UNIVERSITÀ
degli STUDI
di CATANIA

DIPARTIMENTO DI INGEGNERIA ELETTRICA
ELETTRONICA E INFORMATICA
Dottorato di Ricerca in Ingegneria dei Sistemi, Energetica,
Informatica e delle Telecomunicazioni
XXXVIII Ciclo

Sarah Di Grande

Machine Learning for Computational Sustainability: Predictive Models for Energy, Urban Mobility, and Healthcare

Coordinatore:
Chiar.mo Prof. Paolo Pietro Arena
Tutor:
Chiar.mo Prof. Ing. Salvatore Cavalieri
Co-Tutor:
Ing. Roberto Gueli

ANNO ACCADEMICO 2024/2025

A Mamma e Papà per avermi insegnato a muovere i primi, cauti passi e per aver
continuato ad accompagnare e sostenere il mio cammino, infondendomi la forza
e l'entusiasmo per affrontare anche le corse più impegnative.

A Federico, per aver scelto con amore e ferma volontà di camminare al mio fianco in
questo decennio di vita e di sogni. La prospettiva del nostro futuro condiviso e della
famiglia che costruiremo insieme è il mio più grande traguardo
e la mia più dolce motivazione quotidiana.

Alla mia Grande Famiglia, che abbraccia il presente, vive nel ricordo e guarda al futuro.
Siete il ponte saldo che congiunge le mie radici alla direzione del mio cammino.

Summary

Chapter 1: Introduction

1.1 Background and Motivation

Sustainability, is widely recognized as both a global goal and one of the most urgent challenges of the 21st century. As a concept rooted in the Latin word *sustinere*, meaning “to uphold” or “to support” (*Sustain - Etymology, Origin & Meaning*, 2025), it encompasses the ability of humanity to coexist with Earth’s ecosystems over the long term. Despite its widespread use, the term has taken on varying definitions across disciplines and contexts. At its core, sustainability is generally understood as the capacity to meet present needs without compromising the ability of future generations to meet theirs (Scoones, 2007). Beyond its frequent use in political and public discourse, sustainability has become a guiding concept for research and policy, particularly in areas where environmental, social, and economic systems are deeply interdependent.

Although today the term is closely associated with environmental, economic, and social dimensions, its origins can be traced back several centuries. The first systematic articulation of sustainability is attributed to Hans Carl von Carlowitz, who in his 1712 treatise Sylvicultura Oeconomica argued for the prudent and long-term management of forests to ensure their regenerative capacity. For much of its early history, the concept remained confined to resource stewardship, particularly in forestry. However, during the late twentieth century it gained prominence as a broader paradigm for addressing the complex interrelations between the environment, economy, and society. The environmental crises of the 1960s and 1970s highlighted the risks of unchecked industrial expansion and resource depletion, thereby catalysing global concern. It was within this context that the World Commission on Environment and Development, chaired by Gro Harlem Brundtland, published its seminal report Our Common Future in 1987. This report provided the now canonical definition of sustainable development as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (Scoones, 2007; UN. Secretary-General and World Commission on Environment and Development, 1987), marking the consolidation of sustainability as a guiding principle for international policy and research.

Indeed, modern interpretations of sustainability are typically framed around three interdependent pillars: environmental, economic, and social (Purvis et al., 2019). This tripartite model captures the complexity of sustainable development, emphasizing that ecological integrity, economic viability, and human well-being must be pursued together. As climate change accelerates, urban populations grow, and critical resources such as energy, water, and public health systems face mounting pressure, the demand for scalable, evidence-based, and adaptive solutions becomes increasingly urgent.

While environmental concerns, such as pollution, biodiversity loss, and natural resource depletion, often dominate sustainability discourse, the economic and social pillars are equally vital. A society cannot be deemed sustainable if it fails to ensure access to clean water, healthcare, education, or fair economic opportunities, just as no society can thrive on a damaged planet.

1.2 Sustainable Development Goals

The adoption of the United Nations Sustainable Development Goals (SDGs) in 2015 (*Transforming Our World: The 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs*, 2015) provided a global framework to operationalize sustainability. Building on the Millennium Development Goals

(*Millennium Development Goals (MDGs)*, 2018), which focused primarily on poverty reduction and basic human development, the SDGs expanded both the scope and ambition of the global agenda. They consist of 17 goals addressing interconnected challenges that span poverty, health, education, gender equality, clean energy, sustainable cities, climate action, and ecosystem protection. Crucially, they were endorsed by all 193 UN member states, signalling a collective commitment to a universal vision of sustainable development (United Nations, 2015).

Unlike earlier frameworks, the SDGs explicitly recognize the indivisibility of environmental, social, and economic dimensions of sustainability. They also emphasize inclusivity and global responsibility: sustainable development is not only a priority for developing nations but a universal obligation requiring systemic transformation in every country. Progress toward the SDGs is monitored through more than 230 global indicators, which provide measurable benchmarks for accountability, though disparities in data quality and reporting capacity remain persistent challenges.

This thesis situates itself within this global agenda by examining how machine learning (ML) methods can be applied in domains directly tied to specific SDGs. The studies presented in later chapters focus on energy, urban mobility, and healthcare; these areas align most closely with SDGs 3 (Good Health and Well-Being), 7 (Affordable and Clean Energy), 11 (Sustainable Cities and Communities), and 13 (Climate Action). Beyond their substantive contributions, computational approaches such as ML can also support SDGs implementation more broadly by improving monitoring systems, forecasting outcomes, and helping to navigate trade-offs between competing objectives. In this way, the thesis contributes not only to methodological innovation but also to the global pursuit of sustainability as articulated in the SDGs.

1.3 The Role of Machine Learning

In the last two decades, ML has emerged as a transformative tool across scientific and industrial domains. Its strength lies in the ability to model complex systems, make accurate predictions, and uncover insights from large and dynamic datasets to support early intervention, optimize operations, and reduce inefficiencies. These capabilities are particularly well-suited to the sustainability domain, where timely, accurate, and adaptive responses to complex systems are vital.

ML supports a wide range of applications, from energy demand forecasting and water management to urban traffic optimization and preventive healthcare (Mrabet and Sliti, 2024). These models enable a shift from reactive responses to proactive interventions, allowing systems to be designed not just for efficiency, but for resilience, equity, and long-term viability.

Indeed, ML is not inherently aligned with sustainability. To be effective in this context, it must be applied in ways that explicitly support sustainable goals. This is the premise of the emerging interdisciplinary field known as computational sustainability (Gomes, 2009; Gomes et al., 2019).

1.4 Framing the Thesis: Computational Sustainability

Computational sustainability is a research domain at the intersection of computer science, artificial intelligence, and sustainability science. It was formally introduced in 2009 by Carla Gomes and colleagues at Cornell University, with the goal of developing computational methods that promote long-term sustainability across environmental, economic, and social dimensions (Gomes, 2009).

The field addresses a wide range of problems, including:

- Environmental sustainability: climate modelling, energy forecasting, conservation planning, and natural resource management.
- Societal sustainability: equitable healthcare, accessible transportation, public policy design, and education planning.
- Economic sustainability: optimization of supply chains, reduction of operational waste, and development of circular economy systems.

What sets computational sustainability apart is its focus on modelling trade-offs, constraints, and long-term objectives in complex, interdependent systems. Problems in this domain often involve limited or noisy data, multiple stakeholders with competing interests, and high-stakes ethical or policy implications. As a result, the models developed must be not only accurate, but also fair, and robust.

This thesis is grounded in the principles of computational sustainability. Through the application of ML techniques to forecast, monitor, and optimize systems in energy, mobility, and healthcare, it contributes to a growing body of work that demonstrates how data-driven methods can enhance sustainability outcomes. Each study is rooted in a real-world problem and illustrates how ML can reduce environmental impact, improve well-being, or support preventive strategies.

Rather than treating these fields in isolation, this work adopts an integrated view positioning them under the umbrella of computational sustainability to identify shared challenges, methods, and opportunities. While the specific techniques used vary techniques, the unifying goal is to develop models that provide actionable insights for more sustainable systems.

1.5 Domains of Application

This thesis summarizes the research conducted during the Ph.D., emphasizing the application of ML across three sustainability-critical domains. It highlights the results obtained through the methodological approaches adopted, the research gaps addressed in each field, the motivations driving model development, and the potential impact of these contributions in real-world contexts, specifically in:

- *Energy*. Efficient management of energy is essential to both environmental and economic sustainability. This research applies ML to two forecasting challenges:
    - Energy consumption prediction in water infrastructure.
    - Hydropower production forecasting in water infrastructure.

  These forecasts can reduce energy waste, enhance renewable integration, and support smarter infrastructure operation.
- *Urban Mobility*. Transportation systems significantly influence greenhouse gas emissions, air quality, and urban liveability. This thesis explores ML applications in:
    - Traffic flow forecasting, using multi-source data to improve congestion management.
    - Cyclist safety risk prediction, to support safer infrastructure planning and promote active mobility.

  These models contribute to safer, more efficient, and more sustainable urban transportation.
- *Healthcare*. Preventive healthcare is crucial to social and economic sustainability. This research develops a ML–based approach for postural analysis, using low-cost imaging as an alternative to traditional, expensive diagnostic systems. The goal is

to make early detection of musculoskeletal issues more accessible, thereby improving health outcomes and reducing long-term care costs.

Each of these domains is addressed through peer-reviewed studies of which full technical specifications are available in the original publications. An overview of the developed papers to be discussed in this thesis is presented in Table 1.

Table 1 - Summary of the original papers reviewed in the thesis

Year	Title	Summary
2023	A Proactive Approach for the Sustainable Management of Water Distribution Systems (Di Grande et al., 2023b)	Proposed short-term forecasting models for water demand and pump energy consumption in water distribution systems (WDSs).
	A Machine Learning Approach for Hydroelectric Power Forecasting (Di Grande et al., 2023a)	Applied ML to hydropower generation forecasting in WDS-integrated plants.
	Proposal of an AI based approach for Urban Traffic Prediction from Mobility Data (Berlotti et al., 2023)	Proposed a two-level ML framework for 24-hour traffic flow forecasting using clustering and prediction models.
2024	Data Science for the Promotion of Sustainability in Smart Water Distribution Systems (Di Grande, Berlotti, Cavalieri, et al., 2024b)	Developed improved forecasting results for water demand and pump energy consumption in WDSs.
	Optimizing Planning Strategies: A Machine Learning Forecasting Model for Energy Aggregators and Hydropower Producers (Di Grande, Berlotti, Cavalieri, et al., 2024d)	Designed forecasting models for multi-step-ahead hydropower generation forecasting.
	A Machine Learning Approach to Forecasting Hydropower Generation (Di Grande, Berlotti, Cavalieri, et al., 2024a)	Explored hydropower generation forecasting at monthly and bi-weekly horizons, addressing data quality and imputation issues.

	Harnessing Multivariate AI to Enhance Hydropower Generation Forecasting (Di Grande, Berlotti, Cavalieri, et al., 2024c)	Introduced multivariate forecasting models combining hydropower data with weather and hydrological variables for medium-term prediction.
	AI-Powered Urban Mobility Analysis for Advanced Traffic Flow Forecasting (Di Grande, Berlotti, and Cavalieri, 2024)	Developed advanced traffic flow forecasting models using one-year datasets to capture daily, weekly, and yearly seasonal patterns.
	Proposal of a Machine Learning Approach for Traffic Flow Prediction (Berlotti et al., 2024)	Advanced traffic flow forecasting models through a full scientific validation.
	Enhancing Urban Traffic Management Through Machine Learning Prediction Models for Sensor-Less Roads (De Souza Oliveira et al., 2024)	Created prediction models combining sensor and floating car data (FCD) to estimate traffic flows on roads without monitoring infrastructure.
	Biomechanical Posture Analysis in Healthy Adults with Machine Learning: Applicability and Reliability (Roggio et al., 2024)	Applied ML-based pose estimation methods for static postural analysis, including Principal Component Analysis (PCA) and clustering to explore structural patterns.
2025	Data-Driven Prediction of High-Risk Situations for Cyclists Through Spatiotemporal Patterns and Environmental Conditions (Di Grande et al., 2025)	Proposed a classification framework to predict accident risk for cyclists using contextual spatiotemporal and environmental features.
In press.	AI-Driven Hydropower Forecasting: A Multivariate Machine Learning Approach (Di Grande et al., In press.)	Advanced medium-term hydropower forecasting by integrating climatic drivers with lagged production data.

Although the application areas differ, the underlying objective remains the same: to leverage ML in ways that enhance the efficiency, resilience, and sustainability of real-world systems.

1.6 Structure of the Thesis

The thesis is structured as follows:

- Chapter 2 – Machine Learning for Sustainable Energy Systems: Discusses the application of ML models for energy optimization and hydropower production in WDSs. It outlines the motivation, modelling approach, and implications for sustainable resource management.
- Chapter 3 – Machine Learning for Sustainable Urban Mobility: Discusses ML applications in traffic forecasting and cyclist safety. It highlights how predictive models can improve traffic efficiency and prevent accidents, contributing to safer and more sustainable urban environments.
- Chapter 4 – Machine Learning for Postural Analysis in Healthcare: Focuses on the use of ML for postural analysis. It describes a low-cost, accessible solution for postural screening. The chapter considers its accessibility and impact on preventive healthcare.
- Chapter 5 – Conclusions: Synthesizes the lessons learned across all domains, identifies cross-cutting challenges and opportunities, and outlines future directions for expanding the field of computational sustainability.

Chapter 2: Machine Learning for Sustainable Energy Systems

The efficient management of energy resources is essential for advancing sustainability across environmental, economic, and societal domains. Energy underpins nearly every aspect of modern life, from water supply and industrial production to mobility and healthcare (Ahmad et al., 2020). However, the global energy sector faces an unprecedented dual challenge: on the one hand, the continuing rise in demand driven by population growth, urbanization, and technological development; on the other hand, the urgent need to reduce greenhouse gas emissions and mitigate climate change (Aslam et al., 2021; Kerscher and Arboleya, 2022, 2022). These pressures make the optimization of energy systems not only a technical concern but also a critical requirement for meeting international climate targets and achieving the United Nations Sustainable Development Goals (SDGs), particularly SDG 7 (Affordable and Clean Energy) and SDG 13 (Climate Action) (Fuso Nerini et al., 2018).

A central component of this optimization lies in the ability to accurately forecast energy consumption and production. Reliable predictions allow utilities and policymakers to anticipate demand peaks, allocate resources efficiently, integrate renewable sources, and design adaptive operational strategies (Ahmad et al., 2020; Aslam et al., 2021). Forecasting is also essential for maintaining grid stability in increasingly complex energy landscapes, where variable renewable generation such as solar and wind introduces new layers of uncertainty (Kezunovic et al., 2020). Without accurate forecasts, the risk of inefficiency, higher operational costs, and system instability increases substantially (Hernandez-Matheus et al., 2022).

Within this broader energy context, WDSs represent a particularly relevant and often underexplored sector (Yi et al., 2022). Water and energy are deeply interlinked in what is known as the water–energy nexus. Vast amounts of energy are consumed to extract, treat, and distribute water, while water is equally critical for energy production through cooling processes, hydropower, and renewable integration (Sharif et al., 2019). Globally, it is estimated that water supply and distribution account for approximately 7–8% of total energy use, with pumping operations representing the largest share of this consumption (Sharif et al., 2019). In some regions, this figure is even higher, making WDSs one of the most energy-intensive public services. Conversely, WDSs also offer opportunities for renewable generation, for example, through hydropower plants embedded within the network. These plants, which have only been developed relatively recently (Sari et al., 2018), are designed to exploit excess hydraulic pressure in WDSs. Under normal operation, this pressure must be reduced to prevent damage to pipes and leakages, and is therefore usually dissipated through pressure-reducing valves as wasted energy. By replacing these devices with turbines, the same pressure reduction can be achieved while simultaneously generating renewable electricity, thereby transforming what would otherwise be an inefficiency into a valuable source of clean energy.

Despite their dual significance as both major consumers and potential producers of energy, WDSs have received comparatively less attention in the sustainability and ML literature than other sectors such as power grids or building energy systems. Nevertheless, they present distinctive challenges that make them an excellent context for testing and developing advanced forecasting methods. Their consumption patterns are driven not only by physical and hydrological factors but also by complex social and behavioural dynamics, such as variations in population demand, industrial activity, or even seasonal tourism flows (de Souza Groppo et al., 2019). Similarly, hydropower generation within WDSs is shaped by both natural variability (for example rainfall, temperature, and snowmelt) and operational constraints (for example water storage requirements, demand priorities, and regulatory frameworks) (Jung et al., 2021; Zhou et al., 2020). These characteristics create nonlinear, multiscale, and highly interdependent dynamics that traditional statistical models often fail to capture (Mosavi et al., 2019).

In this context, ML offers powerful tools to improve predictive performance and operational efficiency. By leveraging complex data, nonlinear patterns, and correlations across multiple variables, ML models can go beyond the limitations of traditional approaches and provide robust forecasts under uncertain and dynamic conditions (Mosavi et al., 2019). Importantly, the use of ML in WDSs not only enhances technical efficiency but also aligns with broader sustainability goals, including reducing wasted energy, lowering operational costs, supporting the integration of renewable generation, and contributing to climate resilience (Alhendi et al., 2022).

This chapter focuses on two interrelated research areas within sustainable energy systems, both specifically applied to WDSs:

- Forecasting of energy consumption, with particular emphasis on pump energy use and its relation to water demand patterns (Di Grande, Berlotti, Cavalieri, et al., 2024b; Di Grande et al., 2023b);
- Forecasting of hydropower generation, with attention to WDS-integrated plants that recover otherwise wasted hydraulic energy (Di Grande, Berlotti, Cavalieri, et al., 2024a, 2024d, 2024c; Di Grande et al., 2023a, In press.).

