 (PRE2022-101401).

References

[1] Achiam, Josh, et al. Gpt-4 technical report. *arXiv preprint: 2303.08774*, 2023.

[2] Carlini, Nicholas, et al. Extracting training data from large language models. *30th USENIX Security Symposium (USENIX Security 21)*. 2021.

[3] Maini et all. TOFU: A task of fictitious unlearning for LLMs *Conference on Language Modeling*, University of Pennsylvania between October 7-9, 2024.

Emergent Behaviour and Spatial Effects in Microbial Communities

Mattia Mattei *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
mattia.mattei@urv.cat

1 Introduction

Microorganisms, such as bacteria, do not function in isolation; rather, they exist within multi-species communities, cohabiting the same environment and engaging in a wide range of interactions. Different species—or even individuals within the same species—can compete for resources, cooperate through cross-feeding, establish predator-prey relationships, and form intricate networks of interactions. Understanding these complex microbial communities is crucial, as microorganisms coexist with nearly all forms of life. They inhabit our gut, skin, and mouth, influencing our health, and they shape marine and soil ecosystems, determining their stability and function. However, their complexity presents a major challenge: how do the physiological behaviors of individual species give rise to emergent properties such as stability, productivity, and resilience? From this perspective, microbial communities can be regarded as complex adaptive systems.

Despite their complexity, recent studies have revealed universal properties that emerge across microbial communities. However, many of these regularities remain difficult to explain using existing theoretical models. For instance, microbial communities exhibit both high diversity and remarkable long-term stability in composition [3], a phenomenon that contradicts classical theoretical ecology models—such as Lotka-Volterra and Consumer-Resource models —which predict that high diversity should lead to instability. This paradox, particularly evident in microbial ecosystems, echoes the well-known paradox of the plankton. Another intriguing feature is that, while the taxonomic composition of microbial communities varies across individuals and ecosystems, their functionality—defined by the metabolites they produce and consume—remains highly conserved [3]. This suggests the presence of underlying assembly rules that shape microbial communities. Recently, Grilli [2] analyzed data from diverse biomes and identified three macroecological laws that quan-

* PhD advisors: Alex Arenas, Sergio Gómez

titatively describe species abundance fluctuations across communities and over time. Strikingly, these patterns were reproduced by a mathematical model that disregards interactions entirely, relying solely on environmental stochasticity. This result appears to contradict the well-documented importance of microbial interactions, raising fundamental questions about the role of species interactions in shaping microbial community dynamics.

2 The role of space

My research focuses on developing mathematical and computational models of the microbiome that can reproduce emergent patterns and enhance existing theoretical frameworks. In particular, we propose that space is the missing ingredient in current theoretical models. Specifically, incorporating the fact that microbial species are embedded in a physical space—an aspect typically overlooked in mean-field descriptions—is crucial for a more accurate representation. Several factors support this perspective. First, microbial interactions are predominantly short-range, typically occurring over just a few body lengths (a few micrometers). As a result, microorganisms interact only with their closest neighbors, remaining unaware of the broader community. Second, over 80% of bacteria in natural environments exist in biofilms—dense, diverse communities embedded in a self-secreted polymer matrix that often adheres to surfaces. This implies that most microbes are spatially constrained rather than freely moving. Third, bacteria exhibit self-organization: they can sense their environment and adjust their motion and behavior accordingly. By explicitly considering these aspectes and, therefore, spatial structure, our approach aims to provide a more realistic theoretical framework for understanding microbial communities.

2.1 Emergent Coexistence explained by Spatial Segregation

Recently, Chang et al. [1] provided strong experimental evidence that multispecies coexistence is an emergent phenomenon. They isolated organisms from stable synthetic bacterial communities composed of multiple species and systematically tested all possible pairwise competitions. In most cases, one species outcompeted the other, leading to exclusion. From this, they concluded that coexistence in communities cannot be reduced to simple pairwise interactions, leaving open the question of the fundamental mechanisms driving this phenomenon. In our work [5], we investigate the role of spatial self-organization as a key factor underlying emergent coexistence. Using high-performance individual-based simulations, we demonstrate that bacteria can naturally form spatially segregated patches if they have the ability to move away from regions of high competition (figure 1). Incorporating this patchy environment into a Lotka-Volterra framework, we show that spatial segregation

enhances coexistence when multiple species are present but not in two-species systems—thus reproducing the findings of Chang et al.

