UNIVERSITÀ DEGLI STUDI DI CATANIA

Dipartimento di Matematica e Informatica

Dottorato di Ricerca in Informatica (Internazionale)

XXXVIII Ciclo

Daniele Marletta

A Framework for the Management and Optimisation of Renewable and Distributed Energy Resources

---

Ph.D. Thesis

Supervisor: Prof. Emiliano Tramontana

Academic Year 2024 - 2025

Abstract

The increased adoption of renewable energy sources is driven by the commitment to reduce greenhouse gas emissions and enhance energy resilience. However, this global transition also introduces significant challenges, including the necessity of effectively assessing environmental impact and handling the intrinsic variability of renewable generation. The main focus of this doctoral thesis is proposing a novel framework to tackle these issues, focusing both on the management of distributed energy resources and on the optimisation of energy and data flows within communities. The framework introduces a systematic methodology for automatic photovoltaic panel detection in aerial images based on the extraction of characterising colours. This approach was further extended to land use and land cover classification, enabling the identification of suitable sites for photovoltaic panel deployment and confirming its general applicability to remote sensing analysis. Subsequently, the framework was extended to address energy fluctuations through a software architecture that integrates a threshold-based strategy for aggregators with an incentive-based demand response programme enhanced with gamification, thereby promoting coordinated and energy-aware behaviour among users. To provide fairness and trust, a blockchain-based data certification mechanism was developed, enabling a cost-efficient and privacy-preserving strategy to ensure data integrity. Finally, the versatility of this approach was demonstrated by its extension to emergency response systems in the healthcare domain, where it provides timestamp certification in a data-intensive scenario. This research's contributions provide a unified solution to support transparent, sustainable and adaptive energy systems. Its applicability to other domains emphasises the broader relevance of this work.

Contents

Chapter 1

Introduction

There has been a growing adoption of renewable energy sources (RES) in recent years, motivated by the need to reduce greenhouse gas emissions, increase energy resilience and promote sustainable development. Distributed energy resources (DERs), such as rooftop solar photovoltaic (PV) systems and battery storage devices, are reshaping traditional energy systems. At the same time, the rise of energy communities provide novel opportunities for more efficient, decentralised and cooperative energy management. However, these new trends also introduce new challenges, including the need for advanced tools to support environmental planning, methodologies to evaluate the impact of DERs in energy systems, strategies to deal with the intrinsic variability of renewable generation, and mechanisms to ensure data integrity in increasingly complex energy ecosystems [41, 53, 81].

This doctoral thesis proposes a novel framework to manage and optimise distributed energy resources utilising renewable sources. The framework tackles challenges that arise both in the planning and assessment of renewable installations, and in the optimisation of energy and data flow within communities. The first challenge that the framework addresses is the automatic detection of solar PV panels in aerial images [70, 89] using a fast, flexible methodology that does not require

large annotated datasets. This approach forgoes computationally and data-intensive training phases by automatically extracting a set of PV panel characterising colours from pixel-level analysis of only a few images. This strategy achieves high detection performance and simultaneously retains high robustness when applied to different geographical areas, thus overcoming the problem of overfitting to specific regions. Furthermore, the methodology can be configured to detect both individual PV panels and entire PV installation deployment areas, providing stakeholders and policymakers with effective tools to assess solar power generation and its environmental impact.

The characterising colours-based strategy was subsequently extended to enhance the framework with the ability to classify land use and land cover (LULC) in remote sensing images [58, 88]. "Land cover" refers to the natural composition of a landscape, such as vegetation, soil, and water bodies, whereas "land use" encompasses all areas inhabited or utilised by humans, such as buildings and industrial facilities. Within the framework, LULC classification represents a natural extension of the PV panel detection task: in the latter, the goal is to identify a single class (PV panels), whereas in LULC classification the methodology is applied to identify multiple categories (i.e., the types of land). When adapted to the LULC classification task, the characterising colours methodology retains high performance, while maintaining the same key advantages, including fast execution times, no need for large annotated datasets and forgoing lengthy training procedures. Hence, the robustness and flexibility of the proposed approach are demonstrated, and its general applicability to remote sensing image analysis is confirmed. Importantly, since studies indicate that PV panels are most commonly installed on cropland, followed by arid regions and grassland [56], LULC classification further strengthens the framework by providing the capability to identify suitable locations for solar farms. Therefore, the framework represents a

unified solution that supports the planning of distributed energy resources in different environments.

Building on these contributions in asset management, the framework discussed in this thesis was further developed to address the optimisation of energy and data flow from DERs. One of the most significant challenges is represented by the inherent variability of RES, as their output depends on changing natural and meteorological conditions. These unpredictable fluctuations pose several challenges, including maintaining grid stability and balancing electricity supply with demand [130]. In energy communities, an entity such as an aggregator combines the consumption and generation profiles of members, assessing the net energy balance. However, a proper strategy is needed in order to effectively handle fluctuations and optimise energy flows. Moreover, although users may change their energy load in exchange for financial rewards such as incentive-based demand response (DR) programmes, participation is usually on a voluntary basis [103, 114]. Consequently, there is a need to make participation more appealing and to coordinate user responses in a fair and transparent manner.

To address these shortcomings, the framework presented in this thesis was further extended by designing a software architecture that supports the optimisation of energy usage patterns in energy communities. Specifically, the proposed system defines a threshold-based strategy for the aggregator, enabling it to detect anomalous fluctuations hindering grid reliability. To encourage users to modify their usage patterns, the framework integrates an incentive-based DR programme combined with a gamified application that promotes coordinated and energy-aware decisions aligned with the community's welfare.

Whenever financial rewards are distributed, disputes can arise. For example, a

user may contest the reward amount and claim eligibility to a higher remuneration. Moreover, system operators need to ensure that incentives are correctly assigned, and a strategy must be put in place to handle the risk of data tampering by malicious entities. Consequently, it is paramount to obtain a methodology to certify data integrity in such scenarios.

To address these challenges, the framework was extended with the design of a blockchain-based certification mechanism [38, 139]. By storing on-chain hashes of energy and DR-related data, the proposed solution provides an immutable and verifiable audit trail while also protecting user privacy. This solution provides a cost-efficient method to solve disputes and invalidates data manipulation attempts, since tampering is easily detected.

The importance of certifying data integrity is not limited to energy systems. Other domains, such as healthcare, can also greatly benefit from a methodology that ensures correctness and verifiability of data [80, 95]. For example, in emergency response systems there are legal requirements to prove the timeline of events, and a decentralised certification method can provide trust without reliance on a central authority. To tackle these challenges and prove the reusability and versatility of the proposed data integrity certification approach, the blockchain-based methodology of the framework was adapted to handle timestamp certification in emergency response systems. Additional data handling strategies were defined to take into account the increased event variability in this healthcare scenario, and achieve optimal balance between system performance and economic feasibility.

Overall, the framework presented in this doctoral thesis represents a comprehensive solution that supports the planning, deployment, management, monitoring and optimisation of DERs. The research provides a valuable contribution towards

promoting sustainable energy systems and supporting transparent, trustworthy and efficient practices. Moreover, the application of the framework to the healthcare domain illustrates its conceptual and practical relevance beyond the energy sector.

This doctoral thesis is structured as follows. Chapter 2 discusses the literature and works that are related to the contributions of this doctoral thesis. Chapter 3 introduces the framework for DERs management and planning, presenting the PV panel detection approach in Section 3.1 and the LULC classification strategy in Section 3.2. Chapter 4 discusses how the developed framework addresses the optimisation of energy and data flow in renewable energy communities, with Section 4.1 presenting the proposed architecture for a gamified and incentive-based strategy for coordinated and adaptive energy optimisation, and Section 4.2 detailing the blockchain-based approach for providing trust and data integrity in demand response programmes. Chapter 5 discusses how the framework was extended to provide data certification capabilities in the healthcare scenario, considering the emergency response system as a case study. Finally, Chapter 6 draws the conclusions of this doctoral thesis.

1.1 List of Publications

The research described in this thesis has been presented in the following publications and submitted articles:

- D. Marletta, A. Midolo, E. Tramontana, *Detecting Photovoltaic Panels in Aerial Images by Means of Characterising Colours*, 2023, Technologies (DOI)
- D. Marletta, A. Midolo, E. Tramontana, *Automatic Land Use and Land Cover Classification by Means of Characterising Colours*, 2024, IEEE 32nd International WETICE Conference (DOI)
- D. Marletta, A. Midolo, E. Tramontana, *Automatically Injecting Robustness Statements into Distributed Applications*, 2024, Future Internet (DOI)
- D. Marletta, A. Midolo, E. Tramontana, *A Blockchain-Based Strategy for Certifying Timestamps in a Distributed Healthcare Emergency Response Systems*, 2025, Future Internet (DOI)
- D. Marletta, E. Scaletta, E. Tramontana, *User Engagement and Adaptive Optimisation in Renewable Energy Communities*, 2025, International Workshop on Analytics for Software Product and Process Improvement (accepted)
- D. Marletta, A. Midolo, E. Tramontana, *Certifying Energy Data in Demand Response Programmes by means of Blockchain*, 2025, (draft)

Chapter 2

Related Work

This chapter discusses works presented in literature that are relevant to the two main aspect of this doctoral thesis: the management of distributed energy resources through aerial imagery analysis, and the optimisation and certification of energy and data flows within energy communities. For each topic, the goal of these analyses is to identify key results reported by researchers , summarise their strengths and limitations, compare them with our approach and emphasise research gaps that informed the design choices and contributions developed in this thesis.

The remainder of this chapter is organised as follows. Section 2.1 presents related work in the field of PV panel detection. Section 2.2 analyses LULC classification approaches and compares them with our methodology. Section 2.3 surveys the relevant literature for incentive-based strategies to promote user engagement and compliance to requests of load reduction by energy aggregators or operators. Finally, Section 2.4 discusses previous approaches for certifying data integrity using the blockchain technology.

2.1 Photovoltaic Panel Detection in Aerial Images

Estimating the number of PV panels in a region is a complex task due to the insufficiency (or even lack of) official registers. Many papers have proposed approaches to detect PV systems by analysing satellite and aerial images, often using Random Forest (RF) or Convolutional Neural Networks (CNN) classifiers.

A deep neural network model called Faster-RCNN was used to design the identification model of PV panels [31]. The approach consisted of two parts: first, a ResNet-50 classifier was pretrained, then a CNN was fine-tuned for the identification task of rooftop PV panels. Similarly, three convolutional layers and three fully connected layers were used to evaluate the performance of the identification [30]. Moreover, eight 2D convolutional layers were used to detect PV panels in residential areas; to achieve the best performance, thirteen architectures were trained and the most accurate was selected [90]. Other approaches have used InceptionV3, a CNN used for image analysis and object detection, which was fine-tuned for the task of PV panel identification [45, 140]. These approaches were designed to detect PV panels in both residential and non-residential areas; however, due to the lack of PV panel images, data augmentation was performed during the training process. The framework proposed in [140] was employed for the detection of PV panels in Sweden to collect further market statistics [66]. Similarly, an innovative approach was presented to detect rooftop PV panels on the three-dimensional (3D) orientation [77]. This approach employed the InceptionV3 model to classify images; subsequently, segmentation and geocoding steps were performed to analyse the 3D images. ML and deep learning techniques were used for rooftops PV panels detection in [63]. The k-means approach was applied to segment the images in order to define the contours of each rooftop, then a support vector machine (SVM) classifier with a

CNN was integrated to accurately identify solar PV arrays. A Mask-RCNN was used for segmentation and identification in [64, 89, 111]. These approaches applied the object detection technique to reveal PV panels on aerial images, with CNN being fine-tuned to characterise the mask contours used for the arrays. A CNN with the VGG16 encoder was presented in [20]; first, image segmentation was performed to select the suitable portions of solar panels, then the azimuth of the solar arrays was predicted using edge detection and the Hough transform.

Despite the fact that, generally, CNNs obtain high performance, they need a massive amount of labelled data during the training process, and several runs are required to properly fine-tune a model. Moreover, excessive training can lead to overfitting, meaning that the model fails to generalise. Additionally, CNNs are often considered black boxes, that is, hard to interpret and comprehend. Therefore, the debugging and validating process can be complex and time-consuming, and explainability about the model's decisions remains challenging. These approaches are very costly both when processing the dataset and when running the model during the training phase.

A different approach was proposed in [70] to extract image features such as colours, textures, and other patterns from each pixel, then pass them as input to train an RF classifier to identify pixels related to PV arrays. In a similar approach [138], the focus was on the identification of water PV systems (WPV); an RF classifier with 400 trees was trained to extract pixels related to WPV, then post-processing was performed to remove noise and rooftop PV panels. Another pixel-based RF algorithm used the L-8 surface reflectance (SR) product to identify suitable PV panels [148]. The RF classifier was based on the Google Earth Engine (GEE) and used to map PV power plants. Similarly, a RF classifier for an Object-Based Image Analysis

(OBIA) approach used different combinations of multispectral Sentinel-2 imagery and radar backscatter from Sentinel-1 SAR imagery [127]. RF algorithms use a lot of computational power, and need additional resources to build numerous trees and combine their outputs. Furthermore, the training process is time-consuming and resource intensive, as it needs to combine different decision trees to determine the class for the identification.

In [69], RF classification was combined with a CNN. First, the RF was used to assign a confidence value to each pixel in order to determine the possibility of that pixel belonging to a solar PV array; then, a CNN was used to classify 40×40 patches of RGB images to determine whether or not they corresponded to solar PV panels. An innovative deep learning technique called EfficientNet-B7 was employed for PV panel detection in [98], showing better accuracy and efficiency compared to classic CNN approaches. EfficientNet-B7 was used as a backbone encoder to train a U-Net model for segmenting solar panels.

Spectral characteristics have been investigated to detect PV panels from hyperspectral data [18, 49] by focusing on the physical absorption and reflection characteristics of PV panels. To handle the material diversity of PV panel types, these studies applied a tailored image spectral library, which together with the hydrocarbon index mitigated the spectral variance caused by the detection angle. The results of these approaches showed that the shape and the area of the PV panels were not accurate. Moreover, the identification was sensitive to the thresholds set for the spectral bands of different sensors.

Overall, previous machine learning and deep learning methodologies require large annotated datasets and a lot of computational resources. Moreover, the training process for these models is time-consuming, prone to overfitting and lacks

explainability. On the other hand, hyperspectral analysis approaches are complex and exhibit limitations in accurately detecting PV panel surfaces. To overcome these challenges, our approach proposes several innovative aspects: (i) due to the absence of a training process, the execution is fast, it is not computationally intensive, and it does not require a large amount of annotated data; (ii) a set of characterising colours for PV panels is automatically extracted and used to perform the detection, and this set can be easily updated to adapt the detection task for images taken in different geographical areas or for images featuring different types of panels; (iii) it can be used to identify both the PV panel surfaces and the broader PV panels deployment area, with only minimal changes to the algorithm's parameters; (iv) the proposed approach makes it possible to explain the resulting outcomes, that is, the reason behind the detection (or lack of detection) of PV panels.

2.2 Land Use and Land Cover Classification

The Land Use and Land Cover (LULC) classification task has attracted a lot of attention in recent times, due to the changes brought about by climate change. Many articles presented in literature have proposed approaches to analyse satellite and hyperspectral aerial images to classify land categories according to their use and morphology.

A convolutional Machine Learning (ML) model was proposed to map land cover on satellite images, trained on a large dataset of images called BigEarthNet [124]. Different algorithms and classifiers were evaluated, where the surface reflectance of the images in the dataset Landsat 8 was used to classify images according to land use [11]. Similarly, six ML algorithms were used in [121]: random forest,

support vector machines, artificial neural network, fuzzy adaptive resonance theory-supervised predictive mapping, spectral angle mapper and Mahalanobis distance. In [4], a comparison between support vector machine, a maximum likelihood estimator and an artificial neural network was carried out to identify the best algorithm for this specific task. A supervised approach with a maximum likelihood classifier was used in [2]. In [15], a neural network with two convolutional layers and a fully connected layer was trained on multispectral images to classify land cover and land use. Similarly, a pretrained model was presented in [112].

Deep learning (DL) and ML approaches generally obtained high performances; however, a huge amount of annotated data was needed to properly train the models, requiring, in addition, several runs to properly tune the model. However, the risk of overfitting poses a challenge, resulting in the model's inability to generalise when presented with out of distribution data. Moreover, for convolutional neural networks (CNNs) the debugging and validation phases can become intricate and time-consuming, with model-driven decisions being the ultimate determinants. Such approaches incur significant costs both for the dataset pre-processing and during training.

Hyperspectral imaging was widely used for LULC classification tasks [88]. Random forests [58, 101] and support vector machines [52] were widely used for such types of images, showing good performances. In [78], in addition to these two models, a light gradient boosted machine was used to achieve better accuracy and precision for the classification task. DL approaches with CNNs have been employed for analysing hyperspectral images [125, 145]. In [67], a model was proposed to effectively capture long-range dependencies and extract local features from hyperspectral images. This model combines an interactive transformer with a CNN, allowing them to interact

adaptively with each other. In [129], an approach called the cascaded residual dilated network was introduced to tackle the complex task of LULC classification utilising very high spatial resolution images.

In general, DL-based approaches have to deal with high processing time and computational complexity. Additionally, obtaining an adequate number of training images is considered a crucial challenge in DL-based LULC classification, as a substantial amount of images is needed to ensure optimal learning procedures [146]. Furthermore, hyperspectral images required high costs for data gathering, and presented a limited annotated training sample. Besides, spectral bands are often irrelevant and redundant, leading to the common problem of the Hughes phenomenon [37, 87]. This phenomenon occurs when the number of spectral bands increases, since data becomes more sparse, computational complexity rises, and visualisation becomes more challenging. Dimensionality reduction is then often performed to mitigate such challenges, however it results in a more complex approach and the need of further processing and computational resources.

In contrast, our approach introduces several innovative aspects: (i) by eliminating the need for a training process, our method exhibits fast execution times without needing many computational resources, and it does not require have a large annotated image dataset; (ii) the automatic extraction of characterising colour sets enables the approach to be applied to all types of land exhibiting colours that belongs to the corresponding set, and allows to expand the detection capabilities by defining additional land categories; finally, (iii) our proposed algorithm provides explainability, i.e. it gives insights into the classification (or absence thereof) of lands.

2.3 Incentive-based Strategies for Optimal Energy Balance

Several research papers have attempted to address the challenge of achieving energy balance in systems with a high share of renewables, and proposed several strategies to handle the inherent variability of production and demand. Analyses based on periodic behavioural patterns of production and consumption allowed the formulation of criteria to ensure power grid stability and net energy balance in energy communities [47]. Focusing on short-term energy balance, an optimisation model was proposed for hybrid energy hubs, which operates at 15-minute intervals. The approach focused on cost minimisation while ensuring efficient integration of PV and storage at the building level [110]. Although the model achieved cost savings and a reduction in non-renewable energy consumption, strategies for leveraging coordinated response at a broader community level were not defined.

A dynamic energy price strategy in a simulated microgrid was proposed to achieve efficient balance between demand and generation from seven renewable sources [3]. Other works address technical and economic aspects of distributed power systems, such as voltage control [131] and aggregator-based coordination to maximise profit and decrease cost [35]. These works focussed on optimising financial benefits for prosumers while maintaining energy balance, but they lack mechanisms for sustained user engagement, which is an important aspect under dynamic pricing schemes [92].

In addition to these technical contributions, several studies have explored gamification as a tool for promoting energy-related behavioural change. Some works presented strategies for active user participation in smart communities through tailored interfaces and social competition [82], while others analyse consumer participation in

virtual power plants using behavioural models and gamified applications [13]. An interdisciplinary framework was also proposed to promote residential user engagement through gamification techniques [5]. The adopted behavioural model (transtheoretical model) defines the requirements for behavioural change, which are supported by a technical architecture and game design elements classified into five categories. These works highlight the benefit of gamified approaches to promote coordination and energy-aware choices, but the impact of financial incentives and adaptive energy optimisation strategies remains underexplored.

Demand Response (DR) schemes have also been proposed to stimulate end users to modify their energy usage patterns. DR programmes can be broadly divided into two main classes [6]: price-based and incentive-based. In the price-based approach, electricity price is dynamically modified depending on the congestion status of the grid. This strategy attempts to indirectly influence consumers to reduce load demand by making the purchase of energy less economically appealing during the peak utilisation period, when the price becomes significantly higher. On the other hand, incentive-based DR programmes directly offer financial rewards to users who agree to reduce their consumption in response to a signal from system operators. Several approaches have been presented in the literature to design and implement effective DR schemes, including: electricity price forecasting [28, 97], baseline estimation [132, 133], forecasting of residential appliances load demand [123] and forecasting of PV power generation [29, 46]. While these approaches have shown promising results, testing and empirical validation in real-world scenarios is lacking.

Many studies have proposed the integration of public blockchain with DR programmes to enhance and accelerate their adoption [55]. Pop et al. [104] presented a blockchain-based system to configure multiple levels on energy flexibility and to

validate DR terms. Mao et al. [73] proposed a methodology to handle bidding transactions deployed on the Ethereum blockchain, which improves transparency of DR transactions. Park et al. [99] introduced a privacy-preserving scheme called BPPS which demonstrated high resilience against several attacks and ensured data integrity. Wu et al. [135] leveraged blockchain to securely and transparently store energy and pricing data, enabling the definition of a dynamic DR scheme that adapts prices depending on net energy balance. Van Cutsem et al. [126] defined a decentralised system for coordinating prosumers in different buildings. Their approach proved successful in ensuring a net reduction during peak demand. While the results from these works demonstrate the wide applicability of the blockchain technology in DR programmes, more investigations are needed to assess the scalability and efficiency of the proposed energy systems. Moreover, the integration of these approaches with IoT devices such as smart meters remains limited, and a comprehensive cost assessment was not conducted. Finally, blockchain-based approaches that propose to store the complete set of data on-chain incur in potentially high operational costs and threats to user privacy.

Previous approaches have focussed on several aspects related to energy optimisation, investigating the impact of dynamic pricing, gamification, and DR programmes on promoting energy-aware behaviour among customers. However, they lack mechanisms for sustained user engagement and coordinated response on a community scale. Furthermore, the integration of incentives with adaptive energy management strategies was not analysed in depth. Finally, previous blockchain-based models lack a comprehensive cost assessment, and the protection of user privacy remains limited.

To overcome these shortcomings, the framework presented in this doctoral thesis

was extended to provide a unified model combining gamified engagement with data-driven and flexible optimisation strategies in the context of energy communities. Our analysis, conducted on real-world datasets, shows that a coordinated response built on observed user patterns and targeted incentives can achieve optimal energy management while promoting community welfare. Moreover, the blockchain-based system that we designed employs a cost-efficient strategy by storing only a 32-byte data hash on-chain, optimising costs associated with blockchain interactions. Furthermore, user privacy is ensured because plaintext energy data are not stored on the public blockchain. Finally, our cost evaluation conducted on a real-world dataset demonstrated the feasibility of our approach.

2.4 Blockchain-based Data Certification

Several approaches have been proposed to ensure the integrity and authenticity of long-term time-stamping (LTTS) systems, which leverage cryptographic techniques, trusted third parties, and decentralised architectures [12]. While some focus on formal security models, others explore blockchain-based solutions to eliminate central authorities. Meng et al. [85] introduced a formal security model for LTTS schemes based on Message Authentication Codes (MACs), archives, and transient keys, addressing the lack of cryptographic analysis in alternative approaches. Despite enhancing robustness against cryptographic advancements, their solution depends on trusted third parties (e.g., archives and TSAs), incurs storage overhead, and requires frequent cryptographic updates. Similarly, Vigil et al. [128] explored decentralised timestamping for long-term digital archives but still relied on timestamping servers, making their approach vulnerable to centralisation risks.

Blockchain-based timestamping has gained traction as an alternative to centralised solutions. Meng and Chen [83] proposed a scheme that records hash digests on a private blockchain, ensuring long-term verifiability without relying on external authorities. However, a private blockchain cannot provide guarantees as a public one, and the transaction delays introduced by the consensus mechanism and the risk of loss of keys affecting verification pose significant challenges. Wilson [134] extended this concept by implementing a permissioned blockchain for secure document timestamping, integrating off-chain storage and a quarantine mechanism. Although document management was enhanced, centralised control over the blockchain raises concerns regarding governance. Similarly, Zhang et al. [149] leveraged Ethereum blockchain to leverage transactions for timestamping, benefitting from its immutability but introducing privacy risks and high operational costs.

Scalability and computational efficiency remain key concerns in long-term trusted timestamping (LTTS). Bin et al. [105] introduced IoETTS, a decentralised scheme for the Internet of Energy (IoE), leveraging public blockchain to improve data integrity. However, the computational and energy overhead associated with blockchain operations, particularly those relying on Proof of Work consensus, conflicts with the efficiency goals of IoE. Similarly, Zhang et al. [147] proposed a scalable LTTS scheme using commitment signatures and bilinear pairing accumulators to reduce transaction costs, but their approach lacks analysis of real-world deployment. In addition to system implementations, theoretical analyses of LTTS security and efficiency have emerged. Meyer [86] formalised timestamping mechanisms using transitive prefix authentication graphs, providing a structured comparison of efficiency and security trade-offs. However, maintaining authenticated prefix structures introduces storage and computational overhead, limiting large-scale applicability. Meng et al. [84]

examined client-side security risks, identifying vulnerabilities in key management and cryptographic resilience, but did not offer concrete mitigation strategies.

Blockchain technology has been used in several contexts where it is fundamental that data remain unchanged, tamper-proof, and provable [96, 109, 144]. In the healthcare domain, the need for advanced technologies is rapidly increasing, with blockchain playing a key role in the transformation of the sector [39, 71, 122].

Odeh et al. [95] provided an overview of blockchain-based applications in healthcare, focusing on its ability to improve electronic health records (EHR), enhance data security, and support patient monitoring and drug traceability. They emphasised blockchain ability to ensure that only authorised parties can access sensitive medical data. Le et al. [60] proposed a permissioned blockchain system that uses the Hyperledger Fabric platform to allow quick and secure access to health records during emergencies. The system enables patients to define access rules for their data, which are enforced through smart contracts.

Suthaputchakun et al. [120] explored blockchain's potential to improve communication and security in emergency rescue operations, particularly in ambulance-to-everything communications. Their system provides secure data exchange between ambulances, hospitals, and other emergency services, preventing tampering and ensuring authenticity. Ksibi et al. [57] designed a blockchain-based system that integrates the Internet of Vehicles and the Internet of Medical Things to enhance emergency medical responses in smart cities, allowing the real-time transmission of accident victim data to nearby healthcare services.

Yue et al. [143] introduced a public blockchain architecture that empowers patients to control their healthcare data while ensuring privacy and enabling secure sharing among different healthcare providers. This decentralised approach improves

data access and prevents unauthorised access and tampering. Griggs et al. [34] proposed a system that integrates a private blockchain with IoT medical devices, using smart contracts to automate notifications for medical interventions and ensure HIPAA-compliant data handling.

Lin et al. [65] presented a framework designed to address security concerns by ensuring secure mutual authentication and fine-grained access control. This framework leverages blockchain's capabilities to provide confidentiality, auditability, and scalability, making it suitable for dynamic environments such as smart factories. Fernandez-Aleman et al. [26] introduced a blockchain-based system for managing audit logs among healthcare organisations, enabling interoperability and ensuring the integrity of log data. Nugent et al. [93] leveraged blockchain's tamper-resistant features to secure clinical trial data, using Ethereum smart contracts to automate updates and ensure data transparency using a permissioned blockchain.

Benchoufi et al. [14] discussed how blockchain can address challenges in clinical research, such as data reproducibility, sharing, privacy, and patient enrollment. They emphasise blockchain's decentralised, transparent, and tamper-resistant nature, which is crucial for maintaining data integrity in clinical trials. Zheng et al. [151] propose a hybrid blockchain-cloud system for secure sharing of personal health data, using blockchain for transaction management and validation, while cloud storage handles large datasets.

Shi et al. [113] reviewed how blockchain technology was used to enhance the security and privacy of electronic health record (EHR) systems. They highlighted critical limitations of traditional EHR systems, such as data breaches, unauthorised access, and lack of transparency, and proposed that blockchain decentralisation, immutability, and access control capabilities address these challenges. Practical use

cases include secure patient data sharing, tamper-proof audit trails, and privacy enforcement through smart contracts. However, the study acknowledges significant barriers to adoption, including scalability, high computational costs, integration complexity, and regulatory constraints.

Xia et al. [137] proposed MeDShare, a blockchain-based system designed to facilitate medical data sharing in a trust-less environment. MeDShare ensures data provenance, auditing, and control for shared medical data in cloud repositories among big data entities. The system monitors entities accessing data for malicious use and records all data transitions and sharing activities in a tamper-proof manner. It employs smart contracts and an access control mechanism to effectively track data behaviour and revoke access to offending entities upon detecting permission violations. Similarly, Xia et al. [136] addressed the challenges of securely sharing electronic medical records (EMRs) in cloud environments, particularly concerning patient privacy and data breaches. The authors proposed a permissioned blockchain-based data-sharing framework that leverages the immutability and autonomy properties of blockchain technology to improved access control for sensitive data stored in the cloud. Their system utilises a permissioned blockchain, granting access exclusively to invited and verified users, thus ensuring accountability by maintaining a log of all user actions.

Azaria et al. [9] introduced a private blockchain-based system designed to streamline access management and permissions for medical data. It addresses the fragmentation and inaccessibility of medical records by enabling patients to retain control over their data while providing secure, auditable access to healthcare providers. However, MedRec faces several challenges, including scalability limitations, high transaction costs, privacy risks from metadata exposure, and the complexity of patient and

provider adoption. Additionally, regulatory compliance, energy consumption due to Ethereum's Proof of Work, and integration with existing EHR systems pose significant barriers to real-world deployment. Patientory [79] is a blockchain-based peer-to-peer network for securely storing and sharing electronic medical records (EMRs). It integrates with existing EHR systems, allowing patients and healthcare providers to access medical data while ensuring security and compliance through smart contracts and the token economy. However, the approach faces several challenges, including scalability limitations, high transaction costs, privacy risks from off-chain storage, and regulatory compliance issues, particularly with data immutability.

Previous studies, reported above, have explored a wide range of blockchain-based applications in healthcare, primarily focusing on authentication, data privacy, access control, and data integrity. However, privacy risks and scalability limitations remain significant challenges, restricting the real-world applicability of these approaches. In contrast, our approach innovatively leverages blockchain technology to ensure the authenticity and integrity of event timestamps by storing only their hashes. This design choice significantly reduces blockchain costs while enhancing system maintainability, addressing a key limitation of existing research, where escalating costs hinder scalability [48]. Furthermore, our framework introduces event filtering mechanisms, allowing the selection and storage of only relevant events to optimise resource usage. It also incorporates data aggregation techniques to efficiently group multiple event records, further minimising blockchain transaction costs. To validate its practicality, we conducted extensive experiments on real-world datasets, analyzing both transaction costs and the time required to securely store hashes on the blockchain. These strategies ensure that our solution is effective, cost-efficient, and feasible for the targeted medical domain.

Chapter 3

Managing Distributed Energy Resources via Image Analysis

In recent years, there has been a growing adoption of renewable energy sources, whose share is projected to increase from 30% in 2023 to 46% in 2030 [44]. In particular, the global cumulative photovoltaic (PV) solar capacity rose dramatically from 1.6 TW in 2023 to over 2.2 TW at the end of 2024, and annual growth has remained above 35% for the past ten years [43].

In this context, the assessment of total PV capacity using image-based PV panel detection offers several advantages. First, this information can be used by stakeholders in the energy sector for the management and forecasting of electricity flows [41]. Moreover, it can be used to track the adoption of solar PV at local and global scales [66]. Finally, automatic detection facilitates the identification of regions that could benefit from expanded solar PV coverage [8]. This last point is particularly important because Land Use and Land Cover (LULC) analyses show that PV panels are most frequently deployed on cropland, followed by arid regions and grassland [56]. For this reason, LULC classification has been applied to identify suitable locations for solar farms [25, 59, 106].

Many previous studies have proposed Machine Learning and Deep Learning approaches to analyse aerial imagery for PV panel detection [20, 45, 90, 138, 148] and for LULC classification [11, 15, 52, 58, 124]. However, these methodologies typically require large annotated datasets, and their training phases are costly, time-consuming and prone to overfitting, which results in the inability to generalise when applied to different data sources.

Other approaches have used multispectral and hyperspectral data to identify PV panels [18, 49] and LULC classes [67, 88]. However, hyperspectral data are complex because they span numerous spectral bands. Moreover, the process of creating accurate annotations requires expertise and may involve sophisticated analysis techniques, such as spectral classification or feature extraction. This often results in a limited number of available labelled data, which in turn has a negative impact on the performance of deep learning methods.

To address these challenges and provide an efficient and flexible methodology for the management of DERs, we designed a framework for detecting PV panels and LULC categories in aerial images using automatically extracted *characterising colours* (CC). A CC set is a small collection of colours that are most commonly found within a specific category (e.g. PV panels, land covers) and not in others (e.g. background). Therefore, this set of colours can be said to characterise the corresponding category.

To detect PV panels, we designed a deterministic algorithm that analyses the colours of a few annotated aerial images of PV panel installations to identify a specific set of characterising PV panel colours. Subsequently, this set is used to detect the locations of PV arrays in other, unannotated, remote sensing images. The proposed approach includes a pre-processing phase, which analyses and automatically

classifies dataset images into multiple classes according to prominent PV panel colours. Moreover, a post-processing phase enhances detection results by automatically identifying and eliminating incorrectly marked background areas.

The results show that our methodology is fast and moreover it can detect both the surfaces of individual panels and the full deployment areas of PV arrays, requiring only minimal changes to its configuration parameters. Additionally, the CC set can be readily applied to new images without additional tuning, which supports the robustness of the approach.

The same principle of using characterising colours was applied for LULC classification in aerial images. LULC classification represents a natural extension of our PV panel detection approach: in both tasks the goal is to identify specific regions within an image, but in LULC classification the methodology is applied to identify multiple categories rather than a single class.

We developed a robust and explainable approach to classify land cover based on the characterising colours of each land type. In our experiments, we considered three land types: forest, irrigated land, and non-irrigated land. Similar to the PV panel detection approach, the proposed LULC classification methodology begins by extracting and analysing colours from a small set of annotated images to obtain CC sets for each land type. Subsequently, these colour sets are used to mark pixels in unannotated images based on the LULC category they belong to. Pixels are then clustered based on their density, and a final post-processing phase removes incorrect detections and refines the classification results.

The key strengths of our approach based on characterising colours are: (i) it does not require large annotated datasets for creating CC sets; (ii) the colour extraction phase is fast and does not require computationally intensive resources; (iii) the

CC sets can be readily applied to unannotated images and to different datasets without further tuning; (iv) post-processing automatically identifies and handles false negatives and false positives, improving detection results; (v) the methodology is explainable: for each category, the distribution of characterising colours is provided and directly relates to the final classification results.

To evaluate our approach for both the PV panel detection and LULC classification tasks, we conducted experiments on open-source datasets of aerial imagery. We computed standard metrics, including accuracy, recall and precision, to demonstrate the effectiveness of the method. Our findings confirm that a relatively small subset of characterising colours can be used effectively to detect PV panels and specific land covers. The results for PV panel detection were presented in [76], and the results for LULC classification were presented in [75].

The remainder of this chapter is structured as follows. Section 3.1 details our methodology for detecting PV panels in aerial images, while Section 3.2 describes our approach for automatic LULC classification.

3.1 Detecting Photovoltaic Panels in Aerial Images by Means of Characterising Colours

The integration of renewable energy sources, such as solar power harnessed through PV panels, has contributed to diminished reliance on conventional fossil fuel-based power generation facilities [19, 27]. PV systems are one of the most promising low-carbon energy generation methods [115], and data shows that PV energy production has grown rapidly over the last decade, at a rate of more than 35% annually [10, 32].