Both tasks are crucial for improving efficiency, reducing environmental impacts, and supporting long-term strategic planning in the water–energy sector. The analysis presented in this chapter is grounded in several original research studies conducted during the Ph.D. program (Di Grande, Berlotti, Cavalieri, et al., 2024b, 2024d, 2024a, 2024c; Di Grande et al., 2023b, 2023a, In press.), which collectively demonstrate how ML methods can be applied to concrete challenges in energy forecasting. The following sections provide a summary of the methods and results, while the original papers should be consulted for complete technical specifications.

Beyond presenting methodological results, the chapter also highlights the practical implications of these forecasting systems in real-world contexts. These include their integration into decision-support platforms for utilities, their role in supporting energy market participation by hydropower producers, and their potential contribution to sustainability monitoring under EU and UN directives (Fuso Nerini et al., 2018, 2018; Iria and Soares, 2023; Kerscher and Arboleya, 2022). By situating technical findings within their broader societal and environmental context, the chapter illustrates how advances in computational methods can directly support both local operational needs and global sustainability objectives.

2.1 Sustainable Energy Consumption

2.1.1 Background and Motivation

WDSs are among the most critical infrastructures for modern societies, ensuring the continuous and reliable supply of potable water to both urban and rural populations. Their importance extends beyond public health: access to clean and safe water underpins economic development, industrial activity, and social well-being (Di Grande et al., 2023b). For this reason, WDSs are often described as "lifeline systems," whose proper functioning is indispensable to human settlements and to achieving the SDGs (Alhendi et al., 2022; de Souza Groppo et al., 2019).

However, this essential service comes at a high energy cost. The operation of WDSs is energy-intensive, primarily due to pumping systems that must overcome hydraulic constraints such as elevation differences, frictional head losses, and required pressure levels (Di Grande et al., 2023b). In many systems, pumps account for the largest share of operational costs, sometimes representing the 95% of the total energy bill for a utility (Abdelsalam and Gabbar, 2021; Brás et al., 2025). Globally, it is estimated that 7–8% of total energy production is consumed by water production and distribution, with the majority of this energy used for distribution (Sharif et al., 2019). The majority of this energy still comes from fossil fuels, contributing significantly to greenhouse gas emissions and environmental degradation, while also exposing utilities to price volatility in energy markets (Alhendi et al., 2022).

The challenge is compounded by increasing demographic and climatic pressures. Population growth and urban expansion intensify demand for potable water, while climate variability introduces new uncertainties such as prolonged droughts, erratic rainfall, and higher evaporation rates (de Souza Groppo et al., 2019; Esen et al., 2020; Hussain et al., 2022; Leitão et al., 2019; Patil et al., 2022). These stressors force utilities to operate under tighter resource constraints while maintaining service quality (Adedeji et al., 2022; Mesalie et al., 2021). Moreover, aging infrastructure and water losses through leaks further exacerbate energy waste, creating a strong incentive for adopting innovative management approaches that can balance water reliability with energy efficiency (de Souza Groppo et al., 2019).

To respond to these pressures, modern WDSs are increasingly expected to meet sustainability objectives in addition to traditional performance requirements. This means not only guaranteeing continuous delivery of safe water with adequate pressure and flow, but doing so with minimal resource waste and reduced environmental impact (Adedeji et al., 2022; Bolognesi et al., 2014; Mesalie et al., 2021). Achieving such a balance requires moving away from purely reactive operations toward predictive and adaptive strategies supported by digital technologies (Alhendi et al., 2022).

A promising approach lies in optimization strategies such as advanced pump scheduling, which have been shown to reduce energy up to 25% when intelligently implemented (Luna et al., 2019). These savings are achieved by reducing unnecessary or inefficient pumping operations. This can be done, for instance, by shifting pump activity to off-peak tariff periods, when electricity prices are lower, or by preventing simultaneous operation of multiple pumps when storage tanks already contain sufficient reserves. However, the effectiveness of such strategies relies on the ability to anticipate future conditions. Accurate forecasts of both water demand and energy consumption are therefore essential, as they allow utilities to schedule pumping proactively rather than reactively, ensuring that energy efficiency is maximized without compromising service reliability (Alhendi et al., 2022).

Water demand forecasting plays a central role, as it enables utilities to anticipate consumption patterns across different temporal horizons. Short-term forecasts allow operators to adjust pumping schedules hour by hour, while medium- and long-term forecasts inform storage planning, network design, and infrastructure investments. Accurate predictions also support anomaly detection: deviations between observed and predicted demand may signal leakages, unauthorized consumption, or sensor malfunctions (de Souza Groppo et al., 2019).

Energy consumption forecasting complements this task by providing insights into how pumping activities will translate into energy requirements. Such forecasts help utilities in load planning, cost estimation under variable tariffs, and the detection of inefficiencies, such as excessive energy use caused by mechanical faults or suboptimal operating points (Yi et al., 2022). By combining water demand and energy consumption forecasting, utilities can manage the water–energy nexus more effectively, aligning operational decisions with both economic and environmental objectives (Alhendi et al., 2022).

Despite their importance, forecasting approaches traditionally applied in WDSs remain limited. Most rely on local models, where each time series (for example, demand or consumption at a given station) is forecasted independently. While conceptually simple, this approach becomes increasingly impractical as utilities expand sensor coverage, generating dozens or even hundreds of parallel time series. Moreover, local models fail to exploit the correlations that exist across zones, for instance between residential and industrial districts, or between upstream and downstream nodes in a distribution network (de Souza Groppo et al., 2019; Di Grande et al., 2023b; Montero-Manso and Hyndman, 2021). This results in fragmented knowledge, higher computational costs, and less accurate predictions.

To address these limitations, our work (Di Grande, Berlotti, Cavalieri, et al., 2024b; Di Grande et al., 2023b) advances the use of global forecasting models, where a single predictive model is trained on multiple time series simultaneously. Global models are more scalable, as they require fewer parameters relative to maintaining a large collection of local models, and more data-efficient, since they can leverage shared temporal structures across zones. For example, diurnal consumption cycles observed in residential districts can inform predictions in less-instrumented areas, while seasonal variations in one part of the network can provide transferable insights for another.

In addition, the present work (Di Grande, Berlotti, Cavalieri, et al., 2024b; Di Grande et al., 2023b) validates the use of N-BEATS (Oreshkin et al., 2020), a neural architecture originally developed for general time series forecasting but not previously applied to the WDS domain. N-BEATS is designed to capture both short- and long-term temporal patterns in a flexible and interpretable way, making it particularly suitable for handling the nonlinearities and multi-scale dependencies characteristic of WDS data. Furthermore, the proposed approach explores the ability of global models to generalize to unseen series, that is, to deliver accurate predictions even in zones not included in the training set. This is a critical innovation for practical deployment, as utilities often need to extend predictive capabilities to new or sparsely instrumented parts of their networks without retraining models from scratch.

By combining scalability, transferability, and advanced neural architectures, the innovations proposed in this research directly address current gaps in the literature and practical needs in the field. They provide a foundation for more efficient and sustainable resource management in complex water–energy systems, aligning with broader global goals to decarbonize infrastructure, improve resilience, and ensure equitable access to essential services.

2.1.2 Methods and Results

As part of this Ph.D. research, an ML framework was developed for short-term forecasting of water demand and pump energy consumption in a WDS, both being essential components of energy optimization. The work was validated in two peer-reviewed studies: “A Proactive Approach for the Sustainable Management of Water

Distribution Systems" (Di Grande et al., 2023b) and "Data Science for the Promotion of Sustainability in Smart Water Distribution Systems" (Di Grande, Berlotti, Cavalieri, et al., 2024b), both of which explored how predictive intelligence can support the sustainable and efficient operation of water infrastructure.

The dataset was generated through EPANET and WNTR (Klise et al., 2020), simulating hourly operation days in the Milano WDS. Twenty network zones were modelled, each characterized by distinct demand profiles, hydraulic structures, and pumping configurations. For every zone, time series included aggregated water demand and the corresponding pump energy consumption. The forecasting task involved jointly predicting water demand and energy consumption over the next 24 hours, with the objective of supporting operational decisions such as pump scheduling and energy load planning.

Several comparative experiments were conducted to assess the performance and robustness of the models. One important line of investigation was the contrast between local and global forecasting models. Local models were trained separately for each zone, while global models were trained on data from all zones simultaneously. The global N-BEATS model clearly outperformed the local approaches. In fact, training a single global model required less computational time than training 20 separate local models, while delivering almost the same accuracy.

A second set of experiments compared different deep learning architectures available in the Darts time series library (Herzen et al., 2023). Among recurrent and transformer-based designs, the N-BEATS model consistently provided the most accurate and stable results across evaluation metrics, confirming its suitability for structured time series such as those generated by water demand.

Another important trial examined whether a model trained exclusively on water flow data could be reused to forecast pump energy consumption, given the strong physical correlation between these variables. The results confirmed this possibility: a model trained on flowrate was able to deliver forecasts of energy use with accuracy comparable to that of models trained directly on energy data. This demonstrated the potential for cross-domain generalization in WDSs, where the physical coupling between variables allows predictive models to share structure across related tasks.

We also tested whether global models could generalize to previously unseen zones. In leave-one-out experiments, the global N-BEATS model was trained on 19 zones and evaluated on the excluded one. In most cases, the model was able to transfer knowledge successfully, delivering accurate predictions even without direct training data for that zone. Only in two instances did performance degrade substantially, where demand patterns diverged strongly from those observed elsewhere. This suggests that generalization improves as the diversity of the training set increases, reinforcing the importance of heterogeneous data.

Taken together, these trials demonstrate that global N-BEATS models are robust, scalable, and highly accurate for short-term forecasting of both water demand and pump energy use in complex distribution systems. Their ability to exploit multiple time series simultaneously, transfer knowledge across related variables, and generalize to new zones makes them well suited for deployment in smart water networks. Beyond advancing the state of the art in water infrastructure forecasting, this research shows how modern ML tools can directly improve operational efficiency, reduce energy consumption, and support the broader sustainability of critical urban systems.

2.1.3 Practical Implications

The forecasting framework developed in this research has a broad range of applications that extend from immediate operational benefits to long-term strategic planning. By enabling accurate short-term predictions of both water demand and pump energy consumption, the proposed models offer direct value for water utilities, energy planners, and sustainability stakeholders. They not only improve the operational efficiency of WDSs but also contribute to the wider transition toward data-driven infrastructure management.

One of the most impactful applications is in pump scheduling. In many WDSs, pumping stations include several pumps operating in parallel to meet variable demand. With forecasting models that provide a 24-hour outlook, operators can plan pump activity more intelligently, shutting down unnecessary pumps during periods of low demand, such as nighttime, and increasing output during peak hours. This reduces wasted energy, minimizes mechanical wear, and lowers costs associated with peak electricity tariffs. By anticipating demand fluctuations in this way, utilities can shift from rigid operating routines to adaptive, data-driven scheduling.

Forecasts of energy consumption provide additional advantages for load planning and cost estimation. Knowing in advance how much energy will be required allows utilities to anticipate operational costs under different tariff regimes, avoid spikes that trigger penalty charges, and schedule power-intensive operations during off-peak hours. In contexts where energy contracts are tied to demand windows or dynamic pricing, such foresight is particularly valuable, supporting financial as well as technical optimization.

Another important use of the forecasting framework lies in anomaly detection and preventive maintenance. Deviations between predicted and observed consumption often signal potential problems, such as leaks or bursts in the network, sensor malfunctions, or inefficiencies caused by pump degradation. By integrating forecasts into supervisory control and data acquisition systems or digital twins, anomalies can trigger automated alerts and inspections, enabling early intervention that reduces water losses and prevents energy waste.

Integration into digital platforms and dashboards provides further opportunities. Forecasts can be visualized alongside observed demand and energy use in real time, operators can simulate alternative strategies under predicted conditions, and automated alerts can be generated when deviations exceed defined thresholds. Because the models are lightweight and compatible with microservice architectures, they can be easily embedded into existing data acquisition systems or enterprise systems, extending their value without requiring major infrastructure changes.

The models also act as decision-support tools for resource allocation. In large or multi-zone systems, operators must constantly prioritize where to allocate resources such as maintenance teams, mobile pumps, or reservoir refilling. Forecasts help guide these decisions proactively: for example, if multiple zones are predicted to experience low demand simultaneously, maintenance can be scheduled in one of them without compromising overall supply.

A further dimension is the reinforcement of the water–energy nexus. In some systems, energy recovery devices such as hydropower turbines are embedded in pressure-reducing valves. Forecasting both water demand and pump energy use allows utilities to coordinate pumping strategies with opportunities for energy savings, renewable generation, or peak shaving across the utility network. This supports integrated sustainability strategies and creates synergies between water distribution and energy management.

Another strength of the forecasting framework lies in its scalability. Since global models are trained on multiple time series simultaneously, the same system can be extended to additional series that exhibit similar patterns, without the need to develop separate models for each case. This reduces maintenance costs, simplifies deployment, and makes predictive intelligence accessible even in data-scarce areas. Smaller utilities or rural districts, which often lack sufficient historical data, can still benefit from knowledge transferred from larger, better-instrumented systems.

Finally, the forecasting framework supports policy and sustainability objectives. By quantifying indicators such as energy intensity per cubic meter of water delivered, utilities can better report on efficiency and progress under EU and UN sustainability directives. Forecasting also enables scenario analysis for decarbonization strategies, demonstrating how improved operational practices can directly reduce emissions. In this way, predictive modelling strengthens the link between operational improvements at the plant level and broader sustainability goals at regional and international scales.

In summary, the ML models developed in this research provide a practical, scalable, and forward-looking solution for short-term operational planning in water utilities. By enabling predictive control of both water and energy use, the framework advances efficiency, reduces costs, enhances sustainability reporting, and supports the transformation of WDSs into data-driven infrastructures that are resilient, adaptive, and aligned with long-term climate and sustainability goals.

2.2 Energy Production Forecasting

2.2.1 Background and Motivation

Electricity generation from renewable sources is increasingly recognized as a cornerstone of global sustainability strategies, as societies seek to meet rising energy demands while reducing greenhouse gas emissions and mitigating climate change (Kerscher and Arboleya, 2022). Within this transition, hydropower occupies a unique and enduring position. It is not only one of the oldest and most mature renewable energy technologies but also one of the largest contributors to global renewable electricity production . Indeed, in 2024, renewable sources accounted for 46.9% of the European Union's net electricity production. Within this share, wind and hydropower together represented nearly two-thirds, contributing 39.1% and 29.9% respectively, while solar provided 22.4%. The remainder was supplied by combustible renewables (8.1%) and geothermal energy (0.5%) (*Electricity from Renewable Sources Reaches 47% in 2024*, 2025).

The strategic importance of hydropower derives from its operational versatility. As highlighted in (Silva and Castillo, 2021), hydropower plants can provide a stable supply of baseload electricity while also offering rapid regulation services through turbine flexibility and reservoir storage. This enables them to respond effectively to fluctuations in demand and to balance the intermittency of other renewable sources such as wind and solar. In many regions, hydropower plants are operated as multipurpose facilities, not only generating electricity but also contributing to irrigation, flood control, and drinking water supply. These multiple functions reinforce their role in stabilizing modern power grids and in supporting integrated water–energy management strategies.

Beyond electricity production, hydropower facilities frequently serve multiple purposes within integrated water–energy systems. Large reservoirs often play critical roles in irrigation, drinking water supply, and flood control, while smaller distributed units embedded in aqueducts and water distribution systems can convert hydraulic pressure surpluses into clean electricity (Sari et al., 2018). This multifunctional character places hydropower at the heart of the water–energy nexus, where resource management strategies must balance technical, environmental, and societal objectives simultaneously. In contexts such as Mediterranean regions, where water scarcity coincides with rising energy demands, the optimization of WDS-integrated hydropower facilities becomes a strategic priority for both sustainability and resilience (Di Grande et al., In press.).

The effective operation of hydropower plants, however, depends heavily on the ability to forecast future energy generation with high accuracy. Forecasting supports a wide range of decision-making tasks across temporal scales. On short horizons, accurate predictions enable operators to optimize turbine scheduling, reduce unnecessary water releases, and coordinate hydropower with other renewable resources. On medium horizons, forecasts inform reservoir management, helping to balance competing demands for energy generation, irrigation, and ecological flow requirements. On longer horizons, predictive intelligence underpins energy market participation, allowing producers and aggregators to align contractual commitments with expected production, thereby reducing imbalance penalties and improving profitability (Ahmad et al., 2020; Hernandez-Matheus et al., 2022; Q. Wang et al., 2015). In all cases, reliable forecasting reduces risks, lowers operational costs, and supports the integration of hydropower into broader sustainability frameworks.

Conventional hydropower forecasting methods have mainly addressed run-of-river and reservoir-based plants, whereas WDS-integrated hydropower systems have received no attention. Historically, these traditional approaches have relied on physical hydrological models and statistical time-series analyses. Physical models simulate system behaviour by representing processes such as rainfall–runoff dynamics, catchment hydrology, or reservoir hydraulics. While they can provide valuable insights into the underlying mechanisms of water flow and storage, their effectiveness depends on the availability of detailed system parameters, which are often incomplete, outdated, or difficult to measure. Moreover, their calibration can be time-consuming and computationally expensive (Ampas et al., 2025). On the other hand, statistical approaches such as autoregressive integrated moving average (ARIMA) or seasonal decomposition methods have been widely used for short-term forecasts based on historical generation data. These models are easier to implement but frequently struggle with nonlinearities, abrupt changes in conditions, and non-stationary time series, all of which are characteristic of hydropower production influenced by weather variability and operational constraints (Polprasert et al., 2021).

In recent years, ML methods have emerged as a promising alternative to complement and, in some cases, surpass these traditional techniques. Unlike physical or purely statistical models, ML approaches are data-driven: they learn directly from historical records and contextual variables, without requiring explicit knowledge of system parameters. This makes them particularly suitable for complex, dynamic, and nonlinear systems such as WDS-integrated hydropower plants. By incorporating diverse sources of information, including inflows, reservoir head levels, turbine characteristics, meteorological forecasts, and even upstream water demand patterns, ML models can capture intricate dependencies that are difficult to represent with conventional methods (Barzola-Monteses et al., 2022). Moreover, they can adapt to evolving conditions, such as changes in climate patterns, land use, or operational strategies, ensuring their continued relevance in dynamic environments.