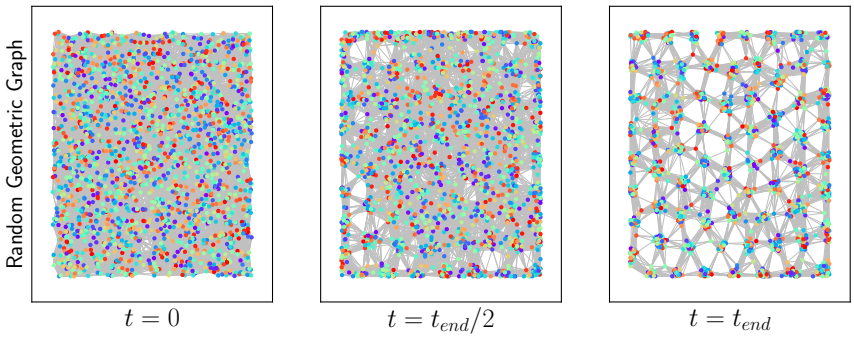

Fig. 1: Example of a simulation outcome showing segregated clusters of bacteria emerging due to competition avoidance. Each point represents an individual bacterium, with colors indicating different species. Interactions between neighboring bacteria are determined by a Random Geometric Graph (RGG), a network in which two nodes are connected only if their distance is below a specified threshold R.

2.2 Biofilm Formation

Currently we are investigating the mechanisms that drive bacteria to form biofilms. The transition from a free-moving (planktonic) state to a surface-bound biofilm is a complex process influenced by multiple factors. Among these, cooperative interactions within the community—providing collective benefits—have been proposed as a key driver of biofilm formation [4]. Using again a combination of individual-based simulations and analytical models, we show that this transition can be triggered by spatial self-organization. With a sufficient number of mutualistic interactions, the system can undergo a sharp phase transition from an almost fully planktonic phase to a state in which all particles find stable favorable locations, driving the formation of the biofilm (figure 2).

2.3 Future work

The final part of my thesis will focus on developing a minimal model that incorporates the effects of spatial structure while summarizing the key findings discussed above. One approach is to consider species interactions as time-

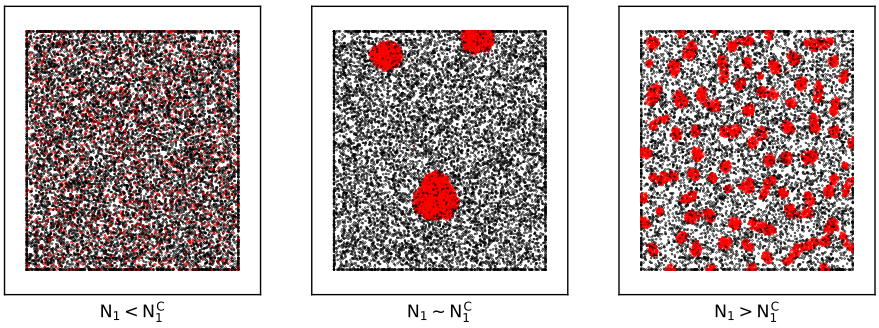

Fig. 2: Simulation of a system with two bacterial species: one capable of cooperation and biofilm formation, and the other purely competitive. When the cooperative species is scarce, both species remain in motion, leading to a mixed spatial distribution (left panel). As its population increases, a critical threshold N_1^C is reached, allowing the cooperative bacteria to cluster and initiate biofilm formation. At this threshold, large, dense clusters emerges (central panel). Further increasing the abundance of the cooperative species results in the formation of multiple smaller clusters (right panel).

dependent rather than fixed within a static interaction network. These temporal dynamics will be governed by the relationships between movement, spatial self-organization, and growth timescales.