At the end of 2022, the world's cumulative installed PV capacity was 1055.03 GW, compared to 589.43 GW in 2019, almost doubling in three years [1].

Estimating the total installed PV capacity and power generation can enhance the ability of policymakers and stakeholders to evaluate progress in terms of sustainability, quantify the actual benefits of green energy, and consider potential future installations [72]. Aerial and satellite images have been analysed to recognise PV panels using Machine Learning (ML) approaches such as such random forests [69, 70, 138, 148], and Deep Learning (DL) methods such as convolutional neural networks (CNN) [20, 30, 31, 45, 63, 64, 89, 90, 111, 140].

However, ML and DL methodologies require a large amount of annotated datasets for their training to be effective, and the corresponding effort to build datasets and perform training is costly and time-consuming. Moreover, many such approaches rely on a training phase that can cause overfitting, resulting in the inability of the ML and DL-based model to generalise when applied to different datasets.

Further approaches have focused on analysing the physical absorption and reflection characteristics of PV panels to identify them from airborne images [18, 49]. The presented results show that the identification of shapes and areas of PV panels is often inaccurate. Additionally, spectral detection is highly sensitive to the thresholds used for the spectral bands of the different sensors.

Within the framework presented in this doctoral thesis, we propose an approach that identifies PV panels by means of a deterministic algorithm that extensively analyses the pixel colours. The proposed approach firstly analyses images to reveal potential overlapping between PV panel colours and colours found in background areas. Based on the results of this analysis, images are classified into multiple categories depending on the prominent PV panel colours. Subsequently, pixel colours

in selected annotated images are extracted and filtered by removing background colours that are typically found in the surroundings of PV panels. Specifically, colour shades commonly found in very dark portions (shadows, ground, etc.) and very bright portions (roads, roofs, etc.) are excluded from the characterising colours of PV panels. The resulting colour selection is then used to analyse colours in unannotated images and identify PV panels. Additionally, to enhance the detection, the density of pixels identified as belonging to the surface of PV panels is evaluated.

Our proposed approach does not require large annotated datasets or computationally intensive training. The characterising colours of PV panels can be obtained from a small dataset using a low amount of computational resources, then readily used to identify PV panels in other images from a similar geographical area. The approach is deterministic, and can be adapted to perform different detection tasks with only minimal changes to its configuration parameters. Moreover, the results of the detection tool are explainable. For example, a lower recall value can be explained by the number of selected characterising colours, threshold for dark pixels, and so on. The approach has been validated on a publicly available annotated dataset of remote sensing images, and several metrics, including accuracy, recall, and precision were computed to demonstrate that the approach manages to locate PV panels more accurately than previous works in recent literature.

The key advantages of our approach are: (i) it has fast execution times; (ii) it does not require a large dataset of annotated data; (iii) the obtained charactering colour sets can be readily applied to perform detection on both unannotated images and different datasets; (iv) it automatically identifies and isolates images featuring PV panels displaying pixels with high colour saturation or low brightness levels, which typically do not yield characterising colours; and (v) it automatically identifies

and handles false negatives and false positives based on the density of pixels marked as belonging to PV panels.

The remainder of this section is structured as follows. Section 3.1.1 provides an overview of our proposed approach for PV panels detection. Section 3.1.2 describes the first phase of the approach, aimed at classifying images based on the prominent colours of PV panels. Section 3.1.3 presents the methodology used to extract the characterising PV colour set. Section 3.1.4 details the detection algorithm. Section 3.1.5 explains the technique employed to handle noise in the images. Section 3.1.6 describes the experiments performed to test and validate our approach. Section 3.1.7 discusses the results. Finally, conclusions are drawn in Section 3.1.8.

3.1.1 Proposed Approach

The main strategy of the proposed approach involves analysing the colours of a few annotated aerial images displaying PV panel installations (hereafter referred to as PV images) to identify a specific set of *characterising* PV panel colours. Following this, the set of characterising colours serves as a reference to detect the location of PV arrays in other, unannotated, remote sensing images. Finally, the detection results are further refined with a post-processing operation that identifies and removes incorrect detections.

Our detection strategy builds upon and extends the methodology presented in [117, 118], where colour analysis and scanning windows were employed to automatically determine green and urban areas in remote sensing images and trace their boundaries. This methodology was extended and adapted for the specific task of PV panel detection and improved to handle noise and refine the detection results.

Figure 3.1 provides an overview of the four main phases of the proposed approach: (i) *classification* of PV images to identify those featuring Detectable PV (DPV) panels, i.e., panels exhibiting colours that do not significantly overlap with those of background areas such as fields, roads and water bodies; (ii) *extraction* of characterising PV (cPV) colours from the previously identified DPV images; (iii) *detection* of PV panels in unannotated aerial images using small square windows that scan the image and mark pixels on the panel surfaces; and (iv) a final *denoising* operation that adapts the scanning window methodology to identify and remove noise, which is defined as incorrect pixel-level detections in background areas. A summary of each phase is presented below, followed by a more in-depth description in subsequent sections.

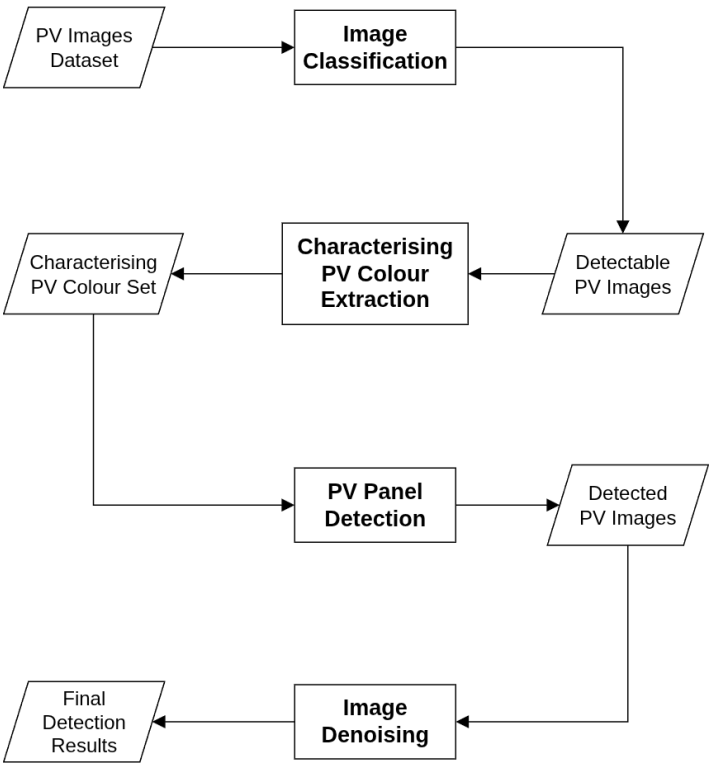


Figure 3.1: Overview of the four main phases (shown in bold) for the proposed PV panel detection methodology.

The purpose of the classification phase is to automatically categorise aerial images into several different classes based on the dominant colours exhibited by PV panels. It has been observed that, due to illumination conditions or the materials used, PV panels may display significant variation in their surface colours. These variations can include shades that are commonly found in background elements such as roads, buildings and terrain. For this reason, we developed a colour extraction and analysis methodology that automatically identifies and selects only those aerial images featuring Detectable PV (DPV) panels, namely panels exhibiting colours that are uncommon in background areas.

In the subsequent characterising PV (cPV) colour extraction phase, annotated DPV images are analysed in-depth to automatically obtain a set of colours that are representative of PV panels. Afterward, in the detection phase, this set of cPV colours is used to identify and mark pixels recognised as belonging to the surface of PV panels in unannotated images. Thereafter, a small square window scans the entire image to cluster previously marked pixels based on proximity and density. To further improve the detection, multiple passes of the moving square window with progressively larger window size are utilised. Finally, a denoising operation processes the detection results to tackle noise, defined as incorrectly marked areas in the background region. For this purpose, the moving window strategy is adapted to evaluate the structure and surroundings of marked tiles and determine whether the previous identification can be confirmed as correct.

The final output of the proposed detection approach consists of images in which PV installations are marked. Depending on the algorithm parameters, it is possible to identify both the surfaces of individual panels and the overall deployment areas of PV arrays, as shown in Figure 3.2.

Figure 3.2: Example final outputs of the proposed approach. The original PV image is shown on the left, the detected area for PV panel deployment is shown in yellow in the middle, and the individually detected PV panels are shown in red on the right.

3.1.2 Image Classification

The image classification phase includes three steps: (i) annotation assessment, which involves evaluating the accuracy and total area of the annotations in the corresponding aerial image; (ii) pixel colour extraction and filtering, which identifies a subset of colours that best characterises the surface of PV panels; and (iii) colour analysis and classification, which determines all images featuring Detectable PV (DPV) panels. Figure 3.3 provides an overview of these three steps. A detailed description of each step follows.

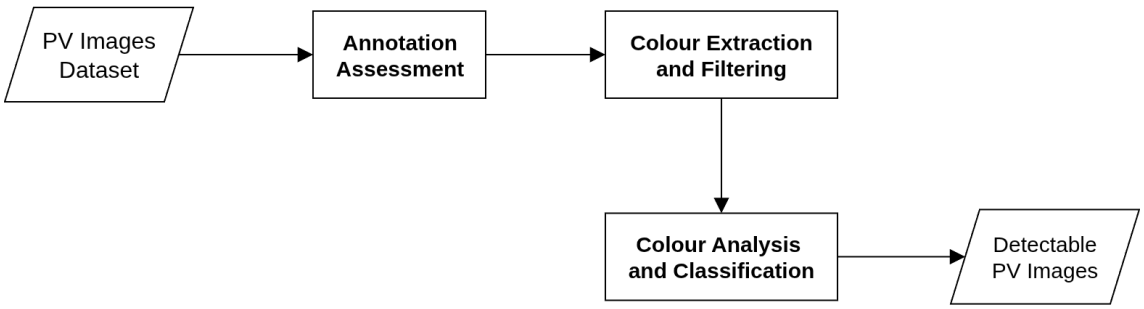


Figure 3.3: Workflow of the dataset image classification phase.

The main purpose of the annotation assessment step is to verify the accuracy of PV panel annotations in the dataset and to evaluate the proportion of the image covered by these annotations. Both aspects directly impact detection performance, as inaccurate or minimal annotations may result in a suboptimal reference colour set for the detection of PV panels and incorrect evaluation metrics. Figure 3.4 shows an example of an incorrect annotation, which fails to include the PV installations located in the bottom-right corner of the image.

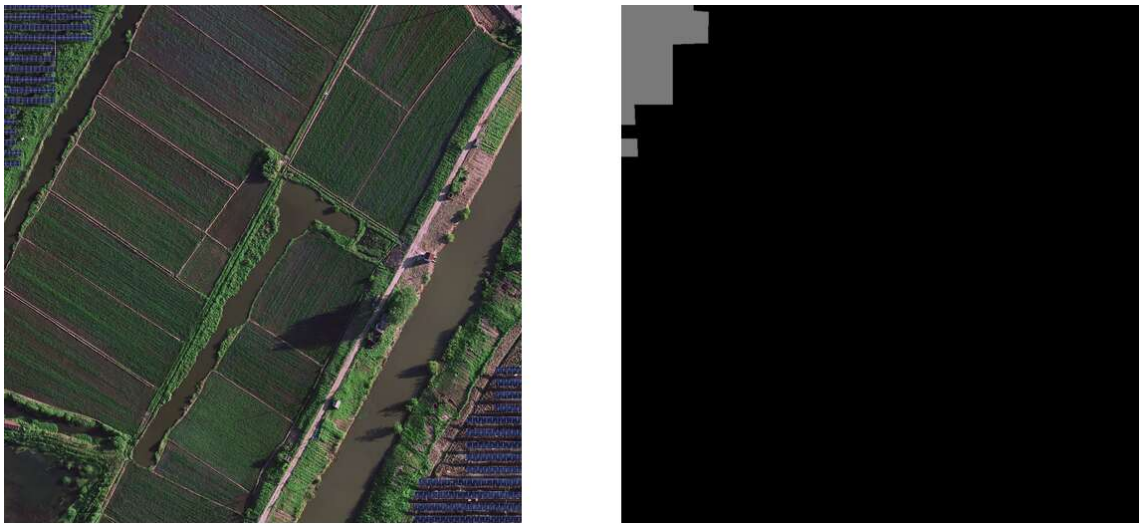


Figure 3.4: An example of an incorrect annotation in the original dataset [50]. PV panels are visible in the top-left and bottom-right corners (image on the left). However, the corresponding annotation (on the right) covers only the top-left corner.

Inaccurate annotations could cause PV panel detections in uncovered areas to be erroneously classified as false positives, leading to incorrect results in the approach's evaluation metrics. For this reason, all incorrect or incomplete annotations were identified, and the corresponding images were manually removed from the dataset. Moreover, PV installations with a limited extension do not provide a significant contribution to the subsequent colour extraction phase, as they yield a relatively

small PV colour set. Consequently, images where annotations cover less than 0.5% of the total image area were automatically identified and excluded from the dataset.

Annotations in the aerial image dataset may extend beyond the PV panel surfaces and include a significant portion of background areas such as terrain, vegetation, roads, shadows, and so on (see Figure 3.5).

Figure 3.5: The region enclosed by the dataset annotations includes both PV panels and background areas.

In such cases, the annotations cannot be reliably utilised to obtain a set of colours that are only found in the panel surfaces. To tackle this challenge, we devised a strategy to analyse all colours within the area enclosed by annotations and automatically exclude the subset of colours that are also commonly encountered outside the annotations, i.e., in the background regions. After this filtering operation, only the colours that more accurately represent PV panel surfaces are retained.

The strategy operates as follows. Initially, annotations are used to partition each aerial image into two disjoint regions: (i) Annotation Region (AR), including all areas delimited by the polygonal annotations; and (ii) Background Region (BR), representing the complementary portion of the image where PV panels are not present. Subsequently, the RGB colour components of all pixels are collected and the *colour count* for each region is computed, indicating the total number of pixels

displaying each colour within the two regions. This information is aggregated from all annotated images in the dataset and is used to build two colour sets: AR colours and BR colours. An overview of this process is shown in Figure 3.6.

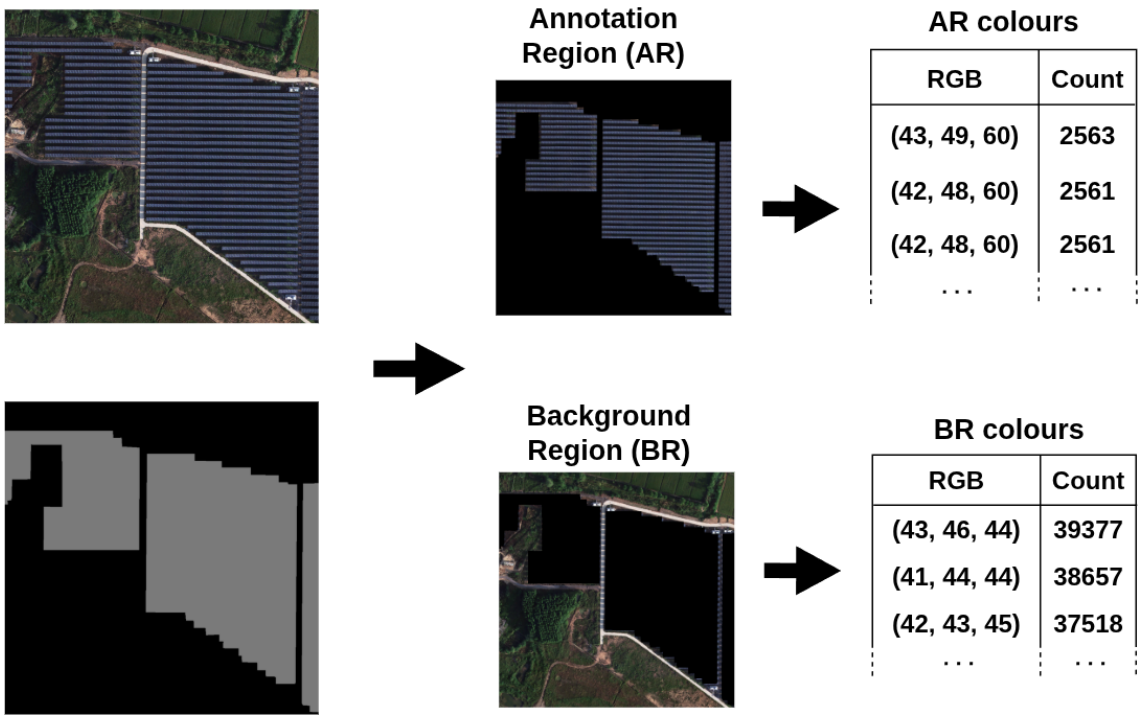


Figure 3.6: Annotations from the dataset are used to partition images into Annotation Region (AR) and Background Region (BR), after which the colours and their counts are determined.

As previously stated, some pixels within the AR may correspond to background rather than PV panels, due to potential inaccuracies in the annotations. It is therefore reasonable to assume that the colours of such pixels will appear frequently in the BR. To tackle this issue, the intersection of colours between the AR and BR is determined and, for each shared colour, the pixel counts in both regions are compared. If the ratio of the BR count to the AR count exceeds a predefined threshold, the colour is determined to be part of the background and is removed from the AR set. The

outcome of this filtering operation is a subset denoted as PV Panel (PVP) colours that more accurately represents the actual colours of PV panel surfaces.

Due to differences in illumination conditions or the materials utilised to build PV panels, the PVP colour set may exhibit significant colour variations. Our proposed methodology addresses this by considering three different classes of PV panels based on the properties of their prominent colours. Figure 3.7 shows examples for the three classes of PV panels.

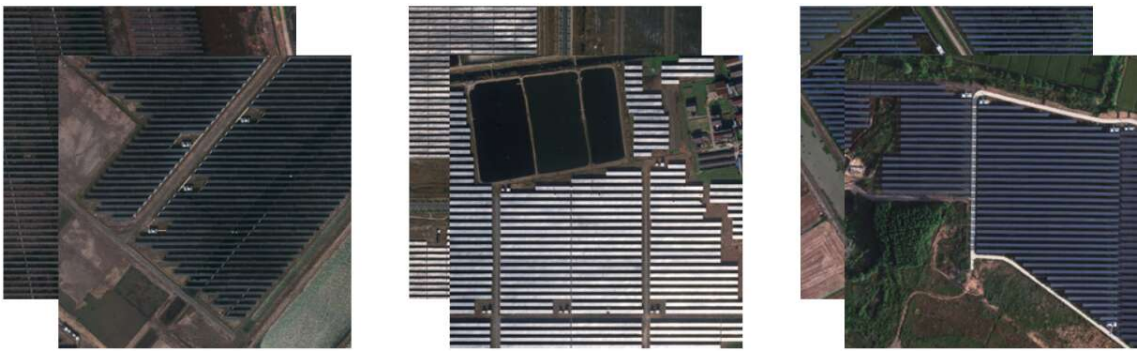


Figure 3.7: Classes of PV panels: dark panels (left), grey–white panels (middle), and detectable panels (right).

The first class is represented by *dark* panels, which are typically found in images captured under low-light conditions. The prevalence of dark shades in these cases limits their usefulness for characterising PV arrays, as similar colours frequently occur in shadows and dark terrain. The second class is denoted as *grey–white* panels. Their shades may result from the panel material or from sunlight reflecting off the surface, and are not distinctive indicators of PV installations. In fact, similar colours are also frequently observed in urban features such as roads and buildings. The third class is called *detectable* panels (DPV), and it includes all images that do not fall into the previous two classes.

We developed a strategy to automate the classification of the dataset images into these three classes. The strategy proceeds as follows: for each image, the AR colour set is intersected with the PVP colour set, resulting in the image-specific PVP colour set (iPVP). The iPVP set comprises all PVP colours found in the image, along with their corresponding pixel counts.

To assess the total amount of dark pixels, the RGB coordinates of all colours in the iPVP set are converted to the HSL (Hue, Saturation, Lightness) colour space [62, 108]. The *lightness* component is then evaluated, and any colour falling below a defined threshold is designated as dark. If the aggregated count of dark colours exceeds a predetermined fraction of the total iPVP pixel count, the image is classified as dark.

To evaluate the amount of pixels displaying white or grey shades, we leverage the fact that such colours can be described as having *low saturation*. Consequently, all colours in the iPVP set are transformed into the HSV (Hue, Saturation, Value) colour space [62, 108], which in our analysis enables better classification based on saturation compared to the HSL space. Similar to the previous step, we sum the count of all colours with a saturation component below a threshold. If this combined count exceeds a predetermined fraction of the total count of the iPVP set, the image is classified as grey–white. Images not categorised as either dark or grey–white are automatically classified as DPV images, and serve as the input for the next phase of our proposed methodology.

3.1.3 Characterising PV Colour Extraction

The goal of this phase is to extract colours from DPV images and isolate the characterising PV colours. Similar to the image classification phase, the first step is

to partition each DPV image into an Annotation Region (AR) and its complementary Background Region (BR). Then, the colours and their corresponding pixel counts for all images are aggregated and recorded separately for each region, resulting in the AR colours and BR colours sets. To remove background colours, the ratio between the BR count and AR count is evaluated for all colours in the intersection of the two sets. As a result of these operations, the PVP colour set is determined.

A further filtering process is then applied to the PVP colour set to identify and remove both dark and low-saturation colours. The reason is that, as discussed earlier, such colours are not representative of PV panels. This filtering operation uses threshold values for the lightness component in the HSL colour space and the saturation component in the HSV colour space. The outcome of this colours processing is the characterising PV (cPV) colour set, which represent a core component of the detection algorithm discussed in the following section.

3.1.4 PV Panel Detection

The detection phase of the proposed approach consists of two steps, shown in Figure 3.8: (i) pixel replacement, in which all pixels displaying cPV colours are marked; and (ii) detection refinement, in which windows of varying sizes scan the image to cluster pixels recognised as belonging to the surface of PV panels. The final output of this phase is a set of images in which the detected PV panel locations are marked.

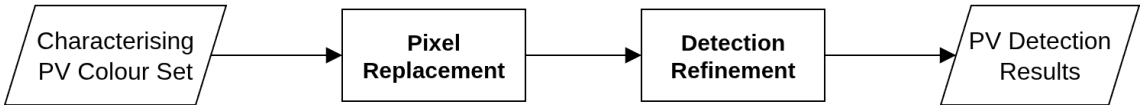


Figure 3.8: Workflow of the PV panel detection algorithm.

Initially, PV images are analysed and all pixels with a colour contained in the cPV set are identified and marked with RGB green. Figure 3.9 shows an example of this pixel replacement operation.

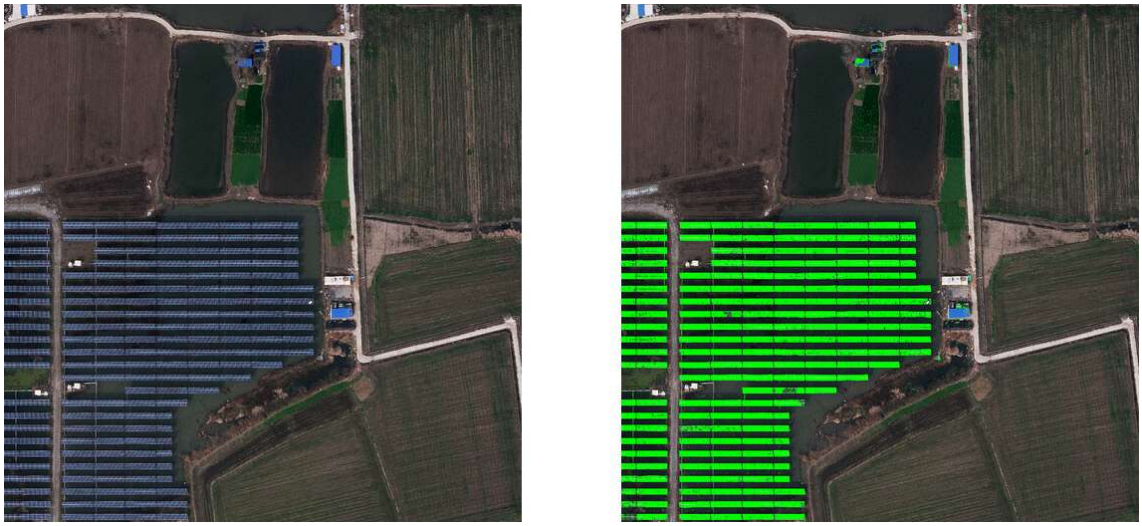


Figure 3.9: An example image with pixels displaying cPV panel colours marked in green (right). The original image is shown on the left for reference.

Although a high concentration of marked pixels is expected to correspond to PV panel surfaces, the filtering process used to generate the cPV colours set may exclude a significant minority of panels pixels because they do not exhibit characterising PV colours. As a consequence, only a fraction on the total panel area is detected at the end of this step. Figure 3.10 illustrates this limitation, showing a zoomed-in section of a PV array where several panel pixels remain unmarked.

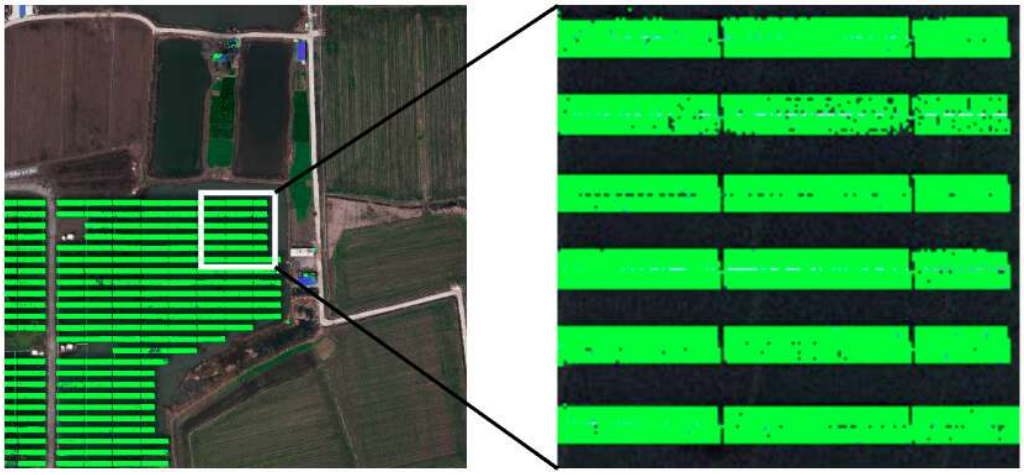


Figure 3.10: The enlarged section (right) shows that not all pixels on the PV panel surfaces were marked after the pixel replacement operation.

To enhance the detection of PV panels, a refinement step is applied. This involves scanning the entire image using a moving window and evaluating the density of previously marked pixels within each window. At the beginning of the detection refinement, three parameters are defined: (i) the window size; (ii) the pixel density threshold required to confirm an existing detection; and (iii) the new colour used to mark the area covered by the window.

In each iteration, the number of marked pixels within the window is assessed. If their total count exceeds a predetermined threshold, for example 75% of the total pixels in the window, the entire tile is marked with a new colour, such as RGB blue. This assessment is performed for the whole image. Depending on the features of PV installations, multiple passes with progressively larger windows can greatly enhance the effectiveness of PV panel detection.

Figure 3.11 shows example results from two consecutive refinement runs using window sizes of 5×5 and 11×11 pixels, respectively. The enlarged sections demonstrate the improvements in PV detection after the successive window passes.

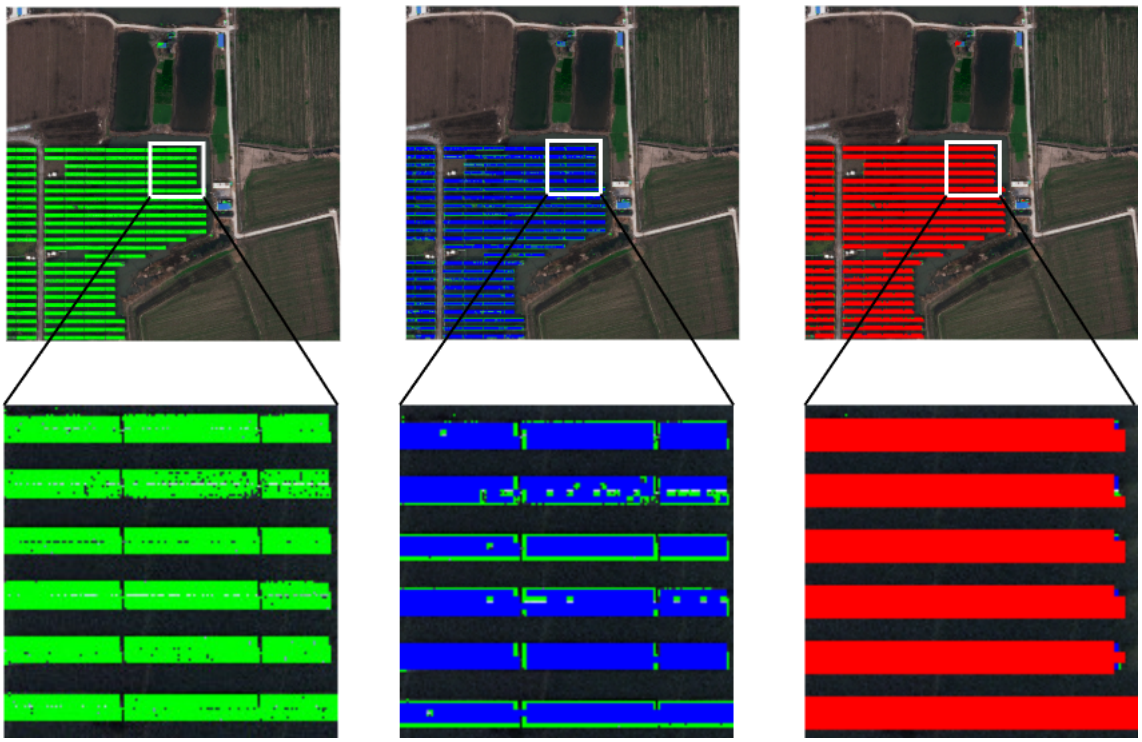


Figure 3.11: Detection results after a 5×5 window pass (middle) performed on the marked pixels image (left), and a subsequent 11×11 window pass (right). The enlarged sections at the bottom show the detection enhancement after successive passes.

The reason for utilising a new marking colour in each refinement run is that each successive pass assesses the density of marked pixels within an increasingly larger area. If a window in a subsequent refinement iteration contains some marked pixels but their cumulative count is not sufficient to trigger the threshold, this indicates that a significant number of pixels within the window do not feature cPV colours. In such cases, it is likely that the previously marked pixels represent incorrect detections. Therefore, no marking is applied to the window, and incorrect detections are addressed during the denoising phase, described in the following section.

3.1.5 Image Denoising

Pixels displaying cPV colours may appear in background regions. Occasionally, the count of such pixels is sufficient to trigger the scanning window threshold, leading to incorrect detections. Examples are shown in Figure 3.12.

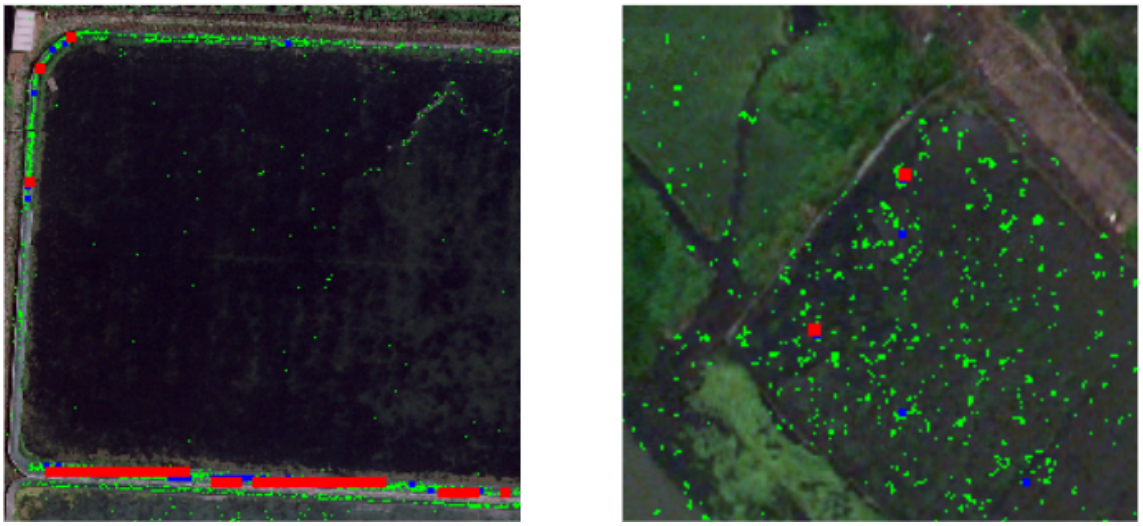


Figure 3.12: Examples of detections in background regions.

To address noise, i.e., occurrences of cPV pixels outside of actual PV panels, we initially note that panel surfaces typically contain large clusters of marked tiles. Conversely, incorrect detections in the background often result in isolated coloured pixels or short coloured stripes. Leveraging this observation, the strategy for identifying and eliminating noise involves evaluating the surroundings of marked pixels.

To achieve this, we apply the same scanning window methodology, adapting it for the denoising task. Adjustments are made to two key parameters: first, to include surrounding pixels, the denoising window size is set larger than the detection window size used in the last run of the algorithm. Then, a threshold is defined such that when the count of marked pixels within the window falls below the threshold, the

pixels are classified as noise and consequently coloured in green, whereas detected panels are coloured with the final marking colour. Figure 3.13 shows the results after the denoising pass. Green pixels are excluded from the detection results and only pixels marked with the final iteration's colour (red in this example) are considered a valid detection.

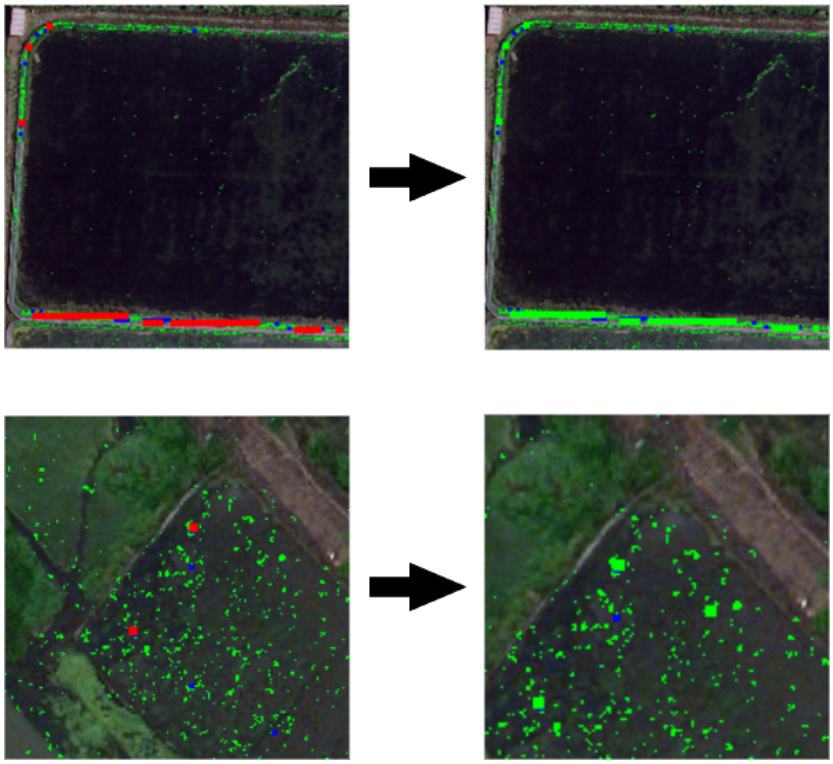


Figure 3.13: Incorrect detections in background areas (the red dots in the left image) are identified as noise and marked in green (right image), effectively reverting the initial detection.

