Doctoral Dissertation
Doctoral Program in Computer and Control Engineering (35^{th}cycle)

Investigation of innovative IoT technologies for complex manufacturing process modelling and optimization

By

Andrea Bellagarda

******

Supervisor(s):
Prof. Enrico Macii

Doctoral Examination Committee:
Prof. Yukai Chen, Referee, Interuniversity Microelectronics Centre (Belgium)
Prof. Donkyu Baek, Referee, Chungbuk National University (South Korea)
Prof. Elisa Ficarra, Referee, Università degli Studi di Modena e Reggio Emilia (Italy)

Politecnico di Torino
2023

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Andrea Bellagarda
2023

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

I would like to dedicate this Doctoral Program to Amedeo: the new one, who has recently joined us and filled our hearts with love and joy; and the old one, who sadly is not with us anymore, but who has always looked after me in all chapters of my academic and professional journey including, I am sure, this one.

Acknowledgements

"People improve their eyes as they learn" (prof. Hajime Yamashina). "Behind every one of [his achievements], is a whole community of people who have grown up with him. Strengthening him over a lifetime of unimaginable challenges. Each one has been crucial, they've been his friends and family, his allies and rivals, inspiring him and pushing him forward. [He] doesn't imagine for one minute that he's gotten this far on his own." (Muten Roshi) It is with the first quote in mind that I embarked on, and with the second one that I concluded, this Doctoral Program journey, which has undoubtedly been the most challenging chapter of my personal, professional and academic life so far.

First of all, I would like to express my gratitude to the Politecnico di Torino. This Doctoral Program is the last of a number of academic and professional achievements which I have been able to accomplish thanks to the endless opportunities which this institute of higher learning offers to its students, opportunities which go well beyond the exceptional quality of its academic education. Most importantly, attending the ASP - Alta Scuola Politecnica over 10 years ago has given me the opportunity to meet a truly extraordinary community of students, professors and alumni, community in which I have developed some of my strongest and closest friendships. ASP has allowed me to surround myself with extraordinary people, whose example and achievements have not only motivated me to pursue this Doctoral Program, but have and continue to inspire me to constantly improve myself as a professional and as an individial. I would therefore like to thank all my friends from ASP, whose positive influence help me become, everyday, a better man.

I would also like to express my gratitude to my supervisor, Prof. Enrico Macii, and to my mentor, Dr. Luciano Massone, who have given me the opportunity to pursue such a challenging experience and who have greatly supported me in times of need. Dr. Massone's mentorship and challenge to strive for excellence have been an invaluable compass in my professional journey, visionary and coherent since our first meeting over fifteen years ago, during which he had already challenged me to consider pursuing a Ph.D. program.

My deepest gratitude also goes to Prof. Edoardo Patti, Prof. Lorenzo Bottaccioli and Prof. Alessandro Aliberti, for guiding me in my research activities, publications and, most importantly, for constantly keeping the pace and selflessly supporting me in times of need. I would particularly like to thank Alessandro, for the moral support and understanding during this extremely challenging experience. I would also like to thank all the B.Sc., M.Sc. and Ph.D. students who have accompanied me in this journey: Donato, Arsene, Claudia and Alessandro, your work has been fundamental for the final completion of my dissertation.

Finally, I would like to thank my cousin Alberto, whose support during the early phases of this Doctoral Program has been invaluable. Although he is far younger than me, he has grown to become a true role model whom I greatly look up to.

Abstract

The aim of this Doctoral research is to investigate innovative solutions aiming to leverage the potential of Internet of Things (IoT) technologies within the manufacturing environment, especially focusing on complex industrial processes. The research can therefore be considered to fall into the realm of Industry 4.0 (I4.0) and, within this framework, if one looks at the I4.0 clusters which were identified by the Italian Government's "Piano Nazionale Industria 4.0" in 2016, it would aim to focus on the "Big Data & Analytics" cluster.

In order to strengthen the link between academic research and industrial applicability, the research focuses on industrial areas which are considered critical for their impact on manufacturing performance (cost, quality, delivery), taking into consideration also aspects such as readiness, cost, robustness, reliability and flexibility during the investigation. If the costs and, especially, losses of an average industrial company are stratified, for example, the main priorities are usually related to manpower. However, if the affinity with the cluster "Big Data & Analytics" is also considered, Energy results being the main priority. Following this prioritization, the research has focused on highly energy consuming processes, because of the weight of the cost of energy in manufacturing systems.

When integrating the energy topic with the I4.0 cluster of Big Data & Analytics, the research naturally falls under the realm of Smart Grids. Within this realm, better coordination between power demand and supply can be achieved by implementing innovative applications, for example real time forecasting or demand response. For such applications to work effectively though, the ability to effectively forecast both power demand and supply with different forecast horizon has become increasingly important. When looking at innovative solutions for forecasting, with the support of Big Data & Analytics, one enters the Machine Learning (ML) realm: ML methods have gaining significant popularity in the forecasting field, particularly those based on

Artificial Neural Network (ANN) models, which present substantial improvements in forecasting modelling compared to benchmarks. The research therefore focuses on developing innovative solutions for the effective forecasting of both power demand and supply, leveraging ML methods and particularly those based on ANN.

The investigation was originally supposed to be carried out on an industrial building case-study, however due to the Covid-19 emergency access to the site was restricted and the research activities were significantly slowed down. Alternative case-studies were therefore identified, which could easily be replicated in the industrial reality.

In terms of forecasting power demand, the research focuses on energy consumption in buildings due to Heating, Ventilation and Air Conditioning (HVAC) systems, and on their relation to in-door air temperature, developing an innovative solution for the effective forecasting of building in-door air temperature and, as a consequence, of power demand. In terms of forecasting power supply, considering current soaring investments in clean electricity and electrification, particularly solar photovoltaic (PV), the research focuses on developing an innovative solution for effective forecasting of photovoltaic power generation. Both solutions are tested and validated on two different real-life case studies. A common methodology applicable to both case studies is developed, leveraging IoT sensors to collect real, but limited, data, and simulators to generate artificial, but accurate and realistic, datasets large enough to effectively train and test different ANNs. The methodology also includes Transfer Learning (TL) to further improve forecasting accuracy.

Contents

List of Figures

List of Tables

Glossary

1D-CNN 1-Dimensional Convolutional Neural Network

3D three dimensional

ADAM Adaptive Moment Estimation

AI Artificial Intelligence

AM Addictive Manufacturing

AMS Advanced Manufacturing Solutions

ANN Artificial Neural Network

APS Announced Pledges Scenario

AR Augmented Reality

BiLSTM Bidirectional LSTM

BIM Building Information Modelling

C-LSTM Constrained LSTM

CAD Computer Aided Design

CC Cloud Computing

CLSTM Contextual LSTM

CNN Convolution Neural Network

CoBot Collaborative Robot

CPS Cyber Physical Systems

DER Distributed Energy Resources

DOE Department of Energy

DR Demand Response

DSM Demand Side Management

ERP Enterprise Resource Planning

EU European Union

FFNN Feed-Forward Neural Network

GHI Global Horizontal Irradiance

GIS Geographic Information System

GRU Gated Recurrent Unit

HVAC Heating, Ventilation and Air Conditioning

I4.0 Industry 4.0

ICT Information and Communication Technology

IEA International Energy Agency

IIoT Industrial Internet of Things

IoT Internet of Things

LED Light Emitting Diode

LSTM Long Short-Term Memory

MAD Mean Absolute Difference

MAE Mean Average Error

MAPE Mean Absolute Percentage Error

MES Manufacturing Execution System

ML Machine Learning

MLP Multilayer Perceptron

MLP-NARX Multilayer Perceptron with Non-linear Autoregressive Exogenous

MLR Multiple Linear Regression

MSE Mean Squared Error

NAR Non-linear AutoRegressive

NLP Natural Language Processing

NZE Net Zero Emissions

OBS Optimal Brain Surgeon

PMV Predicted Mean Vote

PPD Percentage of Person Dissatisfied

PV photovoltaic

R&D Research & Development

$\mathbf{R^2}$ Coefficient of Determination

ReLU Rectified Linear Unit

RFID Radio Frequency IDentification

RMSD Root Mean Square Difference

RMSE Root Mean Square Error

RNN Recurrent Neural Network

SME Small and Medium-Sized Enterprise

STEPS Stated Policies Scenarios

TL Transfer Learning

TMY Typical Meteorological Year

TPM Total Productive Maintenance

USA United States of America

UV ultraviolet

VLSI Very Large Scale Integration

VRE Variable Renewable Energy

WCM World Class Manufacturing

WLAN Wireless Local Area Network

Chapter 1

Introduction

1.1 New challenges for the manufacturing community

The manufacturing production in Italy and Europe has faced substantial strain and obstacles even before the Covid-19 pandemic, as emphasized by the Italian Confindustria in its 2019 report [7], which emphasises how throughout 2018 manufacturing production appeared to slow down globally. This pattern came after a two-year expansion that had already settled on a growth trajectory that was slower than that seen during the peak years of globalization. The report analyses how the slowdown was caused by a number of economic factors that combined to create an environment of growing uncertainty, many of which still remain today. These factors included the more inward-looking trade policies of the United States of America (USA), the ongoing uncertainty surrounding the potential outcomes of Brexit, the economic tensions between the USA and China, and election-related risks in Europe. The redeployment of value chains globally (which had caused a structural increase in the volume of trade for output units), the dizzying growth of China before its physiological slowdown, the spread of multilateralism, once hailed but later seen as a source of escalating inequality, and other structural factors, however, were also reflected in this slowdown.

In addition, industrial companies are still under pressure from a number of competitiveness issues, such as: (i) high energy costs, particularly those companies with energy-intensive operations; (ii) global market competition, particularly in light of the rise of emerging economies and the easing of international trade; and

(iii) technological advancements, which require industries to continuously invest in Research & Development (R&D), adopt new technologies, and upskill their workforces. To increase their competitiveness in the global market, industries must embrace tactics including energy efficiency measures, innovation, diversification, and teamwork.

For instance, as shown by [7], a sizeable portion of the Italian production system has benefited from qualitative upgrading in order to counteract the rising price competition from the developing world by shifting to market segments with a higher added value content. This is a clearly stated plan that served as a genuine "high road" for the repositioning of the Italian industrial system on global markets. This strategic emphasis shifted concurrently on two levels, manifesting as both vertical diversification (improving the quality of already created items) and horizontal differentiation (differentiating production toward more complex types of goods). Another trend identified by [7] is the reliance on industrial system innovation. Manufacturing can be digitized to improve technical and energy efficiency, boost production flexibility, and enrich the industrial offer of new "intelligent" services. These are all significant potential business benefits. With the help of Industry 4.0 (I4.0) technologies, businesses can make choices more quickly and accurately, engage in new types of man-machine interaction, and integrate their entire internal value chain and, potentially, their entire supply chain. However, Europe runs the risk of losing the global race to Asia and North America, especially in terms of leadership in the provision of enabling technologies for the digital transformation of industry and, more specifically, in the case of patent capabilities related to Information and Communication Technology (ICT). Companies that want to go digital need multi-level support from industrial policy, which encourages technology investments, tighter collaboration between academia and business, and talent development and ongoing skill upkeep. Within this framework, Italy has established a medium-long term policy plan in line with worldwide best practices since 2016, "Piano Nazionale Industria 4.0" [2] (”National Industry 4.0 Plan), albeit lagging behind the other major European nations.

1.2 Industry 4.0

1.2.1 Introduction to Industry 4.0

The term "Industry 4.0" was first used in Germany in 2011 during the Hannover Fair, the largest German industry and automation exhibition. It was coined by Henning Kagermann's research group, which was in charge of laying the groundwork for the German government's strategic initiative "Plattform Industrie 4.0," which was launched in 2013 to support the digital transformation of the German manufacturing industry. As explained by [1], this phrase was then used to refer to the "Fourth Industrial Revolution," which occurred in the first decades of the twenty-first century and was characterized by technological progress related to ICT, advanced sensors, and data analysis as driving elements, representing a further advancement of the three previous industrial revolutions:

- 1st Industrial Revolution (second half of the 18th century) - during which the advent of the steam engine and the flying shuttle enabled mechanization of fabric manufacturing first, and later other sectors, giving rise to the factory concept and profoundly transforming people's lives and ideas about work.
- 2nd Industrial Revolution (second half of the nineteenth century) - marked by significant developments and discoveries in the technical and technological fields, including electricity, steel, the chemical industry, and the internal combustion engine, while Taylorism gave birth to a new notion of production, allowing for mass production and the economies of scale.
- 3rd Industrial Revolution (latter decades of the twentieth century) - often known as the digital or computer revolution, it was enabled by the development of electronics and communication technologies (Internet), as well as the widespread adoption of personal computers; in the industrial setting, the use of these technologies enabled the automation of some tasks through the use of robots.

With the introduction of I4.0, a new age of technical developments and revolutions in the industrial sector has begun. I4.0 promises to transform manufacturing and produce a paradigm change in industrial processes, product creation, and customer experiences through the confluence of physical and digital technologies. This will profoundly impact all aspects of industrial companies: from business models to value creation, from work organization to downstream services. The purpose of this introduction is to investigate the significance of I4.0 by exploring its essential components, perspective benefits, and consequences for various stakeholders, and to provide an overview of the relevance of I4.0 and its future consequences by drawing on academic citations and research.

Because the I4.0 has only just been formed and its environment is still evolving, providing an univocal definition for "Industry 4.0" is problematic, as emphasized by [8]. However, Kagerman defines I4.0 as "[...] the technical integration of Cyber Physical Systems (CPS) into manufacturing and logistics, as well as the use of the Internet of Things (IoT) and Services in industrial processes." [9] [10]

In order to fully comprehend the industrial digital transformation goals, an examination of the current scenario and the reasons behind I4.0 is required before listing the items corresponding to the two categories above. The European Union (EU) plan "For an European Industrial Renaissance" [11] provides a first overview, highlighting the importance of industrial activities, particularly in relation to the value chain that includes Small and Medium-Sized Enterprise (SME) in member countries. The European industry must remain competitive by producing high-quality, innovative products and services in an efficient manner using technology tools that enable high levels of productivity and efficient use of resources. Furthermore, the relevance of the internal market has increased since the integration of companies in regional and global value chains is regarded as critical in order to lower input prices and increase both quality and productivity. To meet such challenges, the manufacturing industry can take advantage of the opportunities provided by I4.0 technologies by developing a more flexible production system that allows it to meet market demands in a short period of time and at costs comparable to those of larger economies, as highlighted by [12].

With this in mind, the Fourth Industrial Revolution, often known as "Industry 4.0", can be considered to be the result of successful integration of CPS, the IoT, cloud computing, and Artificial Intelligence (AI) into manufacturing and industrial processes. It builds on the previous industrial revolutions' foundations and strives to create a highly interconnected and intelligent ecosystem of machines, products, and humans. The basic concept of I4.0 is the interconnection and interoperability of multiple components, allowing for real-time data sharing and intelligent decision-making. Common features of I4.0 systems, as presented by [1], include:

- Cyber Physical Systems (CPS): CPS, which combine physical parts with computational power and connectivity, serve as the backbone of I4.0. Sensors, actuators, and embedded systems are combined with networked communication to monitor and control physical processes in real time. According to [13], CPS enable the collection of enormous volumes of data and the completion of complicated tasks with minimal human participation.

- Industrial Internet of Things (IIoT): The Industrial Internet of Things (IIoT) is a network of interconnected physical devices, sensors, and actuators that allow data exchange and communication between machines and systems. In I4.0, the IIoT is critical in connecting multiple components of the production process, enabling real-time monitoring, predictive maintenance, and operational optimization. [14], for example, emphasizes that IIoT-based solutions enable the monitoring of machine conditions, allowing for predictive maintenance that decreases downtime and losses.

- Cloud Computing (CC):Cloud Computing (CC) provides the infrastructure and resources required for data storage, processing, and analysis. It allows for smooth access to data from several sources, as well as collaborative efforts and scalable solutions. CC, according to [9], provides the processing power and storage capacity needed to handle the huge volumes of data created by I4.0 systems, enabling enhanced analytics and decision-making processes.

- Artificial Intelligence (AI): AI approaches such as Machine Learning (ML), deep learning, and cognitive computing provide intellectual capabilities to I4.0 systems. AI algorithms analyze and interpret data, find trends, and make autonomous judgments, resulting in increased efficiency, productivity, and creativity.

Despite the lack of a clear definition, the fundamental concepts of I4.0 are extensively documented and recognised in literature, and I4.0 can be divided into two parts: (i) I4.0 Design Principles and (ii) I4.0 Key Enabling Technologies, as presented by [1].

1.2.2 Design principles for Industry 4.0

I4.0 Design Principles reflect the requirements imposed by current and future scenarios, with the goal of achieving a flexible, integrated, and efficient production system: they can be used as a guide for the implementation of I4.0 describing the key values beyond the phenomenon as proposed by [15] and presented again by [1].

- Interoperability: Interoperability is a key idea in I4.0 which is a complex ecosystem comprised of advanced technologies that have attained maturity at the same time. All of the actors in the manufacturing process, belonging to various technology categories, must communicate with one another in order to transfer data and information and manage the activities required to be carried out in a plant. IIoT solutions enables the connection of CPS that govern the process, but an open and standardized communication protocol is required for complete integration of different technologies.
- Virtualization: Based on the physical-digital correlation, it is represented by the concept of the Digital Twin. Already a reality in many companies, it consists of the creation of a simulated model of both the product and the plant, capable of interacting with its real world version, to allow for new and more effective simulations in all phases of the topic's life cycle.
- Decentralization: It deals with the decision-making process, which is designed to be autonomous in the I4.0 environment in order to develop a responsive system, where the CPS may take decisions on their own, facilitated by edge computing and AI.
- Real-time Capability: One of the primary benefits of I4.0 is the massive volume of data that can be gathered from equipment and then analyzed to obtain critical information about production levels and parameters. Keeping such data under control in real-time can be a deciding factor in a manufacturing company's success, allowing it to respond to changes quickly or take preventive measures.
- Modularity: Modularity and standardization are well-known ideas to those familiar with Lean Manufacturing, and their importance grows in the context of I4.0 given the previously described developments. As a result, having a modular system is critical to achieving flexibility in order to adapt production

to market demands, being able to reconfigure the line in a short time thanks to the advantages supplied by interconnected programmable machines.

- Service Orientation: It refers to the potential created by the industrial Internet by providing services through the web. It promotes the servitization trend, which is defined as "a process that marks the transition from the pure sale of a product based on traditional transactional logic to the sale of a solution that in fact creates a long-term relationship between customer and supplier." [1]

1.2.3 Key Enabling Technologies for Industry 4.0

The absence of a single, stand-alone enabling component is a continuous contrast between the Fourth Industrial Revolution and its preceding ones. Literature presents a range of technologies, varying from four [15] to thirteen [8], which can be identified as the drivers of the ongoing I4.0 transformation, as they merge to form an innovative ecosystem that serves as a powerful mean to increase the productivity of manufacturing systems and create a new production paradigm. As presented by [1], the enabling technologies chosen as the foundation for the construction of this thesis are part of the "Piano Nazionale Industria 4.0" [2], the first policy adopted by the Italian government in 2016 to address the changing manufacturing scenario at the European and global level. This plan and the subsequent ones ("Piano Impresa 4.0" in 2017 and "Piano Transizione 4.0" in 2020) included economic incentives such as over-depreciation for I4.0 technological equipment, tax credits for research, and the establishment of eight Competence Centres, with the goal of supporting the digitalization of Italian companies and I4.0 investments, increasing the competitiveness of all companies - particularly SME - and providing support for adequate training. The essential enabling technologies outlined in Piano Nazionale Industria 4.0 [2] are now briefly described by [1], with examples of practical implementations.

Fig. 1.1 Key Enabling Technologies - adapted by [1] from Piano nazionale Industria 4.0 [2]

- Advanced Manufacturing Solutions (AMS): include all those machines, commonly referred to as CPS, that are able to interact with each other and with the surrounding reality by detecting and communicating real-time data about their operating status and the operations in progress or that must be performed, also interacting with the product being processed and managing it with partial autonomy. Among the innovative manufacturing solutions are Collaborative Robot (CoBot) examples. They enable collaboration between operators and robots by utilizing stop-at-contact technology to eliminate the risk of impact or crushing, hence reducing the need for barriers or fences. These instruments might be useful for repetitive and non-ergonomic operations like material handling or inspection. FANUC, one of the world's largest manufacturers of industrial robots, provides an example of application in the automotive industry by presenting its CoBot CR35-iB as a tool to improve the ergonomics of heavy handling operations traditionally performed by operators, such as lifting spare tires in vehicle assembly lines.

- Addictive Manufacturing (AM): is a manufacturing technology that uses overlapping layers of material to print a Computer Aided Design (CAD) model in three dimensions (three dimensional (3D) printing). It is the inverse of the classic manufacturing idea, in which the finished product is generated by removing material from a solid. 3D printing can be done with a variety of materials, including plastics and polymeric materials, resins, and metals, and different printing procedures can be used depending on the material and other technical constraints. Rapid prototyping is the most popular application of Addictive Manufacturing (AM) in the industrial environment: components can be created starting from their CAD model already developed during the design phase, requiring less time and lower costs than traditional prototyping techniques. Another example of an application is the use of additive manufacturing to create auxiliary components used in manufacturing, such as supports or specific tools that are not accessible on the market and must be designed in collaboration with suppliers. The main benefits of this approach are cost reduction (because the volumes for these types of components and tools are low, the company cannot rely on economies of scale when purchasing them) and complete control over the availability of these components (even though their economic value is very low, auxiliary components are essential in production and their lack may cause slowdowns in the productive line, with significant losses, so being able to produce them internationally is advantageous).

- Augmented Reality (AR): is an innovation that combines the real, visible reality and everyday world with computer-generated information and superimposes it on the user's actual perception through visual contents, noises, or haptic feedbacks [16]. AR can be used to deliver information to technical employees during maintenance activities, hence speeding up procedures and reducing human errors. Furthermore, this program supports the action of technicians who can be remotely guided by maintenance specialists to address problems on machines that require specialized knowledge, significantly lowering the time of intervention. Another example, specific to the automotive industry, is the use of AR glasses to aid in the control of painting defects at the end of the line: this technology assists the operator in identifying any defects, being able to manage different positionings and defect types, resulting in an accurate control that relies on the precision of technology and human experience, which work in tandem.

- Simulation technologies: they are already a reality in many companies. The plant and all of the process elements are mapped and replicated in a CAD model that can be used to analyze the process and layout of the production line prior to its installation. This technology also incorporates the product's CAD model in order to map all of the operations required to produce the physical product, as well as all of the components required at every workstation, in order to coordinate logistics. This technology is extremely cost-effective: it allows you to simulate the entire production process in a virtual environment where the costs for changes are almost zero before putting the process in place, ensuring that no major adjustments are required when the physical line is set. This concept is further enhanced by innovations such as the Digital Twin, where the virtual simulation of the product and process are able to interact with their physical counterpart, allowing for even more accurate simulations.

- Horizontal/Vertical Integration: it refers to the interconnection of all participants in the production process, both within and outside the company (from customers to suppliers).

- Industrial Internet: or IIoT, it is the IoT in the industrial environment and is built on a bidirectional data exchange, encompassing smart and networked items and devices. All industrial devices are linked to the Internet using a standard communication protocol, and they exchange information about their position, activity and status in real time, allowing data to be accessed both from other machines involved in the process and from higher levels for management and control [17].

- Cloud Computing (CC), which allows the administration of vast amounts of data on open structures.

- Cybersecurity: it refers to sufficient safeguards to prevent cyber risks and hacker attacks, which are a critical issue in smart and networked factories. Two major weaknesses can be identified for the digitalized and connected industry: (i) machine interoperability enables for remote setting of operating parameters and programming of machines without being physically present, and malicious attacks could be carried out by raising the operating parameters above a safe level, jeopardizing operators on the line or the machine's integrity; (ii) because data is the currency of the future due to its expanding relevance in product and process optimization, it must be protected from potential leaks caused by hacker assaults.

- Big Data & Analytics: refers to the vast amount of information that may be gathered and processed within a company's activity in order to uncover meaningful trends, correlations, and phenomena that can be defined starting with sensor recordings. Data collection alone is insufficient; primary analysis is required to extract significant information, reduce noise, and correlate the obtained data with actual events, changing plain data into Smart Data that is useable and adds value to the firm.

1.2.4 Smart factory

According to [1], the notion of Smart Factory, which symbolizes the ideal vision of factories in the future, is the most complete manifestation of I4.0. All the industrial processes and the actors in charge of managing and carrying them out are interconnected in the Smart Factory environment due to the integration of ICT across the whole production process. The goal of this type of interconnected system, according to the I4.0 Design Principles, is to provide a flexible system that can manage the arising complexity in production with virtually no waste of resources and a reduced lead time, in order to strengthen the company's competitiveness and increase the level of service. For this to happen, the Smart Factory must be vertically and horizontally integrated.

Horizontal integration refers to what happens within the manufacturing plant, taking into account smart production systems based on CPS as well as smart goods that collaborate and exchange data via IIoT. As a result, the system is driven by the product itself, which is fitted for example with Radio Frequency IDentification (RFID) tags that can interact in real-time with the machines that produce it, providing exact information about its location within the plant, its status, and the manufacturing stages still to be completed. Machines can make their own decisions, using the information embedded in the smart product to generate the desired output, allowing the system to be flexible and responsive, adapting production based on external parameters such as market demand or inventory level, and using sensors to constantly monitor the process. Horizontal Integration in a Smart Factory is defined by the constant and real-time data exchange between all aspects of the production process.

With this in mind, to support data-driven strategic decision making, all of the core activities identified by the Porter Value Chain require information transparency.

Furthermore, the entire value chain is included in a fully horizontal integrated system, and data exchange occurs between external partners who work with the manufacturing company to produce subcomponents or services.

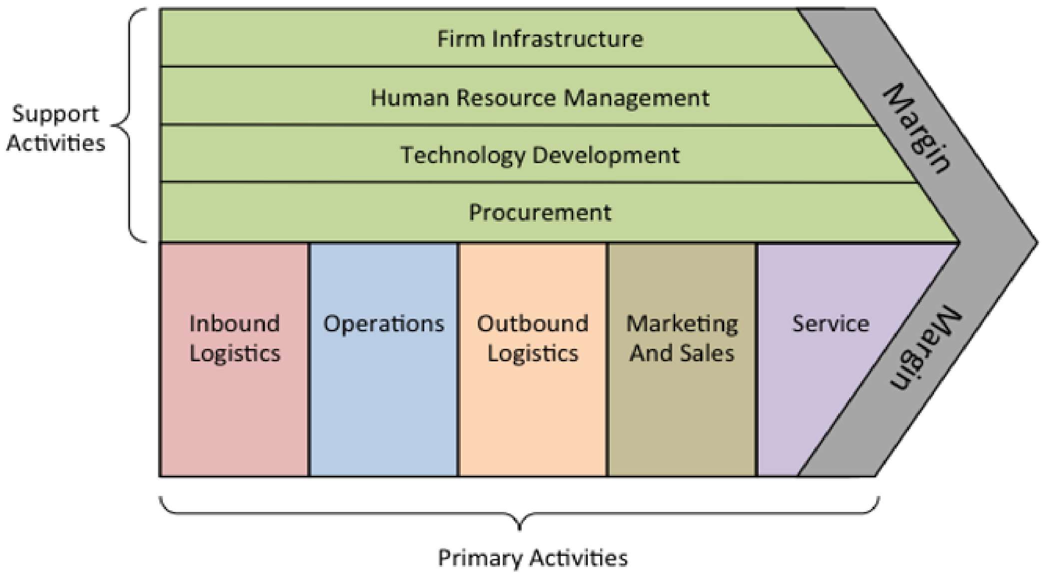


Fig. 1.2 Porter Value Chain - presented by [1] and originally from [3]

Vertical Integration, on the other hand, is a concept that extends beyond the operations boundary, encompassing all support activities within the company: technology development (ICT, R&D), human resources management, procurement, corporate (which is used to identify all other departments of the company, such as accounting, finance, legal, and general administration). Data from production is exchanged throughout all company departments in order to integrate information in the entire industrialization process, beginning with R&D and continuing through every level of the organization. Even if support activities are not directly involved in the generation of value, information transparency and real-time data exchange within departments allow them to perform more efficiently and provide targeted support to key operations.

The demonstrative and research platform established by the non-profit organization "Technology Initiative SmartFactory" in 2005, as reported by [4], is an example of Smart Factory application. The goal of this project was to design and build an independent manufacturing facility where the newest ICT technologies could be used to experiment and develop an I4.0 system, analyze the potential outcomes, and lay the groundwork for the factory of the future. The inaugural project, for instance, consisted of a production line producing colored liquid soap outfitted with sensors and actuators linked via various wireless communication technologies such as Wireless Local Area Network (WLAN), Bluetooth, and RFID. This type of connectivity allowed the machines to be "plug and play" because they did not require any wiring other than the power supply: this followed the Modularity Design Principle of I4.0 making the system extremely versatile and flexible, allowing modules to be easily replaced in the event of line changes. The items, which are made on a special order, drive the process. Each bottle contains an RFID chip that stores the order and product data, as well as the production procedures that must be followed, such as which liquid, lid, and label are required to manufacture the desired product. The chip inserted in the bottle enables product-to-machine and machine-to-machine communication at each station, telling the manufacturing facility which processes must still be completed, while a computer checks whether the bottle was made in accordance with the order specifications. In terms of ICT, the line is integrated within a wireless ecosystem that covers every level of the automation pyramid, beginning with the field level, in which the manufacturing process and all related operations (such as maintenance and logistics) are monitored and carried out by machines and devices using IIoTand ICT technologies. The field and shopfloor levels are then linked at a higher level with the Manufacturing Execution System (MES), the software that manages the plant in an automated manner, integrating data about production levels, key performance indicators, and parameters to control the machines, and allowing product traceability throughout the manufacturing process. Finally, additional connectivity allows such a system to interact with the Enterprise Resource Planning (ERP) software, which unifies all business processes within the company, greatly supporting management.

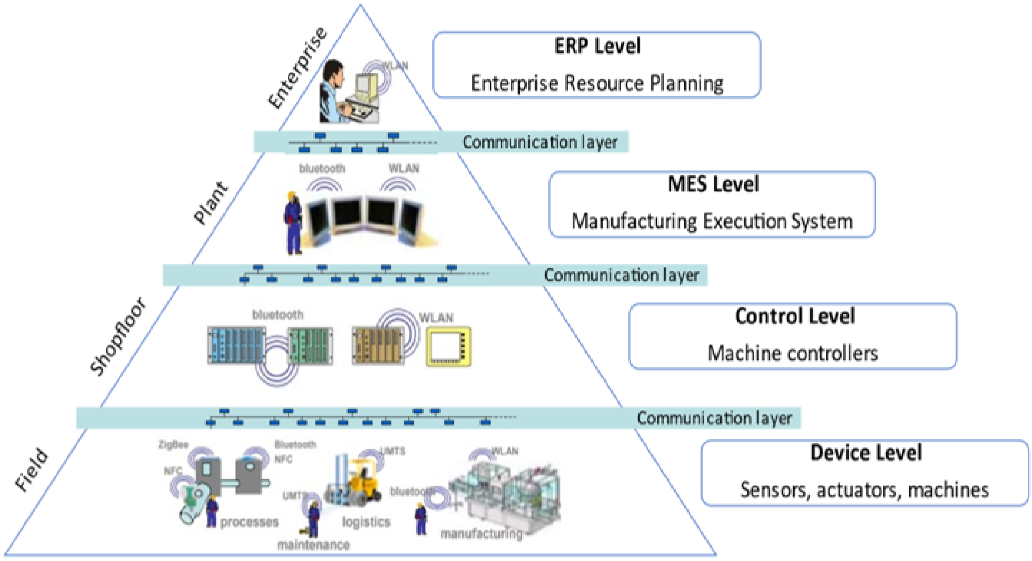


Fig. 1.3 Automation pyramid - presented by [1] and adapted from [4]

1.2.5 Digitization vs Digitalization vs Digital Transformation

Giving a precise description of the terms "Digitization," "Digitalization," and "Digital Transformation" is critical to understanding the meaning of the actions that must be made to transition to the new industrialisation period. These concepts refer to three distinct processes that deal with the shift from traditional production to a fully integrated and digitalized sector. As presented by [1], they can be viewed as sequential steps that can be followed to varying degrees in order to innovate. However, while a definition for "Digitization" and "Digitalization" can be found in Gartner's Information Technology Glossary [18], "Digital Transformation" still does not have a clear definition in literature.

- "Digitization" refers to the process of converting analog data to digital data. It is the first step toward smart data use, but it will not improve industrial processes unless it is accompanied by additional changes in the organization and the way the company uses data. [18]
- "Digitalization" is defined as "[...] the use of digital technologies to change a business model and provide new revenue and value-producing opportunities." It is the transition to a digital business. [18]

- "Digital transformation" instead refers to the phase after digitalization: a new process dealing with people's mindsets and business culture must be initiated in order to transform the organization and maximize the new opportunities provided by the digital approach throughout the entire process and operations. Digital transformation is defined as "changes triggered by digitalisation when implemented strategically and transforms the enterprise's business model, market relationships, and/or organizational arrangements." [19]

The term "digitalization" is therefore used in this work to describe the key step in the process of transitioning from traditional manufacturing to the I4.0 paradigm, primarily referring to the use of data-related technologies that allow the collection and extraction of new information from the manufacturing process and its better use, thus allowing the company to create more value and improve production performance.

To illustrate the distinction between these phases, [1] considers the daily process monitoring operations carried out in a manufacturing plant to track the productivity of equipment or labor. Paper supports have traditionally been used to record the major events that occur during working shifts, as well as information regarding productivity and the amount of manufactured goods. The first and most important step toward change is thus digitalization, which involves replacing physical supports with digital ones. This first step, digitization, in and of itself, is insufficient; it must be supported by the digitalization process, which consists of establishing a robust data collection system and a wise management system to exploit it, with the goal of bringing benefits to the company in terms of productivity, competitiveness, and cost-efficiency. Finally, the digital transformation begins when the company sets strategic initiatives to fit the new possibilities enabled by digitalization (e.g., the introduction of remote control of the process, which can only be achieved if data is entirely collected digitally and proper systems are present within the company in terms of ICT infrastructure, competences and data management).

1.2.6 Potential Benefits of Industry 4.0

One of the most significant benefits that I4.0 may bring to those who choose to invest in this transformation path is clearly increased efficiency and production. Automation, real-time monitoring, and predictive analytics are enabled by I4.0 technologies, resulting in enhanced efficiency and productivity. Machines can interact and optimize operations on their own, saving downtime and operational expenses. According to [20], I4.0 enables the integration of data from many sources, allowing for real-time process optimization and efficient resource usage, ultimately leading to increased production. Improved quality and customisation are also advantages: the interconnection of I4.0 systems enables real-time quality monitoring and feedback loops, ensuring constant product quality. Numerous case-studies show that I4.0 enables the installation of intelligent quality control systems capable of detecting errors and deviations early in the manufacturing process, resulting in higher product quality and reduced waste. Furthermore, the adaptability of these technologies enables customized production, to satisfy particular consumer requirements. Also, by providing end-to-end supply chain visibility and optimization through real-time data interchange, predictive analytics, and demand-driven production, I4.0 solutions can also contribute also to Supply Chain optimization, by enabling better inventory management, shorter lead times, and higher customer satisfaction. I4.0, according to [21], enables real-time coordination, synchronization, and optimization of processes across the whole value chain.

Another advantage is expedited innovation and time-to-market. I4.0 promotes an innovative culture by providing the tools and technologies required for rapid prototyping, simulation, and virtual testing. I4.0 decreases time-to-market and promotes faster innovation cycles by encouraging collaboration and agile product development procedures. I4.0 enables virtual design and simulation, allowing businesses to test and enhance product concepts prior to actual production, avoiding traditional losses due to those processes which take longer and costs more money, such as traditional prototyping.

However, the implications and requirements of implementing I4.0 solutions must be thoroughly recognized, particularly when it comes to the transformation of organizational competencies. I4.0 causes major changes in the workforce, necessitating the acquisition of new skill sets and capabilities. While routine operations are increasingly mechanized, there is a growing demand for people that are digitally literate, data analysts, and problem solvers. This creates a continuous requirement for workforce upskilling and continuous learning in order to adapt to the changing needs of I4.0 and to ensure that the staff can properly exploit the available technology.

As a result, I4.0 is a revolutionary force with enormous implications for the manufacturing and industrial sectors. I4.0 enables increased efficiency, higher quality, and faster innovation by merging CPSs, IoT, CC, and AI. However, implementing I4.0 also implied difficulties, such as personnel upskilling, business model reinvention, and data security. Nonetheless, the potential benefits of I4.0 still encourages companies to embrace and harness its revolutionary power in order to remain competitive and thrive in the digital age.

Within this framework, the research of this Doctoral Program focuses mainly on applications in the spheres of IIoT and Big Data & Analytics.

1.3 Aligning Industry 4.0 benefits with company strategy

In order to strengthen the link between academic research and industrial applicability, the research focuses on industrial areas which are considered critical for their impact on manufacturing performance (cost, quality, delivery), taking into consideration also aspects such as readiness, cost, robustness, reliability and flexibility during the investigation.

First of all, it is important to highlight the potential benefit which I4.0 applications, and particularly digital transformation, can bring to increasing an industrial system's competitiveness. Lessons learned from past projects and applications have demonstrated that unless I4.0 solutions are developed within a clear continuous improvement framework (such as Lean Production, TPM, WCM, . . .), rather than increasing the competitiveness of the industrial system they can have the opposite effect, for example by digitalizing its losses. On the other hand, if applied on top of a clear continuous improvement framework, I4.0 can help fine-tune improvements by identifying losses which were previously impossible to identify, or by attacking losses which were previously impossible to attack. For example, as presented by Figure 1.4 and [5] highlights how, in terms of performance, industrial continuous improvement programs can generate savings up to 10% a year, onto which I4.0 and Digitalization can add a further 2%. [5] presents, in this case, data related to an industrial continuous improvement program, World Class Manufacturing (WCM), applied in a global automotive company [22].

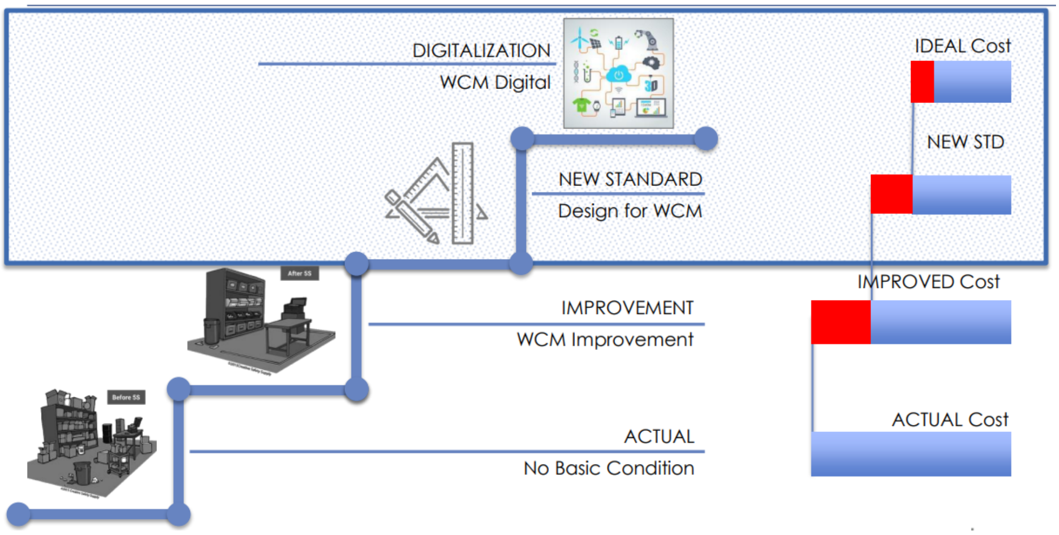


Fig. 1.4 WCM evolution steps and digital, as presented by [5]

If the costs and, especially, losses of an average industrial company are stratified, for example, the main priorities are usually related to manpower, as presented by Figures 1.5 and Figure 1.6. However, if the affinity with the cluster "Big Data & Analytics" is also considered, Energy results being the main priority. Following this prioritization, the research area focuses on highly energy consuming processes, because of the weight of the cost of energy in manufacturing systems. The investigation therefore aims to leverage I4.0 solutions for Smart Energy applications in the manufacturing environment.