The flexibility of ML also enables the integration of multiple temporal and spatial scales within a single framework. For example, short-term forecasts can be generated to support daily turbine scheduling, while medium-term predictions may incorporate lagged meteorological variables to anticipate seasonal inflow variations. Multivariate models can capture interactions between hydropower generation and external drivers such as rainfall, temperature, or snowmelt, offering predictive horizons that extend weeks or even months into the future. In this way, ML not only enhances forecasting accuracy but also broadens the scope of actionable insights for operators, utilities, and energy markets.

As stated earlier, traditional approaches to hydropower forecasting have primarily focused on run-of-river or reservoir-based plants, while hydropower units integrated within WDSs have received no attention. These WDS-integrated plants differ substantially from traditional run-of-river or reservoir-based systems because their production patterns are shaped by a dual dependency. On one side, natural drivers such as rainfall, temperature, and snowmelt determine how much water is collected and available within the network. On the other, anthropogenic factors, mainly the water consumption of municipalities served by the system, directly affect how much of that water remains available for power generation. As municipal demand fluctuates across days and seasons, the residual flow reaching the turbines changes accordingly, making the generation profile far more dynamic and complex than that of conventional hydropower plants.

Taken together, these advances underscore the potential of ML to play a transformative role in hydropower forecasting. By overcoming the limitations of purely physical or statistical approaches, our research provides the first contribution to the forecasting of WDS-integrated hydropower plants, a largely unexplored field characterized by complex interactions between natural and human-driven processes. Data-driven methods developed in this work deliver robust and scalable predictive intelligence, enabling hydropower to fulfil its strategic role as both a backbone of renewable energy systems and a key enabler of the global transition toward sustainability.

2.2.2 Methods and Results

As part of the research conducted during the Ph.D., we developed ML–based forecasting models for hydropower generation in WDS-integrated plants. This work was carried out and validated in five peer-reviewed studies: “A Machine Learning Approach for Hydroelectric Power Forecasting” (Di Grande et al., 2023a), “A Machine Learning Approach to Forecasting Hydropower Generation” (Di Grande, Berlotti, Cavalieri, et al., 2024a), “Optimizing Planning Strategies: A Machine Learning Forecasting Model for Energy Aggregators and Hydropower Producers” (Di Grande, Berlotti, Cavalieri, et al., 2024d), “Harnessing Multivariate AI to Enhance Hydropower Generation Forecasting” (Di Grande, Berlotti, Cavalieri, et al., 2024c), and “AI-Driven Hydropower Forecasting: A Multivariate Machine Learning Approach” (Di Grande et al., In press.). The studies focused on proposing forecasting models for hydropower generation in WDS-integrated plants, with such forecasts crucial for various purposes, including capacity planning, energy management, decision-making, and anomaly detection.

The dataset used in this research was collected from the Alcantara 1 Hydroelectric Plant, located in Taormina, Sicily, and operated by Siciliacque S.p.A. (*Siciliacque*, 2025), the public company responsible for the island’s large-scale drinking water supply. Siciliacque manages an extensive network of dams, wells, pumping and treatment stations, aqueducts, and renewable energy facilities, with the dual objective of guaranteeing reliable water distribution and improving the efficiency of energy use across the system. Within this context, Alcantara 1 represents a particularly interesting case study. Unlike conventional river-based hydropower stations, this facility is integrated directly into a WDS, making its operation dependent not only on natural hydrological factors but also on the daily and seasonal variability of municipal water demand. The aqueduct that feeds the plant extends for about 65 km and maintains a continuous flow of approximately 600 liters per second. To manage the excess pressure that would otherwise threaten the integrity of the pipelines, Siciliacque installed Pelton turbines capable of converting these hydraulic jumps into electricity. The choice of vertical-axis turbines with three jets was well suited to the characteristics of the system, where relatively modest flows combine with high heads. In this way, a risk that traditionally had to be dissipated through tanks and valves is now transformed into a source of renewable energy.

The Alcantara 1 hydroelectric plant has a rated power of about 1.1 MW, with all electricity injected into the medium-voltage grid and incentivized under national tariff schemes. However, its primary function remains embedded within the broader WDS. Water is collected from sources on the slopes of Mount Etna and first allocated to satisfy the upstream needs of municipalities. Only the residual flow, after these demands are met, is diverted through the turbines to generate electricity, before the same water continues downstream to supply additional municipalities along the Ionian side of Messina. Consequently, the inflow to the plant, and thus its electricity production, depends on a dual dependency: on the one hand, natural factors such as snowmelt, rainfall, and seasonal temperature fluctuations determine the volume of water collected; on the other hand, municipal consumption patterns shape how much of that volume remains available for hydropower generation. These consumption patterns themselves vary across daily, weekly, and seasonal cycles and are also influenced by weather conditions. The result is a highly dynamic and non-linear generation profile, markedly different from conventional river-based or reservoir-based hydropower plants. This unique interplay between natural inflows and societal demands makes Alcantara 1 an ideal case for testing advanced forecasting techniques.

Siciliacque's commitment to energy efficiency further enhances the relevance of this dataset. The company was among the first in Italy to obtain ISO 50001 (ISO 50001, 2018) certification for energy management and has invested heavily in measures to reduce costs and environmental impact, including the installation of high-efficiency motors, variable-speed drives for pumps, photovoltaic fields, and small hydro units. Between 2018 and 2022, five hydropower facilities were commissioned on the island's aqueducts, yielding significant economic and environmental savings. Alcantara 1 plays a central role in this strategy, not only generating renewable energy but also providing valuable operational data. Since 2019, Siciliacque has systematically recorded detailed measurements of hydroelectric output, hydraulic conditions, and meteorological variables. These data are used internally for planning and financial evaluation, but also for identifying anomalies in plant performance and tracking progress toward renewable energy goals. In this thesis, the dataset forms the basis for the development and testing of ML forecasting models, enabling a systematic exploration of how predictive intelligence can support the sustainable operation of WDS-integrated hydropower systems.

The first distinction among the trials conducted concerns the development of univariate versus multivariate ML models. In the univariate approach, forecasts are generated using only a single variable, namely the target variable, whereas in the multivariate approach, multiple variables are incorporated to improve predictive accuracy.

The starting point was the development of monthly univariate forecasting models, the description and results of which are presented in (Di Grande et al., 2023a), using hydropower data from January 2019 to May 2023. Statistical analysis of the 5-minute flow measurements revealed a clear seasonal pattern, with fluctuations repeating consistently across the year. This finding was validated by domain experts at Siciliacque, who confirmed that the observed dynamics reflected the actual hydrological behaviour of the system. Because the variability was driven primarily by seasonal rather than short-term fluctuations, the high-frequency 5-minute data were aggregated into monthly averages. This transformation emphasized the annual cycle and provided a more appropriate basis for developing one-step-ahead forecasting models at the monthly scale. Several algorithms were evaluated, with ARIMA serving as the statistical benchmark and N-BEATS and Temporal Convolutional Networks (TCN) representing more advanced ML approaches. To preserve temporal dependencies, walk-forward validation was adopted instead of traditional splits. Results showed that TCN captured the seasonal structure most effectively, clearly outperforming ARIMA and N-BEATS, and demonstrating the value of deep learning for this forecasting task.

Once the feasibility of univariate monthly forecasting had been established, the next step was to extend the approach toward multi-step forecasting (Di Grande, Berlotti, Cavalieri, et al., 2024d). While (Di Grande et al., 2023a) addressed one-step-ahead prediction, (Di Grande, Berlotti, Cavalieri, et al., 2024d) explored whether models could produce forecasts for several months simultaneously, an ability essential for energy aggregators and hydropower producers who need multi-month planning for market bidding and resource allocation. The dataset, again aggregated from 5-minute records, was modelled with ARIMA, N-BEATS, and TCN. The results reinforced the earlier findings: TCN consistently delivered the most accurate and stable forecasts, followed by N-BEATS, while ARIMA lagged behind. This confirmed that convolutional models are particularly well-suited to capture the seasonal and nonlinear structures that characterize WDS-integrated hydropower production. At this stage, the research established that univariate ML could support both short- and medium-horizon planning, but also revealed limitations when data irregularities and finer-scale fluctuations came into play.

These limitations motivated the next study (Di Grande, Berlotti, Cavalieri, et al., 2024a), which introduced shorter time intervals, data quality considerations and other ML models, with an extended dataset until November 2023 so allowing for a full year to be used as validation set. Here, forecasts were developed at both monthly and bi-weekly horizons, with the goal of capturing intra-month variability that monthly aggregation might smooth out. The research also addressed the practical challenge of missing data, comparing as techniques the linear interpolation with seasonal-trend decomposition using LOESS (STL). Experiments confirmed that STL imputation was superior, as it better preserved seasonal cycles, thereby reducing forecast errors. In terms of algorithms, Random Forest (RF) performed best at the monthly level, handling smaller aggregated datasets effectively, whereas TCN excelled at the bi-weekly scale, leveraging its ability to model long-range temporal dependencies. An interesting outcome emerged when the models were trained on data aggregated at a two-week resolution and then applied to predict monthly values using a two-step-ahead forecasting strategy. In this setup, the model first generated predictions for consecutive two-week periods, which were then combined to obtain monthly forecasts. Surprisingly, this indirect approach produced more accurate results than models trained directly on monthly data. The improvement can be attributed to the fact that two-week data preserve finer-grained fluctuations within each month, enabling the model to capture short-term variations that would otherwise be smoothed out in monthly aggregates. As a result, forecasts based on higher-resolution inputs provided richer information, leading to more reliable predictions at the coarser monthly scale. This stage of the research added both robustness and granularity to the forecasting pipeline.

With these foundations in place, the research expanded toward multivariate forecasting (Di Grande, Berlotti, Cavalieri, et al., 2024c), motivated by the fact that Alcantara 1 production depends not only on its past history but also on exogenous climatic and hydrological drivers. To reflect this, the dataset was enriched with meteorological variables such as temperature, solar radiation, humidity, wind speed, and hydrometer levels, collected across the hydropower plant. The original 5-minute hydropower records were aggregated to daily values, a deliberate choice aimed at smoothing high-frequency noise and aligning the data with the temporal resolution of available weather inputs, while still retaining the seasonal and weather-driven signals essential for medium-term forecasting. This choice was also motivated by the need to test whether reliable forecasting performance could be achieved when working at a higher precision, since daily aggregation preserves more fine-grained information compared to monthly averages. By doing so, the study not only harmonized hydropower and meteorological datasets but also explored the potential of daily resolution as a basis for accurate long-horizon predictions. To design a realistic forecasting horizon, we first performed a correlation analysis between hydropower production and lagged weather and hydrological variables over a window of 1–180 days. The correlation analysis revealed that the hydrometer level lagged by 101 days exhibited the strongest relationship with hydropower output ($r > 0.9$), making it the closest and most influential predictor in temporal terms. This result guided the choice of a 100-day forecasting horizon: by setting the horizon just below the strongest correlated lag, the model was able to incorporate the hydrometer information together with other variables that displayed significant but more distant correlations, such as temperature (lagged by 127 days). In practice, this approach ensured that each predictor was included at its most informative lag while maintaining a consistent and operationally meaningful forecasting horizon. Thus, the 100-day horizon represents not only a methodological compromise but also a theoretically grounded choice that balances correlation strength, temporal proximity, and practical applicability for hydropower forecasting. This finding directly motivated the choice of a 100-day ahead forecasting task, designed to test whether ML models could effectively capture these long-lag dependencies. Models were then developed using Linear Regression (LR), RF, and multilayer perceptron (MLP). Results demonstrated that while LR captured general seasonal tendencies, it failed to account for nonlinearities and lag effects. Both RF and MLP achieved considerably lower errors, with RF outperforming MLP across all tested metrics. To evaluate the added value of external features, univariate RF models were trained on lagged hydropower alone; these consistently performed worse than the multivariate versions, highlighting that weather and hydrological inputs are indispensable for reliable medium-term forecasting in WDS-integrated hydropower systems.

Building on the promising results of the fourth study, the fifth contribution (Di Grande et al., In press.) advanced the forecasting framework in two key directions: first, by testing model generalizability on a longer and hydrologically more diverse dataset, and second, by embedding internal system memory into the feature set. A new dataset was collected from the Alcantara 1 hydroelectric facility, extending the temporal coverage up to November 2024 and thus capturing a wider range of climatic and operational variability. Compared to earlier work, which was constrained to January 2022–June 2023, this expanded dataset comprised 901 consecutive daily observations, of which 382 were used for training (up to June 2023) and 519 were reserved exclusively for testing (July 2023–November 2024). This separation enabled a temporal holdout validation, offering a much more realistic evaluation than the internal train–validation splits employed in prior studies. Temporal holdout mirrors real-world deployment, where models trained on past data must operate under unseen future conditions potentially influenced by anomalies such as droughts or shifts in consumption patterns.

Feature engineering followed the earlier methodology: hydropower production records were aggregated to daily resolution, and exogenous meteorological drivers (including air temperature, solar radiation, wind speed, humidity, and hydrometer levels) were enriched with lagged versions determined through correlation analysis. In this extended work, a new endogenous feature was introduced: hydropower production lagged by 101 days. This variable was specifically chosen because, for a 100-day forecast horizon, only lags beyond 100 days are valid predictors, and lag 101 exhibited the highest autocorrelation with future output. By incorporating this lagged production feature, the model could embed internal system memory, capturing persistent operational cycles and latent seasonality that external climatic variables alone could not fully explain.

The experimental results confirmed the superiority of RF over the MLP. On the unseen test set, RF consistently outperformed MLP, achieving lower errors across all metrics (for instance, Symmetric Mean Absolute Percentage Error (SMAPE) 18.6 versus 29.8, Mean Absolute Percentage Error (MAPE) 22.4 versus 36.9, Root Mean Square Error (RMSE) 117 versus 153, and Normalized Root Mean Square Error (NRMSE) 0.2 versus 0.3). When the lagged hydropower feature was added, the RF model's performance improved substantially, with SMAPE dropping to 11.2, MAPE to 13.0, RMSE to 88.9, and NRMSE to 0.1. This represented error reductions of over 40% compared to the original feature set, demonstrating the value of embedding system memory. The forecasting results of the optimized RF model on the training and test sets are shown in Figure 1, where predictions align closely with observed production.

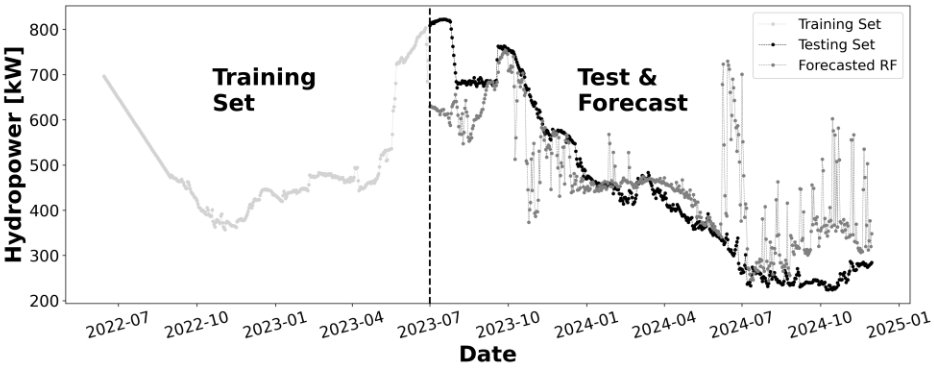

Figure 1 - Actual and forecasted values by the best RF model (Di Grande et al., In press.).

By contrast, the corresponding forecasts produced by the MLP model, reported in Figure 2, show larger deviations from the actual series.

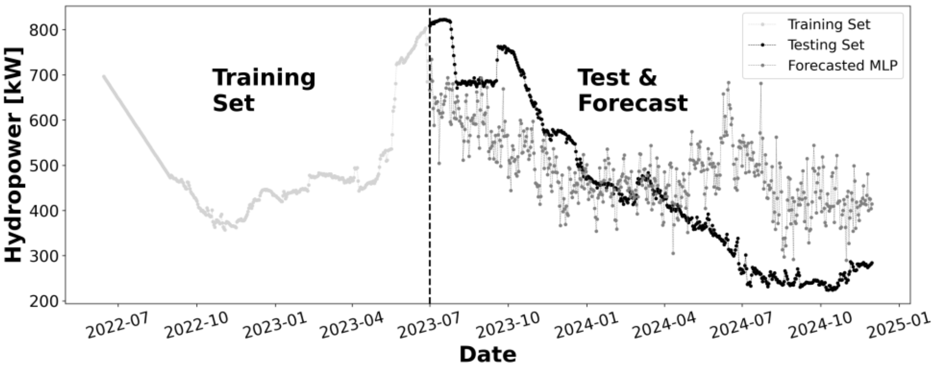

Figure 2 - Actual and forecasted values by the best RF model (Di Grande et al., In press.).

The advantages of the enhanced RF model with the lagged hydropower feature are illustrated in Figure 3, which highlights how the predictions better reproduce seasonal fluctuations and peaks (see ellipse 1 in Figure 3).