3.1.6 Experiments and Results

We tested and evaluated our proposed approach on the aerial imagery dataset presented in [50]. We selected images with a resolution of 0.3 m/px and a size of 1024×1024 from five distinct sub-datasets, namely: *Cropland*, *Grassland*, *SalineAlkali*, *Shrubwood*, and *WaterSurface*. The selected images constituted the ground dataset,

comprising $2,072$ images in total. This ground dataset served as the input for the first phase of the proposed approach, classification, which analysed annotations and pixel colours to identify DPV images.

After the annotation assessment, a total of 71 incorrect annotations and 115 annotations covering less than 0.5% of the image area were found. Consequently, the corresponding images were removed from the ground dataset. Next, colours and counts were extracted from both AR and BR, and the background colours were filtered out using a threshold of 1.0 on the colour count ratio between BR and AR, to form the PVP colour set.

The PVP colour set analysis proceeded as follows. Firstly, to identify dark panels, the pixel RGB coordinates were projected in the HSL colour space. A threshold of 0.20 (in the range $[0, 1]$) was utilised for the lightness component, and a threshold of 50% for the dark pixel count was used to identify and mark images as belonging to the dark PV panels class. Following this, to detect grey–white panels, the RGB coordinates were transformed into the HSV colour space. To filter colours based on the saturation component, a threshold of 0.25 (in the range $[0, 1]$) was applied, and images were categorised into the grey–white class if the threshold of 33% for low saturation pixel count was exceeded. Table 3.1 shows a summary the image classification parameters.

Parameter	**Threshold**
Annotation Area	0.005
BR/AR Colour Count	1.00
Lightness Component (HSL)	0.20
Dark Pixel Count	50%
Saturation Component (HSV)	0.25
Low-Saturation Pixel Count	0.33%

Table 3.1: Summary of parameters used in the classification phase of the aerial image dataset displaying PV panels.

After completing the PVP colour set analysis, the dataset was partitioned in the following classes: 174 dark PV panels images, $1,254$ grey–white PV panels images, and 458 Detectable PV panels (DPV) images. The DPV images were selected and served as the input for the subsequent phase. Initially, 80% of the DPV images were randomly chosen to extract characterising PV (cPV) colours, while the remaining 20% were used as the test set to evaluate the performance of the proposed approach.

The experimental results of colour extraction on the DPV images were as follows: an aggregated $1,133,935$ unique colours for the AR, and a total of $1,836,190$ unique colours for the BR. After filtering background colours, a total of $403,713$ (cPV) colours remained, which accounted for 35.60% of the total AR colours.

After obtaining the cPV set, the PV panel detection algorithm was executed on the test images. In the pixel replacement step, any pixel with a colour found within the cPV set was marked in green. In the following detection refinement step, two scanning windows were utilised, one of size 5×5 pixels with a detection colour of RGB blue and another of size 11×11 pixels with a detection colour of RGB red.

The detection threshold for both windows was set to 0.5. Figure 3.14 shows an image for which the two-pass detection was performed, together with the original dataset image and the corresponding annotation.

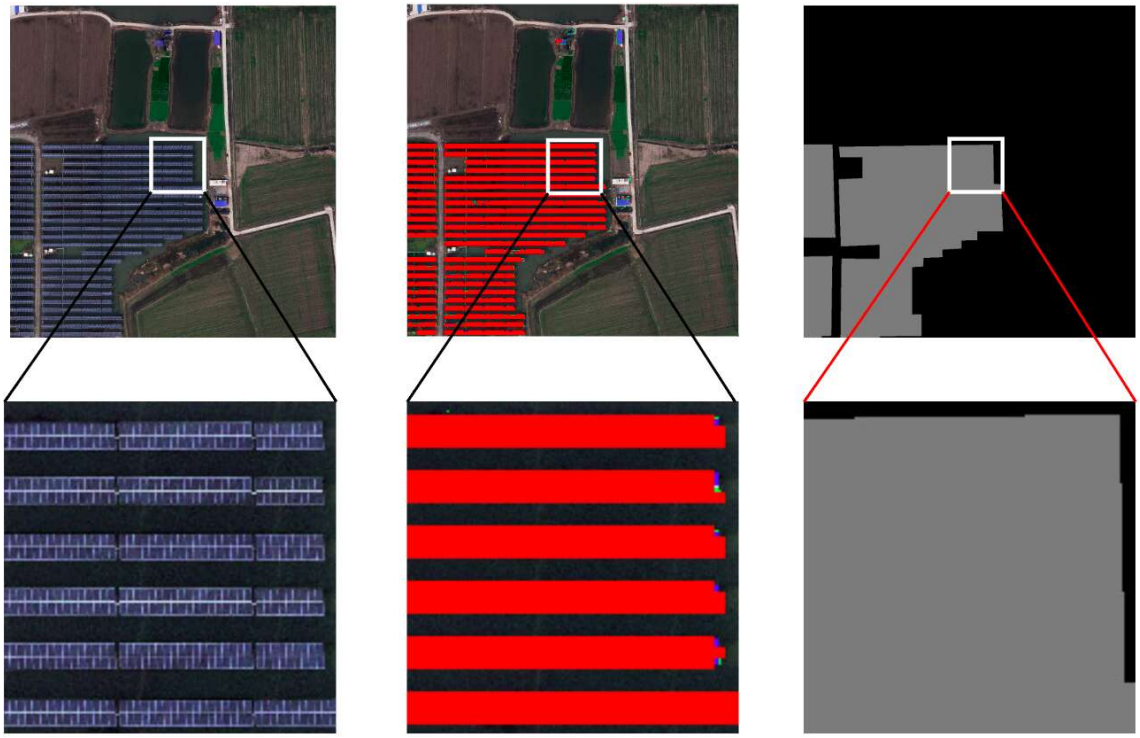


Figure 3.14: An example PV image (left) and the results of PV detection after two successive detection passes (middle). The corresponding dataset annotation is shown on the right. At the bottom, a zoomed-in section of each of the three images is included.

The output image (middle) shows that the panel surfaces were identified and the background portions between rows of PV installations were not marked with the detection colour (red). This confirms that the colour filtering approach was able to distinguish between PV-characterising colours and background colours with very high accuracy.

To detect the whole PV panels installation area, which in the annotated dataset includes the background between individual panels, an additional detection pass with a window of size 21×21 pixels was used. Compared to the previous iterations,

a lower detection threshold of 0.25 was set to account for the high number of colours not included in the cPV set that appeared in the background portions between PV panel rows. The marking colour for this third iteration was set to RGB yellow. Table 3.2 shows the parameters used for the three detection passes, while Figure 3.15 presents the detection results.

Detection Iteration	**Scanning Window Size**	**Detection Threshold**	**Marking Colour (RGB)**
1	5×5	0.50	Blue (0, 0, 255)
2	11×11	0.50	Red (255, 0 , 0)
3	21×21	0.25	Yellow (255, 255, 0)

Table 3.2: Parameters used for PV area detection.

Figure 3.15: Results of PV panel region detection (middle) compared with the original dataset label (right). The original image is shown on the left for reference.

In the final phase, a post-processing denoising operation was applied to the images. The window size was 63×63 pixels and the denoising threshold was set to 1/3 of the window area. All coloured tiles identified as incorrect detections were marked in RGB green, as shown in Figure 3.16.

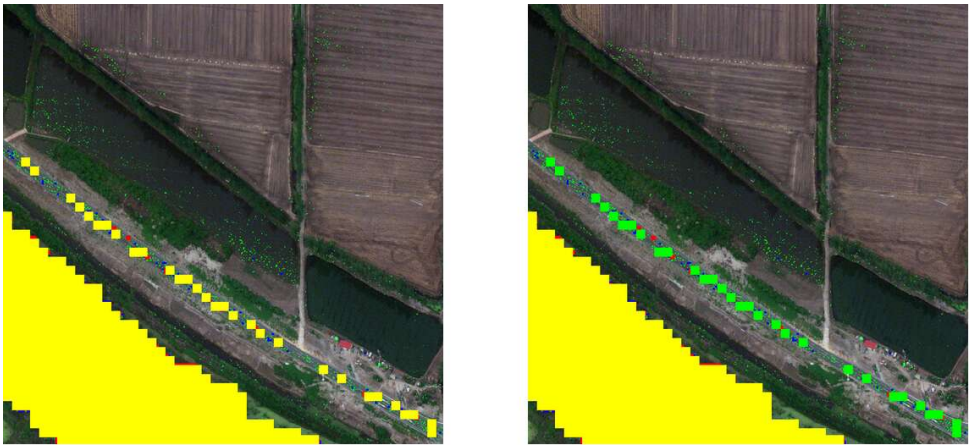


Figure 3.16: The denoising process reverts false positive detections in the background (yellow tiles) by marking them as green tiles. The final result is shown in the image on the right.

We assessed the detection results of the proposed approach by evaluating five metrics: accuracy, precision, recall, F1 score, and intersection over union (IoU). Table 3.3 presents the results for the DPV ground dataset as well as the five individual sub-datasets: Cropland, Grassland, SalineAlkali, Shrubwood, and WaterSurface. The last row of Table 3.3 displays the values reported by the authors who released the original dataset in [50].

Dataset	**Size**	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**IoU**
DPV Ground	458	0.963	0.918	0.901	0.897	0.833
Cropland	146	0.971	0.931	0.938	0.931	0.877
Grassland	42	0.952	**0.953**	0.881	0.912	0.844
SalineAlkali	53	**0.982**	0.913	**0.958**	0.934	0.878
Shrubwood	77	0.975	0.934	0.956	**0.944**	**0.897**
WaterSurface	140	0.974	0.915	0.928	0.917	0.852
Max in [50]		0.981	0.960	0.903	0.931	0.877

Table 3.3: Evaluation metrics for the DPV Ground dataset and the five sub-datasets.

To evaluate the detection performance of the PV panel surface, as opposed to the whole area covered by PV installations, we manually created polygonal annotations for five test images. These new annotations exclusively enclose the panel surfaces without including any background areas. Figure 3.17 shows a comparison between the PV panels annotations and the original dataset annotations.

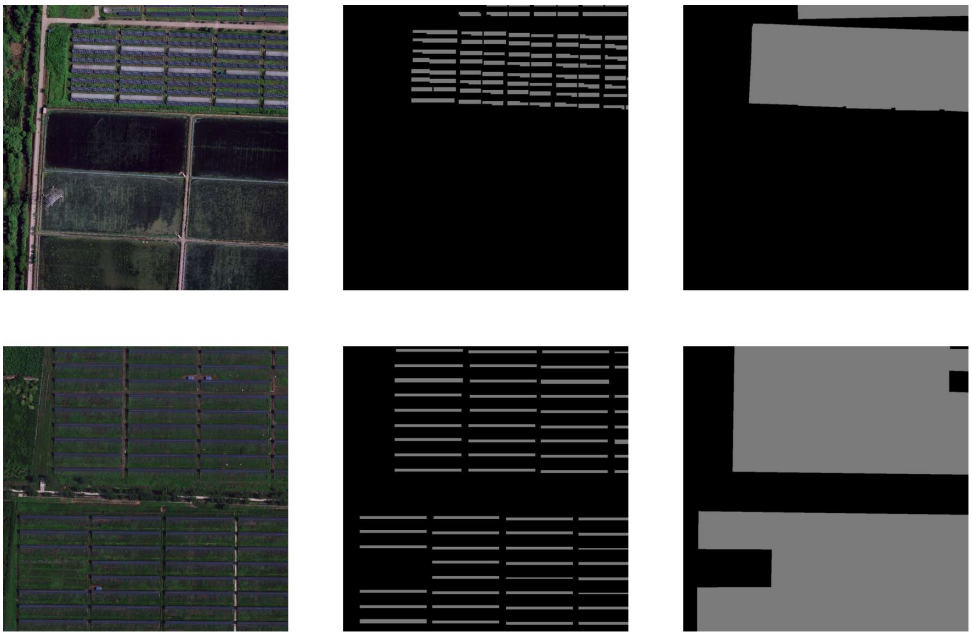


Figure 3.17: Updated annotations covering only the panel surfaces (middle) alongside the original dataset annotations (right). The corresponding PV images are shown on the left.

The detection algorithm was then executed on the DPV Ground dataset using the parameters shown in Table 3.4.

Detection Iteration	**Scanning Window Size**	**Detection Threshold**	**Marking Colour (RGB)**
1	3×3	0.75	Blue (0, 0, 255)
2	5×5	0.5	Red (255, 0 , 0)

Table 3.4: Parameters used for detecting PV panel surfaces.

The experiments revealed that smaller window sizes produce better results when detecting only PV panel surfaces. Moreover, since the background areas between panels do not need to be considered, a higher threshold of 0.75 is utilised for clustering pixels displaying cPV colours. For the same reason, only two iterations instead of three were sufficient to complete the detection. The final denoising phase was executed with a denoising window size of 15×15 and a threshold of $1/3$ of the window area. Example results of the PV panel surface detection are shown in Figure 3.18.

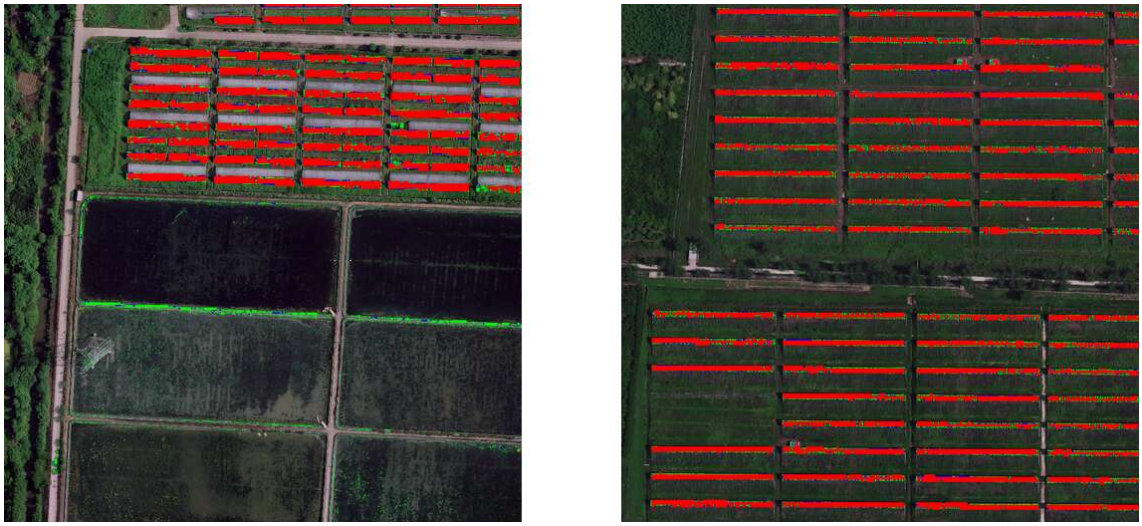


Figure 3.18: Example results of PV panel surface detection.

The recall and precision metrics for the test images were calculated for both the PV panel surface and original annotations. The results are displayed in Table 3.5, where the *PV/Annot.* columns represent the ratio between the area covered by the PV panel surfaces and the total annotation area.

	PV Surface Annotation			Original Annotation		
Image ID	**Precision**	**Recall**	**PV/Annot.**	**Precision**	**Recall**	**PV/Annot.**
325711_1198541	0.988	0.916	100%	0.984	0.436	47%
318008_1197612	0.948	0.918	100%	0.979	0.330	35%
331352_1190202	0.955	0.932	100%	0.969	0.292	31%
331764_1180482	0.936	0.949	100%	0.991	0.309	31%
335603_1188533	0.950	0.870	100%	0.998	0.166	18%
Average	0.955	0.917		0.984	0.307	

Table 3.5: Comparison of recall and precision scores between the updated and original annotations.

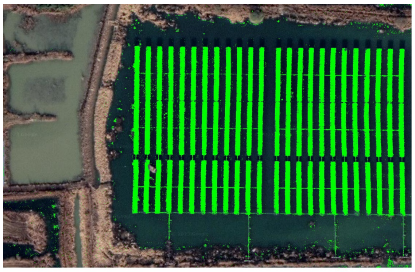
To evaluate the effectiveness of our cPV colour-based approach in detecting PV panels in unannotated aerial images that are not part of the ground dataset, we acquired three new images displaying PV panels from Google Maps™ [33], in the same geographical areas indicated by the authors of the original dataset [50]. These three images were named based on the type of land they depict: `green_bg`, `water_bg` and `rocky_bg` (where `bg` denotes 'background'). Table 3.6 shows the geographical coordinates of the images.

Image	**Latitude**	**Longitude**
`green_bg`	33.1926	118.0222
`water_bg`	33.2296	117.9829
`rocky_bg`	33.1879	118.0231

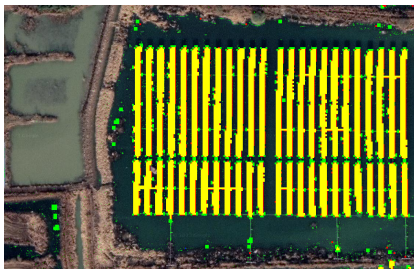
Table 3.6: Geographical coordinates of the three PV images.

To perform the panel detection on these new images, we employed the same cPV colour sets that were extracted from the DPV images at the end of the second phase of our proposed approach. Each characterising colour set matched the corresponding geographical area of the new images, that is, Greenland, WaterSurface

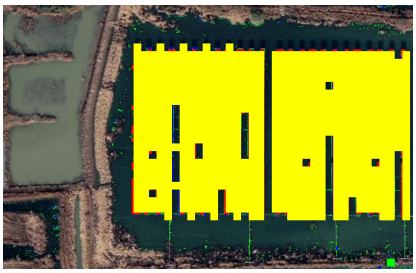
(a) Original `green_bg` image.



(b) Output of pixel replacement.



(c) PV panel surface detection.



(d) PV panel area detection.

Figure 3.19: Detection results on the `green_bg` image.

and SalineAlkali, respectively. The detection results for the `green_bg` image are shown in Figure 3.19, and the complete results are available in the publicly accessible supplementary materials [119]. The evaluation metrics for the detection of PV panel surfaces and areas are shown in Table 3.7 and Table 3.8, respectively.

Image	Precision	Recall	F1 Score
`green_bg`	0.872	0.945	0.907
`water_bg`	0.829	0.980	0.898
`rocky_bg`	0.829	0.917	0.870

Table 3.7: Evaluation metrics for PV panel surface detection.

Image	Precision	Recall	F1 Score
`green_bg`	0.960	0.899	0.928
`water_bg`	0.948	0.943	0.945
`rocky_bg`	0.937	0.946	0.941

Table 3.8: Evaluation metrics for PV panel area detection.

3.1.7 Discussion

We evaluated the performance of our approach based on the following five metrics: accuracy, precision, recall, F1 score, and intersection over union (IoU). Table 3.3 shows the results, with the highest value marked in bold, and the metrics reported in [50] are also displayed in the last row of the table for comparison.

Analysis of the results reveals that our approach consistently performed better, except for precision. However, it should be noted that the difference between the two precision values is just 0.007. Recall exhibits the highest variation across different datasets, with a minimum value of 0.881 for the Grassland dataset and a maximum value of 0.958 for the SalineAlkali dataset. This variation can be explained by the presence of a large number of common colours between PV panels and their surroundings in Grassland images. In fact, our approach filters out PV panel colours with a high count in background regions, resulting in a reduction of the cPV colours set size in Grassland areas. In general, the detection performance is directly influenced by the dissimilarity in colours between the PV panel surfaces and the background areas.

The detection methodology presented in [50] and other deep learning-based approaches identifies the broad deployment area of PV panels, which typically includes a significant portion of background. In our analysis, a significant number of images in the original dataset feature annotations containing more than 50% of background pixels relative to the total annotation pixel count. In such cases, it is not possible to reliably estimate the actual PV panel size. On the other hand, our approach is able to identify both PV panel deployment areas and PV panel surfaces, with only minimal changes to the algorithm's parameters, namely the size and threshold for the detection windows. To show the effectiveness of our methodology,

we created accurate annotations for panel surfaces in a small subset of images from the ground dataset, and executed the algorithm.

Table 3.5 shows the performance comparison between the updated annotations and the original annotations. When using the updated annotations, a mean recall score of 0.917 was achieved, demonstrating our algorithm's effectiveness in detecting PV panel surfaces. Conversely, when utilising the original annotations, the mean recall score dropped to 0.307. This difference can be explained by the fact that the PV surface comprises only a fraction of the total annotated area, as indicated in the PV/Annot. column. Indeed, in all selected images, the background accounts for more than 50% of the original annotation area. The high recall value with the updated annotations was obtained while also attaining an average precision score of 0.955. The higher value obtained with the original annotations can be explained by the fact that most of the pixels surrounding the panel surface are considered true positives when using the dataset annotations, despite actually representing the background.

Overall, while the precision scores remain high for both annotation types, the recall varies significantly and is consistently much higher for the updated annotations. Therefore, our approach reliably identifies PV panels, whereas other approaches are useful only for the broad PV array deployment area.

In contrast to CNN-based techniques, our method forgoes a training phase, offering several advantages. First, it requires only a limited number of annotated images to obtain the characterising colours of PV panels, whereas ML and DL approaches require large datasets of annotated data for training. This training process can be time-consuming and error-prone, and incorrect annotations can pose significant challenges. Second, training phases in deep learning typically require substantial computational power and a large amount of storage capacity, in addition

to prolonged execution times for model training. In contrast, the execution time of our approach is solely dependent on the number of images to be analysed. Finally, along with its shorter execution time, our approach achieves high performance, as shown in Table 3.3, which is comparable to, if not better than, that of a CNN-based approach.

Table 3.9 shows the average execution times of each phase in our proposed approach, for each image. The algorithm was run on a PC with the following characteristics: (i) Linux Ubuntu operating system kernel version 6.2.0-36-generic; (ii) 2.0 GHz AMD Ryzen 5 2500U 4-core processor; (iii) AMD Radeon Vega 8 Graphics; (iv) RAM memory 8 GB 2400 MHz DDR4.

Phase	**Execution Time per Image**
Image Classification	2.42 s
cPV Colours Extraction	1.63 s
PV Panels Detection (3 passes)	9.25 s
Image Denoising	3.16 s

Table 3.9: Average execution times per image for each phase.

The cPV colour extraction phase yields a set of colours that are used to mark pixels on the panel surface in the aerial images. This extraction is performed only once and the resulting cPV colour set can be used to analyse many different images from the same geographical area. However, the colours of PV panels can be influenced by several factors, including the panel's model (e.g., the materials used), the intensity of incident light (which can either lighten or darken the panel), and the panel's inclination and orientation. Accordingly, PV panels with colours that closely resemble the background or exhibit darker shades pose challenges for detection. Therefore,

overall detection results are more reliable for images displaying fewer very dark colours and where the background or surrounding objects exhibit different colours.

In our experiments, we utilised remote sensing imagery with a pixel resolution of 0.3 metres per pixel. When using lower resolutions, the risk of lower detection performance increases due to the coarseness of the figure edges and the blending of pixel colours. The availability of additional annotations for roads, fields, buildings and other entities commonly displayed in the dataset images enables the creation of a characterising colour set for each such category. This allows for more advanced colour analysis and comparison between cPV colours and characterising colours of other classes, potentially leading to enhanced detection results.

The versatility of our approach was shown in Figure 3.19, where we employed the cPV colour set obtained from the DPV dataset on new images from similar geographic regions but not belonging to the original dataset. Tables 3.7 and 3.8 show high values for the evaluation metrics for both the PV panel surface and area detection, proving that the set of characterising colours does not need to be tuned for new images to be processed, and hence our proposed approach is highly robust.

3.1.8 Summary

In this study, we have presented an approach that accurately and automatically detects PV panels in remote sensing images. The proposed methodology consists of four main phases: the first phase, classification, involves analysing pixel colours and partitioning the image dataset into multiple classes depending on the illumination conditions of the PV panels. In the following phase, a small number of annotated images are used to identify the PV panels deployment area, and the set of characterising PV (cPV) panel colours is determined by evaluating the ratio of PV panel

colours with respect to the colours found in background areas. Next, the detection of PV panels is automatically performed on unannotated images by marking pixels using the cPV colour set as a reference. To refine detection, a scanning window is utilised to cluster marked pixels based on their density. A final denoising phase identifies and removes incorrect detections.

To validate the effectiveness of our approach, experiments were performed using an open-source dataset of aerial images displaying PV panels. The results showed that the proposed approach can reliably and consistently detect PV panels, achieving high precision and recall values. Furthermore, a comparison with a Convolutional Neural Network (CNN)-based approach revealed superior performance in terms of recall and accuracy for our approach. Notably, unlike many CNN methods, our approach is capable of identifying both the areas and the precise surfaces of PV panels, offering a dual identification capability.

Thanks to the initial classification phase in our approach, datasets for different regions can be analysed successfully, e.g., images that have been captured at different times of the day are automatically grouped based on illumination conditions, and a subset of images that are best suited for PV panel detection is identified. Finally, our approach exhibits fast execution times, as no training phase is required, and the automatic extraction of characterising PV panel colours from images of different regions provides a high level of flexibility when applying our methodology to images with widely varying colours of panels and surrounding areas. In particular, our results demonstrated that the proposed approach can be successfully applied to detect PV panels in unannotated images that do not belong to the dataset used to generate the cPV colour set.

3.2 Automatic Land Use and Land Cover Classification by Means of Characterising Colours

There is a pressing need for concerted efforts towards systematic management and planning to optimise the utilisation of agricultural lands [81]. In this context, "land cover" refers to all uninhabited areas untouched by humans, including rivers, seas, lakes, oceans, trees, forests, agricultural areas, deserts, and bare lands. Conversely, "land use" encompasses all areas inhabited or utilised by humans, such as buildings, industrial facilities, hospitals, universities, schools, and service facilities [150].

Many previous studies have attempted Land Use and Land Cover (LULC) classification with promising results [88, 102, 125]. Aerial and satellite images were analysed with machine learning techniques [121], such as random forests [58, 101] and support vector machines (SVM) [4, 52], as well as deep learning techniques [11, 15, 112, 124]. However, such approaches require large datasets to train their models, and a considerable effort to build and annotate datasets. Additionally, solely relying on a training phase carries the risk of overfitting, resulting in poor generalisation for the classification task.

Other approaches used multispectral and hyperspectral images to perform LULC classification. However, this methodology requires a significant amount of time and resources to gather [67, 129]. Hyperspectral data is high-dimensional and complex, since it contains information across numerous spectral bands. Moreover, the process of creating accurate annotations requires expertise and may involve sophisticated analysis techniques, such as spectral classification or feature extraction. These challenges often result in a limited amount of available labelled data.

To extend the functionality of the framework presented in this doctoral thesis,

we propose a robust and explainable approach to classify land cover based on the most prominent colours of each land type. The proposed methodology builds upon and extends our previous research [76], adapting it for the task of detecting several different land cover categories in remote sensing images. More specifically, our analysis focuses on three different land types: forest, irrigated land and non-irrigated land. The proposed approach initially extracts all unique pixel colours for each land type from a small set of annotated images. Subsequently, a filtering operation is applied to these colour sets to remove all colours exhibiting a low lightness level, i.e. shades typically found in shadow areas. Next, the sets are further analysed and filtered by comparing the colour counts across all land categories. The output of these colour processing operations is represented by three sets that contain characterising colours for each land type. These sets are then utilised by the detection algorithm to mark pixels in unannotated images and identify land cover categories. A final post-processing phase handles incorrectly marked pixels and refines the detection, improving the classification performance.

The main strengths of our approach are: (i) only a small annotated image dataset is needed to obtain the characterising colours of land cover types; (ii) the colour extraction algorithm is fast and does not require computationally intensive resources; (iii) results are deterministic, allowing for further tuning of the sets of characteristic colours for each land type; (iv) the classification methodology is explainable, i.e. for each land cover type, the distribution of characterising colours is provided, which determines the final classification results.

The remainder of this section is structured as follows. Section 3.2.1 presents an overview of the proposed approach. Section 3.2.2 explains the methodology for extracting the characterising colour sets for each land cover type. Section 3.2.3 details

the land cover classification phase. Section 3.2.4 elaborates on the denoising strategy applied to the remote sensing images. Section 3.2.5 describes our experiments and discusses the results. Finally, conclusions are drawn in Section 3.2.6.

3.2.1 Proposed Approach

The main methodology of the proposed LULC classification algorithm consists of analysing the colours of a small remote sensing imagery dataset to obtain a set of *characterising colours* for each LULC category. These sets are then utilised to identify the location of the LULC categories in other, unannotated aerial images. Figure 3.20 shows an overview of the main phases of the proposed methodology.

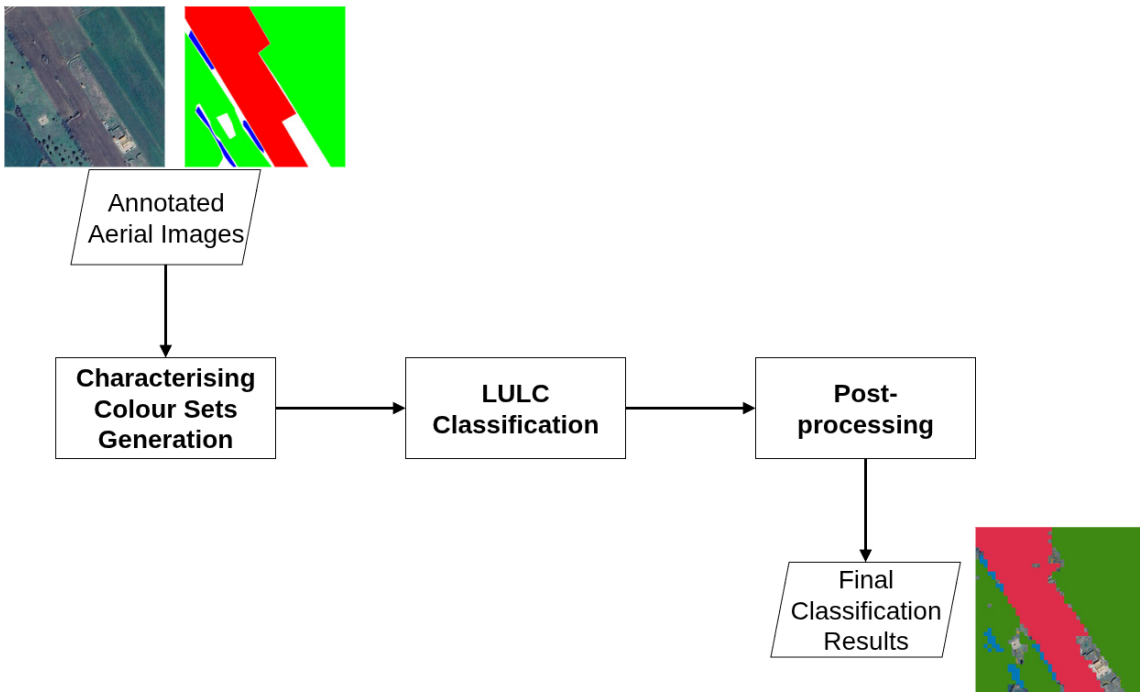


Figure 3.20: Overview of the proposed approach, featuring the three main phases: characterising colour sets generation, Land Use and Land Cover (LULC) classification, and post-processing.

The three main phases are: (i) characterising colour sets generation from annotated images; (ii) LULC classification; (iii) post-processing, which handles incorrect detections and enhances the classification results. A summary of each phase is provided below, and a more detailed discussion is given in the following sections.

In the characterising colour sets generation phase, colours are extracted from previously annotated aerial images, followed by an analysis and frequency comparison to identify characterising colours for each LULC category. In the subsequent LULC classification phase, pixels in unannotated images are marked based on the LULC category they belong to; afterwards, a small square window is employed to scan the entire image and cluster marked pixels together based on their density. Finally, the post-processing phase removes incorrect detections and further refines the classification results. The final output of the LULC classification algorithm are images in which land types are detected and marked accordingly. The following sections detail the operations for each phase.

3.2.2 Characterising Colour Set Generation

The first phase of the proposed approach is the characterising colour (CC) sets generation, which includes three steps (see Figure 3.21): (i) colour extraction, (ii) colour filtering, and (iii) colour classification.

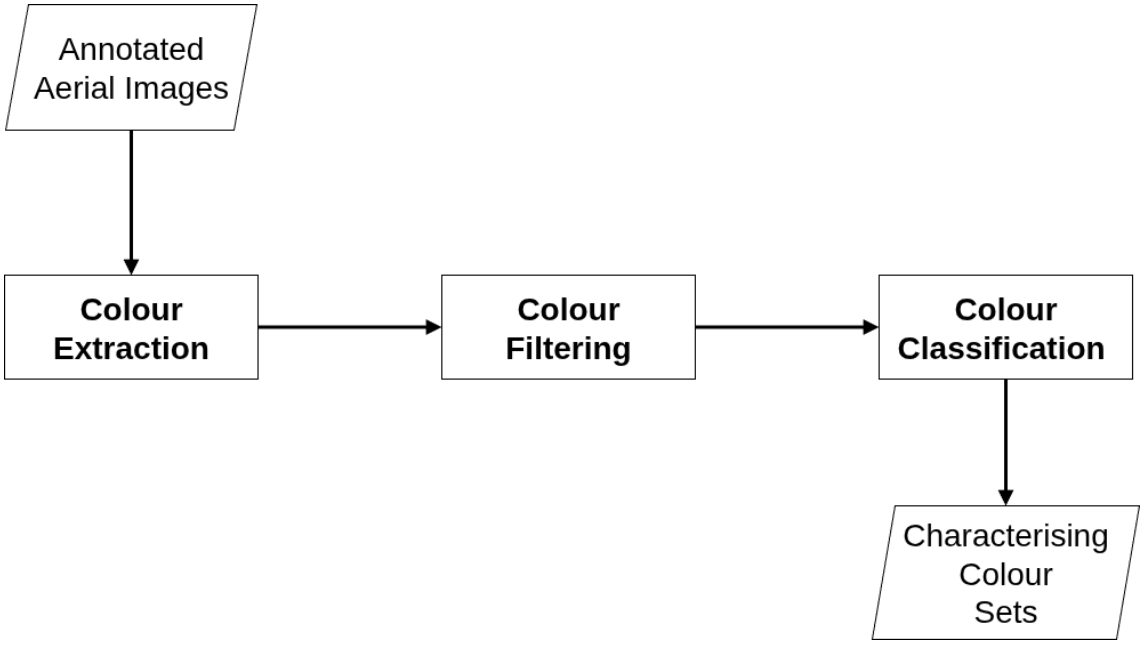


Figure 3.21: The three steps of the characterising colour set generation phase. Starting from a small set of annotated images, colours are extracted and analysed to determine the set of characterising colours for each LULC category.

It is assumed that annotations in the input aerial imagery dataset feature a unique annotation colour for each LULC category. Pixels that are not found within the boundaries of any annotation are automatically assigned to the *background* category. The overall goal of this phase is to obtain a set of CC colours for each category, which are subsequently used by the classification algorithm to identify LULC types.

In the initial step, annotations are used to select all pixels belonging to a specific category, and the RGB colour components are extracted. Next, the *colour counts* are determined, indicating the overall number of pixels displaying each individual colour. This process is repeated for each LULC category, and the background.

This data is collected from all input images and aggregated to build the Land Colour Table (LCT). Example LCT entries for three LULC categories and the background are shown in Table 3.10.