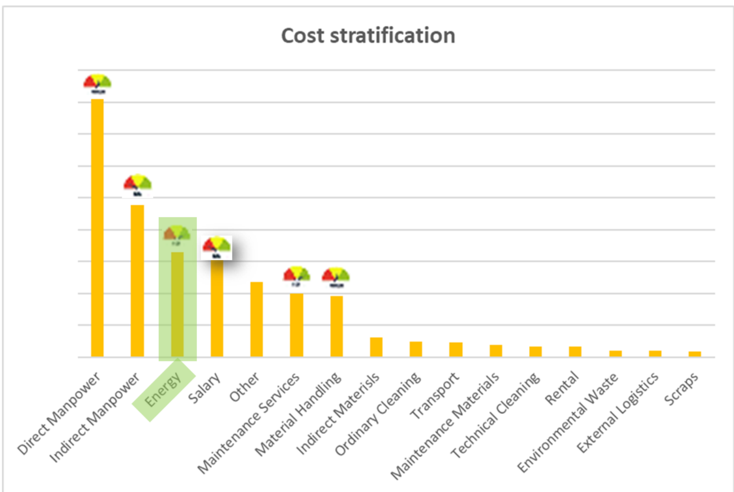


Fig. 1.5 Example of manufacturing costs stratification

Fig. 1.6 Example of manufacturing losses stratification

By deciding to work in the energy sphere, the investigation of this Doctoral investigation can potentially benefit industrial companies not only from the economic and competitiveness point of view, but also in terms of environmental sustainability. According to [23], the environmental problems directly related to energy production and consumption include air pollution, climate change, water pollution, thermal pollution, and solid waste disposal. For example, the emission of air pollutants from fossil fuel combustion is the major cause of urban air pollution, and burning fossil fuels is also the main contributor to the emission of greenhouse gases. As presented by [24], energy derived from fossil fuels contributes significantly to global climate change, accounting for more than 75% of global greenhouse gas emissions and approximately 90% of all carbon dioxide emissions. According to [25], due to their high energy density, fossil fuels are the primary energy source worldwide; however, fossil fuel combustion produces greenhouse gases; approximately 35% of greenhouse gases are emitted by existing power plants. [26], on the other hand, presents that China's coal-fired power plants emit 42% of nitrous oxides and 38% of sulphur dioxides, for a total of 40% of the heat-trapping greenhouse gases, thereby increasing global temperature. As presented by [24], over 300 natural disasters were caused by climate change in 2018, affecting more than 68 million people and causing approximately $131.7 billion in economic losses, with storms, wildfires, floods, and droughts accounting for 93%.

In order to reduce the environmental impact of their energy needs, companies and societies as a whole can follow the following paths: (i) reducing energy demand by reducing and/or optimizing energy consumption; (ii) better synchronize energy demand and supply, to avoid energy waste and thus energy over-production; and (iii) migrating to more sustainable energy sources. The topics investigated by this Doctoral Research will potentially be able to allow industrial companies to improve their performance in all three areas, therefore reducing the environmental impact of their energy needs, and thus potentially benefiting them not only from the economic and competitiveness point of view, but also in terms of environmental sustainability.

The investigation was originally supposed to be carried out industrial case-studies, however due to the Covid-19 emergency access to such sites was restricted and the research activities were significantly slowed down. An alternative was then found by using other case studies for which some data was already available, and which could easily be replicated in the industrial reality.

1.4 Energetic challenges and smart energy systems

1.4.1 Energetic difficulties

As mentioned in earlier sections, one of the major issues that industrial organizations have been facing is rising energy expenses, particularly those with energy-intensive activities. According to [7], energy is a major cost issue in Europe (the EU Commission estimates that wholesale power prices are around 30% higher than in the US, and gas prices are more than 100% higher). Because the industrial sector is so reliant on energy supplies for its operations, energy prices are a crucial competitiveness factor. Energy costs have steadily risen over the last decade, consequent to a variety of causes such as rising demand, geopolitical conflicts, and environmental laws [27].

Increased global energy consumption, geopolitical conflicts and supply interruptions, environmental laws, and the transition to renewable energy sources are all factors leading to increased energy costs. According to [28], the rapid rise of emerging economies and the expanding global population have also resulted in an increase in energy consumption. The increase in energy prices is of course strongly correlated to increasing demand, putting a considerable strain on sectors with energy-intensive operations. Furthermore, political insecurity, conflicts, and disruptions in energy-producing countries have an impact on the supply and availability of energy supplies, driving up prices. Historical data clearly shows how large oil-producing countries, for example, have periodically caused supply shortages and price surges. Finally, as illustrated by [29], stringent environmental restrictions aiming at lowering greenhouse gas emissions have compelled the use of cleaner energy sources, frequently at greater costs. The move from fossil fuels to renewable energy necessitates significant investment and infrastructural modifications, all of which add to the overall energy cost burden.

The growth of energy prices suffocates economic growth and development by diverting resources away from productive investments and limiting the potential to expand industrial activity. Industries lose competitiveness, stifling job creation and innovation. Rising energy costs reduce business profit margins, particularly in energy-intensive industries such as manufacturing, mining, and transportation. Higher production costs hinder enterprises' capacity to compete on a global scale, potentially resulting in job losses and economic stagnation. Energy costs also disproportionately affect SMEs, which frequently lack the financial resilience to tolerate significant cost rises. Such enterprises are at risk of closing or reducing their operations, undermining entrepreneurship and innovation.

To alleviate the impact of these energy difficulties, businesses can turn towards process optimization and, especially, more sophisticated technologies and energy management systems, which can result in significant energy and cost savings. Furthermore, encouraging energy source diversification, including increased use of renewable energy, can ensure long-term cost stability for businesses. Governments and industry stakeholders, on the other hand, must work together to accelerate the development and deployment of sustainable energy technologies while assuring their affordability and reliability. Through regulatory interventions and incentives, governments play a critical role in minimizing the impact of energy costs on companies. Tax breaks, subsidies for energy-efficient devices, and funding for research and development activities can all help to create a more sustainable and economical energy landscape. According to [7], the current EU Commission has made the creation of an energy union one of its main goals, with five objectives: i) to ensure the supply of all types of energy sources (especially oil and gas); ii) to develop an integrated and competitive energy market; iii) to promote energy efficiency; iv) to reduce carbon dioxide emissions; and v) to support innovation in the European energy sector.

Energy cost increases offer significant issues for enterprises, influencing profitability, competitiveness, and overall economic growth. Understanding the elements that contribute to this problem is critical for developing effective mitigation methods. Industries can navigate the complicated energy landscape and maintain sustainable growth by prioritizing energy efficiency, diversifying energy sources, and enacting supportive legislation. Collaboration among governments, corporations, and research institutes is critical in developing new solutions to the problem of rising energy costs.

1.4.2 Smart energy systems

Smart energy management provides businesses with numerous chances to optimize their energy consumption and cut costs. Energy monitoring and analytics are two examples of smart energy management opportunities for industries. Implementing real-time energy monitoring systems and advanced analytics can help industries identify energy usage patterns, detect inefficiencies, and make informed decisions for energy optimization, as demonstrated by [30]. Demand Response (DR) programs, which allow enterprises to modify their energy usage in response to grid circumstances and price signals, are another intriguing application. According to [31], this can enable them to minimize peak demand, optimize energy use, and win financial incentives. [32] also recommends incorporating energy storage technologies, such as batteries or flywheels, in smart grids, which allow industries to store excess energy during off-peak periods and use it during peak periods, lowering energy costs and providing grid stability. Renewable energy integration, which occurs when industries install on-site renewable energy generation systems, such as solar panels or wind turbines, to offset their reliance on the grid, lower energy costs, and reduce environmental impact, can also be a solution, according to [33]. Finally, as cited by [34], integrating energy-efficient technologies and practices such as Light Emitting Diode (LED) lighting, energy-efficient motors, and improved Heating, Ventilation and Air Conditioning (HVAC) systems can dramatically cut energy consumption and operating costs for industries.

Companies are facing the difficulty of efficiently managing electricity usage in smart industries and smart cities [35]. This frequently results in the installation or integration of efficient smart grids [36]. Smart grids enable users to assess and manage electricity consumption in real time. Indeed, they can identify periods of high demand and manage energy distribution accordingly using data analytics and ML techniques. To balance energy needs, sources of alternative energy generation, such as solar energy sites, can be erected. Furthermore, thanks to the utilization of sensors, predictive maintenance and remote controls are conceivable.

Smart grids and the technological advancements they incorporate can therefore improve the administration and stability of power grids. Also, innovative solutions such as communication networks and modern distributed computing facilities can further be integrated to achieve this goal [37]. Within the framework of a smart grid, the implementation of innovative applications such as real-time forecasting or DR, can be leveraged so that power demand and supply can be better coordinated, and power demand and consumption can be adjusted to align with the available power demand, and vice versa [38]. For such applications to function effectively, however, it has become increasingly essential to forecast power demand and supply with varying time horizons.

When integrating this need, forecasting of both power demand and supply, and its challenges with the I4.0 cluster of Big Data & Analytics mentioned in the previous sections, one can clearly understand how technology and innovation can support industries to overcome these challenges. When looking at innovative solutions for forecasting, with the support of Big Data & Analytics, one enters the ML realm. As presented by [39], ML methods are gaining popularity in the forecasting field. Of the different ML methods, [39] highlights how particular attention must be given to those based on Artificial Neural Network (ANN) methods, which present substantial improvements in forecasting modelling compared to benchmarks.

The research therefore focuses on developing innovative solutions for the effective forecasting of both power demand and supply, leveraging ML methods and particularly those based on ANNs.

1.5 Energy demand and supply forecast applications

An important part of the research activity is to select the best case-studies where to apply the investigation, in order to strengthen the link between academic research and industrial applicability. As previously mentioned, the investigation was originally supposed to be carried out industrial case-studies, however due to the Covid-19 emergency access to such sites was restricted and the research activities were significantly slowed down. An alternative was then found by using other case studies for which some data was already available, and which could easily be replicated in the industrial reality.

In terms of forecasting power demand, the natural choice would be to focus on those manufacturing processes which require the most energy to operate. However, since due to the Covid-19 emergency access to industrial sites, and therefore their processes, was restricted, new viable solutions had to be identified. After an initial study, it was found that in the EU, more than 40% of the overall energy consumption and about 36% of total pollution is attributable to buildings, as reported by the EU Directive on the *Energy Performance of Buildings* [40]. Furthermore, with this sector undergoing a constant expansion, its energy consumption and CO_2 emissions are also consistently rising [40]. Further investigations revealed that there are still strong margins for improvement in energy management in this sector, since centralized heating systems with out-of-date control strategies are still present in many building structures [41]. Also, almost half of the overall amount of building energy consumption can be attributed to the HVAC systems [42]. These systems are primarily controlled by the interior air temperature, with the ventilation system acting during the summer months, while the heating system acts in the winter months. If an indoor air-temperature forecast system is available, with the capability of taking into consideration variables such as building occupancy, environmental and external conditions, seasonal change and other influencing factors, a significant contribution to the application of smart energy management for building HVAC systems can be made, thereby improving the energy efficiency of the building. Effective indoor air-temperature prediction systems, for instance, can be used to predict energy and cost requirements in advance and as input for load forecasting problems [43]. An interesting opportunity arose to carry out the investigation on a real-world building located in Turin (Italy). As a consequence, the first part of the Doctoral research focused on developing an innovative solution for the effective forecasting of building in-door air temperature and, as a consequence, of power demand, leveraging ML methods and particularly those based on ANNs.

On the other hand, in terms of forecasting power supply, the challenge was again to identify those areas which could bring the most benefit to industrial realities. As highlighted by the International Energy Agency (IEA), in its 2022 World Energy Outlook Report [44], with the world in the midst of a global energy crisis, and faced with energy shortfalls and high prices, there has been a rush to try and secure alternative energy sources and supplies, and an acceleration in the flow of new renewables projects. Investments in clean electricity and electrification, particularly solar photovoltaic (PV), have been soaring. Within this framework, the increasing significance of sources of renewable energy such as PV energy is evident. PV energy falls under the category of Variable Renewable Energy (VRE) sources due to its fluctuating power output derived from solar energy. The PV power's variable character poses a challenge to its use in power grids, which are very dependent on the relation between the generation and consumption of energy for their stability, because of its lack of reliability. According to the IEA, the stability of a power system depends on its ability to respond rapidly, within economic constraints, to large fluctuations in supply and demand by ramping down production when demand decreases and rising it again when demand increases, for scheduled and unscheduled events [45]. However, the concept of power systems and their flexibility must now be redefined due to the increased presence of power production coming from variable and difficult to predict sources, such as wind and solar, which creates uncertainty from a supply point of view [46]. Smart grids and the technological advancements they incorporate can however improve the administration and stability of power grids. Also, innovative solutions such as communication networks and modern distributed computing facilities can further be integrated to achieve this goal [37]. Within this framework, the implementation of innovative applications such as real-time forecasting or demand response can be leveraged so that power demand and supply can be better coordinated, and power consumption can be adjusted to align with the available power demand, and vice versa [38]. For such applications to function effectively, however, it has become increasingly essential to forecast power demand and supply with varying time horizons. If the ability to effectively predict PV power generation becomes fundamental [47], according to [48], [49] and [50], ML for PV and wind power generation forecasting have become increasingly effective. An interesting opportunity arose to carry out the investigation on a real-world PV installation located in Turin (Italy). As a consequence, the second part of the PhD research focused on developing an innovative solution for the effective forecasting of PV power generation, leveraging ML methods and particularly those based on ANNs.

Chapter 2

Neural Networks

2.1 Time-series forecasting

As presented by [6], the availability of vast quantities of data is the lowest common denominator among all application disciplines. These data, when properly organized, processed, and analyzed, permit the enhancement of existing services and the expansion into unexplored paradigms. The study of available data to predict the future is unquestionably one of the most promising methods in the discipline. Future prediction entails the capacity to implement new and more effective control strategies.

The past must be comprehended to uncover the future. Since the beginning of humankind, we have sought laws that explain the behavior of observed phenomena. There are numerous examples, ranging from the operation of the heart and the irregularity of a pulse to the simulation of the financial market's volatility. In other words, human nature is constantly attempting to foretell the future by attempting to comprehend and model the past [51].

If there is a well-established underlying understanding of deterministic equations, argues [6], they can typically be solved to predict the outcome of an experiment. In contrast, if the equations are unknown, we must determine both the rules regulating system evolution and the current state of the system in order to make a prediction. For instance, behaviours such as a pendulum's movement contain the potential to predict its future movements based on the understanding of how it oscillates, without requiring understanding of the overall apparatus. On the other hand, occurrences such as the air temperature of a structure or the power generation of a photovoltaic (PV) installation are influenced by numerous external factors which conventional systems can model with extreme difficulty. These necessitate the investigation and comprehension of underlying mechanisms. Humans have developed two methods for analyzing time series with which they are unfamiliar: the series can be either i) "understood" or ii) "learned from". To comprehend the series is to demonstrate an explicit mathematical comprehension of how systems behave. Rather, learning from the series involves adopting algorithms, methods and tools that can mimic time series' anatomy. The two methods pursue the same objective of so-called time-series analysis [51] to explain observations. Time series analysis typically has three objectives, as presented by [52] and by [6]:

- forecasting - to foresee, in the most accurate way, how the system will evolve (so with the least probable error);
- modelling - to find a parametrization capable of accurately describing and replicating the characteristics of the system's behavior;
- characterization - to determine fundamental features, such as a system's level of randomization or number of degrees of freedom, with little or no a priori knowledge.

Before the 1920s, the global fit was used to extrapolate the series in order to effectively forecast [51]. Then, the autoregressive technique to forecast the annual number of sunspots was invented by Yule in 1927, thereby initiating the modern era of time-series prediction [53]. Through this method, the next value cab be predicted based on a weighted sum of the series' preceding values. Then, as [6] highlighted, two significant developments occurred around 1980, both of which were enabled by the widespread availability of powerful computers that made it possible to record much lengthier time series, exploit them with algorithms of increasing complexity, and interactively visualize their outputs and results. The first was the reconstruction of state space by time-delay embedding [51]. Based on differential topology and dynamical systems, this provides a method for determining when deterministic governing equations have produced a time-series. The second was the development of the Machine Learning (ML) field, exemplified by neural networks that can investigate a vast variety of possible models in an adaptive manner [54].

With data-driven methods attracting the change in artificial intelligence, and the availability of instruments that store an amount of data numerous orders of magnitude greater than before, [6] highlights how time series were able to be analyzed using machine-learning techniques that require relatively large data sets. This has enabled the development of increasingly efficient and robust methods, and has allowed the uncovering of concealed and non-linear relationships between data, thereby generating new ideas.

Some contexts which can gain the greatest benefits from these instruments are smart cities and smart factories. In these realities, vast quantities of data are collected and stored, the majority of which correspond to time series. We are now able to discover new correlations and dependencies by leveraging time series methods to effectively accumulate and analyze this data. This enables us to gain a deeper understanding of the context (i.e., the factory and all its actors), allowing us to generate novel hypotheses and control policies [55]. In the intelligent energy environment, for instance, methods such as Demand Response (DR) [56] and Demand Side Management (DSM) [57] allow to continuously improve the management of energy supply and demand; furthermore, these new methodologies allow energy resources to be transformed and managed based on requirements in a smart environment. This occurs in all sectors, including energy, health, mobility, and safety. Smart contexts are made possible by the ideal combination of data availability and ML techniques, such as neural networks. In turn, these constitute the ideal laboratory for continuous technological innovation.

2.2 Time-series forecasting with machine learning

The urge to understand and predict the future has fascinated humans since the beginning of time. Additionally, the capacity to acquire new skills through experience is a component of human development. Indeed, when a person is born, he has no knowledge and cannot provide for himself. But as he acquires experience, he becomes increasingly independent and capable of performing increasingly complex tasks every day. The combination of these two abilities, along with the advancement of knowledge and technology, gave rise to advanced methodologies. One of the results of these advances is ML, a collection of methods which leverage algorithms and statistical models, enabling set actions to be carried out by computers without explicit instructions, exploiting patterns and inference instead [58]. Consequently, as presented by [6], these advanced techniques, which combine statistics and computer science, enable computers to understand how a given activity can be performed without requiring a pre-existing and specific program, just as a person does during when learning and developing. A person's ability to recognize the difference between a dog and a cat, for example, is developed as the person recognizes the features that characterize the "dog" and "cat", understanding the differences between their distinctive characteristics (e.g., shape of head or body, size, wars, etc.) and, as a result, having the ability to identify the two.

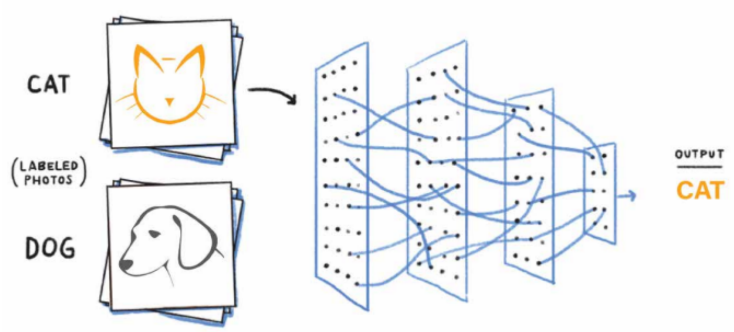


Fig. 2.1 Diagram of ML approach - Cat and dog recognition, presented by [6]

In figure 2.1, a basic example of ML is presented by [6], with the identification of two classes of images occurring by leveraging different layers of filters. The left side shows the inputs to the system. In this example, the inputs are images of the two animals with their corresponding labels. The system is taught, through these images, to distinguish the two species. The main section portrays the actual ML system, which through filters, weights and other instruments, understands how to identify the figures in images. When the training is complete, the algorithm is able to distinguish, with a precision error, each the different figures in the image (e.g., cat and dog) independently.

In a system where it is autonomously learning, a machine that has been programmed to learn looks inside the data to discover statistical patterns. Such capability enables it to identify classes independently [59]. As explained by [6], by studying numerous images of cats and dogs, for instance, the machine could extrapolate knowledge from the image data and understand autonomously that cats have shorter faces, while the size of dogs can be very variable. Such characteristics are capable of being coded in a dataset, where fundamental mathematical elements such as matrices and vectors are used to organize data. Every dataset is typically characterized by a number of variables that describe the features (or observable properties) of an object. Feature vector refers to the list of characteristics that characterize an object in ML systems. Furthermore, the data present in the samples often carries a wealth of information, which must be reorganized in order to prevent redundancy, a phenomenon in which multiple features provide the same information. This is undesirable since it can reduce the precision and dependability of these ML systems.

When describing the ML phase, [6] explains how trends within the features have to be identified by the machine, and a function for recognizing future input must also be established. In addition, considerations such as the right dataset to select, or how to implement past learning into future decision-making, have to be completed by the programmer. There will be blunders during this phase, but the more the learning algorithm is able to receive data, the greater the accuracy with which it will be able to predict or classify. Consequently, it is crucial that machine learns from a dataset as large and as feasible as possible, in order to increase generalization capabilities, but not so large that the algorithm cannot apply the "laws" that govern it.

Therefore, according to [6], ML represents a revolutionary shift in the paradigm of computation. Extraction of knowledge from examples and experience, interpreted through examination of real data, is the epitome of learning process based on induction. Currently, ML cab be characterized into three major types, as presented by [60] and [6]:

- supervised learning;
- unsupervised learning;
- reinforcement learning.

As explained by [6] supervised learning aims to approximate the mapping function between input and output, in such a way that it is possible to determine the output variables for new input data. Two categories of problems are addressed by supervised learning: regression and classification [61]. In contrast, in unsupervised learning, the data used are not labeled, and thus there is no prior knowledge. As a result, data is only used to identify correlations, patterns, or structures within the information itself. Clustering and dimension reduction are the most frequently employed techniques [62]. In clustering, [6] explains how classification is achieved by discretely categorizing data based on similarity in the same class and differences between separate classes, whereas dimensionality decrease achieves the same for continuous data. With a reinforced learning environment, the training phase concludes with the system receiving feedback on its actions [63]. Such learning approach, explains [6], leverages environmental interaction to discover different optimal sets and actions by storing and managing past experiences. The algorithm adopts decisions (i.e., prediction or classification) based on the actions deemed virtuous or detrimental, using the feedback received during the learning phase. In general, [6] explains how a robust ML model requires the following:

- the available data must be analyzed in-depth;
- optimal pre-processing of the dataset must be completed;
- the dataset must be carefully split into a training set, a test set and a validation set;
- the right algorithms most suitable for the task must be chosen;
- scaliability of the model must be considered.

Currently, companies collect and store vast quantities of data, which are regarded as a true economic asset [64]. With dedicated ML methods intended to extracting knowledge from these datasets, frequently with real-time applications, businesses can conduct their activities and take decisions more quickly and accurately, thereby gaining a variety of benefits. Even high-tech entities, such as smart cities or smart factories, enjoy numerous advantages. As described in the preceding paragraphs, these sophisticated laboratories generate a vast quantity of valuable data. They provide particularly timely information. This vast quantity of data, when combined with the most recent ML techniques, paves the way for new research and development opportunities. There are numerous applications of ML, including energy, manufacturing, transportation, government and public administration, health care, advertising, marketing and financial services [65].

2.3 Neural Networks

As highlighted by [6], Artificial Neural Network (ANN) models are one of the applications of Machine Learning (ML) that have received the most research. An Artificial Neural Network (ANN) is a computing system modeled after the biological minds of animals [58]. Typically, a neural network is composed of connected computational elements representing neurons. Synapses are represented by the connections in a biological brain. Typically, neurons are structured through layers that are interconnected. A feedback link can be present in such an organization, allowing a signal to be propagated back in order to reduce it. ANNs are an interface between disciplines such as neurology, psychology, and artificial intelligence [66]. Through the use of software and hardware, they simulate the synaptic transmission behaviors that occur in the brain when information (signals) is being acquiring and transmitted. ANNs belong to the field of ML. As a result, they employ an approach to ML which, as one can deduce from their name, takes inspiration from the neuron, the fundamental component in our brain with which information is processed. A biological neuron has a structure comprised of three essential components, presented by [6]:

- cellular body - the core in which the nucleus is located, which directs all a neuron's activities;
- axon - a single fibre, very long, through which messages originating in the cellular body and directed to other neurons' dendrites are transmitted;
- dendrites - fibres that obtain the other neurons' incoming messages and send them to the cellular body.

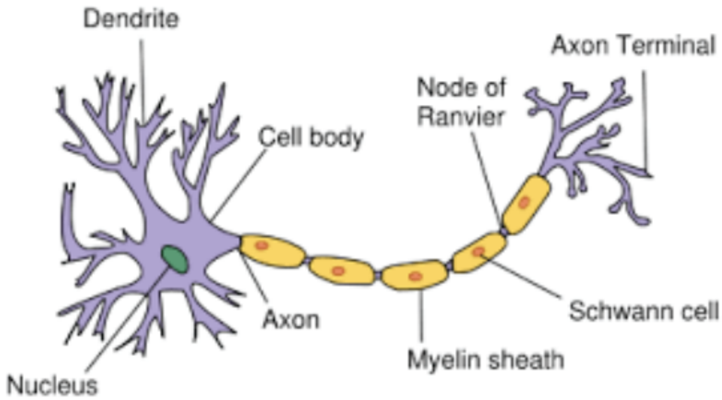


Fig. 2.2 Representation of a biological neuron structure, presented by [6]

A simplified diagram of a neuron structure is shown in Figure 2.2. As explained by [6], neurons transmit information as an electrical impulse through synapses: electrical inputs are received by neurons through their dendrites, then energy is assimilated and released towards other neurons through output channels known as axons. Signals are able to facilitate activating the neuron receiving them. When this happens, the synapse is excitatory, whereas it is inhibitory during contraction. A period of time, known as the refractory period, must then pass before the neuron can generate another impulse, demonstrating how the nature of this simple, but vital, biological element is binary [67]. In a similar way, an ANN is constructed by the interdependence of simple processing units, referred to as artificial neurons, which can extract knowledge from data and store it using synaptic weights. ANNs are typically distinguished by two key characteristics, as presented by [6]:

- ANNs are capable of approximating any function, linear or nonlinear. In fact, linearity or nonlinearity depends solely on the activation functions, which can be linear or nonlinear, and on the learning process;
- Synaptic weights can be updated by ANNs during training. The ANN receives the samples, and calculations and modifications are applied to the weights so that the minimum distance between the desired output and the actual output is minimized. The ANN is thus able to construct an map of the input-output of the system, beginning with training examples.

Multiple varieties of ANNs, ranging from simple feed-forward networks to more complex recurrent structures, have been created by modern researchers. Some of these models, particularly the ones specialized in forecasting time series, will be explained more in depth in the following sections.

2.3.1 History

As presented by [6], the ANNs first appear in 1943, when McCulloch and Pitts first postulated the artificial neuron concept [68], which is the fundamental building block of all ANNs. Hebb developed a theory of learning based on neural plasticity in 1949, which is now known as "Hebbian learning" [69] [70]. Hebbian learning is a type of unsupervised learning. According to Hebb, a synapse between two neurons is strengthened, when the outputs of the neurons on either side of the synapse (input and output) are highly correlated. In other words, when an input neuron fires, if it frequently leads to the firing of the output neuron, the synapse is strengthened. Following the analogy to an artificial system, the tap weight is enhanced when there is a strong correlation between two consecutive neurons. Equation 2.1 describes the mathematical concept in such a way:

$$w_{ij}[n+1] = w_{ij}[n] + \eta x_i[n] x_j[n] \tag{2.1}$$

where the coefficient of the learning rate is η, and the outputs of the ith and jth elements at time step n are, respectively, $x_i[n]$ and $x_j[n]$.

Working on such a construct, Farley and Clark proposed the first ANN in 1954 [71]. Rosenblatt invented the perceptron in 1958 [72]. Minsky and Papert discovered in 1969 that perceptrons are incapable of processing exclusive-or circuits. In addition, they argued that computers lacked the computational capability to simulate a large network [73]. This had a negative impact on ANN research. The development of the backpropagation algorithm in 1974 rekindled an interest in ANNs [74]. This algorithm enabled the efficient training of multi-layer ANNs. In recent decades, the exponential expansion of computer processing capacity has significantly allowed this field of study to progress.

Today, a variety of tasks in numerous disciplines leverage ANNs, including game playing, machine translation, image classification, facial recognition, speech recognition and, of course, time series forecasting.

The Perceptron

In 1958, the concept of the Perceptron was proposed by Frank Rosenblatt. In his research [72], he presents the Perceptron, the precursor to modern ANNs in its ability to recognize and classify structures, to interpret a biological systems' general organization. The Rosenblatt probabilistic model was intended for the analysis through mathematical methods of functions for information storage, and on how pattern recognition is impacted by them. Figure 2.3 illustrates a the composition of a perceptron, as described by [6].

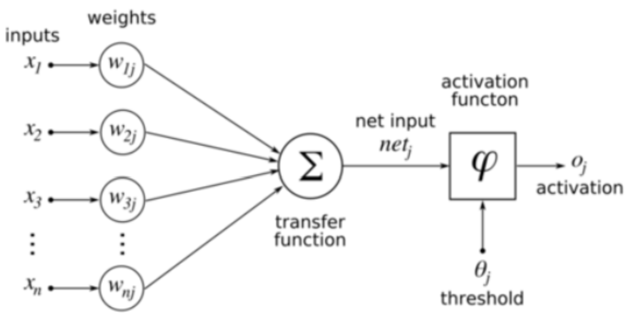


Fig. 2.3 Structure of a perceptron by Rosenblatt, presented by [6]

The perceptron is defined as a structure with a layer of input and a layer of output, and a rule for learning which leverages error minimization (i.e. a function for back-propagation of the error) that modifies the weights of the connections (synapses) as the difference between the actual and desired output, where the actual output is generated by the network after a specific input. As a linear classifier, the simplest Feed-Forward Neural Network (FFNN) is the single-layer perceptron.

Multilayer Perceptron

The basic perceptron can then be expanded to the Multilayer Perceptron (MLP)). Similar to the normal perceptron, the MLP consists of nodes or neurons (units) organized in an input layer, with one or more concealed layers, and an output layer.

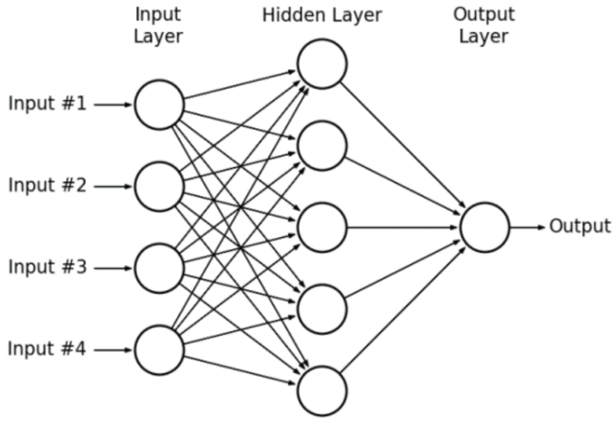


Fig. 2.4 MLP schema, presented by [6]

Figure 2.4 depicts the fundamental structure of an MLP, as described by [6]. The MLP is characterized by a feed-forward architecture, with interconnected layers. Weights, or parameters which can be adjusted, characterize inter-unit connections. This relates to the link intensity between two nodes [58]. Each neuron computes the activation function, which is a function of the sum of the weighted inputs. The functional model is represented by Equation 2.2:

$$\hat{y}_i(w,W) = F_i(\sum_{j=1}^{q} W_{ij}h_j + W_{i0}) = F_i(\sum_{j=0}^{q} W_{ij}f_i(\sum_{l=1}^{m} W_{jl}u_l + w_{j0}) + W_{i0}) \tag{2.2}$$

The matrices $W = [W_{ij}]$ and $w = [w_{jl}]$ specify the weights; where W_{ij} scales the connection between the hidden unit j and the output unit i. The connection between the hidden unit j and the input unit l are instead scaled by w_{jl}. W_{i0} and w_{j0} are the corresponding biases. These weights are vectorized in a vector θ. The vector $u(t)$ represents the input units, while the hidden neuron outputs are represented by the vector h. The training process, during which the parameters are determined, requires a training set Z^n, composed of a set of inputs, $u(t)$, and corresponding desired outputs, $y(t)$, specified by:

$$Z^n = [u(t), y(t)], t = 1, ..., N \tag{2.3}$$

Training permits the determination of a mapping between the set of training data and the set of potential weights:

$$Z^N \to \hat{\theta} \tag{2.4}$$

$\hat{y}(t)$ is predicted by the network, and can then be compared to the true output $y(t)$. The prediction error method, on the other hand, relies on the introduction of a measure of proximity based on the criteria of the mean square error:

$$V_N(\theta, Z^N) = \frac{1}{2N} \sum_{t=1}^{N} [y(t) - \hat{y}(t|\theta)]^T [y(t) - \hat{y}(t|\theta)] \tag{2.5}$$

Then the weights can be found as:

$$\hat{\theta} = arg_{\theta} min V_N(\theta, Z^n) \tag{2.6}$$

with an iterative minimization scheme:

$$\theta^{i+1} = \theta^i + \mu^i + f^i \tag{2.7}$$

where θ^i specifies the current iteration, f^i the search direction and μ^i the step size. In the study of time-series-based systems and ANN families, MLP can be considered as extremely effective and potent [67].

2.3.2 Neural Networks for time-series forecasting

As presented by [6], ANNs are one of the most effective methods for predicting time series [75]. This is a result of their adaptability and ability to model a wide variety of systems, thereby reducing development time and improving performance [51]. As previously stated, there are numerous varieties of ANNs. These vary in architecture and function. There are various varieties of ANNs specific for time-series forecasting. In this section, the most popular ANNs for time-series forecasting will be described.

Recurrent Neural Networks

Recurrent Neural Network (RNN) models, as presented by [6] and as depicted in Figure 2.5, are characterized by multiple layers of recurrent units that share the same parameters and loops that enable information to propagate back to the same computational units. A mechanism is enabled by this type of neural architecture, in which the decision made at time step $t-1$ influences the decision made at time step t [76]. Consequently, two input sources exist for RNNs, the present and the recent past, and they influence how they react to new data. In this manner, each computational phase considers not only the current input at time t, but also what has been learned from previous inputs. Such a process is known as memory [77]. Memory-enabled ANNs extrapolate information from the sequence itself. As a consequence, RNNs are ideally suited for time-series forecasting [78].

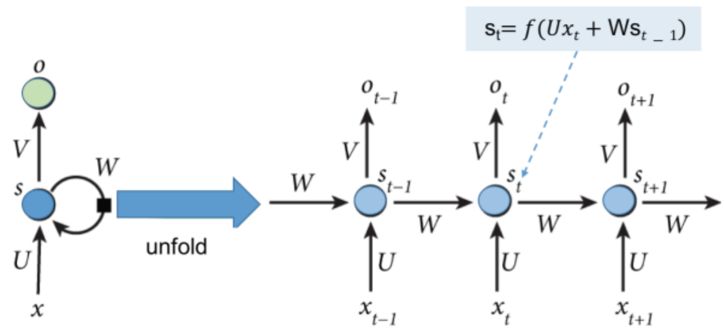


Fig. 2.5 Recurrent Neural Network unit, presented by [6]

An unfolding RNN unit is depicted in detail in Figure 2.5. x_t is presented as the input at time step t, while the hidden state, also called memory of the network (initialised to 0), is s_t. The non-linear activation function is f and, finally, the output is o_t. The following equation describes mathematically the process of carrying memory forward:

$$st = f(Ux_t + Ws_{t-1}) \tag{2.8}$$

where the function of the input at the same time step x_t is known as hidden state, modified by a weight matrix U added to the hidden state of the previous time step h_{t-1} multiplied by its own hidden-state-to-hidden-state matrix W. The weight matrices are filters that determine the relative weight of the current input and the previous hidden state. The error they produce will be returned via backpropagation and used to modify their weights until the error can no longer be reduced.

One of the fundamental characteristics of a time series is the precise dependence between a value at a given time t and the values in an interval preceding t. RNNs are used to address this issue [79]. In RNNs, there is a one-to-one relationship between the specific position in the sequence and the layers (known as the time-step). Nonetheless, RNNs present a significant training challenge due to the vanishing and exploding gradient problems [80], which cause them to have a good short-term memory but a poor long-term one. This is because when the subject of their learning is short sequences, and thus a small number of layers is required, they perform well.

Long-Short Term Memory

The instability of long-term predictions due to vanishing or exploding gradient problems [81] [80], is a well-known limitation of traditional RNN architectures in which the parameters of a large number of layers are learned via backpropagation, as presented by [6]. This causes them to have a good short-term memory and a poor long-term one, since they perform well when learning with short sequences, which require a small number of layers. During the training of a deep ANN, when the error gradients are propagated back in time to the first layer via continuous matrix multiplications, such issues arise. As the initial layers are approached by the gradients, if the value of the gradients is small (much less than 1), they undergo an exponential decrease until they disappear, rendering the model incapable of learning. Similarly, extremely large gradient values (greater than 1) continue to grow and can cause the model to collapse. Long Short-Term Memory (LSTM) ANNs can circumvent such limitations [82], as explained by [6]. In fact, it is an evolution of traditional recurrent models. While the system's structure is substantially similar to that of RNN, a distinct function is used to calculate the concealed state [83]. Repeating modules in an RNN consist of a single layer with a tangential activation function. In LSTM models, memory is implemented as cells (see Figure 2.6), with specific gating functions determining whether information should be retained or erased at each time step. The key to this process is the cell state (green horizontal line in Figure 2.6), which transmits information to the next cell. Input, output, and forget gates (respectively) consist of sigmoidal activation functions coupled with pointwise multipliers to control the power of removing or adding information to the cell state. Each sigmoid output's 0 to 1 values modulate the proportion of the corresponding signal that should pass [83]. Indeed, regardless of the network's depth, the LSTM network's mechanisms allow it to remember values passed through gates in a single state, as explained by [6]. In the LSTM ANN architecture (2.6), the gate mechanism is introduced, thus modifying recurrence conditions on how the hidden states are propagated.

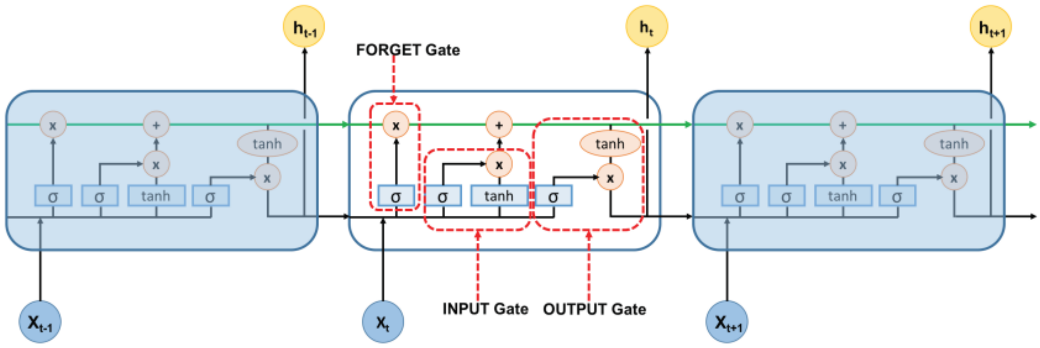


Fig. 2.6 LSTM anatomy, presented by [6], where the input of the cell is x_t and the output is x_t.

As presented by [6], this makes the LSTM model intrinsically resistant to gradient vanishing and explosion. The neglect gate specifies which data should be eliminated from the long-term state. The input gate then determines the data that is added to the long-range state. Lastly, the output gate determines the data to be received and emitted during the present time phase. Figure 2.7 illustrates the internal mechanisms of an LSTM cell.