Acknowledgement. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 945413 and from the Universitat Rovira i Virgili (URV).

References

[1] Chang-Yu Chang *et al.* Emergent coexistence in multispecies microbial communities. *Science* 381, 343-348 (2023).

[2] J. Grilli. Macroecological laws describe variation and diversity in microbial communities. *Nat Commun 11*, 4743 (2020).

[3] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*, 486, 207-214 (2012).

[4] KK. Jefferson. What drives bacteria to produce a biofilm? *FEMS Microbiol Lett* 236(2): 163-73, 2004.

[5] M. Mattei and A. Arenas. Exploring spatial segregation induced by competition avoidance as driving mechanism for emergent coexistence in microbial communities. *Phys. Rev. E* 110, 014404, 2024.

Enhancing Personal Tourist Experiences with Context-Aware, Explainable, and Sustainable Strategies Using Hierarchical Multi-Criteria Recommendation Methods

Monir Yahya Salmony *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
moniryahyaali.salmony@urv.cat

1 Abstract

Recommendation systems have advanced significantly in several industries, including tourism, due to the quick development of information technology and the growing need for individualized services. Recommendation systems play a critical role in influencing the experiences of tourists by recommending attractions, activities, and itineraries that suit personal interests. However, despite widespread adoption, current recommendation systems often neglect important criteria such as sustainability and explainability. This research aims to bridge this gap by developing novel recommendation strategies that are context-aware, personalized, sustainable, transparent, and grounded in hierarchical multi-criteria decision-making. Moreover, the study uses data related to point of interest in Tarragona city, Spain that scraped from TripAdvisor, therefore ensuring that the suggested models are grounded in actual data.

2 Background and Motivation

Tourism is one of the largest and most dynamic industries globally, with a profound impact on local economies, cultures, and environments [1,2]. The emergence of digital channels has revolutionized the way visitors plan their travel, transforming recommendation systems into an essential mean for decision-making [2]. Traditional recommendation systems, however, frequently focus solely on matching user preferences with existing alternatives, neglecting the broader implications of their suggestions. For instance, many systems prioritize popular attractions, leading to overcrowding and strain on local resources, which can undermine the sustainability of tourism [1,3]. Moreover, the lack of

* PhD advisor: Antonio Moreno

transparency in generating recommendations can erode user trust and limit the potential for meaningful user interaction with the system. These concerns underscore the need for multifaceted approaches, such as hierarchical multi-criteria decision-making that concurrently considers user preferences, sustainability goals, and contextual factors [3].

3 Gap in Previous Work

Previous research in recommendation systems has made significant strides in personalization, leveraging techniques such as collaborative filtering, content-based filtering, and hybrid approaches to improve recommendation accuracy [2]. However, these systems often overlook two critical dimensions: sustainability and explainability. Sustainability, in the context of tourism, refers to the balance between economic, social, and environmental considerations [3]. Existing recommendation systems rarely account for the ecological impact of their suggestions or the distribution of tourist activities, which can lead to over-tourism and resource depletion [1,2]. Similarly, explainability remains underexplored; many users are presented with recommendations without insight into the reasoning or criteria used to generate them [2]. Hierarchical multi-criteria decision-making can address this gap by systematically evaluating multiple dimensions—such as environmental impact, user satisfaction, and cost—within a single, unified framework [3,1].

4 Research Objectives

The primary objective of this research is to integrate context-awareness, sustainability, explainability, and hierarchical multi-criteria decision making into a unified recommendation system. Specifically, the research aims to:

- **Develop Context-Aware Recommendation Strategies:** Design algorithms that includes contextual variables, user preferences and real-time information to deliver personalized suggestions.
- **Incorporate Sustainability Considerations:** Develop mechanisms to distribute tourists evenly, encouraging visits to lesser-known places to ease pressure on popular attractions while promoting economic diversification and preserving natural and cultural heritage.
- **Enhance Explainability:** Implement features that offer clear explanations for the recommendations made. This will boost tourists trust allowing to make informed decisions about their travel plans.
- **Embed Hierarchical Multi-Criteria Decision Making:** Employ a structured, multi-level framework to weigh different criteria such as preferences, context and sustainability.