Colour (R, G, B)	**Count**			
	Forest	**Irrigated**	**Non-irrigated**	**Background**
(55, 68, 100)	314	25	84	55
(64, 70, 106)	88	0	2262	70
(84, 124, 134)	35	112	4	136
(91, 122, 125)	14	1102	0	755

Table 3.10: Example entries from the Land Colour Table. The LULC categories include forest, irrigated land, and non-irrigated land. Colour count values are recorded for each category, as well as for the background.

The LCT is structured as follows: the first column comprises the unique RGB components of the collected colours and the subsequent columns correspond to the LULC categories, plus the background. In each row, the first element contains the RGB colour components and all other values are the corresponding counts for every category.

The second step, colour filtering, implements strategies for identifying and removing colour subsets that could potentially hinder detection performances. In our approach we mainly focus on detecting and isolating dark colours, which typically correspond to shadows, and therefore are not representative of any category. As a consequence, a strategy has been designed to identify and filter dark colours from the LCT: initially, the RGB coordinates are projected into the HSL (Hue, Saturation, Lightness) colour space. Next, the lightness component is evaluated and, if it falls below a predetermined threshold, the colour is marked as dark and removed from the LCT.

In the last step, characterising colours for each land type are identified by comparing their normalised count values across all categories. More specifically, a colour is assigned to the category corresponding to the maximum count if both these conditions hold true: (i) no other count in the row exceeds a predetermined fraction (e.g. 75%) of the maximum count, and (ii) the cumulative count of the other rows

does not exceed a set threshold (e.g. 1.5 times the maximum count). Such conditions ensure that a given colour is prominently found within a specific category, therefore qualifying it as a characterising colour for that category. Conversely, colours that do not meet both conditions are designated as *unassigned colours* (UCs) and constitute the UC set.

The final output of this phase is represented by the CC sets, one set for each LULC category plus the background, and the UC set. These sets are used to identify LULC categories in the classification phase, discussed in the next section.

3.2.3 Land Use and Land Cover Classification

The LULC classification is performed in unannotated images, and comprises two steps (see Figure 3.22): (i) pixel marking, which assigns pixels to a specific LULC category using information from the CC and UC sets, and (ii) pixel aggregation, which utilises windows with increasingly larger sizes to cluster pixels that belong to the same LULC category.

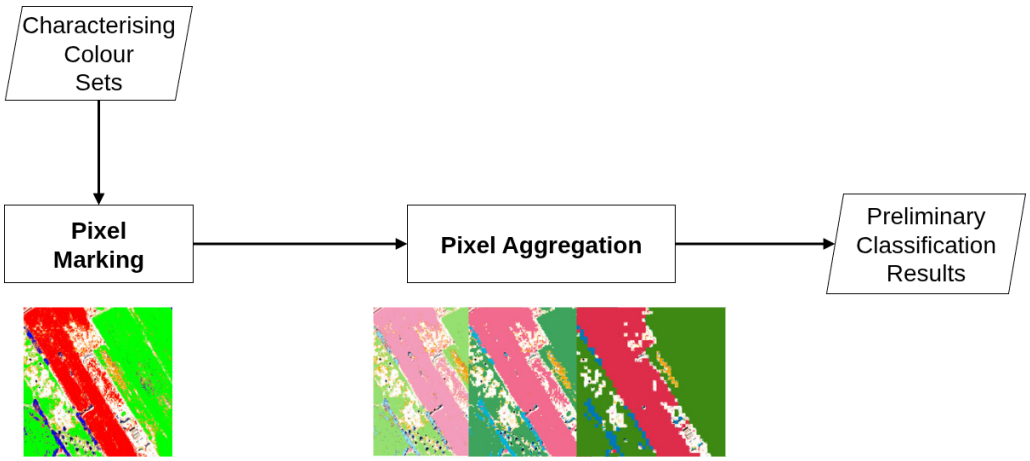


Figure 3.22: A visual representation of the two steps in the LULC classification phase. Firstly, pixels are marked based on their category. Then, pixels are aggregated based on their proximity.

Initially, a unique marking colour is defined for each CC set and the UC set. Then, images are analysed and pixels are marked based on their category. Figure 3.23 shows an example output from this operation.

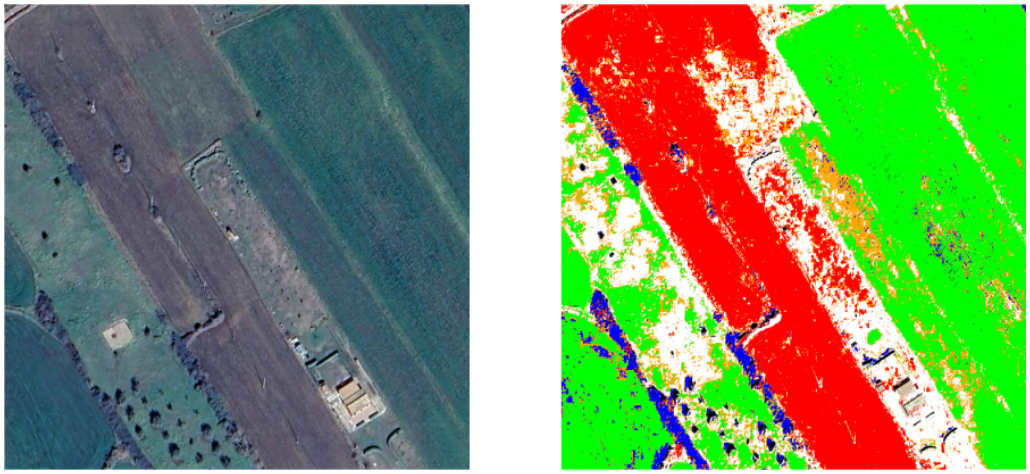


Figure 3.23: Example results for the pixel marking operation. The original aerial image is shown on the left, while in the image on the right three LULC categories are visible: irrigated land (marked in green), non-irrigated land (marked in red), and forest (marked in blue). Additionally, pixels displaying colours in the UC set are marked in orange, and pixels belonging to the background category are marked in white.

The reason for marking colours in the UC set is as follows: unassigned colours, by definition, appear in multiple categories with comparable frequency, and LULC categories typically occupy a large portion of the image without substantial intermixing. As a consequence, it is a reasonable assumption that whenever pixels marked as UCs are detected near a cluster of colours belonging to a specific category, they can ultimately be assigned to that category. The proposed approach takes advantage of this observation and utilises moving square windows to aggregate pixels based on their density, a strategy that has proven effective in enhancing the classification process [76]. When initiating a pixel aggregation run, three parameters are set: (i) the size of the scanning window; (ii) the threshold density that determines the colouring of the entire window; and (iii) a new marking colour for every category.

The aggregation strategy operates as follows. In each iteration, the number of marked pixels within the window is evaluated. If the total count of exactly one category exceeds a threshold, e.g., 50% of the total window pixels, then the entire window is coloured with the corresponding new marking colour. Conversely, if no individual category count exceeds the threshold, the count of UC pixels within the window is also included into the count for every category. If, after the updated count values, exactly one category exceeds the threshold, the entire window area is marked with the corresponding colour. Otherwise, the area remains unaltered because it was not possible to unambiguously assign it to a specific category.

This assessment is performed for the whole image and, depending on the characteristics of the examined regions, employing multiple aggregation passes with increasingly larger window sizes can substantially enhance the classification's performance. Figure 3.24 shows the detection results after three consecutive iterations utilising window sizes of 3×3, 9×9 and 25×25, respectively.

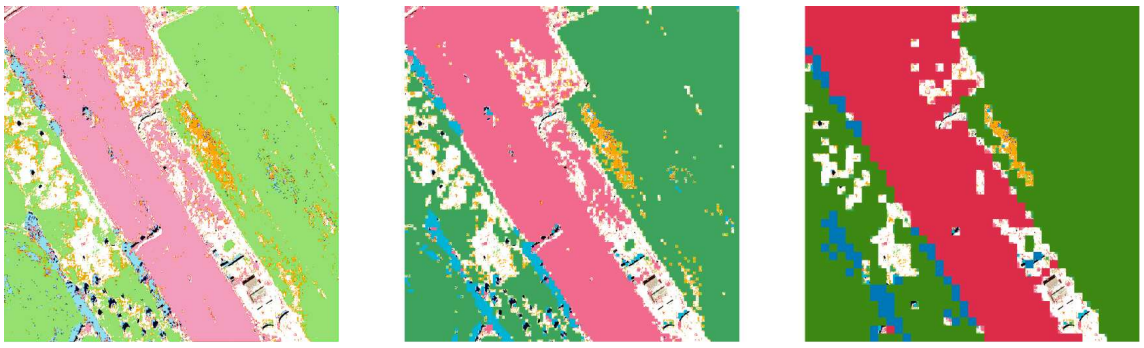


Figure 3.24: LULC classification results obtained after three consecutive scanning window iterations, utilising window sizes of 3×3 (left), 9×9 (middle) and 25×25 (right).

The reason for employing a new marking colour in each classification refinement pass is based on the fact that each subsequent iteration evaluates the marked colours within a progressively wider area. If a window in a successive pass includes previously

marked pixels but their cumulative count is not sufficient to trigger the threshold, this indicates that a significant number of pixels within the larger window do not feature characterising colours. Therefore, in this situation, the area enclosed by the window remains unmarked. When assessing the classification results, only the marking colours from the final refinement run are considered.

The last phase of the proposed approach, post-processing, is performed on these preliminary classification results to tackle incorrect detections and further enhance the classification, as discussed in the next subsection.

3.2.4 Image Post-processing

The goal of the post-processing phase is to improve the previous results by using a twofold approach: (i) denoising, which removes incorrectly marked tiles *outside* LULC categories, and (ii) closing, which strengthens classification *within* marked LULC areas. Figure 3.25 provides an overview of post-processing.

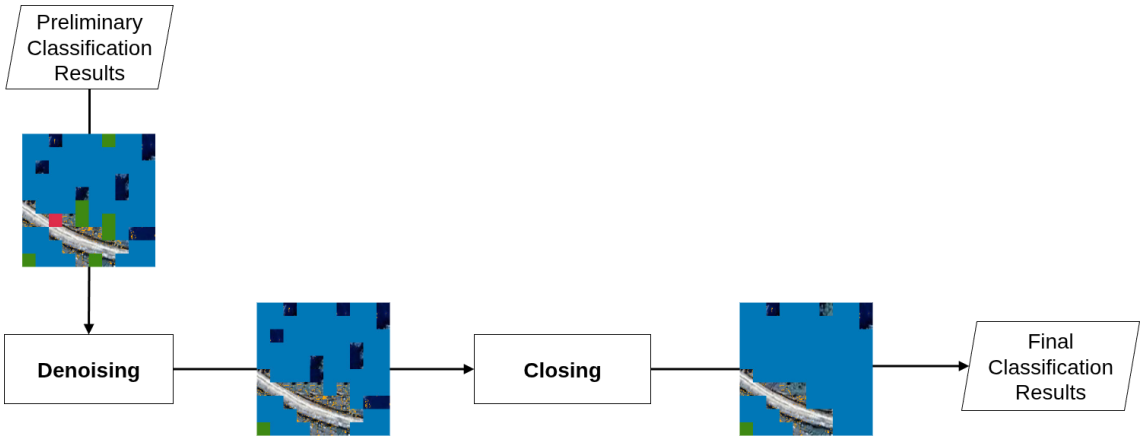


Figure 3.25: Overview and visual representation of the two key steps in the final post-processing phase: the denoising operation removes incorrectly marked tiles, and the closing operation refines the detection within LULC regions.

Occasionally, characterising colours of a given LULC category are found and marked within the area of a different LULC category. There are several possible explanations for this phenomenon. Firstly, relatively small regions of a certain LULC type can be enclosed by a different and much larger LULC class. For example, a small patch exhibiting characterising colours of irrigated land can be found within a broader forest area. Additionally, background regions may also contain limited areas that display characterising colours of LULC types.

In all these cases, the key observation is that correctly identified LULC tiles typically extend over significantly large areas, while misclassified areas are often isolated or cover regions of relatively limited extension. Therefore, the denoising operation employs a moving window that expands whenever it encounters a coloured tile to include and evaluate the surroundings. If, within the enlarged window, the total count of pixels having the same colour as the original tile is below a set threshold, the marking operation is determined to be incorrect and it is reverted.

An additional classification refinement can be performed when unmarked tiles are found within marked areas. This situation can happen because sometimes LULC categories contain small patches of land that exhibit non-characterising colours. The strategy to assign the unmarked tiles to the encompassing LULC category is once again based on the scanning window, and the assessment logic is the opposite of denoising. More specifically, whenever an unmarked tile is encountered, the window expands to include the surroundings and, if in the enlarged window the count of a LULC category is above a threshold, the tile is coloured with the corresponding category colour. Figure 3.26 shows example results after applying the post-processing.

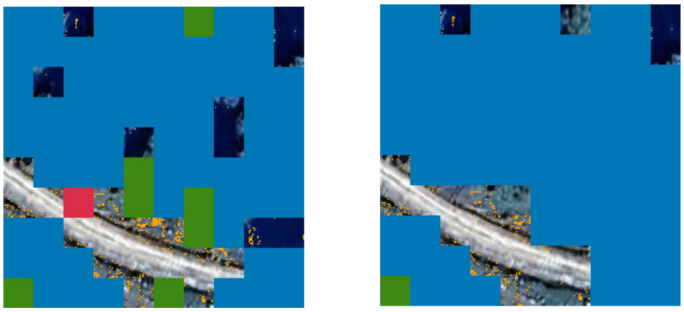


Figure 3.26: A zoomed-in section of an image where LULC classification has been performed (left), and the results after applying post-processing (right). After denoising, incorrectly marked green and red tiles are removed, and after the closing operation unmarked tiles are assigned to the surrounding blue category.

After post-processing, classification results are finalised. In the next section, we present the experiments conducted on a set of aerial images and discuss the results to show the effectiveness of our proposed approach.

3.2.5 Results and Discussion

To perform our experiments, we collected a total of thirty-eight aerial images from a rural area situated in the southeastern region of Sicily, Italy. The image resolution is 0.3 m/px and the size is 1024×1024 pixels. We defined three LULC categories: forest, irrigated land and non-irrigated land. The corresponding annotation colours are blue, green and red, respectively.

Initially, 80% of the aerial images were randomly selected for the extraction of characterising colours (CC), while the remaining 20% were used to assess the performance of the approach. Next, colours and counts were extracted for each LULC category, plus the background, and the Land Colour Table was built. To identify dark colours, the RGB coordinates were projected into the HSL colour space and a threshold of 0.16 (in the range $[0, 1]$) was defined for the lightness component. Any colour that fell below the threshold was removed from the table.

Subsequently, the CC sets and UC set were generated by applying a threshold of 75% for the fraction of the maximum count and a threshold of 1.2 times the maximum count for the sum of the other counts. Table 3.11 shows colour statistics after the CC set generation phase, while pixel counts statistics are displayed in Table 3.12.

LULC Category	**Total Colours**	**Colour Ratio (%)**		
		Dark	**UC**	**CC**
Forest	172,607	11.53%	11.75%	47.01%
Irrigated	87,487	10.53%	23.18%	34.86%
Non-irrigated	55,404	6.03%	36.61%	33.70%

Table 3.11: Results of the characterising colour set generation phase. For each LULC category, the total number of colours, and the corresponding percentage of dark colours, unassigned colours (UC) and characterising colours (CC) are displayed.

LULC Category	**Total Pixels**	**Pixel Counts (%)**		
		Dark	**UC**	**CC**
Forest	6,328,713	33.71%	7.71%	46.45%
Irrigated	12,273,030	0.44%	5.38%	88.00%
Non-irrigated	6,451,967	0.08%	3.70%	92.09%

Table 3.12: Pixel counts for dark colours, unassigned colours (UC) and characterising colours (CC). For each LULC category, the total number of pixels, and the fraction of pixels covered by the corresponding colour sets are shown.

After obtaining the CC sets for each category and the UC set, we proceeded to perform the classification on the test images. Pixels were marked based on the information in the colour sets, and were subsequently aggregated by using three scanning windows, of size 3×3, 9×9 and 25×25, respectively. In all three iterations, the threshold for colouring the window was set to 0.5.

The results were post-processed by using: (i) a denoising 75×75 window, with a threshold set to 1/3 of the window area; (ii) a closing 75×75 window with threshold set to 2/3 of the window area. Figure 3.27 shows the final classification results on a test image.

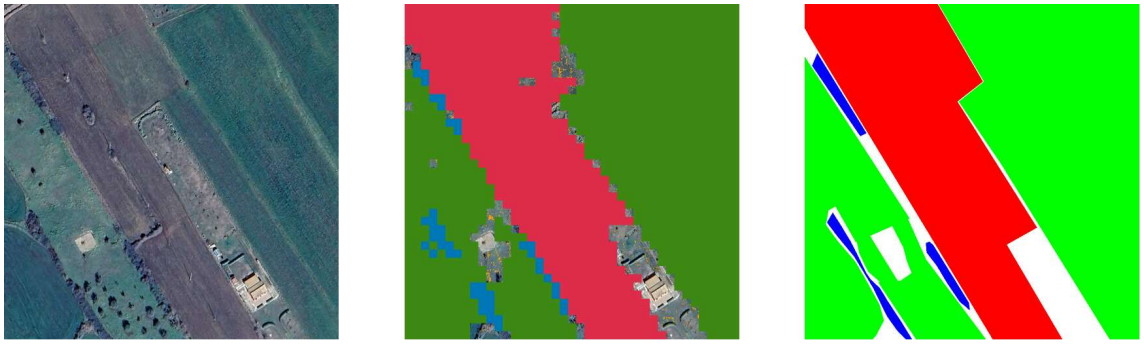


Figure 3.27: Final classification results on a test image after executing the proposed algorithm (image in the middle). The original image is shown on the left, and the corresponding annotations are shown on the right.

We evaluated the performance of the proposed approach by assessing the following metrics: accuracy, precision, recall, and F1 score. Table 3.13 presents the aggregated metrics for the LULC classification task executed over the test dataset of seven images.

LULC Category	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Forest	0.898	0.664	0.742	0.694
Irrigated	**0.948**	**0.798**	0.886	0.826
Non-irrigated	0.926	0.771	**0.971**	**0.850**

Table 3.13: Evaluation metrics for the proposed LULC classification algorithm.

In Table 3.13, the highest recorded value for each metric is indicated in bold. Irrigated land achieves the highest accuracy and precision scores, with values of 0.948 and 0.798, respectively. Non-irrigated land attains the highest recall and F1 scores, with respective values of 0.971 and 0.850.

As shown in Table 3.11, the CC set comprises just over one third of the total colours for both irrigated and non-irrigated land, with ratios of 34.86% and 33.70%, respectively. Table 3.12 illustrates that these CC sets cover the vast majority of irrigated and non-irrigated areas, representing 88.00% and 92.09% of total pixels, respectively.

This analysis demonstrates that a relatively small subset of characterising colours can be effectively employed for the LULC classification task. Unlike most ML- and DL-based methodologies, our algorithm does not rely on large annotated datasets: CC sets are automatically generated by analysing a small number of images. Furthermore, training and fine-tuning operations are unnecessary, resulting in quicker execution times and reduced demands on memory and computational resources.

The lower performance in the forest category can be explained by the following two observations. Firstly, depending on the illumination conditions of the aerial images, shadows cast by trees may cover considerably large areas, often resulting in dark regions within the forest itself. Table 3.12 shows that dark colours account for about one third (33.71%) of the total forest pixels. The proposed algorithm filters out these shades, with the forest CC set ultimately representing less than half of the total pixels (46.45%). Although the high prevalence of dark colours may prevent reaching the threshold for marking tiles in forested areas, the closing operation in the post-processing phase has proven effective at handling darker regions, as shown in Figure 3.25. The second observation is that forest annotations in the aerial dataset include relatively small patches or stripes of trees, often situated in boundary regions between LULC categories (examples of forest annotation stripes can be observed in Figure 3.27). To tackle this issue, we leverage the flexibility of our approach, enabling us to dynamically modify the elements in the colour set for marking windows. Specifically, whenever the marking threshold is not reached, CC sets are extended to include the unassigned colours. These colours are typically found in border regions, where the shades of one category gradually transition into another category. As illustrated in Table 3.12, UCs account for a significant fraction of forest pixels (7.71%), indicating their effectiveness in enhancing classification results. This

proves the robustness of our approach, which is capable of evaluating and handling areas displaying greater colour heterogeneity.

3.2.6 Summary

This study presented an innovative and robust approach to classify Land Use and Land Cover (LULC) types from aerial images. The proposed methodology analyses a small set of images to extract colours from pixels and, through a systematic process, builds sets of characterising colours for each land category. The automated LULC classification was accomplished by leveraging the inherent land characterising colours and their respective densities to improve detection performance. This effectively reduces false positives and false negatives by quantifying colour densities within image tiles, all without requiring prior manual analysis or annotation.

The approach was tested on a set of unannotated images. The results showed that the approach accurately classifies lands according to their cover, achieving high precision and recall values. Our proposed approach exhibits fast execution times, as it forgoes a computationally intensive training phase and operates efficiently without relying on a large dataset of annotated images. The extracted characterising colours were used for images from diverse regions, thereby showing the versatility and robustness of the proposed approach. This methodology can be applied to several domains, such as assessing crop health, tracking deforestation and identifying suitable locations for the deployment of PV installations.

Chapter 4

Optimising Energy and Data Flow in Renewable Energy Communities

The increasing adoption of Renewable Energy Sources (RES), particularly distributed solar PV panels, is changing how power systems operate. Energy Communities (ECs) represent a new paradigm for coordinating local generation, consumption and storage, but the intrinsic RES variability poses challenges for grid stability and optimal energy balancing. These challenges require both effective mechanisms to promote coordinated and energy-aware user responses and reliable methods to certify the integrity of energy data.

The first section of this chapter presents a novel strategy for user engagement and adaptive energy optimisation. The framework discussed in this thesis was extended to integrate a software architecture where the local aggregator monitors net energy balance, applies threshold-based rules to detect peak events, and elicits a response from members via a gamified, incentive-based mobile application. Extensive evaluations were conducted on real-world datasets to demonstrate that this combined approach can effectively reduce peak consumption and strategically employ energy storage systems, while coordinating energy-aware behaviour among participants.

The second section of this chapter describes a further extension of the framework to provide proof of integrity for data in the context of incentive-based Demand Response (DR) programmes. These programmes offer financial rewards to consumers who agree to reduce their load demand in response to signals from aggregators or system operators. To resolve disputes and detect data tampering, we propose a dual-layered storage strategy. The aggregator assembles data batches, requests the writing of 32-byte hashes on a public blockchain, and retains the complete data records off-chain, in cloud storage. This methodology ensures user privacy, minimises blockchain costs and provides an immutable audit trail. Our empirical evaluation on a real-world large-scale DR trial demonstrates that our approach achieves cost-efficient data certification capabilities.

These two extensions of the framework provide an innovative and scalable approach to managing both the behavioural and technical aspects of community energy systems. This chapter is organised as follows. Section 4.1 presents the energy optimisation system that promotes user engagement and coordination, and Section 4.2 details the blockchain-based energy data certification system.

4.1 Engagement and Adaptive Optimisation in Energy Communities

In recent years, there has been a growing adoption of renewable energy sources (RES), mainly wind and solar photovoltaic (PV), whose share is projected to increase from 30% in 2023 to 46% in 2030 [44]. This trend reflects global efforts to decarbonise energy systems and reduce dependence on fossil fuels. However, RES are inherently variable, as their output depends on changing natural and meteorological conditions.

These unpredictable fluctuations pose several challenges, including maintaining grid stability, balancing electricity supply with demand, and handling the inaccuracy of energy forecasts [130].

In this context, a promising solution is represented by Energy Communities (ECs), defined by European Union directives as open and voluntary organisations of citizens engaged in energy-related activities [23]. The new prosumer role was introduced, as the evolution of the traditional consumer, who is also able to generate electricity using RES, mainly solar PV and Energy Storage Systems (ESS) [61]. PV-equipped prosumers can feed surplus energy into the electrical grid, and ESS-enabled prosumers can strategically charge or discharge their batteries to maximise self-consumption [68]. Data measuring the power consumption, generation and storage of EC members are remotely managed by the local aggregator, an entity that acts as a central orchestrator of distributed RES within the community, and coordinates energy optimisation strategies to support grid operations [53].

To realise the full potential of ECs, active and sustained member participation is paramount. This can be fostered by methodologies and stimuli that promote energy-aware decisions. One such methodology is Demand Response (DR), where financial compensation is offered to users who adjust their energy usage in response to requests from an aggregator or system operator [40]. In addition, gamification has been applied in the energy domain [51]. The main idea is to apply elements of game design, such as challenges and leaderboards, to stimulate changes in energy consumption or production. Although these approaches have shown promising results, the integration of games with financial rewards and community-wide energy optimisation strategies is still limited [94].

We enhanced the framework presented in this doctoral thesis by integrating an

innovative software system that combines gamified engagement with data-driven optimisation strategies, derived from the community's aggregated energy profile. Financial rewards and targeted challenges are proposed to EC members via a mobile application. The integration of incentives and socially engaging features encourage users to align with optimal energy use, benefiting both participants and grid operators.

The main contributions of this work are: (i) we propose a novel software architecture for adaptive energy management in ECs, which combines monitoring, flexible threshold-based strategy definition, and member coordination via a mobile application; (ii) we introduce a gamified incentive mechanism that integrates financial rewards and targeted challenges to promote energy-aware decisions aligned with the community's welfare; (iii) we simulate and evaluate our approach on multiple EC configurations using a publicly available dataset based on real-world data; (iv) we show that our proposed approach can effectively tackle energy fluctuations by means of a coordinated response that builds on observed user patterns to achieve optimal management and harness the potential of renewable energy infrastructure.

The remainder of this work is structured as follows. Section 4.1.1 presents our proposed software architecture. Section 4.1.2 describes the experiments we conducted on real-world data. Finally, Section 4.1.3 draws the conclusions.

4.1.1 System Architecture

We propose an innovative software system for the management of Energy Communities (ECs), designed to dynamically monitor energy flows, assess the net energy balance, and coordinate adaptive energy optimisation strategies. Figure 4.1 illustrates the system architecture, showing the main entities and data flows.

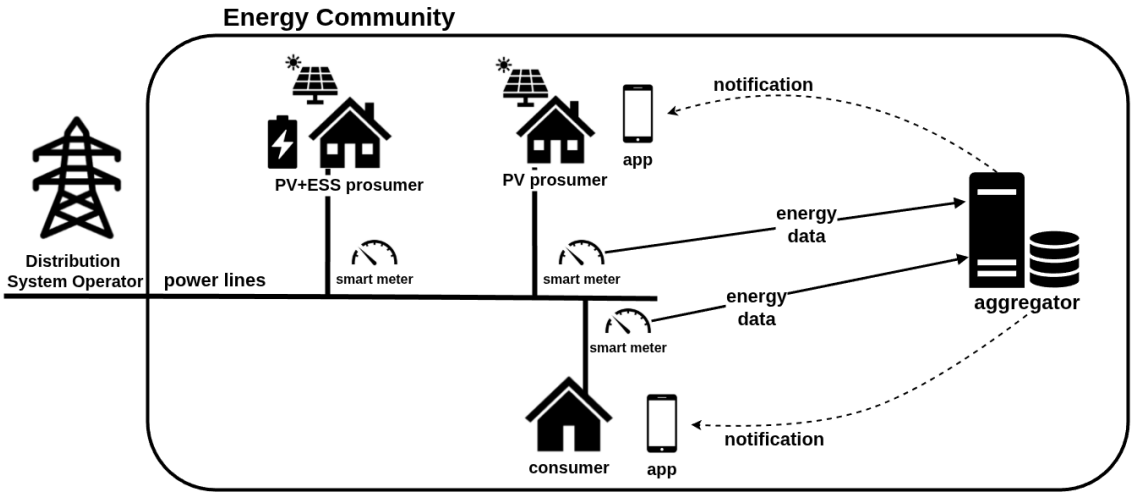


Figure 4.1: Architecture diagram of the proposed software system for adaptive energy management in Energy Communities.

The Distribution System Operator (DSO) manages the electricity distribution networks and infrastructure and is responsible for delivering power to consumers and prosumers. In our system, we consider two types of prosumers: those equipped with PV panels and those with both PV panels and Energy Storage Systems (ESS). Smart Meters (SMs) are embedded IoT devices installed on users' premises, integrating sensing and communication capabilities. These devices measure the incoming or outgoing flow of electricity and periodically transmit the recorded data, over the Internet, to the energy aggregator, which stores them in a local database.

The aggregator is a cloud-based service that collects and handles energy data for all community members. By combining data from SMs, the aggregator actively monitors the net energy balance of the EC. Based on such data, it forecasts energy use patterns and readily detects abnormal fluctuations by comparing observed values to predefined thresholds. These thresholds are either derived from optimal consumption profiles or agreed upon with the DSO to maintain grid stability. When net energy exceeds the thresholds, the aggregator sends a notification to users requesting changes

in their energy consumption patterns. For example, if monitoring data reveal that net energy falls below a predefined consumption threshold, users are signalled to reduce their energy demand, and ESS-enabled prosumers are requested to discharge their batteries into the grid. Conversely, in overproduction scenarios, ESS-enabled prosumers are notified to redirect their energy surplus to charging their batteries.

Aggregator's requests are delivered to users via in-app notifications. The app includes gamification features and financial incentives, defined by the aggregator, to promote user participation. Users receive rewards for compliance and can engage in goal-oriented tasks that promote collective optimisation. For example, the app may include challenges such as "Reduce your evening energy use" or "Charge your battery before sunset". Leaderboards and achievement badges displayed in-app further incentivise continuous engagement and the adoption of energy-aware choices that align with the community's goals. The adaptive strategies of the aggregator, combined with financial incentive schemes, enable a coordinated response to address energy fluctuations, illustrating the key role of the proposed software system in realising the full potential and flexibility of renewable energy infrastructure.

4.1.2 Experiments and Discussion

To evaluate the feasibility of our proposed software system, we conducted an analysis and assessment of energy data utilising a residential prosumers dataset publicly available at Figshare [141] and presented in [142].

The dataset was built using real-world energy data from 2,000 consumers living in the same neighbourhood, aggregated with renewable energy sources data such as the generation of PV panels and ESS operations. Specifically, the PV profiles were created using the solar ninja tool [100], and the ESS profiles were obtained from

real-world industrial specification and data [142]. Each row in the dataset contains hourly energy consumption/production data in kWh, labelled with a season.

Real-world ECs greatly vary among different countries, ranging from fewer than 50 members to over 2000 [54]. In our analysis, we considered the scenario of a medium-sized EC comprising 400 participants. In total, we designed and analysed ten different EC configurations. Two types of prosumers were considered: (i) PV-equipped prosumers, and (ii) PV+ESS-equipped prosumers. For each type, we modeled five ECs with prosumer ratios of 0%, 25%, 33%, 50%, and 75%, respectively. For each configuration, hourly energy profiles were obtained as follows. Based on the predetermined prosumer ratio, entries were randomly extracted from the consumer and prosumer datasets. To account for the variability introduced by randomness, the sampling process was repeated 500 times for each configuration. The final energy values were obtained by averaging the hourly data across all 500 iterations.

Figure 4.2 compares the seasonal energy profiles of ECs with 75% prosumers. Negative values on the y-axis indicate that the energy is imported from the grid, while positive values represent energy that is exported to the grid.

Both EC configurations exhibit a similar trend, with surplus energy during daylight hours and net imports in the evening. However, the PV+ESS scenario features a smoother profile, with lower midday peaks and reduced evening demand. This suggests that ESS charge during peak solar generation and discharge their stored energy later in the day.

Table 4.1 and Table 4.2 show daily net energy balances (in MWh), for all seasons and different ratios of PV prosumers and PV+ESS prosumers, respectively.

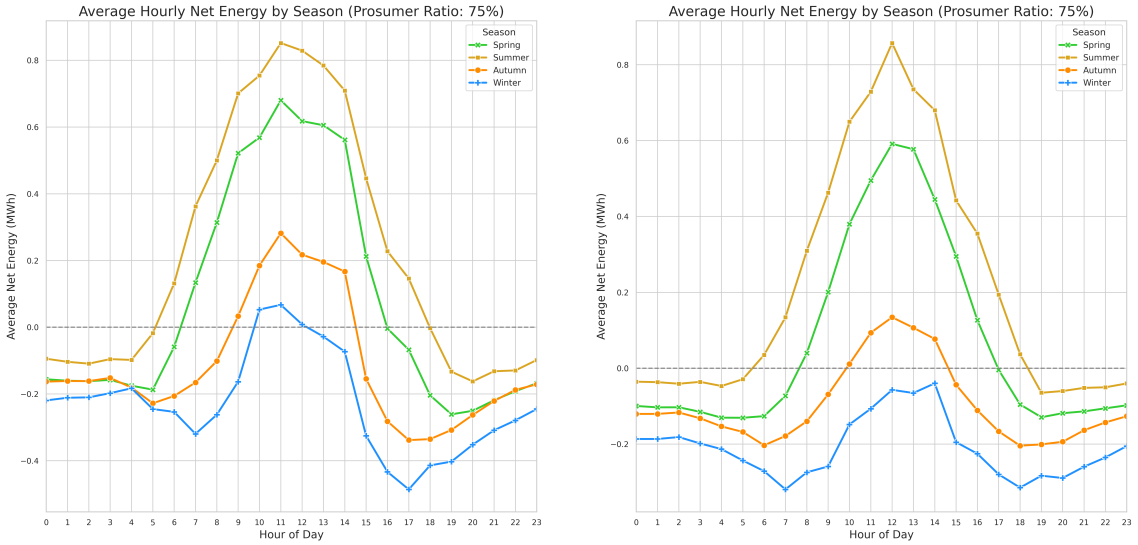


Figure 4.2: Average hourly net energy by season for a community of 400 members. The plot for 75% prosumers with PV is shown on the left, while the plot for 75% prosumers with PV and Energy Storage Systems (ESS) is shown on the right.

	PV Prosumer Ratio				
Season	**0%**	**25%**	**33%**	**50%**	**75%**
Autumn	−4.94	−4.17	−3.93	−3.44	−2.69
Winter	−6.33	−6.06	−5.95	−5.77	−5.49
Spring	−5.05	−2.77	−2.03	−0.47	+1.81
Summer	−3.70	−0.71	+0.25	+2.28	+5.24

Table 4.1: Net balance of Energy Community for various PV prosumer ratios (in MWh).

	PV+ESS Prosumer Ratio				
Season	**0%**	**25%**	**33%**	**50%**	**75%**
Autumn	−4.94	−4.08	−3.80	−3.20	−2.31
Winter	−6.32	−5.90	−5.75	−5.49	−5.06
Spring	−5.06	−2.82	−2.11	−0.57	+1.62
Summer	−3.70	−0.75	+0.18	+2.20	+5.12

Table 4.2: Net balance of Energy Community for various PV+ESS prosumer ratios (in MWh).

The analysis of these tables reveals that the inclusion of prosumers significantly reduces the overall energy consumption in the EC. This effect is particularly noticeable in spring and summer, where a prosumer ratio of only 25% leads to a consumption decrease of approximately 45% and 80%, respectively. Both spring and summer feature a net daily surplus when the prosumer ratio is 75%, and summer in particular experiences a dramatic increase from the net consumption of -3.7 MWh (0% prosumers) to the net production of over 5 MWh. In seasons with less solar irradiance, while the net balance remains negative across all configurations, it can be observed that PV+ESS prosumers provide consistently lower values than the PV prosumers case.