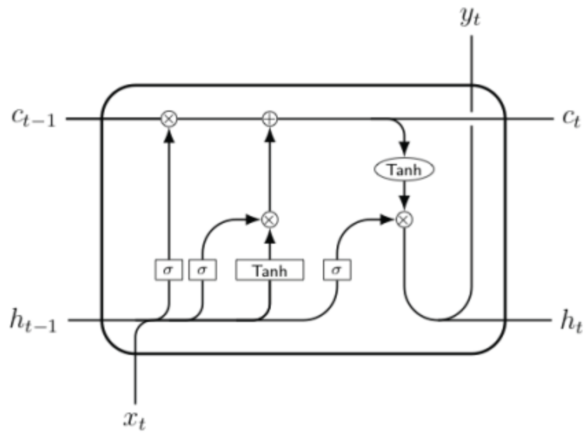


Fig. 2.7 LSTM cell mechanisms in greater depth, presented by [6].

From a mathematical perspective, the short-term state h_t of the LSTM cell is computed, together with its long-term state c_t and its output y_t at each time step t, with the following equations presented in vectorial form:

$$
\begin{aligned}
i_t &= \sigma(W^{xi}x_t + W^{hi}h_{t-1} + b_i)\\
f_t &= \sigma(W^{xf}x_t + W^{hf}h_{t-1} + b_f)\\
o_t &= \sigma(W^{xo}x_t + W^{ho}h_{t-1} + b_o)\\
g_t &= tanh(W^{xg}x_t + W^{hg}h_{t-1} + b_g)\\
c_t &= ft \otimes c_{t-1} + i_t \otimes + g_t\\
y_t &= h_t = ot \otimes tanh(c_t)
\end{aligned}
\tag{2.9}
$$

where the weight matrices of the connections to the input vector are x_t W^{xi}, W^{xf}, W^{xo}, W^{xg} , the weight matrices of the connections to the previous short-term state vector h_{t-1} are W^{hi}, W^{hf}, W^{ho}, W^{hg}, and the bias terms are b_i, b_f, b_o and b_g.

Convolutional Neural Networks

The Convolution Neural Network (CNN), a regularized variant of MLP motivated with biological processes, was initially investigated by Hubel and Wiesel [84]. As presented by [6], this type of ANN is the current standard for image classification and pattern recognition. In fact, it is applicable to image and video recognition, image classification, and natural language processing [85]. As a consequence, CNNs are a type of network widely used in Computer Vision for image classification and object detection [86, 87]. A CNN has an input layer, an output layer, and several concealed intermediate layers. Typically, the concealed layers consist of a series of convolutional layers activated by a Rectified Linear Unit (ReLU) function, accompanied by a flatten layer and dense layers that are fully connected. Figure 2.8 depicts the topology of a standard CNN. The convolutional layer is the fundamental component of a CNN. This layer is composed of a set of filters that, when applied to the entire dataset, create feature maps from a subset of the input data. To obtain a new feature, a learned kernel is used to convolve the resultant feature maps, and the result is fed into a nonlinear activation function (such as ReLU). Between the convolutional and dense layers is the flatten layer, which takes a two-dimensional matrix of features and transforms it into a vector that can typically be passed within an ANN with full connectivity. The dense layer connects every neuron of the preceding layer to each and every neuron of the current layer (usually one or more wholly connected layers). The final dense layer is then followed by an output layer, as explained by [6].

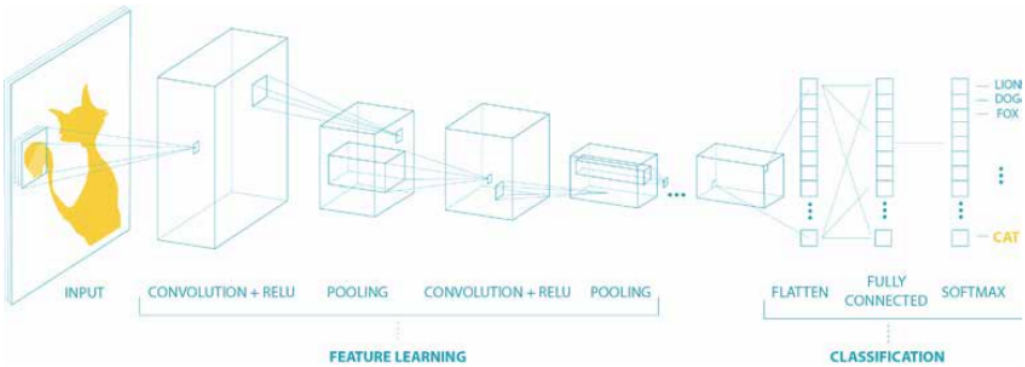


Fig. 2.8 Convolutional Neural Network topology, presented by [6].

The primary benefit of employing a CNN is that it can autonomously detect significant features at computational cost which is acceptable, allowing it to deliver excellent performance in a variety of applications.

1D-Convolutional Neural Network

As previously stated, CNNs are optimal for image processing applications such as image recognition and classification. The 1-Dimensional Convolutional Neural

Network (1D-CNN) is a special form of CNN [88]. The 1D-CNN architecture is ideally adapted for the analysis of timeseries data, such as sensor measurements, signals, or Natural Language Processing (NLP), particularly when the input data contains features that depend on brief consecutive subsequences. As shown in Figure 2.9, the sliding operation is performed from top to bottom and there is no horizontal sliding in 1D-CNNs. Other than that, 1D-CNNs function similarly to conventional CNNs.

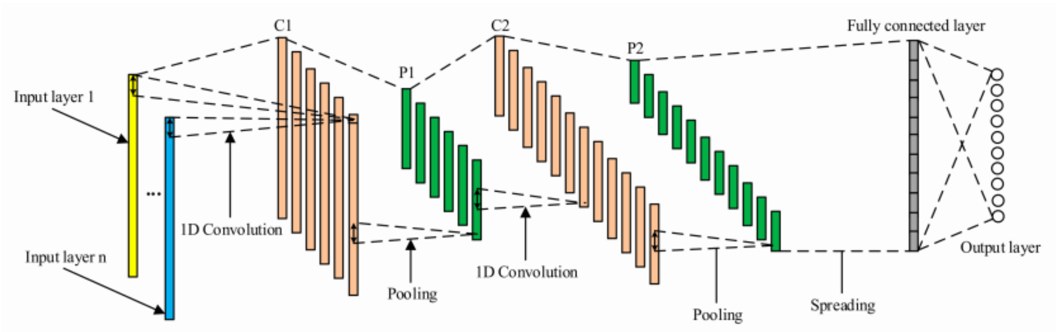


Fig. 2.9 1D-CNN architecture

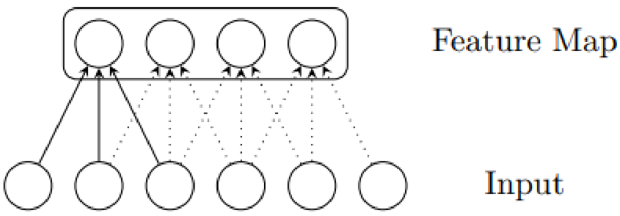


Fig. 2.10 Filter representation, presented by [6].

As presented by [6], for applications based on time series, researchers have recently devised the 1D-CNN, a CNN variant designed specifically for modeling one-dimensional inputs [89]. Currently, 1D-CNN models obtain top performance in a variety of signal processing applications [89] due to their exceptional ability to take sequential data and extract meaningful features. The element responsible for the feature selection procedure is the filter, consisting of a feature detector capable of taking the input data and learning to identify a particular sequence. As depicted in Figure 2.10, after traversing the input sequence the filter activates the neuron which corresponds to the feature map whenever a pattern alignment is discovered. Also, the map of the features denotes where this particular pattern was discovered and provides information regarding the feature's location. Convolution is the action of taking a filter and sliding it across the input data, and it consists of a series of multiplications between the input values and the filter. In general, convolutional layers include multiple filters operating across multiple input channels, allowing the network to recognize multiple patterns in the input data.

In addition, [6] also explains how convolutional layers can be layered to form a deep architecture in which the features detected by the preceding layer are used to build each following layer. The max pooling layer reduces the amount of inputs coming from the preceding convolutional layer. The pooling layer overlays the feature map with a movable window, selecting only the maximum value of each block. At the conclusion of the network, one or more fully connected layers are added to interpret the extracted features. Figure 2.11 depicts a comprehensive example of a generic 1D-CNN architecture.

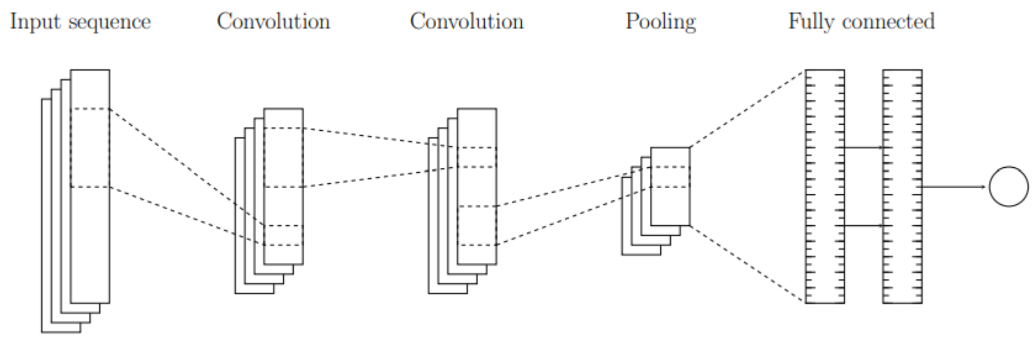


Fig. 2.11 1D-CNN topology, presented by [6].

2.3.3 Model development

As presented by [6], a rigorous and precise method is commonly followed when approaching ANN applications. This is necessary in order to be effective and avoid getting confused in the maze of available tools. Prof. Nrgaard proposes in his book [90] a comprehensive method for developing the ANN models. This procedure is comprised of four stages, as shown in Figure 2.12.

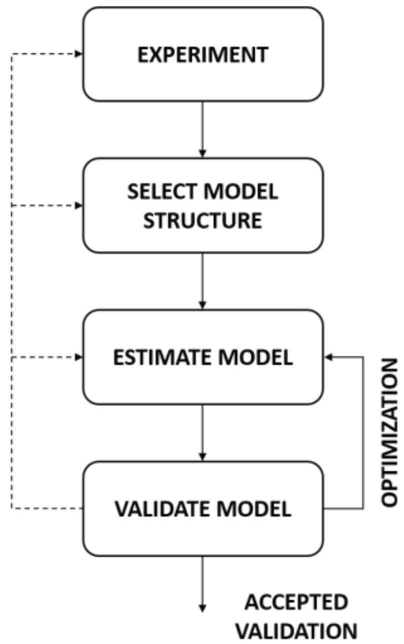


Fig. 2.12 Model development procedure, presented by [6].

The initial stage, the Experiment, focused on problem analysis. [6] explains how, typically, a researcher will approach a problem by identifying its primary characteristics and anticipated outcomes. Simultaneously, sampling and data collection are conducted to capture a reliable data set. Once the scope of ANN applications has been determined, a sufficient quantity of data is required. In general, a greater number of data enables more accurate forecasting [91]. The available data must then be separated into two distinct sets: the training set and the validation set. These data sets are utilized during the training and validation phases of the neural network, which are depicted in Fig. 2.12. as the Estimate Model and Validate Model steps, respectively.

As presented by [6], the Selection of the Model Structure is the second stage. This phase enables the accurate architecture model to be identified [90]. This is a crucial stage because using the incorrect instrument can alter the anticipated results [92]. To accomplish this, the system regressor must be investigated. These regressors identify independent variables able to influence dependent variables in mathematical modeling. In time series, these regressors therefore represent prior samples in relation to the predicted ones [75]. Thus, the optimal neural structure can be selected.

[6] also explains how once one has identified the model of the network and the number of regressors, the implementation and then training of the network is carried out in this step. This step is known as Model Estimation. In time series scenario, training an ANN is needed to provide:

- the vector containing the desired output data;
- the number of regressors to define the prediction;
- the vector containing the weights of both input-to-hidden and hidden-to-output layers;
- the data structure containing the parameters associated with the chosen training algorithm.

The training phase concludes with the production of a training error that represents the network performance index [93].

As explained by [6], Model Validation verifies the trained neural network. Validating a network enables the evaluation of its capabilities [94]. Cross-validation of the test set is the most common method for validating time-series predictions, which involves analyzing the residuals (i.e., prediction errors) in the test set. This method permits the execution of a series of tests, including the autocorrelation function of residuals and the crosscorrelation function of controls and residuals. This analysis provides the test error [93], which is a generalization of the estimation error index. This index should not be excessively high relative to training error. Consequently, the network may overfit the training set.

[6] also explains how, typically, when the training set is overfit by the network, the identified model structure includes an excessive number of weights. The model structure then undergoes final validation and Network optimization. The procedure must revert to the Estimate Model stage in order to modify and redefine a number of structural parameters in order to optimize the entire architecture. To this end, the Optimal Brain Surgeon (OBS) strategy, which is one of the fundamental optimization strategies [95], must be utilized to remove excess mass. Therefore, the network architecture must be revalidated once the new weights have been assigned.

As presented by [6],when moving from the validation block to the previous stage (represented by the dashed line), the paths imply that the execution of the entire procedure is carried out in an iterative fashion. For instance, it may be necessary to retrace steps to determine at least one alternative models, or in the worst case scenario, to repeat the experiment. In general, returning to model estimation signifies that the problem has multiple local minima and it can be difficult to identify the global minimum. The feedback path additionally conceals an enhancement of the criterion known as weight decay or regularization. Instead, returning to model structure selection indicates that the structure of the network is inadequate for its intended purpose. Typically, it is oversized. Consequently, it is typical to employ the prevalent strategy of pruning. Usually, a large enough first structure of the model capable to adequately characterize the system is determined, and a progressive reduction then occurs until the optimal structure is attained. Finally, returning to the experimentation phase entails that the dataset does not reflect certain operating range regimes, necessitating the conduct of additional experiments to acquire more information about the absent regimes.

2.3.4 Forecast horizons configuration

When analyzing and designing a neural network-based problem, it is necessary to precisely configure the output forecast time horizons, as explained by [6]. There are various configurations of neural architectures depending on the desired outcomes and frequently also the available resources. Such configurations can typically be sorted into the following groups:

- iterative prediction;
- multi-output ANNs;
- dedicated networks for each forecast horizon.

Iterative prediction is utilized primarily for short-term (i.e., a few steps ahead) prediction horizons [96]. Such methodologies rely on ANN models that have been trained for single-step predictions (i.e., with a single output). As a consequence, the forecast from the previous phase is used as input for each subsequent process. This strategy generally has drawbacks related to the tendency to accumulate prediction errors.

ANNs with multiple outputs require using a singular network with n outputs, where n represents the number of predicted steps ahead. When implementing the ANN architecture, contrary to the prior method, one must precisely identify the amount of future actions that must be predicted. Nonetheless, particularly for extended forecast horizons, this technique outperforms the iterative method [97].

In conclusion, distinct ANNs can be utilized for each forecast horizon in the analysis. This strategy requires the most resources. Moreover, this method ought to yield superior short-term horizons results (e.g., a few steps ahead), whereas the multi-output method outperforms it for lengthier horizons [96].

2.3.5 Transfer learning

Non Linear Models and its techniques belonging to the fields of artificial intelligence and ML, more specifically to the domain of ANNs, present one of their greatest challenges in that they require substantial amount of data to make sure they are effectively trained and produce acceptable accuracy. In their investigation on the application of CNN, LSTM, and CLSTM, Wang et al. [98] recommend selecting a data length of at least three years.

When approaching the acquisition of new duties, humans utilize their prior knowledge. The greater the similarity between their prior knowledge and the new task, the simpler it is for them to acquire it. Transfer Learning is a branch of ML that consists of utilizing the knowledge gained from one or more source tasks in order to learn a new related target task more accurately and rapidly. In recent years, Transfer Learning has been investigated more and more as a potential remedy to the problem posed by the lack of large and reliable enough data in numerous domains. It has also been investigated for use in PV power forecasting, but, to the best of our knowledge, there is still a dearth of research on this topic.

This is accomplished by continuing to train a model that has already been trained on a specific field using another training set. This has numerous benefits, including:

- **The new task / dataset required less data to be trained**: if a first dataset is used to train a model to predict a certain output from it, and a different dataset with the same type of data is then used as input for the model for further prediction, the original model already has the ability to recognize common patterns in such datasets, and thus requires significantly less data to be tuned and accurately predict the outcome from the second dataset.

- **Training is faster**: since fewer training samples are provided as input, the training duration is drastically reduced.
- **Performance is improved**: compared to a brand-new model, the transferred knowledge model (trained on analogous data from previously) performs better initially on the new objective task. Consequently, after the training of the transferred model on the new dataset is completed, its performance is also superior to that of a model trained from zero.

As demonstrated by [99] and [100], when the data available to train a ANN from scratch is limited, a viable alternative is represented by transfer learning. This is because it enables the creation of a predictive model capable of producing accurate predictions even with a small dataset. Therefore, transfer learning is also utilized in this endeavor, and three techniques are employed:

- **Soft Start**: it involves retraining the model with the target dataset and initializing the weights with the source dataset pre-trained.
- **Fine tune - last layer unfrozen**: leveraging the pretrained model, the weights of all layers except the final one are frozen.
- **Fine tune - first layer unfrozen**: leveraging the pretrained model, the weights of all layers except the first one are frozen.

2.3.6 Models Evaluation

Generally, when working with time-series predictions, it is advisable to analyze the prediction results in order to obtain a comprehensive evaluation of the employed methods, as explained by [6]. This comprehensive analysis includes:

- Analytical assessment - This analytical approach evaluates the prediction performance from a regression analysis perspective, by calculating different metrics commonly used to quantify the similarity between a discrete time-series and a reference ground truth.

- Qualitative evaluation - This analytical approach qualitatively evaluates the prediction performance (e.g., a clinical perspective when predicting blood glucose levels). Utilizing indicators specifically intended to qualitatively evaluate measurements and their outputs, would be advantageous. Obviously, these metrics vary based on the application and closely depend on the type of dataset being analyzed.

Typically, these two evaluations are conducted on a test set which is completely distinct from the one with which the prediction models are developed, trained, and optimized. The following paragraphs present the analytical metrics which are most commonly used when analyzing time series. On the other hand, the qualitative metrics will be appropriately described in the following chapters in relation to the context under consideration.

Analytical assessment

[6] presents how, to quantify the similarity between predicted and observed time-series for the purpose of evaluating the prediction accuracy of the models, it is possible to resort to a number of metrics commonly employed in descriptive statistics and regression analysis. Specifically, the research presents those metrics which are utilized more frequently in time-series research [101]:

<u>Root Mean Square Error</u>

Root Mean Square Error (RMSE, also known as RMSD), is the difference between the predicted and the actual values. It is the most frequently employed prediction error index in scientific literature. Represented as a percentage, its mathematical expression is as follows:

$$RMSD = \frac{100}{\bar{y}_{test}} \sqrt{\frac{\sum_{i=1}^{n} (y_{pred,i} - y_{test,i})^2}{n}} \tag{2.10}$$

<u>Coefficient of Determination</u>

The Coefficient of Determination, R^2, is defined as the square of the correlation between the predicted and actual values (R). Consequently, it ranges from 0 (no correlation) to 1 (perfect correlation). Its mathematical expression is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{test,i} - y_{pred,i})^2}{\sum_{i=1}^{n}(y_{test,i} - \bar{y}_{test})^2} \tag{2.11}$$

Mean Absolute Difference

The Mean Absolute Difference (MAD, also known as MAE), is calculated as the Mean Absolute Difference between the predicted and actual values. Its mathematical expression is as follows:

$$MAD = \frac{100}{\bar{y}_{test}} \frac{\sum_{i=1}^{n}|y_{pred,i} - y_{test,i}|}{n} \tag{2.12}$$

Chapter 3

Summary of works

3.1 Applications

As previously mentioned, the main focus of the research of this Doctoral Program is mainly on applications in the spheres of Industrial Internet of Things (IIoT) and Big Data & Analytics. In order to strengthen the link between academic research and industrial applicability, the research focuses on industrial areas which are considered critical for their impact on manufacturing performance (cost, quality, delivery), and aspects such as readiness, cost, robustness, reliability and flexibility must be taken into consideration during the investigation. In order to understand what areas of manufacturing can most benefit from the application of Industry 4.0 (I4.0) solutions, it is important to proceed with prioritization. If the costs and, especially, losses of the company are stratified, and if the affinity with the cluster “Big Data & Analytics” of the "Piano Nazionale Industria 4.0" [2] is also considered, Energy results being the main priority. According to [7], energy is a major cost issue in Europe (the Commission estimates that wholesale power prices are around 30% higher than in the United States of America (USA), and gas prices are more than 100% higher), and one of the major issues that industrial organizations have been facing is rising energy expenses, particularly those with energy-intensive activities.

To alleviate the impact of these energy difficulties, businesses can engage in sophisticated technology, process optimization, and energy management systems, which can result in significant energy savings and cost savings. Smart energy management, for example, provides enterprises with numerous chances to optimize their energy consumption and cut costs. Energy monitoring and analytics are two examples of smart energy management opportunities for industries. Implementing real-time energy monitoring systems and advanced analytics can help industries identify energy usage patterns, detect inefficiencies, and make informed decisions for energy optimization, as demonstrated by [30]. Better coordination in demand and supply of power within the smart grid framework can be achieved with the implementation of innovative applications [38].

When integrating this need, forecasting of both power demand and supply, with the I4.0 cluster of Big Data & Analytics mentioned in the previous sections, one can clearly understand how technology and innovation can support to overcome these challenges. When looking at innovative solutions for forecasting, with the support of Big Data & Analytics, one enters the Machine Learning (ML) realm. As presented by [39], ML methods are gaining popularity in the forecasting field. Of the different ML methods, [39] highlights how particular attention must be given to those based on neural networks, which presented substantial improvements over benchmarks in the modelling of forecast uncertainty. The research therefore focuses on developing innovative solutions for the effective forecasting of both power demand and supply, leveraging ML methods and particularly those based on neural networks.

An important part of the investigation was to select the best case-studies where to apply the investigation, in order to strengthen the link between academic research and industrial applicability. As previously mentioned, the investigation was originally supposed to be carried out industrial case-studies, however due to the Covid-19 emergency access to such sites was restricted and the research activities were significantly slowed down. An alternative was then found by using other case studies for which some data was already available, and which could easily be replicated in the industrial reality.

Heating, Ventilation and Air Conditioning (HVAC) systems account for nearly half of the total energy consumption of buildings [42]. These systems are primarily controlled by the interior air temperature, with the ventilation system acting in the summer and the heating system acting in the winter. If a valid indoor air-temperature forecasting system in available, and able to effectively consider variables such as building occupancy, external and environmental conditions, seasonal change and other variables, significant contributions can be made to smart energy management applications for building HVAC systems, thereby improving the energy efficiency of the building. Effective indoor air-temperature prediction systems, for instance, can be used to predict energy requirements and costs in advance and as input for load forecasting problems [43]. An interesting opportunity arose to carry out the investigation on a real-world building located in Turin (Italy). As a consequence, the first part of the PhD research focused on developing an innovative solution for the effective forecasting of building in-door air temperature and, as a consequence, of power demand, leveraging ML methods and particularly those based on neural networks.

In terms of power supply forecasting, the challenge was to identify those areas that would provide the greatest benefit to industrial realities. Recent trends show a race to secure alternative energy sources and supplies, an acceleration of new renewables projects, and a rise in clean electricity and electrification investment, particularly related to solar photovoltaic (PV). Within this framework, it is evident that the importance of renewable energy sources such as PV energy is growing. Nevertheless, PV energy can be categorized as Variable Renewable Energy (VRE) source due to its fluctuating power output based on solar energy. PV power's variability poses a challenge to its use in power systems as a stable energy source, since the equilibrium between energy generation and consumption is a strong factor in defining their stability. However, the introduction of smart grids and other technological innovations can overcome these challenges. Within this framework, effective PV power generation prediction can receive special consideration [47], and according to [48], [49] and [50], ML techniques have become an excellent tool for wind generation and PV generation forecasting. An interesting opportunity arose to carry out the investigation on a real-world PV installation located in Turin (Italy). As a consequence, the second part of the Doctoral research focused on developing an innovative solution for the effective forecasting of PV power generation, leveraging ML methods and particularly those based on neural networks.

As previously mentioned, the purpose of the Doctoral research was to develop innovative solutions for the effective forecasting of both power demand and supply, leveraging ML methods and particularly those based on Artificial Neural Network (ANN)s. During the investigations, different ANNs were tested for innovative applications, and their performance evaluated. However, one of the main challenges posed by the application of ML and Artificial Intelligence (AI), particularly those related to ANNs, for forecasting problems, is that large quantities of data are necessary to effectively train them and produce acceptable accuracy. Using different simulators capable of accurately modeling the identified case-studies, and leveraging them to create an artificial, but accurate and realistic dataset large enough to effectively train and test different ANNs, the Doctoral research investigated innovative methods to overcome these obstacles. The resulting ANN models were then applied to the actual, but limited, data from the case studies to make the desired prediction. The application of Transfer Learning (TL) has also been investigated during the Doctoral research, as it has been investigated more and more in recent years as a potential solution to the problem posed by the lack of large and reliable enough data in a variety of disciplines.

3.2 Common methodology

The Doctoral research focuses on developing innovative solutions for the effective forecasting of both power demand and supply, leveraging ML methods and particularly those based on neural networks. These solutions were applied on forecasting of indoor air temperature, and also on forecasting of PV power generation. However, it was important to develop a common methodology which could then be applied successfully to all different applications.

The techniques belonging to the fields of AI and ML, more specifically to the domain of ANNs, present one of their greatest challenges in requiring a substantial amount of data to make sure they are effectively trained and produce acceptable accuracy.

Wang et al. [98] recommend selecting a data length of at least three years. However, availability of large enough datasets of real data, complete, accurate and reliable is still hard to achieve. As a consequence, the research focused on developing innovative solution for effective forecasting even when faced with limited real data availability, by leveraging accurate simulators and Transfer Learning (TL). Also, it focused on investigating the applicability of neural network models which, to the best of our knowledge, had previously seen little application in the chosen forecast domains.

Part of the novelty of these investigations therefore lies in the common methodology used for forecasting through ANNs, presented in Figure 3.1. In all applications a small, real, but limited dataset of real data is available. For each application a simulator, accurately modelled to replicate the real environment, was found and leveraged to generate more data. These simulators, which accurately model the case-study environments, are used to create an artificial, but accurate and realistic, dataset large enough to effectively train and test different ANN models. It is worth noting that, for both case studies, the real and simulated environments are accurately coincident, since the identified simulators have been developed to accurately replicate precisely the two environments which were chosen for the two different applications. The accuracy of such simulators are presented by [102] and [103], and by [104], as described in Sections 4.2.2 and 5.2. Therefore, although using simulated data to train the ANN models clearly introduces some form of bias and limitation in the accuracy of the models' predictions, the use of such accurate simulators, which have been developed to accurately replicate precisely the two environments which were chosen for the two different applications, and whose accuracy is presented by [102] and [103], and by [104], significantly reduces this limitation, allowing to create an artificial, but accurate and realistic, dataset large enough to accurately train the different ANN models.

Furthermore, in order to evaluate the prediction performance of these models with the highest possible level of accuracy, while they are trained and tested on the artificial, but accurate and realistic, dataset, they are then exploited only a portion of the real, but limited, dataset of real data. Similarly, also the different Transfer Learning (TL) techniques which are used to tune the ANN models are applied only on the remaining portion of the real dataset. By using only the real dataset to exploit the ANN models and to apply the TL techniques, any bias and limitations coming from using a simulated, artificial dataset to train the model are intercepted when the model's prediction accuracy is evaluated.

As already mentioned, the whole methodology has been tested and validated on different real-life case studies.

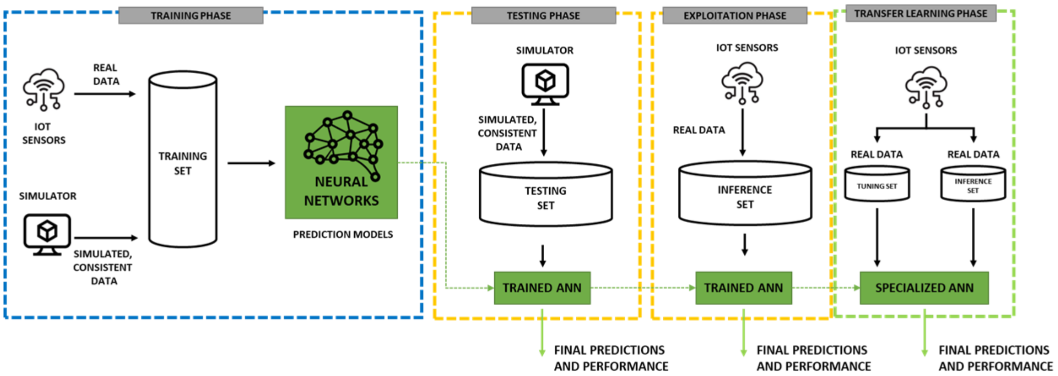


Fig. 3.1 Common methodology

3.3 Forecasting indoor air temperature

As previously mentioned, the first part of the Doctoral research focused on developing an innovative solution for the effective forecasting of building indoor air temperature and, as a consequence, of power demand, leveraging ML methods and particularly those based on neural networks.

During this investigation, an innovative methodology to support the energy management of HVAC systems, through Smart Building indoor air temperature forecast, is proposed. The study is conducted on a public school building in Turin, Italy. The methodology explores the applicability of state-of-the-art ANNs, more specifically 1-Dimensional Convolutional Neural Network (1D-CNN)s and Long Short-Term Memory (LSTM) ANNs, for time-series predictions. These ANNs are first trained on a large, artificial, but realistic dataset based on Building Information Modelling (BIM) simulations with real meteorological data. The inference phase is then carried out on a second dataset collected by Internet of Things (IoT) devices previously installed in the corresponding real-world building, to compensate for the lack of real data. They then undergo an optimization process for the tuning of all their hyperparameters. Finally, Transfer Learning (TL) techniques are exploited to improve the performances of the ANNs' predictions and their ability to generalize. The experimental results are further validated by applying Fanger's model of indoor thermal comfort and show consistent levels of accuracy and comfort even in the face of limited data availability.

3.4 Forecasting photovoltaic power production

The second part of the Doctoral research then focused on developing an innovative solution for the effective forecasting of PV power generation, leveraging ML methods and particularly those based on neural networks.

This study presents an innovative method for forecasting PV power generation with ANNs when only a limited quantity of actual data is available. Initially, feature selection is used to investigate various meteorological features and their potential impact on enhancing the prediction of PV power generation. Then, a simulator that accurately replicates an actual PV installation is used to construct an artificial, but accurate and realistic dataset of PV power generation that is large enough to effectively train and test various ANNs. The ANN models trained and evaluated on the simulated dataset are then applied to a portion of the real, but limited, dataset of the real power generated by the real PV installation upon which the simulator is based in order to evaluate their prediction performance against real data. Different transfer learning techniques are then used to fine-tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, and their ability to improve the prediction performance of PV power generation against the same real data is investigated. The methodology has been tested and validated on a real-life PV installation located on the rooftop of a building of a university campus in Turin, Italy.

Chapter 4

Indoor air temperature forecast

Starting in 2007, the European Union (EU) has set targets to improve its energy efficiency and has focused on sectors with the highest energy-saving potential, such as buildings. 40% of total European energy consumption is attributable to buildings, with heating and hot water accounting for 79% of energy use in EU households. Information and Communication Technology (ICT) play a key role towards identifying innovative opportunities for energy efficiency, especially for forecast of energy consumption, which can be integrated with new control policies in order to reduce energy waste, such as policies for Demand Response (DR) and Demand Side Management (DSM). However, such technologies must also overcome important challenges such as lack of large amounts of accurate historic data on which to base these predictions.

This section of the Doctoral research proposes an innovative methodology to support the energy management of Heating, Ventilation and Air Conditioning (HVAC) systems, through Smart Building indoor air-temperature forecast. The study is conducted on a public school building in Turin, Italy. The methodology explores the applicability of state-of-the-art neural networks, more specifically 1-Dimensional Convolutional Neural Network (1D-CNN) and Long Short-Term Memory (LSTM) Artificial Neural Network (ANN)s, for time-series predictions. These neural networks are first trained on a large, artificial, but realistic dataset based on Building Information Modelling (BIM) simulations with real meteorological data. The inference phase is then carried out on a second dataset collected by Internet of Things (IoT) devices previously installed in the corresponding real-world building, to compensate for the lack of real data. They then undergo an optimization process for the tuning of all their hyperparameters. Finally, Transfer Learning (TL) techniques are exploited to improve the performances of the neural networks' predictions and their ability to generalize. The experimental results are further validated by applying Fanger's model of indoor thermal comfort and show consistent levels of accuracy and comfort even in the face of limited data availability.

4.1 Introduction

In the EU, more than 40% of the overall energy consumption and about 36% of total pollution is attributable to buildings, as reported by the EU Directive on the *Energy Performance of Buildings* [40]. Furthermore, with this sector undergoing a constant expansion, its energy consumption and CO_2 emissions are also consistently rising [40]. Therefore, it is imperative to reduce this sector's energy consumption in order to reduce both the consistent increase in electricity demand and the sector's impact on climate change. As a consequence, research has become increasingly focused on the design and implementation of energy-efficient buildings (whether existing or new), advancing the Smart Building perspective [105].

There is still a lot of room for improvement in energy management in this sector, as many buildings still use control strategies which are out-of-date, for their centralized heating systems [41]. The HVAC systems account for nearly half of a building's total energy consumption [42]. These systems are primarily controlled by the interior air temperature, with the ventilation system acting in the summer and the heating system acting in the winter. The availability of a good indoor air temperature prediction system, able to effectively consider all variables affecting it, such as building occupancy, environmental conditions, and seasonal change, can make a significant contribution to smart energy management applications for building EU systems, thereby improving the energy efficiency of the building. Effective interior air temperature prediction systems, for instance, can predict energy requirements and costs in advance and serve as input for load forecasting problems [43].

When integrated in buildings connected to the district heating system, the benefits of such applications can be enhanced by harmonizing the energy demand and optimizing the planning of energy supply. Within this framework, innovations such as DR [106] and DSM [107], which also consider ambient comfort [108], can thus be introduced in building heating systems. Moreover, by predicting energy profiles, the peaks in thermal energy requirements can be minimized and the consumption of thermal energy can be transformed, at least for district heating applications [109]. As analyzed by Bianchini et al. [110], DR is a program consisting of mechanisms that endeavors to modify the demand energy profile of the consumer over time so as to reduce demand during peak periods and align overall capacity of energy supply with the energy demand. EU is a particularly essential participant in this program. As a result, DR also allows for the reduction of operating costs associated with costly generators, and promotes the introduction of renewable energy sources and of a Distributed Energy Resources (DER), reducing the emissions of CO_2 [111].

4.1.1 Building and thermal modelling

To control an EU system, a model must be developed that precisely predicts the building's response to operational and environmental changes. Recent literature provides numerous methodologies for constructing modelling [112–114]. White-box building models, for instance, can be utilized in the construction industry. In fact, it is possible to construct three-dimensional representations of buildings that include all variables affecting thermal response, such as materials and insulation of walls, roof, and windows, window size, ventilation scheduling, solar exposure, etc.

After the development of the model, the building's thermal behavior can be forecasted through a simulation software. In the realm of building thermal energy performance simulation, both during design or refurbishment phases, EnergyPlus [115] and TRNSYS [116] are considered reference simulation tools. EnergyPlus, whose development is funded by the Department of Energy (DOE) of the United States of America (USA) [117], is a building energy simulation program for modeling building heating, cooling, lighting, ventilating, and other energy flows [118], and is commonly used by designers, architects and researchers [119].

However, despite the fact that these software produce extremely robust and accurate results, they are extremely resource-intensive in terms of computational power [102]. New methodologies capable of balancing computational costs and the accuracy of thermal estimations are described in academic literature as means of overcoming these limitations. Such methodologies begin with an accurate building model, and then develop a more compact approximation through i) model order reduction, ii) model aggregation, or iii) ad-hoc dynamics extraction [120, 121].

Alternately, an alternative modeling strategy in which the structure is represented as a grey-box model is proposed by Bacher et al. [122]. The model is constructed using the resistor–capacitor representation of the building's thermal dynamic, meaning resistances and capacitors consistitute its main components. However, these simulations require numerous assumptions regarding their variables, such as ventilation scheduling, and the calibration of these variables; consequently, they are rarely representative of the actual environment. Additional challenges include correctly accounting for surrounding buildings, heat sources, and airflow obstructions [100], in addition to simulations of Computational Fluid Dynamics, and their the high computational expense. To overcome these limitations, ANNs are currently extensively used to model the thermal response of buildings, including the prediction of interior air temperature, with promising results.

Other approaches to developing a compact thermal model emphasize Very Large Scale Integration (VLSI) techniques. This method's solutions frequently employ matrix pencil [123] and subspace identification [124], which disregard physical constraints and render the compact model extremely flexible. Solutions which are based on VLSI-based carry out a comprehensive analysis of numerical simulations or sampled real-world data, resulting in a highly accurate training phase for the entire model. However, they are typically inadequate for dealing with the nonlinearity of a building's entire thermal system. Again, Machine Learning (ML) techniques such as ANNs can be utilized to effectively address this issue.

4.1.2 Black-box modelling with Artificial Neural Networks

Artificial Neural Network (ANN) are data-driven methods that can model the fundamental structure of a system by utilizing only particular inputs and outputs. In this context, they can be used to model and predict the indoor air temperature using only previous measurements of the indoor air temperature as inputs. As a result, ANNs require significantly less data than traditional modelling and simulation methods, which require a wide range of parameters, such as building and environmental parameters, to accurately model and simulate indoor air temperature behavior. Furthermore, the transferability of ANNs is a significant advantage of employing them in this context. As a result, the use of ANNs in various building energy management applications has increased dramatically over the past years [125] [126].

Mateo et al. [127], Comparing traditional ML models for time-series forecasting, such as Autoregressive, Multiple Linear Regression (MLR), and Robust MLR, to ANN models, such as Multilayer Perceptron with Non-linear Autoregressive Exogenous (MLP-NARX) and Extreme Learning Machine, in their investigation of indoor air temperature prediction problems. All models are created using one year of TeKton three dimensional (3D) software-generated simulated data. The results demonstrate that neural models outperform other ML models, with the MLP-NARX ANN demonstrating notably strong performance for predicting interior air temperature.

This ANN is then further investigated in a research conducted by Aliberti et al. [128], introducing the Non-linear AutoRegressive (NAR) ANN architecture, which is capable of basing its prediction on a large number of regressors, and emphasizing an overall good performance in predicting interior air temperature values up to two hours in advance. To increase the applicability of the methodology for Smart Energy Management in buildings, the research leaves open the opportunity to investigate the possibility of extending the forecasting horizon. Xu et al. [129] initially investigate LSTM ANNs for predicting interior air temperature in a public structure using sensor data sampled every 5 minutes. The performance of the LSTM ANN is evaluated for two time horizons, five minutes (one step ahead prediction) and thirty minutes (six steps ahead prediction), and then compared to the Back Propagation ANN, Support Vector Machine, and Decision Tree models. The results indicate that the LSTM ANN performs marginally better than all other networks in both instances, with the authors speculating that the positive difference in performance could be even greater if not for the limited quantity of data.

An LSTM model for indoor air temperature prediction was also developed by Kamel et al. [130], with the aim to detect the most impacting inputs, demonstrating that a small number of inputs are the most effective at producing an acceptable level of accuracy.

Another interesting input variable was also added by Deihimi et al. [43] to their networks: they added a variable highlighting if the data is related to a working day or a holiday, with their results showing greater prediction accuracy when only working days are considered.