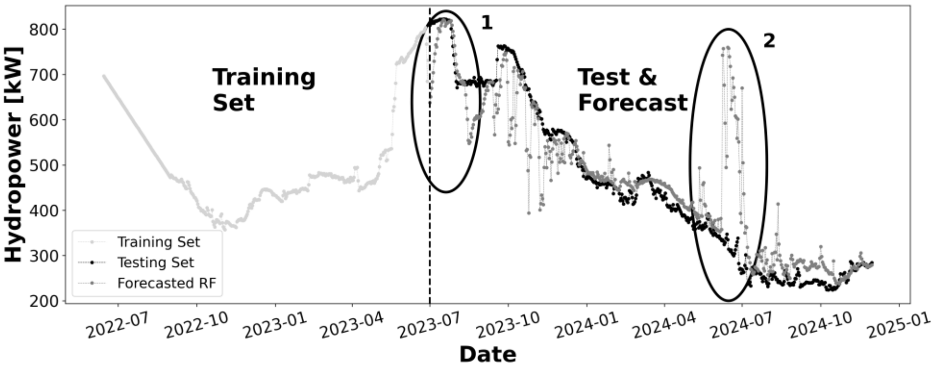

Figure 3 - Actual and forecasted values by the best RF model with hydropower variable (Di Grande et al., In press.).

To provide further context, Figure 4 reports the historical series of hydropower production from 2019 to 2024, showing how the year 2024 diverged from typical seasonal peaks due to exceptional drought conditions.

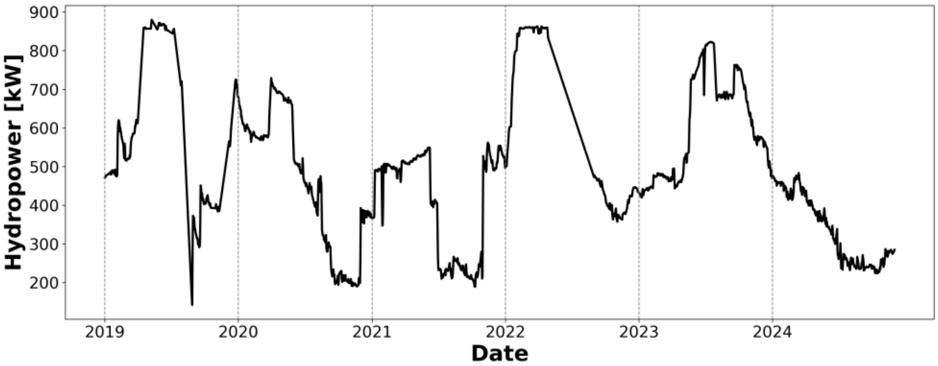

Figure 4 - Historical hydropower production pattern from 2019 to 2024 (Di Grande et al., In press.).

In June 2024, both the baseline and enhanced models predicted a positive seasonal peak in production, consistent with past behaviour, but this peak failed to materialize. Severe drought, coupled with high summer temperatures and reduced snow accumulation on Mount Etna, disrupted the usual inflow cycles. Water availability was further constrained by municipal consumption priorities and regulatory restrictions, leaving less surplus for electricity generation. While the models partially adapted, the unprecedented hydrological stress underscored some limits of the model in case of extreme anomalies (see ellipse 2 of Figure 3).

Overall, this study validated that RF remains the most robust algorithm for medium-term hydropower forecasting in WDS-integrated plants. More importantly, it demonstrated that incorporating lagged production as a proxy for system memory significantly enhances predictive stability and accuracy. By combining exogenous climatic drivers with endogenous historical signals, the framework delivers forecasts that are not only accurate but also resilient to the complexities of WDS-integrated hydropower dynamics.

These results not only advance methodological understanding but also demonstrate the practical implementation and value of predictive intelligence for sustainable hydropower management.

2.2.3 Practical Implications

The practical value of the hydropower forecasting studies developed in this Ph.D. lies in their ability to transform raw production data into actionable intelligence for utilities, energy aggregators, and sustainability planning. Each stage of the research contributed complementary insights, gradually expanding the scope of applications from plant-level operations to long-term market participation and strategic decision-making.

At the most fundamental level, the forecasting models enable plant-level operational improvements. Accurate short- and medium-term forecasts allow operators like Siciliacque to optimize the scheduling of turbines, better align production with expected inflows, and detect anomalies when predicted generation diverges from actual output. Such anomalies may signal inefficiencies or mechanical issues, prompting preventive maintenance before failures occur. By integrating forecasts into routine plant management, utilities can reduce both operational costs and environmental impacts, while ensuring the reliable delivery of renewable energy.

Beyond internal operations, forecasting becomes a key enabler of energy and financial planning. For prosumers operating within WDSs, such as Siciliacque, generation forecasts support revenue estimation from energy sales to the grid, improving budget planning and cost recovery. This predictive capability also aids in investment decisions, as long-term generation profiles help evaluate the economic viability of new hydropower facilities or upgrades to existing infrastructure.

The research also demonstrated the relevance of forecasting for energy aggregators and market strategies. In deregulated energy systems, where multiple renewable producers participate in electricity markets, accurate multi-step forecasts are critical for bidding strategies and risk reduction. By extending forecasting horizons to several months and showing that advanced models such as TCN outperform simpler baselines, this work directly supports the integration of hydropower into aggregated renewable portfolios. Aggregators can use these models to optimize trading decisions, balance renewable availability with demand, and improve the stability of their portfolios, thereby strengthening the role of hydropower in mixed-energy systems.

A further set of applications arises from the use of multivariate forecasting, which integrates climatic and hydrological drivers into the models. The ability to predict hydropower production up to 100 days ahead based on weather variables has significant operational and strategic implications. For utilities, this horizon provides the opportunity to schedule maintenance activities during predicted low-generation periods, minimizing lost revenue. For aggregators, it enables long-term market participation and hedging strategies, reducing exposure to variability in renewable supply. More broadly, the integration of weather and hydrological variables allows forecasts to reflect climate variability, making them a valuable tool for sustainability reporting and adaptation planning.

Finally, the advanced work on model generalizability and system memory represents a critical step toward real-world deployment. By moving from walk-forward validation to temporal holdout testing on entirely future data, the models were evaluated under conditions that mirror actual operational use. This ensures that forecasts remain robust even when hydrological dynamics shift due to seasonal anomalies or climate change. The incorporation of lagged production values further enhances reliability by encoding persistent seasonal cycles directly into the model. These innovations provide utilities and aggregators with forecasting tools that are not only as accurate as possible but also resilient to unexpected environmental changes, thereby strengthening their role in long-term energy planning and investment strategies.

Taken together, the practical implications of this research extend across multiple scales. At the plant level, forecasts support daily operations, preventive maintenance, and efficiency improvements. At the organizational level, they enable revenue planning, cost reduction, and informed investment. At the system level, they provide the predictive foundation for market bidding, portfolio optimization, and climate adaptation. By progressively addressing challenges from short-term variability to long-term generalizability, the studies demonstrate that ML–based hydropower forecasting is not only a methodological advancement but also a strategic asset for achieving sustainable energy management.

Having explored energy and water systems as a crucial component of sustainability, the thesis now shifts focus to urban mobility, another sector where predictive ML can yield both efficiency and safety gains.

Chapter 3: Machine Learning for Sustainable Urban Mobility

Transportation systems are indispensable to the functioning of modern cities, enabling the circulation of people, goods, and services that sustain economic activity and social life. At the same time, they represent one of the largest sources of unsustainability in contemporary societies. Road transport alone accounts for one quarter of global CO_2 emissions, making it a critical driver of climate change (*Greenhouse Gas Emissions from Energy Data Explorer – Data Tools*, 2024). In addition to carbon impacts, road vehicles contribute substantially to local air pollution, releasing nitrogen oxides, particulate matter, and other pollutants that directly affect public health (Doğan Güzel and Alp, 2024; EEA, 2019; European Environment Agency, 2016). Urban transport is also a major source of noise pollution, which has been linked to stress, sleep disruption, and cardiovascular risks (EEA, 2025).

The economic costs of congestion add another layer of urgency. Inefficient travel networks are estimated to cost EU Member State economies approximately €110 billion annually (European Court of Auditors, 2019). The social dimension is equally pressing: beyond time lost in traffic, road systems contribute to unequal accessibility, with disadvantaged groups often facing longer commutes and reduced mobility options (Liao et al., 2025). Moreover, traffic environments remain inherently dangerous. Vulnerable road users such as cyclists and pedestrians bear a disproportionate share of accident risks, with collisions frequently resulting in severe injuries or fatalities (WHO, 2023). Indeed, sustainable modes such as cycling remain underutilized, despite their well-documented benefits for health, the environment, and urban liveability. Cycling produces zero direct emissions, promotes physical activity, reduces congestion, and requires less infrastructure than motorized alternatives. Yet safety concerns remain the most significant barrier to widespread adoption. Infrastructure improvements such as bike lanes and safer intersections have mitigated some risks, but persistent safety concerns limit modal shift (Asgarzadeh et al., 2017; Reynolds et al., 2009). Ensuring efficient traffic flow while protecting cyclists thus emerges as a dual priority for sustainable mobility systems.

Traditional approaches to traffic management have primarily relied on infrastructure-based monitoring systems such as inductive loops, video cameras, and radar sensors. While these technologies provide valuable local data, they suffer from limitations: high installation and maintenance costs, restricted spatial coverage, and vulnerability to adverse weather (Antoniou et al., 2011; Deng et al., 2025). As a result, operators are often forced into interventions only once congestion has already formed.

Recent advances in data availability and computational techniques create new opportunities to address these challenges. ML provides powerful tools to move beyond the limitations of traditional monitoring by exploiting heterogeneous data sources and uncovering patterns that classical statistical methods often fail to capture (Rasaizadi et al., 2021). Urban mobility now generates vast amounts of data, from fixed sensors and FCD to public transport records, accident databases, and contextual information such as weather (Gheorghe and Soica, 2025; Zhang et al., 2025). By combining these sources, ML models can learn nonlinear, spatiotemporal relationships across the mobility system, enabling more accurate forecasting and risk prediction.

In the context of traffic flow, ML forecasting supports a shift from reactive to proactive management. Accurate short-term predictions allow operators to anticipate congestion before it fully develops, enabling dynamic signal control, adaptive rerouting, and other mitigation strategies (Moraga et al., 2025; Skoropad et al., 2025). Longer-horizon forecasts inform infrastructure planning, logistics optimization, and policy design, helping cities adapt to both daily variations and long-term shifts in demand (Mystakidis et al., 2025).

Equally important is the potential of ML to improve cycling safety. By analysing accident records alongside contextual variables such as lighting, weather, and road geometry, predictive models can estimate the likelihood of high-risk conditions for cyclists in space and time (Bassani et al., 2020; Ding et al., 2024; Zhu, 2021). This enables preventive interventions, ranging from targeted infrastructure investments to real-time warnings. By reducing both perceived and actual risks, ML contributes to making cycling a safer, more attractive, and sustainable mode of urban transport.

The research presented in this chapter explores these opportunities along two complementary lines. The first focuses on traffic flow forecasting, addressing challenges such as multi-horizon prediction, multi-source time series integration, and sensor-less network coverage (Berlotti et al., 2023, 2024; De Souza Oliveira et al., 2024; Di Grande, Berlotti, and Cavalieri, 2024). The second focuses on cycling safety, developing predictive frameworks that identify spatiotemporal high-risk conditions for cyclists based on accident data and environmental variables (Di Grande et al., 2025). Together, these contributions demonstrate how ML can transform urban mobility systems, making them not only more efficient but also safer, more inclusive, and more sustainable.

The following sections provide a summary of the main methods and results; the original papers should be consulted for the complete technical specifications.

3.1 Car Traffic Forecasting

3.1.1 Background and Motivation

Congestion is a pervasive challenge for urban mobility systems worldwide, with wide-ranging consequences for sustainability, quality of life, and economic competitiveness. It is not merely an inconvenience; congestion represents a systemic inefficiency with profound environmental, economic, and social costs.

Mitigating congestion requires more than simply expanding road capacity or adding new infrastructure, approaches that have historically proven unsustainable (Anciaes et al., 2025). Instead, modern strategies increasingly emphasize predictive and adaptive traffic management. Anticipating traffic flow before it develops allows operators to implement proactive interventions, such as dynamic signal control, adaptive rerouting, congestion pricing, or temporary restrictions on specific road segments (Gheorghe and Soica, 2025). In this sense, forecasting becomes a central capability: only by predicting future traffic states can operators move from reactive to proactive management and reduce congestion in a way that aligns with sustainability goals.

Accurate traffic flow forecasting is therefore recognized as one of the most effective solutions for containing congestion. As said before, traditional approaches to monitoring rely primarily on fixed infrastructure such as inductive loops, radar counters, or video cameras, which provide valuable localized data on vehicle counts and flows. However, while useful, these systems remain insufficient for understanding city-wide mobility dynamics. Their coverage is often limited to major intersections or arterial roads, leaving large portions of the network unobserved. Furthermore, their data are inherently reactive, reporting traffic conditions after they occur rather than anticipating them. In growing metropolitan areas, where traffic patterns are influenced by multiple interacting factors such limited data prove inadequate.

ML forecasting models address these gaps by offering a means to integrate diverse data sources, capture complex spatiotemporal dependencies, and transfer predictive knowledge to unsensed locations. By learning from both sensor-based traffic counts and complementary data streams such as FCD from navigation providers (e.g., TomTom, Google Maps), weather conditions, and historical patterns, ML models can generalize beyond instrumented road segments and provide forecasts across entire networks (Xing et al., 2022). This capacity to generate predictive intelligence even for roads without direct monitoring is especially important for medium-sized cities, which often lack the resources for dense sensor deployments but face mobility challenges comparable to larger metropolitan areas.

The research presented in this thesis applied these ideas to the case of Catania, a city that exemplifies the need for effective traffic flow forecasting. Catania is situated on the eastern coast of Sicily and forms the core of a larger metropolitan area that includes the main municipality and 26 surrounding urban centres. The metropolitan region is characterized by a highly centralized structure: most of the economic, educational, and cultural activities are concentrated in the main city, which attracts substantial daily commuting flows from the surrounding municipalities.

As a result, the transport system of Catania is heavily dependent on road traffic, with private vehicles accounting for the majority of mobility demand. Public transport options are limited, and despite recent efforts to expand metro and bus services, the modal split remains skewed toward individual motorization. This dependency has led to a steady increase in the number of vehicles circulating in the urban area, exacerbating congestion on infrastructure that is already operating near its maximum capacity.

The consequences are felt daily. Central areas of the city experience recurrent congestion during peak hours, creating bottlenecks that ripple outward to peripheral districts. The impact is not limited to delays: elevated vehicle density has contributed to a rise in air, noise, and visual pollution, degrading the urban environment and negatively affecting public health.

Moreover, climatic conditions further complicate mobility management. The hot Mediterranean summers increase reliance on private vehicles for comfort and speed, while occasional heavy rainfall events disrupt traffic flows by reducing road capacity and safety. Seasonal tourism adds additional pressures, introducing fluctuating demand that varies significantly between summer and winter months. These multiple drivers converge to make Catania a particularly challenging environment for mobility planning, where predictive traffic flow models are not only useful but essential.

In this context, ML forecasting becomes a strategic enabler of sustainable urban mobility. By anticipating demand fluctuations, and extending predictive intelligence to sensor-less roads, ML models provide the foundation for a more adaptive and resilient transport system. The ability to forecast traffic flows over different horizons also opens the door to applications ranging from short-term traffic management to long-term infrastructure planning and policy evaluation.

The Ph.D. research conducted on this case study progressed through four interrelated studies, each extending the capabilities of predictive models and addressing increasingly complex challenges: “Proposal of an AI based approach for Urban Traffic Prediction from Mobility Data” (Berlotti et al., 2023), “AI-Powered Urban Mobility Analysis for Advanced Traffic Flow Forecasting” (Di Grande, Berlotti, and Cavalieri, 2024), “Proposal of a Machine Learning Approach for Traffic Flow Prediction” (Berlotti et al., 2024), and “Enhancing Urban Traffic Management Through Machine Learning Prediction Models for Sensor-Less Roads” (De Souza Oliveira et al., 2024).

Together, these contributions illustrate a progressive research trajectory: from localized forecasting to city-wide predictive intelligence, from short-term horizons to multi-scale prediction, and from sensor-dependent methods to approaches that extend beyond direct observation. In doing so, the research demonstrates not only the technical feasibility of applying ML to urban traffic forecasting but also its strategic importance for advancing sustainable mobility in cities such as Catania.

3.1.2 Methods and Results

The first study (Berlotti et al., 2023) aimed to address the problem of short-term traffic forecasting on roads without sensors by proposing a two-level ML approach. The aim of the study was to develop a ML model to forecast the traffic from the next time step (1h) to 24h later with hourly aggregated data.

The traffic data used in this study were collected through a network of 21 microwave traffic counters (MobilTraf 300, FAMAS) installed across the Catania urban area. These devices operate autonomously, combining radar-based detection at 24 GHz with onboard electronics and a hybrid power system of batteries and photovoltaic panels, ensuring continuous operation with low energy consumption. For each passing vehicle, the counters record the date and time of transit, the travel direction, and the lane of passage with high measurement accuracy. Data from the sensors were retrieved using the MobilTraf Manager software, which allows export of traffic flow measurements aggregated. Data from fully operational counters between January 1 and March 31, 2023, were selected for this research.

In deep, the proposed approach enables predictions for roads lacking sensor data by utilizing a limited subset of available data and assigning roads to appropriate clusters. A clustering unsupervised model has been created to cluster roads with similar patterns and divide roads with different patterns. In this way, data will be organized in different clusters. Then, for each cluster, has been created a supervised ML model that perform the prediction of traffic flow for roads inside that clusters. See Figure 5 for a description of the architecture of the proposed two-stage framework for traffic flow prediction. On the input side, the method begins with traffic flow data collected across multiple road segments. These time series are then processed by a clustering algorithm, which groups together roads that exhibit comparable temporal patterns of traffic demand (for instance, clusters C1, C2, C3, and C4). Once the clusters are defined, a dedicated forecasting model is trained for each group, relying only on the time series belonging to that cluster. In this way, every forecasting model is tailored to the specific dynamics of its cluster and is used to generate predictions of traffic flows for the next 24 hours for all the corresponding road segments.