- **Evaluate the Impact of Recommendations:** Assess the effectiveness of the proposed system concerning user satisfaction, sustainability outcomes, and explainability. The evaluation will involve new metrics that capture the multidimensional nature of the recommendations.

5 Methodology

The research adopts a comprehensive and iterative approach to develop and evaluate the recommendation system. Figure 1, shows the methodology will consist of the following steps:

- **Literature Review:** Conduct an extensive review of existing recommendation systems, with special attention to their sustainability and explainability components, as well as the core principles of hierarchical multi-criteria decision making.
- **Data Collection and Scraping:** Gather diverse datasets on tourist activities, geographical locations, and user preferences. This will include data scraping of points of interest from TripAdvisor for Tarragona, Spain—a setting chosen for its rich cultural heritage and tourism challenges. Additional data will be sourced from publicly available repositories, local tourism authorities, and user interactions with digital platforms.
- **Model Development:** Adapt, and develop suitable algorithms that blend machine learning techniques with hierarchical multi-criteria decision making for context-aware, sustainable, and explainable recommendations.
- **Training and Validation:** Train and apply cross-validation for the models on the collected data, to ensure robustness and generalizability. Model performance will be measured by accuracy, precision, in addition to new metrics designed to capture sustainability and explainability.
- **Iterative Refinement:** Integrate feedback from users and stakeholders to refine the recommenders. Adjustments will be made to accommodate new data inputs, emerging trends, and evolving sustainability objectives.
- **Deployment and Monitoring:** Implement the recommendation system in a real-world context, tracking its performance and long-term impacts. Continuous feedback loops will allow ongoing improvements to maintain alignment with tourist preferences, sustainability goals, and transparency standards.

6 Expected Outcomes and Contributions

This research aims to advance recommendation systems, particularly in sustainable and explainable tourism, by incorporating hierarchical multi-criteria

decision-making. The proposed system will provide balanced travel recommendations that address multiple objectives, guiding tourists to less-visited destinations to reduce over-tourism while promoting sustainable practices. By offering clear, user-friendly explanations, the system fosters trust and encourages responsible travel. Additionally, new evaluation metrics will establish a robust framework for assessing tourism recommendation systems, setting a standard for future research. Also, this study will pave the way for next-generation systems that enhance travel experiences while supporting local communities through fairness , transparency, and environmentally conscious practices.

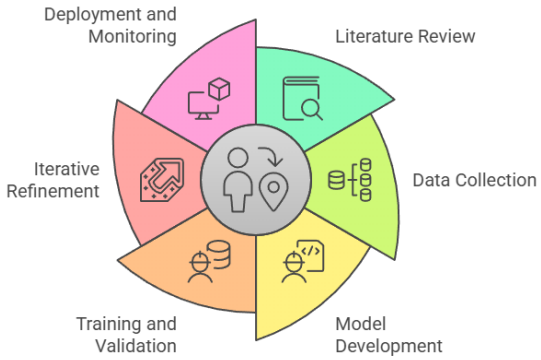

Fig. 1: General recommendation system methodology steps.

Acknowledgement. This work is supported by the Programa Martí Franquès under the grant 2023PMF-PIPF-18

References

[1] A. Banerjee, T. Mahmudov, and W. Wörndl. Green destination recommender: A web application to encourage responsible city trip recommendations. *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*, pages 486–490, 2024.

[2] D. Massimo and F. Ricci. Building effective recommender systems for tourists. *AI Magazine*, 43(2):209–224, 2022.

[3] L. V. Nguyen. OurSCARA: Awareness-based recommendation services for sustainable tourism. *World*, 5(2):471–482, 2024.