Previous studies in this field, such as [131] and [128], using ANNs, were able to devise accurate models for predicting interior air temperature, but training these networks required a vast quantity of data. The accumulation of such a large dataset with acceptable quality levels (clean data, absence of time leaps, absence of bogus measurements, etc.) remains challenging, as it requires the implementation of multiple technologies and the ability to orchestrate them together effectively. Consequently, the application of ANNs for indoor air temperature prediction and energy management in buildings is still hindered by a deficiency of large and reliable data, despite the fact that research in the various sectors is continually improving and the availability of data is expanding.

Most recently, Cifuentes et al. [132] reviewed the literature on ML strategies for indoor air temperature forecasting, highlighting their pros and cons and identifying their research gaps. The research notes that Deep Learning-based approaches, for instance, necessitate the acquisition of extended time series or complementary simulation systems in order to generate sufficient samples for the training-validation process. To accurately predict interior air temperature, the training-set and test-set must contain at least three and one years of sampling data, respectively. This is also confirmed in [100, 79, 131, 128]. In the work of Alawadi et al. [133], the algorithms of 36 ML models, capable of being used for indoor air temperature prediction in a smart building, are compared, highlighting how data noise strongly affects the majority of them, including ANN, and how the accuracy of the best model remains valid even after the forecasting time is increased.

4.1.3 Black-box model transferability through leverage of Transfer Learning

In recent years, Transfer Learning (TL) has been investigated more and more as a potential remedy to the problem posed by a lack of large and reliable enough data in numerous disciplines, including Smart Energy Management for buildings. In [134], Jiang and Lee present the use of Transfer Learning (TL) on an LSTM network for Thermal Dynamic Modeling of interior temperature evolution and energy consumption in buildings. Their research demonstrates that the TL method is effective for adapting the pretrained LSTM model from one building to another, demonstrates superior prediction performance compared to the traditional model applied to a limited amount of data, and demonstrates that LSTM ANNs are a suitable model on which to base such an application. However, their investigation pre-trains the initial model on a large quantity of data collected from the initial building and does not evaluate the model's efficacy with end-user comfort in mind. In [135], Gao et al. instead investigate a TL based multilayer perceptron model to address the common data-shortage issue and improve the thermal comfort prediction performance. Their research provides a valid application of TL to the prediction of thermal comfort, but includes the latter as one of the features of their ANN model and, therefore, rather than focusing on prediction indoor air temperature it directly forecasts thermal comfort. Xu et al. [136], instead, present a novel TL-based approach to overcome the challenge of long training time for data-driven learning methods, and to effectively transfer a controller trained for a source building to a controller for a target building with minimal effort and enhanced performance. Their method presents an innovative TL approach by decomposing the ANN that is the basis for the controller into two sub-networks (a front-end network that can be directly transferred because it is building-independent, and a building-specific back-end network that can be efficiently trained with offline supervise learning), however the replicability of this method on indoor air temperature prediction is not demonstrated. Deng and Chen [137] use TL to transfer the logic and a portion of the occupant behavior model structure between buildings with different control systems, as well as to transfer a model from office buildings to residential buildings by modifying how occupant behavior is affected by air temperature. Again, despite the fact that their research demonstrates that in the context of Smart Energy Management for buildings, TL can have positive results, their method is not explicitly applied to the prediction of indoor air temperature.

When looking at indoor air temperature, Chen et al. in [100] investigate indoor air temperature forecasting by employing different TL techniques. Using EnergyPlus, they construct two datasets: a multi-year dataset for a room in Beijing, which is used as the source dataset, and a 15-days-only dataset for a room in Shanghai, which is used as the target dataset. The two rooms have distinct attributes. The source dataset is then utilized to train a Multilayer Perceptron (MLP) ANN and then, using the resulting weights, TL is applied on the target dataset with three different techniques: i) re-train the entire ANN, ii) re-train only the final layer, and iii) re-train only the initial layer. All three TL techniques achieve a good prediction on the target dataset, but the third technique (re-train only the first layer) consistently outperforms the other two for all 10-minute, 30-minute, 60-minute, and 90-minute prediction horizons. By comparing the research carried out in [99] and [100], both investigations use the technique named "soft-start" in [99], where all the ANN is re-trained, while [100] examines the two other methods, which only retrain the first or last layer of the ANN. Again, this study shows interesting results, but it also leaves open the option of looking into whether or not the prediction window could be increased, to make the method more suitable for Smart Energy Management in buildings.

4.1.4 Contribution

The research aims to investigate and evaluate the predictive accuracy of ANNs for indoor air temperature. LSTM and LSTM are the most advanced ANNs for analyzing time series data, given that the investigation is conducted on such data. However, while LSTM has been effectively utilized in the literature to predict indoor air temperature, LSTM has, to the best of our knowledge, not yet been utilized for this purpose, but is well-suited for the analysis of time series data. Furthermore, the application of TL on both LSTM and LSTM is novel because, to our knowledge, this method has not previously been used to predict interior air temperature. In addition, the research examines the possibility of extending the forecasting horizon for interior air temperature prediction in order to enhance the applicability of the methodology for Smart Energy Management in buildings.

The ANNs are first pre-trained on a simulated dataset. A BIM of a real-world case-study building is developed, and subsequently used as an input to EnergyPlus, together with real meteorological data instead of Typical Meteorological Year (TMY) data, following the research carried out in [102]. This generates the simulated dataset, composed of 6 years of data, large and reliable enough to train the ANNs. The ANNs are then fine-tuned, through the application of TL, on a real dataset, generated by IoT devices installed in the real-world building and used to collect one-month of indoor air temperature measurements. All three TL methods applied in [100] are used, in order to further evaluate how to improve the prediction accuracy.

Following the work carried out by [43], the ANNs are differentiated according to whether they regard the day of the week as an additional variable. All of these distinct ANN architectures are employed and their results are contrasted in order to determine which architecture is most capable of delivering a model capable of accurately predicting the interior air temperature. The hyperparameters of each ANN are calibrated using the same methodology; however, each ANN is optimised separately in order to identify the unique set of parameters that allows each ANN to determine the interior air temperature most accurately. This method strengthens the comparison because it seeks to maximize the accuracy of each ANN and makes the accuracy of the results dependent on the architecture of the ANN, rather than on the influence of inadequately tuned hyperparameters.

To test our methodology, the proposed solutions were applied to a real-world school located in Turin (Italy). The experimental results are further validated by applying Fanger's model of thermal comfort [138], a method for evaluating thermal comfort in a building based on environmental and occupants, which is part of both ASHREE [139] and ISO 7730 [140] standards.

The rest of the paper is organized as follows. Section 4.2 presents the proposed methodology and its application in a real-world case-study building, which is a secondary school in Turin, Italy. Section 4.3 discusses the experimental results. Finally, Section 4.4 provides concluding remarks.

4.2 Material and methods

4.2.1 Enabling Technologies

This section will now provide a concise overview of the various ANN architectures investigated in this study, as well as the various methodologies used to train and then apply Transfer Learning (TL) to these architectures.

Artificial Neural Networks and Transfer Learning

A detailed description of the different types of ANNs, and of the Transfer Learning (TL) methodology, can be found in the previous section and will therefore not be repeated. Of the different available ANNs previously presented, this investigation will be leveraging:

- 1-Dimensional Convolutional Neural Network (1D-CNN) Artificial Neural Network (ANN)

- Long Short-Term Memory (LSTM) Artificial Neural Network (ANN)

Both ANNs can be considered novel for this application since, to the best of our knowledge, they have not yet been applied to predicting indoor air temperature.

TL is also applied also in this work, and the three techniques previously described are used:

- **Soft Start**: it consists of re-train the model with the target dataset, and having the pretraining on the source dataset initialize the weights.
- **Fine tune - last layer unfrozen**: using the pretrained model, the weights of all the layers are frozen except the ones of the last layer.
- **Fine tune - first layer unfrozen**: using the pretrained model, the weights of all the layers are frozen except the ones of the first layer.

4.2.2 Case study

This investigation is carried out on a primary school building of about 14,500 m^2, distributed on two floors, located in Turin, north-western Italy. The building is connected to the district heating system, which regulates the on/off status of the heating system, and lacks an air conditioning system. The building's windows are double-glazed and installed on masonry wall facades. Solar radiation contributes significantly to the thermal energy content of both the east- and west-facing facades. The selection of this particular building type enables a large-scale evaluation of potential retrofit actions on public assets. The energy management and simulation infrastructure designed to predict interior air temperature in extant buildings seeks to incorporate heterogeneous data sources in accordance with a challenging methodology that integrates Building Information Modelling (BIM), meteorological data, and IoT devices. In recent years, a growing number of public procurement activities have been managed with BIM tools. In order to simulate the energy behavior of buildings in the Smart City scenario, we therefore strategically consider the utility of employing digital models that accurately represent the actual characteristics of buildings. Regarding the architectural and energy modeling of the building, we have adopted a BIM-based procedural workflow established through prior research [103, 102] to optimize the activities and make the method replicable. Using Autodesk Revit model authoring software and Design Builder, the virtualization of the building was accomplished. In contrast, the EnergyPlus engine was used to simulate energy in the dynamic regime. The BIM model is created using archive documentation validated by an on-site survey to determine the current condition of the asset. Although the graphical representation must be simplified for energy purposes, a BIM model has the potential to reduce the number of misinterpretations and inaccurate approximations of the building geometry that occur in conventional practice. To make the model a repository of useful, mutually consistent information for the analysis, it must include: i) accurate characterizations of building envelope in terms of correct stratigraphy, thermal and physical properties; ii) materials nomenclature standards, iii) thermal zones using rooms entity; iv) facility management information (such as room type and occupants). The energy analysis model can be generated directly from the modeling software and exported in green building eXtensible Markup Language (also known as gbXML) format. With the current state of software development, a complete exchange is not possible; therefore, a variety of modeling solutions have been developed to maximize interoperability and accelerate the subsequent re-entry of data. The following are some detailed instructions on how to standardize the BIM: i) removal of all elements not required for simulation, including decorations, staircases, and furniture; ii) walls must be correctly oriented and wall joints must be configured; iii) selection of a representative stratigraphy for the external envelope and the internal partitions; iv) replace structural elements with architectural ones; v) the room bounding of architectural pillars must be disabled; vi) floor objects must be modelled following the centreline of Regarding the systems, only the dimensions of the water radiators were utilized to determine the nominal capacity and flow rate for each radiator via schedules. The energy analysis model was validated by comparing the BIM component physical surfaces (i.e. walls, floors, roof, doors, and windows) to the analytical surfaces using an iterative optimization strategy to achieve deviations of less than 10%. Design Builder is therefore utilized as the most user-friendly program interface to complete the required simulation input data populating. Specifically, the most important energy-related inputs to manage involve materials that must be reassigned based on the software library, transparent surfaces that may not be correctly recognized, and systems. Data regarding the heating system and energy consumption must be manually entered. All of this data generates the building energy model via the IDF file, which is the Design Builder output. It is used as input for the EnergyPlus engine with actual meteorological data rather than the default TMY to generate a large 6-year synthetic data set. The third-party weather data source from the closest weather station (i.e. solar radiation, outdoor air temperature, and humidity) has been utilized to enable more precise forecasts, as demonstrated by the research carried out in [102].

The internal rooms of the building are evaluated, to identify those which best suit the study. The following three rooms are chosen, based on factors such as room shape, symmetry, and regular internal distribution:

- East Room - classroom facing east;
- West Room - a classroom facing west;
- Corridor - the corridor in the main entrance.

As far as these spaces are concerned, it is essential to highlight both their architectural and occupant characteristics. During business days, the corridor's population fluctuates considerably. The chambers to the west and east, on the other hand, are enormous and feature numerous apertures and glazed windows. The model of the building is shown in Figure 4.1.

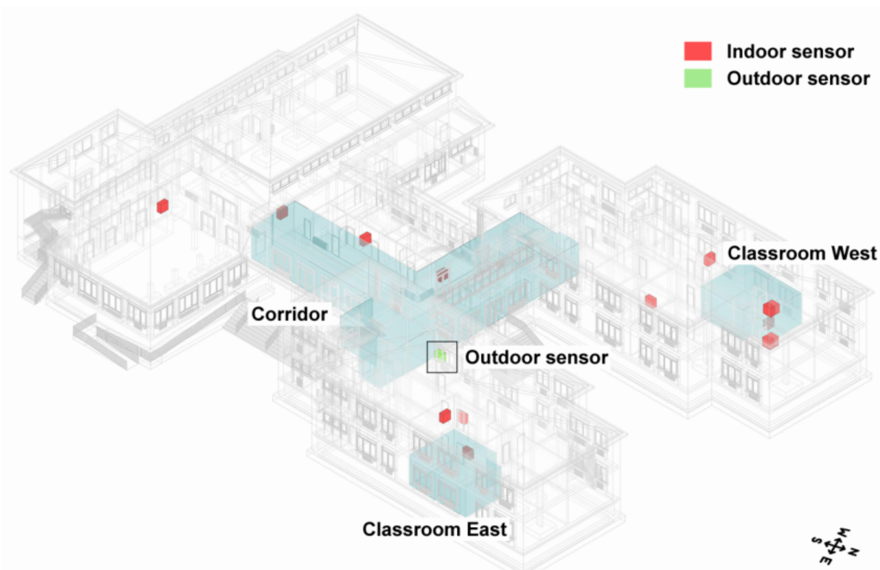


Fig. 4.1 BIM model of the case-study building (reference rooms are highlighted in blue)

Each of the exhibited chambers generates its own dataset, which is then applied with the methodology. In addition, in order to conduct a more exhaustive study, a specific dataset titled "Whole Building" is created to account for the building as a whole. The methodology used to generate these datasets will be described below. The research is therefore conducted in the following four environments:

- Whole Building;
- East Room;
- West Room;
- Corridor.

Utilizing the BIM model, a unique dataset is created for each room: West Room, East Room, and Corridor. The BIM model of the entire structure is then utilized to generate the fourth dataset, titled "Whole Building". First, simulated temperature datasets are created for every chamber in the structure. Then, inconsequential areas where students and staff are not expected to frequent on a daily basis, such as the location under the roof and subterranean archives and maintenance rooms, are omitted. The "Whole Building" temperature dataset is then produced by aggregating all other simulated temperature datasets.

In order to capture one month of actual data during the winter season, for the real dataset, a mesh wireless sensor network with 12 prototype nodes was deployed in the building. The prototype nodes, described in [141], was created using off-the-shelf devices from the STM32 NUCLEO ecosystem by STMicroelectronics:

(i) a NUCLEO-L152RE board equipped with a 32-bit microcontroller running at up to 32MHz;

(ii) an X-NUCLEO-IKS01A1 motion MEMS and an HTS221 temperature and humidity sensor;

(iii) an X-NUCLEO-IDS01A4 radio transceiver board equipped with the transceiver SPIRIT1 (868MHz); and

(iv) a STEV AL-ISV021V1 energy harvester board, equipped with converter and battery charger SPV1050 MPPT, a 26.5 cm^2 solar panel and a 120mAh Lithium coin cell battery.

To collect the information, we used the distributed software infrastructure presented in [102]. It is a cloud-based software architecture that incorporates heterogeneous hardware and software components for monitoring and simulating the energy behavior of buildings. The infrastructure utilizes ubiquitous IoT devices with geo-referenced data to capture environmental and energy-related data. Utilizing middleware technologies, IoT devices are associated with BIM, Geographic Information System (GIS), and weather data. It enables energy simulations to evaluate both the energy efficacy of buildings and the effects of potential renovations. In addition, the data provided by IoT devices can be used to (near)-real-time monitor and control environmental and energy parameters of buildings.

Figure 4.2 represents the "simulated" and "real" datasets. The structure of both datasets is identical: data, data type, and sampling time are identical. Both the simulated and actual datasets contain interior temperature measurements collected at 15-minute intervals. The sensor datasets from the selected rooms are then used to execute the TL application. The actual "Whole Building" dataset is generated by averaging the temperature readings from all sensors installed in the school, not just those in the selected rooms.

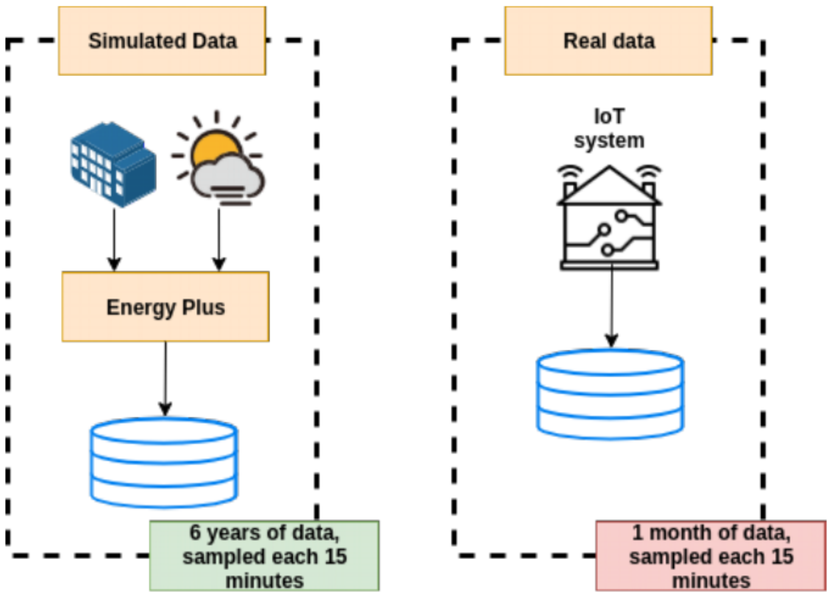


Fig. 4.2 Datasets

After the behaviour of the indoor air temperature has been observed (Figure 4.3), one can appreciate a definite distinction between weekdays and weekends. This is due to the fact that the HVAC system is turned down on weekends, resulting in reduced absolute values and variations of interior air temperature. In addition, Monday behaviour is distinct from that of the other weekdays: it takes longer for the domestic air temperature to reach the required level on Monday, because the HVAC system is turned off over the weekend. As indicated, this conduct is taken into account during the investigation.

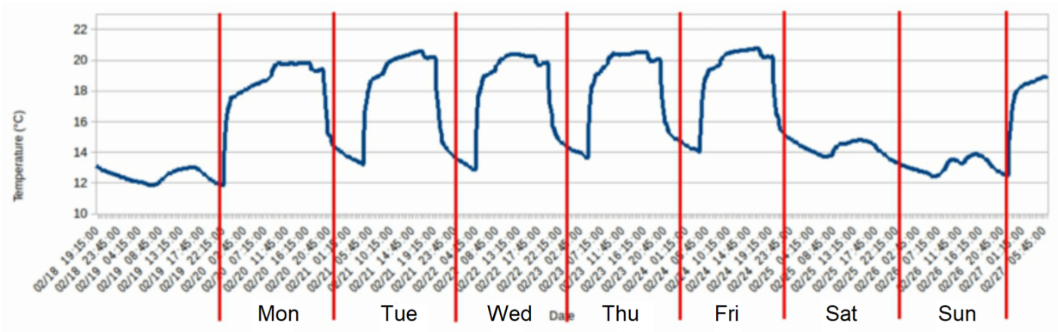


Fig. 4.3 Day of the week behaviour

As previously described, each of the aforementioned environments generates its own dataset. Therefore, the various ANNs, or more particularly their hyperparameters, are optimized for each environment, and there is a version of each ANN variety for each environment. There will be additional ANN variants based on whether the weekday and weekend differences in behavior are taken into account.

Lastly, it is essential to comprehend why the investigation generates unique ANNs for each chamber and the building as a whole. On a room-by-room basis, the efficacy of ANNs can be evaluated in terms of their potential applications in terms of capillary control applied to individual structures, and is thus related to Smart Building applications [105]. At building level, on the other hand, the results of the investigation can be evaluated for applications of DR [106] or DSM [107] at city level, such as District Heating [109].

4.2.3 Evaluation metrics

When evaluating the performance of ANNs, it is essential to comprehend the application's objective and to establish evaluation metrics that can effectively evaluate the performance in comparison to the application's objective. The purpose of this study is to determine whether ANNs are capable of forecasting future internal air temperature within acceptable thermal comfort ranges (comfort analysis is explained in greater detail below). The objective of the research is not the accomplishment of exceedingly accurate predictions, but rather the prediction's ability to remain within a certain threshold. Following these considerations, the Mean Average Error (MAE) and the Predicted Mean Vote (PMV) are used to evaluate the ANNs' performance, with $\hat{y}_t$ as the predicted values and y_t the observed ones. In addition, as a result of the disparity between the size of the data set and the number of steps preceding the prediction, the MAEs and PMVs are analyzed by calculating their mean, median, maximum, one standard deviation, and two standard deviations, respectively.

In order to evaluate if the indoor air temperature forecasts remain within ranges which can be considered acceptable for thermal comfort, the parameters studies by Fanger [138], the PMV, which estimates the state of well-being of a group of individuals, and the Percentage of Percentage of Person Dissatisfied (PPD), which expresses the percentage of people that are uncomfortable with thermal condition of the room, are used, following the logic used by current thermal comfort evaluation standards, both based on Fanger's studies, ASHREE [139] and ISO 7730 [140].

Fanger's studies on thermal comfort determine that in order to maintain the perceived thermal comfort of an occupant, the operative temperature should not vary more than 1.1 °C over 15 minutes and no more than 2.2 °C over 1 hour [138, 142]. As a consequence, the acceptable MAE threshold for acceptable prediction accuracy is set at 2 °C. Following the standards identified in literature, the PMV is therefore analyzed as an absolute value and its acceptable threshold for acceptable comfort levels is set at 0.5.

4.2.4 Methodology

The purpose of this section is to discuss the methodology adopted during the implementation of the different ANN models, from the data collection phase to the final testing phase, together with explaining the motivations that led to certain choices during the implementation.

As represented in Figure 4.4, this paper investigates the application on two different ANNs: 1D-CNN and LSTM ANN. After the initial training, the ANNs undergo TL. Three different TL techniques, presented in Section 4.2.1, are applied in order to investigate the best performance in terms of prediction.

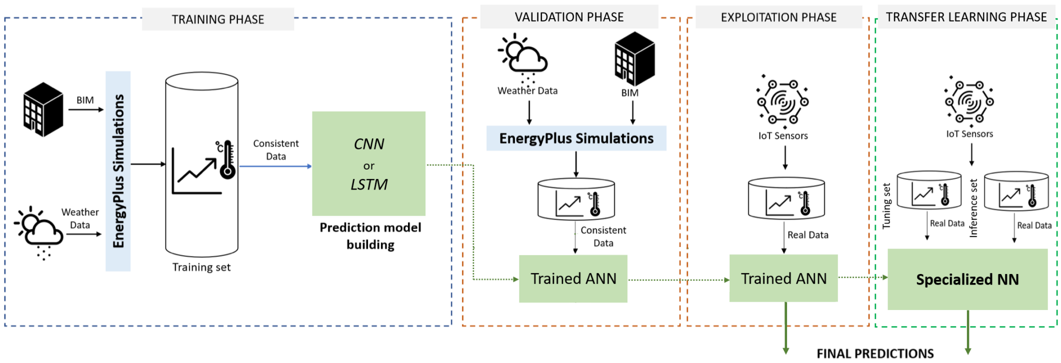


Fig. 4.4 Schema of the proposed methodology

As introduced in section 4.1.4, the ANNs are first trained on a synthetic dataset, composed of 6 years of data, large and reliable enough to train the ANNs. To generate this synthetic dataset, which is referred to as "simulated", a BIM of a real-world case-study building is developed, and subsequently used as an input to EnergyPlus, together with real meteorological data instead of TMY data, following the research carried out in [102]. The testing and TL phases are then carried out using the real dataset, which is referred to as “real”. This real dataset is generated by IoT devices installed in the real-world building and used to collect one-month of indoor air temperature measurements.

The investigation’s objective is to predict the indoor air temperature when the HVAC system is operating. The case study building is supplied with a heating system but not with an air conditioning system. For this reason, only datapoints from days when the heating system is operational are regarded within the datasets. The building’s heating system is directly controlled by the district heating system, which is connected to the building’s heating system. Between the 15th of October and the 15th of April, the district heating system is activated annually. Table 4.1 provides a comprehensive summary of the datasets, their extent, the interval considered for each year, and the size of these intervals. The simulated dataset is then divided into a training set and a validation set, whereas the actual dataset is divided into a tuning set and an inference set, so that the traditional ANN training can be performed first, followed by the fine tuning through TL. Section 4.2.4 provides additional information on the partitioning of the datasets.

Table 4.1 Size of datasets

Year	Type of data	Number of datapoints	Data interval considered	Number of datapoints in the interval
2010	Simulated	35.040	01 Jan - 15 Apr; 15 Oct - 31 Dec	17.472
2011	Simulated	35.040	01 Jan - 15 Apr; 15 Oct - 31 Dec	17.472
2012 *	Simulated	35.136	01 Jan - 15 Apr; 15 Oct - 31 Dec	17.568
2013	Simulated	35.040	01 Jan - 15 Apr; 15 Oct - 31 Dec	17.472
2014	Simulated	35.040	01 Jan - 15 Apr; 15 Oct - 31 Dec	17.472
2015	Simulated	35.040	01 Jan - 15 Apr; 15 Oct - 31 Dec	17.472
2016 *	Real	35.136	01 Jan - 15 Apr; 15 Oct - 31 Dec	17.472
2017	Real	35.040	01 Jan - 15 Apr; 15 Oct - 31 Dec	17.472

* Leap year

Following the previously described data collection, the four main phases required to develop any new predictive model are: i) data pre-processing, ii) training, iii) validation and iv) testing phase. Each phase is explained in more detail in the rest of this section.

Data pre-processing

Any dataset must be preprocessed before it can be used to train an ANN or model. During preprocessing, the data are cleansed, outliers are removed, the data are scaled, and then they are divided into training, validation, and testing datasets. Errors in the data are primarily the result of sensor malfunction, which generates singular out-of-scale readings (also known as outliers) or no readings at all. The afflicted samples are replaced using linear interpolation between their prior and subsequent values. Data pre-processing has to be applied only to the real dataset, since the simulated dataset is complete and consistent.

The data is then subsequently scaled, as suggested by Aggrawal in [143], and brought within a range of -1 and 1. Finally, in order to allow first the traditional ANN training and then the fine tuning through TL, the datasets are split in the following way:

- **Training set**: simulated dataset, years 2010 to 2013;
- **Validation set**: simulated dataset, years 2014 to 2015;
- **Tuning set**: 80% real data;
- **Inference set**: 20% real data.

Training the neural network

The ANNs are then trained to make indoor air temperature predictions based on an input window of indoor air temperature measurements. These predictions can be made one step ahead, predicting only the next temperature value after the last temperature input, or multiple steps ahead, predicting multiple time steps after the last temperature input. Forecast horizon refers to the number of stages or period of time in the future of the prediction. The purpose of this study is to develop a model capable of predicting interior air temperature values with the longest possible forecast horizon, taking into account a certain Fanger-determined threshold for acceptable prediction error.

When carrying out a multi-step prediction, three different strategies can be used: i) *Joint method*, i.e. a single network is used to predict all forecast horizons; ii) *Independent method*, i.e. a dedicated network is used to predict each forecast horizon; iii) *Iterative method*, i.e. a single step-ahead model is used to iteratively generate forecasts. According to Kline's study [144], the Joint Method delivers better prediction accuracy for long forecast horizons (more than 4 steps-ahead). This investigation therefore uses the Joint method.

In order for the ANN to be compatible with supervised learning algorithms, the input time series data must be processed. This sequence of time series data is transformed into input and output sequences. Since the Joint method is employed, the sliding window algorithm is used to map each input sequence to an output sequence (Figure 4.5).

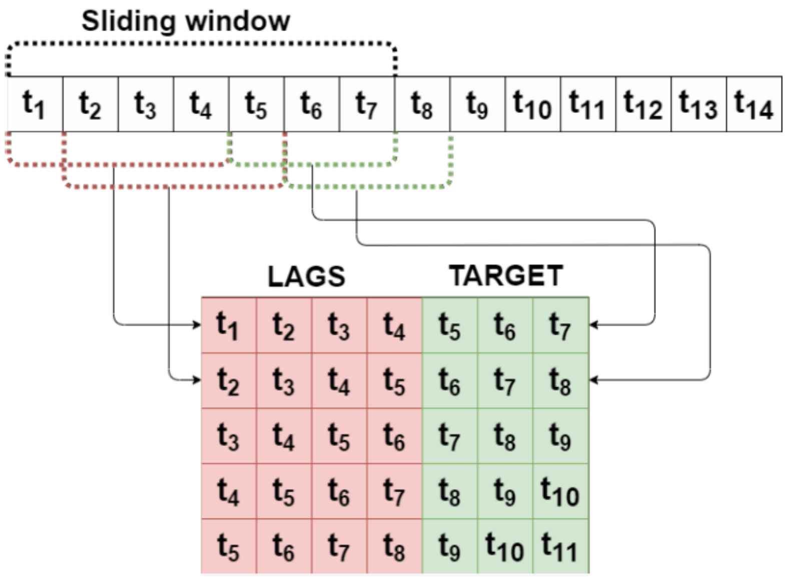


Fig. 4.5 Reframed samples

After input data has been processed and made appropriate for network training, the network must be prepared. The hyperparameters of the various ANNs must be fine-tuned in order to create network structures that are optimally suited to the prediction problem under investigation.

One-Dimensional Convolutional Neural Network

The chosen 1D-CNN architecture is based on the one proposed by Mehrkanoon in [145] and it is represented in Figure 4.6. It has the following layers:

1. Input: It accepts a dimension input *(steps, features)*, where *steps* is the prediction lag (the number of previous temperature values that are used as regressors) and the only feature is temperature.
2. Convolutional: It has a convolutional layer vector of size 2, which after the convolution operation generates an output of length equal to input dimension minus 1. The second dimensionality is equal to the number of filters, which is chosen based on the result of different trials.
3. Pooling: The pooling strategy used is the Max Pooling with pool size equal to 2, with the dimension of the output being halved as a consequence.

4. Flatten: Because of the multiple filters used there is a multidimensional output, so a flatten layer is used. In order to feed the data to the dense layer, the flatten layer is used to "unroll" the data.

5. Dense: In order to improve the learning ability of the network, an intermediate dense layer is used. The number of units of this layer is decided in the tuning phase.

6. Output: The output layer consists of a number of neurons equal to the number of forecasting horizons.

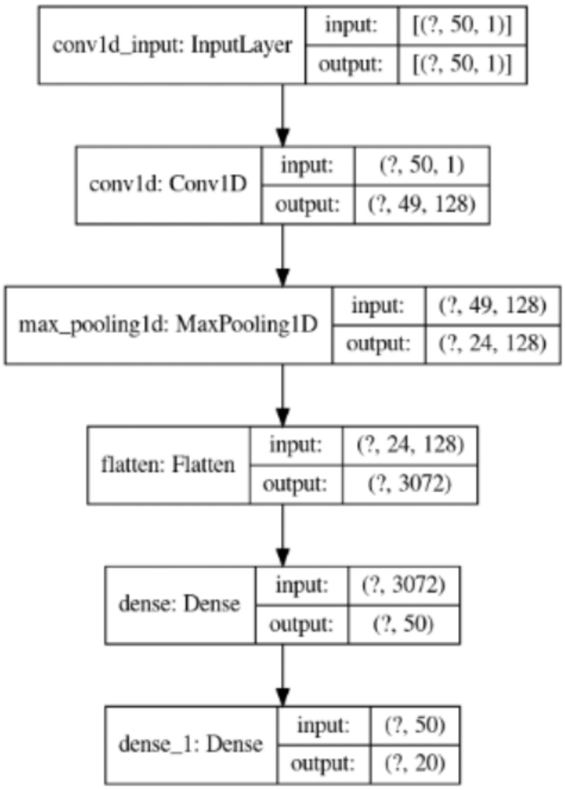
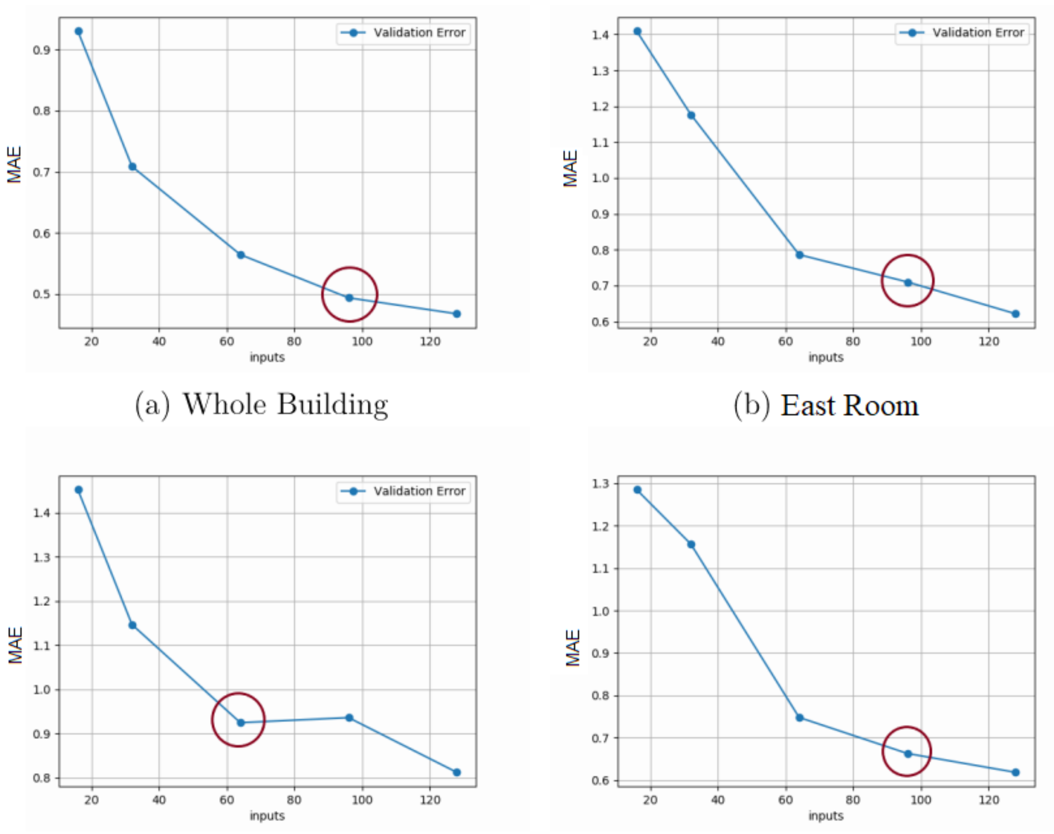


Fig. 4.6 Architecture of the 1D-CNN

Furthermore, Early Stopping technique with a patience of *8 epochs* and a delta of *0.003* is used to avoid overfitting. The activation functions chosen are Rectified Linear Unit (ReLU) for the Convolutional and Dense layers, while linear for the Output.

During the phase of tuning hyperparameters, the optimal set of parameters for the ANN with this dataset and objective is sought. Using a trial-and-error approach, these hyperparameters are tuned in the following order: prediction latency (inputs), number of hidden neurons in the dense layer, batch size, learning rate, and number of filters (filter optimization). After selecting the set of parameters, the model is evaluated on the test set for predictions up to 1 day (96 steps) in advance.

In this ANN, the first parameter to be optimized is the number of regressors. The ANN is evaluated with timesteps of 16, 32, 64, 96, and 128. To evaluate the efficacy of this parameter, the following values are maintained for the other parameters: (i) batch size is 128; (ii) filters are 64; (iii) the forecasting horizon is 7.5 hours; and (iv) the number of hidden neurons is 100. The results are shown in Figure 4.7. In addition to evaluating the accuracy of the prediction and the risk of overfitting, network complexity and training time are also considered. The goal is to determine the optimal balance between network precision and training time. Clearly, the greater the input size, the more precise the prediction. However, the maximum number of inputs is limited to 120 (30 hours in advance), as a larger value would introduce too much complexity to the network, thereby increasing both the training duration and the danger of overfitting. It is fascinating to note that although 120 neurons guarantee greater accuracy than the previous input size, the improvement is negligible when compared to the increase in training time. Following these considerations and the presented results, the input size for all environments except West Room is set to 96, while the input size for West Room is set to 64.

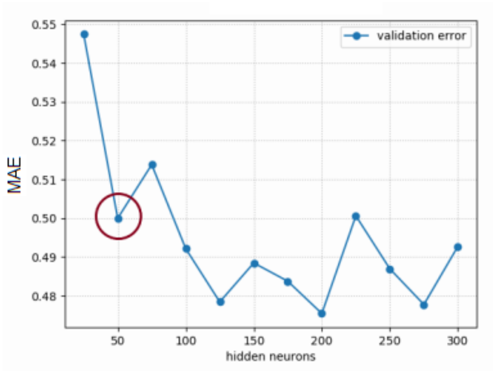


(a) Whole Building (b) East Room

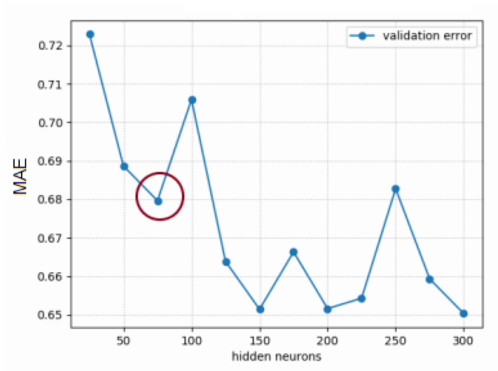
(c) West Room (d) Corridor

Fig. 4.7 1D-CNN, MAE variation against number of regressors

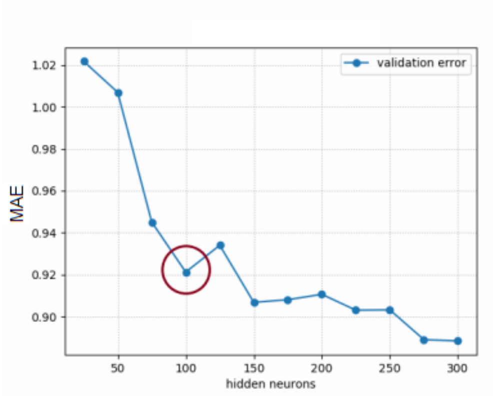
In this ANN, the second parameter to be optimized is the number of hidden neurons in the first Dense layer. The other parameters established during the optimization of the inputs remain unaltered during this optimization. The outcomes are shown in Figure 4.8. Again, the compromize between training time and prediction accuracy is a crucial factor in determining this parameter's value. The number of neurons for the various environments is kept between 50 and 100, as the results demonstrated that increasing the number of neurons does not result in a significant increase in accuracy, as evidenced by the fluctuating error value of 0.03 °C.