Overall, the results of our experiments indicate that storage systems help mitigate peak consumption and suggest that ESS-enabled residential users actively engage in optimising battery usage, revealing an awareness of the strategic value of these storage devices at the individual level. The aggregator can capitalise on this trend by offering financial incentives that further promote and scale behaviours that align with optimal energy management. Its access to aggregated data enables it to define adaptive

strategies from a community-wide perspective. Consequently, the gamified app can present users with targeted and carefully designed challenges, such as "Maintain your battery charge above 50% throughout the week" or "Reduce your energy use by 20% between 6 p.m. and 8 p.m.". These challenges, supported by financial rewards and social features such as leaderboards, further reinforce user engagement and align with community-level optimisation.

4.1.3 Summary

This work presented an innovative and adaptive software system that enables coordinated energy management in energy communities. By collecting data from smart meters installed on consumer and prosumer premises, the community aggregator actively monitors the overall energy balance. When predefined usage thresholds are exceeded, the aggregator notifies users via a gamified app that promotes optimal energy use through financial incentives and targeted challenges.

The feasibility of our proposed system was evaluated by conducting simulations on a publicly available dataset based on real-world energy profiles. The results show that the adaptive behaviour of the aggregator can be effectively employed to tackle energy fluctuations with flexible strategies that fully harness the potential of renewable energy sources. Future developments in this research include the deployment of the proposed software system to assess its computational and economic viability. Moreover, further investigations are required to design effective targeted challenges for all categories of users. Finally, developing and integrating strategies for identifying and handling outliers and faulty values in energy measurements represents an important direction for future research.

4.2 Certifying Energy Data in Demand Response Programmes

The adoption of Renewable Energy Sources (RES) has experienced steady growth, driven by the need to reduce reliance on fossil fuels. Solar photovoltaic (PV) power in particular has emerged as one of the key sustainable technologies, with an average annual growth rate exceeding 35% over the past decade [10, 115]. Global PV capacity almost doubled between 2019 and 2022, and the overall share of RES in the electricity mix is projected to rise from 30% in 2023 to 46% by 2030 [1, 44]. Despite this rapid growth, RES are characterised by intrinsic variability, as their generation depends on meteorological conditions. This unpredictability creates significant challenges for energy systems, including ensuring grid stability and balancing net energy [130].

In recent years, Demand Response (DR) programmes have experienced a growing interest as a methodology for tackling the unpredictable power generation output from renewable energy sources [7]. Broadly, there are two main categories of DR strategies: (i) price-based, and (ii) incentive-based [40]. Price-based approaches employs dynamic energy pricing to encourage end users to modify their consumption patterns. However, studies have shown that efficacy of this approach can be low [92]. On the other hand, incentive-based DR programmes offer rebates to consumers who agree to reduce their load demand in response to signals from system operators.

However, when financial incentives are awarded, disputes may arise. For example, a customer's claim can contest the reward amount and ask for a higher remuneration. In such cases, it is paramount to obtain proof of integrity for recorded data and implement strategies that can deter data tampering activities by malicious entities. In this context, we propose an approach that leverages the distributed blockchain

network to foster trust among all energy system entities. Specifically, data hashes are stored on-chain to: (i) provide proof of integrity for the associated data, and (ii) detect data tampering that results from a hash mismatch.

Numerous approaches in the literature have proposed the integration of public blockchain with DR programmes to enhance and accelerate their adoption [55, 73, 99, 104, 126, 135]. However, these approaches exhibit limitations such as high costs, significant delays in finalising the large amount of transactions, a lack of user data privacy, or a missing validation on real-world datasets. In contrast, our dual-storage methodology that integrates cloud storage with blockchain transactions overcomes these limitations and provides a reliable and cost-effective data certification service.

We further developed the framework presented in this doctoral thesis by defining a decentralised software system that provides proof of integrity for energy measurements sent by smart meters during DR events. In our architecture, the main actor is the community aggregator, that continuously collects energy data, assembles them in batches of predetermined size, and requests the writing of data batches' hashes on the public blockchain. Once finalised, the receipt of the writing operation is associated with the original data, which are eventually committed to a remote cloud server for long-term storage. This dual-layered storage strategy ensures verifiability, scalability and security, while minimising the costs of interacting with the blockchain [38, 48, 139]. The correctness of the recorded energy data can be proved by comparing the hash of the data on the cloud storage with the corresponding hash retrieved from the blockchain [42].

The key contribution of this research is the design and implementation of a decentralised system that employs the immutability property of blockchain storage to provide proof of integrity for recorded energy data, while also ensuring data

privacy. Our solutions provide transparent and traceable data management, fostering trust without relying on a centralised authority for data certification. Moreover, our comprehensive cost analysis demonstrates that the total cost for using blockchain during a complete DR event represents only a small fraction of the incentive plan, thereby showing the feasibility of our approach in real-world scenarios.

The remainder of this section is structured as follows. Section 4.2.1 discusses the motivation and presents the proposed software architecture for certifying DR events. Section 4.2.2 describes our experiments conducted on a real-world DR dataset. Section 4.2.3 discusses the results. Finally, Section 4.2.4 draws the conclusions.

4.2.1 System Architecture

Energy Communities (ECs) were introduced in European Union directives as voluntary organisations of citizens involved in energy-related activities [23]. As an evolution of the traditional consumer, the new prosumer role was introduced to describe end users who can also generate electricity using RES, primarily solar PV and Energy Storage Systems (ESS) [61]. PV prosumers can inject surplus energy into the electrical grid, while ESS-enabled prosumers can offer flexibility services by strategically charging or discharging their batteries [68]. The local aggregator is an entity that remotely manages the data of EC members. It acts as a central orchestrator of distributed energy resources and coordinates energy optimisation strategies to support grid operations [53]. Figure 4.3 illustrates the main entities in energy communities.

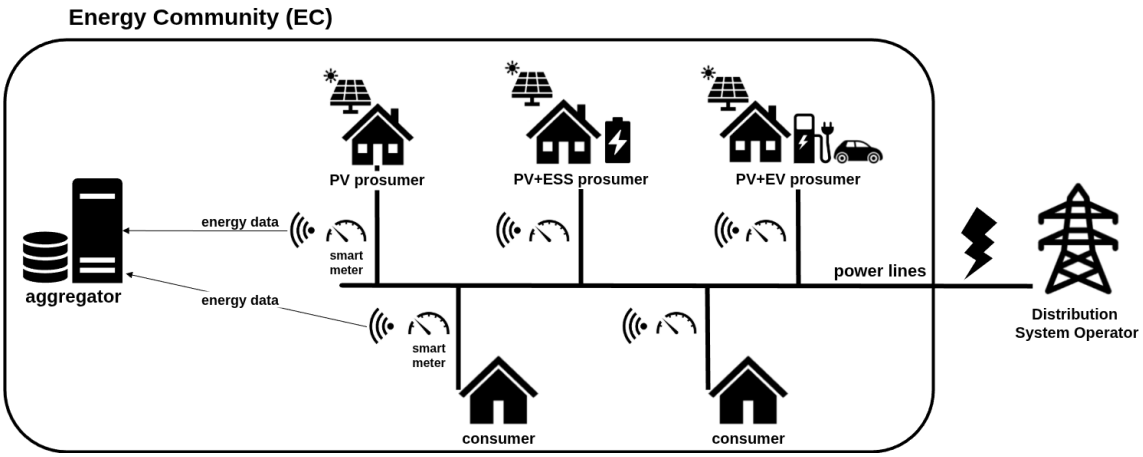


Figure 4.3: Main entities in Energy Communities.

The Distribution System Operator (DSO) is responsible for managing the power distribution network and infrastructure, and delivers electricity to end users. In our system, we consider two types of energy users: (i) traditional consumers, and (ii) prosumers, who can both consume and produce electricity. Prosumers own one or more distributed RES, including rooftop solar PV panels, ESS, and electric vehicles (EV). Each user of the EC is equipped with a Smart Meter (SM), an embedded IoT device that measures the incoming or outgoing flow of electrical power. The SMs periodically send energy data over the Internet to the community aggregator, which stores them in a local database.

The aggregator is an agent that collects and handles energy data for all community members. Using data transmitted by SMs, it continuously monitors the EC's net energy balance and forecasts short-term energy profiles for the community. Anomalies and peaks are identified by comparing observed and predicted values to predetermined thresholds, which are defined to ensure compliance with optimal energy profiles or to promote grid stability in agreement with the DSO. When these thresholds are exceeded, the aggregator requests users to modify their energy utilisation pattern to

mitigate or altogether prevent undesirable deviations from the baseline profile. For example, if monitoring data indicate a situation of overconsumption, users are notified to reduce demand. Additionally, ESS-enabled and Vehicle-to-Grid (V2G) prosumers are signalled to discharge their batteries into the electricity network. Conversely, when net energy exceeds the predetermined production threshold, prosumers are prompted to use the energy surplus to charge their batteries.

To improve members' compliance with energy usage modification requests, the aggregator can issue an incentive-based Demand Response (DR) programme, where users are rewarded for reducing their energy load. However, the distribution of remuneration inherently introduces the risk of disputes, such as a user contesting their reward amount or claiming eligibility for a higher reward. To address this challenge, system operators must implement mechanisms to ensure that incentives are correctly assigned and protect against data tampering by malicious entities.

In this context, the blockchain technology represents an innovative and decentralised solution for ensuring data integrity. In essence, blockchain acts as a distributed ledger that leverages cryptographic functions to securely record data in a chain of blocks. Blockchain properties ensure that content in a finalised block remains immutable, preventing deletion and modification. Therefore, blockchain fosters trust for entities that do not rely on a central authority and require integrity for the stored data. Ethereum and Ethereum Virtual Machine (EVM)-compatible blockchains have further developed this technology by enabling the definition of programmable smart contracts. The execution of these contracts is triggered by a set of predetermined conditions, and their source code is committed to and stored on the blockchain, ensuring transparency.

We leveraged the properties of blockchain technology to design an innovative

architecture that ensures secure and transparent certification of energy data during DR programmes. The proposed system certifies critical energy data, promoting trust and accountability. The architecture diagram in Figure 4.4 illustrates the key entities and the main data flow of our proposed system.

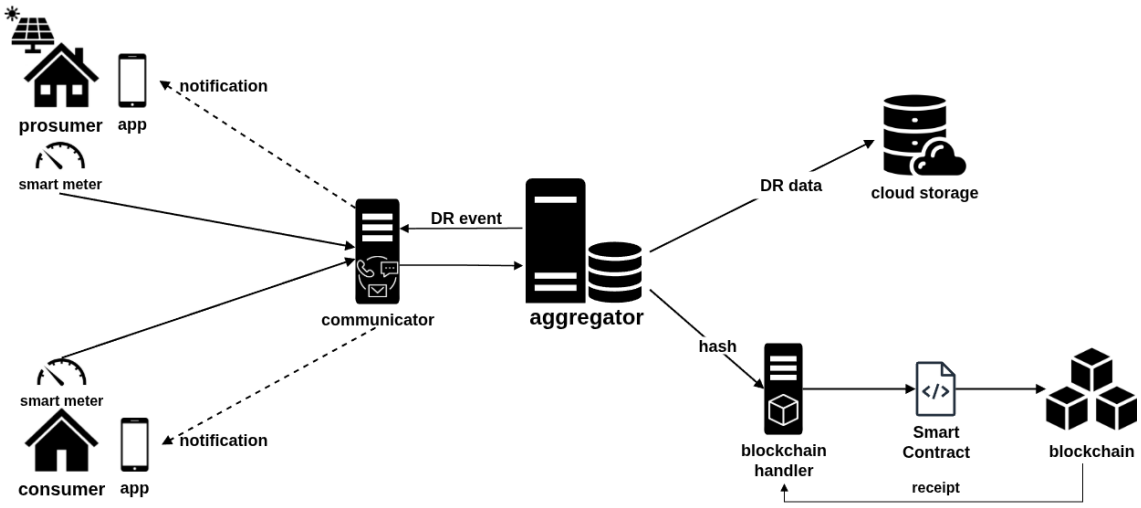


Figure 4.4: Main components and data flow of the proposed architecture for certifying energy data in demand response programmes.

The main agent in our proposed system is the EC aggregator, which is responsible for starting DR events, managing energy data sent from smart meters, and orchestrating a dual-layer storage strategy that certifies data integrity. The aggregator continuously monitors the community's energy usage, and employs predefined thresholds to ensure that the net energy profile does not diverge from optimal patterns. When the aggregator detects or forecasts that the reference thresholds are exceeded, it defines and activates a new DR event. DR parameters set by the aggregator include: the start time, the event duration, and the financial incentive per unit of reduction awarded to complying users. This information is transmitted to the communicator service, which is responsible for delivering DR event data to EC members' smartphones using messages, SMS or in-app notifications, depending on

the users' preferences. Each user independently evaluates the conditions and rewards associated with the event, and replies to the communicator with the decision to participate or not in the programme.

The communicator sets a timeout immediately after notifying users and, since participation to DR programmes is on a voluntary basis, it is assumed that users who do not reply within the predefined time interval will not take part in the DR event. Consequently, unless users explicitly opt-in for the event, either by sending a reply or configuring the app, their energy data are not evaluated, and they will not receive a reward even if they met the reduction target during the event. However, as long as the DR event is still active (i.e., before its scheduled end time), a user can decide to opt-in by notifying the communicator. In this situation, users eligibility for the financial reward is evaluated only for energy data transmitted after their reply and until the end of the event.

After collecting user replies, the communicator forwards them to the aggregator, which determines the subset of EC members who agreed to modify their energy usage patterns in exchange for a financial reward. Consequently, the aggregator updates the DR event record in its local database to set up the energy data certification procedure for all participating users. Smart meters send timestamped energy measurements to the communicator, which validates and forwards them to the aggregator. The aggregator applies a filtering operation to collected energy data to extract measurements of participating users, which are inserted in the DR event record. The process of collecting energy measurements and storing them in the local database continues until the end of the DR event. Therefore, the aggregator acts as a reliable short-term storage system, ensuring that all DR event data are securely kept in its local database until the event concludes.

The aggregator leverages blockchain technology to provide a transparent and immutable record of hashes for DR-related energy data. This blockchain integration enables the certification of data integrity and therefore invalidates tampering with original data. When the number of data points reaches a predetermined threshold, the aggregator assembles them in a data batch and forwards them to the blockchain module. This module is responsible for computing the corresponding hash and ensuring that the hash is correctly stored on the blockchain for immutable storage by invoking a smart contract. Once finalised, the receipt of the writing operation is associated with the original data, which are eventually committed to a remote cloud server for long-term storage. Figure 4.5 illustrates the sequence diagram for the entire DR event workflow.

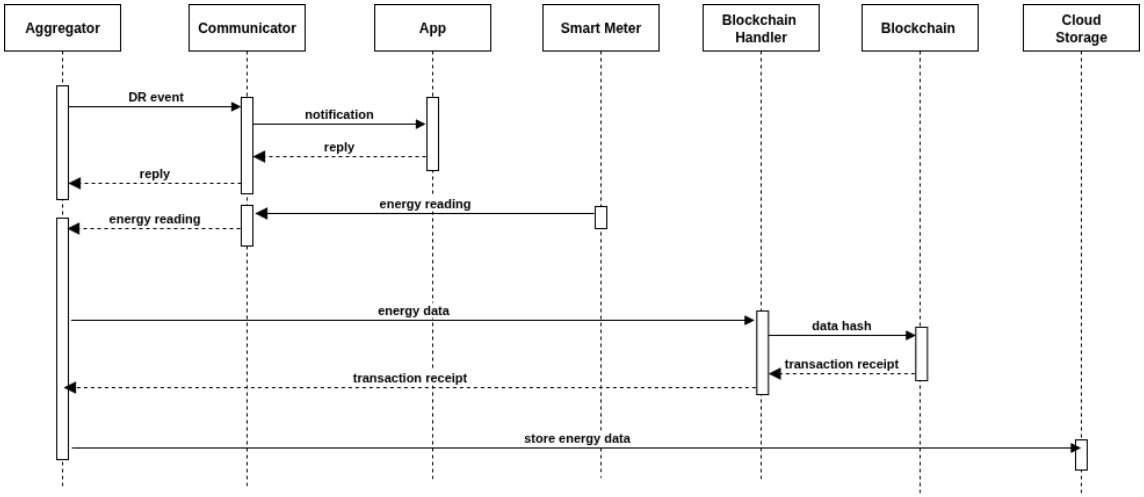


Figure 4.5: Sequence diagram of the demand response event workflow.

The hashes serve as proof of integrity for the corresponding data or, if a mismatch occurs between the hash of data in the storage and the corresponding hash retrieved from the blockchain, they reveal that the storage data were modified or are incorrect. Consequently, malicious data tampering activities on the stored data become futile, as they will be revealed by comparing the hashes. If a technical fault prevents the

finalisation of the transaction, the aggregator requests the blockchain module to retry the hash writing operation. In these situations, the energy data remain in the local database and are not sent to the remote cloud storage until the receipt is generated. The approach ensures that each data batch is always associated with an immutable record, even in the presence of temporary faults. This strategy of integrating traditional cloud storage with the properties of blockchain achieves an optimal balance between system scalability and security. The immutability property of blockchain provides a reliable proof of data integrity, promoting trust between all entities in the energy system. Moreover, by storing only the hash in the distributed public ledger, the privacy of user data is ensured. To provide strong security, the SHA-256 algorithm is used to compute hashes, which represent a unique digital fingerprint for the corresponding data. By storing the hashes in the blockchain, a public, transparent and immutable record is established, which can be utilised by stakeholders and authorities to verify that the original data have not been altered.

In our proposed approach, each traceable data entry consists of the following information: the ID of the DR event, the user ID, the energy measurement and the associated timestamp. Since storing a hash for each individual data point could result in high transaction costs for using the blockchain, data are aggregated in batches based on predetermined thresholds. Subsequently, the hash function is applied to the data batch, and the output is sent to an EVM-compatible blockchain for immutable storage. Hashes can later be retrieved from the distributed ledger to confirm the integrity of the corresponding energy data maintained in the remote cloud storage.

The `DataHashStorage` smart contract defines and exposes the two key functions for writing and reading hashes on the blockchain: `storeDataHash(bytes32`

`_dataHash)` and `getBlockTimestampFromHash(bytes32 _dataHash)`. The blockchain module executes the `storeDataHash` function, providing the hash of a data batch as the input argument. This transaction results in saving the hash in a mapping defined by the contract, which associates each hash with the timestamp of the block containing the transaction. The block timestamp represents the time interval in seconds between the Unix epoch and the minting of the block. After the transaction is finalised, the corresponding receipt is returned to the aggregator, concluding the data batch processing. Periodically, the aggregator sends all data for which there is a confirmed hash registration to the remote cloud server for permanent storage.

The `getBlockTimestampFromHash` function is used to obtain proof of integrity for a specific data batch. The data hash is provided as input for the function, which accesses the smart contract's internal mapping to recover the corresponding block timestamp. The existence of such a timestamp proves that the hash is recorded on the blockchain, thus validating the original data batch. Conversely, if no match is found in the internal mapping for the provided hash, the function returns the default value of zero. This proves that the provided input hash was not stored on the blockchain, and therefore the corresponding data batch was either incorrect or altered.

Disputes regarding energy data and corresponding financial remuneration may occur when an EC member claims different values for their energy usage during the DR event compared to the ones that are stored in the system. When such situations arise, data hashes retrieved from the blockchain enable to verify or disprove the user's claim. This verification process unfolds as follows: firstly, the complete energy data batch is recovered from the cloud storage and the corresponding hash is determined. Subsequently, a second hash is computed on data that includes the

disputed energy values and timestamps. Finally, the `getBlockTimestampFromHash` function is invoked to confirm the presence of these hashes on the blockchain storage. A successful hash retrieval proves the integrity and validity of the corresponding data, whereas failing to confirm the existence of the hash demonstrates its incorrectness, thereby concluding the dispute.

Notably, our proposed approach uses a public blockchain, enabling independent data validation by authorised entities. When facing legal or regulatory proceedings, it is critical to ensure external stakeholders that recorded data remain unaltered and are not compromised by unauthorised modifications. The immutability property of the blockchain guarantees that successfully stored hashes cannot be tampered with, providing a secure and verifiable audit trail. The use of public blockchain, as opposed to an Ethereum sidechain or a private blockchain or network controlled by the aggregator, provides a trustless and decentralised record which is resistant to centralised tampering.

This dual-layer storage strategy represents a scalable and cost-efficient solution, providing several advantages. Storing on blockchain only hashes derived from one-way cryptographic functions ensures that private information is not exposed on the public distributed ledger, thereby enforcing compliance with privacy regulations. Moreover, the complete set of data is securely maintained in the off-chain cloud storage, which can only be accessed by authorised entities. Furthermore, the data validation process is streamlined, requiring only the retrieval of the corresponding immutable hash from the blockchain and its comparison with the hash derived from the data from the cloud. An exact hash match proves the integrity of the original data, confirming that they were not altered by malicious entities who obtained an unauthorised access to the remote cloud storage. Finally, a blockchain transaction involving only a 32-byte

hash leads to a significant cost decrease compared to storing a large amount of energy data records on-chain.

Smart meters typically transmit energy measurements at regular time intervals (e.g., every 30 minutes). To achieve an optimal balance between data granularity and blockchain costs, the aggregator uses a set of predetermined thresholds to determine the frequency of hash writing transactions. When the total amount of collected measurements reaches a threshold, the aggregator assembles them into a data batch and requests the blockchain module to store the hash and return the receipt. The thresholds are defined to maintain a consistent size for the data batch, regardless of the number of participants in the DR event, which can differ from event to event. The next section presents the experimental assessment of our proposed approach, evaluated using a real-world energy dataset.

4.2.2 Experiments and Results

To evaluate the performance of our proposed architecture, we conducted experiments using an open-source DR dataset presented in [91] and publicly available on the Smart Grid Smart City (SGSC) website [116]. The SGSC project represents one of the largest real-world commercial-scale evaluations of Smart Grid technologies, and the publicly released data include information about customer load profiles, electrical network properties, Demand Response modelling and EV trials. [91].

To use only data that are relevant to our analysis of DR programmes, we selected the following datasets: (i) electricity use interval reading dataset, containing half-hourly interval measurements of consumption and generation sent by smart meters of users taking part in the SGSC programme; (ii) peak events dataset, detailing the time and duration of several demand response events issued during the SGSC trial; (iii)

peak event response dataset, including the actual energy usage and the corresponding baseline load for users who agreed to the consumption reduction request.

Table 4.3 shows sample entries from the peak events dataset. Each event is uniquely identified by its corresponding key (i.e., an ID), and it is associated with two timestamps, indicating the event start and end time, respectively. The SGSC trial defined two categories of DR events: the tariff-based dynamic peak pricing (DPP) and the incentive-based dynamic peak rebate (DPR), as indicated in the "Event Type" column.

Event ID	**Event Type**	**Start Timestamp**	**End Timestamp**
1000241	DPR	25/01/2013 13:00	25/01/2013 17:00
1000242	DPR	08/02/2013 13:30	08/02/2013 17:30
1000243	DPP	26/02/2013 13:00	26/02/2013 15:00

Table 4.3: Sample entries from the Peak Events dataset.

Table 4.4 shows sample entries from the peak response dataset, where customers' identities were anonymised by using a numerical ID. Similarly, the DR event customers participated in is indicated by the event ID. The remaining columns show the measured consumption value during the event (in kWh) and the reference baseline, defined as the customer's energy use before the event (in kWh). Finally, the "rebate amount" column displays the financial incentive, in Australian dollars, which was awarded to customers who actually reduced their consumption.

Customer ID	**Event ID**	**Actual kWh**	**Baseline kWh**	**Rebate (AUD)**
10233684	1000250	1.34	5.94	20.70
10233698	1000248	1.86	5.33	15.60
10233698	1000266	2.10	7.90	26.10

Table 4.4: Sample entries from the Peak Response dataset.

Initially, we applied pre-processing operations to clean and evaluate data quality, detailed as follows. Firstly, since in our proposed system the aggregator offers financial incentives to users who reduce consumption, we extracted only data related to DPR events. In fact, DPR events in the dataset are associated with a rebate amount, whereas DPP events adopt the strategy of dynamically changing the electricity price, therefore providing an indirect financial advantage for reducing energy load. The analysis of the peak events dataset revealed that the SGSC project defined a total of 18 DPR events. However, the peak response dataset contains data for only 15 events, and among these, complete data are available for only 13 events.

Subsequently, our analysis evaluated the energy and rebate values. To determine the amount of demand reduction during the DR events, we subtracted the consumption values from the baseline values. The results showed that three DPR events exhibit a ratio of more than 85% of participants who did not reduce their demand. This unusually high ratio suggests that either the events were not properly notified to customers or there was an error in data collection and reporting. Consequently, we did not consider these events in our analysis, resulting in a total of 10 DPR events. Following this, we identified and removed all entries in which the rebate amount is positive even though the corresponding user did not actually reduce consumption. This occurrence also suggests an error in recording data, as the financial incentive is

only given when the customer complies with the reduction request.

A final filtering operation was applied to remove outliers. Firstly, all entries with a consumption or baseline value lower than 0.01 kWh were removed, as such low values likely correspond to an erroneous baseline estimation, faulty measurements, or a vacant household. Subsequently, to identify outliers featuring high energy values, we observed that only about 3% of entries correspond to a consumption value of more than 10 kWh during the event. Given that the average daily consumption for a household in Australia has been estimated between 20–25 kWh [24, 107], and the duration of DPR events is 3–4 hours, it can be assumed that consumption or baseline values higher than 10 kWh are either incorrect or do not correspond to residential users. For this reason, we identified and removed them from the dataset.

After these pre-processing operations, we obtained a complete and clean set of data for 10 DPR events, involving $1,206$ unique customers and with a total of $7,339$ entries. Table 4.5 shows the statistical indices for baseline, consumption, reduction and rebate values for the cleaned peak response dataset.

	Baseline (kWh)	**Consumption (kWh)**	**Reduction (kWh)**	**Rebate (AUD)**
mean	4.32	2.02	2.30	10.69
std	2.55	1.73	2.24	9.52
min	0.01	0.01	−7.66	0.00
25%	2.14	0.79	0.59	2.70
50%	3.96	1.48	1.82	8.20
75%	6.30	2.73	3.74	16.80
max	9.99	9.93	9.51	42.80

Table 4.5: Summary statistics for the cleaned Peak Response dataset.

In the next phase of our analysis we determined the total number of participants

for each DPR event and, based on the results, we classified the size of the set of customers into three categories: (i) small, for users up to 200 members; (ii) medium, for users between 200 and 900 members; and (iii) large, for users over 900 members. Figure 4.6 shows the bar plot of the number of participants for each event and the corresponding size.

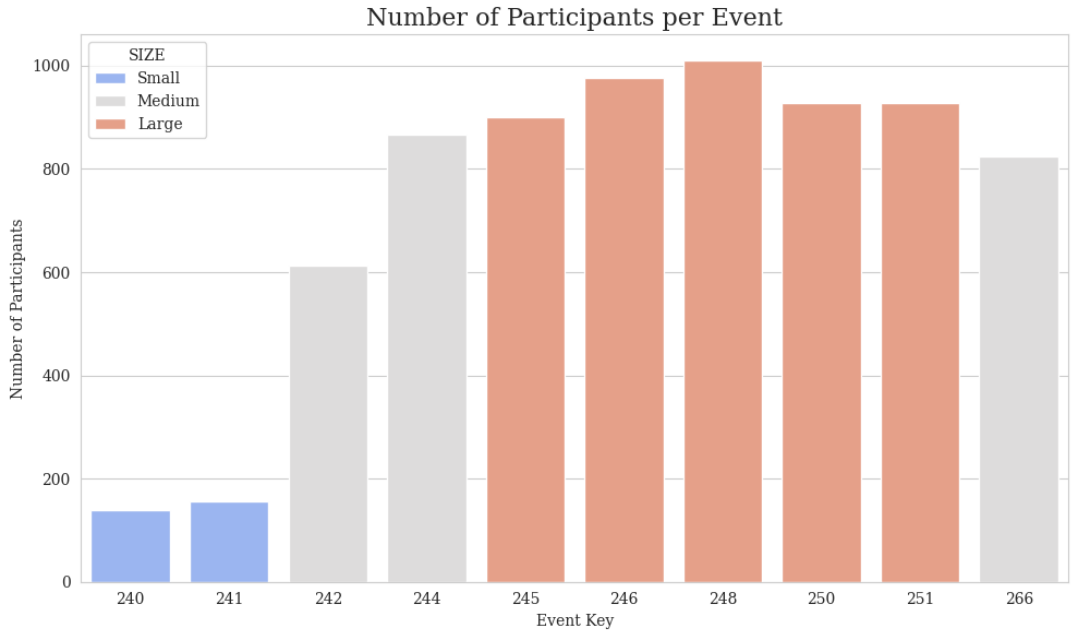


Figure 4.6: Number of participants for each DPR event, and corresponding community size.

Subsequently, our analysis focused on temporal information. The ten DPR events were issued over the span of just over one year, from January 2013 to February 2014. Moreover, their total duration was either three or four hours, and they were predominantly called at midday. Finally, we observed that all events took place on weekdays (Wednesday, Thursday and Friday), and they were more frequently called in summer and autumn. This temporal data are visualised in Figure 4.7, which shows the time spans for each DPR event, and Figure 4.8 which displays the distribution of DPR events by weekday, season and duration.

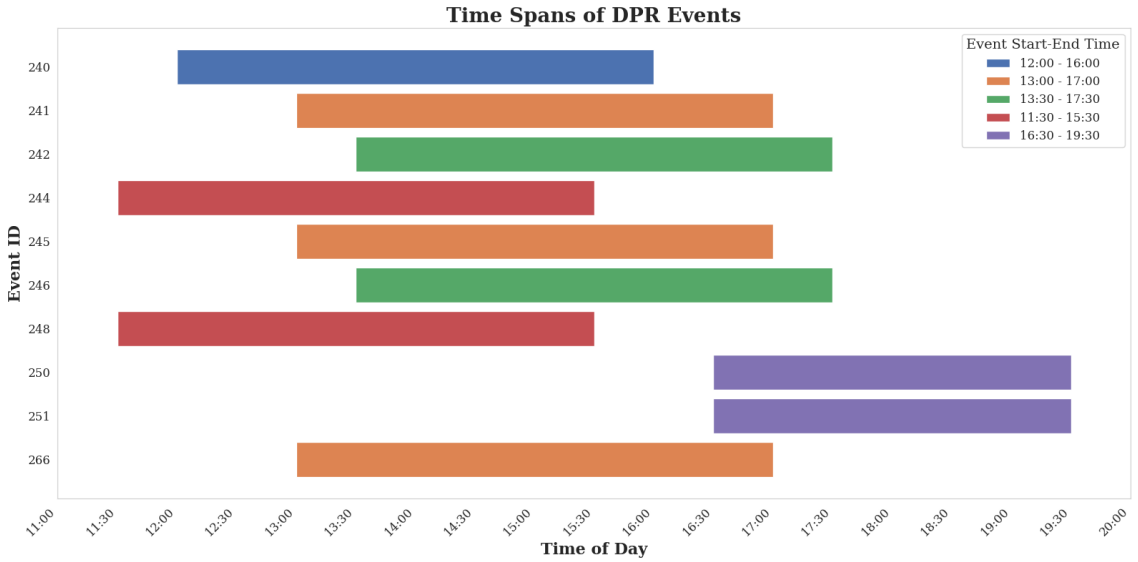


Figure 4.7: Time spans for the DPR events: bars having the same colour correspond to events activated during the same interval of time. The majority of DPR events were issued between 13:00 and 17:30.

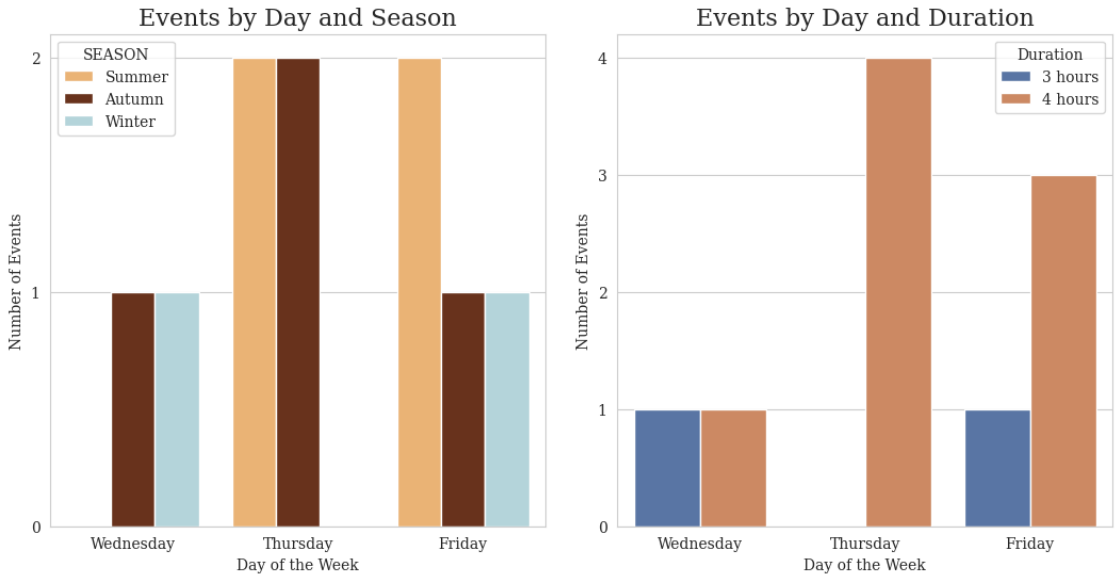


Figure 4.8: Count of DPR events by season and day of the week (left), and by duration (right).

With reference to the data in the cleaned dataset, timestamped information that our proposed system tracks and for which it certifies integrity includes: start and

end timestamp of the DPR event, and for each community member: (i) timestamp of when they sent their reply to participate in the programme, and (ii) their timestamped energy measurements for the entire duration of the event. In the SGSC trial, smart meters transmitted timestamped energy readings at 30-minute intervals. Consequently, based on the number of participants and the event duration, we determined the total amount of timestamped data to certify for each DPR event. This information is shown in Table 4.6, where the event ID was shortened to the last three digits for conciseness. Additionally, the percentage of users who, according to the dataset, actually reduced their load demand is also reported. Figure 4.9 visualises the key information from the Table: for each DPR event, the total amount of timestamped data to certify and the ratio of participants who actually reduced and did not reduce load demand, respectively.

Event ID	**Date**	**Duration**	**Participants**	**Reducing**	**Timestamps**
240	17/01/2013	4 hours	138	74%	1244
241	25/01/2013	4 hours	155	87%	1397
242	08/02/2013	4 hours	612	91%	5510
244	07/03/2013	4 hours	866	94%	7796
245	13/03/2013	4 hours	901	95%	8111
246	22/03/2013	4 hours	976	89%	8786
248	28/03/2013	4 hours	1010	86%	9092
250	24/07/2013	3 hours	928	93%	6498
251	02/08/2013	3 hours	929	95%	6505
266	13/02/2014	4 hours	824	94%	7418

Table 4.6: Summary of DRP events and key metrics.