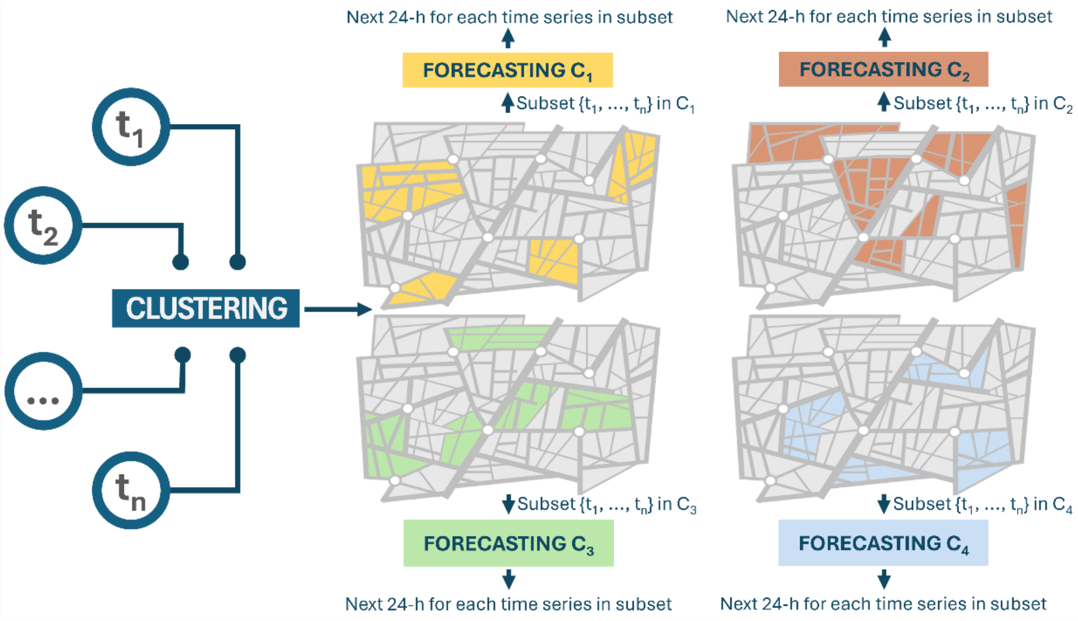

Figure 5 - Two-level ML approach for traffic flow prediction (Berlotti et al., 2024).

According to the proposed approach, prediction on roads where sensors are not present is possible, using this two-level approach, because the model can be applied also to new data from alternative sources such as Google, Tom-Tom, or through on-site data collection. Indeed, roads can be assigned to one of the existing clusters, and subsequently, the corresponding forecasting model, trained with comprehensive and high-precision sensor data over an extended period, can predict traffic flow for roads with only a limited dataset. Different models were tested, with CatBoost (Dorogush et al., 2018) being the best one for the traffic forecasting of the next 24 hours.

The second and third studies (Berlotti et al., 2024; Di Grande, Berlotti, and Cavalieri, 2024) built on the initial pilot by employing a full year of traffic sensor data (until December 2023) rather than the three-month dataset used previously. The availability of a longer observation period enabled the models to capture richer and more complex temporal dynamics, including not only daily and weekly cycles but also broader seasonal variations across the year. This shift from a short dataset to an annual one significantly strengthened the robustness of the models, allowing them to generalize better to irregular conditions such as holidays and atypical traffic events. In these experiments, CatBoost continued to achieve strong predictive performance, but XGBoost (Chen and Guestrin, 2016) emerged as the most accurate and reliable algorithm overall, demonstrating superior robustness across different forecasting horizons and traffic scenarios.

Finally, the research addressed the challenge of sensor-less roads (De Souza Oliveira et al., 2024) in a different way. The aim was to develop forecasts based solely on FCD provided by TomTom. To achieve this, a model was trained using both FCD and sensor-based measurements, with the sensor data serving as the ground-truth target variable against which the predictions were calibrated. In this way, the study sought to develop a ML framework capable of predicting traffic flows on roads without direct monitoring infrastructure. Thus, the approach relies on combining two complementary data sources: fixed sensors, which provide accurate flow counts but are available only on selected road segments, and FCD, which cover the entire urban network but represent only a fraction of total vehicles. Since FCD are generated only by vehicles equipped with the TomTom application or mobile phones actively transmitting data, they offer wide spatial coverage but limited representativeness. On average, FCD account in Catania for about 5% of road users, though this proportion varies from street to street and can fluctuate on the same street over the course of a single day.

The challenge for the model is to learn the relationship between these two data sources: sensor data that reflect the complete set of vehicles on a given road, and FCD that reflect only a subset of users but are available city-wide. These relationships are inherently non-linear and vary by location and time of day, making them difficult to capture with traditional methods. To address this, the study combined sensor-based measurements of traffic flows on instrumented roads with FCD obtained from the TomTom Move platform (*TomTom Move*, n.d.). By training ML models to link these two data streams, it became possible to extrapolate reliable flow estimates across the entire network. This enables traffic monitoring and forecasting not only on sensor-equipped roads but also on streets without direct measurement, thereby extending predictive coverage to the whole city. The experiments confirmed that LSTM was well-suited to the task, as its ability to model sequential dependencies allowed it to capture the complexities of traffic time series. Through iterative feature selection and hyperparameter tuning, the model was optimized to achieve high predictive accuracy, even in scenarios with limited data availability.

3.1.3 Practical Implications

The forecasting models developed in this research translate into a wide range of operational and strategic applications for urban mobility management. At the operational level, short-term forecasts allow traffic control centres to anticipate congestion before it materializes. By acting proactively rather than reactively, operators can adjust traffic signal timings, deploy dynamic rerouting strategies, or regulate access to saturated corridors. These measures not only reduce delays but also lead directly to lower fuel consumption and emissions.

Another significant application lies in the extension of predictive capabilities to sensor-less roads. Traditionally, traffic intelligence is limited to areas with monitoring infrastructure, but the models developed here can generalize to unsensed parts of the network by exploiting spatial and temporal correlations. This enables municipalities to obtain city-wide traffic forecasts without having to invest in costly large-scale sensor deployments, an advantage that is particularly relevant for mid-sized or resource-constrained cities.

The integration of multi-source time series further broadens the impact of these forecasting systems. By combining TomTom navigation data with fixed count sensors, the models capture traffic patterns at both arterial and secondary street levels. This makes it possible to plan interventions at the scale of entire urban areas rather than isolated intersections, enhancing the coordination of infrastructure works, road closures, and special events.

In addition to immediate operational benefits, longer-horizon forecasts support more strategic functions. City planners and logistics operators can rely on these predictions to optimize delivery schedules, synchronize public transport timetables with expected road demand, and even design low-emission zones that take into account forecasted traffic density. By embedding predictive intelligence into daily operations, transport policies can become more adaptive and responsive to evolving conditions.

Ultimately, the cumulative effect of these applications is to generate clear sustainability benefits. Proactive traffic management not only improves the efficiency of mobility systems but also reduces congestion-related emissions, improves air quality, and contributes to healthier urban environments. In this way, the models move beyond technical accuracy to provide a tangible contribution to the environmental, economic, and social pillars of sustainability.

3.2 Cycling Risk Forecasting

3.2.1 Background and Motivation

While efficiency is often presented as the central goal of urban mobility systems, true sustainability cannot be achieved without addressing safety and inclusivity. A transport system that reduces emissions and congestion but fails to protect its most vulnerable users cannot be considered sustainable. Among all transport modes, cycling stands out as one of the most environmentally friendly and health-promoting options.

It produces zero tailpipe emissions, requires minimal infrastructure investment compared to road networks, promotes physical activity, and alleviates traffic congestion (Rimano et al., 2015). For these reasons, cycling is increasingly promoted as a cornerstone of sustainable urban transport strategies across the world.

Yet, despite its many benefits, the widespread adoption of cycling continues to be hindered by a persistent barrier: safety. Cyclists are among the most exposed road users, with little physical protection compared to motor vehicle occupants. Their risk in collisions is shaped not only by traffic volumes and speeds but also by contextual and environmental factors, including road geometry, surface conditions, visibility, lighting, and weather (Macioszek and Granà, 2022). These multi-factorial risks create uncertainty for cyclists and reinforce the perception that cycling is more dangerous than driving, a perception that continues to discourage many potential riders from adopting this sustainable mode of transport.

Traditional safety interventions have primarily focused on infrastructure-based solutions, including the construction of dedicated bicycle lanes, the redesign of intersections, traffic calming measures, and clearer road signage. Infrastructure improvements are necessary but not sufficient: the persistence of high accident rates, increased by 53% over a ten-year period from 2014 to 2023 (*Bicycle Deaths*, 2025), indicates that complementary approaches are needed, especially those that can anticipate risk conditions before accidents happen.

The emergence of large-scale mobility datasets and advances in ML offer a new set of tools for addressing this challenge. Much of the existing research on cycling safety has concentrated on post-event analyses, examining accident severity, causes, and contributing factors after incidents have already occurred (Bassani et al., 2020; Ding et al., 2024; Zhu, 2021). Techniques such as spatial clustering and classification have been applied to accident datasets to identify high-risk locations and understand correlating variables (Bassani et al., 2020; Birfir et al., 2023; Brito et al., 2024; Ding et al., 2024; Lu et al., 2022; C. Wang et al., 2019; Zhu, 2021). These retrospective approaches are valuable for long-term planning and evaluation, but their reactive nature limits their effectiveness for real-time safety management.

A second line of research has explored real-time safety detection through wearable technologies and on-bike sensors, such as accelerometers, GPS units, or video-based hazard detection systems (Lehmann et al., 2022; Schnee et al., 2021; Tabei et al., 2021). These approaches are capable of capturing near-miss events, monitoring rider behaviour, or detecting hazardous riding conditions as they occur. However, their scalability is limited by practical constraints. They require widespread adoption of devices by individual cyclists, continuous data collection, and substantial infrastructure to process and analyse high-frequency time series data. Such requirements make them challenging to deploy at the scale of entire cities or national transport systems.

This gap between retrospective hotspot analyses and real-time detection opens the door for a third, complementary approach: predictive modelling of cyclist risk based on readily available contextual variables. The originality of the present research (Di Grande et al., 2025) lies in the development of a ML framework that leverages historical accident records together with spatiotemporal and environmental features to estimate the likelihood that a given traffic situation poses elevated risk for cyclists. By focusing on features that are universally observable (such as time of day, road surface conditions, lighting status), the proposed framework achieves a level of generalizability that does not depend on specialized devices or intrusive monitoring.

This ensures scalability across different cities and countries, including contexts with limited resources or monitoring infrastructure.

The forward-looking perspective of predictive risk modelling shifts the focus from explaining past crashes to anticipating future high-risk scenarios. Rather than asking "why did a crash occur here?" the model asks "what is the likelihood of a cyclist being involved in an accident under these conditions?". This subtle but important change allows the development of tools that can be deployed in real-time or for scenario-based planning. For urban planners and policymakers, predictive intelligence supports the proactive design of safer infrastructure and more targeted safety campaigns. For navigation applications and cycling support systems, it enables the generation of real-time risk maps, helping cyclists avoid dangerous road segments under specific conditions. Importantly, by reducing both actual and perceived risks, such systems can encourage more people to adopt cycling as a safe, reliable, and sustainable mode of transport.

In this sense, predictive ML models become not only a methodological innovation but also a key enabler of sustainable, inclusive, and equitable mobility systems. Traditional infrastructure investments are complemented by a dynamic, data-driven layer of safety intelligence. By bridging the gap between retrospective crash analyses and real-time detection, this approach provides a practical and scalable solution to one of the central challenges of modern urban mobility: ensuring that cycling is not only efficient and sustainable, but also safe and attractive for all.

3.2.2 Methods and Results

The study titled "Data-Driven Prediction of High-Risk Situations for Cyclists Through Spatiotemporal Patterns and Environmental Conditions" (Di Grande et al., 2025) developed a predictive ML framework designed to improve cyclist safety by identifying conditions under which accidents involving bicycles are more likely to occur. The work was carried out during a period of research abroad at the University of Porto, Faculty of Engineering (FEUP), Portugal, and benefited from collaboration within that academic environment. The model was trained on historical traffic accident records of the entire Germany, available from 2016 to 2023 (data.europa.eu, 2018), where each entry represented a documented incident. Within this dataset, some accidents involved bicycles while others did not, allowing the task to be formulated as a binary classification problem in which the target variable indicated whether a cyclist was involved (1) or not (0).

To ensure generalizability, the input features were restricted to contextual variables that are observable independently of whether an accident has already happened. These included temporal attributes such as time of day, as well as environmental and infrastructural factors such as lighting conditions and the status of the road surface. Furthermore, an additional feature was introduced to incorporate localized accident history into the dataset: bike accident density. By focusing on these universally available features, the model can be applied not only to analyse past incidents but also to estimate the likelihood of cyclist involvement in new or hypothetical scenarios.

This approach transforms traditional retrospective accident analysis into a forward-looking risk estimation tool. The model addresses the question: “Given that an accident occurs at a particular location and time, under certain environmental conditions, what is the probability that a cyclist will be involved?”. In doing so, it enables conditional risk assessment that can be applied in real time or for planning future scenarios, even in places where no bicycle-related accidents have yet been recorded.

Among the algorithms tested, CatBoost proved to be the most effective, consistently delivering balanced and accurate predictions. Its capacity to maintain strong performance across both class-specific metrics and overall accuracy makes it especially suitable for a task where identifying minority cases, such as bicycle-related accidents, is critical. The model's predictive quality is illustrated for example in Freiburg im Breisgau, where comparisons between observed (Figure 6) and predicted (Figure 7) accidents demonstrated the reliability of the approach in capturing local risk patterns.

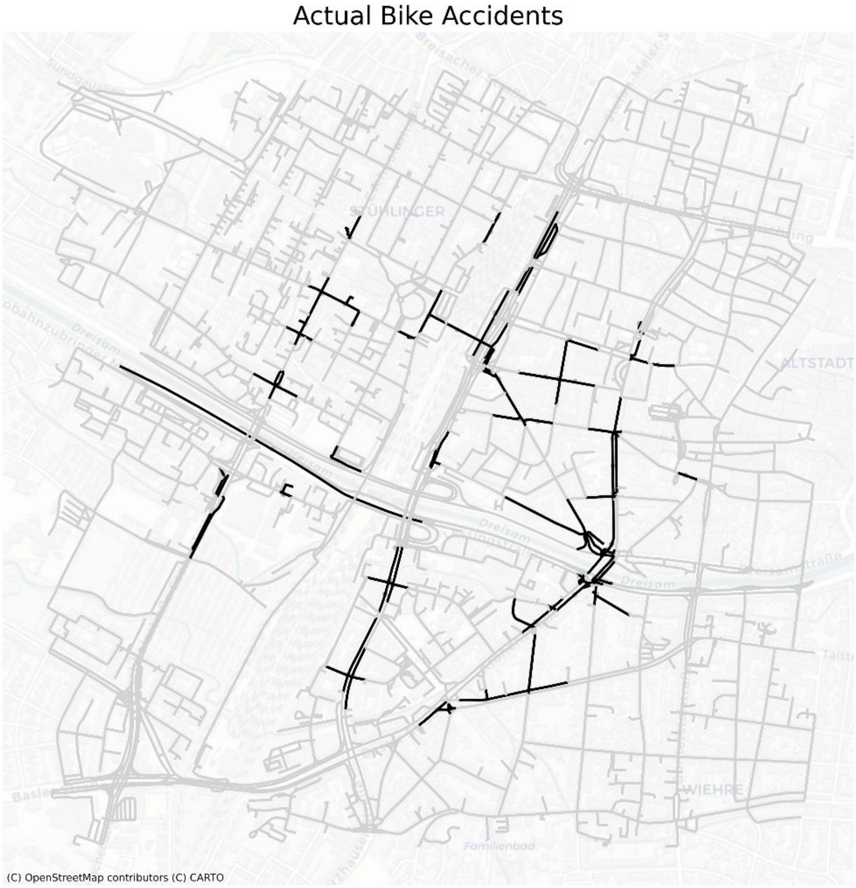

Figure 6 - Actual Bike Accidents in an urban area of Freiburg im Breisgau, Germany (Di Grande et al., 2025).

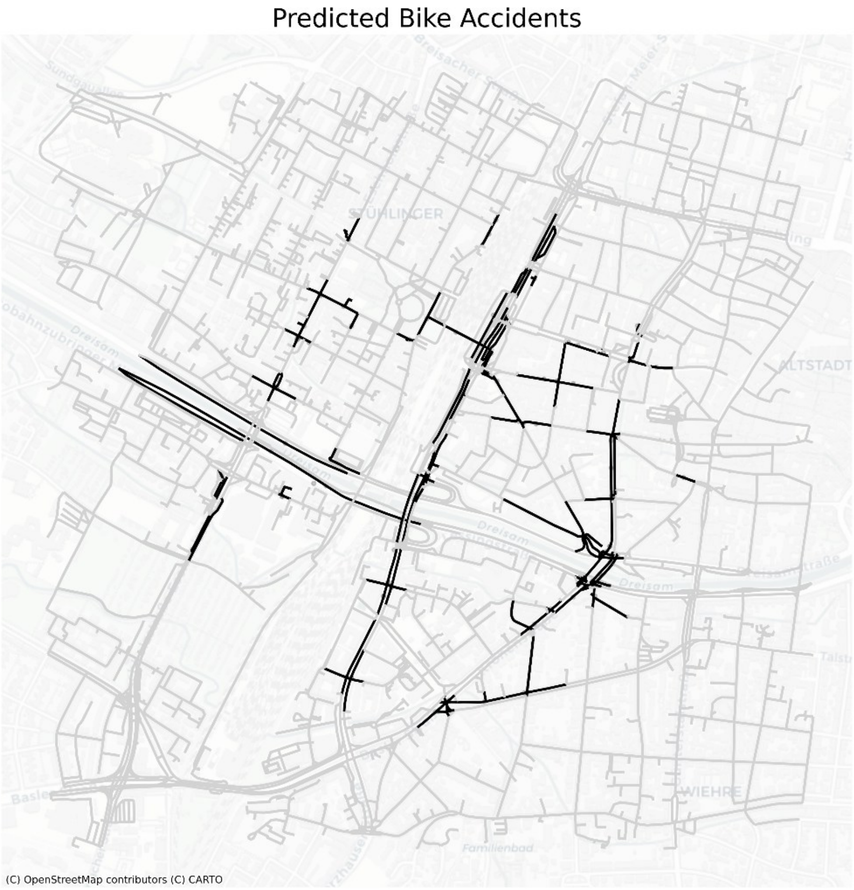
Predicted Bike Accidents

Figure 7 - Predicted Bike Accidents by CatBoost model in an urban area of Freiburg im Breisgau, Germany (Di Grande et al., 2025).