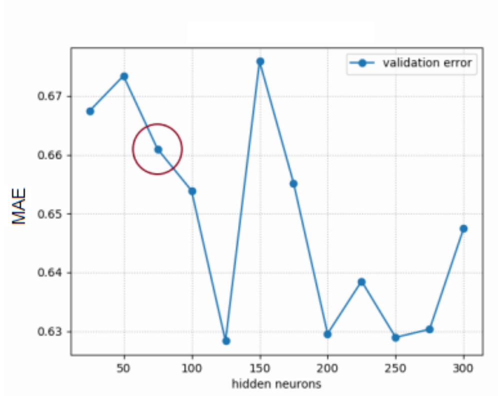
(a) Whole Building



(b) East Room



(c) West Room



(d) Corridor

Fig. 4.8 1D-CNN, MAE variation against number of neurons

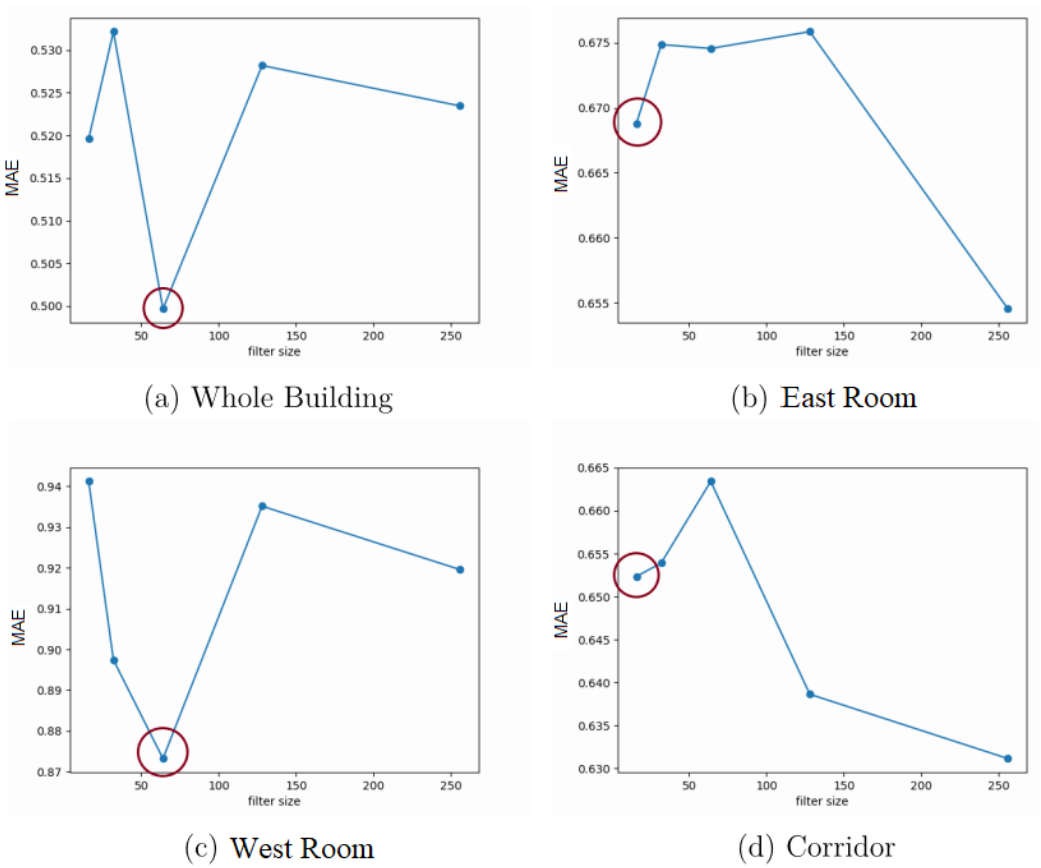
The next stage involves optimizing two hyperparameters simultaneously: sample size and learning rate. Batch size is one of the most crucial hyperparameters to adjust in contemporary deep learning systems. When training ANNs, one must take into account the behavior of such parameters in non-convex optimization. In this investigation, batch size and learning rate are considered together due to their close relationship: batch size effects the number of iterations, while learning rate affects the convergence's "pace". One would therefore anticipate that a small batch size with a low learning rate would produce comparable results to a large batch size with a high learning rate. In addition, the advantages of reduced group sizes are anticipated to include improved generalization and increased efficiency. This decrease in batch size cannot be exorbitant, however, because if the batch size is too small (for example, 1), the model will be unable to generalize and risk overfitting. Figure 4.9 highlights the results. In this instance, the selection of the optimal hyperparameters is driven by accuracy. Once comparable accuracy values have been determined, training time is considered as the second choice factor. The parameters allowing for the simplest (and quickest to train) model are therefore chosen, along with the largest group size and learning rate allowing for the same degree of accuracy to be maintained.

(a) Whole Building (b) East Room

(c) West room (d) Corridor

Fig. 4.9 1D-CNN, MAE variation against batch and learning rate

The final phase in the optimization of this ANN's hyperparameters is the optimization of the number of convolutional filters. As shown in Figure 4.10, the performance of the ANN is assessed with varying numbers of filters, as all other parameters have already been determined. Similar to the optimizations of the preceding parameters, the forecasting accuracy of the ANN is considered in conjunction with network complexity and training time. It is worth noticing that for some environments, such as the West Room and Whole Building, the performance decreases as the number of filters increases, possibly because of overfitting.



(a) Whole Building (b) East Room

(c) West Room (d) Corridor

Fig. 4.10 1D-CNN, MAE variation against number of filters

Table 4.2 resumes the parameters for this ANN chosen for the 4 environments.

Table 4.2 1D-CNN parameters

Zone	Regressors	Neurons	Batch	Learning rate	Filters
Whole Building	96	50	64	0.001	64
East Room	96	75	32	0.001	16
West Room	64	100	32	0.001	64
Corridor	96	75	256	0.001	16

Long Short Term Memory neural network

The architecture chosen for this ANN, represented in Figure 4.11, is the stateful vanilla model, which consists of a single hidden LSTM layer. The following layers make up the network:

- Input: It accepts a shape of *(batch size, lags, number of features)*.
- LSTM: Composed of LSTM cells, whose number is decided in the tuning phase.
- Output: The dimensionality of the LSTM layer matches the length of the forecasting horizons. Its activation function is the linear function.

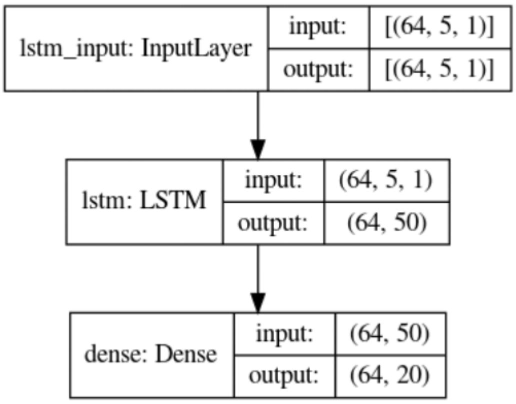
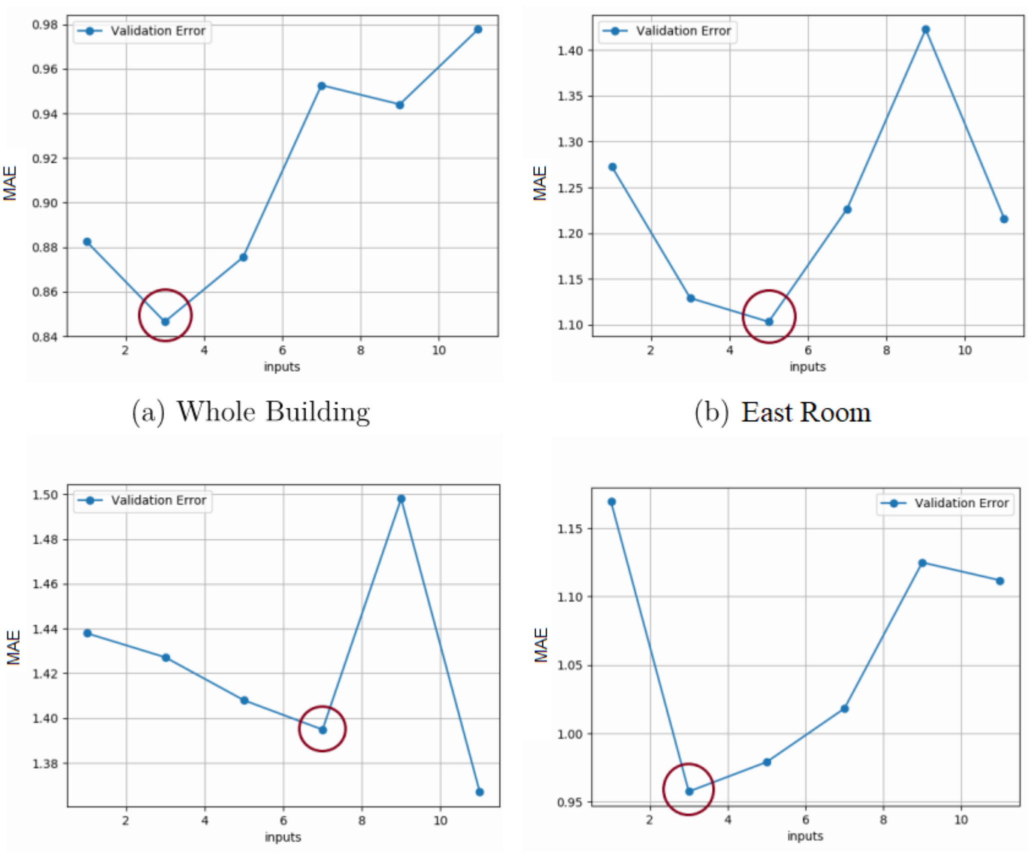


Fig. 4.11 Architecture of the LSTM ANN

This ANN architecture is being tuned for batch size, number of LSTM cells, number of regressors, and learning rate. As with the other ANN, the calibration of these hyperparameters is based on trial and error, with parameters adjusted in a particular order. (i) group size: 256; (ii) forecasting horizon: 7.50 hours; (iii) training epochs: 100; and (iv) number of LSTM cells: 50. This is the initial configuration, after which various inputs are tested to determine the optimal inputs. First, the number of regressors (inputs) is optimized, followed by the number of epochs, by analyzing the behaviour of loss during training and halting training when there is no further significant improvement. Due to their relationship, the sample size and the learning rate are optimized jointly by creating a grid with their values coupled *(learning rate, batch size)*.

This section describes the methodology used to optimize the LSTM ANN's hyperparameters. The same considerations discussed during the 1D-CNN optimization process pertaining to the selection of the most appropriate inputs, segments, and learning rate also apply to LSTM.

The quantity of regressors used by the ANN is the first parameter to be optimized. The LSTM structure is distinguished by a cell mechanism that can "remember" the state of previous cohorts, allowing for fewer inputs than 1D-CNN networks. Therefore, the network was assessed for 1, 3, 5, 7, 9, and 11 timesteps. To evaluate the efficacy of this parameter, the following values are maintained for the other parameters: (i) the batch size is 256; (ii) the forecasting horizon is 7.5 hours; (iii) the number of training epochs is 100; and (iv) the number of LSTM cells is 50. This initial configuration is replicated for all environments, and the performance of the ANN in terms of prediction accuracy is assessed with varying input sizes, as demonstrated by the results in Figure 4.12.

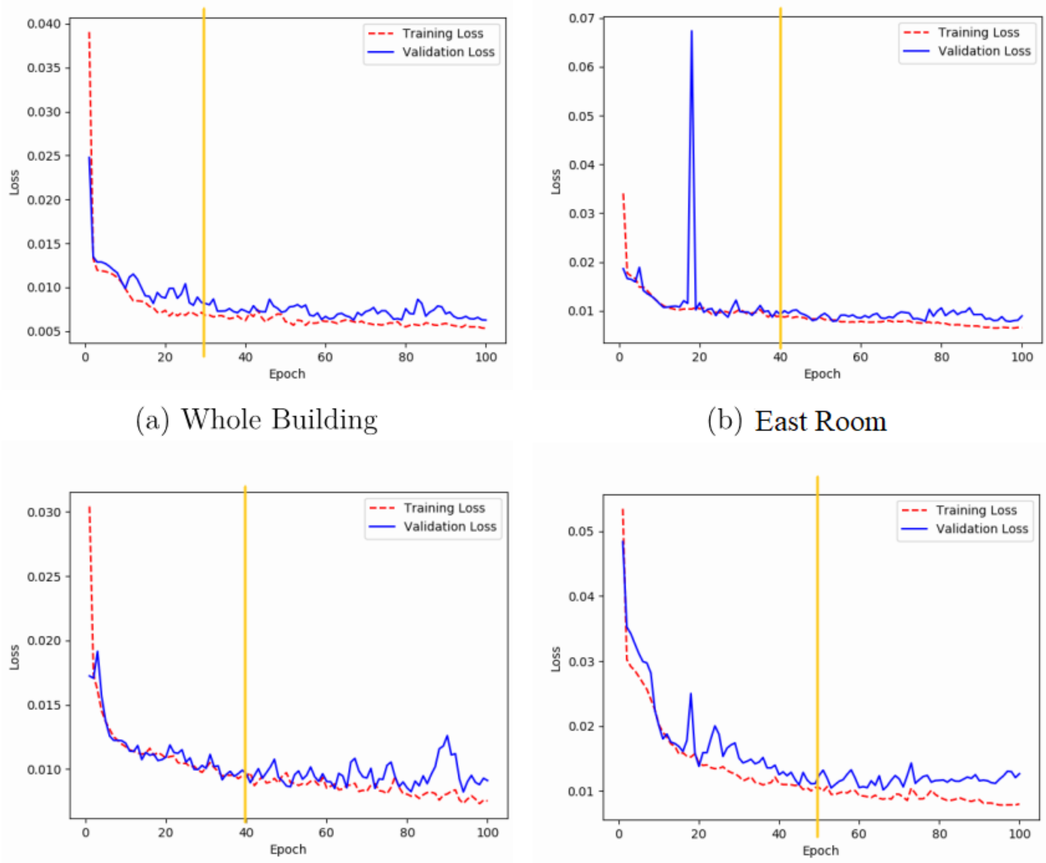


(a) Whole Building (b) East Room

(c) West Room (d) Corridor

Fig. 4.12 LSTM, MAE variation against number of regressors

Both the risk of overfitting and the enhancement in network learning are taken into account when determining the optimal value for the epochs parameter. As shown in Figure 4.13, the training can be interrupted and the number of epochs can be calculated when the loss at each epoch value becomes negligible. This optimization is conducted with a batch size of 256, which has been defined as the highest allowed batch size, to ensure a conservativepoint, because alternatively more iterations would be allowed with a smaller batch size.

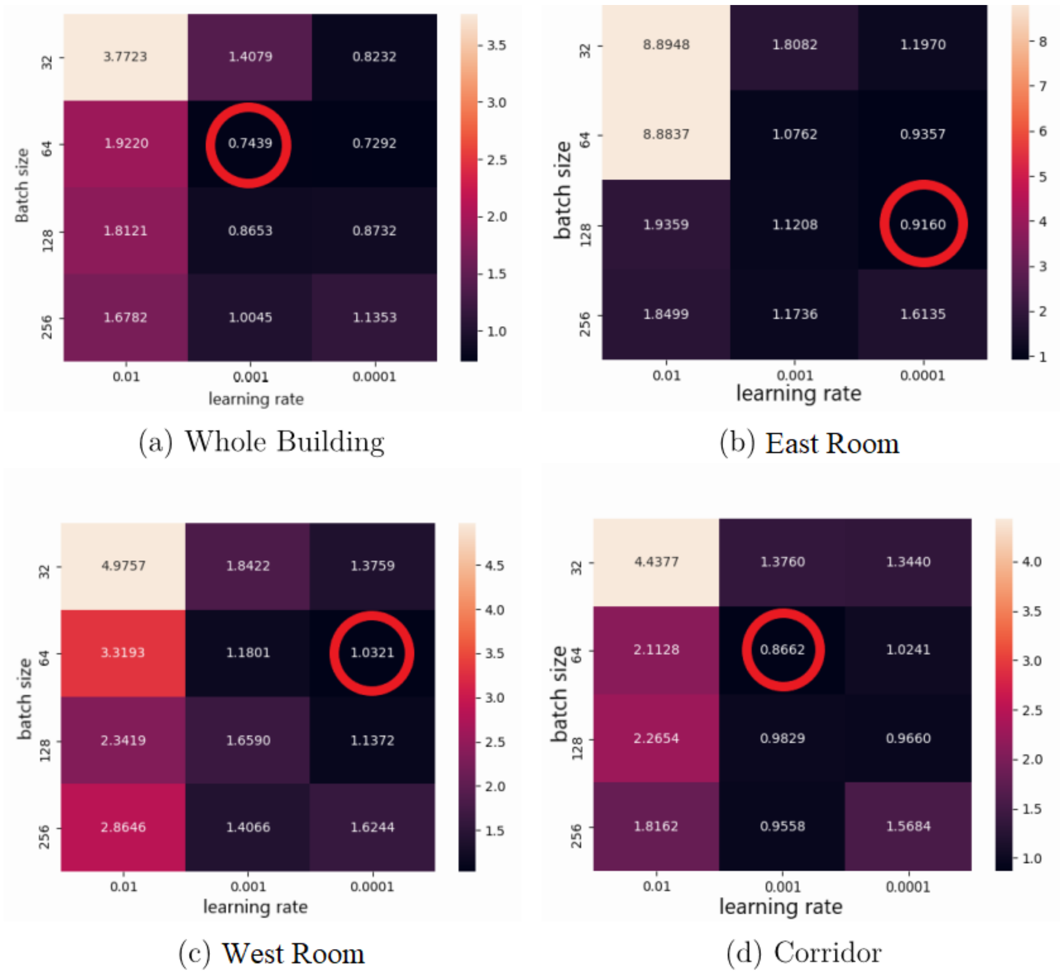


(a) Whole Building (b) East Room

(c) West Room (d) Corridor

Fig. 4.13 LSTM, epochs loss

The next step in optimizing the hyperparameters involves optimizing the batch size and learning rate, which are again considered concurrently due to their impact on the number of iterations and convergence "speed", respectively. As stated previously, batch size optimization is a crucial stage due to its influence on the model's speed, precision, and generalizability. Again, similar results are anticipated between small batch size with a small learning rate and large batch size with a large learning rate. However, smaller batch sizes are anticipated to provide better generalization, and care is taken to avoid an excessive decrease in batch size in order to maintain the model's ability to generalize and avoid over-fitting. The batch size values being evaluated are 32, 64, 128, and 256, while the learning rate values are 0.01, 0.001, and 0.0001. The results are presented in Figure 4.14, which shows that there is a high variability in the accuracy. In this case, the best accuracy performance is therefore achieved with small learning rates combined with small batch sizes.

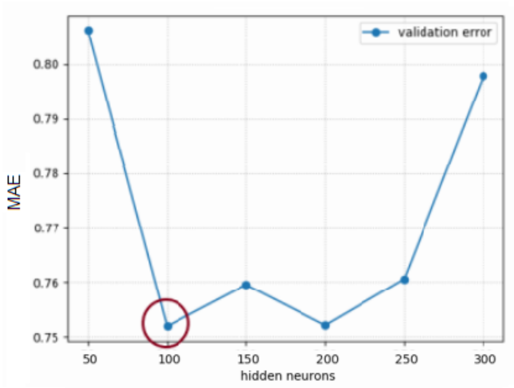


(a) Whole Building (b) East Room

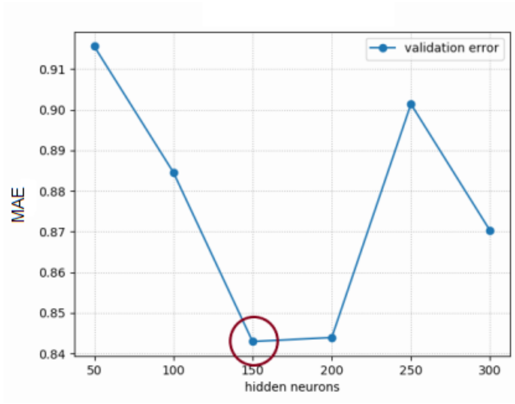
(c) West Room (d) Corridor

Fig. 4.14 LSTM, MAE varition against batch and learning rate

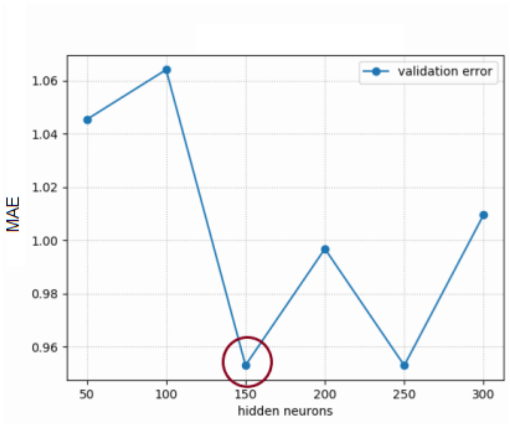
The final phase of the optimization of the hyperparameters is the optimization of the neurons, or number of LSTM cell units. All the other parameters are already identified, so the performance of the ANN is evaluated with different number of units: 50, 100, 150, 200, 250, 300. The results are shown in Figure 4.15: it is interesting to highlight how as the number of LSTM units increase beyond a certain value, the error also increases, probably because of overfitting.



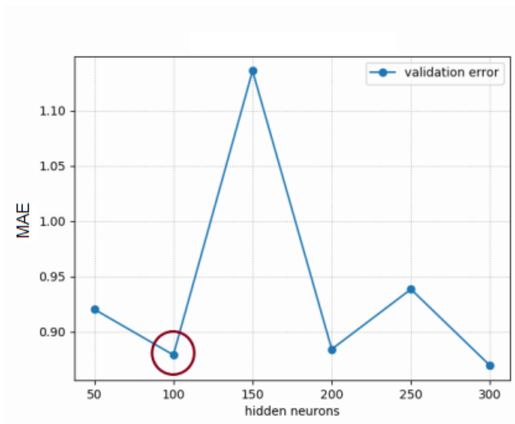
(a) Whole Building



(b) East Room



(c) West Room



(d) Corridor

Fig. 4.15 LSTM, MAE variation against cell number

Table 4.3 resumes the parameters for this ANN chosen for the 4 environments.

Table 4.3 LSTM parameters

Zone	Regressors	Epochs	Batch	Learning rate	Units
Whole Building	3	30	64	0.001	100
East Room	5	40	128	0.001	150
West Room	7	40	64	0.001	150
Corridor	3	50	64	0.001	100

The multivariate approach

The previously demonstrated methodology to be applied to the selected ANNs is being implemented with interior air temperature as the sole input feature. The purpose of this section is to introduce a new input variable, namely the day of the week.

Previous works, such as the one carried out by Deihimi et a. [43], show how if the day is also added as input variable to the ANNs, it can result in greater prediction accuracy. As highlighted previously, after observing the indoor air temperature's behaviour (see Figure 4.3), there is a discernible contrast between weekday and weekend behavior. The goal is to "assist" the ANNs in recognizing this pattern by assigning a unique designation to each day of the week. This application has been executed with both the 1D-CNN and LSTM architectures, while the remaining parameters of the original ANNs have not been modified. Due to the fact that the purpose of this application is to comprehend the effect of the second variable, the identical same architectures and parameters must be utilized. The ANNs applied to datasets that do not account for the day of the week are known as "univariate" cases, whereas the ANNs applied to datasets that include a unique label for each day of the week are known as "multivariate" cases.

Fine Tuning using real data

Starting from the models trained according to the procedure explained in the previous paragraphs, the model is then retrained through TL in three different ways: i) retrain all the layers; ii) retrain only the input layer; iii) retrain only the output layer. As discussed in [146], and in [147], the primary effects of the various TL retraining approaches are on precision and computation time. If none of the model layers are frozen, the trained model will be more accurate, but it will also require more computational time. Alternatively, if all layers except the final one are locked, the model only requires backpropagation and modifying the weights of the final layer, resulting in a significant reduction in computational time. Therefore, the various solutions are investigated.

For the purposes of this investigation, only the precision of the various TL approaches has been considered. At the outset of the investigation, it was observed that the fine-tuning of ANNs using the various TL techniques did not differ significantly in terms of time. What has been observed, however, is the time difference between training the original ANNs and fine-tuning them using TL. TL allowed for the completion of fine-tuning in a matter of hours, whereas the initial training required many months to complete. This allowed for the initiation of a greater number of TL simulations with different approaches and the collection of a greater number of results regarding the effects of various TL methods in the same amount of time. However, the underlying objective of the study was to determine whether fine-tuning using TL on a much smaller dataset (2 years as opposed to 6 years for the initial training) was capable of producing a model with an acceptable level of accuracy.

4.3 Results and discussions

Before applying Transfer Learning (TL) to the actual datasets, the initial stage in evaluating the investigation's findings is to verify the performance of the initial ANNs trained on the simulated datasets. Both the 1D-CNN and LSTM ANN are trained on each environment individually: Whole Building, East Room, West Room, and Corridor. In addition, each ANN in each environment is applied to both the univariate and multivariate cases, with the multivariate case including the day of the week label as additional input. As described in Section 4.2, following Fanger's studies [138] and the current standards to evaluate thermal comfort based on such studies [139] [140], the MAE and PMV thresholds for acceptable prediction accuracy and comfort levels are set at 2 °C (MAE) and 0.5 (PMV). The forecast horizon (in hours), that each ANN is able to achieve while maintaining within such thresholds before the application of TL, is presented in Table 4.4.

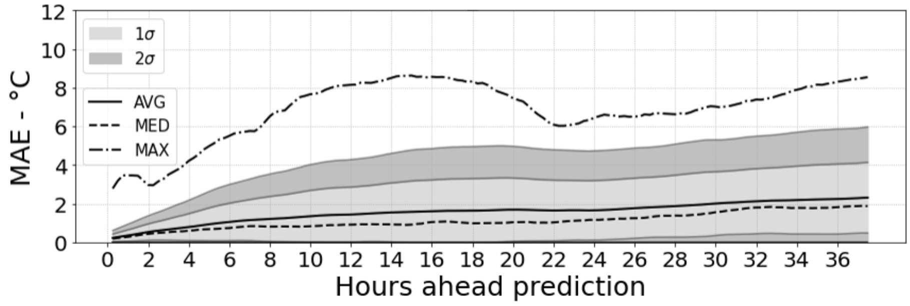
Table 4.4 Performance of original ANNs before the application of TL: the best performance for each environment is highlighted in grey

			ORIGINAL NEURAL NETWORKS	
Environment	**Case**	**Network**	**Max forecast horizon (h) with MAE <= 2°C**	**Max forecast horizon (h) with PMV <= 0.5**
Whole building	Univariate	1DCNN	32	32
		LSTM	35	36
	Multivariate	1DCNN	25	10
		LSTM	11	17
East Room	Univariate	1DCNN	36	34
		LSTM	27	36
	Multivariate	1DCNN	24	36
		LSTM	29	36
West Room	Univariate	1DCNN	28	36
		LSTM	29	36
	Multivariate	1DCNN	7	32
		LSTM	5	36
Corridor	Univariate	1DCNN	30	36
		LSTM	31	36
	Multivariate	1DCNN	25	35
		LSTM	28	36

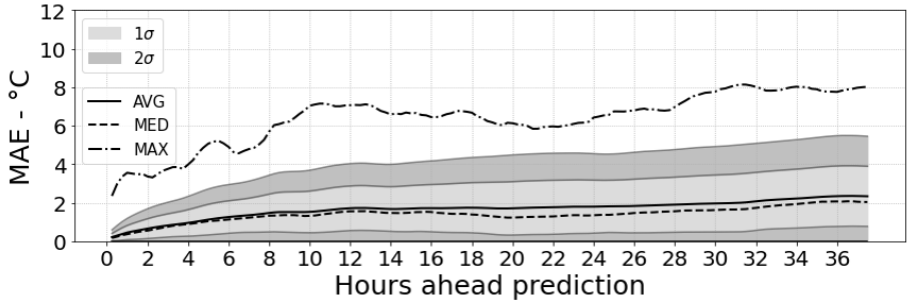
The aforementioned results provide initial, intriguing insight into the performance of ANNs. First, the MAE and PMV do not have the same forecast horizons. This disparity is nevertheless predictable, given the distinct natures of the two metrics: The MAE evaluates the ANN's precision (the difference between the $\hat{y}_t$ predicted values and the y_t real values), while the PMV evaluates the estimated state of well-being of a potential group of individuals given the $\hat{y}_t$ predicted values. Consequently, the actual forecast horizon for each ANN application is the shorter of the two forecast horizons. The preceding results indicate that, on average, the PMV forecast horizon is longer than the MAE forecast horizon. In all environments, when comparing the performance of the univariate and multivariate cases, the latter obviously underperforms the former. In terms of ANN type, LSTM ANNs appear to perform better than 1D-CNN ANNs, attaining the longest forecast horizon in all environments except the East Room.

In addition to witnessing the utmost forecast horizon attainable by the various ANN applications, it is essential to observe the contour of the graphs produced by plotting the MAE and PMV metrics versus the forecast horizon. These graphs exhibit the mean, median, maximum, 1 standard deviation, and 2 standard deviations of the MAE and PMV metrics, which provide additional insight into the performance variability of ANNs.

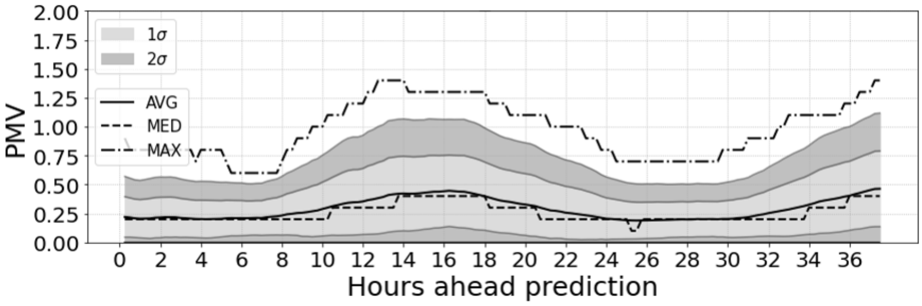
For example, Figure 4.16 depicts the univariate case graph of the metrics for ANNs in the Corridor environment. Despite the fact that the MAE graphs for the 1D-CNN and LSTM ANNs appear to be quite similar in shape, the PMV plots for the two ANNs demonstrate radically distinct behaviors. The 1D-CNN exhibits a sinusoidal trend, with increasing divergence and consequently degrading performance as the forecasting horizon lengthens, whereas the LSTM ANN exhibits a converging trend, with its performance appearing to improve as the forecasting horizon lengthens. These trends are even more significantly portrayed in Figure 4.17, which zooms under the MAE and PMV thresholds. Lastly, it is essential to observe not only the behavior of the mean, but also the median and 1 standard deviation. The 1 standard deviation limits appear equitably distributed around the mean line in both the 1D-CNN and LSTM graphs, indicating a consistent distribution in the prediction of the various outcomes for each forecasting horizon. In addition, the median line is consistently lower than the average line, particularly for the MAE diagrams, indicating that the distribution of the predictions of the various outcomes for each forecasting horizon tends to be biased toward lower, acceptable values. If the median line, as opposed to the mean line, was used to evaluate the forecasting horizon achievable by the ANN, then this horizon would increase.



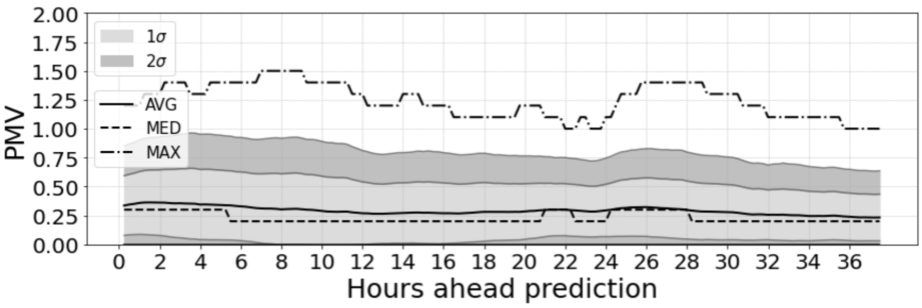
i) Corridor temperature predicted by 1D-CNN univariate original network with n_out = 150



ii) Corridor temperature predicted by LSTM univariate original network with n_out = 150

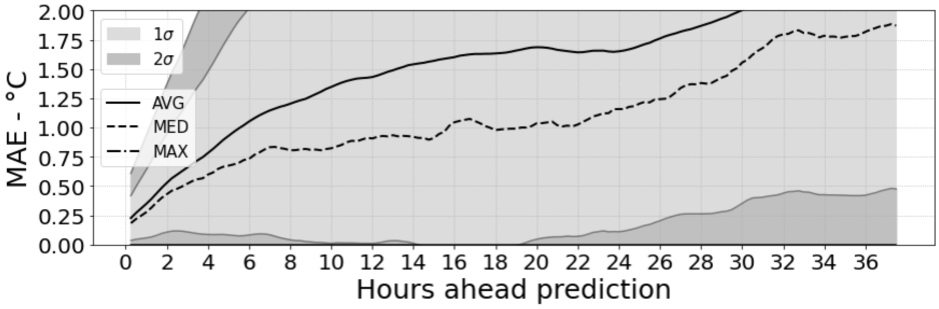


iii) PMV for Corridor temperature predicted by 1D-CNN univariate original network with n_out = 150

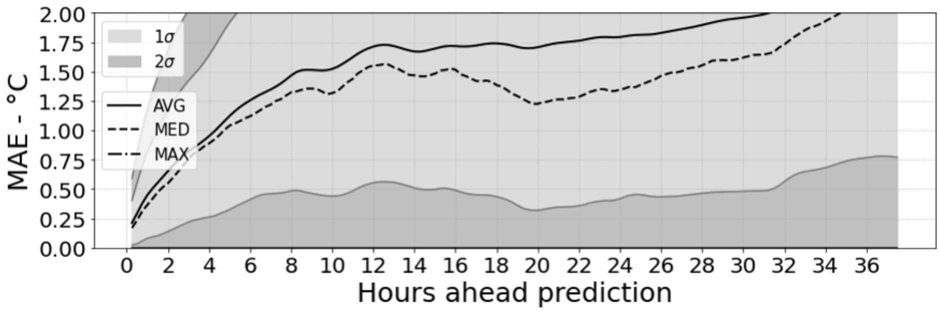


iv) PMV for Corridor temperature predicted by LSTM univariate original network with n_out = 150

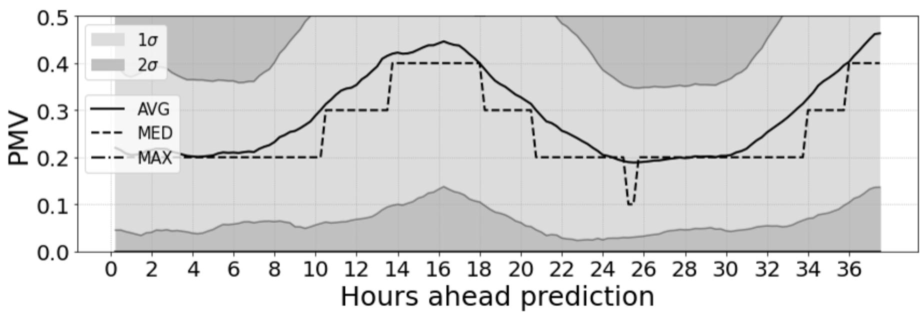
Fig. 4.16 Performance of the Corridor univariate original ANNs before the application of TL



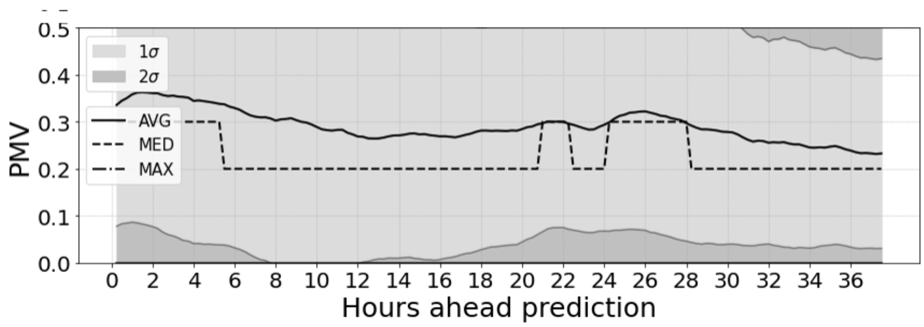
i) Corridor temperature predicted by 1D-CNN univariate original network with n_out = 150



ii) Corridor temperature predicted by LSTM univariate original network with n_out = 150



iii) PMV for Corridor temperature predicted by 1D-CNN univariate original network with n_out = 150



iv) PMV for Corridor temperature predicted by LSTM univariate original network with n_out = 150

Fig. 4.17 Performance of the Corridor univariate original ANNs before the application of TL - zoom on MAE and PMV thresholds

The performance of the different ANN is not always aligned to expectations, as shown in Table 4.4. Both ANNs of the West Room environment multivariate case, for instance, are unable to go beyond 10h when the MAE is taken into account, despite achieving over 30h of forecast horizon while maintaining the PMV value below the desired threshold. The MAE graphs for the West Room multivariate case (Figures 4.18 and 4.19), resemble those of the previously analyzed Corridor univariate case in terms of their overall shapes. The median line also stays below the average line, showing that the distribution of the predictions of the different outcomes for each forecasting horizon for these ANNs tends to be skewed toward lower, acceptable values. The 1 standard deviation limits are also evenly spread around the mean line, showing a consistent distribution in the predictions of the different outcomes for each forecasting horizon. However, these West Room multivariate graphs are much wider than the Corridor univariate ones, showing a greater variability in the difference between the $\hat{y}_t$ predicted values and the y_t real values. Furthermore, when zooming under the MAE acceptable threshold for the West Room multivariate case (Figure 4.19), and comparing the graphs to the ones of the Corridor univariate case (Figure 4.17), the graphs clearly show how the West Room multivariate MAE quickly grows beyond the set threshold, therefore indicating the ANNs' lack of acceptable accuracy.

Analyzing the graphs of the PMV values produced by the applications of multivariate ANNs in the West Room can yield similarly intriguing insights. The multivariate 1D-CNN case for the West Room exhibits a comparable sinusoidal trend as the univariate 1D-CNN case for the Corridor. In contrast to the Corridor's univariate case, the West Room's multivariate 1D-CNN case does not appear to diverge as the forecast horizon lengthens, while its oscillation amplitude increases. In contrast to the univariate instance of the Corridor, the LSTM ANN for the multivariate West Room data also exhibits sinusoidal behavior. When peering into the PMV graphs and observing their behavior below the PMV threshold, it is fascinating to note that the 1D-CNN is able to stay below the threshold for 32h of forecasting horizon, but not consistently. It actually exceeds the set limit after 8h, similar to when its MAE exceeds its threshold, and remains above such threshold for 12h before falling below the set 0.5 PMV at 20h and staying there for another 12h until the 32h forecast horizon. In contrast, the LSTM PMV value for the West Room multivariate case exhibits a more linear behavior, decreasing slightly as the forecasting horizon lengthens, just as it does for the same network type in the Corridor univariate case, albeit with higher values.