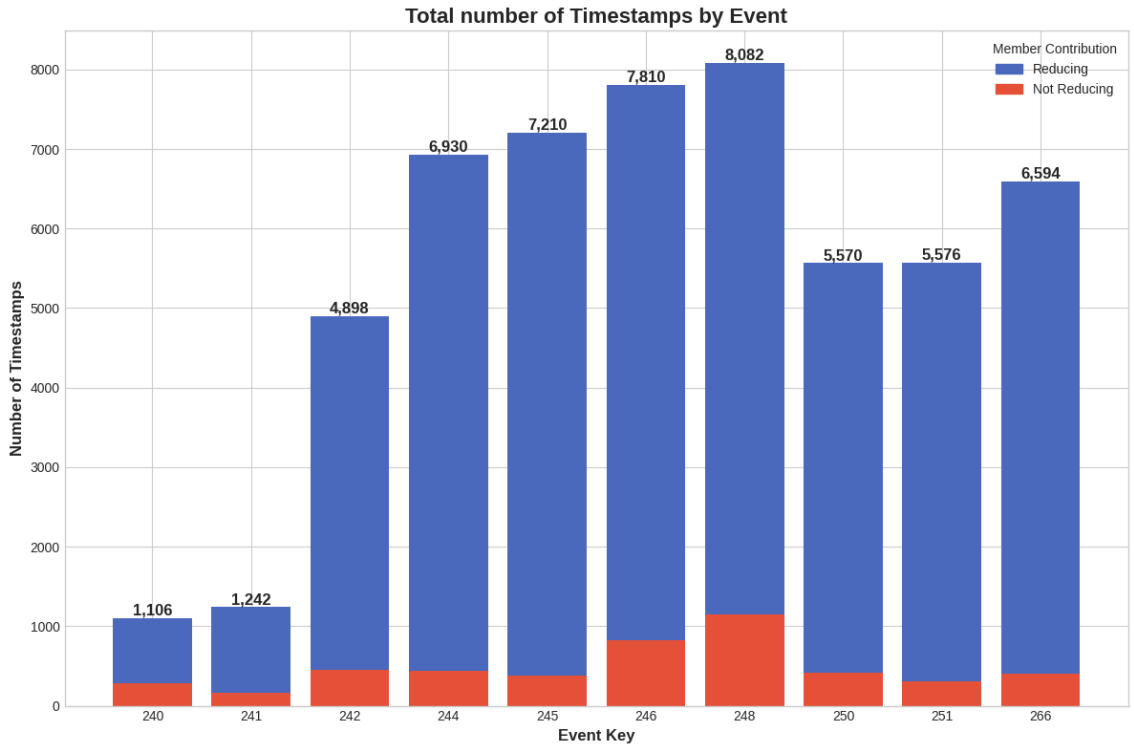


Figure 4.9: Total number of timestamps per demand response event.

In our simulation, the aggregator executed on a central server, and activated the DPR events by sending a request to the communicator. The communicator operated on a cloud node, for resilient and reliable interaction with the community members, and transmitted users' replies back to the aggregator. Energy data readings were received and stored in the local database by the aggregator. Subsequently, based on predetermined strategies, the aggregator created energy data batches and invoked the blockchain module that was responsible for generating the data hash, storing the hash on the blockchain and returning the transaction receipt to the aggregator. Finally, the aggregator submitted the energy data, the corresponding hash value and the receipt to the cloud database.

In our experiments, the data batches served as the input for the SHA-256 hash

function, and the resulting hash value was transmitted to the `DataHashStorage` smart contract for immutable storage on the blockchain. The contract was deployed on the Ethereum Sepolia testnet, and it is publicly available at the following address: $0x2a5A789beBeE29c8fc824d76a82D3aE787b6d1F5$. The source code of the `DataHashStorage` contract and the full history of transactions can be reviewed on the corresponding Sepolia Etherscan page[1]. The contract implements the `storeDataHash(bytes32 _dataHash)` function, which takes as input the SHA-256 value and records it in an internal mapping. This mapping provides a persistent storage of key–value pairs, where each hash is associated with the timestamp of the block containing the hash writing transaction.

To assess blockchain fees and block confirmation times, we conducted a total of 1,093 hash writing transactions across the Sepolia network. We requested the transactions over different weekdays, with a success rate of 100%, and evaluated the total costs associated with blockchain interactions by extracting the transaction fee from the receipt, in Gwei (1 Gwei $= 1 \times 10^{-9}$ ETH). To derive AUD fees from the corresponding Gwei amounts we used CoinGecko[2], a cryptocurrency data aggregator exposing a publicly available API which can be used for obtaining real-time exchange rates, including ETH–AUD. The results of our experiments are reported in Figure 4.10, which displays the transaction fee distribution in Gwei and AUD, and Table 4.7, which shows the key statistical indices for block confirmation times and fees.

[1] https://sepolia.etherscan.io/address/0x2a5a789bebee29c8fc824d76a82d3ae787b6d1f5

[2] https://www.coingecko.com

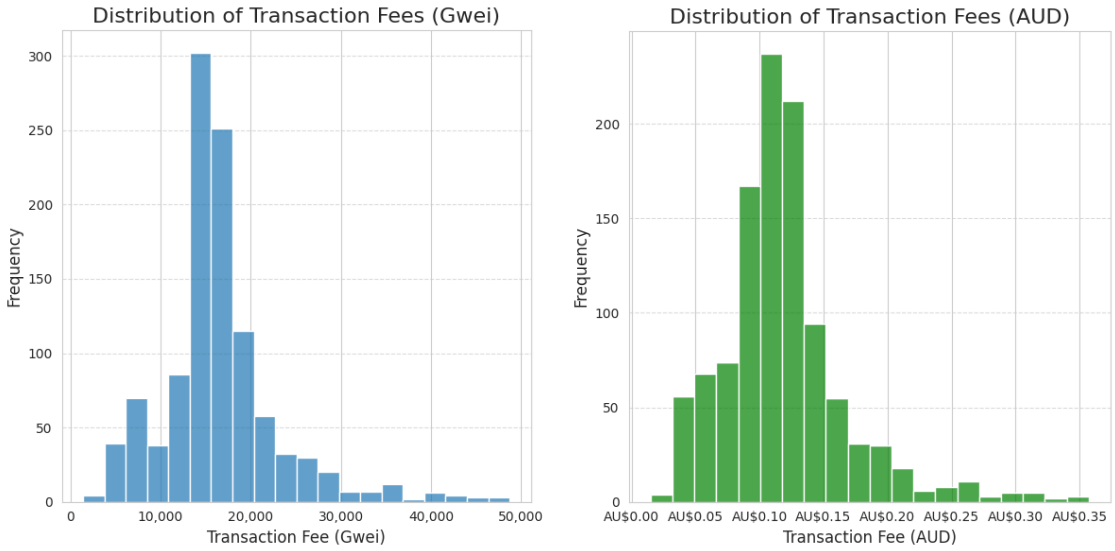


Figure 4.10: Distribution of blockchain transaction fees in Gwei (left plot) and Australian dollars (right plot).

	Transaction		
Statistics	**Confirmation Time (s)**	**Fee (Gwei)**	**Fee (AUD)**
mean	13.38	16 305	0.120
std	5.77	6591	0.049
min	2.34	1435	0.015
25%	11.50	13 335	0.097
50%	13.19	15 662	0.116
75%	13.31	18 254	0.134
max	108.67	48 717	0.357

Table 4.7: Descriptive statistics for blockchain transaction data.

To optimise hash writing requests, balancing cost and performance, the aggregator used thresholds to determine the total number of entries to include in a single energy data batch. In our simulation, we considered four different thresholds: 50, 75, 125 and 175. The corresponding total transaction fees for each DPR event were derived, and are displayed in Table 4.8. For reference, the total and mean values for rebate amounts are also provided. Since the SGSC trial took place from 2013 to 2014,

to account for inflation the reported rebate values were updated using the official Consumer price index (CPI) rates provided by Australian government portals [16, 17].

ID	Rebate Amount (AUD)		Total Cost by Threshold (AUD)			
	Total	**Mean**	**50**	**75**	**125**	**175**
240	825	5.98	2.88	1.92	1.08	0.84
241	2159	13.92	3.24	2.16	1.32	0.84
242	8491	13.88	13.20	8.76	5.28	3.72
244	12584	14.53	18.60	12.36	7.44	5.28
245	14315	15.89	19.44	12.96	7.68	5.52
246	13561	13.89	21.00	14.04	8.40	6.00
248	11285	11.17	21.72	14.52	8.64	6.12
250	14945	16.10	15.48	10.32	6.12	4.44
251	16682	17.96	15.60	10.32	6.24	4.44
266	12840	15.58	17.76	11.76	7.08	5.04

Table 4.8: Summary of blockchain total costs by threshold. Rebate values, adjusted for inflation, are provided for reference.

4.2.3 Discussion

A total of 1,093 hash writing transactions were requested and finalised on the Ethereum Sepolia blockchain. To reproduce the same temporal conditions of the SGSC trial, transactions were executed during weekdays (see Figure 4.8) and over the same time interval, spanning from 11:30 to 19:30 in the Australian timezone (see Figure 4.7). Although the SGSC trial was in effect in the years 2013–2014, it is still

a realistic assumption that DR events are requested in similar time intervals, when energy consumption is typically higher. Moreover, during midday the energy production from solar photovoltaic panels reaches its peak, which is significant especially in summer, when the majority of DPR events were requested (see Figure 4.8). A high availability of energy output from renewable energy sources ensures more options for community members willing to participate in the DR programme, especially prosumers equipped with ESS and EV charging stations (in the V2G scenario).

Table 4.5 reveals that, on average, customers achieved a reduction of 2.30 kWh, with 75% of all entries reporting a value as high as 3.74 kWh. Since the average baseline value, i.e. the estimation of user consumption without DPR events, is 4.32 kWh, the results in the table confirm that the DPR request and associated rewards had a significant impact in modifying energy consumption patterns, and ensuring that grid congestion is mitigated. While the presence of negative reduction values signifies that some users increased their demand load, our analysis in the cleaned datasets demonstrated that the vast majority of customers who agreed to take part in the DR programme actually reduced consumption during the DPR event (see Table 4.6). Specifically, the average ratio of complying users was 89.8%, with percentage exceeding 90% for 6 out of 10 events. This data represent an important confirmation that DR programmes are successful, and that financial rewards make an important contribution to ensuring that users respond to the request of modifying their energy usage patterns. It is important to note that our approach continues to track and certify energy data for all users who agreed to the DR request, even when some of them do not actually reduce consumption during the DR event. This guarantees the system can prove that non-complying users are not entitled to a reward, should a dispute arise.

The statistical analysis of block confirmation times and transaction fees for the 1,093 hash writing operations provides a solid basis for evaluating performance and economic feasibility in a real-world scenario. The experimental results show that the mean time to finalise a block is 13.38 seconds, with 75% of all transactions taking at most 13.31 seconds (see Table 4.7). While we observed the presence of some outliers, with one hash writing operation taking approximately 109 seconds to complete, the corresponding data remains securely stored in the aggregator's local database. Moreover, since in the SGSC dataset energy readings are transmitted by smart meters every 30 minutes, a delay of even a few minutes does not have a significant impact on system performance. Additionally, whenever the transactions are not confirmed because of, for example, network faults, the aggregator resubmits the hash writing request to the blockchain module for the same data batch. Finally, when the transaction is completed and the corresponding receipt is returned, the data batch becomes associated with an immutable record that certifies energy data integrity. This property is essential for verifiability and for complying with legal requirements if disputes arise. In fact, any malicious or unauthorised modification to data stored in the database can be detected, because it would result in the hash of modified data not matching with the corresponding hash retrieved from the blockchain. Conversely, identical hashes confirm data integrity, enabling dispute resolution.

Costs for executing transactions on the Ethereum blockchain are paid in ETH. A transaction fee distribution originates from the same hash writing operation because the fee for using the blockchain depends on the state of the network, resulting in higher prices during periods of congestion. The difference between the Gwei and corresponding AUD distributions (see Figure 4.10) is due to the fact that ETH/AUD exchange rates are continuously subject to fluctuations. In our

simulations, we observed that the average cost for finalising a blockchain writing operation is AUD 0.12, and 75% of transactions' costs were less than AUD 0.135 (see Table 4.7). Moreover, the lowest and highest recorded costs were AUD 0.357 and AUD 0.015, respectively. We used the mean fee value to estimate the total cost needed to provide integrity certification for energy data in every DPR event, and the results are shown in Table 4.8, divided by batch size threshold. The table analysis reveals that the total cost for providing proof of integrity for the entire set of energy data during a DPR event is lower than the average rebate amount, adjusted for inflation, in the vast majority of the scenarios. Specifically, the total costs remains consistently lower with the 125 and 175 thresholds, and exceeds the average in only two and five occurrences for the 75 and 50 thresholds, respectively. Overall, total cost is less than the average rebate in 80% of the cases, and even when the cost is higher than the average, it represents only a small fraction of the total rebate amount awarded to participants (see Table 4.8). Certification cost-to-total rebate ratios range from 0.03% to 0.35%, and remains consistently lower than 0.10% for the 125 and 175 thresholds.

Higher thresholds correspond to larger data batches and less frequent hash writing requests, resulting in lower overall costs. The threshold choice depends on the identification of an optimal trade-off between cost savings and data overhead when retrieving information from the cloud during dispute resolution, and ultimately it is determined by system requirements. To address this challenge, our approach provide a high level of flexibility: depending on the number of participants (that can significantly vary from event to event, see Figure 4.6) and blockchain costs, the thresholds can be dynamically changed to adapt to different conditions and achieve an optimal balance between cost and performance.

While the costs associated with hash writing operations in our experiments remain economically feasible, strategies can be employed to control price fluctuations and further optimise costs. For example, blockchain costs can be monitored via a polling service, and transactions are requested only when the block fee falls below a predetermined threshold. In the dataset we used, measurements were sent every 30 minutes, and therefore this mechanism can be effectively employed to handle short-term cost fluctuations. In energy systems where measurements are transmitted using significantly shorter time intervals, e.g. every five minutes, timeouts can be set to ensure that the hash writing transactions are finalised before receiving the new set of energy readings.

Overall, our approach of storing on-chain only 32-byte hashes rather than the complete record of data achieves a significant reduction in blockchain utilisation costs while preserving data confidentiality. The empirical results demonstrated that transaction fees remain comparatively low with respect to DPR events reward amounts, despite cost fluctuations due to blockchain congestion. Additionally, hashes are stored on a public blockchain, enabling independent data integrity verification by authorised entities when disputes arise. This mechanism promotes trust among all EC members and represents a key advantage in contrast to third-party and commercial certification systems, which can define arbitrary costs and terms of use for their services and therefore cannot be fully trusted.

In our experiments we used publicly available datasets officially released by the Australian government, which correspond to a real-world implementation of an incentive-based DR programme, involving thousands of households over the course of more than a year. Consequently, our cost analysis can be effectively used as a basis to obtain realistic cost estimations for certifying data integrity in DR programmes.

Additionally, these estimations enable cost projection for different DR configurations, with varying number of participants, duration and frequency of DR events. This informed analysis represents a key aspect for optimal investment planning and the definition of DR programme parameters. Results analysis shows that our proposed approach achieves feasible operational costs even in the presence of high variability for blockchain transaction fees. The strategy of aggregating energy data in batches and committing on the blockchain only the corresponding hash enables proof of integrity for the data at a minimal cost. Finally, the definition of thresholds for batch sizes enables the optimisation of writing requests, that avoids blockchain congestion and provides a high degree of flexibility to system operators.

4.2.4 Summary

This study presented a framework that leverages blockchain properties to provide proof of integrity for data in energy communities participating in incentive-based demand response programmes. The proposed approach achieves data certification using the decentralised blockchain network, tackling a significant limitation of centralised approaches, which require trust in an external authority. The proposed system requests on-chain hash writing transactions of energy data in the context of demand response events. The hashes recorded on the blockchain represent an immutable record that can be used to certify the validity of energy data and detect malicious data alteration in the database.

Experimental assessment of the proposed system was conducted on a real-world energy dataset, publicly released by government agencies that designed and implemented demand response trials spanning over one year. Results show that our approach achieves efficient and cost-effective data management. The definition of

configurable thresholds enables a high degree of flexibility and allows for tackling fluctuations in the number of participants and blockchain costs. It also enables an efficient use of system resources.

Crucially, the insights derived from our study pave the way for broader applicability of blockchain-based data certification. In the context of distributed data systems, having a reliable and cost-effective strategy for proving data integrity and obtaining an independently verifiable audit trail is paramount, and it promotes trust among stakeholders without relying on a central authority. To demonstrate the generalisability of our framework in other domains, we applied the blockchain-based approach presented in this study to certify timestamps of critical data in healthcare systems, as discussed in the next chapter.

Chapter 5

Certifying Data in a Distributed System

The results presented in the previous chapter demonstrated how the blockchain technology can be leveraged to provide proof of data integrity and address disputes over financial rewards, detect tampering attempts by malicious entities, and provide a verifiable audit trail while ensuring user privacy. While we designed and adopted this strategy in the context of the framework presented by this thesis, the need for reliable and traceable data recording is not limited to the energy domain.

Distributed systems where critical data are produced, transmitted and stored can greatly benefit from methodologies that can ensure integrity, transparency and verifiability. To assess the performance and flexibility of our blockchain-based data certification strategy, we considered the case study of emergency response systems in the healthcare domain. These systems are typically data-intensive and need to comply with strict legal requirements to verify the timeline of events. For this reason, we adapted our framework to handle timestamp certification in emergency response and conducted a comprehensive cost and performance evaluation on a real-world dataset, considering several data optimisation strategies. The results of this research

were presented in [74].

The remainder of this chapter is organised as follows. Section 5.1 introduces the case study. Section 5.2 presents the proposed system architecture. To test and validate our approach, we conducted experiments on a real-world emergency response dataset. These experiments and the three hash storage strategies we designed are described in Section 5.3, Section 5.4, and Section 5.5. The results are discussed in Section 5.6. Finally, conclusions are drawn in Section 5.7.

5.1 Certifying Timestamps in a Distributed System: a Case Study on Emergency Response Systems

In healthcare emergency response systems, all activities performed must be promptly recorded for potential further investigations. In such cases, external authorities can challenge healthcare service providers regarding the timing of critical events, such as the arrival time of an ambulance at the designated location, the starting time of a required analysis, and other critical events. Generally, the activities follow a well-established practice and, thus, they are not at the discretion of the responding personnel. However, the timing of such activities can be subject to scrutiny. Disagreements about the proper timing of some events can escalate into disputes, and the healthcare service provider must demonstrate that the appropriate sequence of actions was implemented at the correct time.

Generally, a healthcare provider records events and their timestamps, as they occur, on a dedicated server. However, the integrity of such timestamps may be questioned when the authority responsible for recording the data is one of the

disputing parties. For example, the challenging authority could argue that the data were tampered with in certain circumstances, making it appear that the ambulance arrived on time. In such cases, the defendant would not be able to demonstrate otherwise. Although a centralised database would be under the control of some authority, when no authority can be considered trustworthy, blockchain technology is well-suited to ensure data integrity and access control, thanks to its mechanisms based on cryptographic signatures and hashes that confirm that the recorded data remain unmodifiable [22].

Several blockchain applications have been proposed for healthcare systems, including the management of electronic health records (EHR) and patient data [60, 80, 95], emergency medical services [57, 120], patient-centric data sharing [143, 151], integration with IoT devices [34], clinical research [14, 93], healthcare security and audit processes [26, 65, 113, 136], and medical data sharing [9, 136, 137]. In all these scenarios, ensuring patient confidentiality and secure data exchanges is paramount [36]. Unlike methods that depend on trusted third parties or centralised timestamping servers, our framework leverages a blockchain model to eliminate such dependencies. Our solution uses blockchain-stored hashes that either certify data integrity or make any alteration evident and easily detectable. Previous blockchain-based timestamping approaches have been limited by issues such as transaction delays, high costs, and privacy. Our dual-layered design, which combines real-time cloud storage with blockchain recording, tackles these issues by ensuring both efficiency and robust security even under high data loads. Importantly, our approach uniquely ensures non-repudiation of recorded events, a critical capability in healthcare incident response that, to the best of our knowledge, previous solutions have not addressed.

To demonstrate the reusability and versatility of the framework developed in

this doctoral thesis, we successfully applied the blockchain-based data integrity methodology, presented in Section 4.2, to the healthcare domain. Specifically, we designed and implemented an architecture to certify the integrity and authenticity of timestamped data for events in healthcare emergency response systems. The proposed system includes a health emergency aggregator (HEA), a software component that collects incident data, records such timestamped data on a cloud server, and hashes them using SHA-256, as they occur. Data certification is then ensured by storing hashes on an EVM-compatible blockchain via smart contracts. This dual-layered approach is a well-known off-chain data storage technique that provides scalability, security, and transparency, while reducing the costs related to the blockchain [38, 48, 139]. Later, the authenticity of recorded events can be verified by comparing the hashes computed on the data stored on a cloud server with those stored on the blockchain [42]. To assess the practicality of our solution, we performed a comprehensive evaluation of our prototype software system using real-world data from the incident dispatch dataset provided by the NYC OpenData initiative [21]. Our findings demonstrate the efficacy of our approach, making it a cost-effective strategy for aggregating the numerous data points and timestamps that would otherwise necessitate more recordings per day.

The main contribution of this work is the development of a decentralised framework that leverages blockchain technology to certify the integrity and authenticity of incident data. By employing a client-server and blockchain-based software architecture, our framework guarantees that all recorded events are certified, thus providing a proof of their recorded timing. This solution facilitates secure and transparent data handling and enhances trust among actors. Additionally, we performed a thorough cost assessment to demonstrate the framework's economic viability, highlighting

its cost-effectiveness in real-world scenarios. This combination of features provides emergency system authorities and patients with a trusted third-party mechanism to certify incident-related data efficiently and effectively.

5.2 System Architecture

Current limitations in healthcare emergency response systems highlight a critical need for solutions that can effectively certify the authenticity and integrity of events throughout the incident lifecycle. To address this shortcoming, we propose a novel architecture that integrates blockchain technology to provide a secure and transparent mechanism for event certification in emergency response systems. This architecture ensures that critical data related to each incident are certified, fostering trust and accountability. Figure 5.1 presents an overview of the proposed architecture, outlining the key actors involved and the technologies employed.

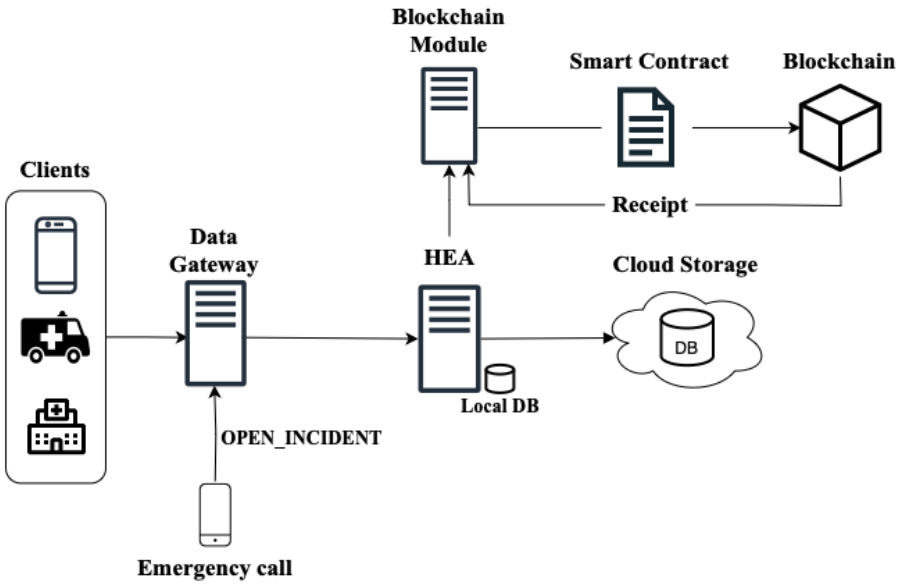


Figure 5.1: Overview of the proposed architecture, featuring the HEA, data gateway, blockchain integration for secure incident data management, and cloud storage for persistent data storage.

At the heart of the architecture is the health emergency aggregator (HEA), the main component responsible for managing incidents and triggering the activities necessary to safeguard data integrity. The lifecycle of an incident begins with the emergency call, which leads to the creation of a new incident record. This call goes to the data gateway, a component that acts as a facade and handles all the data from the clients and filters and forwards incident data to the HEA, which will store them in a local database. The HEA collects all the data related to an incident until it is resolved and can be archived in the cloud for long-term storage. The HEA functions as a reliable storage system, ensuring that all incident-related data are securely held without undergoing further processing or analysis. This capability provides an accurate historical record of emergency events.

To further enhance the security, a mechanism that can demonstrate the authenticity of incident data has been implemented. Critical events identified by the HEA are submitted to the blockchain module, where they are permanently recorded. This blockchain integration allows for verification of data integrity and renders any tampering activities futile, since it would be easily exposed in case of disputes. The benefit of this approach is twofold: data are protected from potential tampering, and a transparent audit trail is maintained for future reference. Timestamped events are periodically sent to the blockchain while they occur, whereas the analysis of data is performed when needed, well after the incident has been handled, e.g., when further analyses are started to gather statistics on the healthcare service, or to check whether proper activities were performed in a timely manner.

The workflow followed by the HEA is presented in Algorithms 1 and 2, which outline a systematic process of data handling and aggregation.

Algorithm 1 Healthcare Emergency Aggregator (HEA) workflow when receiving data.

```
1: ordinary_data ← []
2: critical_data ← []
3: while isListening do
4:     data_payload ← receiveFromDataGateway()
5:     switch (data_payload.type)
6:         case open:
7:             incident_id ← create_incident(data_payload)
8:             sendToDataGateway(incident_id)
9:             store(incident_id, data_payload)
10:        break
11:        case close:
12:            incident_id ← data_payload.incident_id
13:            incident_data ← getIncidentData(incident_id)
14:            agg_data ← aggregateData(incident_data, data_payload)
15:            ordinary_data.append(agg_data)
16:        break
17:        case ordinary:
18:            incident_id ← data_payload.incident_id
19:            store(indicent_id, data_payload)
20:        break
21:        case critical:
22:            incident_id ← data_payload.incident_id
23:            store(indicent_id, data_payload)
24:            agg_data ← aggregateData(incident_data, data_payload)
25:            critical_data.append(agg_data)
26:        break
27: end while
```

Algorithm 1 shows that payloads received from a data gateway are categorised

into one of four types: open, closed, ordinary, or critical. These categories then trigger specific operations to manage the incidents and their associated data. For open payloads, a new incident record is created with a unique identifier, and then stored. For closed payloads, data related to a specific incident are retrieved and aggregated; then, the results of their management are appended to a list of ordinary data. When processing ordinary and critical payloads, received data are stored; for critical cases, data are immediately aggregated and the results are added to a dedicated critical data list.

Algorithm 2 Healthcare Emergency Aggregator (HEA) workflow when timeouts occur.

```
1: while isUp do
2:     switch (timeout)
3:         case critical:
4:             agg_data ← aggregateData(critical_data)
5:             trans_receipt ← sendToBlockchain(agg_data)
6:             store(critical_data, trans_receipt)
7:             critical_data ← []
8:         break
9:         case ordinary:
10:            agg_data ← aggregateData(ordinary_data)
11:            trans_receipt ← sendToBlockchain(agg_data)
12:            store(ordinary_data, trans_receipt)
13:            ordinary_data ← []
14:        break
15: end while
```

Algorithm 2 shows a timeout mechanism for periodically aggregating and securing both ordinary and critical data. Hashes of aggregated data are sent to a blockchain to ensure integrity and traceability, and transaction receipts are stored for verification

purposes. This workflow ensures efficient data management, secure storage, and streamlined incident handling in dynamic and high-stakes healthcare environments.

The HEA serves as a central hub for managing all incidents and their associated data, ensuring seamless coordination and integration within the emergency response system. For each incident, the aggregator receives comprehensive data from the data gateway, which initiates the incident record. This record contains the time of occurrence, the situational context, and detailed information about the stakeholders assigned to the incident. Such stakeholders are ambulances, medical professionals (such as doctors and paramedics), healthcare facilities (such as hospitals), and other emergency response units.

The data gateway abstracts system complexity by providing a unified interface. Clients interact exclusively with the gateway, which consolidates incident data and forwards them to HEA for processing and storage. Centralising incident data through HEA facilitates efficient resource allocation and real-time decision making. This approach is critical for prompt emergency responses and maintaining a comprehensive repository for post-incident analysis and system optimisation. Emergency assistance requests typically begin when a patient contacts the designated emergency telephone number. Upon receiving a call, the operator collects relevant information about the patient and their situation. This data is used to create a standardised incident record in the system, which HEA manages. Each incident is assigned a unique identifier to ensure seamless tracking and management throughout its lifecycle.

Figure 5.2 presents a sequence diagram to illustrate the interactions among the proposed components throughout the lifecycle of an incident. By showing these interactions, we provide details of communication handling (requests and answers), data filtering (separating ordinary and critical events), data transformation

(aggregation and computing hashes) and permanent data storage (hashes on the blockchain and data on the cloud server).

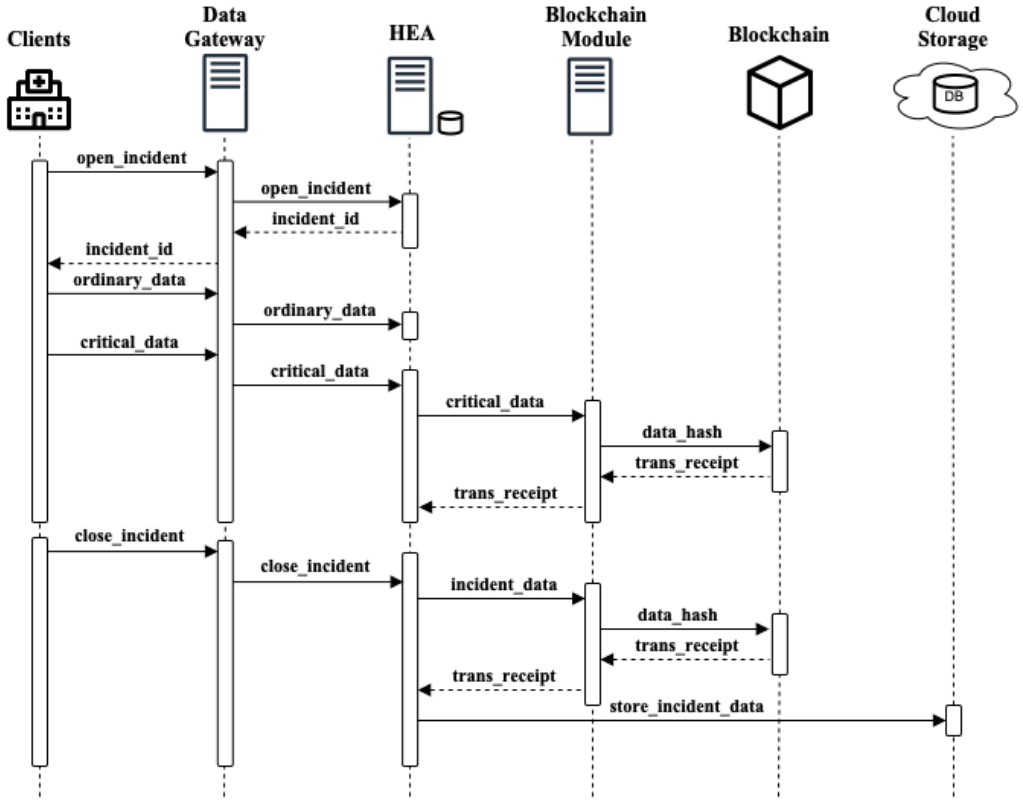


Figure 5.2: The lifecycle of an incident highlights interactions among components and the HEA's role in logging critical events to the blockchain and storing incident data in the cloud servers.

The lifecycle begins with the formal creation of the incident (`open_incident`), initiated by a starting event that triggers the sequence of actions required for resolution. HEA assigns a unique identifier (`incident_id`) to each incident and communicates this ID to clients, enabling consistent reference during subsequent data transmissions related to the incident. As the incident progresses, updates on status changes, resource allocation, and priority adjustments are recorded, along with communication traces arising from coordination within emergency response teams. The lifecycle ends with a final event (`close_incident`) marking the resolution or

closure of the incident. This process provides a comprehensive chronological record of the incident's progression, offering valuable insights for accountability and continuous improvement.

The data collected during the lifecycle are categorised into two priority levels: ordinary and critical. Ordinary data encompass ancillary information such as the geographical dispatch area, postal code, and descriptive situational details. As these pieces of data provide contextual understanding, they do not directly influence real-time decision making. Critical data, in contrast, consist of timestamped events that document pivotal moments in the incident's lifecycle, such as dispatch times, arrival times at the scene, and specific interventions performed. These data are essential for real-time coordination and post-incident analysis. To ensure operational efficiency, legal compliance, and process improvement, the integrity and accuracy of critical data must be maintained rigorously.

The HEA employs a multi-layered approach to incident data management that ensures both system responsiveness and data integrity. A local database serves as a temporary cache for incident data, enabling rapid access and updates during an incident's active phase. When an incident is closed, the HEA consolidates all related information, including timestamps, resource allocations, stakeholder actions, and blockchain transaction receipts, into a coherent record. These data are then transferred to remote cloud storage, which serves as the permanent repository for historical data analysis and reporting, offering scalability, redundancy, and accessibility.

To further enhance data security and trust, HEA uses blockchain technology to save hashes of incident data. Such hashes, being immutable, prove whether the corresponding data were the original data. Conversely, any hash mismatch proves

that the data were altered. After receiving critical data, the information is stored in the local database and then passed to the blockchain module for subsequent aggregation and submission. Likewise, when an incident is closed, all associated data are aggregated and their corresponding hashes are submitted to the blockchain. In the event that a blockchain transaction does not receive confirmation, the affected data batch and its hash remain securely stored in the local database, and the HEA subsequently reissues the hash-writing request to the blockchain module. This mechanism leverages the inherent resilience of blockchain technology, ensuring that, despite any transient failures, the hash writing is ultimately and reliably completed. Once a blockchain transaction is completed, a receipt is generated and attached to the incident record to facilitate later verification. This dual-layered storage strategy, combining cloud storage with blockchain solutions, balances scalability with security. It also reinforces the reliability and integrity of the emergency response system, fostering stakeholder confidence and ensuring compliance with legal and regulatory requirements.

Blockchain has emerged as an innovative technology for secure and decentralised data management, offering properties particularly relevant to sensitive domains such as the healthcare sector. At its core, blockchain is a distributed ledger system that records transactions in a series of cryptographically linked blocks. These blocks are designed to be immutable, ensuring that data, once added, cannot be modified or deleted, thus providing a high level of security and trust. Ethereum and other Ethereum Virtual Machine (EVM)-compatible blockchains have further advanced this technology by enabling programmable smart contracts. These contracts operate as self-executing code triggered by predefined conditions, eliminating the need for intermediaries while ensuring transparency.

In ambulance dispatch and patient data management, blockchain provides a secure and reliable framework to validate the critical events, while safeguarding sensitive information. The proposed system uses an Ethereum Virtual Machine (EVM)-compatible blockchain to certify the records of incident data stored in a cloud storage.

For enhanced security, incident data are hashed using the SHA-256 algorithm, generating a unique digital fingerprint that guarantees their integrity. This hash is then immutably recorded on the blockchain, establishing a transparent and unalterable record of all system transactions. This mechanism secures data and facilitates independent verification, as the blockchain serves as a reliable audit trail that can be referenced to confirm that the recorded data have not been tampered with.

The proof of the existence of an event generated by the client (mobile application) is represented by the concatenated hash of three components: a description of the event (e.g., a patient requesting an ambulance), its timestamp, and an optional document associated with the event (e.g., a picture of a wound, a patient's ID cards). The resulting hash is stored on the blockchain to guarantee the event's proof of existence while preserving its integrity and validity. Hash writing and retrieval operations are performed by the `DataHashStorage` smart contract deployed on an EVM-compatible blockchain. The contract provides two core functions: (i) `storeDataHash(bytes32 _dataHash)`, and (ii) `getBlockTimestampFromHash(bytes32 _dataHash)`.

The `storeDataHash` function takes as input the hash sent by the blockchain module and records it on the blockchain using an internal mapping. This mapping associates each hash with the timestamp of the block in which it was recorded. The block timestamp, represented as a `uint256` type in Solidity, is measured in seconds since the Unix epoch. The HEA retrieves the block timestamp from the transaction

receipt and associates it with the corresponding data batch in its local database.