The road network illustrated is limited to streets accessible to cyclists. Segments are highlighted in black if at least one bicycle accident, either observed or predicted, occurred in their vicinity, while all other streets are displayed in light grey for contextual reference. Figure 6 presents the distribution of recorded accidents, whereas Figure 7 depicts the locations predicted by the model. A comparison of the two reveals that the model is able to capture many of the high-risk areas, particularly at major intersections and along primary cycling corridors. Only a small number of streets with observed accidents were not identified by the model. Conversely, certain segments were classified as high-risk despite having no documented accidents, which may reflect model conservatism or possible underreporting of incidents.

The study demonstrated that ML models can effectively identify spatiotemporal hotspots of cycling risk, predicting not only where accidents are more likely but also under what conditions. This capability supports preventive safety measures and contributes to the broader promotion of active mobility.

3.2.3 Practical Implications

The ML framework developed for cycling risk prediction offers a broad range of practical applications that extend from real-time support for individual cyclists to long-term policy and infrastructure planning. By functioning as a proactive risk estimation tool, the model moves beyond retrospective accident analysis and enables forward-looking interventions, anticipating where and when dangerous conditions are most likely to occur.

One of the most immediate applications is its integration into navigation and route-planning services. The model can combine a cyclist's current location, the time of travel, and contextual features such as weather and lighting to provide localized safety predictions for nearby road segments. This capability was demonstrated in a simulation conducted in Berlin, where the system was used to evaluate road safety conditions within a 500-meter radius of a cyclist's position (Di Grande et al., 2025). Each road segment was evaluated at its start, midpoint, and end, and risk was classified as high when all three points exceeded the threshold, moderate when one or two points were flagged, and low when none were flagged.

Figure 8 illustrates the output of this process. In the example, carried out on a Monday afternoon in June 2023, the model produced a map where high-risk segments are displayed in black, moderate-risk in grey, and low-risk in dashed light grey, with the cyclist's location indicated by an icon. This visualization shows how cyclists could be provided with a clear, intuitive safety map in real time, allowing them to adjust their routes and avoid the most hazardous areas. Such functionality could easily be embedded into cycling apps, connected navigation systems, or wearable devices, offering immediate safety benefits to riders without the need for intrusive surveillance or large-scale sensor deployments.

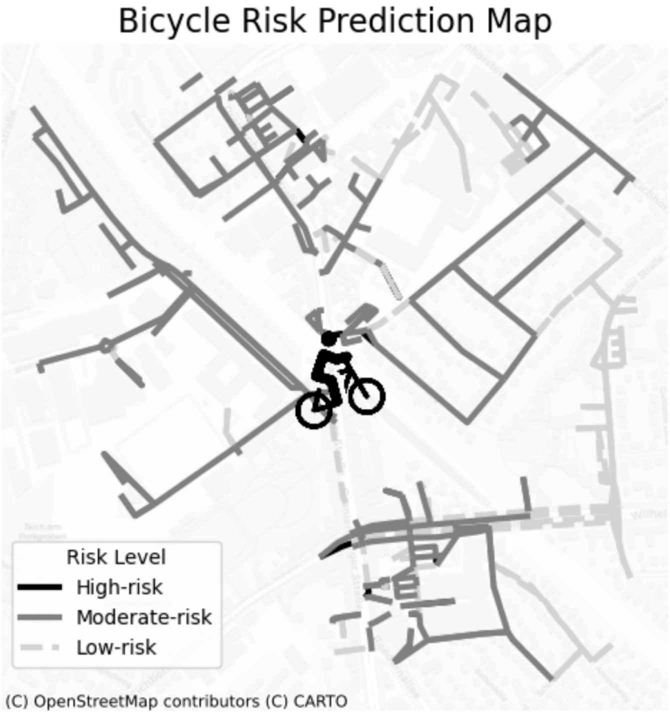

Figure 8 - Map with roads safety predictions (Di Grande et al., 2025).

In this way, the model can be integrated into real-time warning systems, for example by triggering alerts under adverse conditions such as rain, darkness, or poor surface quality. Warnings could be communicated directly to cyclists via mobile applications, increasing awareness and reducing accident likelihood in the most critical situations. In the same mobile application, a cyclist could view the safest route to reach their destination based on security considerations.

Beyond real-time navigation, the framework has substantial implications for urban planning and infrastructure prioritization. By identifying recurring high-risk zones, the model can help municipalities decide where to invest in protective measures, such as building dedicated bike lanes, improving street lighting, or redesigning intersections. Unlike traditional approaches, which depend solely on historical accident counts, this predictive system enables forward-looking planning that anticipates risk before accidents occur, ensuring that interventions are both efficient and preventive.

At the policy level, predictive risk maps generated by the model can be used to support long-term safety strategies. They allow authorities to evaluate the expected safety impact of proposed measures, such as the expansion of cycling networks or the implementation of speed reduction zones, before they are implemented. Over time, predicted risk levels can be compared against actual accident data to monitor the effectiveness of safety policies.

From a broader sustainability perspective, the framework contributes to promoting a modal shift toward active mobility. Safety concerns remain one of the strongest deterrents to cycling, and by reducing both perceived and actual risks, predictive intelligence can encourage more people to adopt the bicycle as a daily mode of transport. Increased cycling uptake, in turn, reduces congestion, lowers emissions, and contributes to public health, thereby reinforcing the environmental, economic, and social pillars of sustainability.

In summary, the proposed model demonstrates how ML can be translated into concrete tools for safer urban mobility. Whether through real-time route guidance, targeted infrastructure investments, policy evaluation, or safety alerts, the framework has the potential to make cycling not only safer but also more attractive and sustainable. By transforming historical accident data into proactive risk maps, it bridges the gap between data analysis and actionable safety measures, supporting both individual cyclists and city authorities in the pursuit of safer, healthier, and more sustainable urban transport systems.

While energy and mobility emphasize environmental and infrastructural sustainability, healthcare highlights the social dimension, where ML can support preventive strategies and improve well-being.

Chapter 4: Machine Learning for Postural Analysis in Healthcare

Postural analysis plays a central role in musculoskeletal health, ergonomics, and rehabilitation, making it an important area of study not only for clinicians but also for public health, preventive medicine and sustainable healthcare systems. Musculoskeletal disorders (MSDs) are among the leading causes of disability worldwide, with the World Health Organization estimating that approximately 1.71 billion of people suffer from MSDs worldwide (WHO, 2022). Poor posture, particularly when sustained over long periods in work or daily activities, is a major contributing factor to these conditions. It can lead to discomfort, chronic pain, reduced mobility, and long-term structural changes in the musculoskeletal system, all of which negatively affect individual well-being and place significant burdens on healthcare systems and economies (GBD 2021 Low Back Pain Collaborators, 2023). In fact, MSDs are among the most common causes of lost workdays and productivity reduction, in both developed and developing nations (Dianat et al., 2015; Soares et al., 2020). Within this context, postural evaluation is not only a diagnostic procedure but also a cornerstone of preventive healthcare, offering a cost-effective means to detect early musculoskeletal imbalances and reduce the long-term clinical and economic impacts of MSDs.

Traditionally, several methods have been used for posture evaluation, each with specific advantages and limitations (Roggio et al., 2021). Two primary approaches dominate: visual inspection and motion capture technologies. Visual inspection remains the most widely adopted method because of its simplicity and low cost. However, it is inherently subjective, with results varying according to the clinician's training, experience, and even momentary judgment, which leads to inconsistencies across evaluators and over time (Fedorak et al., 2003). In contrast, optoelectronic motion capture systems, considered the gold standard, offer unparalleled precision in quantifying joint angles and body inclinations. These systems, however, are prohibitively expensive, require specialized laboratory setups, and are largely restricted to research or elite clinical environments (Corazza et al., 2010). Their reliance on markers, cameras, and controlled conditions further limits their feasibility for routine assessments, large-scale screenings, or everyday applications such as workplaces or schools. This trade-off between accessibility and precision has long represented a barrier to making high-quality postural analysis a standard, scalable healthcare practice.

This dichotomy between subjective but accessible visual inspection and objective but costly laboratory-based systems has long represented a bottleneck to making high-quality postural analysis widely accessible.

Recent advances in ML and computer vision provide a unique opportunity to overcome this barrier by democratizing access to accurate, objective, and low-cost postural evaluation. Pose estimation algorithms such as MediaPipe (Google, 2020), OpenPose (Cao et al., 2021), and MoveNet (Google, 2024) can automatically detect anatomical landmarks in two- and three-dimensional space from ordinary photos or videos. These landmarks enable the computation of joint angles, segment alignments, and body inclinations with reliability approaching that of motion capture systems but without requiring specialized hardware (Hii et al., 2023). A simple camera or smartphone is sufficient to capture postural information, opening the way to scalable, reproducible, and minimally invasive methods for posture assessment.

From a healthcare perspective, these technologies directly contribute to the digital transformation and sustainability of medical services. By decoupling postural analysis from specialized laboratories, ML-based approaches enable new applications in preventive healthcare, ergonomics, education, sports science, and rehabilitation. Clinicians can obtain objective measures during routine visits; employers can monitor worker ergonomics without costly studies; athletes can benefit from continuous feedback on posture in training environments; and individuals can track their own musculoskeletal health using home-based tools. Such capabilities align with emerging trends in telemedicine and digital health, supporting the transition from reactive treatment to proactive monitoring. Importantly, these solutions align with the broader goals of sustainable healthcare by reducing costs, improving accessibility, and enabling early detection and intervention. Preventing musculoskeletal issues before they escalate reduces the need for long-term treatments, rehabilitation programs, or surgeries, thereby alleviating both individual suffering and systemic healthcare expenditures.

Within the broader healthcare landscape, posture analysis and musculoskeletal prevention are increasingly recognized as priorities for public health systems, occupational health programs, and aging populations. The integration of ML-based posture assessment into digital health platforms could support large-scale screening in clinical, workplace, and community settings, directly contributing to the objectives of SDG 3 (“Good Health and Well-Being”). In this sense, the proposed framework does not only address a technical challenge but also responds to an emerging need for scalable, data-driven tools in preventive healthcare.

The research presented in this chapter (Roggio et al., 2024) builds on this opportunity by designing and validating a ML–based pipeline for postural assessment in healthy adults. The study had two main objectives. First, it aimed to evaluate the applicability and reliability of ML-based pose estimation for static posture analysis, systematically comparing its performance with that of traditional approaches. The focus was on whether pose estimation could consistently extract reliable anatomical measurements across individuals and sessions, an essential requirement for clinical or ergonomic adoption. Second, the study explored the use of unsupervised ML techniques, including PCA and clustering, to investigate whether new structural patterns in posture could be discovered beyond conventional distinctions such as sex or basic anthropometric characteristics. This exploratory element acknowledges that posture is a multidimensional phenomenon that may not be fully explained by existing categorical distinctions, and that ML may uncover latent structures or "types" of posture previously overlooked.

By providing normative datasets, testing reproducibility, and demonstrating methodological robustness, this research contributes to the development of scalable, affordable, and reliable postural analysis tools. Its implications extend far beyond academic curiosity. In healthcare, the approach supports preventive screening and telemedicine, enabling clinicians to monitor patients remotely and consistently. In ergonomics, it can guide workplace interventions to reduce the risk of repetitive strain injuries and improve occupational health. In sports science, it may inform training strategies and injury prevention programs. Finally, in rehabilitation, it offers a practical way to monitor patient progress over time, providing quantitative evidence of recovery or setbacks.

By integrating ML into posture analysis, this research aligns with the principles of computational sustainability: it leverages data-driven methods to improve human well-being, reduce healthcare burdens, and create more equitable access to advanced diagnostic tools. In doing so, it contributes to a future where musculoskeletal health is monitored not only in specialized labs but also in everyday environments, ensuring that preventive care becomes accessible to all.

The following section provide a summary of the main methods and results; the original paper should be consulted for the complete technical specifications.

4.1 Methods and Results

The study titled "Biomechanical Posture Analysis in Healthy Adults with Machine Learning: Applicability and Reliability" (Roggio et al., 2024) confirmed that ML-based pose estimation can provide a reliable and practical tool for static postural assessment in healthy adults. Using standardized photographs and the MediaPipe Pose framework, we extracted anatomical landmarks and derived postural indicators such as joint angles, body inclinations, and vector lengths. The analysis revealed clear differences between male and female participants, for example in shoulder and hip alignment, confirming that sex-related variability in posture can be consistently captured by ML methods.

A central goal of the work was to assess reproducibility, and the test–retest procedure demonstrated that the majority of postural parameters remained highly stable when participants were re-evaluated after one week. This finding shows that the proposed methodology does not simply capture noise or momentary fluctuations but is capable of producing consistent measurements over time.

Correlation analysis further highlighted how anthropometric features, such as height, naturally influence postural parameters, with taller individuals tending to present larger vector lengths in the torso and legs. While expected, these relationships validated that the system was sensitive to underlying structural differences, reinforcing its credibility as an objective assessment tool.

Beyond conventional comparisons, the use of unsupervised learning methods opened up new perspectives. PCA reduced the complexity of the dataset while retaining most of its variability, and clustering revealed two distinct postural groups that were not explained by sex alone. These groups differed in structural proportions, particularly in torso and shoulder–hip relations, suggesting that posture may be meaningfully categorized into types resembling somatotypes, such as more compact versus more elongated structures. This is an important insight, as it demonstrates that ML can uncover latent categories that traditional assessment approaches might overlook.

Taken together, these results show that the ML-based pipeline is both applicable and reliable, capable of producing stable and meaningful postural indicators, while also generating new insights into structural variability. Instead of requiring expensive laboratory-based equipment, this approach offers an accessible, low-cost solution that maintains scientific rigor and opens opportunities for broader use in preventive healthcare, ergonomics, and rehabilitation.

4.2 Practical Implications

The findings of this study have several concrete applications across healthcare, ergonomics, sports, and preventive medicine.

In clinical practice, the method provides a low-cost, objective alternative to subjective visual inspections. By reducing reliance on clinician judgment, it supports more consistent diagnoses and monitoring of postural deviations, enabling early identification of risk factors for musculoskeletal disorders. The demonstrated high reproducibility suggests that this approach can be reliably integrated into clinical workflows, including remote telehealth services where patients submit photographs for automated evaluation.

In ergonomics and workplace health, predictive posture analysis can be applied to detect musculoskeletal load risks in workers. Employers could use simple photographs or video recordings to identify individuals at higher risk of repetitive strain or poor ergonomic conditions, and intervene with targeted adjustments to workplace design or training.

In sports science, posture plays a crucial role in performance optimization and injury prevention. The unsupervised clustering approach, which identified distinct postural categories, can inform athlete profiling, linking body structure and posture to sport-specific demands. Coaches could use such tools to tailor training programs, correct imbalances, or monitor the effects of long-term conditioning.

In rehabilitation and physiotherapy, this methodology offers an efficient and non-invasive way to monitor progress in patients recovering from injuries or surgeries. By providing objective measurements of angles, inclinations, and vector lengths, clinicians can quantify improvements over time and adjust rehabilitation exercises accordingly.

Finally, from a public health and sustainability perspective, the adoption of ML-driven posture analysis contributes to reducing healthcare costs by enabling preventive screening. By making reliable postural evaluation accessible outside specialized labs, the method supports early interventions, reducing the burden of chronic musculoskeletal disorders. This aligns directly with the sustainability goals of improving population health, reducing inequities in access to diagnostic tools, and promoting healthier lifestyles.

Chapter 5: Conclusions

This thesis shows how ML can serve as a catalyst for advancing Computational Sustainability across three domains crucial to contemporary society: energy, urban mobility, and healthcare. While each of these sectors faces domain-specific challenges, they are inherently interdependent. Energy efficiency shapes mobility systems, urban air quality influences public health outcomes, and healthcare infrastructures contribute to the broader sustainability of communities. The works presented in this dissertation reflect this interconnectedness and demonstrate that sustainability should be treated not as an aggregation of isolated efforts, but as a network of co-evolving systems.

The concept of Computational Sustainability provides the unifying framework for integrating these interrelated tasks. It addresses the balance between environmental resource management, economic viability, and social equity within computational decision-making processes. Each study in this Ph.D. research embodies this principle through the design of predictive and adaptive systems that optimize real-world operations while advancing SDGs. In the energy domain, predictive models enhance the efficiency of water distribution and hydropower generation, reducing waste and supporting renewable integration. In urban mobility, forecasting frameworks improve traffic fluidity and cyclist safety, contributing to cleaner and more liveable cities. In healthcare, posture analysis models promote preventive, accessible medicine: an essential pillar of social sustainability.

Viewed together, these contributions reveal a coherent computational ecosystem. The same methodological strategies recur across domains, making them transferable solutions rather than domain-specific tools. For instance, spatiotemporal representation learning developed for traffic prediction can be repurposed for energy demand forecasting, where seasonal and behavioural cycles mirror commuting dynamics. This interconnection underscores one of the central messages of this thesis: advances in one area of Computational Sustainability can generate systemic benefits across others.

5.1 Machine Learning in Computational Sustainability

ML represents both the methodological nucleus and the conceptual bridge across the three areas investigated. The computational challenges in these domains (e.g., nonlinear dynamics, data scarcity, uncertainty, and multi-objective optimization) demand adaptive, data-centric strategies. Throughout the thesis, a diverse set of ML approaches, including time series forecasting and clustering, have been employed to address these challenges.