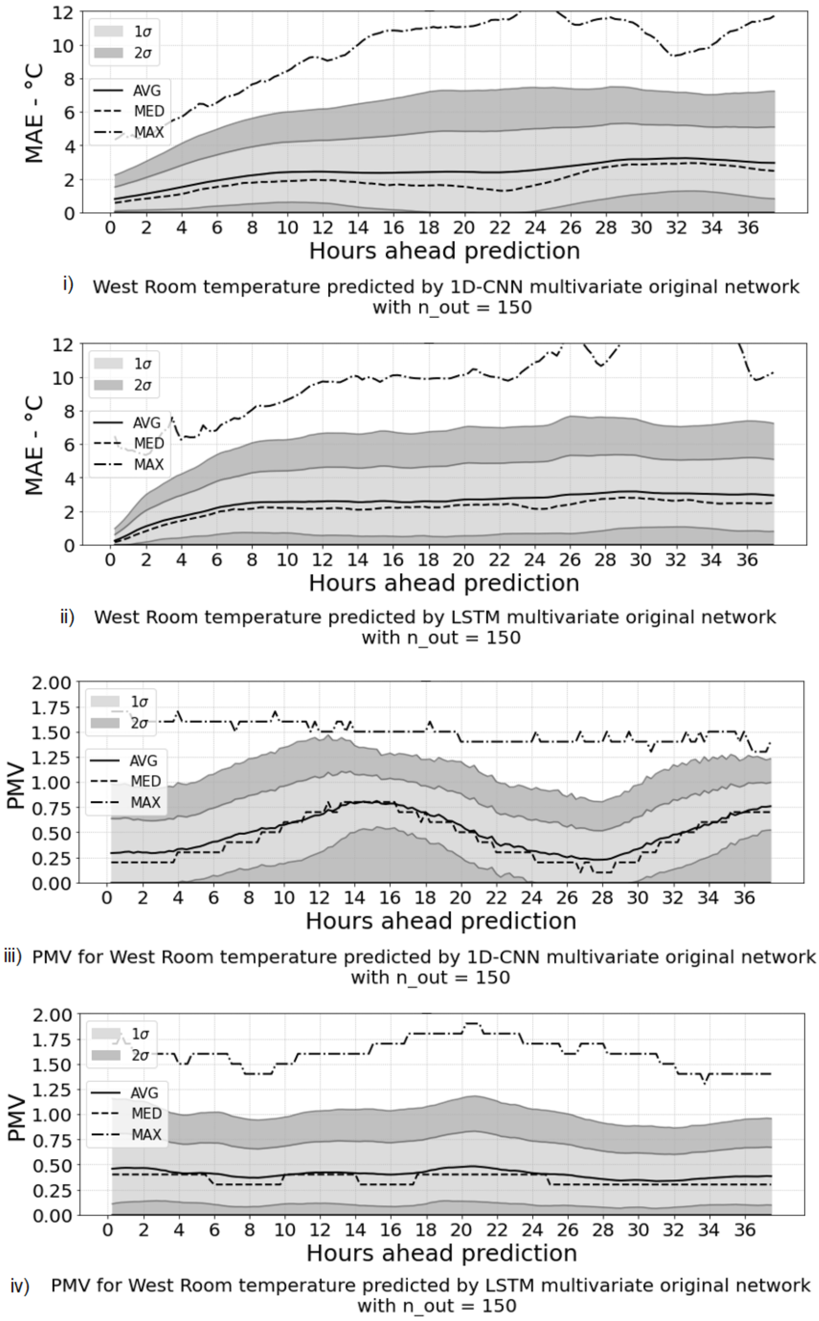


Fig. 4.18 Performance of the West Room multivariate original ANNs before the application of TL

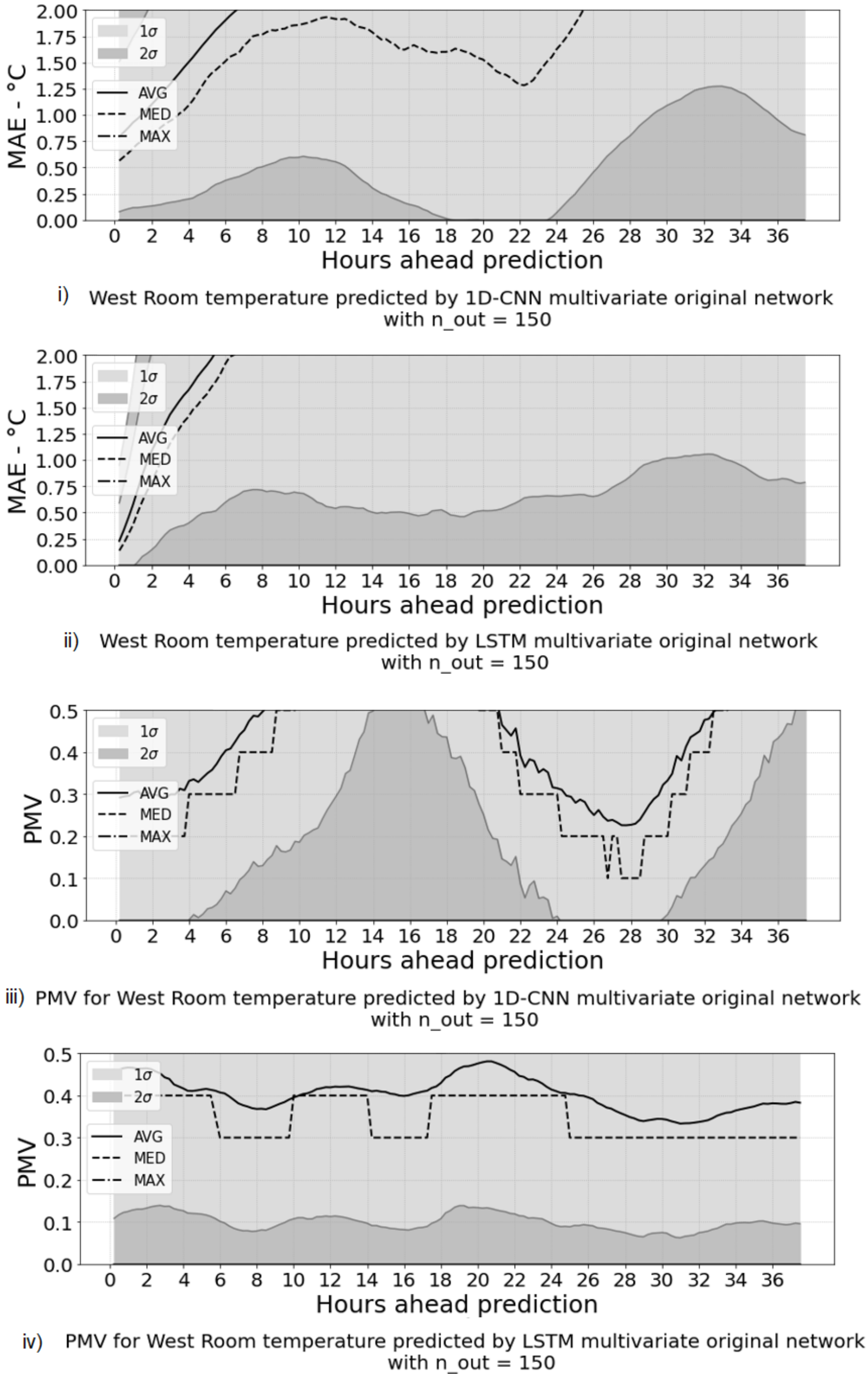


Fig. 4.19 Performance of the West Room multivariate original ANNs before the application of TL - zoom on MAE and PMV thresholds

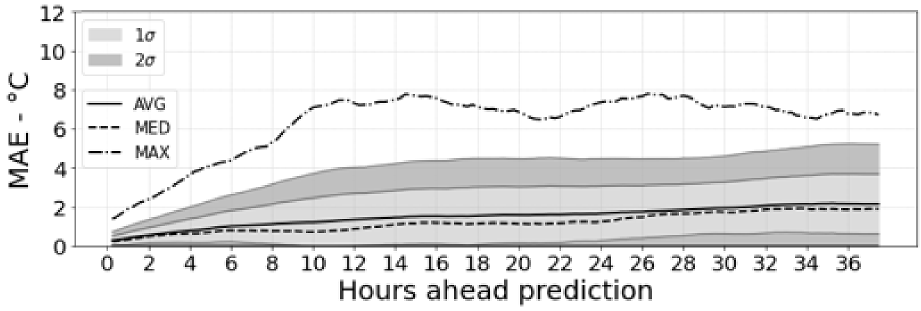
After verifying and analyzing the performance of the initial ANNs trained on simulated datasets, the various TL techniques are implemented and their results contrasted and analyzed further. The outcomes are shown in Table 4.5.

Table 4.5 Performance of ANNs after the application of TL

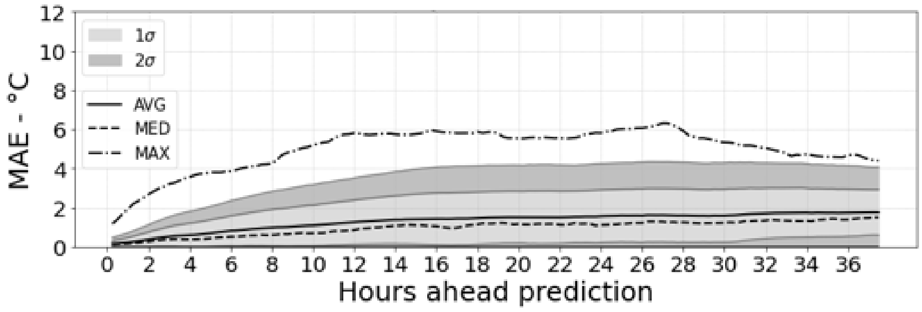
			ORIGINAL NEURAL NETWORKS		TRANSFER LEARNING WITH LAST LAYER RE-TRAINED		TRANSFER LEARNING WITH FIRST LAYER RE-TRAINED		TRANSFER LEARNING WITH ALL LAYERS RE-TRAINED	
Environment	Case	Network	Max forecast horizon (h) with MAE <= 2°C	Max forecast horizon (h) with PMV <= 0.5	Max forecast horizon (h) with MAE <= 2°C	Max forecast horizon (h) with PMV <= 0.5	Max forecast horizon (h) with MAE <= 2°C	Max forecast horizon (h) with PMV <= 0.5	Max forecast horizon (h) with MAE <= 2°C	Max forecast horizon (h) with PMV <= 0.5
Whole building	Univariate	1DCNN	32	32	36	36	36	34	23	36
		LSTM	35	36	36	25	36	25	36	32
	Multivariate	1DCNN	25	10	36	36	36	35	36	16
		LSTM	11	17	36	8	36	13	36	8
East Room	Univariate	1DCNN	36	34	36	34	36	34	36	36
		LSTM	27	36	36	20	36	18	36	22
	Multivariate	1DCNN	24	36	36	36	36	35	36	34
		LSTM	29	36	36	3	36	3	36	5
West Room	Univariate	1DCNN	28	36	33	34	28	36	14	36
		LSTM	29	36	34	36	36	36	36	36
	Multivariate	1DCNN	7	32	36	36	36	36	36	36
		LSTM	5	36	36	29	36	14	36	8
Corridor	Univariate	1DCNN	30	36	36	36	23	36	15	36
		LSTM	31	36	36	29	33	36	23	36
	Multivariate	1DCNN	25	35	19	36	36	36	36	36
		LSTM	28	36	36	25	36	23	36	24

In a heterogeneous manner, the preceding results demonstrate that the implementation of TL improves the performance of ANNs. Specifically, in 79% of cases, the MAE observes an increase in forecast horizon while remaining within the acceptable threshold, with an average increase of 13.4 hours. In contrast, the PMV experiences an increase in forecast horizon while remaining within the acceptable threshold in 27After applying one of the TL strategies, ten of the original sixteen ANNs exhibit an increase in forecast horizon for both MAE and PMV. Such improvements are, however, heterogeneous; there is no TL technique that consistently enhances the efficacy of all ANNs. In addition, there appears to be no correlation between TL approaches, the environment/case or network of the original ANN, and the performance boost.

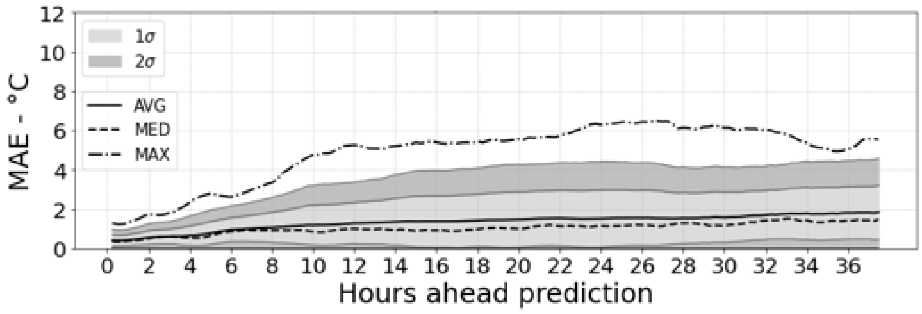
Similar to the analysis of the original ANNs prior to the application of TL, the shape of the graphs produced by juxtaposing the MAE and PMV metrics against the forecast horizon must be observed in order to evaluate the efficacy of the methodology. Figures 4.20, 4.21, 4.22 and 4.23 depict a successful application of TL to 1D-CNN univariate ANNs in the Whole Building environment. Regarding the MAE, the results demonstrate that the application of TL causes the shape of the curve to compress, thereby reducing the MAE and indicating an improvement in the network's accuracy, but only for two of the three TL techniques, when either the last layer or the first layer is retrained. The MAE graph actually deteriorates when all layers are retrained. This is further shown in Figures 4.22 and 4.23, which zoom beneath the MAE and PMV thresholds, illustrate how ANNs that apply TL and retrain only the last or first layer experience an increase in the forecast horizon from 32h to 36h, whereas when all layers are retrained, the forecast horizon is reduced from 32h to 22h. The PMV graphs then provide additional insightful information. Similar to the analysis of the original ANNs prior to TL, these PMV graphs exhibit a sinusoidal pattern. In contrast to the MAE graphs, the ANN that retrained all layers yielded the greatest results. This ANN observes a significantly reduced amplitude in the sinusoidal pattern of the PMV graph, which averages half the threshold value and never exceeds it. In addition, the 1 standard deviation shadow also appears to have shrunk substantially, remaining almost entirely below the threshold value. When only the first layer of the ANN is retrained, the graph exhibits some improvement but still exceeds the predetermined threshold, as shown in the graphs pertaining to ANNs that employ other TL techniques. When only the final layer is retrained, however, the graph significantly improves, maintaining the PMV value well below the threshold until 36h of forecast horizon and also the majority of the 1 standard deviation area. The amplitude of the sinusoidal pattern in this ANN with the final layer retrained is greater than in the network with all layers retrained, but it proves to be the most effective when both PMV and MAE improvements are taken into account.



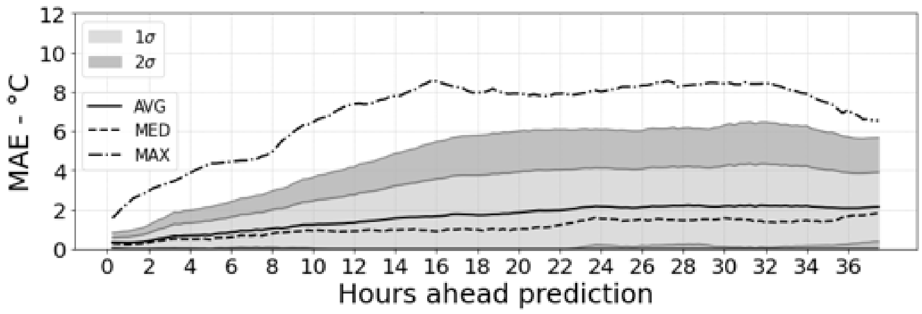
i) Building Average temperature predicted by 1D-CNN univariate original network with n_out = 150



ii) Building Average temperature predicted by 1D-CNN univariate new network with TL and last layer retrained with n_out = 150

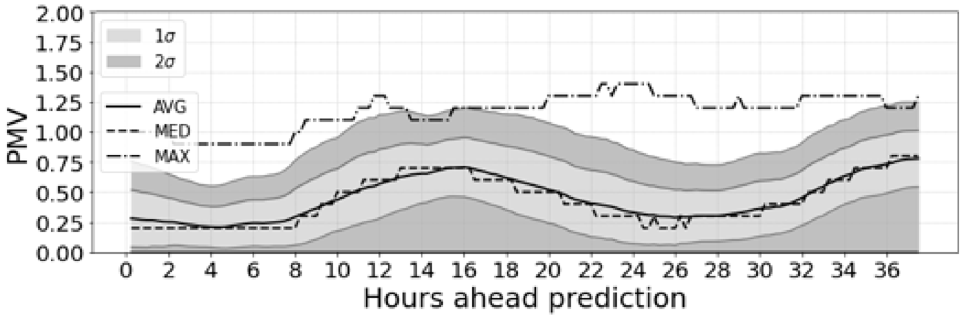


iii) Building Average temperature predicted by 1D-CNN univariate new network with TL and first layer retrained with n_out = 150

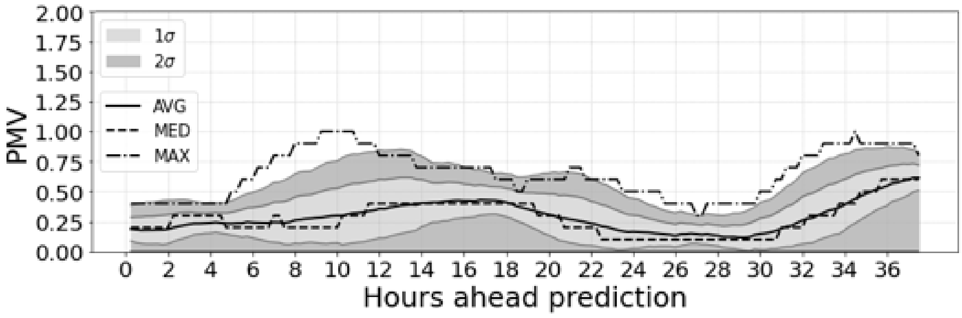


iv) Building Average temperature predicted by 1D-CNN univariate new network with TL and all layers retrained with n_out = 150

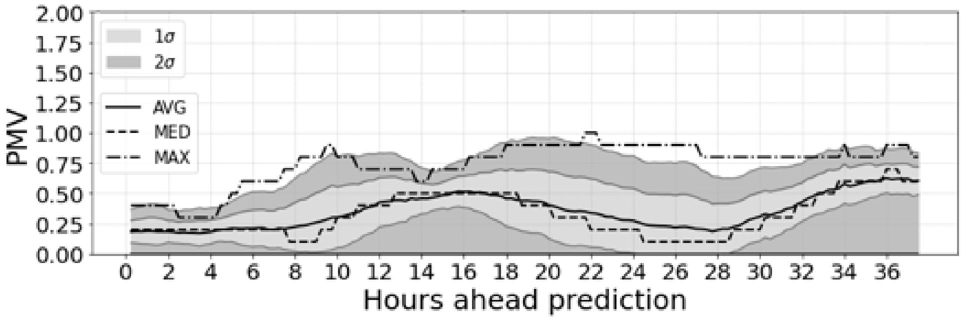
Fig. 4.20 Comparison of the different 1D-CNN univariate ANNs for the Whole Building environment - MAE



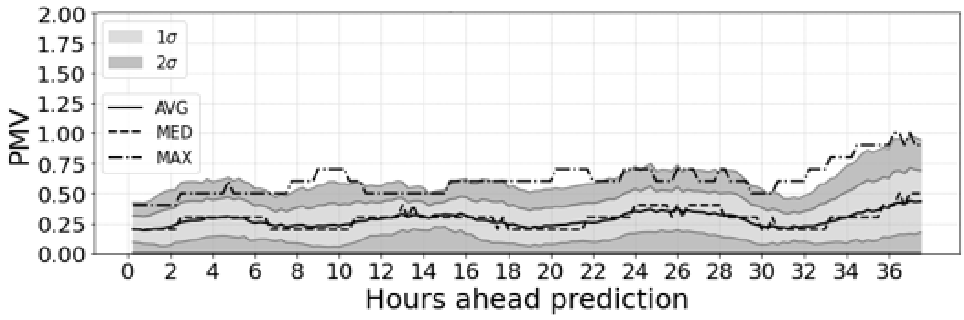
i) PMV for Building Average temperature predicted by 1D-CNN univariate original network with n_out = 150



ii) PMV for Building Average temperature predicted by 1D-CNN univariate new network with TL and last layer retrained with n_out = 150

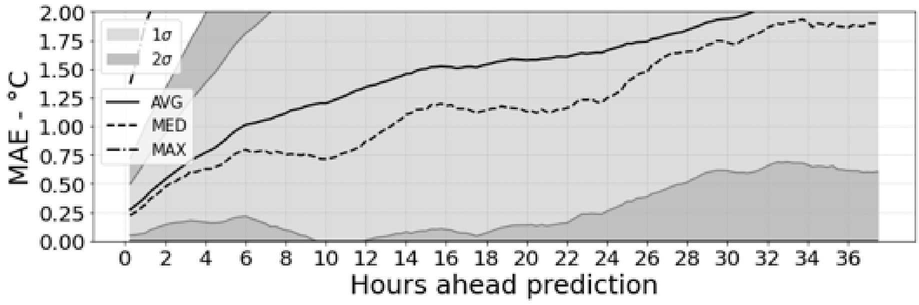


iii) PMV for Building Average temperature predicted by 1D-CNN univariate new network with TL and first layer retrained with n_out = 150

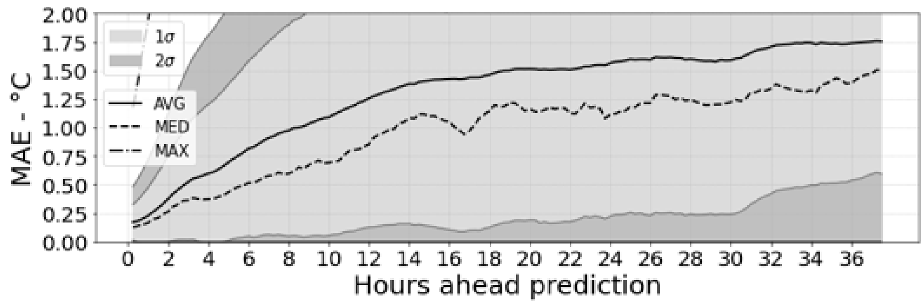


iv) PMV for Building Average temperature predicted by 1D-CNN univariate new network with TL and all layers retrained with n_out = 150

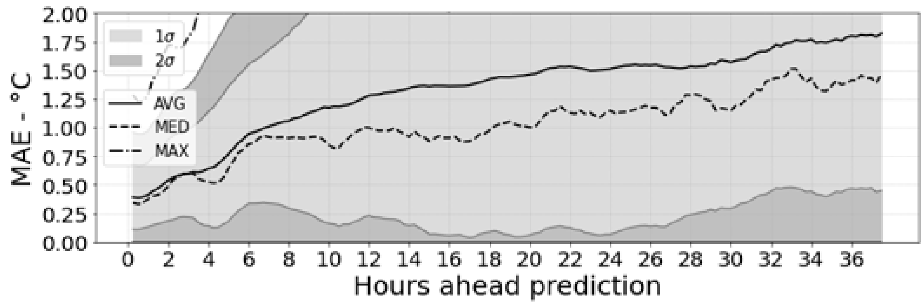
Fig. 4.21 Comparison of the different 1D-CNN univariate ANNs for the Whole Building environment - PMV



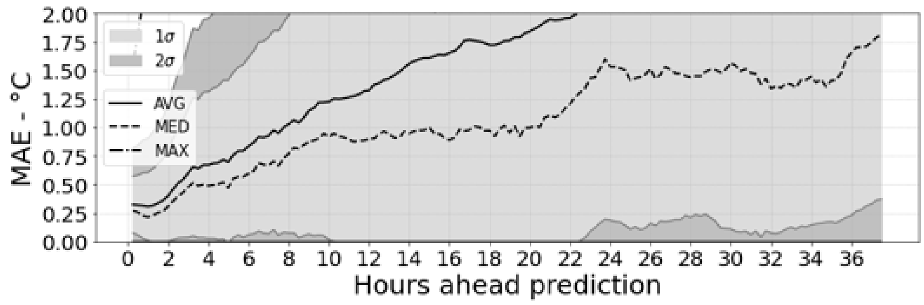
i) Building Average temperature predicted by 1D-CNN univariate original network with n_out = 150



ii) Building Average temperature predicted by 1D-CNN univariate new network with TL and last layer retrained with n_out = 150

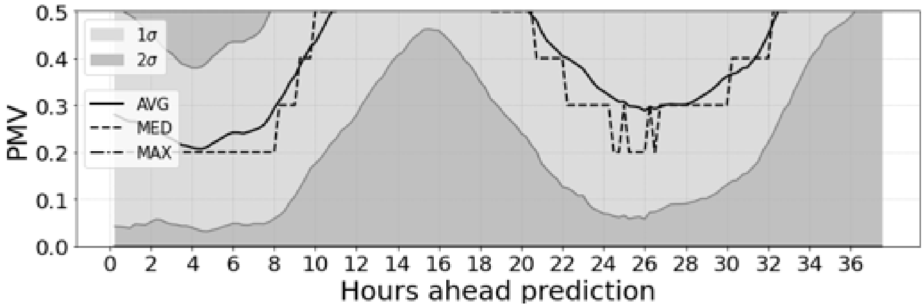


iii) Building Average temperature predicted by 1D-CNN univariate new network with TL and first layer retrained with n_out = 150

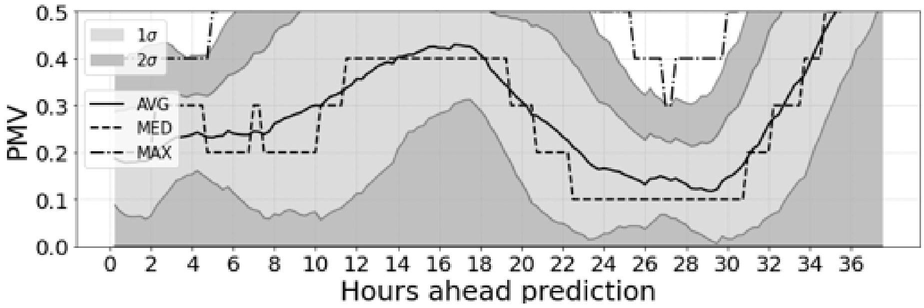


iv) Building Average temperature predicted by 1D-CNN univariate new network with TL and all layers retrained with n_out = 150

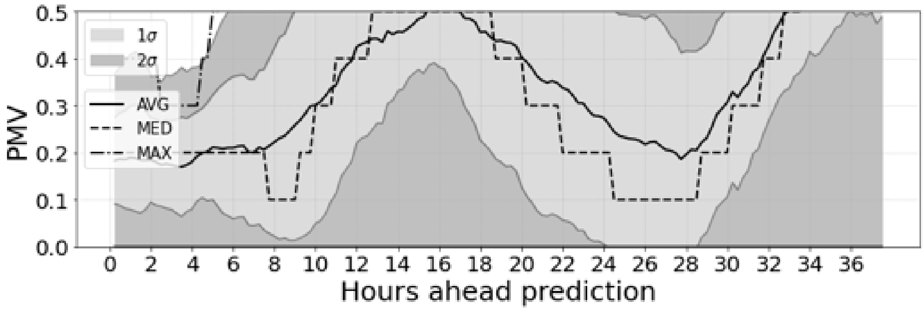
Fig. 4.22 Comparison of the different 1D-CNN univariate ANNs for the Whole Building environment - zoom on MAE thresholds



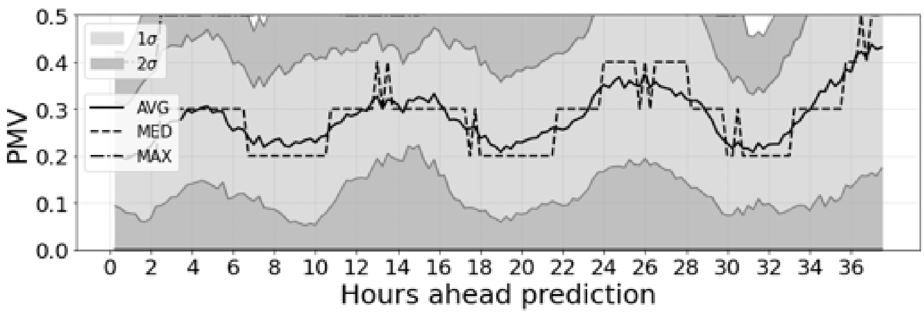
i) PMV for Building Average temperature predicted by 1D-CNN univariate original network with n_out = 150



ii) PMV for Building Average temperature predicted by 1D-CNN univariate new network with TL and last layer retrained with n_out = 150



iii) PMV for Building Average temperature predicted by 1D-CNN univariate new network with TL and first layer retrained with n_out = 150



iv) PMV for Building Average temperature predicted by 1D-CNN univariate new network with TL and all layers retrained with n_out = 150

Fig. 4.23 Comparison of the different 1D-CNN univariate ANNs for the Whole Building environment - zoom on PMV thresholds

The results of the study demonstrate how, as a consequence of TL, ANNs significantly improve their performance. In cases where the forecast horizon can remain within the MAE threshold for at least 36 hours, the situation is less intriguing. However, the most intriguing finding was that the MAE values tend to level off well below the MAE threshold. This allows for a longer forecasting horizon while maintaining acceptable accuracy in predicting the interior air temperature. However, these enhancements were heterogeneous; there is no TL technique that enhances the performance of all ANNs consistently. In addition, there is no discernible relationship between the TL approaches, the environment/case or network of the original ANN, and the performance boost.

4.4 Conclusions

The results of this study demonstrate that 1D-CNN and LSTM can effectively predict interior air-temperature even when trained on simulated datasets, and that their performance can be further enhanced through the application of transfer learning on limited real datasets. Specifically, in 79% of cases, the MAE observes an increase in forecast horizon while remaining within the acceptable threshold, with an average increase of 13.4 hours. In contrast, the PMV experiences an increase in forecast horizon while remaining within the acceptable threshold in 27% of cases, with an average increase of 8.6 hours. After applying one of the transfer learning strategies, ten of the original sixteen ANNs exhibit an increase in forecast horizon for both MAE and PMV. Such improvements are, however, heterogeneous; there is no transfer learning technique that consistently enhances the efficacy of all ANNs. In addition, there appears to be no correlation between transfer learning approaches, the environment/case or network of the original ANN, and the performance boost.

When reflecting on the validity of these results, a potential bias in the model was identified since the temperature prediction would be affected by the external temperature, especially in summer and winter. However, as explained in Section 4.2.4, the investigation's objective is to predict the indoor air temperature when the HVAC system is operating. The case study building is supplied with a heating system but not with an air conditioning system. For this reason, only datapoints from days when the heating system is operational are regarded within the datasets. The building's heating system is directly controlled by the district heating system, which is connected to the building's heating system. Between the 15th of October and the 15th of April, the district heating system is activated annually. So the data is collected only between the 15th of October and the 15th of April, which can be considered the winter period (with a little bit of autumn at the beginning, and spring at the end). Since only the data during this time interval is collected, the model is already trained with data which is embedded with the impact of lower external temperature. One could further argue that, as a consequence, the model is biased since it is trained only with data from the winter months, and would therefore be ineffective in predicting indoor air temperature when the external temperature is warmer, such as during late spring, summer and early autumn. However, as explained in Section 4.2.4, the investigation's objective is to develop a model capable of accurately predicting the indoor air temperature when the HVAC system is operating. Since the building's heating system is directly controlled by the district heating system, which is activated annually only between the 15th of October and the 15th of April, then there is no need for the model to be trained also on data coming from the warmer months.

As explained in Sections 4.1.4, 4.2.3 and 4.2.4, the contribution of this investigation is not, like for most papers related to smart building systems, the estimation of the power consumption by the HVAC to maintain the room temperature as well as the prediction of the temperature. The novelty of this investigation lies in the prediction of the indoor air temperature and on the evaluation of the accuracy of this prediction by using Fanger's model (the other novelty is using an accurate simulator, precisely designed to replicate the case-study environments to create an artificial, but accurate and realistic, dataset large enough to effectively train and test different ANN models). The methodology developed for this investigation can therefore be integrated with traditional methodologies related to smart building systems, to better evaluate not just the estimated power consumption required for the HVAC to maintain the desired room temperature, but also the effectiveness in maintaining that desired room temperature by measuring the thermal comfort for the people in the room.

As future work for this investigation, we intend to investigate potential applications in environments similar to the current one. This investigation was conducted on a building that, based on its structural characteristics, can be either public or private. Future research could examine the applicability of this methodology to industrial structures, whose structural characteristics and environmental factors differ substantially. Its applicability in the context of Smart Buildings could be the subject of a future study. Thus, DR control strategies can be evaluated in a Smart Building environment.

As discussed by Yu in [148], digital age efficiency requirements are one of the many challenges power grids have faced in recent years, and an increasing number of artificial intelligence and big data-driven solutions are emerging to advance smart-grid development. Thus, the proposed solution can make a significant contribution to the ongoing endeavor to reduce energy consumption, which is a topic of great interest in contemporary society.

Chapter 5

Forecasting photovoltaic power production

Artificial Neural Network (ANN) models can be effectively incorporated into intelligent models for energy prediction, but their training requires large data sets. This section of the dissertation presents an innovative method for forecasting photovoltaic (PV) power generation with ANNs when only a small quantity of actual data is available; the method has been tested and validated on a real PV installation. Feature selection determines which meteorological factors have the greatest impact on PV energy production. A simulator that accurately replicates an actual PV installation is used to generate an artificial, but accurate and realistic dataset of power generation that is large enough to effectively train and test various ANNs. Then, these are applied to a portion of an actual, but limited, dataset of power generated by the real PV installation that the simulator models. The remaining portion of the actual, but limited, dataset of PV power generation is used to fine-tune the ANN models using transfer learning techniques.

5.1 Introduction and State of the Art

As highlighted by the International Energy Agency (IEA), in its 2022 World Energy Outlook Report [44], in the midst of a global energy crisis and in the face of energy shortages and high prices, governments have raced to secure alternative energy sources and supplies while expediting the development of new renewable energy projects. Electricity accounts for approximately 20% of the world's total ultimate energy consumption, but its proportion of energy services is higher due to its efficacy. Investments in sustainable electricity and electrification, as well as expanded and modernized infrastructure, offer clear and cost-effective opportunities to reduce emissions more rapidly while simultaneously reducing electricity prices from their current highs. In the most affected regions of this energy crisis, for instance, it appears that lower electricity costs were closely related to higher percentages of renewable energy, and further, although not enough, benefits occurred to those customers with more energy-efficient homes and heat powered by electricity. If current growth rates for the deployment of solar PV and wind power are maintained, this will result in a much quicker transformation than anticipated by the Stated Policies Scenarios (STEPS), although this will require supportive policies not only in the dominant markets for these technologies, but globally. Some supply chains for essential technologies, such as PV, are expanding at rates that support greater global ambition. If all announced expansion plans for solar PV manufacturing come to fruition, manufacturing capacity would exceed deployment levels in the Announced Pledges Scenario (APS) by approximately 75% in 2030 and approach levels required by the Net Zero Emissions (NZE) Scenario. These renewable energy supply channels are a major source of employment development, with clean energy jobs already exceeding those in fossil fuels on a global scale, and the APS predicting that the number of clean energy jobs will increase from approximately 33 million today to nearly 55 million by 2030. Finally, it is intriguing to note that the World Energy Outlook Report also highlighted the fact that demand-side measures have generally received less attention, despite the fact that increased efficiency is a crucial component of both the short- and long-term response.

Within this framework, the increasing significance of renewable energy sources such as PV energy is evident. PV energy falls under the category of Variable Renewable Energy (VRE) sources due to its fluctuating power output derived from solar energy. PV power's variable character poses a challenge to its use as a reliable energy source in power systems, whose stability is highly dependent on the equilibrium between energy generation and consumption. According to the International Energy Agency (IEA), a power system is flexible if it can respond rapidly, within economic constraints, to large fluctuations in supply and demand by ratcheting down a generation when demand decreases and ramping up when demand increases for scheduled and unscheduled events [45]. However, the flexibility of power systems has become a concept that needs to be redefined, due to the increasing penetration levels of power generation from variable and hardly predictable sources such as wind and solar energy, which generate uncertainty on the supply side [46].

These obstacles can be surmounted with the aid of technological innovations such as smart grids, which improve the administration and stability of existing power grids by integrating them with modern distributed computational facilities and communication networks [37]. Within the framework of the smart grid, innovative applications can be implemented to better coordinate power demand and supply, such as real-time forecasting or demand response [38], which can modify power consumption to align power demand and supply. In order for such applications to function properly, however, it has become increasingly essential to forecast power demand and supply with varying horizons. Within this framework, an emphasis can be placed on the accurate forecasting of PV power generation [47].

According to [48], [49] and [50], artificial intelligence techniques have evolved into an outstanding forecasting instrument for wind and PV generation. [149] discusses how ANN-based forecasting is one of the most effective methods for PV generation forecasting, while also highlighting ANNs' main drawbacks, such as the large amount of data required for their training process, the random initial dataset required which could potentially jeopardize the prediction reliability, and the difficulty and time required for accurate development of the model architecture. However, various ANN solutions for long-, medium-, and short-term PV generation forecasting have been investigated.

For long-term prediction horizons, different methods to forecast PV power from 0h to 48h in advance were used by [150], based on spatial clustering of the PV fleet and an ensemble of Multilayer Perceptron (MLP)s using satellite and numerical weather prediction data. For mid-term prediction horizons, [151] and [152] investigate various approaches for forecasting PV power generation up to 24 hours in advance, each employing distinct ANNs and meteorological inputs, with results demonstrating how the latter improve prediction performance. Forecasting PV power generation up to 24h in advance using PV simulation software to generate the data inputs for their model was also investigated by[153], with results demonstrating that their generalized model could effectively forecast PV power generation on normal (clear-sky) and abnormal (cloudy or rainy) days, as well as in different seasons and weather conditions. For shorter prediction horizons, Long Short-Term

Memory (LSTM), Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), Bidirectional GRU, Convolution Neural Network (CNN), and hybrid architecture were used by [154] to examine one step and multi step ahead PV power generation forecasting for 1 minute, 5 minutes, 30 minutes, and 60 minutes. Various model architectures are compared to a Non-linear AutoRegressive (NAR) and an Elman recurrent ANN in terms of their performance. Results indicate that LSTM and GRU-based models obtain the highest levels of performance, with LSTM exhibiting the highest levels of precision and implementation simplicity.

With LSTM exhibiting intriguing performance when applied to PV power generation forecasting, various authors have investigated the performance of this ANN. [155] experiment with different model parameters (number of hidden nodes, activation function, number of input variables) and also by altering the division of the dataset, with their results demonstrating good performance in forecasting the daily PV power generation, while highlighting that increasing the number of input variables does not produce better results. [98] introduces CNN, LSTM, and Contextual LSTM (CLSTM) as three distinct models for PV power forecasting. All models demonstrate positive performance, with CLSTM outperforming the others in terms of accuracy and LSTM exhibiting the minimum training time. Lastly, [156] also presents an investigation utilizing ANNs for PV power prediction, including LSTM, which offers the highest prediction accuracy.

[157] examine PV power forecasting using Global Horizontal Irradiance (GHI) measured by sensors and GHI under clear sky conditions. The authors assess the efficacy of exogenous inputs in conjunction with various Machine Learning (ML) ANN models (Feed-Forward Neural Network (FFNN), Echo State, 1-Dimensional Convolutional Neural Network (1D-CNN), LSTM, and Random Forest) for short-term solar radiation forecasting. The investigation's findings indicate that the best model is LSTM, and that exogenous inputs considerably improve the forecasting performance for prediction horizons greater than 15 minutes, while the improvements are inconsequential for very brief prediction horizons (i.e. 15 minutes).

Regardless of the scope of prediction, one of the primary challenges posed by techniques belonging to the fields of ML and artificial intelligence, particularly those related to ANNs, is that a substantial amount of data is required to make sure they are effectively trained and produce acceptable accuracy [98], for example, in their work with CNN, LSTM and CLSTM, recommend to select a data length of at least 3 years.

In recent years, Transfer Learning (TL) has been investigated as a potential remedy to the problem posed by the scarcity of large and reliable enough data in numerous domains. It has also been investigated for use in PV power forecasting, but, to the best of our knowledge, there is still a dearth of research on this topic.

[158], for instance, propose a method for transferring the knowledge gained from historical solar irradiance data to the prediction of PV power output using an LSTM model that is trained with historical solar irradiance data and then fine-tuned with PV output data. The author employs six months of historical solar irradiance data and forty-five days of historical output data with a 10-minute sampling interval. Transfer Learning (TL) improves the performance of the prediction, with Mean Absolute Percentage Error (MAPE) improving by an average of 23% and Root Mean Square Error (RMSE) improving by an average of 10%, for forecasts up to 40 minutes in advance.

In another investigation led by [159], using a Constrained LSTM (C-LSTM) model in conjunction with two parameter-transferring strategies, they combine TL and deep learning models to address the challenge of effectively executing hourly day-ahead PV power generation in newly constructed PV plants. The results demonstrate that C-LSTM models outperform standard LSTM models in terms of forecasting accuracy and that the proposed combination of C-LSTM and TL strategies can improve variability and accuracy issues caused by various sky conditions.

[160] devised a model based on TL to predict PV power generation, with experimental results demonstrating that the proposed TL model outperforms traditional learning methods. The model they developed uses the variation of solar altitude angles throughout the year to identify the season, combines it with the meteorological factors hidden in the data collected from a PV system, and uses it as input for online learning models based on both traditional and TL approaches to predict power generation.