Analytical Insights into Structural Super-Diffusion and Indirect Influence in Random Networks

Lluís Torres-Hugas *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili
Tarragona, Spain
lluis.torres@urv.cat

The interplay between network structure and dynamical processes on networks remains a central topic in network science, with diffusion dynamics serving as a key process for studying this relationship due to the model's simplicity. However, despite its importance, few theoretical results are available. In our research, we explore the role of network structure in determining algebraic connectivity, characterized by the second-smallest eigenvalue of the Laplacian matrix, which quantifies diffusion times in complex networks.

In [1], an intriguing behavior was observed in diffusive processes on multiplex networks: under certain conditions, the diffusion time of a multiplex network can be shorter than that of any of its isolated layers. This phenomenon, termed super-diffusion, has motivated extensive efforts to understand and characterize it. For the first time, in [3], we provided an analytical description of super-diffusion, offering new insights and establishing bounds for the super-diffusion region in the connectivity space $k_1 - k_2$ of random duplex networks.

Building on this result, our research extends to a comprehensive analysis of diffusion in networks with indirect connections [2]. Indirect connections enable the spread of information between two nodes, i and k, that are not directly connected but are linked through a third node j in a path of length two. This additional diffusion channel allows information to propagate between nodes i and k without necessarily involving j. In our study, we derive an analytical description of the phase diagram for indirect influence—a novel parameter quantifying how the introduction of indirect diffusive channels enhances the diffusion process—using perturbation theory (see Fig. 1). Furthermore, we identify a new topological phase transition in the abundance of indirect channels at a critical value $p_c \sim N^{(d-1)/d}$.

* PhD advisors: Jordi Duch, Sergio Gómez

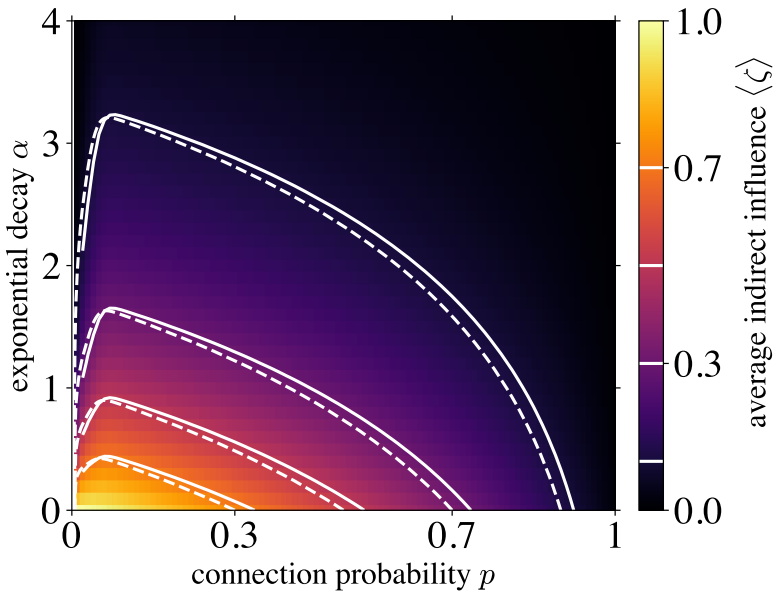

Fig. 1: Experimental phase diagram of indirect influence as a function of the network's connection probability and the exponential decay of indirect connection diffusion rates, governed by $d^{-\alpha}$, where d denotes the shortest path length between nodes. Solid white lines represent the experimental contours of the color map, while dashed white lines indicate the corresponding theoretical predictions.

Acknowledgement. L.T. acknowledges financial support from Universitat Rovira i Virgili and Diputació de Tarragona, Spain (2023PMF-PIPF-21).
This work was supported by Ministerio de Ciencia e Innovación, Spain (PID2021-128005NB-C21, RED2022-134890-T), Generalitat de Catalunya, Spain (2021SGR-633), MICIU/AEI/10.13039/501100011033 FEDER EU (PID2022-142600NB-I00), and Universitat Rovira i Virgili, Spain (2023PFR-URV-00633).