Beyond technical performance, a key insight emerging from this research is the importance of scalability and generalizability. Global models in energy and mobility sectors have proven more efficient and robust than localized ones, enabling shared learning across contexts. This principle mirrors the broader goal of Computational Sustainability: to generalize computational insights to diverse socio-environmental challenges.

Moreover, these ML-based systems demonstrate how computational methods can act as instruments of sustainability themselves. Computation actively contributes to the optimization of energy use, mobility flows, and healthcare processes. Integrating ML into sustainable systems therefore amplifies the feedback loop between technological innovation and societal well-being, moving from predictive modelling toward prescriptive decision-support for sustainable policies.

An emerging frontier lies in integrated data infrastructures, frameworks where models from different sectors interact dynamically. An urban management system that combines hydropower forecasts with mobility analytics, for example, could reduce emissions by synchronizing energy production with transportation demand. Similarly, combining environmental monitoring with health data could enable early-warning systems for pollution-related diseases. The methodological consistency demonstrated across the energy, mobility, and health domains creates the foundation for such holistic, computationally sustainable architectures.

5.2 Future Perspectives

The contributions of this thesis open exciting avenues for future research in the rapidly evolving field of Computational Sustainability. One promising direction lies in the development of hybrid models that combine traditional physical process knowledge with modern deep-learning techniques. Such integration would not only enhance the interpretability and stability of predictive models but also ground them in well-established domain expertise, particularly important in areas like energy systems and hydrology where physical laws govern system behaviour. This fusion has the potential to reduce uncertainty and improve the robustness of sustainability predictions.

Another important research frontier concerns the ability to generalize insights across domains. For instance, patterns learned from forecasting energy consumption might inform models predicting urban mobility dynamics, thereby fostering efficient knowledge transfer and reducing the need for extensive data collection in every sector. Exploring such cross-domain generalization would accelerate innovation and promote resource-efficient computing in sustainability applications.

Ethics will take centre stage as computational methods increasingly impact public welfare. Future work must prioritize the development of machine learning systems that are not only powerful but also transparent, fair, and explainable. Embedding these values into sustainability-focused AI will be essential to securing societal trust and ensuring that technological advances benefit all communities equitably.

To maximize real-world impact, the predictive models introduced here could evolve into decision-support platforms directly accessible to policymakers, utility managers, and healthcare providers. Translating academic research into operational tools is essential for data-driven sustainable governance.

Looking further ahead, this research envisions a future where fully integrated computational ecosystems dynamically coordinate sustainability efforts at both micro and macro scales, from individual infrastructure systems to regional policy frameworks. Such intelligent ecosystems will underpin the smart cities, circular economies, and carbon-neutral societies of tomorrow.

In closing, the work presented in this thesis shows that ML is not merely a tool for technical optimization but a catalyst for sustainability. By reducing inefficiencies in energy and water use, by enabling more efficient and safer transportation, and by expanding access to preventive healthcare, ML can contribute directly to environmental protection, economic resilience, and social well-being. These contributions align with the vision of the SDGs and affirm the central potential of computational sustainability as a field capable of integrating scientific innovation with societal needs. Much remains to be done, but the results presented here provide evidence that artificial intelligence, guided by the principles of sustainability, can help humanity navigate the complexities of the twenty-first century and move toward a future where development is not achieved at the expense of the planet or of future generations, but in harmony with them.

References

Abdelsalam, A. A., and Gabbar, H. A. (2021). Energy saving and management of water pumping networks. *Heliyon*, *7*(8). https://doi.org/10.1016/j.heliyon.2021.e07820

Adedeji, K. B., Ponnle, A. A., Abu-Mahfouz, A. M., and Kurien, A. M. (2022). Towards Digitalization of Water Supply Systems for Sustainable Smart City Development—Water 4.0. *Applied Sciences*, *12*(18), Article 18. https://doi.org/10.3390/app12189174

Ahmad, T., Zhang, H., and Yan, B. (2020). A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. *Sustainable Cities and Society*, *55*, 102052. https://doi.org/10.1016/j.scs.2020.102052

Alhendi, A. A., Al-Sumaiti, A. S., Elmay, F. K., Wescaot, J., Kavousi-Fard, A., Heydarian-Forushani, E., and Alhelou, H. H. (2022). Artificial intelligence for water–energy nexus demand forecasting: A review. *International Journal of Low-Carbon Technologies*, *17*, 730–744. https://doi.org/10.1093/ijlct/ctac043

Ampas, H., Refanidis, I., and Ampas, V. (2025). Hybrid Hydrological Forecasting Through a Physical Model and a Weather-Informed Transformer Model: A Case Study in Greek Watershed. *Applied Sciences*, *15*(12), 6679. https://doi.org/10.3390/app15126679

Anciaes, P., Cheng, Y., and Watkins, S. J. (2025). Policy measures to reduce road congestion: What worked? *Journal of Transport & Health*, *41*, 101984. https://doi.org/10.1016/j.jth.2025.101984

Antoniou, C., Balakrishna, R., and Koutsopoulos, H. N. (2011). A Synthesis of emerging data collection technologies and their impact on traffic management applications. *European Transport Research Review*, *3*(3), 139–148. https://doi.org/10.1007/s12544-011-0058-1

Asgarzadeh, M., Verma, S., Mekary, R. A., Courtney, T. K., and Christiani, D. C. (2017). The role of intersection and street design on severity of bicycle-motor vehicle crashes. *Injury Prevention*, *23*(3), 179–185. https://doi.org/10.1136/injuryprev-2016-042045

Aslam, S., Herodotou, H., Mohsin, S. M., Javaid, N., Ashraf, N., and Aslam, S. (2021). A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renewable and Sustainable Energy Reviews*, *144*, 110992. https://doi.org/10.1016/j.rser.2021.110992

Barzola-Monteses, J., Gómez-Romero, J., Espinoza-Andaluz, M., and Fajardo, W. (2022). Hydropower production prediction using artificial neural networks: An Ecuadorian application case. *Neural Computing and Applications*, *34*(16), 13253–13266. https://doi.org/10.1007/s00521-021-06746-5

Bassani, M., Rossetti, L., and Catani, L. (2020). Spatial analysis of road crashes involving vulnerable road users in support of road safety management strategies. *Transportation Research Procedia*, *45*, 394–401. https://doi.org/10.1016/j.trpro.2020.03.031

Berlotti, M., Di Grande, S., and Cavalieri, S. (2024). Proposal of a Machine Learning Approach for Traffic Flow Prediction. *Sensors*, *24*(7), 2348. https://doi.org/10.3390/s24072348

Berlotti, M., Di Grande, S., Cavalieri, S., Torrisi, V., and Inturri, G. (2023). Proposal of an AI based approach for Urban Traffic Prediction from Mobility Data. *2023 IEEE International Conference on Big Data (BigData)*, 2570–2577. https://doi.org/10.1109/BigData59044.2023.10386509

Bicycle Deaths. (2025). Injury Facts. https://injuryfacts.nsc.org/home-and-community/safety-topics/bicycle-deaths/

Birfir, S., Elalouf, A., and Rosenbloom, T. (2023). Building machine-learning models for reducing the severity of bicyclist road traffic injuries. *Transportation Engineering*, *12*, 100179. https://doi.org/10.1016/j.treng.2023.100179

Bolognesi, A., Bragalli, C., Lenzi, C., and Artina, S. (2014). Energy Efficiency Optimization in Water Distribution Systems. *Procedia Engineering*, *70*, 181–190. https://doi.org/10.1016/j.proeng.2014.02.021

Brás, M., Moura, A., and Andrade-Campos, A. (2025). Cost efficiency in water supply systems: An applied review on optimization models for the pump scheduling problem. *European Journal of Operational Research*, *323*(1), 1–19. https://doi.org/10.1016/j.ejor.2024.07.039

Brito, B., Costa, D. G., and Silva, I. (2024). Geospatial Risk Assessment of Cyclist Accidents in Urban Areas: A K-means Clustering Approach. *2024 IEEE 22nd Mediterranean Electrotechnical Conference (MELECON)*, 744–749. https://doi.org/10.1109/MELECON56669.2024.10608791

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *43*(1), 172–186. https://doi.org/10.1109/TPAMI.2019.2929257

Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Corazza, S., Mündermann, L., Gambaretto, E., Ferrigno, G., and Andriacchi, T. P. (2010). Markerless Motion Capture through Visual Hull, Articulated ICP and Subject Specific Model Generation. *International Journal of Computer Vision*, *87*(1), 156–169. https://doi.org/10.1007/s11263-009-0284-3

data.europa.eu. (2018). *Data download for road accidents in Germany*. https://data.europa.eu/data/datasets/5d7cb58c-c4f4-4998-a7fd-2cb3abb0e512?locale=en

de Souza Groppo, G., Costa, M. A., and Libânio, M. (2019). Predicting water demand: A review of the methods employed and future possibilities. *Water Supply*, *19*(8), 2179–2198. https://doi.org/10.2166/ws.2019.122

De Souza Oliveira, T., Di Grande, S., Berlotti, M., Cavalieri, S., Torrisi, V., Calabrò, G., and Inturri, G. (2024). Enhancing Urban Traffic Management Through Machine Learning Prediction Models for Sensor-Less Roads. In *Emerging Cutting-Edge Applied Research and Development in Intelligent Traffic and Transportation Systems* (pp. 102–111). IOS Press. https://doi.org/10.3233/ATDE241185

Deng, J., Jin, L., Wang, H., Zhang, Z., Liu, Y., Meng, F., Wang, J., Li, Z., and Wu, J. (2025). Distributed Acoustic Sensing for Road Traffic Monitoring: Principles, Signal Processing, and Emerging Applications. *Infrastructures*, *10*(9), 228. https://doi.org/10.3390/infrastructures10090228

Di Grande, S., Berlotti, M., and Cavalieri, S. (2024). AI-Powered Urban Mobility Analysis for Advanced Traffic Flow Forecasting: *Proceedings of the 13th International Conference on Smart Cities and Green ICT Systems*, 57–64. https://doi.org/10.5220/0012625900003714

Di Grande, S., Berlotti, M., Cavalieri, S., and Costa, D. G. (2025). Data-Driven Prediction of High-Risk Situations for Cyclists Through Spatiotemporal Patterns and Environmental Conditions. *14th International Conference on Data Science, Technology and Applications*, 677–684. https://www.scitepress.org/Link.aspx?doi=10.5220/0013646400003967

Di Grande, S., Berlotti, M., Cavalieri, S., and Gueli, R. (2023a). A Machine Learning Approach for Hydroelectric Power Forecasting. *2023 14th International Renewable Energy Congress (IREC)*, 1–6. https://doi.org/10.1109/IREC59750.2023.10389561

Di Grande, S., Berlotti, M., Cavalieri, S., and Gueli, R. (2023b). A Proactive Approach for the Sustainable Management of Water Distribution Systems: *Proceedings of the 12th International Conference on Data Science, Technology and Applications*, 115–125. https://doi.org/10.5220/0012121200003541

Di Grande, S., Berlotti, M., Cavalieri, S., and Gueli, R. (2024a). A Machine Learning Approach to Forecasting Hydropower Generation. *Energies*, *17*(20), 5163. https://doi.org/10.3390/en17205163

Di Grande, S., Berlotti, M., Cavalieri, S., and Gueli, R. (2024b). Data Science for the Promotion of Sustainability in Smart Water Distribution Systems. In O. Gusikhin, S. Hammoudi, and A. Cuzzocrea (Eds.), *Data Management Technologies and Applications* (pp. 50–72). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-68919-2_3

Di Grande, S., Berlotti, M., Cavalieri, S., and Gueli, R. (2024c). Harnessing Multivariate AI to Enhance Hydropower Generation Forecasting. *2024 AEIT International Annual Conference (AEIT)*, 1–6. https://doi.org/10.23919/AEIT63317.2024.10736754

Di Grande, S., Berlotti, M., Cavalieri, S., and Gueli, R. (2024d). Optimizing Planning Strategies: A Machine Learning Forecasting Model for Energy Aggregators and Hydropower Producers. *26th International Conference on Enterprise Information Systems*, 490–501. https://doi.org/10.5220/0012626100003690

Di Grande, S., Berlotti, M., Cavalieri, S., and Gueli, R. (In press.). AI-Driven Hydropower Forecasting: A Multivariate Machine Learning Approach. *2025 AEIT International Annual Conference (AEIT)*, 1–6.

Dianat, I., Kord, M., Yahyazade, P., Karimi, M. A., and Stedmon, A. W. (2015). Association of individual and work-related risk factors with musculoskeletal symptoms among Iranian sewing machine operators. *Applied Ergonomics*, *51*, 180–188. https://doi.org/10.1016/j.apergo.2015.04.017

Ding, H., Wang, R., Chen, T., Sze, N. N., Chung, H., and Dong, N. (2024). A hybrid approach for modeling bicycle crash frequencies: Integrating random forest based SHAP model with random parameter negative binomial regression model. *Accident Analysis & Prevention*, *208*, 107778. https://doi.org/10.1016/j.aap.2024.107778

Doğan Güzel, T., and Alp, K. (2024). The effects of technological developments in transportation vehicles on air pollution mitigation of metropolitan cities: A case study of Istanbul. *Science of The Total Environment*, *912*, 168996. https://doi.org/10.1016/j.scitotenv.2023.168996

Dorogush, A. V., Ershov, V., and Gulin, A. (2018). *CatBoost: Gradient boosting with categorical features support* (arXiv:1810.11363). arXiv. https://doi.org/10.48550/arXiv.1810.11363

EEA. (2019). *Air quality in Europe 2019*. https://www.eea.europa.eu/en/analysis/publications/air-quality-in-europe-2019

EEA. (2025, February 27). *Health impacts of exposure to noise from transport in Europe*. https://www.eea.europa.eu/en/analysis/indicators/health-impacts-of-exposure-to-1

Electricity from renewable sources reaches 47% in 2024. (2025, March 19). Eurostat. https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20250319-1

Esen, Ö., Yıldırım, D. Ç., and Yıldırım, S. (2020). Threshold effects of economic growth on water stress in the Eurozone. *Environmental Science and Pollution Research*, *27*(25), 31427–31438. https://doi.org/10.1007/s11356-020-09383-y

European Court of Auditors. (2019). *Urban mobility in the EU*.

European Environment Agency. (2016). *Explaining road transport emissions: A non technical guide*. Publications Office. https://data.europa.eu/doi/10.2800/71804

Fedorak, C., Ashworth, N., Marshall, J., and Paull, H. (2003). Reliability of the visual assessment of cervical and lumbar lordosis: How good are we? *Spine*, *28*(16), 1857–1859. https://doi.org/10.1097/01.BRS.0000083281.48923.BD

Fuso Nerini, F., Tomei, J., To, L. S., Bisaga, I., Parikh, P., Black, M., Borrion, A., Spataru, C., Castán Broto, V., Anandarajah, G., Milligan, B., and Mulugetta, Y. (2018). Mapping synergies and trade-offs between energy and the Sustainable Development Goals. *Nature Energy*, *3*(1), 10–15. https://doi.org/10.1038/s41560-017-0036-5

GBD 2021 Low Back Pain Collaborators. (2023). Global, regional, and national burden of low back pain, 1990-2020, its attributable risk factors, and projections to 2050: A systematic analysis of the Global Burden of Disease

Study 2021. *The Lancet. Rheumatology*, *5*(6), e316–e329. https://doi.org/10.1016/S2665-9913(23)00098-X

Gheorghe, C., and Soica, A. (2025). Revolutionizing Urban Mobility: A Systematic Review of AI, IoT, and Predictive Analytics in Adaptive Traffic Control Systems for Road Networks. *Electronics*, *14*(4), 719. https://doi.org/10.3390/electronics14040719

Gomes, C. (2009). *Computational Sustainability: Computational Methods for a Sustainable Environment, Economy, and Society*. *The Bridge, 39*(4), 5–13.

Gomes, C., Dietterich, T., Barrett, C., Conrad, J., Dilkina, B., Ermon, S., Fang, F., Farnsworth, A., Fern, A., Fern, X., Fink, D., Fisher, D., Flecker, A., Freund, D., Fuller, A., Gregoire, J., Hopcroft, J., Kelling, S., Kolter, Z., … Zeeman, M. L. (2019). Computational sustainability: Computing for a better world and a sustainable future. *Communications of the ACM, 62*(9), 56–65. https://doi.org/10.1145/3339399

Google. (2020). *Mediapipe*. https://chuoling.github.io/mediapipe/

Google. (2024). *MoveNet: Ultra fast and accurate pose detection model*. TensorFlow. https://www.tensorflow.org/hub/tutorials/movenet

Greenhouse Gas Emissions from Energy Data Explorer – Data Tools. (2024). IEA. https://www.iea.org/data-and-statistics/data-tools/greenhouse-gas-emissions-from-energy-data-explorer

Hernandez-Matheus, A., Löschenbrand, M., Berg, K., Fuchs, I., Aragüés-Peñalba, M., Bullich-Massagué, E., and Sumper, A. (2022). A systematic review of machine learning techniques related to local energy communities. *Renewable and Sustainable Energy Reviews*, *170*, 112651. https://doi.org/10.1016/j.rser.2022.112651

Herzen, J., Lässig, F., Piazzetta, S. G., Neuer, T., Tafti, L., Raille, G., Van Pottelbergh, T., Pasieka, M., Skrodzki, A., Huguenin, N., Dumonal, M., Kościsz, J., Bader, D., Gusset, F., Benheddi, M., Williamson, C., Kosinski, M., Petrik, M., and Grosch, G. (2023). Darts: User-friendly modern machine learning for time series. *The Journal of Machine Learning Research*, *23*(1), 124:5442-124:5447.