The `getBlockTimestampFromHash` function verifies the presence of a specific hash in the blockchain storage. It takes the hash as input and uses it to query the internal mapping, and retrieves the corresponding block timestamp if the hash is found. This timestamp serves as proof of the hash's existence on the blockchain, confirming the associated data's registration. From the block timestamp, the existence of the block and the hash writing transaction can be verified. Conversely, if the function returns the default value of zero, it indicates that the provided hash is not present in the blockchain storage, indicating that it has not been recorded.

Disputes regarding data accuracy may arise when external parties (e.g., patients) report timestamps for critical events (e.g., the arrival time of an ambulance) that differ from those recorded by the system. In these cases, blockchain-stored data hashes are used to verify the integrity and accuracy of the timestamps. The process involves retrieving the complete incident record from remote cloud storage and computing two hashes: (i) the system-recorded data hash, and (ii) the hash of the data with the disputed timestamp substituted. The smart contract function `getBlockTimestampFromHash` is then called to verify the existence of these hashes on the distributed ledger. A matching hash confirms the data's integrity and the correct timestamp, whereas a mismatch indicates invalid data, thereby resolving the dispute.

It is important to note that although the HEA manages incident data, our design incorporating a public blockchain provides independent external verification. In situations such as legal disputes or regulatory audits, external stakeholders require assurance that recorded event timestamps and data integrity have not been subject to unilateral modification. By using a public blockchain rather than an Ethereum

sidechain or a private blockchain under HEA's control, the audit trail remains decentralised, trustless, and immune to centralised tampering. The immutability of the blockchain guarantees that once a hash is recorded, it remains unalterable, establishing a robust, time-stamped record.

Only storing the data's hash on-chain offers several advantages. The hash serves as an exact digital fingerprint of the original data, preserving its integrity and confidentiality. When data are retrieved from secure off-chain (cloud-based) storage, the system recomputes the hash and compares it with the corresponding blockchain entry, ensuring that the data remain unaltered. This process provides a cost-effective and scalable record of incident data. Moreover, because the hash is generated using a one-way cryptographic function, no sensitive information is revealed, thus ensuring compliance with privacy requirements. The blockchain's immutability further guarantees that once a hash is recorded, it remains unaltered, creating a reliable audit trail that verifies the data's existence at a specific time. Finally, only storing the hash, rather than the complete dataset, provides significant cost savings by avoiding the substantial expenses associated with on-chain storage of large data volumes.

To efficiently manage storage of critical and ordinary data, the system employs two main strategies: time-based and threshold-based hash storage. The time-based strategy ensures that blockchain updates occur at regular intervals, aggregating critical data every 10 minutes while appending ordinary data less frequently, such as every 60 minutes. This strategy provides predictability and consistency, ensuring that data are recorded regardless of fluctuations in event frequency. In contrast, the threshold-based strategy dynamically adapts the frequency of blockchain write operations based on real-time data flow. A new hash is computed only when the

number of accumulated critical records surpasses a predefined threshold, allowing this method to adaptively respond to variations in the frequency of event occurrence. Unlike the time-based approach, which maintains a fixed schedule, the threshold-based method optimises resource usage by writing data only when necessary, reducing suboptimal transaction requests.

The choice between these strategies depends on the specific system requirements: the time-based strategy is best suited to scenarios where regular and predictable updates are required, while the threshold-based approach is more effective in scenarios where event frequency is highly variable and periodic writing may lead to inefficiencies. For enhanced flexibility and efficiency, the system can also adopt a hybrid strategy that dynamically combines time-based and threshold-based approaches. This combined strategy allows the system to remain cost-effective when the frequency of event occurrence varies, providing configurable parameters to specify the lower and upper bounds for event aggregation before writing.

To evaluate the performance of our proposed architecture, we conducted experiments on the publicly available Incident Dispatch dataset published on the NYC OpenData website [21]. Each entry in the dataset contains detailed information about a specific incident, spanning from the time it was opened in the dispatch system to the moment it was marked as closed. At the time of writing, the dataset includes entries up until 3 October 2024.

To model critical data in our proposed approach, we extracted all dataset fields containing timestamps, which were generated and recorded by the dispatch system whenever significant events occurred. A complete list of these fields (7 out of 31 columns in the dataset), along with their descriptions, is provided in Table 5.1.

Field Name	**Field Description**
`Incident_Datetime`	Incident creation in the dispatch system
`First_Assignment_Datetime`	First unit is assigned to the incident
`First_Activation_Datetime`	First unit is enroute to the incident location
`First_On_Scene_Datetime`	First unit arrives at the incident location
`First_To_Hosp_Datetime`	First unit is enroute to the hospital
`First_Hosp_Arrival_Datetime`	First unit arrives at the hospital
`Incident_Close_Datetime`	Incident closes in the dispatch system

Table 5.1: Name and description of key fields in the incident dispatch dataset, used to identify critical data within the proposed approach.

The dataset contains 24 additional fields that provide supplementary details about each incident, such as the neighbourhood name, dispatch location zip code, and police precinct code. These fields, categorised as ordinary data in our system, are collected by HEA until the corresponding incident is officially closed. Table 5.2 presents a representative subset of these fields, including example values to illustrate the data format.

Field Name	**Description**	**Example 1**	**Example 2**
`CAD_ID`	Unique incident identifier	50011297	50132272
`Call_Type`	Type of emergency call	INJURY	OTHER
`Severity_Code`	Call priority level	5	6
`Disposition_Code`	Final incident outcome	CANCEL	90
`Borough`	Incident borough	BRONX	QUEENS
`Zipcode`	Incident zip code	10467	11369
`Police_Precinct`	Incident police precinct	47	115
`Standby_Indicator`	Indicates if units were on standby	N	Y

Table 5.2: Representative subset of fields categorised as ordinary data in the incident dispatch dataset, with descriptions and example values. The full list of fields and descriptions is available on the dataset homepage [21].

HEA operated continuously on a central server, receiving and storing incoming incident data in its local database. The data gateway was deployed on a cloud node, ensuring robust client interaction and seamless forwarding of incident data to HEA. The blockchain module, operating on a separate node, processed the incident data by generating a hash and recording it on the blockchain according to a predefined storage strategy. Once the transaction was successfully written on the blockchain, a transaction receipt was generated and returned to HEA for logging it onto the local database. Subsequently, both the receipt and the corresponding incident data, along with the computed hash, were transmitted to the cloud database. In our experiments, we evaluated three storage strategies: (i) time-based hash storage, (ii) threshold-based hash storage, and (iii) combined hash storage strategy. These strategies are detailed in the following sections.

5.3 Time-Based Hash Storage

In the time-based strategy, the blockchain module selected data at fixed time intervals to form a transaction batch. We set a time interval of 10 minutes to select critical data, and a time interval of 60 minutes to select the complete record of closed incidents, which comprises all the ordinary data. Subsequently, the transaction batch was processed using the SHA-256 hash function, and the resulting data hash was submitted to the `DataHashStorage` smart contract for permanent storage on the blockchain. The contract was deployed on the Ethereum Sepolia testnet, and is accessible at the following address: $0x2a5A789beBeE29c8fc824d76a82D3aE787b6d1F5$. The contract code, along with the complete transaction history, is accessible on the corresponding Sepolia Etherscan page[1]. The `DataHashStorage` contract implements the function `storeDataHash(bytes32 _dataHash)`, which accepts the SHA-256 hash as input and stores it in an internal mapping. This mapping persistently associates each hash with the timestamp of the block in which it was included, consuming some portion of the storage space in the Sepolia network.

We conducted extensive experiments to evaluate the transaction fees and the time required to write hashes to the blockchain. In total, we executed 2,050 transactions on the Sepolia network, spanning different days of the week and different times of the day. The success rate for executing transactions was 100%. Figure 5.3 presents a histogram illustrating the distribution of block confirmation times.

[1] https://sepolia.etherscan.io/address/0x2a5a789bebee29c8fc824d76a82d3ae787b6d1f5

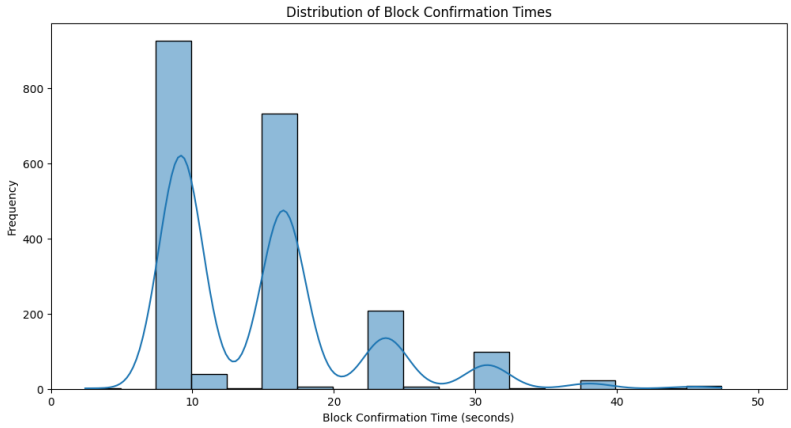


Figure 5.3: Histogram illustrating the distribution of block confirmation times, highlighting the variability in transaction processing speeds within the blockchain network.

Our analysis of the data distribution revealed that block confirmation times tend to cluster. To identify the time intervals corresponding to each cluster, we utilised the kernel density estimate (KDE) inferred from the data. Specifically, we extracted the relative minima of the multimodal KDE plot (see Figure 5.3) and used the corresponding x-axis values to define the boundaries for each time cluster. Table 5.3 presents the reference metrics and count values for each cluster.

Confirmation Time (s)	**Mean**	**Median**	**Count**	**Ratio**
00.00–12.82	9.18	9.16	967	47.17%
12.82–20.50	16.43	16.39	739	36.05%
20.50–27.51	23.69	23.63	213	10.39%
27.51–35.19	30.85	30.78	99	4.83%
35.19–41.97	38.09	38.08	22	1.07%
41.97–50.00	45.45	45.49	10	0.49%
Overall	14.83	15.88	2,050	100%

Table 5.3: Summary of cluster analysis metrics for block confirmation times.

To evaluate the costs of hash writing, we analysed the transaction receipts and calculated the fees in Gwei (1 Gwei $= 1 \times 10^{-9}$ ETH). The total fee was calculated by multiplying the gas price at the time of block creation by the gas units consumed. In our experiments, we set a gas limit to 44,500 units and a maximum priority fee of 0.001 Gwei per transaction. To convert fees from Gwei to USD, we used CoinGecko[2], a publicly accessible cryptocurrency data aggregator that provides real-time exchange rates, including ETH-USD.

To evaluate the effect of the time of day on block confirmation times, we grouped transactions into three equally-sized timespans, each spanning eight hours. Table 5.4 reports the metrics for each transaction group. Figure 5.4 shows the results of this analysis as a plot of the distribution of transaction fees in Gwei and USD. Figure 5.5 presents boxplots of transaction execution times, while Figure 5.6 displays boxplots of transaction fees.

Timespan	**Mean Time**	**Median Time**	**Mean USD**	**Median USD**	**Count**
00:00–08:00	16.59	16.23	0.21	0.17	677
08:00–16:00	13.29	9.56	0.41	0.29	707
16:00–00:00	14.69	16.01	0.22	0.17	666
Overall	14.83	15.88	0.29	0.19	2050

Table 5.4: Comparison of key transaction metrics across different timespans.

[2] https://www.coingecko.com

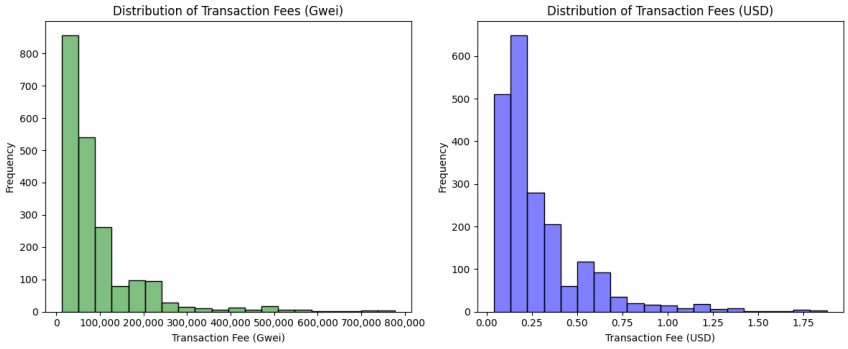


Figure 5.4: Comparative distribution of transaction fees, presented in both Gwei (left plot) and USD (right plot), illustrating cost variability in blockchain-based data recording.

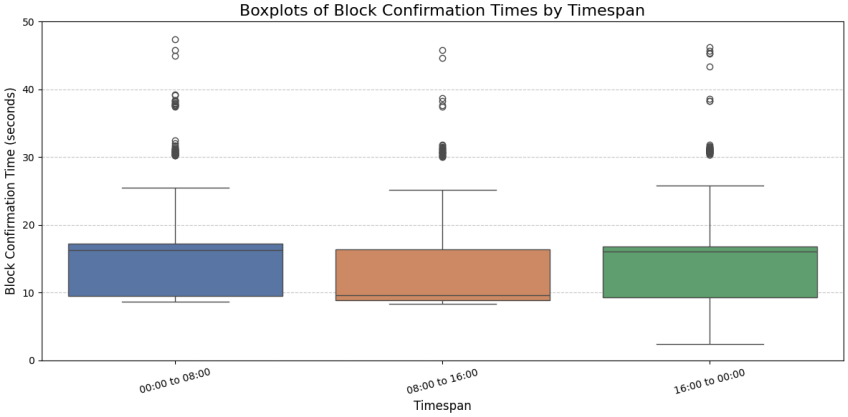


Figure 5.5: Boxplots representing the distribution of block confirmation times across timespans.

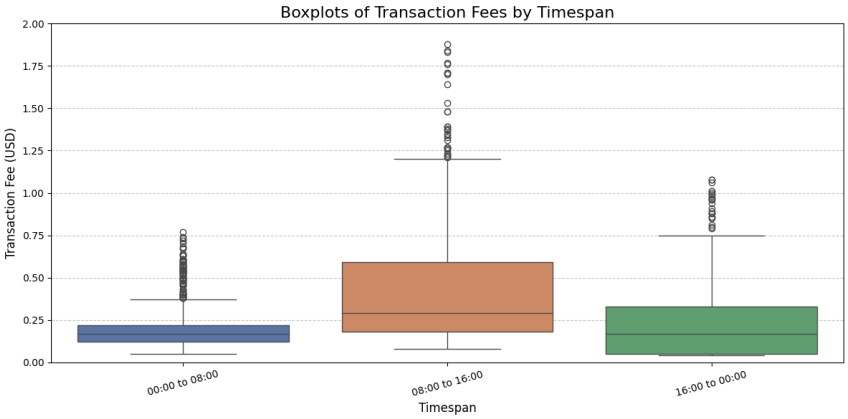


Figure 5.6: Boxplots illustrating transaction fee variations across different timespans.

5.4 Threshold-Based Hash Storage

In the threshold-based strategy, the blockchain module selected critical data from its local database to compute a hash only when a predefined threshold for the number of critical data entries was reached. The handling of ordinary data remained unchanged: Every 60 minutes, complete records of closed incidents were included in the transaction batch.

To determine the frequency of critical events in a real-world scenario, we conducted a comprehensive analysis of the Incident Dispatch Data dataset. The dataset was cleaned and filtered as follows: Firstly, records covering the span of one year, from 1 January 2023 to 31 December 2023, were extracted; subsequently, duplicates and incidents marked as cancelled were removed; and finally, entries lacking an incident conclusion timestamp were discarded. The resulting clean dataset consisted in 1,457,802 unique incident entries.

The analysis of the clean dataset revealed consistent daily incident counts throughout the year, with an average of 3,994 incidents per day and a median of 4,010. Each incident had a maximum of seven critical events counted (see Table 5.1), though not all entries include every critical timestamp. For example, when an incident concludes with patient treatment at the location, timestamps for `First_To_Hosp_Datetime` and `First_Hosp_Arrival_Datetime` were not recorded and the incident was subsequently marked as closed.

For our experiments, we selected 3 October 2023, the day with the highest recorded number of critical data entries. On this day, 4,499 unique incidents were recorded, corresponding to a total of 28,119 critical timestamps. We evaluated the system's performance for writing data hashes while using four threshold values for critical data: 100, 150, 175, and 200. Figure 5.7 illustrates the hourly costs of writing

transactions for each threshold, with the mean hourly cost of the time-based strategy included as a reference.

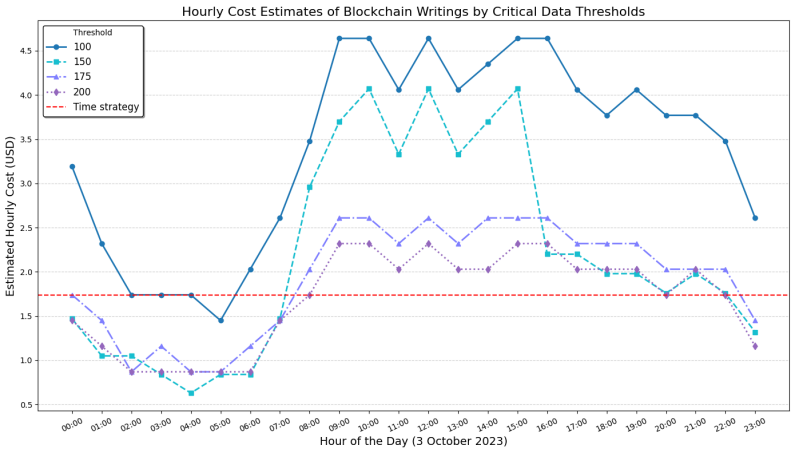


Figure 5.7: Analysis of cost variations based on different data threshold values on 3 October 2023, the day with the highest recorded volume of critical incident data.

Table 5.5 presents the total number of hash writing operations and the corresponding estimated costs. The estimates are based on a mean transaction fee of USD 0.29 (see Table 5.3). For comparison, the total values for the time-based strategy are also included at the bottom of the table.

Threshold	**Total Transactions**	**Mean Cost (USD)**	**Mean Transactions**	**Max Transactions**
100	281	81.49	11.7	16
150	187	54.23	7.8	11
175	160	46.4	6.67	9
200	140	40.60	5.83	8
Time strategy	144	41.76	6	6

Table 5.5: Blockchain hash writing operations and their associated costs for different thresholds, on 3 October 2023, the day with the highest recorded volume of critical incident data.

5.5 Combined Hash Storage Strategy

To provide a more accurate cost assessment and demonstrate the system's flexibility, we extended our analysis over one month and developed a hash writing strategy that combines both time-based and threshold-based approaches. This strategy defines a minimum and maximum threshold, which allows for an adaptive storage mechanism. At fixed 10-minute intervals, the system evaluates the volume of accumulated critical data. If the volume falls below the minimum threshold, hash computation and writing are skipped, thereby reducing unnecessary transactions and optimising cost efficiency. Conversely, if the maximum threshold is reached before the next scheduled interval, the system immediately performs a hash computation and writes the critical data to the blockchain. This ensures timely data aggregation while maintaining an efficient balance between cost and performance.

We conducted a detailed cost analysis of the combined strategy using various threshold pairs. The analysis used data from July 2023, the month with the highest recorded volume of critical data. In July 2023, 128,296 unique incidents were recorded, corresponding to a total of 799,008 critical timestamps. Figure 5.8 illustrates the estimated daily costs for four threshold configurations (125–175, 115–190, 135–195, and 145–205), compared to the time-based strategy.

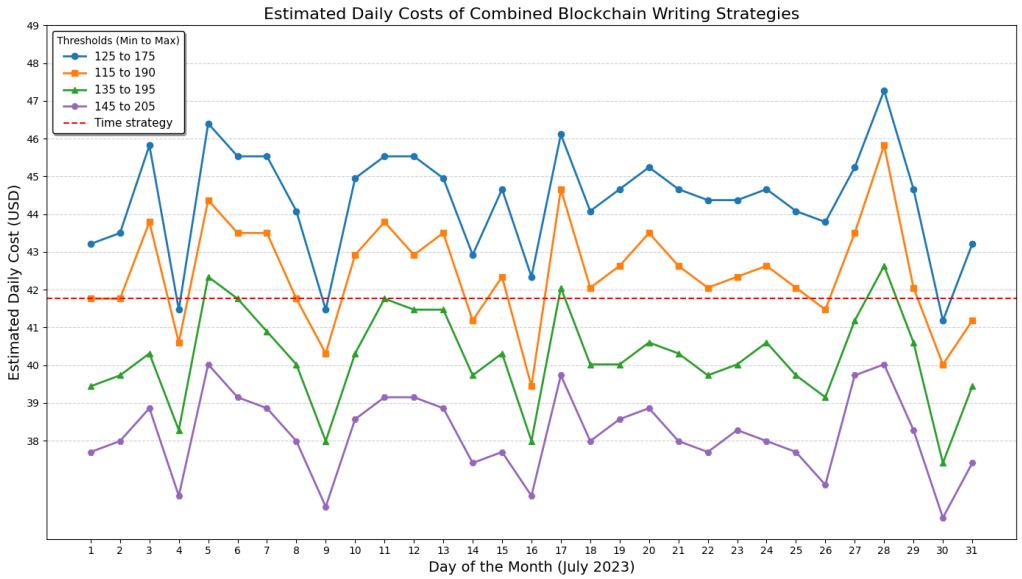


Figure 5.8: Cost analysis of combined blockchain data writing strategies over July 2023, the month with the highest volume of critical data.

Table 5.6 summarises the total number of hash-writing operations and the estimated costs associated with each threshold pair. These estimates are based on an average transaction fee of USD 0.29 (see Table 5.3). For comparison, the corresponding values for the time-based strategy are also included.

Thresholds	**Mean Daily Cost (USD)**	**Mean Daily Transactions**	**Max Daily Transactions**
125–175	44.37	153	163
115–190	42.45	146.4	158
135–195	40.23	138.7	147
145–205	38.19	131.7	138
Time strategy	41.76	144	144

Table 5.6: Hash writing operations and estimated costs for different pairs of threshold values in July 2023, the month with the highest volume of critical data.

5.6 Discussion

A total of 2,050 hash writing requests were confirmed on the Ethereum Sepolia blockchain, spanning various days and times of the week. The time-based strategy executed 144 daily transactions, whereas the threshold-based strategy ranged from 140 to 281 daily writing operations, depending on threshold values (see Table 5.5). Consequently, the experiments covered over two weeks of operation for the time-based strategy and one to two weeks for the threshold-based strategy. This transaction volume, designed to assess the performance, timing, and cost efficiency of blockchain-based hash writing under realistic conditions, provides a sufficient basis for meaningful economic feasibility estimations.

The combined strategy was evaluated using comprehensive data spanning a full month, ensuring a more extensive analysis of its adaptability and cost efficiency under varying operational conditions. This broader timeframe allowed us to assess long-term performance trends, capturing fluctuations in data volume and evaluating threshold configurations accordingly. Our experimental results indicate that nearly half (47.17%) of the total transactions were written on the blockchain in less than 13 s (see Table 5.3), with a mean confirmation time of 9.18 s and a median of 9.16 s. Moreover, the vast majority of transactions (93.61%) were confirmed within 30 s, with overall mean and median confirmation times of 14.83 s and 15.88 s, respectively. The confirmation time boxplots in Figure 5.5 show that transactions were confirmed more rapidly between 08:00 and 16:00, with a mean time of 13.29 s and a median of 9.56 s (see Table 5.4). In contrast, slower confirmation times were observed during the 00:00–08:00 and 16:00–00:00 timespans, with median values of 16.23 s and 16.01 s, respectively. This represents an increase of approximately 67–70% compared to the 08:00–16:00 timespan.

The integration of blockchain technology significantly augments the system by providing a verifiable timeline of events. By permanently recording event hashes on the blockchain, any change to events stored in the database would be detected; hence, a dispute on recorded timestamps could be resolved. A match between the hash computed on recorded timestamps and the blockchain hash would demonstrate that timestamps were unaltered, while a difference would show that they had been altered. This is vital for accountability and legal verification in emergency response scenarios. Notably, in our experiments, all blockchain transactions were confirmed, underscoring the reliability of established platforms such as Ethereum for secure data storage. In cases where a transaction is not confirmed, the affected data remain securely stored in the local database and the hash writing request is reissued, leveraging the inherent resilience of blockchain technology to ensure reliable completion. Collectively, these features enhance data integrity and fault tolerance and provide an indisputable record of incident timelines.

In every strategy we tested, the data aggregation performed by the HEA prevents congestion by avoiding overly frequent writing requests on the blockchain. In the transaction fee distributions (Figure 5.4), we observed that 83.61% transactions incurred costs below USD 0.50. The difference between the Gwei and USD distributions is due to the actual fee in Gwei, which depends on the congestion of the blockchain network at the time of the transaction, and the ETH/USD exchange rate, which is subject to continuous fluctuations. To handle price fluctuations, some strategies can be employed. For example, a polling mechanism can be activated some minutes before the timeout to monitor the current gas cost and request a transaction whenever the cost falls below a set threshold. Depending on the system requirements, either the transaction cost or the timeout to send the transaction can be prioritised.

The analysis of the transaction fee boxplots in Figure 5.6 revealed that the 08:00–16:00 timespan was the most expensive, with a significant number of outliers: the median fee was USD 0.29 and the mean fee was USD 0.41 (see Table 5.4). Conversely, the 00:00–08:00 and 16:00–00:00 timespans exhibited lower and less variable fees, with median values of USD 0.17 and mean values of USD 0.21 and USD 0.22, respectively. We estimated the total daily transaction cost using the mean transaction fee of USD 0.29 (see Table 5.4). In the time-based strategy, 144 daily transactions amounted to an average cost of USD 41.76. In the threshold-based strategy, daily costs ranged from USD 40.60 for a threshold of 200 to USD 81.49 for a threshold of 100 (see Table 5.5). In the combined strategy, estimated daily costs varied from USD 38.19 for the threshold pair of 145–205 to USD 44.37 for the threshold pair of 125–175 (see Table 5.6).

The effectiveness of the three strategies can be assessed through their impact on writing frequency and cost efficiency. The time-based strategy enforced a fixed update interval of one hash every 10 minutes, offering consistency but potentially incurring unnecessary transaction costs during low-activity periods. In contrast, the threshold-based strategy responded dynamically to data flow, increasing the frequency of writing operations when critical data accumulation exceeded predefined thresholds. Even in the worst-case scenario (threshold = 100), the system maintained an average of 12 writing operations per hour, demonstrating a balance between responsiveness and cost (see Table 5.5).

The combined strategy leveraged both approaches to optimise cost-effectiveness while ensuring timely data recording. Notably, its daily writing frequency remained closely aligned with that of the time-based approach (see Table 5.6). The 125–175 threshold pair configuration resulted in 153 average daily writing operations, peaking

at 163, reflecting only moderate increases of 6% and 13%, respectively, relative to the time-based model. Meanwhile, the 145–205 threshold pair maintained a lower transaction volume than the time-based approach, with a maximum of 138 daily writing operations, compared to 144 in the time-based strategy. This configuration also yielded a mean daily cost of USD 38.19, representing a daily operational cost reduction of USD 3.57 while retaining the combined strategy's capacity for dynamic adaptation to fluctuating data loads.

Our approach significantly reduces the high costs typically associated with on-chain storage of large data volumes. In our system, only a 32-byte hash is written to the blockchain; hence, recording a fixed-length hash incurs only a small fraction of the cost required for storing complete data records. The experimental results confirm that transaction fees for hash writing operations remain low even under variable gas prices. These findings highlight the inherent trade-offs between predictability, cost-efficiency, and adaptability. The time-based strategy ensures a consistent recording schedule but may result in redundant blockchain interactions. Conversely, the threshold-based approach offers greater adaptability but introduces variability in update timing. The combined strategy, particularly with the 145—205 configuration, demonstrates an optimal balance, achieving cost savings while maintaining responsiveness to fluctuations in critical data volume.

Compared to third-party and commercial timestamp certification systems, our approach leverages the inherent properties of a public blockchain to ensure the critical function of proving data integrity remains within the system rather than being outsourced to external entities that may impose arbitrary conditions for service access and utilisation. By recording data hashes on a public blockchain, the proof of integrity can be independently verified by authorised entities in the event of

disputes. This internal verification process enhances system trust and differentiates our approach from conventional timestamping methods, which rely on external certification providers.

We used a publicly available dataset of the New York metropolitan area, encompassing approximately eight million residents, and conducted experiments to derive average transaction cost values. These values, derived from analysing multiple threshold configurations over various time intervals, offer a practical baseline for expense estimation. This baseline, supported by empirical data, allows for the projection of costs over different durations, thereby enabling more informed decisions regarding investment and operational expenses. The results demonstrate that our approach maintains feasible costs even amid blockchain fee variability. By aggregating data and employing a threshold-triggered strategy, the system avoids unnecessary network congestion and minimises operational expenses. Consequently, these cost estimates underscore the scalability and applicability of our system in densely populated, data-intensive environments.

5.7 Summary

This study presented a blockchain-based framework designed to allow healthcare service providers to hold (and, when necessary, exhibit) proof of the integrity and authenticity of timestamped events, in case of disputes. The designed system ensures accurate and blockchain-certified timestamps for critical events, addressing a major limitation of centralised architectures that depend on a single authority. This capability is achieved through blockchain technology, which provides a transparent and verifiable record of events. The proposed framework permanently records event hashes on the blockchain, making the timeline of incidents (as stored in the cloud)

verifiable. Therefore, the tampering of cloud-stored data becomes futile, as it would be easily detected in case of disputes.

Our experimental evaluation, based on a real-world incident dataset, confirmed that the proposed framework is feasible and effective for the management of secure incidents. The integration of regular, time-based recording with an adaptive threshold strategy enables the system to record and secure critical event data reliably, ensuring that the incident timeline is maintained accurately. This balanced approach ensures that essential information is logged without excessive redundancy, thus reducing resource utilisation and maintaining system efficiency.

The dynamic nature of the combined strategy allows the framework to adapt seamlessly to varying operational conditions, ensuring that the level of data capture aligns closely with the actual flow of incident information. Such a continuous adaptation ensures that the data are kept accurate and verifiable, which is essential in scenarios where swift response and strict accountability are paramount. In addition, the insights derived from our study highlight the broader applicability of blockchain technology in incident management systems. By combining server-side timestamping with blockchain-backed verification, the system guarantees the auditable history of emergency events, fostering trust among stakeholders such as emergency responders, healthcare providers, and patients.

Chapter 6

Conclusions

This doctoral thesis has presented a framework for the management and optimisation of renewable and distributed energy resources. The research addressed several challenges emerging in the transition towards a greener, more sustainable and decentralised energy systems: the need for effective resource planning and assessment, and the requirement for efficient optimisation and verifiable integrity for data flow in energy communities.

On the management side, the framework introduced an innovative and systematic approach for automatic detection of solar photovoltaic panels in aerial imagery. The presented methodology, based on the extraction of a set of characterising colours, demonstrated high detection performance and robustness across different geographical areas without relying on computationally intensive training phases nor large annotated datasets.

This methodology was further extended for the task of land use and land cover classification, providing a unified solution that supports both the identification of suitable sites for new solar installations and the detection of already deployed photovoltaic panels. Together, these contributions provide stakeholders and policy-makers valuable insights for planning and assessing the impact of distributed energy

resources.

On the optimisation side, the framework tackled challenges posed by the fluctuations of renewable energy sources generation, in the domain of energy communities. A software system was designed and integrated into the framework, combining a threshold-based strategy for aggregators with an incentive-based demand response programme leveraging gamified application to enhance user participation. This combined approach supports the balancing of energy demand and supply, improves grid reliability by preventing peak consumption and promotes coordinated behaviour aligned with the welfare of the community.

To ensure fairness and trust, the framework was extended with a blockchain-based certification workflow that guarantees data integrity and provides a cost-effective verifiable record of energy data. The robustness and flexibility of this approach were demonstrated through its extension to timestamp certification in emergency response systems, highlighting its generalisability beyond the energy domain.

In conclusion, the contributions presented in this doctoral thesis advance both the conceptual understanding and the practical implementation of management and optimisation strategies for distributed energy resources in renewable energy communities. Based on empirical validation conducted on real-world datasets, the framework developed as part of this research demonstrated that efficient, robust and privacy-preserving methods can support the planning, deployment, monitoring and optimisation of renewable energy systems, while also providing a solid basis of applicability in other domains.

6.1 Future Work

The contributions presented in this thesis provide the basis for further developments. Future work on PV panel detection could focus on refining the methodology to obtain accurate assessments of power generation potential. Additionally, the charactering colours-based approach could be adapted to analyse temporal image series, enabling the monitoring of PV installations growth and the evaluation of the effectiveness of incentive plans designed to promote their adoption.

On the optimisation side, further research is needed to validate the proposed architecture in large-scale, real-world energy communities, including the deployment of gamified demand response strategies in diverse socio-economic contexts. Moreover, further investigations are needed to evaluate the trade-off between performance, scalability and economic viability in blockchain-based data integrity certification, especially in larger and more data-intensive communities. Finally, the extension of the data certification capabilities of the framework to other critical domains beyond healthcare, such as transport or supply chain management, represents a promising direction to further demonstrate its broader applicability.

Bibliography

[1] International Renewable Energy Agency (IRENA). *Renewable Energy Statistics*. 2023. URL: https://www.irena.org/Publications/2023/Jul/Renewable-energy-statistics-2023 (visited on 10/15/2025).

[2] Akhtar Alam, M Sultan Bhat, and M Maheen. "Using Landsat satellite data for assessing the land use and land cover change in Kashmir valley". In: *GeoJournal, Springer* 85.6 (2020), pp. 1529–1543. DOI: 10.1007/s10708-019-10037-x.

[3] Fahad R Albogamy et al. "An optimal adaptive control strategy for energy balancing in smart microgrid using dynamic pricing". In: *IEEE Access* 10 (2022), pp. 37396–37411. DOI: 10.1109/ACCESS.2022.3164809.

[4] Kishwar Ali et al. "Land usage analysis: a machine learning approach". In: *International Journal of Computer Applications, Foundation of Computer Science* 141.12 (2016), pp. 23–28. DOI: 10.5120/ijca2016909936.

[5] Tarek AlSkaif et al. "Gamification-based framework for engagement of residential customers in energy applications". In: *Energy Research & Social Science* 44 (2018), pp. 187–195. DOI: 10.1016/j.erss.2018.04.043.

[6] Ioannis Antonopoulos et al. "Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review". In: *Renewable*

and Sustainable Energy Reviews 130 (2020), p. 109899. DOI: 10.1016/j.rser.2020.109899.

[7] Ioannis Antonopoulos et al. "Data-driven modelling of energy demand response behaviour based on a large-scale residential trial". In: *Energy and AI* 4 (2021), p. 100071. DOI: 10.1016/j.egyai.2021.100071.

[8] Edoardo Arnaudo et al. "A comparative evaluation of deep learning techniques for photovoltaic panel detection from aerial images". In: *IEEE Access* 11 (2023), pp. 47579–47594. DOI: 10.1109/ACCESS.2023.3275435.

[9] Asaph Azaria et al. "MedRec: Using Blockchain for Medical Data Access and Permission Management". In: *International Conference on Open and Big Data (OBD)*. 2016, pp. 25–30. DOI: 10.1109/OBD.2016.11.

[10] N.J. Bartie et al. "The resources, exergetic and environmental footprint of the silicon photovoltaic circular economy: Assessment and opportunities". In: *Resources, Conservation and Recycling* 169 (2021), p. 105516. ISSN: 0921-3449. DOI: 10.1016/j.resconrec.2021.105516.