References

[1] Gómez, Sergio, et al. Diffusion dynamics on multiplex networks. *Physical review letters.*, 110.2:028701, 2013.

[2] Miranda, Manuel, et al. Indirect social influence and diffusion of innovations: An experimental approach. *PNAS nexus.* 3.10:409, 2024.

[3] Torres-Hugas, Lluís, et al. Structural prediction of super-diffusion in multiplex networks. *Chaos, Solitons & Fractals*, 186:115265, 2024.

Perceptually Robust Iterative Similarity Momentum Attack

Younas Khan *

Department of Computer Engineering and Mathematics, Universitat Rovira i Virgili Tarragona, Spain
younas.khan@urv.cat

The widespread sharing of facial images on social media platforms has enabled unauthorized facial recognition (FR) systems to massively track and profile individuals without their consent. We introduce PRISMA (Perceptually Robust Iterative Similarity Momentum Attack), whose artificial intelligence contribution is a novel two-stage adversarial attack framework that combines momentum-based optimization with perceptual constraints. The first stage computes robust target features across multiple classes, while the second stage uses these features along with momentum-based optimization to generate visually imperceptible perturbations. The engineering application of PRISMA is privacy protection against unauthorized FR systems. Our empirical evaluation reveals a critical vulnerability in existing protection methods: they can be circumvented through reverse attacks that achieve recognition rates up to 74.64%. In contrast, PRISMA significantly outperforms state-of-the-art methods by reducing FR accuracy to below 26.28% while maintaining high perceptual quality (peak signal-to-noise ratio-human visual system-metric of 35.82 decibels). Our method demonstrates strong transferability across different model architectures and commercial services, which makes it a practical solution for protecting users' privacy.

1 Indications

We present PRISMA (Perceptually Robust Iterative Similarity Momentum Attack), a novel method that fundamentally reimagines adversarial image protection. Unlike existing methods that incur unacceptable trade-offs between privacy and quality, PRISMA uses momentum-based optimization with perceptual constraints to generate visually imperceptible perturbations that nonetheless render images unusable for FR training.
Our empirical results demonstrate that PRISMA significantly outperforms existing approaches in privacy protection while maintaining similar percep-

* PhD advisor: Josep Domingo-Ferrer, David Sánchez

tual quality scores. Our experiments also uncover a critical vulnerability in existing approaches, *i.e.*, Fawkes, and Lowkey: they can be bypassed through reverse attacks that achieve recognition rates up to 69.56%, nearly matching their protected accuracies. Finally, to support the evaluation of privacy protection mechanisms across diverse age groups, we contribute a carefully curated dataset of 416 facial images from 53 underage public figures.

Acknowledgement. This research was funded by the European Commission (project H2020-871042 "SoBigData++"), the Government of Catalonia (ICREA Acadèmia Prizes to J.Domingo-Ferrer and to D. Sánchez), MCIN/AEI/ 10.13039/501100011033 and "ERDF A way of making Europe" under grants PID2021-123637NB-I00 "CURLING" and PRE2019-089210, the EU's NextGenerationEU/PRTR via INCIBE (project "HERMES" and INCIBE-URV cybersecurity chair). Y.K. acknowledges financial support from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 945413 and from the Universitat Rovira i Virgili. The authors thank Dr. Emanuela Podda for helpful discussions on children's privacy protection. This paper reflects only the authors' view and the European Research Executive Agency is not responsible for any use that may be made of the information it contains.

This book contains the proceedings of the 10th Doctoral Workshop in Computer Science and Mathematics - DCSM 2025. It was celebrated in Universitat Rovira i Virgili (URV), Campus Sescelades, Tarragona, on 2025. The aim of this workshop is to promote the dissemination of ideas, methods, and results developed by the students of the PhD program in Computer Science and Mathematics from URV.

Departament d'Enginyeria [DEIM] Informàtica i Matemàtiques Universitat Rovira i Virgili

Escola Tècnica Superior d'Enginyeria Universitat Rovira i Virgili