In this context, the objective of this study is to present an innovative method for forecasting PV power generation using ANNs when only a limited quantity of actual data is available. Feature selection is initially employed to investigate various meteorological features, such as GHI, humidity, air temperature, etc., in order to identify those that have the greatest influence on the accuracy of data prediction forecasts. The PV power generation simulator presented by [104], which accurately simulates actual PV installations, is then utilized. As a case study, we chose a PV system erected on the rooftops of a number of structures on the Turin, Italy, campus of our university. Consequently, the PV power generation simulator has been applied to these rooftops in order to generate an artificial, but accurate and realistic dataset of PV power generation sufficient for training and testing various ANNs. As a second dataset, we also obtained measurements of the actual power generated by these PV systems in the real world. Notable is the coincidence between actual and simulated PV installations. In the proposed methodology, the simulated dataset and previously selected meteorological features are used for the initial training and evaluation of ANNs. The ANN models resulting from training and testing on the simulated dataset are then applied to a portion of the actual dataset to assess their prediction performance using real data. Different TL techniques are then employed to fine-tune the ANN models with the remaining portion of the real dataset in an effort to enhance the prediction performance of PV power generation against the same real data. As stated previously, the entire methodology was verified and validated on an actual PV installation on our university's campus.

The novelty of this study lies in the use of a PV power generation simulator, which accurately models a real PV installation, to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test different ANN models, which are then utilized on a portion of the real, but limited, dataset of the real power generated by the real PV installation on which the simulator is based. The application of different TL techniques to tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, evaluating their efficacy to enhance the prediction performance of PV power generation always against the same real data, adds to the novelty of the study. LSTM, which is a well-established ANNs in PV power generation forecasting, and 1D-CNN, which is a variation of CNNs that have been intermittently used in PV power generation forecasting, are the ANNs utilized in this study.

After feature selection, the variables that have the greatest impact on power prediction are global horizontal irradiance, humidity, temperature, dew point, ultraviolet index, sunshine duration, and time of day; LSTM ANNs provide the best prediction performance up to 4 hours; TL techniques can successfully improve short-term forecasting performance up to 2 hours.

The rest of this manuscript is organized as follows. Section 5.2 introduces the case study. Section 5.3 presents the proposed methodology. Section 5.4 discusses our experimental results. Finally, Section 5.5 provides our concluding remarks.

5.2 Case Study

The methodology presented in this work, aiming to forecast PV power generation with ANNs, when only a limited amount of real data is available, tested and validated on a real-life PV installation located on the rooftop of a building of our university campus in Turin, Italy, as shown in Figure 5.1.

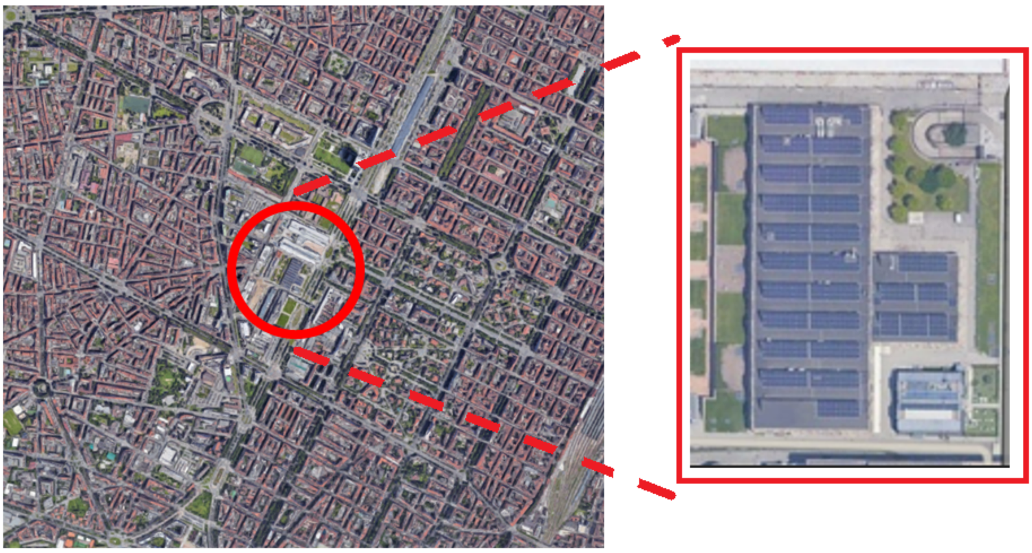


Fig. 5.1 University campus location in Turin, Italy, and its PV system under analysis

The PV installation has a total surface area of almost 3,000 m^2, and the characteristics described in Table 5.1.

Table 5.1 Technical specifications of the PV system

INSTALLATION YEAR	2016	
SIZE	631 kWp	
# OF PV CELLS	96	
POWER PV CELLS	3,406 Wp	
BRAND PV CELLS	BEN-Q	
# OF PV MODULES	1836	
BRAND PV MODULES	BENQ SOLAR	
MODEL PV MODULES	SunForte PM096B00	
MODULES' TECHNOLOGY	MONO-CRISTALLINO	
MODULES' NOMINAL POWER	327 W	
MODULES' SURFACE AREA	1,63 m2	
# OF INVERTERS	8	19
MAX POWER INVERTERS	20 kW	25 kW
BRAND INVERTERS	SMA	
MODEL INVERTERS	SUNNY TRIPOWER	

Specific sensors were used to mesaure the PV power output of this installation, and data is collected every 15 minutes. These sensors collect data for the years 2018 through 2020, for a total of 105,216 data points.

The PV power generation simulator presented by [104], which replicates the actual PV installation located on the roof of our university campus in Turin, Italy, is then used to generate an artificial, but accurate and realistic dataset of PV power generation large enough to effectively train and test various ANNs. This artificial, but accurate and plausible dataset, along with the previously selected meteorological features, is used for the initial training and evaluation of ANNs. In addition, the simulated dataset provides PV power production every 15 minutes. The simulator was used to generate a dataset consisting of 210,336 data points encompassing the years 2010 to 2015.

The artificial and real datasets will then be further subdivided, with each subset being utilized in a distinct phase of the training, testing, exploitation, and refining of the ANN models and TL.

The simulated data is divided and used to train and validate the various ANN models. The years 2010 to 2014 (175,296 data points) are used as the training set, while 2015 is used as the assessment set (35,040 data points).

The ANN models are then exploited and their prediction performance against real data evaluated, by using the real data set as input, on which different TL techniques are applied to tune the ANN models and improve the prediction performance of PV power generation against the same real data. The year 2020 (35,136) is utilized to evaluate the prediction performance of ANN models using actual data (inference set). Years 2018 and 2019 (70,080 data points) are used as the training set to tune ANN models with TL (tuning set). Year 2020 (inference set) is used as the test set for the TL models, and their performance is compared to that of the original ANN models used in the same year (test set).

5.3 Methodology

This section aims to present the proposed method for forecasting photovoltaic power generation when only a limited amount of real data is available, by utilizing different meteorological data, a PV power generation simulator that accurately models a real PV installation, various ANNs, and TL techniques.

As shown in Figure 5.2, the different meteorological data features are first collected and then analyzed through feature selection methodologies in order to identify those which most impact the accuracy of data prediction forecast. The PV power generation simulator presented by [104], which accurately models the real PV installation located on the roof of our university campus in Turin, Italy, is then used to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test various ANNs. This artificial, but accurate and realistic dataset, along with the previously selected meteorological features, is used for the initial training and testing of two distinct ANNs: 1D-CNN, and LSTM. The ANN models trained and evaluated on the simulated dataset are then applied to a portion of the real, but limited, dataset of the real power generated by the real PV installation upon which the simulator is based in order to evaluate their prediction performance against real data. Different TL techniques are then used to fine-tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, and their ability to improve the prediction performance of PV power generation against the same real data is investigated.

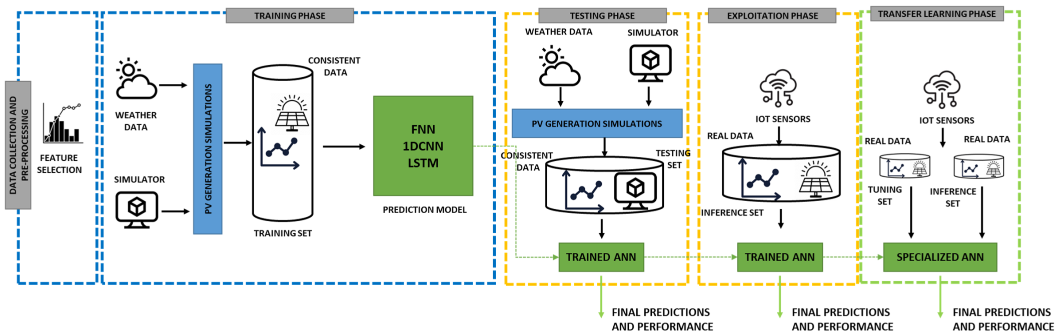


Fig. 5.2 Scheme of the proposed methodology

5.3.1 Data collection, preprocessing and Feature Selection

In addition to the actual PV generation data, a dataset containing various meteorological characteristics must be compiled for the specified time period. The data must then be preprocessed so that it is coherent and consistent across the entire dataset. After determining which of these features are most correlated with PV power generation (Feature Selection), these features are fed as inputs to ANN models.

Data collection and preprocessing

The initial raw dataset, which is presented in the following list, is composed by 22 features and 385,718 rows, referring to years 2010 to 2020. All features are collected with 15 minute intervals.

- Real PV power generation, i.e. real power (W) generated from the PV installation (data available from 2018 to 2020);
- Global Horizontal Irradiance (GHI) (W/m^2);
- Relative Humidity, ranging between 0 and 1;
- Air temperature (°C);
- Air temperature - day (°C);
- Wind speed (m/s);
- Wind direction;

- Atmospheric pressure (hPa);
- ultraviolet (UV) index;
- Temperature (°C);
- Apparent temperature (°C);
- Humidity, ranging between 0 and 1;
- Wind speed (m/s) (wind speed Dark Sky);
- Wind bearing, the direction that the wind is coming from in degrees, with true north at 0° and proceeding clockwise;
- Dew point, the point at which dew can form (°C) below the atmospheric temperature (it changes with respect to pressure and humidity);
- Precipitation intensity, the intensity of precipitation at the given time;
- Precipitation probability, the probability of precipitation occurs, between 0 and 1;
- Cloud cover, the percentage of sky occluded by clouds, between 0 and 1;
- Simulated PV power (W) (simulated power);
- Sunset time, the unix timestamp when sun will set during a given day (s);
- Sunrise time, the unix timestamp when sun will rise during a given day (s);
- Sunshine duration, the difference between sunrise and sunset time (calculated);
- Day;
- Hour;
- Minute.

Real PV power generation is the actual power (W) generated by the PV installation on our university campus (data available from 2018 to 2020), as measured by the sensors present in the PV installation. The meteorological characteristics are derived from various sources. The weather station on our university campus, which is very near to the case study's building (see Section 5.2), provides GHI and atmospheric pressure data. Dark Sky [161] is a software company specializing in weather observations and visualization that provides global historical weather data from multiple sources (in Turin, the nearest station for data collection is near the city's airport, 14.5 kilometers north of the case study's building). The simulated power is obtained via the simulator [104], sunset and sunrise times are obtained via the pvlib python library [162], sunshine duration is calculated by subtracting sunrise time from subset time, and day, hour, and minute are embedded within the data.

After obtaining the data for the meteorological features, data cleansing and feature engineering were performed to clear the dataset.

Data cleaning

Table 5.2 presents a summary of the different features present in the dataset, highlighting the missing datapoints. The minimum and maximum values are also presented, after abnormal values for minimum and maximum were identified and replaced through linear interpolation.

	Data Type	Min	Max	# of Missing Samples	Source
Real PV power generation	float64	0.00	562.20	0	Campus PV installation sensors
GHI	float64	0.00	1139.20	21.032	Campus Weather Station
relative humidity	float64	0.09	1.00	8.670	
air temperature	float64	-10.90	36.20	880	
air temperature - day	float64	-6.00	33.20	772	
wind speed	float64	0.00	12.40	41.798	
wind direction	float64	0.00	360.00	0	
atmospheric pressure	float64	800.20	1008.40	41.798	
UV index	float64	0.00	10.00	12.666	Dark Sky
temperature	float64	-11.60	36.20	5.791	
apparent temperature	float64	-13.50	36.20	5.791	
humidity	float64	0.05	1.00	4.390	
wind speed Dark Sky	float64	0.00	13.59	39.191	
wind bearing	float64	0.00	359.00	43.731	
dew point	float64	-22.25	24.32	1.890	
precipitation intensity	float64	0.00	12.954	45.968	
precipitation probability	float64	0.00	1.00	45.968	
cloud cover	float64	0.00	1.00	48.325	
simulated power	float64	0.00	9.32	0	PV simulator
sunrise time	int64	1262329669.00	1609398464.00	0	Embedded with other data
sunset time	int64	1262361479.00	1609430242.00	0	
sunshine duration	int64	31532.00	56258.00	0	

Table 5.2 Raw dataset summary information

Between the 6th of September 2016 and the 31st of December 2017, the weather station on our campus was not operational, resulting in irregular or nonexistent data collection. This issue affects the GHI, wind speed, atmospheric pressure, and wind direction features. As a result, the data for that period (a total of 21,032 values) is discarded, not just for the compromised features but, to be consistent, for the entire dataset, to make it self-consistent, and will not be used in any subsequent training and/or validation of ANNs and TL.

During the night, all attributes related to PV power generation were set to 0 (Real PV power generation, GHI, UV index, simulated power). This was achieved by taking advantage of the sunrise and twilight periods. Then, specific lacking values were identified and, whenever possible, replaced. A strong, positive correlation between UV index and GHI was identified, so absent UV index values were filled in using existing UV index values for comparable GHI. Finally, random absent values in various features were substituted using linear interpolation.

Feature engineering

Feature engineering is the process of transforming unprocessed data into ML-compatible features. In particular, features scaling is required when the variables have vastly different magnitude orders and must therefore be normalized prior to being input into the models. Normalized features all possess the same magnitude, which accelerates the training of ML models. Applying the min-max normalization, each variable is scaled between 0 and 1.

Feature Selection

The objective of feature selection is to identify those features that have the greatest impact on the ability of models to predict PV power generation, with the intention of reducing the number of input variables and, by extension, computational effort. Guyon et al. [163] identify three primary feature selection categories:

- filter method: filtering is done using the correlation matrix and is most commonly carried out using Pearson Correlation.
- wrapper methods: it requires a ML algorithm and uses its performance as evaluation criteria.
- embedded methods: an iterative process, during which each step of the training process of the model is analyzed, in order to identify the features that most contribute to the training.

In the following sections these techniques applied to our dataset are presented in detail.

Feature Selection - Filter methods

With this method, the process of identifying the features does not depend from any ML algorithm, with correlation criteria (test scores) serving as the selection criteria. This technique avoids overfitting because it does not rely on a ML algorithm, but the selected subset of features is not optimal and may contain redundant variables. Our methodology employs Pearson's correlation criteria and mutual information as ranking criteria.

Correlation Criteria

The Pearson's correlation coefficient is used as a measure to quantify linear dependence between two variables. Its values vary from -1 to 1, where 1 means maximum positive linear correlation, -1 maximum negative linear correlation and 0 means no correlation.

Pearson's correlation formula for two variables x and y is presented in the following equation:

$$R = \frac{\sum_{i=1}^{n}(x_\mathrm{i} - \overline{x})(y_\mathrm{i} - \overline{y})}{\sqrt{(x_\mathrm{i} - \overline{x})^2 (y_\mathrm{i} - \overline{y})^2}} \tag{5.1}$$

Figure 5.3, presents a confusion matrix with the results of the correlation coefficients computed between each variable, excluding the simulated power (since it is equivalent to the real power). The confusion matrix reveals a positive linear correlation of 85% between UV index and GHI features, of 50% between real PV generation and UV index, and of 40% between real PV generation and GHI. Strong correlation (80%) appears to exist between sunshine duration and temperature, and between dew point and temperature, but such correlations are not useful for this investigation. The correlation between real PV generation and all other features appears to be very low, as that between GHI and all other features.

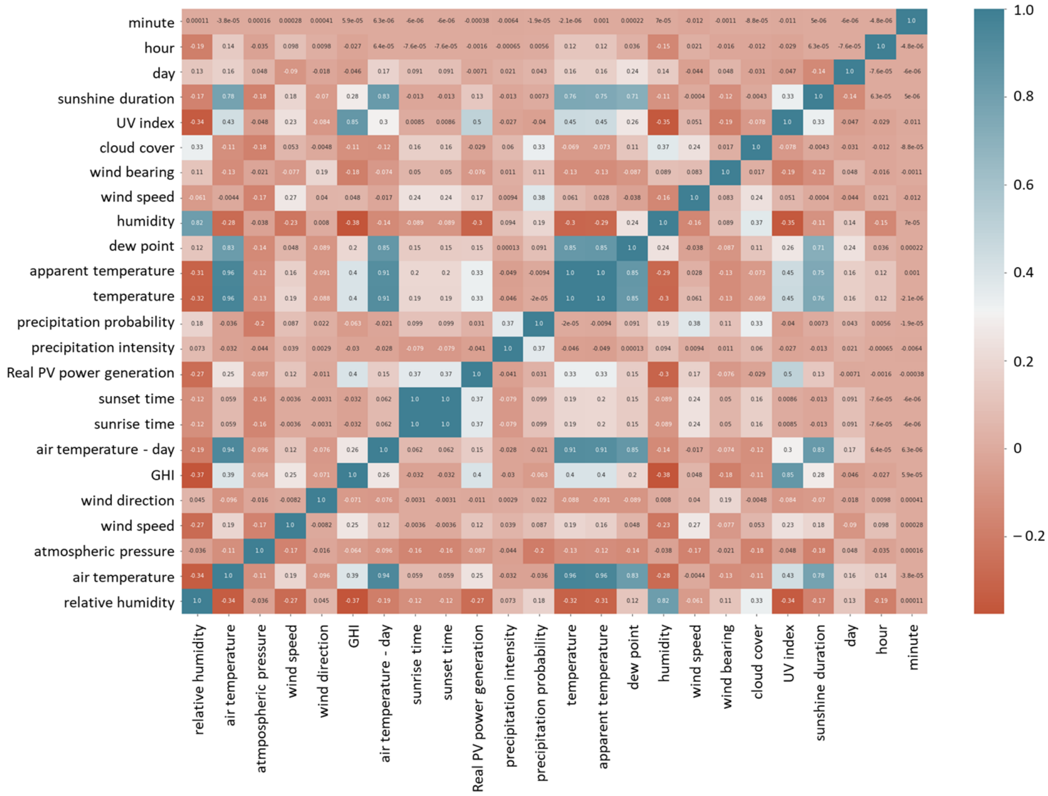


Fig. 5.3 Confusion matrix presenting correlation coefficients computed between each variable

Since Pearson's correlation can be positive or negative, a new coefficient is adopted, the coefficient of determination Coefficient of Determination (R^2), which is defined as the square of the Pearson's coefficient R. R^2 provides a more interpretable measure in order to compare the linear correlation between variables, and is used to evaluate the correlation between the different features and the real PV power generation, as presented in Figure 5.4. According to these results, UV index is by far the feature with the highest correlation to PV power generation, with its R^2 coefficient at 0.25, almost 60% greater than the second feature, GHI. The next eight features (in decreasing order, GHI, sunset time, sunrise time, temperature, apparent temperature, humidity, relative humidity, and air temperature) fall between an R^2 range of 0.15 and 0.06. The next seven features (wind speed Dark Sky, dew point, air temperature - day, sunshine duration, wind speed, atmospheric pressure, wind bearing) present an R^2 between of 0.03 and 0.01, and the last features (precipitation intensity, precipitation probability, cloud cover, wind direction, day, hour, minute) present an R^2 of 0 a slightly above.

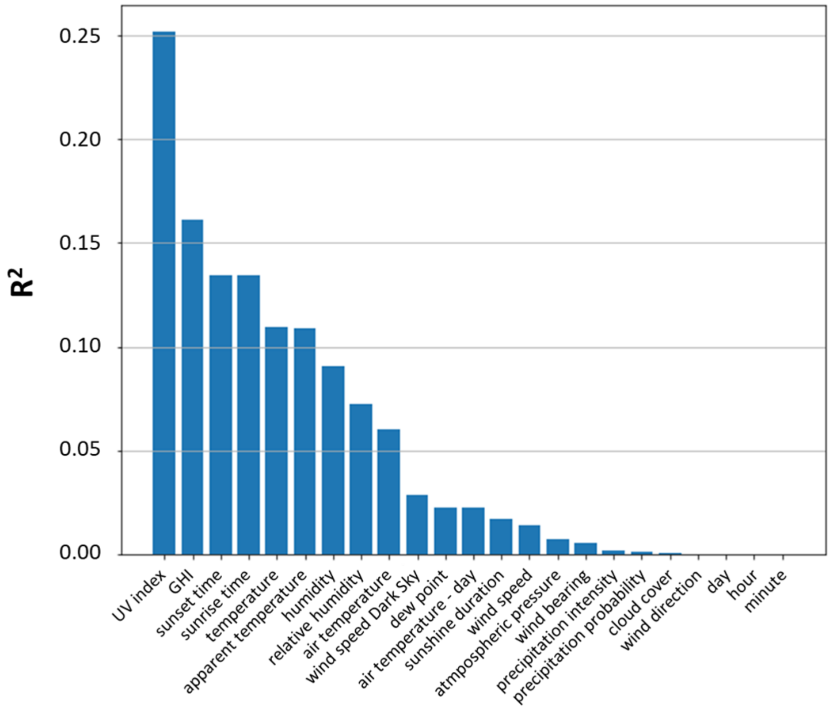


Fig. 5.4 Correlation criteria results between variables and P-PV Cittadella

Mutual information

The mutual information index measures how much can be learned from one variable by observing the other variable (or mutual dependence between two variables) [164], represented by the statistical dependence between the density of a variable x and the density of a variable y.

The formula for mutual information is:

$$MI = \int_x \int_y p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) dxdy \tag{5.2}$$

The probability densities of x and y are $p(x)$ and $p(y)$ respectively, with the joint probability density given by $p(x,y)$. Compared to the Pearson's correlation, the mutual information index is also able to recognize non linear correlations between variables. Figure 5.5 presents the features sorted by their mutual information compared to the real PV power generation (P-PV Cittadella), with results showing how the best feature is by far the GHI, with a mutual information greater than 1.5, over twice that of the following features. UV index and hour are then respectively the second and third features, with mutual information around 0.7. The remaining twenty features have significantly lower mutual information, ranging between 0.3 and 0.

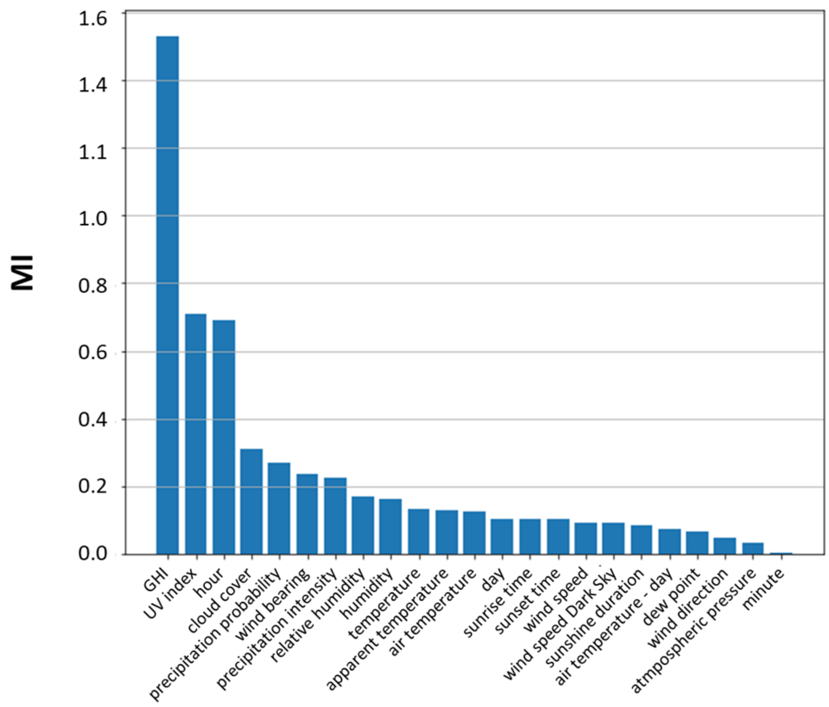


Fig. 5.5 Mutual information criteria results between variables and Real PV power generation

Feature Selection - Wrapper Method

In the Wrapper Method, a specific ML algorithm guides the feature selection process, with the objective of solving an optimization problem by evaluating all possible combinations of features using a technique known as greedy search, and identifying the best ones in comparison to the actual PV power generation. This method improves efficacy and may reduce the possibility of data overfitting. However, as the number of features increases, so does the complexity. Sequential forward and backward selection are the most commonly employed algorithms.

Sequential forward selection

The SFS (Sequential Forward Selection) algorithm begins with a null model and, at each stage, adds the performance-maximizing features. This process will continue until adding new features begins to degrade performance. This procedure does not take into account dependencies between variables, which is a drawback. This study employs a Linear Regression model. To train the model, the training and testing portions of the dataset must be separated. This division will be unique to the training of this model for this form of feature selection, and therefore will not adhere to the described data division from Section 5.2: this training set consists of two years' worth of data, from 2018 to 2019, whereas the test uses the most recent year of data (2020). After normalizing the features with min-max normalization, the algorithm is fed the variables and their Mean Squared Error (MSE) on the test set is evaluated. Figure 5.6 presents the results compared to the real PV power generation, with the most important features being GHI, sunshine duration and humidity, while the worst feature is the wind speed.

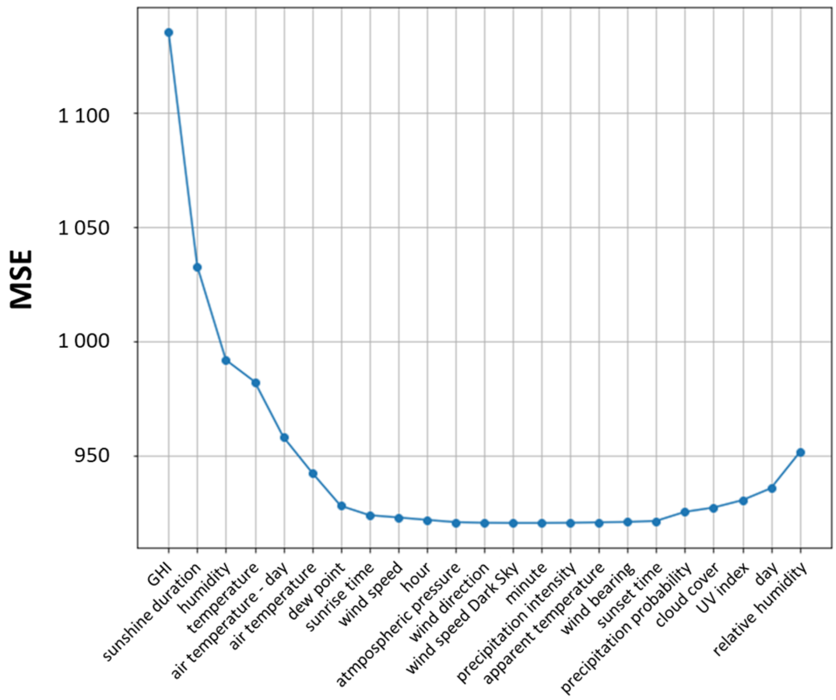


Fig. 5.6 SFS result criteria results between variables and Real PV power generation

Sequential backward selection

Unlike the sequential forward selection method, which begins with an empty model, the SBS (Sequential Backward Selection) algorithm begins with a complete set of features and, at each stage, removes those that decrease prediction performance. This method has the advantage of evaluating features in the presence of other variables, thereby eliminating unnecessary features. The training and evaluating sets are identical in the case of sequential forward selection. The algorithm deletes those features whose deletion produced the lowest MSE on the test set. Figure 5.7 depicts the results in comparison to the actual PV power generation, with GHI, air temperature, and air temperature - day being the most significant factors and precipitation probability being the worst.

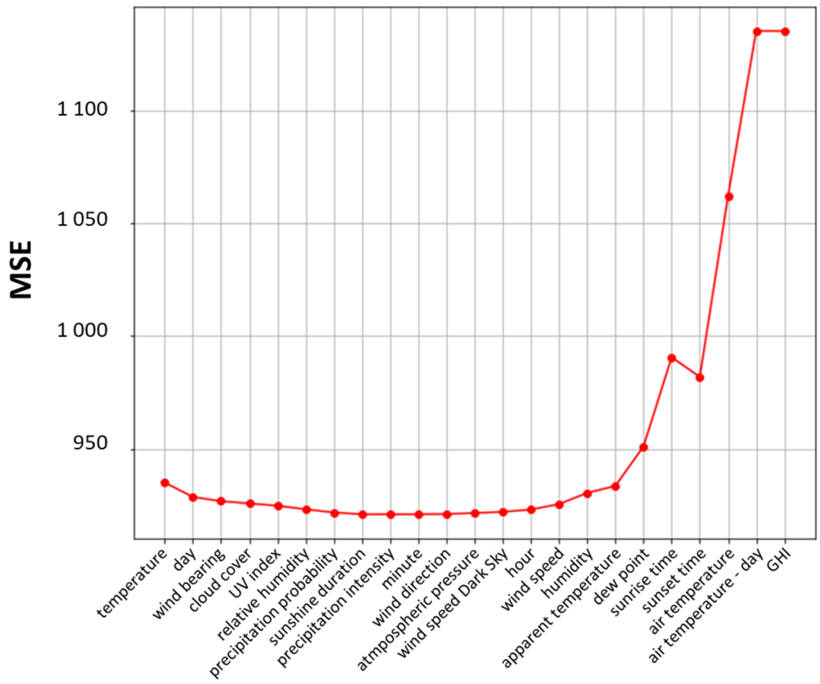


Fig. 5.7 SBS results between variables and Real PV power generation

Feature Selection - Embedded methods

In embedded methods, the feature selection algorithm is implemented into the model's learning algorithm, incorporated into the training procedure, and incorporates the characteristics of filter and wrapper methods. Embedded methods do not require splitting the dataset into training and test sets, and LASSO Regression and Random Forest are the most common embedded techniques.

LASSO Regression

Least Absolute Shrinkage and Selection Operator (LASSO) regression is a Linear Regression that uses l_1, and adds a regularization term called alpha to the cost function.

The cost function is defined as:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(\theta^T x_i - y_i)^2 + \alpha\sum_{i=1}^{n}|\theta_i| \tag{5.3}$$

where the number of training instances is given by m ,the number of input variables is n, x_i is the input vector, y_i is the target, the parameter vector of the model is θ and α is the regularization hyperparameter.

LASSO Regression is similar to features selection because l_1 regularization sets the weight of least significant features to zero. This regularization is done by varying the regularization of the hyperparameter $alpha$. The training set is composed by two years worth of data, from 2018 to 2020, with the features being normalized between 0 and 1 by using min-max normalization. The best model with lowest MSE value obtained from the training set is identifying by varying the model's hyperparameters, with $alpha$ equal to 0.001 being the best. Figure 5.8 presents the features sorted by weight, with the best ones compared to the real PV power generation being GHI and temperature.

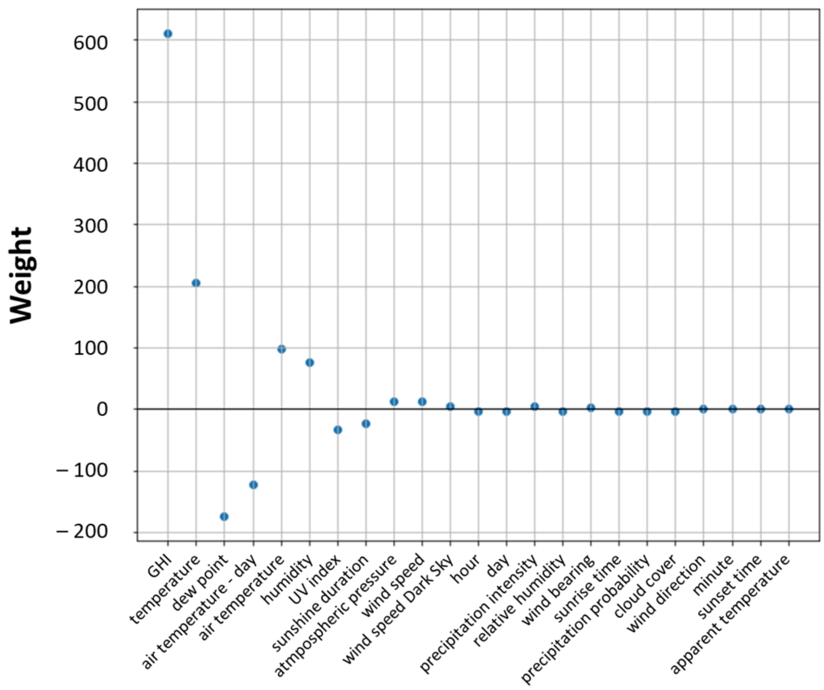


Fig. 5.8 LASSO Regression results between variables and Real PV power generation

Random Forest

Random forest is an algorithm for ML that integrates the results of multiple decision trees into a single output. Different random subsets of the training set are used to train each of these decision trees. The most significant characteristics are located near the tree's root, while the least significant ones are near the foliage. The disadvantage of this method is that the features may overfit the decision tree algorithm, resulting in the elimination of key features. In this study, 50 estimators are used for the years 2018 to 2020. Figure 5.9 presents the results compared to the real PV power generation, with GHI being by far the most important feature with a value greater than 0.8.

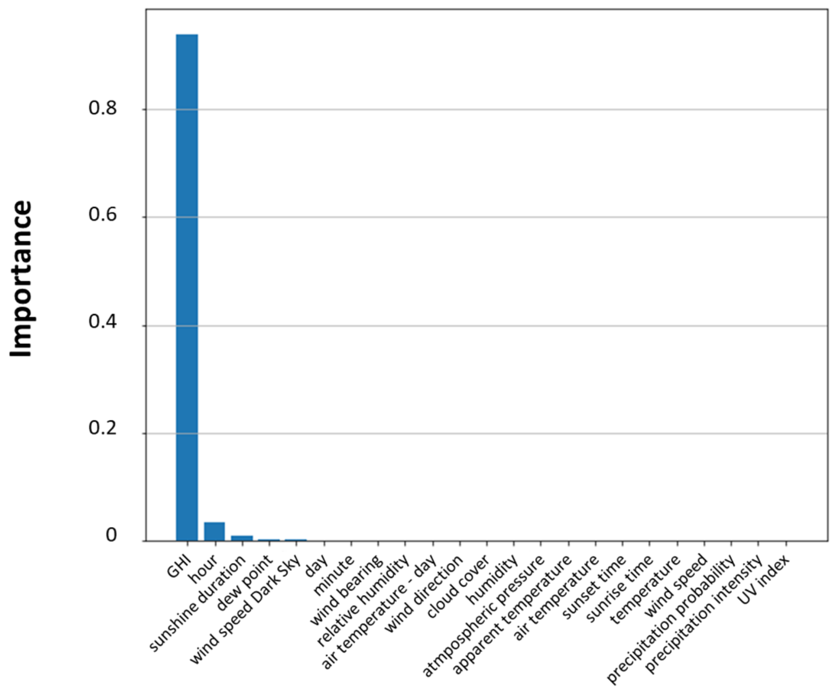


Fig. 5.9 Random Forest results between variables and Real PV power generation

Results of features selection

The purpose of feature selection is to identify the most pertinent features relative to the actual PV power generation, which are then used as inputs to the ML models, along with the PV power generation data. Table 5.3 displays the results for each feature selection method, with features ordered from most to least significant. The ranking is determined by calculating the mean of each test's results, while the threshold between retained and discarded features was determined by trial and error during model evaluation.

GHI, humidity, temperature, dew point, uvindex, sunlight duration, and hour are the variables selected for this investigation. It is fascinating to observe how the classification of features by the various feature selection techniques varies significantly. With the exception of the GHI, which is consistently ranked first or second by all feature selection techniques, the other features are evaluated quite differently by the various methodologies. If one selects the top three characteristics for each technique, they do not all appear in Table 5.3. In the sequential backward selection technique, for instance, only the first feature (GHI) was included in the final selection, while the rest were excluded. Similarly, the second and third features for LASSO regression position sixth and seventh overall, narrowly making the cut. In contrast, two to three of the final selected features rank within the top ten across the various feature selection methodologies.

	Features	Methods						Rank
		R^2	MI	SFS	SBS	L	RF	
Selected	GHI	2	1	1	1	1	1	1
	humidity	7	9	3	9	6	13	2
	temperature	5	10	4	8	2	10	3
	dew point	11	19	11	6	3	4	4
	uV index	1	2	9	16	7	23	5
	sunshine duration	13	18	2	19	8	3	6
	hour	22	3	15	11	12	2	7
Discarded	air temperature	9	12	7	3	5	16	8
	air temperature - day	12	20	5	2	4	10	9
	relative humidity	8	8	6	17	15	9	10
	sunrise time	4	15	13	4	17	18	11
	day	21	14	8	10	13	6	12
	wind speed Dark Sky	10	17	18	13	11	5	13
	cloud cover	19	4	10	15	19	12	14
	wind bearing	16	6	22	14	16	8	15
	sunset time	3	13	23	5	22	17	16
	apparent temperature	6	11	21	7	23	15	16
	wind speed	14	16	14	12	10	20	17
	precipitation probability	18	5	12	18	18	21	18
	atmospheric pressure	15	22	16	20	9	14	19
	precipitation intensity	17	7	20	23	14	22	20
	wind direction	20	21	17	21	20	11	21
	minute	23	23	19	22	21	7	23

Table 5.3 Feature selection results between variables and Real PV power generation

5.3.2 Training, testing and exploitation of the Neural Networks

The purpose of this section is to present the methodology adopted during the development of the different ANN models. As presented in Figure 5.2, the main phases required to develop any new predictive model are: i) training (which includes training and validation, ii) testing.

The PV power generation simulator presented in [104], which accurately models the real PV installation located on the roof of our university campus in Turin, Italy, is utilized to generate an artificial, but accurate and realistic dataset of PV power generation large enough to effectively train and test various ANNs. Together with the meteorological features previously selected in Section 5.3.1, this artificial but accurate and realistic dataset is used for the initial training and testing of the two ANNs: 1D-CNN and LSTM. The ANN models trained and evaluated on the simulated dataset are then applied to a portion of the real, but limited, dataset of the real power generated by the real PV installation upon which the simulator is based in order to evaluate their prediction performance against real data.

In order to identify the best architecture, the networks' performance is evaluated through three statistical indicators proposed by [165], being:

- the Mean Absolute Difference (MAD), which measures the absolute difference between the prediction and the observed value;
- the Coefficient of Determination (R^2), which measures the proportion of variance between the observed and the predicted values;
- and the Root Mean Square Difference (RMSD), which measures the standard deviation of the difference between predicted and observed values.

All ANNs in this study are trained with Adaptive Moment Estimation (ADAM) at a learning rate of 0.001. Adaptive moment estimation is advantageous because it eliminates the burdensome process of hyperparameters modifying the learning rate dynamically based on past gradient values. Mean Squared Error (MSE) is the loss function that has to be minimized during the training procedure. In addition, to prevent overfitting and reduce training time, the early halting criterion with a 10-epoch patience is implemented.

1D-CNN best architecture

The 1D-CNN is a subtype of CNN distinguished by the input dimension and the manner in which the filter traverses the data. The CNN's filter (its ability to automatically detect key features) and its relatively low cost make it a highly versatile model that can be applied to a wide range of tasks. However, its implementation is recommended when the dataset available for training is large enough to prevent overfitting.

Also in this instance, trial and error was used to determine the optimal architecture. These hyperparameters were tested:

- number of one-dimensional convolution layers: varied between 1 and 2
- number of units: varied between 10 and 200
- filter size: varied between 50 and 200
- kernel size: varied between 1 and 3
- activation functions: linear and hyperbolic, with and without a flatten layer and dense layer
- epochs: varied between 250 and 500
- batch size: varied between 100 and 400

Another aspect which was investigated was the effect of adding an entirely connected layer with varying units to the end of the network using a hyperbolic tangent. In addition, a pooling strategy (Max Pooling) with a pool size of 2 is employed, resulting in the output dimension being cut in half.