Hii, C. S. T., Gan, K. B., Zainal, N., Mohamed Ibrahim, N., Azmin, S., Mat Desa, S. H., van de Warrenburg, B., and You, H. W. (2023). Automated Gait Analysis Based on a Marker-Free Pose Estimation Model. *Sensors*, *23*(14), 6489. https://doi.org/10.3390/s23146489

Hussain, Z., Wang, Z., Wang, J., Yang, H., Arfan, M., Hassan, D., Wang, W., Azam, M. I., and Faisal, M. (2022). A comparative Appraisal of Classical and Holistic Water Scarcity Indicators. *Water Resources Management*, *36*(3), 931–950. https://doi.org/10.1007/s11269-022-03061-z

Iria, J., and Soares, F. (2023). An energy-as-a-service business model for aggregators of prosumers. *Applied Energy*, *347*, 121487. https://doi.org/10.1016/j.apenergy.2023.121487

ISO 50001. (2018). *Energy management systems—Requirements with guidance for use. ISO 50001:2018(E).*

Jung, J., Han, H., Kim, K., and Kim, H. S. (2021). Machine Learning-Based Small Hydropower Potential Prediction under Climate Change. *Energies*, *14*(12), Article 12. https://doi.org/10.3390/en14123643

Kerscher, S., and Arboleya, P. (2022). The key role of aggregators in the energy transition under the latest European regulatory framework. *International Journal of Electrical Power & Energy Systems*, *134*, 107361. https://doi.org/10.1016/j.ijepes.2021.107361

Kezunovic, M., Pinson, P., Obradovic, Z., Grijalva, S., Hong, T., and Bessa, R. (2020). Big data analytics for future electricity grids. *Electric Power Systems Research*, *189*, 106788. https://doi.org/10.1016/j.epsr.2020.106788

Klise, K., Hart, D., Bynum, M., Hogge, J., Haxton, T., Murray, R., and Burkhardt, J. (2020). *Water Network Tool for Resilience (WNTR). User Manual, Version 0.2.3* (SAND-2020-9301R; EPA/600/R-20/185). Sandia National Lab. (SNL-NM), Albuquerque, NM (United States). https://doi.org/10.2172/1660790

Lehmann, M., Mair, D., and Guehring, G. (2022). DANGER DETECTION FOR CYCLISTS WITH MACHINE LEARNING (IN THE CITY OF COPENHAGEN). *International Journal for Traffic and Transport Engineering (IJTTE)*, *12*, 272–290. https://doi.org/10.7708/ijtte2022.12(2).09

Leitão, J., Simões, N., Sá Marques, J. A., Gil, P., Ribeiro, B., and Cardoso, A. (2019). Detecting urban water consumption patterns: A time-series clustering approach. *Water Supply*, *19*(8), 2323–2329. https://doi.org/10.2166/ws.2019.113

Liao, Y., Gil, J., Yeh, S., Pereira, R. H. M., and Alessandretti, L. (2025). Socio-spatial segregation and human mobility: A review of empirical evidence. *Computers, Environment and Urban Systems*, *117*, 102250. https://doi.org/10.1016/j.compenvurbsys.2025.102250

Lu, W., Liu, J., Fu, X., Yang, J., and Jones, S. (2022). Integrating machine learning into path analysis for quantifying behavioral pathways in bicycle-motor vehicle crashes. *Accident; Analysis and Prevention*, *168*, 106622. https://doi.org/10.1016/j.aap.2022.106622

Luna, T., Ribau, J., Figueiredo, D., and Alves, R. (2019). Improving energy efficiency in water supply systems with pump scheduling optimization. *Journal of*

Cleaner Production, *213*, 342–356. https://doi.org/10.1016/j.jclepro.2018.12.190

Macioszek, E., and Granà, A. (2022). The Analysis of the Factors Influencing the Severity of Bicyclist Injury in Bicyclist-Vehicle Crashes. *Sustainability*, *14*(1), 215. https://doi.org/10.3390/su14010215

Mesalie, R. A., Aklog, D., and Kifelew, M. S. (2021). Failure assessment for drinking water distribution system in the case of Bahir Dar institute of technology, Ethiopia. *Applied Water Science*, *11*(8), 138. https://doi.org/10.1007/s13201-021-01465-7

Millennium Development Goals (MDGs). (2018). https://www.who.int/news-room/fact-sheets/detail/millennium-development-goals-(mdgs)

Montero-Manso, P., and Hyndman, R. J. (2021). Principles and algorithms for forecasting groups of time series: Locality and globality. *International Journal of Forecasting*, *37*(4), 1632–1653. https://doi.org/10.1016/j.ijforecast.2021.03.004

Moraga, Á., de Curtò, J., de Zarzà, I., and Calafate, C. T. (2025). AI-Driven UAV and IoT Traffic Optimization: Large Language Models for Congestion and Emission Reduction in Smart Cities. *Drones*, *9*(4), 248. https://doi.org/10.3390/drones9040248

Mosavi, A., Salimi, M., Faizollahzadeh Ardabili, S., Rabczuk, T., Shamshirband, S., and Varkonyi-Koczy, A. R. (2019). State of the Art of Machine Learning Models in Energy Systems, a Systematic Review. *Energies*, *12*(7), Article 7. https://doi.org/10.3390/en12071301

Mrabet, M., and Sliti, M. (2024). Integrating machine learning for the sustainable development of smart cities. *Frontiers in Sustainable Cities*, *6*. https://doi.org/10.3389/frsc.2024.1449404

Mystakidis, A., Koukaras, P., and Tjortjis, C. (2025). Advances in Traffic Congestion Prediction: An Overview of Emerging Techniques and Methods. *Smart Cities*, *8*(1), 25. https://doi.org/10.3390/smartcities8010025

Oreshkin, B. N., Carpov, D., Chapados, N., and Bengio, Y. (2020). *N-BEATS: Neural basis expansion analysis for interpretable time series forecasting* (arXiv:1905.10437). arXiv. https://doi.org/10.48550/arXiv.1905.10437

Patil, R., Alandikar, P., Chaudhari, V., Patil, P., and Deshpande, Prof. S. (2022). Water Demand Prediction Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, *10*(12), 122–128. https://doi.org/10.22214/ijraset.2022.47797

Polprasert, J., Hanh Nguyên, V. A., and Nathanael Charoensook, S. (2021). Forecasting Models for Hydropower Production Using ARIMA Method. *2021 9th International Electrical Engineering Congress (iEECON)*, 197–200. https://doi.org/10.1109/iEECON51072.2021.9440293

Purvis, B., Mao, Y., and Robinson, D. (2019). Three pillars of sustainability: In search of conceptual origins. *Sustainability Science*, *14*(3), 681–695. https://doi.org/10.1007/s11625-018-0627-5

Rasaizadi, A., Sherafat, E., and Seyedabrishami, S. (2021). Short-term prediction of traffic state, statistical approach versus machine learning approach. *Scientia Iranica*. https://doi.org/10.24200/sci.2021.57906.5469

Reynolds, C. C., Harris, M. A., Teschke, K., Cripton, P. A., and Winters, M. (2009). The impact of transportation infrastructure on bicycling injuries and crashes:

A review of the literature. *Environmental Health*, *8*, 47. https://doi.org/10.1186/1476-069X-8-47

Rimano, A., Piccini, M. P., Passafaro, P., Metastasio, R., Chiarolanza, C., Boison, A., and Costa, F. (2015). The bicycle and the dream of a sustainable city: An explorative comparison of the image of bicycles in the mass-media and the general public. *Transportation Research Part F: Traffic Psychology and Behaviour*, *30*, 30–44. https://doi.org/10.1016/j.trf.2015.01.008

Roggio, F., Di Grande, S., Cavalieri, S., Falla, D., and Musumeci, G. (2024). Biomechanical Posture Analysis in Healthy Adults with Machine Learning: Applicability and Reliability. *Sensors (Basel, Switzerland)*, *24*(9), 2929. https://doi.org/10.3390/s24092929

Roggio, F., Ravalli, S., Maugeri, G., Bianco, A., Palma, A., Di Rosa, M., and Musumeci, G. (2021). Technological advancements in the analysis of human motion and posture management through digital devices. *World Journal of Orthopedics*, *12*(7), 467–484. https://doi.org/10.5312/wjo.v12.i7.467

Sari, M. A., Badruzzaman, M., Cherchi, C., Swindle, M., Ajami, N., and Jacangelo, J. G. (2018). Recent innovations and trends in in-conduit hydropower technologies and their applications in water distribution systems. *Journal of Environmental Management*, *228*, 416–428. https://doi.org/10.1016/j.jenvman.2018.08.078

Schnee, J., Stegmaier, J., and Li, P. (2021). A probabilistic approach to online classification of bicycle crashes. *Accident Analysis & Prevention*, *160*, 106311. https://doi.org/10.1016/j.aap.2021.106311

Scoones, I. (2007). Sustainability. *Development in Practice*, *17*(4–5), 589–596. https://doi.org/10.1080/09614520701469609

Sharif, M. N., Haider, H., Farahat, A., Hewage, K., and Sadiq, R. (2019). Water–energy nexus for water distribution systems: A literature review. *Environmental Reviews*, *27*(4), 519–544. https://doi.org/10.1139/er-2018-0106

Siciliacque. (2025). Siciliacque. https://www.siciliacque.it/

Silva, S. N., and Castillo, J. Á. del. (2021). An Approach of the Hydropower: Advantages and Impacts. A Review. *Journal of Energy Research and Reviews*, *8*(1), 10–20. https://doi.org/10.9734/jenrr/2021/v8i130201

Skoropad, V. N., Deđanski, S., Pantović, V., Injac, Z., Vujičić, S., Jovanović-Milenković, M., Jevtić, B., Lukić-Vujadinović, V., Vidojević, D., and Bodolo, I. (2025). Dynamic Traffic Flow Optimization Using Reinforcement Learning and Predictive Analytics: A Sustainable Approach to Improving Urban Mobility in the City of Belgrade. *Sustainability*, *17*(8), 3383. https://doi.org/10.3390/su17083383

Soares, C. O., Pereira, B. F., Pereira Gomes, M. V., Marcondes, L. P., de Campos Gomes, F., and de Melo-Neto, J. S. (2020). Preventive factors against work-related musculoskeletal disorders: Narrative review. *Revista Brasileira de Medicina Do Trabalho*, *17*(3), 415–430. https://doi.org/10.5327/Z1679443520190360

Sustain—Etymology, Origin & Meaning. (2025). Etymonline. https://www.etymonline.com/word/sustain

Tabei, F., Askarian, B., and Chong, J. W. (2021). Accident Detection System for Bicycle Riders. *IEEE Sensors Journal*, *21*(2), 878–885. IEEE Sensors Journal. https://doi.org/10.1109/JSEN.2020.3021652

TomTom Move. (n.d.). Retrieved 30 August 2025, from https://move.tomtom.com/login

Transforming our world: The 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs. (2015). https://sdgs.un.org/2030agenda

UN. Secretary-General and World Commission on Environment and Development (Eds.). (1987). *Report of the World Commission on Environment and Development 'Our Common Future'*. UN. https://digitallibrary.un.org/record/139811

United Nations. (2015, September 25). Historic New Sustainable Development Agenda Unanimously Adopted by 193 UN Members. *United Nations Sustainable Development*. https://www.un.org/sustainabledevelopment/blog/2015/09/historic-new-sustainable-development-agenda-unanimously-adopted-by-193-un-members/

Wang, C., Kou, S., and Song, Y. (2019). Identify Risk Pattern of E-Bike Riders in China Based on Machine Learning Framework. *Entropy*, *21*(11), Article 11. https://doi.org/10.3390/e21111084

Wang, Q., Zhang, C., Ding, Y., Xydis, G., Wang, J., and Østergaard, J. (2015). Review of real-time electricity markets for integrating Distributed Energy Resources and Demand Response. *Applied Energy*, *138*, 695–706. https://doi.org/10.1016/j.apenergy.2014.10.048

WHO. (2022). *Musculoskeletal health*. https://www.who.int/news-room/fact-sheets/detail/musculoskeletal-conditions

WHO. (2023). *Road traffic injuries*. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Xing, J., Wu, W., Cheng, Q., and Liu, R. (2022). Traffic state estimation of urban road networks by multi-source data fusion: Review and new insights. *Physica A: Statistical Mechanics and Its Applications*, *595*, 127079. https://doi.org/10.1016/j.physa.2022.127079

Yi, S., Kondolf, G. M., Sandoval-Solis, S., and Dale, L. (2022). Application of Machine Learning-based Energy Use Forecasting for Inter-basin Water Transfer Project. *Water Resources Management*, *36*(14), 5675–5694. https://doi.org/10.1007/s11269-022-03326-7

Zhang, C., Zhou, Y., Zhang, M., Wang, B., and Nie, Y. (2025). Review and prospect of floating car data research in transportation. *Journal of Traffic and Transportation Engineering (English Edition)*, *12*(4), 752–771. https://doi.org/10.1016/j.jtte.2024.09.005

Zhou, F., Li, L., Zhang, K., Trajcevski, G., Yao, F., Huang, Y., Zhong, T., Wang, J., and Liu, Q. (2020). Forecasting the Evolution of Hydropower Generation. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2861–2870. https://doi.org/10.1145/3394486.3403337

Zhu, S. (2021). Analysis of the severity of vehicle-bicycle crashes with data mining techniques. *Journal of Safety Research*, *76*, 218–227. https://doi.org/10.1016/j.jsr.2020.11.011

Acknowledgements

Al termine di questo percorso di dottorato, mi è ancora più chiaro quanto sia determinante, nella vita, incontrare persone disposte ad essere buoni maestri. Il percorso personale richiede impegno e dedizione, ma sono le buone guide a trasformarlo in un'occasione di crescita profonda. Si può arrivare al traguardo da soli, ma farlo nel modo migliore diventa possibile quando qualcuno cammina accanto a noi, offrendo supporto e ispirazione. Per questo sento il bisogno di ringraziare tutti coloro che, in modi diversi, hanno contribuito a questo risultato.

Un ringraziamento sentito al mio supervisor, **Prof. Salvatore Cavalieri**, per avermi trasmesso i pilastri della buona ricerca scientifica, la passione per questo lavoro e tutto ciò che è stato determinante per la mia crescita professionale.

Desidero esprimere la mia gratitudine al team di EHT. Grazie a **Roberto**, per aver saputo spiegarmi con chiarezza e semplicità ciò che per me era del tutto nuovo nel campo dell'idraulica e dell'energia: senza il tuo aiuto, molte sfide sarebbero state più complesse da affrontare. Grazie a **Vladimiro**, per il supporto costante, per aver contribuito alla mia crescita professionale e personale e per l'umanità che trasmetti nel lavoro, un valore oggi più che mai necessario. Un grazie ad **Alessio**, **Luca** e **Antonio**, per i consigli, il sostegno, la presenza discreta ma autentica, e per l'appoggio sia lavorativo che emotivo.

Grazie ai miei colleghi, e ormai amici, di dottorato. **Salvo** e **Claudio**, grazie per i momenti di leggerezza tra un board game e una relazione semestrale da compilare, per la complicità costruita in breve tempo e per tutti i ricordi condivisi che porterò sempre con me. Grazie ad **Elena**, perché nei pochi mesi trascorsi insieme mi hai insegnato quanto una persona possa dare e ricevere in breve tempo: i nostri aggiornamenti annuali sono la prova che certi legami resistono sempre, con la stessa confidenza e affetto. Ringrazio profondamente **Mariaelena** per la collaborazione e l'intesa che hanno reso più leggere le nostre giornate lavorative e per tutto ciò che abbiamo condiviso, dai tanti impegni in sede alle trasferte, e soprattutto per i sei mesi vissuti insieme nella splendida città di Porto. Il tuo sostegno è stato prezioso, nei momenti difficili come nelle piccole e grandi conquiste quotidiane. La tua presenza costante e

la capacità di costruire insieme obiettivi e ricordi hanno rappresentato per me un punto fermo in questi tre anni.

Ringrazio i miei amici di sempre, **Annalisa**, **Alice**, **Aurora**, **Marta**, **Gabriele** e **Alessio**, per la comprensione mai scontata e per non aver mai sminuito il valore di questo mio percorso. Siete le mie certezze, i miei migliori insegnanti su quanta cura bisogna mettere nello scegliere le persone giuste al proprio fianco. Siamo il villaggio che abbiamo costruito e che ci salverà dalla tempesta, l'esempio che anche strade professionali e città diverse non allontaneranno mai chi si vuole bene davvero.

Ringrazio **Sofia**, per l'intima condivisione delle nostre vite, per la volontà di mantenere saldo questo rapporto nonostante la lontananza. Grazie per i consigli e il sostegno in questi anni. Per me sei un grande esempio e sei famiglia.

Un ringraziamento di cuore ad **Alessandra** e **Andrea**. Grazie per l'amicizia sincera e per il dono inestimabile di avermi voluta al vostro fianco come madrina del piccolo Tommaso. Questo gesto non è solo un onore, ma un grande insegnamento sul valore della responsabilità, della condivisione e dell'amore incondizionato. Grazie per avermi concesso di essere una presenza indelebile nella vita della vostra famiglia.

Un ringraziamento speciale ai miei suoceri **Marinella** e **Ferdinando**, per la loro costante disponibilità, per avermi sempre fatto sentire accolta, e per aver reso ancora più prezioso questo viaggio con il loro affetto.

Un pensiero speciale e affettuoso è rivolto ai miei nonni **Aurora**, **Ottavio** e **Domenico**. Pur non essendo più fisicamente con me, percepisco costantemente la loro presenza, la loro protezione e il loro sostegno, sia nei momenti di gioia sia nelle difficoltà.

Infine, ringrazio profondamente nonna **Liliana**, per il suo supporto costante e instancabile. Per la generosità, la premura e per essersi letteralmente "fatta in quattro" per me nel corso degli anni. Sei l'esempio più concreto e vivo di affetto autentico e dedizione che si possa desiderare.

Questo traguardo non appartiene solo a me, ma a tutte le persone che, in modi diversi, hanno reso possibile questo percorso. A voi va la mia gratitudine più profonda.