[11] Sana Basheer et al. "Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques". In: *Remote Sensing, MDPI* 14.19 (2022). ISSN: 2072-4292. DOI: 10.3390/rs14194978.

[12] Dave Bayer, Stuart Haber, and W. Scott Stornetta. "Improving the Efficiency and Reliability of Digital Time-Stamping". In: *Sequences II*. New York, NY: Springer New York, 1993, pp. 329–334. ISBN: 978-1-4613-9323-8. DOI: 10.1007/978-1-4613-9323-8_24.

[13] Behnaz Behi et al. "Consumer engagement in virtual power plants through gamification". In: *2020 5th international conference on power and renewable*

energy (ICPRE). IEEE. 2020, pp. 131–137. DOI: 10.1109/ICPRE51194.2020.9233110.

[14] Mehdi Benchoufi and Philippe Ravaud. "Blockchain technology for improving clinical research quality". In: *Trials* 18.1 (2017), pp. 1–5. DOI: 10.1186/s13063-017-2035-z.

[15] Manuel Carranza-García, Jorge García-Gutiérrez, and José C. Riquelme. "A Framework for Evaluating Land Use and Land Cover Classification Using Convolutional Neural Networks". In: *Remote Sensing, MDPI* 11.3 (2019). ISSN: 2072-4292. DOI: 10.3390/rs11030274.

[16] *Consumer price index (CPI) rates*. Australian Government - Australian Taxation Office. URL: https://www.ato.gov.au/tax-rates-and-codes/consumer-price-index (visited on 10/15/2025).

[17] *Consumer Price Index, Australia*. Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/economy/price-indexes-and-inflation/consumer-price-index-australia (visited on 10/15/2025).

[18] Daniel W. Czirjak. "Detecting photovoltaic solar panels using hyperspectral imagery and estimating solar power production". In: *Journal of Applied Remote Sensing* 11.2 (2017), p. 026007. DOI: 10.1117/1.JRS.11.026007.

[19] Sonam Dorji et al. "An Extensive Critique on Smart Grid Technologies: Recent Advancements, Key Challenges, and Future Directions". In: *Technologies* 11.3 (2023). ISSN: 2227-7080. DOI: 10.3390/technologies11030081.

[20] Ayobami S. Edun et al. "Unsupervised azimuth estimation of solar arrays in low-resolution satellite imagery through semantic segmentation and Hough

transform". In: *Applied Energy* 298 (2021), p. 117273. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2021.117273.

[21] *EMS Incident Dispatch Data — NYC Open Data*. URL: https://data.cityofnewyork.us/Public-Safety/EMS-Incident-Dispatch-Data/76xm-jjuj (visited on 10/15/2025).

[22] Gabriel Estevam et al. "Accurate and decentralized timestamping using smart contracts on the Ethereum blockchain". In: *Information Processing & Management* 58.3 (2021), p. 102471. ISSN: 0306-4573. DOI: 10.1016/j.ipm.2020.102471.

[23] European Parliament. *Directive (EU) 2018/2001 of the European Parliament and of the Council of 11 December 2018 on the promotion of the use of energy from renewable sources*. 2018. URL: https://eur-lex.europa.eu/eli/dir/2018/2001/oj (visited on 10/15/2025).

[24] Hua Fan, IF MacGill, and Alistair B Sproul. "Statistical analysis of driving factors of residential energy demand in the greater Sydney region, Australia". In: *Energy and Buildings* 105 (2015), pp. 9–25. DOI: 10.1016/j.enbuild.2015.07.030.

[25] Farnaz, Narissara Nuthammachot, and Rabia Shabbir. "Evaluating site selection for optimal photovoltaic installations and CO_2 emission reduction in selected districts of khyber pakhtunkhwa". In: *Scientific Reports* 15.1 (2025), p. 6635. DOI: 10.1038/s41598-025-86713-5.

[26] José Luis Fernández-Alemán et al. "Security and privacy in electronic health records: A systematic literature review". In: *Journal of Biomedical Informatics* 46.3 (2013), pp. 541–562. ISSN: 1532-0464. DOI: 10.1016/j.jbi.2012.12.003.

[27] Hossam A. Gabbar et al. "Demonstration of Resilient Microgrid with Real-Time Co-Simulation and Programmable Loads". In: *Technologies* 10.4 (2022). ISSN: 2227-7080. DOI: 10.3390/technologies10040083.

[28] Christian Giovanelli et al. "Towards an aggregator that exploits big data to bid on frequency containment reserve market". In: *IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society*. IEEE. 2017, pp. 7514–7519. DOI: 10.1109/IECON.2017.8217316.

[29] Faranak Golestaneh, Hoay Beng Gooi, and Pierre Pinson. "Generation and evaluation of space–time trajectories of photovoltaic power". In: *Applied Energy* 176 (2016), pp. 80–91. DOI: 10.1016/j.apenergy.2016.05.025.

[30] Vladimir Golovko et al. "Convolutional neural network based solar photovoltaic panel detection in satellite photos". In: *IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*. Vol. 1. 2017, pp. 14–19. DOI: 10.1109/IDAACS.2017.8094501.

[31] Vladimir Golovko et al. "Development of Solar Panels Detector". In: *International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T)*. 2018, pp. 761–764. DOI: 10.1109/INFOCOMMST.2018.8632132.

[32] F. J. Gómez-Uceda et al. "Benchmarking Analysis of the Panorama of Grid-Connected PV Installations in Spain". In: *Technologies* 10.6 (2022). ISSN: 2227-7080. DOI: 10.3390/technologies10060131.

[33] Google LLC. *Google Maps*. URL: https://www.google.com/maps (visited on 10/15/2025).

[34] Kristen N Griggs et al. "Healthcare blockchain system using smart contracts for secure automated remote patient monitoring". In: *Journal of medical systems* 42 (2018), pp. 1–7. DOI: 10.1007/s10916-018-0982-x.

[35] Mirna Gržanić and Tomislav Capuder. "Coordinated scheduling of renewable energy balancing group". In: *International Journal of Electrical Power & Energy Systems* 125 (2021), p. 106555. DOI: 10.1016/j.ijepes.2020.106555.

[36] Abid Haleem et al. "Blockchain technology applications in healthcare: An overview". In: *International Journal of Intelligent Networks* 2 (2021), pp. 130–139. ISSN: 2666-6030. DOI: 10.1016/j.ijin.2021.09.005.

[37] Jonnadula Harikiran and Tatireddy Reddy. "An outlook: machine learning in hyperspectral image classification and dimensionality reduction techniques". In: *Journal of Spectral Imaging, IM Publications Open LLP* 11.1 (2022), a1. ISSN: 2040-4565. DOI: 10.1255/jsi.2022.a1.

[38] Thomas Hepp et al. "On-chain vs. off-chain storage for supply- and blockchain integration". In: *it - Information Technology* 60.5-6 (2018), pp. 283–291. DOI: doi:10.1515/itit-2018-0019.

[39] Marko Hölbl et al. "A Systematic Review of the Use of Blockchain in Healthcare". In: *Symmetry* 10.10 (2018). ISSN: 2073-8994. DOI: 10.3390/sym10100470.

[40] Mohammad Esmaeil Honarmand et al. "An overview of demand response: From its origins to the smart energy community". In: *IEEE Access* 9 (2021), pp. 96851–96876. DOI: 10.1109/ACCESS.2021.3094090.

[41] Julian de Hoog et al. "Using satellite and aerial imagery for identification of solar PV: State of the art and research opportunities". In: *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*. 2020, pp. 308–313. DOI: 10.1145/3396851.3397681.

[42] Marco Iansiti, Karim R Lakhani, et al. "The truth about blockchain". In: *Harvard business review* 95.1 (2017), pp. 118–127.

[43] International Energy Agency - Photovoltaic Power Systems Program. *Shapshot of Global PV Markets 2025*. DOI: 10.69766/PBHV9141.

[44] International Energy Agency (IEA). *Renewables 2024*. Paris. URL: https://www.iea.org/reports/renewables-2024 (visited on 10/15/2025).

[45] Konstantinos Ioannou and Dimitrios Myronidis. "Automatic Detection of Photovoltaic Farms Using Satellite Imagery and Convolutional Neural Networks". In: *Sustainability* 13.9 (2021). ISSN: 2071-1050. DOI: 10.3390/su13095323.

[46] José Iria, Filipe Soares, and Manuel Matos. "Optimal bidding strategy for an aggregator of prosumers in energy and secondary reserve markets". In: *Applied Energy* 238 (2019), pp. 1361–1372. DOI: 10.1016/j.apenergy.2019.01.191.

[47] Saher Javaid, Mineo Kaneko, and Yasuo Tan. "Energy balancing of power system considering periodic behavioral pattern of renewable energy sources and demands". In: *IEEE Access* 12 (2024), pp. 70245–70262. DOI: 10.1109/ACCESS.2024.3397710.

[48] Jayapriya Jayabalan and N. Jeyanthi. "Scalable blockchain model using off-chain IPFS storage for healthcare data security and privacy". In: *Journal of Parallel and Distributed Computing* 164 (2022), pp. 152–167. ISSN: 0743-7315. DOI: 10.1016/j.jpdc.2022.03.009.

[49] Chaonan Ji et al. "Solar photovoltaic module detection using laboratory and airborne imaging spectroscopy data". In: *Remote Sensing of Environment* 266 (2021), p. 112692. ISSN: 0034-4257. DOI: 10.1016/j.rse.2021.112692.

[50] Hou Jiang et al. "Multi-resolution dataset for photovoltaic panel segmentation from satellite and aerial imagery". In: *Earth System Science Data* 13.11 (2021), pp. 5389–5401. DOI: 10.5194/essd-13-5389-2021.

[51] Daniel Johnson et al. "Gamification and serious games within the domain of domestic energy consumption: A systematic review". In: *Renewable and Sustainable Energy Reviews* 73 (2017), pp. 249–264. DOI: 10.1016/j.rser.2017.01.134.

[52] Shahab Eddin Jozdani, Brian Alan Johnson, and Dongmei Chen. "Comparing Deep Neural Networks, Ensemble Classifiers, and Support Vector Machine Algorithms for Object-Based Urban Land Use/Land Cover Classification". In: *Remote Sensing, MDPI* 11.14 (2019). ISSN: 2072-4292. DOI: 10.3390/rs11141713.

[53] Selina Kerscher and Pablo Arboleya. "The key role of aggregators in the energy transition under the latest European regulatory framework". In: *International Journal of Electrical Power & Energy Systems* 134 (2022), p. 107361. DOI: 10.1016/j.ijepes.2021.107361.

[54] Maksym Koltunov et al. "Mapping of energy communities in Europe: status quo and review of existing classifications". In: *Sustainability* 15.10 (2023), p. 8201. DOI: 10.3390/su15108201.

[55] Paraskevas Koukaras et al. "Integrating blockchain in smart grids for enhanced demand response: Challenges, strategies, and future directions". In: *Energies* 17.5 (2024), p. 1007. DOI: 10.3390/en17051007.

[56] Lucas Kruitwagen et al. "A global inventory of photovoltaic solar energy generating units". In: *Nature* 598.7882 (2021), pp. 604–610. DOI: 10.1038/s41586-021-03957-7.

[57] Amel Ksibi et al. "Secure and Fast Emergency Road Healthcare Service Based on Blockchain Technology for Smart Cities". In: *Sustainability* 15.7 (2023). ISSN: 2071-1050. DOI: 10.3390/su15075748.

[58] Keerti Kulkarni and PA Vijaya. "NDBI based prediction of land use land cover change". In: *Journal of the Indian Society of Remote Sensing, Springer* 49.10 (2021), pp. 2523–2537. DOI: 10.1007/s12524-021-01411-9.

[59] Mohammed Hussien Yadem Lamien and Hooman Farzaneh. "Integrated land use land cover-energy modeling framework for solar energy planning in the future expansion areas; the case of Luxor city Region, Egypt". In: *Energy Conversion and Management: X* 25 (2025), p. 100874. DOI: 10.1016/j.ecmx.2025.100874.

[60] Hai Trieu Le et al. "Patient-Chain: Patient-centered Healthcare System a Blockchain-based Technology in Dealing with Emergencies". In: *Parallel and Distributed Computing, Applications and Technologies*. Cham: Springer International Publishing, 2022, pp. 576–583. ISBN: 978-3-030-96772-7. DOI: 10.1007/978-3-030-96772-7_54.

[61] Walter Leal Filho et al. "Prosumers and sustainable development: An international assessment in the field of renewable energy". In: *Sustainable Futures* 7 (2024), p. 100158. DOI: 10.1016/j.sftr.2024.100158.

[62] H. Levkowitz and G.T. Herman. "GLHS: A Generalized Lightness, Hue, and Saturation Color Model". In: *CVGIP: Graphical Models and Image Processing* 55.4 (1993), pp. 271–285. ISSN: 1049-9652. DOI: 10.1006/cgip.1993.1019.

[63] Qi Li et al. "SolarFinder: Automatic Detection of Solar Photovoltaic Arrays". In: *ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 2020, pp. 193–204. DOI: 10.1109/IPSN48710.2020.00024.

[64] SiMing Liang et al. "Mask R-CNN based segmentation method for satellite imagery of photovoltaics generation systems". In: *Chinese Control Conference (CCC)*. 2020, pp. 5343–5348. DOI: 10.23919/CCC50068.2020.9189474.

[65] Chao Lin et al. "BSeIn: A blockchain-based secure mutual authentication with fine-grained access control system for industry 4.0". In: *Journal of Network and Computer Applications* 116 (2018), pp. 42–52. ISSN: 1084-8045. DOI: 10.1016/j.jnca.2018.05.005.

[66] Johan Lindahl, Robert Johansson, and David Lingfors. "Mapping of decentralised photovoltaic and solar thermal systems by remote sensing aerial imagery and deep machine learning for statistic generation". In: *Energy and AI* 14 (2023), p. 100300. ISSN: 2666-5468. DOI: 10.1016/j.egyai.2023.100300.

[67] Yaoting Liu et al. "Interactformer: Interactive Transformer and CNN for Hyperspectral Image Super-Resolution". In: *Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–15. DOI: 10.1109/TGRS.2022.3183468.

[68] Iraide López et al. "European energy communities: Characteristics, trends, business models and legal framework". In: *Renewable and Sustainable Energy Reviews* 197 (2024), p. 114403. DOI: 10.1016/j.rser.2024.114403.

[69] Jordan M. Malof et al. "A deep convolutional neural network and a random forest classifier for solar photovoltaic array detection in aerial imagery". In: *IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*. 2016, pp. 650–654. DOI: 10.1109/ICRERA.2016.7884415.

[70] Jordan M. Malof et al. "Automatic detection of solar photovoltaic arrays in high resolution aerial imagery". In: *Applied Energy* 183 (2016), pp. 229–240. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2016.08.191.

[71] Valerio Mandarino, Giuseppe Pappalardo, and Emiliano Tramontana. "A Blockchain-Based Electronic Health Record (EHR) System for Edge Computing Enhancing Security and Cost Efficiency". In: *Computers* 13.6 (2024), p. 132. DOI: 10.3390/computers13060132.

[72] Hongzhi Mao et al. "Advances and prospects on estimating solar photovoltaic installation capacity and potential based on satellite and aerial images". In: *Renewable and Sustainable Energy Reviews* 179 (2023), p. 113276. ISSN: 1364-0321. DOI: 10.1016/j.rser.2023.113276.

[73] Xiaobo Mao et al. "Centralized bidding mechanism of demand response based on blockchain". In: *Energy Reports* 8 (2022), pp. 111–117. DOI: 10.1016/j.egyr.2022.02.145.

[74] Daniele Marletta, Alessandro Midolo, and Emiliano Tramontana. "A Blockchain-Based Strategy for Certifying Timestamps in a Distributed Healthcare Emergency Response Systems". In: *Future Internet* 17.5 (2025), p. 210. DOI: 10.3390/fi17050210.

[75] Daniele Marletta, Alessandro Midolo, and Emiliano Tramontana. "Automatic Land Use and Land Cover Classification by Means of Characterising Colours". In: *2024 32nd International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*. IEEE. 2024, pp. 146–151. DOI: 10.1109/WETICE64632.2024.00034.

[76] Daniele Marletta, Alessandro Midolo, and Emiliano Tramontana. "Detecting Photovoltaic Panels in Aerial Images by Means of Characterising Colours". In: *Technologies, MDPI* 11.6 (2023). ISSN: 2227-7080. DOI: 10.3390/technologies11060174.

[77] Kevin Mayer et al. "3D-PV-Locator: Large-scale detection of rooftop-mounted photovoltaic systems in 3D". In: *Applied Energy* 310 (2022), p. 118469. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2021.118469.

[78] Dakota Aaron McCarty, Hyun Woo Kim, and Hye Kyung Lee. "Evaluation of Light Gradient Boosted Machine Learning Technique in Large Scale Land Use and Land Cover Classification". In: *Environments, MDPI* 7.10 (2020). ISSN: 2076-3298. DOI: 10.3390/environments7100084.

[79] Chrissa McFarlane et al. "Patientory: A healthcare peer-to-peer EMR storage network v1". In: *Entrust Inc.: Addison, TX, USA* 3 (2017), p. 19. URL: https://patientoryassociation.org/wp-content/uploads/2018/11/patientory_whitepaper.pdf (visited on 10/15/2025).

[80] Thomas McGhin et al. "Blockchain in healthcare applications: Research challenges and opportunities". In: *Journal of Network and Computer Applications* 135 (2019), pp. 62–75. ISSN: 1084-8045. DOI: 10.1016/j.jnca.2019.02.027.

[81] Wolde Mekuria et al. "Effects of Long-Term Land Use and Land Cover Changes on Ecosystem Service Values: An Example from the Central Rift Valley, Ethiopia". In: *Land, MDPI* 10.12 (2021). ISSN: 2073-445X. DOI: 10.3390/land10121373.

[82] Juana Isabel Méndez et al. "Empower saving energy into smart communities using social products with a gamification structure for tailored Human–Machine Interfaces within smart homes". In: *International Journal on Interactive Design and Manufacturing (IJIDeM)* 17.3 (2023), pp. 1363–1387. DOI: 10.1007/s12008-022-01141-3.

[83] Long Meng and Liqun Chen. "A Blockchain-Based Long-Term Time-Stamping Scheme". In: *Computer Security – ESORICS 2022*. Ed. by Vijayalakshmi Atluri et al. Cham: Springer International Publishing, 2022, pp. 3–24. ISBN: 978-3-031-17140-6. DOI: 10.1007/978-3-031-17140-6_1.

[84] Long Meng and Liqun Chen. "Analysis of Client-Side Security for Long-Term Time-Stamping Services". In: *Applied Cryptography and Network Security*. Ed. by Kazue Sako and Nils Ole Tippenhauer. Cham: Springer International Publishing, 2021, pp. 28–49. ISBN: 978-3-030-78372-3. DOI: 10.1007/978-3-030-78372-3_2.

[85] Long Meng and Liqun Chen. "Long-Term Time-Stamping Schemes Based on MACs, Archives, and Transient Keys". In: *2024 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing and*

Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics. 2024, pp. 535–542. DOI: 10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics62450.2024.00101.

[86] Aljoscha Meyer. "SoK: Authenticated Prefix Relations–A Unified Perspective On Relative Time-Stamping and Append-Only Logs". In: *arXiv preprint arXiv:2308.13836* (2023). DOI: 10.48550/arXiv.2308.13836.

[87] Mohammed Abdulmajeed Moharram and Divya Meena Sundaram. "Dimensionality reduction strategies for land use land cover classification based on airborne hyperspectral imagery: a survey". In: *Environmental Science and Pollution Researchm, Springer* 30.3 (2023), pp. 5580–5602. DOI: 10.1007/s11356-022-24202-2.

[88] Mohammed Abdulmajeed Moharram and Divya Meena Sundaram. "Land use and land cover classification with hyperspectral data: A comprehensive review of methods, challenges and future directions". In: *Neurocomputing, Elsevier* 536 (2023), pp. 90–113. DOI: 10.1016/j.neucom.2023.03.025.

[89] Amir Mohammad Moradi Sizkouhi et al. "Automatic Boundary Extraction of Large-Scale Photovoltaic Plants Using a Fully Convolutional Network on Aerial Imagery". In: *IEEE Journal of Photovoltaics* 10.4 (2020), pp. 1061–1067. DOI: 10.1109/JPHOTOV.2020.2992339.

[90] Matthew Moraguez et al. "Convolutional Neural Network for Detection of Residential Photovoltalc Systems in Satellite Imagery". In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. 2020, pp. 1600–1603. DOI: 10.1109/IGARSS39084.2020.9324245.

[91] Omid Motlagh, Greg Foliente, and George Grozev. "Knowledge-mining the Australian smart grid smart city data: A statistical-neural approach to demand-response analysis". In: *Planning support systems and smart cities*. Springer, 2015, pp. 189–207. DOI: 10.1007/978-3-319-18368-8_10.

[92] Deepan Muthirayan et al. "A minimal incentive-based demand response program with self reported baseline mechanism". In: *IEEE Transactions on Smart Grid* 11.3 (2019), pp. 2195–2207. DOI: 10.1109/TSG.2019.2949263.

[93] Timothy Nugent, David Upton, and Mihai Cimpoesu. "Improving data transparency in clinical trials using blockchain smart contracts". In: *F1000Research* 5 (2016). DOI: 10.12688/f1000research.9756.1.

[94] Mikko Nykyri, Salla Annala, Pertti Silventoinen, et al. "Review of demand response and energy communities in serious games". In: *IEEE Access* 10 (2022), pp. 91018–91026. DOI: 10.1109/ACCESS.2022.3202013.

[95] Ammar Odeh, Ismail Keshta, and Qasem Abu Al-Haija. "Analysis of Blockchain in the Healthcare Sector: Application and Issues". In: *Symmetry* 14.9 (2022). ISSN: 2073-8994. DOI: 10.3390/sym14091760.

[96] Oluwaseun Priscilla Olawale and Sahar Ebadinezhad. "Cybersecurity Anomaly Detection: AI and Ethereum Blockchain for a Secure and Tamperproof IoHT Data Management". In: *IEEE Access* 12 (2024), pp. 131605–131620. DOI: 10.1109/ACCESS.2024.3460428.

[97] Shalini Pal and Rajesh Kumar. "Price prediction techniques for residential demand response using support vector regression". In: *2016 IEEE 7th Power India International Conference (PIICON)*. IEEE. 2016, pp. 1–6. DOI: 10.1109/POWERI.2016.8077427.

[98] Poonam Parhar et al. "HyperionSolarNet: solar panel detection from aerial images". In: *arXiv preprint arXiv:2201.02107* (2022). DOI: 10.48550/arXiv.2201.02107.

[99] KiSung Park et al. "BPPS: Blockchain-enabled privacy-preserving scheme for demand-response management in smart grid environments". In: *IEEE Transactions on Dependable and Secure Computing* 20.2 (2022), pp. 1719–1729. DOI: 10.1109/TDSC.2022.3163138.

[100] Stefan Pfenninger and Iain Staffell. "Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data". In: *Energy* 114 (2016), pp. 1251–1265. DOI: 10.1016/j.energy.2016.08.060.

[101] Thanh Noi Phan, Verena Kuch, and Lukas W. Lehnert. "Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition". In: *Remote Sensing, MDPI* 12.15 (2020). ISSN: 2072-4292. DOI: 10.3390/rs12152411.

[102] Darius Phiri and Justin Morgenroth. "Developments in Landsat Land Cover Classification Methods: A Review". In: *Remote Sensing, MDPI* 9.9 (2017). ISSN: 2072-4292. DOI: 10.3390/rs9090967.

[103] Pierre Pinson, Henrik Madsen, et al. "Benefits and challenges of electrical demand response: A critical review". In: *Renewable and Sustainable Energy Reviews* 39 (2014), pp. 686–699. DOI: 10.1016/j.rser.2014.07.098.

[104] Claudia Pop et al. "Blockchain based decentralized management of demand response programs in smart energy grids". In: *Sensors* 18.1 (2018), p. 162. DOI: 10.3390/s18010162.

[105] Bin Qian et al. "IoETTS: A Decentralized Blockchain-based Trusted Timestamping Scheme for Internet of Energy". In: *Journal of Internet Technology* 24.2 (2023), pp. 519–529. DOI: 10.53106/160792642023032402027.

[106] MM Shah Porun Rana and Md Moniruzzaman. "Demarcation of suitable site for solar photovoltaic power plant installation in Bangladesh using geospatial techniques". In: *Next energy* 3 (2024), p. 100109. DOI: 10.1016/j.nxener.2024.100109.

[107] Mike B Roberts et al. "Characterisation of Australian apartment electricity demand and its implications for low-carbon cities". In: *Energy* 180 (2019), pp. 242–257. DOI: 10.1016/j.energy.2019.04.222.

[108] G. Saravanan, G. Yamuna, and S. Nandhini. "Real time implementation of RGB to HSV/HSI/HSL and its reverse color space models". In: *International Conference on Communication and Signal Processing (ICCSP)*. 2016, pp. 0462–0466. DOI: 10.1109/ICCSP.2016.7754179.

[109] Arijet Sarker et al. "Blockchain Handshaking with Software Assurance: Version++ Protocol for Bitcoin Cryptocurrency". In: *Electronics* 13.19 (2024). ISSN: 2079-9292. DOI: 10.3390/electronics13193857.

[110] Rebecka Savolainen and Risto Lahdelma. "Optimization of renewable energy for buildings with energy storages and 15-minute power balance". In: *Energy* 243 (2022), p. 123046. DOI: 10.1016/j.energy.2021.123046.

[111] Maximilian Schulz, Bilel Boughattas, and Frank Wendel. "DetEEktor: Mask R-CNN based neural network for energy plant identification on aerial photographs". In: *Energy and AI* 5 (2021), p. 100069. ISSN: 2666-5468. DOI: 10.1016/j.egyai.2021.100069.

[112] Grant J. Scott et al. "Training Deep Convolutional Neural Networks for Land–Cover Classification of High-Resolution Imagery". In: *Geoscience and Remote Sensing Letters, IEEE* 14.4 (2017), pp. 549–553. DOI: 10.1109/LGRS.2017.2657778.

[113] Shuyun Shi et al. "Applications of blockchain in ensuring the security and privacy of electronic health record systems: A survey". In: *Computers & Security* 97 (2020), p. 101966. ISSN: 0167-4048. DOI: 10.1016/j.cose.2020.101966.

[114] Pierluigi Siano. "Demand response and smart grids—A survey". In: *Renewable and sustainable energy reviews* 30 (2014), pp. 461–478. DOI: 10.1016/j.rser.2013.10.022.

[115] Aljoša Slameršak, Giorgos Kallis, and Daniel W O'Neill. "Energy requirements and carbon emissions for a low-carbon energy transition". In: *Nature communications* 13.1 (2022), p. 6932. DOI: 10.1038/s41467-022-33976-5.

[116] *Smart-Grid Smart-City Customer Trial Data*. Australian Government - Department of Climate Change, Energy, the Environment and Water. URL: https://data.gov.au/data/dataset/smart-grid-smart-city-customer-trial-data (visited on 10/15/2025).

[117] Roberto Spina and Emiliano Tramontana. "An automated classification system for urban areas matching the 'city country fingers' pattern: the cases of Kamakura (Japan) and Acireale (Italy) cities". In: *Journal of Urban Ecology* 7.1 (Sept. 2021), juab023. ISSN: 2058-5543. DOI: 10.1093/jue/juab023.

[118] Roberto Spina and Emiliano Tramontana. "An Image-Processing Approach for Computing the Size of Green Areas in Cities". In: *9th International*

Conference on Computer and Communications Management. Singapore, 2021, pp. 59–65. ISBN: 9781450390071. DOI: 10.1145/3479162.3479171.

[119] *Supplementary Materials*. URL: https://github.com/damarletta/detecting-pv-panels (visited on 10/15/2025).

[120] Chakkaphong Suthaputchakun and Yue Cao. "Blockchain-based Secure Ambulance-to-Everything Communications in Emergency Rescue Operations". In: *International Conference on Ubiquitous and Future Networks (ICUFN)*. 2022, pp. 174–179. DOI: 10.1109/ICUFN55119.2022.9829619.

[121] Swapan Talukdar et al. "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review". In: *Remote Sensing, MDPI* 12.7 (2020). ISSN: 2072-4292. DOI: 10.3390/rs12071135.

[122] Anushree Tandon et al. "Blockchain in healthcare: A systematic literature review, synthesizing framework and future research agenda". In: *Computers in Industry* 122 (2020), p. 103290. ISSN: 0166-3615. DOI: 10.1016/j.compind.2020.103290.

[123] Wen-Jun Tang, Yi-Syuan Wu, and Hong-Tzer Yang. "Adaptive segmentation and machine learning based potential DR capacity analysis". In: *2018 18th International Conference on Harmonics and Quality of Power (ICHQP)*. IEEE. 2018, pp. 1–5. DOI: 10.1109/ICHQP.2018.8378922.

[124] Priit Ulmas and Innar Liiv. "Segmentation of satellite imagery using u-net models for land cover classification". In: *arXiv preprint arXiv:2003.02899* (2020). DOI: 10.48550/arXiv.2003.02899.

[125] Ava Vali, Sara Comai, and Matteo Matteucci. "Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral

Earth Observation Data: A Review". In: *Remote Sensing, MDPI* 12.15 (2020). ISSN: 2072-4292. DOI: 10.3390/rs12152495.

[126] Olivier Van Cutsem et al. "Cooperative energy management of a community of smart-buildings: A Blockchain approach". In: *International Journal of electrical power & energy systems* 117 (2020), p. 105643. DOI: 10.1016/j.ijepes.2019.105643.

[127] Job Rosier Veerle Plakman and Jasper van Vliet. "Solar park detection from publicly available satellite imagery". In: *GIScience & Remote Sensing* 59.1 (2022), pp. 462–481. DOI: 10.1080/15481603.2022.2036056.

[128] Martin Vigil et al. "An efficient time-stamping solution for long-term digital archiving". In: *IEEE 33rd International Performance Computing and Communications Conference*. 2014, pp. 1–8. DOI: 10.1109/PCCC.2014.7017099.

[129] Naftaly Wambugu et al. "A hybrid deep convolutional neural network for accurate land cover classification". In: *International Journal of Applied Earth Observation and Geoinformation, Elsevier* 103 (2021), p. 102515. ISSN: 1569-8432. DOI: 10.1016/j.jag.2021.102515.

[130] Wei Wang et al. "Application of energy storage in integrated energy systems—A solution to fluctuation and uncertainty of renewable energy". In: *Journal of Energy Storage* 52 (2022), p. 104812. DOI: 10.1016/j.est.2022.104812.

[131] Irena Wasiak et al. "Innovative energy management system for low-voltage networks with distributed generation based on prosumers' active participation". In: *Applied Energy* 312 (2022), p. 118705. DOI: 10.1016/j.apenergy.2022.118705.

[132] Yang Weng and Ram Rajagopal. "Probabilistic baseline estimation via gaussian process". In: *2015 IEEE power & energy society general meeting*. IEEE. 2015, pp. 1–5. DOI: 10.1109/PESGM.2015.7285756.

[133] Yang Weng, Jiafan Yu, and Ram Rajagopal. "Probabilistic baseline estimation based on load patterns for better residential customer rewards". In: *International Journal of Electrical Power & Energy Systems* 100 (2018), pp. 508–516. DOI: 10.1016/j.ijepes.2018.02.049.

[134] Kelce S Wilson. *Blockchain with daisy chained records, document corral, quarantine, message timestamping, and self-addressing*. US Patent 11,444,776. Sept. 2022. URL: https://patents.justia.com/patent/11444776 (visited on 10/15/2025).

[135] Xigao Wu et al. "M2m blockchain: The case of demand side management of smart grid". In: *2017 IEEE 23rd international conference on parallel and distributed systems (ICPADS)*. IEEE. 2017, pp. 810–813. DOI: 10.1109/ICPADS.2017.00113.

[136] Qi Xia et al. "BBDS: Blockchain-Based Data Sharing for Electronic Medical Records in Cloud Environments". In: *Information* 8.2 (2017). ISSN: 2078-2489. DOI: 10.3390/info8020044.

[137] Qi Xia et al. "MeDShare: Trust-Less Medical Data Sharing Among Cloud Service Providers via Blockchain". In: *IEEE Access* 5 (2017), pp. 14757–14767. DOI: 10.1109/ACCESS.2017.2730843.

[138] Zilong Xia et al. "High-resolution mapping of water photovoltaic development in China through satellite imagery". In: *International Journal of Applied Earth*

Observation and Geoinformation 107 (2022), p. 102707. ISSN: 1569-8432. DOI: 10.1016/j.jag.2022.102707.

[139] Cheng Xu et al. "SlimChain: scaling blockchain transactions through off-chain storage and parallel processing". In: *Proc. VLDB Endow.* 14.11 (July 2021), pp. 2314–2326. ISSN: 2150-8097. DOI: 10.14778/3476249.3476283.

[140] Jiafan Yu et al. "DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States". In: *Joule* 2.12 (2018), pp. 2605–2617. ISSN: 2542-4351. DOI: 10.1016/j.joule.2018.11.021.

[141] Rui Yuan et al. *A Synthetic Dataset of Danish Residential Electricity Prosumers*. June 2023. DOI: 10.6084/m9.figshare.c.6383862.v1.

[142] Rui Yuan et al. "A synthetic dataset of danish residential electricity prosumers". In: *Scientific Data* 10.1 (2023), p. 371. DOI: 10.1038/s41597-023-02271-3.

[143] Xiao Yue et al. "Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control". In: *Journal of medical systems* 40 (2016), pp. 1–8. DOI: 10.1007/s10916-016-0574-6.

[144] Bowei Zhang et al. "A Blockchain and Zero Knowledge Proof Based Data Security Transaction Method in Distributed Computing". In: *Electronics* 13.21 (2024). ISSN: 2079-9292. DOI: 10.3390/electronics13214260.

[145] Ce Zhang et al. "Joint Deep Learning for land cover and land use classification". In: *Remote Sensing of Environment, Elsevier* 221 (2019), pp. 173–187. ISSN: 0034-4257. DOI: 10.1016/j.rse.2018.11.014.

[146] Ce Zhang et al. "Scale Sequence Joint Deep Learning (SS-JDL) for land use and land cover classification". In: *Remote Sensing of Environment, Elsevier* 237 (2020), p. 111593. ISSN: 0034-4257. DOI: 10.1016/j.rse.2019.111593.

[147] Sichao Zhang et al. "An Efficient Blockchain-Based Time-Stamping Scheme Using Commitment Signatures". In: *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*. 2022, pp. 2591–2596. DOI: 10.1109/GLOBECOM48099.2022.10001631.

[148] Xunhe Zhang et al. "Texture Is Important in Improving the Accuracy of Mapping Photovoltaic Power Plants: A Case Study of Ningxia Autonomous Region, China". In: *Remote Sensing* 13.19 (2021). ISSN: 2072-4292. DOI: 10.3390/rs13193909.

[149] Yuan Zhang et al. "Chronos: An Accurate Blockchain-Based Time-Stamping Scheme for Cloud Storage". In: *IEEE Transactions on Services Computing* 13.2 (2020), pp. 216–229. DOI: 10.1109/TSC.2019.2947476.

[150] Bei Zhao, Yanfei Zhong, and Liangpei Zhang. "A spectral–structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery". In: *Journal of Photogrammetry and Remote Sensing* 116 (2016), pp. 73–85. ISSN: 0924-2716. DOI: 10.1016/j.isprsjprs.2016.03.004.

[151] Xiaochen Zheng et al. "Blockchain-based Personal Health Data Sharing System Using Cloud Storage". In: *IEEE International Conference on e-Health Networking, Applications and Services (Healthcom)*. 2018, pp. 1–6. DOI: 10.1109/HealthCom.2018.8531125.