LSTM best architecture

LSTM ANN is a Recurrent Neural Network (RNN), useful for modelling sequential data. LSTM and RNNs also contain backward connections, meaning that at a given time t they receive the current state input x_t plus its own output at the previous time step y_{t-1}.

Also in this case a trial and error approach was used to identify the best architecture. The following hyperparameters were tested:

- number of layers: varied between 1 and 3
- number of units: varied between 10 and 150 (also varying between layers)
- activation functions: linear and hyperbolic (also varying between layers)
- epochs: varied between 100 and 500
- batch size: varied between 100 and 400

5.3.3 Transfer learning

The final section of the methodology presented in Figure 5.2 aims to evaluate the effectiveness of Transfer Learning (TL) in supporting PV power generation forecast. Different TL techniques are used to tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, investigating their effectiveness to improve the prediction performance of PV power generation always against the same real data. As explained in Section 5.2, the data used for TL is now only the real, limited, dataset: years 2018 and 2019 (70.080 data points) are used for training (tuning), and year 2020 (35.136 data points) is used for testing.

Following the methodologies identified in the state of the art, TL can be applied through three main approaches:

- Retrain only the first layer
- Retrain only the second layer
- Retrain all the layers

As discussed in [166], and in [167], the primary effects of the various TL retraining approaches are on precision and computation time. If all of the model's layers are retrained (no layers are locked), the trained model will be more accurate, but it will take longer to compute. Alternatively, if only the first or second layer is retrained (all other layers are frozen), only the unfrozen layer's weights must be backpropagated and updated, resulting in a significant reduction in computation time. Therefore, the various solutions are investigated.

For the purposes of this investigation, only the accuracy of the various TL approaches has been considered, as no appreciable difference in computation time was observed between the fine tuning of ANNs using the various TL techniques. The difference in computation time required to train the original ANNs versus their fine tuning using TL has been observed. The computation time required for the initial training of the ANNs remained on the order of magnitude of a number of months and necessitated a very large database that was only accessible by utilizing the simulator presented by [104], which was used to generate an artificial, but accurate, dataset of PV power generation large enough to effectively train and validate different ANNs (which were then tested on a portion of the real dataset). In contrast, the fine-tuning of the ANNs via TL was completed in a matter of hours using only the actual, limited dataset. This allowed for the initiation of a greater number of TL simulations with different approaches and the collection of a greater number of results regarding the effects of various TL methods in the same amount of time. However, the underlying objective of the study was to determine whether fine-tuning using TL on a much smaller dataset (2 years as opposed to 6 years for the initial training) was capable of producing a model with an acceptable level of accuracy.

5.4 Results

In this section we report and discuss, for each ANN: the testing prediction performance of the most effective architectures in PV generation forecasting after their training with the artificial, but accurate and realistic, dataset of PV power generation created by the PV power generation simulator presented by [104]; the prediction performance of the same ANN models on a portion of real, but limited, dataset of the real power generated by the real PV installation on which the simulator is modeled, to evaluate their prediction performance against real data; and the effectiveness of different transfer learning techniques to tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, to improve the prediction performance of PV power generation always against the same real data.

After training and testing the model with the artificial, but accurate and realistic, dataset of PV power generation created by the PV power generation simulator presented in [104],, the optimal architecture for 1D-CNN consists of two 1-dimensional convolution layers with filter size equal to 170, kernel size equal to 2, a hyperbolic tangent as activation function, followed by a flatten and dense layer. Using a hyperbolic tangent, a completely connected layer with 100 units is inserted at the conclusion of the network. The output layer of the 1D-CNN consisted of sixteen outputs, 500 epochs, and 200 batches. Figures 5.10, 5.11 and 5.12 illustrate the performance of the most effective 1D-CNN architectures in terms of MAD, R^2, and RMSD. One can observe that there is little variation between the various varieties of 1D-CNN architectures, and that all three indicators improve for the first four steps (up to one hour) before their performance declines and levels off after 24 steps (6 hours). The overall efficacy of the 1D-CNN is inferior to that of the LSTM, both in terms of diminished variance and absolute value. The R^2 has comparable values for both ANNs at the 4th step (1 hour), 0.98, 8th step (2 hours), 0.88, and 12th step (3 hours), 0.85.

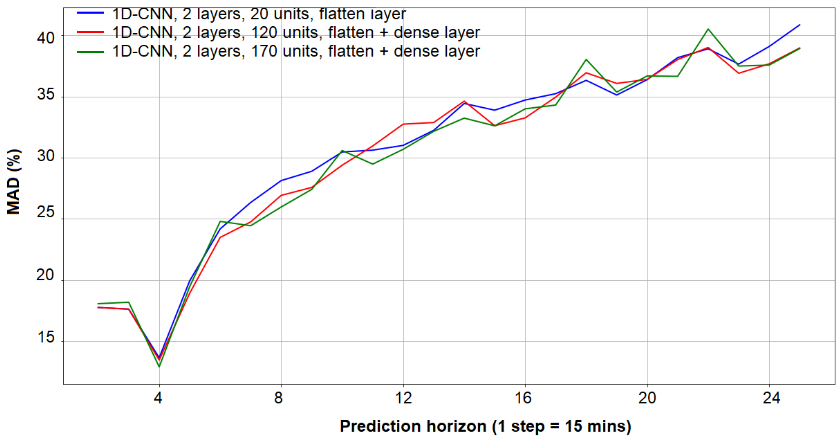


Fig. 5.10 Best 1D-CNN architecture MAD

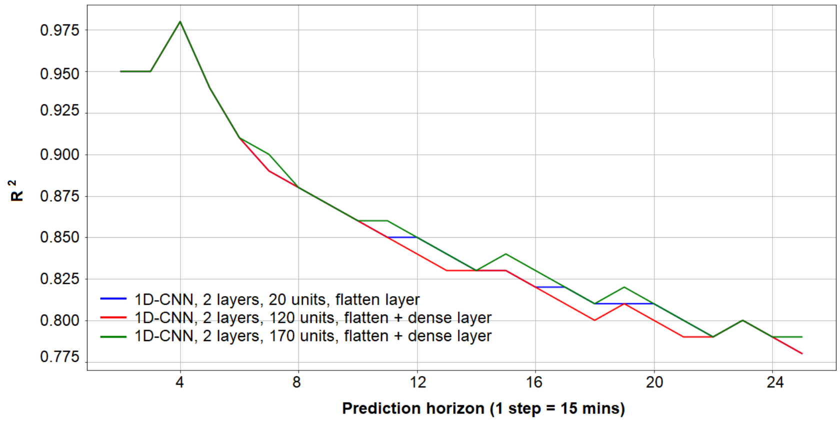


Fig. 5.11 Best 1D-CNN architecture R^2

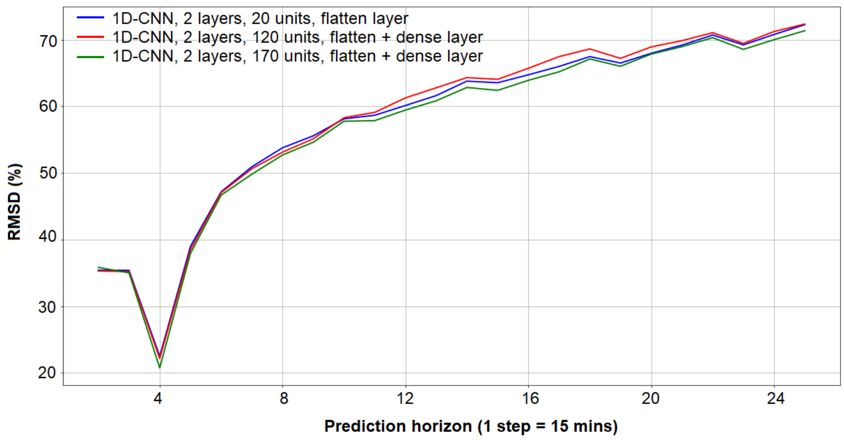


Fig. 5.12 Best 1D-CNN architecture RMSD

For LSTM, after training and testing the model with the artificial, but accurate and realistic, dataset of PV power generation generated by the PV power generation simulator presented in [104], the best identified architecture consists of three recurrent layers, where the first two are composed by 100 units, with a hyperbolic activation function, and the third layer is composed of 24 units with a tanh activation function, and the output layer is composed of 100 units with a Figures 5.13, 5.14 and 5.15 depict the efficacy of the most effective LSTM architectures in terms of MAD, R^2, and RMSD. One can observe that there is little variation between the various types of LSTM architectures, and that all three indicators improve for the first four steps (up to one hour) before their performance declines and appears to level off after 24 steps (six hours). The aggregate LSTM performance is superior to the 1D-CNN performance. At the fourth step (1 hour), the RMSD is approximately between 21 and 23 for 1D-CNN and 20 for LSTM; at the eighth step (2 hours), the RMSD is approximately 53 for 1D-CNN and 54 for LSTM; and at the twelfth step (3 hours), the RMSD is approximately 60 for both ANNs.

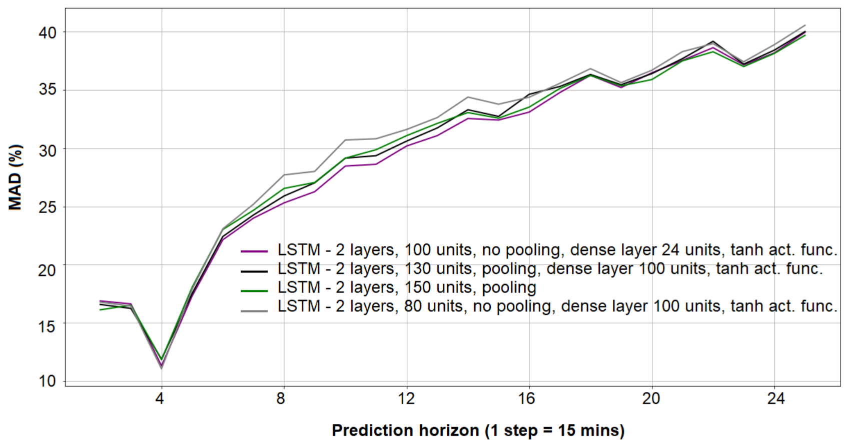


Fig. 5.13 Best LSTM architecture MAD

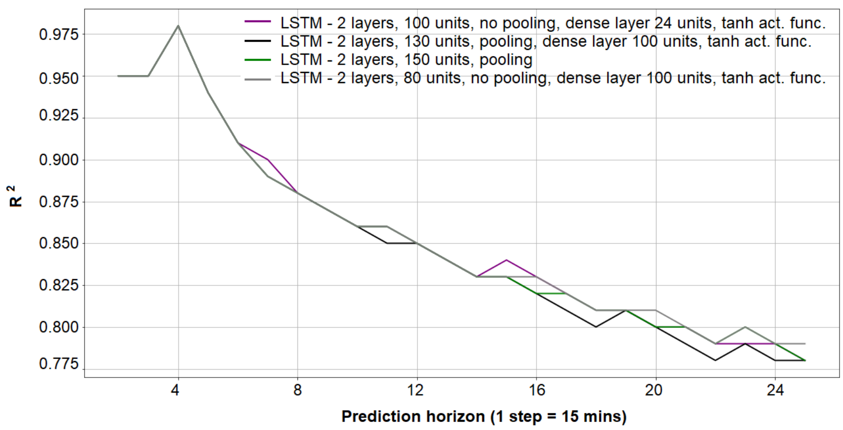


Fig. 5.14 Best LSTM architecture R^2

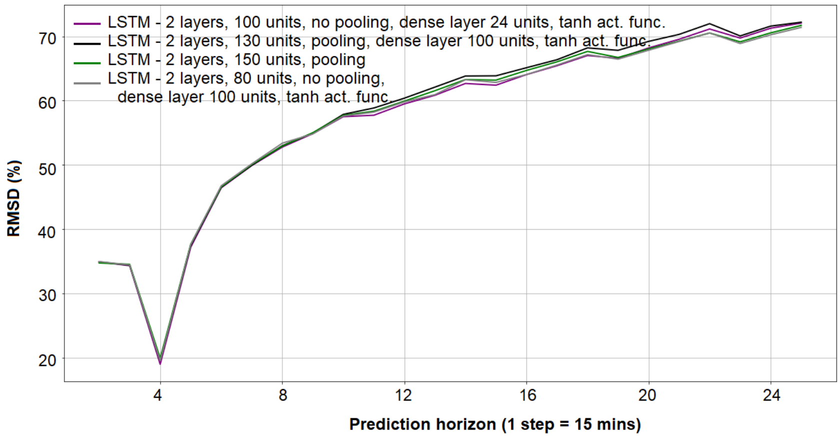


Fig. 5.15 Best LSTM architecture RMSD

Table 5.4 compares the prediction performance of the testing phase for the best of each of these ANNs (in terms of MAD, R^2 and RMSD), for different prediction horizons.

Pred. horizon	MAD		R^2		RMSD	
	1D-CNN	LSTM	1D-CNN	LSTM	1D-CNN	LSTM
15 mins	**19.27**	19.63	0.95	**0.95**	35.70	**35.38**
30 mins	**17.87**	17.95	0.95	**0.95**	35.08	**34.69**
45 mins	12.87	**12.80**	0.98	**0.98**	20.94	**20.42**
60 mins	19.26	**18.86**	0.94	**0.94**	38.03	**37.69**
75 mins	24.96	**23.41**	0.91	**0.91**	46.95	**46.71**
90 mins	24.89	**24.30**	0.89	**0.89**	50.28	**50.01**
105 mins	27.22	**26.10**	0.88	**0.88**	53.73	**52.92**
120 mins	28.33	**27.61**	0.87	**0.87**	55.61	**54.78**

Table 5.4 MAD, R^2 and RMSD comparison for the best ANN architectures

Regarding MAD, one can observe that the relative efficacy of the various ANNs fluctuates depending on the prediction horizon. In the first 30 minutes, 1D-CNN outperforms LSTM by a small margin. From 15 to 45 minutes, the performance of both LSTM and 1D-CNN improves; at 45 minutes, they both reach their peak performance and LSTM begins to marginally outperform 1D-CNN. After 60 minutes, the efficacy of both networks degrades without significant differences. The R^2 results, In contrast, R results demonstrate a much more homogeneous decline in performance for both ANNs, with LSTM consistently outperforming the other by a matter of integers. Again, the efficacy of LSTM and 1D-CNN improves from 15 to 45 minutes, reaches its peak at 45 minutes, and then declines from 60 minutes onward. LSTM performance is consistently marginally preferable to that of 1D-CNN and 1D-CNN networks in terms of RMSD results. Again, the effectiveness of LSTM and 1D-CNN increases from 15 to 45 minutes, peaks at 45 minutes, and then declines after 60 minutes. The results demonstrate that the LSTM model provides the most accurate predictions. The 1D-CNN also provides excellent prediction performance, and its initial MAD values are superior to those of the LSTM model. However, the LSTM model outperforms the 1D-CNN model in the majority of forecasting horizons, for MAD, R^2 and RMSD.

As described in 5.3.2 and in accordance with the methodology presented in Figure 5.2, the ANN models trained and tested on the simulated dataset are then applied to a portion of the real, but limited, dataset of the real power generated by the real PV installation on which the simulator is based, in order to evaluate their prediction performance against real data. On the inference set derived from the actual dataset, the finest ANN model architectures presented previously are utilized. Table 5.5 displays the prediction performance on the true dataset for the 1D-CNN model, whereas Table 5.6 displays the results for the LSTM model. The outcomes plainly demonstrate how the performance of both models declines when applied to a real-world dataset. For instance, the MAD degrades between 40% and 47% for 1D-CNN and between 42% and 46% for LSTM. R^2 degrades by 15% to 20% for 1D-CNN and by 16% to 21% for LSTM, but more crucially, it degrades promptly below 0.8. RMSD deteriorates considerably, between 70% and 83% for 1D-CNN and between 74% and 88% for LSTM.

Pred. horizon	MAD		R^2		RMSD	
	testing	exploitation	testing	exploitation	testing	exploitation
15 mins	19,27	27,36	0,95	0,80	35,70	62,48
30 mins	17,87	25,02	0,95	0,81	35,08	59,64
45 mins	12,87	18,28	0,98	0,78	20,94	37,27
60 mins	19,26	27,16	0,94	0,78	38,03	67,69
75 mins	24,96	35,44	0,91	0,78	46,95	83,57
90 mins	24,89	36,09	0,89	0,77	50,28	90,50
105 mins	27,22	40,01	0,88	0,76	53,73	98,33
120 mins	28,33	41,65	0,87	0,75	55,61	101,77

Table 5.5 1D-CNN MAD, R^2 and RMSD comparison between testing and exploitation prediction performance

Pred. horizon	MAD		R^2		RMSD	
	testing	exploitation	testing	exploitation	testing	exploitation
15 mins	19,63	28,07	0,95	0,79	35,38	62,27
30 mins	17,95	25,49	0,95	0,80	34,69	60,36
45 mins	12,80	18,18	0,98	0,77	20,42	36,76
60 mins	18,86	26,97	0,94	0,76	37,69	68,22
75 mins	23,41	33,48	0,91	0,76	46,71	84,55
90 mins	24,30	34,75	0,89	0,75	50,01	91,02
105 mins	26,10	37,58	0,88	0,74	52,92	96,84
120 mins	27,61	40,31	0,87	0,73	54,78	101,34

Table 5.6 LSTM MAD, R^2 and RMSD comparison between testing and exploitation prediction performance

As explained in Section 5.3.2, and following the methodology presented in Figure 5.2, the ANN models are tuned by leveraging different transfer learning techniques, with the remaining portion of the real, but limited, dataset of PV power generation as input, and their efficacy to improve the prediction performance of PV power generation against the same real data is investigated. The tuning set of the real dataset is used to re-train the ANN models with the transfer learning methodology, whereas the inference set is used for comparing and assessing the prediction performance on real data of transfer learning to that of the exploited models.

As described in Section 5.3.3, various transfer learning techniques are employed, including one in which only the first layer is retrained, one in which only the last layer is retrained, and one in which all layers are retrained. The following figures illustrate, for each ANN, the prediction performance of the original models presented in the exploitation phase, without transfer learning, and the prediction performance of the three distinct transfer learning techniques. For the 1D-CNN, Figure 5.16 presents the MAD, Figure 5.17 presents the R^2 and Figure 5.18 presents the RMSD. The same results for LSTM are then presented by the following figures: Figure 5.19 for MAD, Figure 5.20 for R^2 and Figure 5.21 for RMSD.

The performance of the 1D-CNN model presented in the test, exploitation and best transfer learning phases, are compared in Table 5.7. Similarly, for LSTM the achieved results are presented by Table 5.8. The difference in variation of performance of the 1D-CNN and LSTM models from exploitation to transfer learning are summarized in Table 5.9. A positive variation indicated an improvement in the application of transfer learning, while negative variation indicates a worse performance of transfer learning.

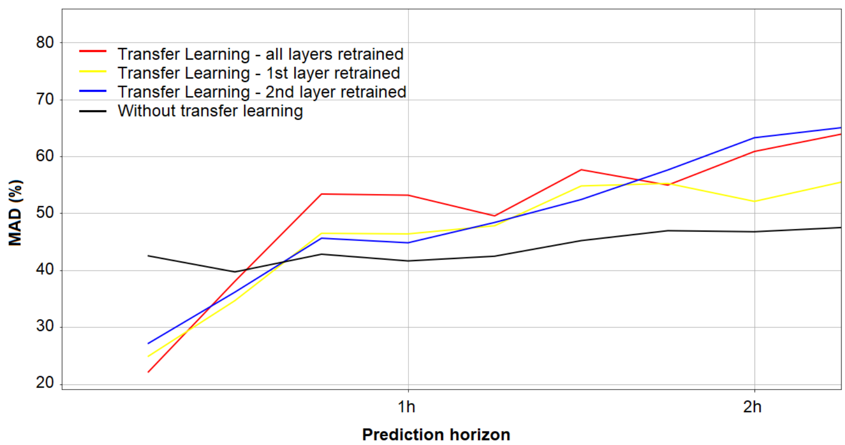


Fig. 5.16 1D-CNN MAD comparison with transfer learning

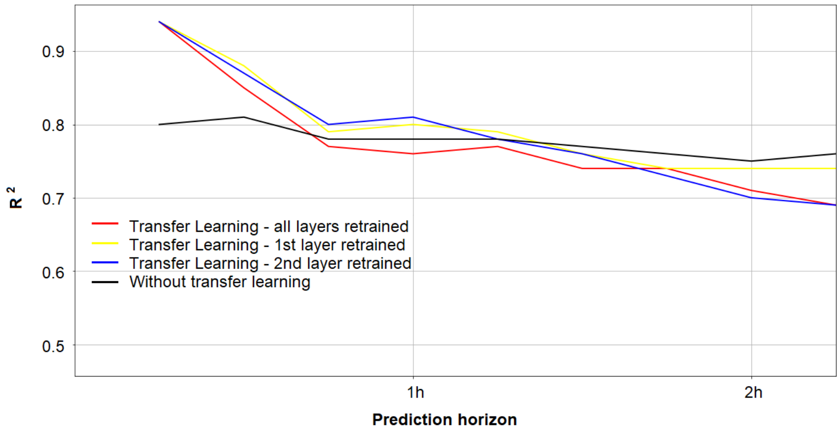


Fig. 5.17 1D-CNN R^2 comparison with transfer learning

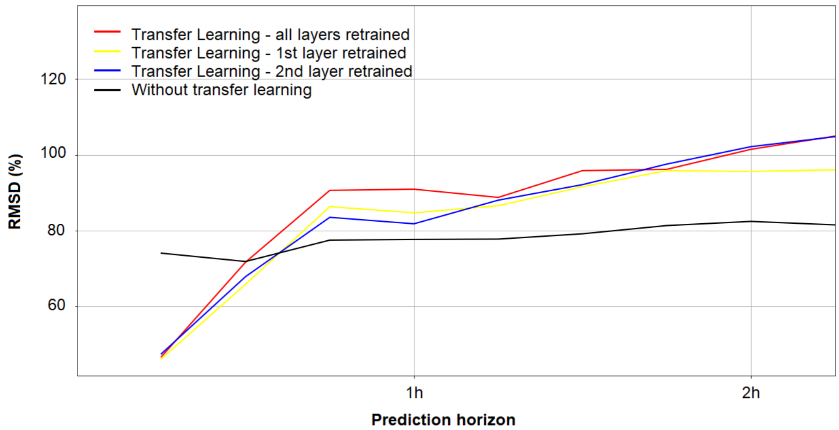


Fig. 5.18 1D-CNN RMSD comparison with transfer learning

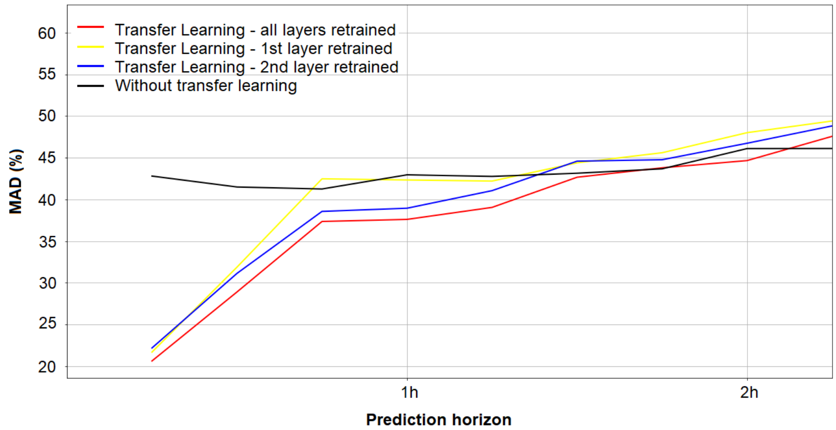


Fig. 5.19 LSTM MAD comparison with transfer learning

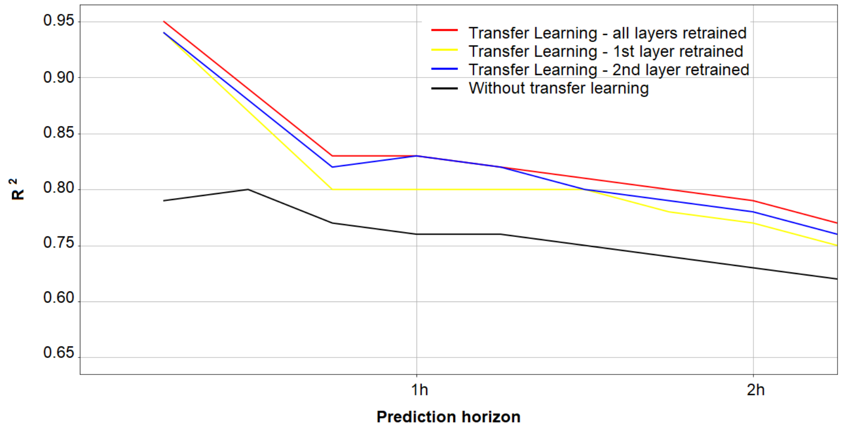


Fig. 5.20 LSTM R^2 comparison with transfer learning

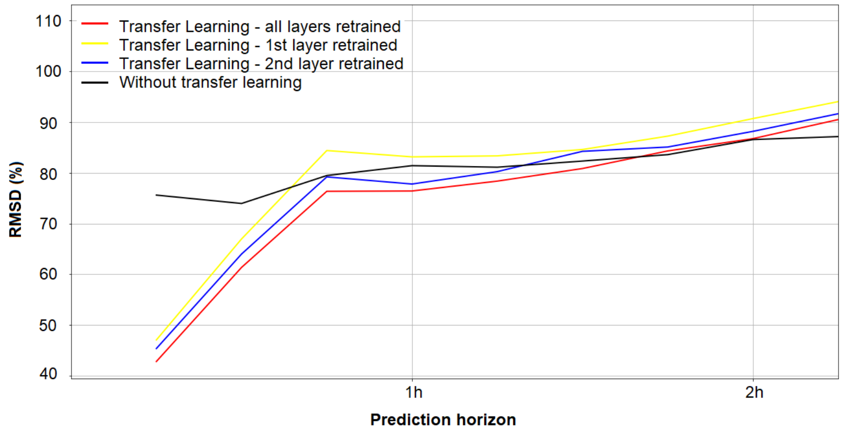


Fig. 5.21 LSTM RMSD comparison with transfer learning

Pred. horizon	MAD			R^2			RMSD		
	testing	exploitation	TL	testing	exploitation	TL	testing	exploitation	TL
15 mins	19,27	27,36	24,09	0,95	0,80	0,98	35,70	62,48	51,77
30 mins	17,87	25,02	24,12	0,95	0,81	0,88	35,08	59,64	57,88
45 mins	12,87	18,28	18,79	0,98	0,78	0,81	20,94	37,27	38,74
60 mins	19,26	27,16	28,12	0,94	0,78	0,82	38,03	67,69	69,98
75 mins	24,96	35,44	37,19	0,91	0,78	0,78	46,95	83,57	86,86
90 mins	24,89	36,09	38,58	0,89	0,77	0,76	50,28	90,50	95,53
105 mins	27,22	40,01	42,19	0,88	0,76	0,72	53,73	98,33	104,77
120 mins	28,33	41,65	43,06	0,87	0,75	0,68	55,61	101,77	108,44

Table 5.7 1D-CNN MAD, R^2 and RMSD comparison between testing, exploitation and best TL prediction performance

Pred. horizon	MAD			R^2			RMSD		
	testing	exploitation	TL	testing	exploitation	TL	testing	exploitation	TL
15 mins	19,63	28,07	23,75	0,95	0,79	0,95	35,38	62,27	50,95
30 mins	17,95	25,49	22,98	0,95	0,80	0,91	34,69	60,36	56,20
45 mins	12,80	18,18	17,54	0,98	0,77	0,85	20,42	36,76	35,94
60 mins	18,86	26,97	25,84	0,94	0,76	0,85	37,69	68,22	66,33
75 mins	23,41	33,48	32,54	0,91	0,76	0,84	46,71	84,55	83,14
90 mins	24,30	34,75	34,51	0,89	0,75	0,83	50,01	91,02	90,52
105 mins	26,10	37,58	37,58	0,88	0,74	0,82	52,92	96,84	97,37
120 mins	27,61	40,31	40,03	0,87	0,73	0,81	54,78	101,34	101,34

Table 5.8 LSTM MAD, R^2 and RMSD comparison between testing, exploitation and best TL prediction performance

Pred. horizon	1D-CNN variation			LSTM variation		
	MAD[%]	R^2[%]	RMSD[%]	MAD[%]	R^2[%]	RMSD[%]
15 mins	12.0	21.9	17.1	15.4	20.6	18.2
30 mins	3.6	9.1	2.9	9.9	14.1	6.9
45 mins	-2.8	3.3	-3.9	3.5	10.1	2.2
60 mins	-3.5	4.9	-3.4	4.2	12.1	2.8
75 mins	-4.9	0.0	-3.9	2.8	10.4	1.7
90 mins	-6.9	-1.7	-5.6	0.7	10.7	0.5
105 mins	-5.4	-5.2	-6.6	0.0	11.0	-0.5
120 mins	-3.4	-8.9	-6.6	0.7	11.3	0.0

Table 5.9 Variation in performance variation from exploitation to TL for 1D-CNN and LSTM

All transfer learning approaches enhance the prediction performance of 1D-CNN and LSTM, as illustrated in the following figures. However, for 1D-CNN, the greatest improvement is obtained by retraining only the second layer of the second transfer learning model. For LSTM, the model with the best performance is the third transfer learning model in which all layers are retrained.

When applying transfer learning to the 1D-CNN, at the first 15 minutes forecast the MADimproves by 12.0%, R^2 by 21.9% and RMSD by 17.1%. At the 30 minutes forecast, the MADimprovement decreases to 3.6% , R^2 to 9.1% and RMSD to 2.9%. After 30 minutes however, the MADand RMSD show a worsening in performance, while the R^2 continues to remain better until 75 minutes. After 90 minutes, all three indicators underperform with transfer learning.

For the LSTM on the other hand, transfer learning appears to be even more effective. For the 15 minutes forecast the MADimproves by 15.4% , R^2 by 20.6% and RMSD by 18.2%. For the 30 minutes forecast the MADremains at 9.9% , R^2 at 14.1% and RMSD at 6.9%. Performance remains better until 90 minutes. After 90 minutes, the MADand RMSD appear to converge, with differences in performance between exploitation and transfer learning levelling out around 0. R^2 with transfer learning, on the other hand, continues to overperform up to 2 hours.

The aforementioned results demonstrate that transfer learning has a limited effect on 1D-CNN but a consistent effect on LSTM, which even after transfer learning remains the top performing ANN.

5.5 Conclusion

This study aimed to present a novel method for forecasting PV power generation using ANNs when only a small amount of actual data is available. The novelty of this study lies in the use of a PV power generation simulator that accurately models a real PV installation to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test different ANN models, which are then applied to a portion of the real, but limited, dataset of the real power generated by the real PV installation on which the simulator is based. The application of various transfer learning techniques to tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, evaluating their efficacy to improve the prediction performance of PV power generation always against the same real data, contributes to the originality of the study.

Various meteorological data features are analyzed using feature selection methodologies in order to identify those that have the greatest impact on the accuracy of data prediction forecasts. GHI, humidity, temperature, dew point, UV index, solar duration, and time of day are the factors that have the greatest impact on power prediction. The PV power generation simulator presented by [104], which is accurately modeled to replicate real PV installations, is then used to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test different ANNs. This artificial, but accurate and realistic dataset, together with the meteorological features previously selected, is used for the initial training and testing of the two different ANNs: 1D-CNN, and LSTM.

The ANN models trained and evaluated on the simulated dataset are then applied to a portion of the real, but limited, dataset of the real power generated by the real PV installation upon which the simulator is based in order to evaluate their prediction performance against real data. Different transfer learning techniques are then used to fine-tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, and their ability to improve the prediction performance of PV power generation against the same real data is investigated. On the rooftop of a structure on our university campus in Turin, Italy, an actual PV installation was utilized to test and validate the methodology.

The optimal LSTM architecture consists of three recurrent layers, the first two of which are composed of 100 units with hyperbolic activation functions, the third of 24 units with tanh activation functions, and the output layer of 100 units with linear activation functions. In addition, transfer learning applied to this LSTM architecture improves performance, particularly when the third transfer learning model is implemented and all layers are retrained. The enhancement is substantial in the short-term (up to 30 minutes) and modest in the long-term (up to 120 minutes).

The optimal architecture for 1D-CNN consists of two 1-dimensional convolution layers with filter sizes of 170 and 2, a hyperbolic tangent activation function, followed by a flatten and dense layer. Using a hyperbolic tangent, a completely connected layer with 100 units is inserted at the conclusion of the network. The output layer was comprised of sixteen outputs, 500 epochs, and 200 batches. When only the second layer of this 1D-CNN architecture is retrained using the second transfer learning model, performance can be improved using transfer learning. This enhancement is however minimal and only temporary (up to 30 minutes).

The outcomes demonstrate that 1D-CNN and LSTM ANNs can effectively predict PV power generation even when trained on simulated data. In addition, the implementation of transfer learning techniques has improved the efficacy of PV power forecasting in both the short-term (15-30 minutes for 1D-CNN) and the medium-term (up to 2 hours for LSTM).

Chapter 6

Conclusions

The main focus of the research of this Doctoral Program is on applications in the spheres of Industrial Internet of Things (IIoT) and Big Data & Analytics. In order to strengthen the link between academic research and industrial applicability, the research focuses on industrial areas which are considered critical for their impact on manufacturing performance (cost, quality, delivery), and aspects such as readiness, cost, robustness, reliability and flexibility must be taken into consideration during the investigation. In order to understand what areas of manufacturing can most benefit from the application of Industry 4.0 (I4.0) solutions, it is important to proceed with prioritization. If the costs and, especially, losses of the company are stratified, and if the affinity with the cluster “Big Data & Analytics” of the "Piano Nazionale Industria 4.0" [2] is also considered, Energy results being the main priority. According to [7], energy is a major cost issue in Europe (the Commission estimates that wholesale power prices are around 30% higher than in the US, and gas prices are more than 100% higher), and one of the major issues that industrial organizations have been facing is rising energy expenses, particularly those with energy-intensive activities.

To alleviate the impact of these energy difficulties, businesses can engage in sophisticated technology, process optimization, and energy management systems, which can result in significant energy savings and cost savings. Smart energy management, for example, provides enterprises with numerous chances to optimize their energy consumption and cut costs. Energy monitoring and analytics are two examples of smart energy management opportunities for industries. Implementing real-time energy monitoring systems and advanced analytics can help industries identify energy usage patterns, detect inefficiencies, and make informed decisions for energy optimization, as demonstrated by [30]. Innovative applications can be implemented within the smart grid framework to better coordinate power demand and supply [38].

When integrating this need, forecasting of both power demand and supply, with the I4.0 cluster of Big Data & Analytics mentioned in the previous Sections, one can clearly understand how technology and innovation can support to overcome these challenges. When looking at innovative solutions for forecasting, with the support of Big Data & Analytics, one enters the Machine Learning (ML) realm. As presented by [39], ML methods are gaining popularity in the forecasting field. Of the different ML methods, [39] highlights how particular attention must be given to those based on Artificial Neural Network (ANN)s, which presented substantial improvements over benchmarks in the modelling of forecast uncertainty. The research therefore focuses on developing innovative solutions for the effective forecasting of both power demand and supply, leveraging ML methods and particularly those based on ANNs.

An important part of the investigation was to select the best case-studies where to apply the investigation, in order to strengthen the link between academic research and industrial applicability. As previously mentioned, the investigation was originally supposed to be carried out industrial case-studies, however due to the Covid-19 emergency access to such sites was restricted and the research activities were significantly slowed down. An alternative was then found by using other case studies for which some data was already available, and which could easily be replicated in the industrial reality.

In terms of forecasting power demand, the natural choice would be to focus on those manufacturing processes which require the most energy to operate. However, since due to the Covid-19 emergency access to industrial sites, and therefore their processes, was restricted, new viable solutions had to be found. After an initial study, the choice fell on Heating, Ventilation and Air Conditioning (HVAC) systems and in-door air temperature forecasting, and an interesting opportunity arose to carry out the investigation on a real-world building located in Turin (Italy). The first part of the Doctoral investigation focused on developing an innovative solution for the effective forecasting of building in-door air temperature and, as a consequence, of power demand, leveraging ML methods and particularly those based on ANNs.

In terms of forecasting power supply, the challenge was again to identify those areas which could bring the most benefit to industrial realities. With investments in clean electricity and electrification, particularly solar photovoltaic (PV), soaring, the significance of renewable energy sources such as PV energy is clearly growing. An interesting opportunity arose to carry out the investigation on a real-world PV installation located in Turin (Italy). The second part of the Doctoral investigation focused on developing an innovative solution for the effective forecasting of PV power generation, leveraging ML methods and particularly those based on ANNs.

The techniques belonging to the fields of Artificial Intelligence (AI) and ML, more specifically to the domain of ANNs, present one of their greatest challenges in that they require substantial amount of data to make sure they are effectively trained and produce acceptable accuracy. Wang et al. [98] recommend selecting a data length of at least three years. However, availability of large enough datasets of real data, complete, accurate and reliable is still hard to achieve. As a consequence, the research focused on developing innovative solution for effective forecasting even when faced with limited real data availability, by leveraging accurate simulators and Transfer Learning (TL). Also, it focused on investigating the applicability of ANN models which, to the best of our knowledge, had previously seen little application in the chosen forecast domains.

Part of the novelty of these investigations therefore lies in the common methodology used for forecasting through ANNs, presented in Figure 3.1. In all applications, a small, real, but limited dataset of real data is available. A simulator, accurately modeled to replicate the real environment, was found and leveraged for both applications. These simulators, which accurately model the case-study environments, are used to create an artificial, but accurate and realistic, dataset large enough to effectively train and test different ANN models. It is worth noting that real and simulated environments of both case studies are coincident. After the different ANN models are trained and tested on the artificial, but accurate and realistic, dataset, they are then exploited on a portion of the real, but limited, dataset of real data on which the simulator is modeled, to evaluate their prediction performance against real data. Different TL techniques are then employed to fine-tune the ANN models with the remaining portion of the real data set, evaluating their efficacy in enhancing the prediction performance of PV power generation against the same real data set. As stated previously, the entire methodology has been verified and validated using different real-world case studies.

During the investigation focusing on developing an effective forecasting of building in-door air temperature, an innovative methodology to support the energy management of Heating, Ventilation and Air Conditioning systems, through Smart Building indoor air-temperature forecast, is proposed. The study is conducted on a public school building in Turin, Italy. The methodology explores the applicability of state-of-the-art ANNs, more specifically 1-Dimensional Convolutional Neural Network (1D-CNN) and Long Short-Term Memory (LSTM) ANNs, for time-series predictions. These ANNs are first trained on a large, artificial, but realistic dataset based on Building Information Modelling simulations with real meteorological data. The inference phase is then carried out on a second dataset collected by Internet-of-Things devices previously installed in the corresponding real-world building, to compensate for the lack of real data. They then undergo an optimization process for the tuning of all their hyperparameters. Finally, TL techniques are exploited to improve the performances of the ANNs' predictions and their ability to generalize. The experimental results are further validated by applying Fanger's model of indoor thermal comfort and show consistent levels of accuracy and comfort even in the face of limited data availability.

This study demonstrates that even when simulated datasets are used to train them, 1D-CNN and LSTM can accurately forecast interior air-temperature, and that their performance can be further improved by applying TL to a subset of real datasets. In particular, 79% of the time, the MAE observes an increase in the prediction horizon while remaining within the acceptable range, with an average increase of 13.4 hours. In contrast, the PMV forecast horizon expands by an average of 8.6 hours while remaining within the acceptable range in 27% of cases.

After implementing one of the TL strategies, ten of the original sixteen ANNs exhibit an increase in their forecast horizon for both MAE and PMV. However, these enhancements are not homogeneous; there was no specific TL techniques capable of consistently improving the performance of all ANNs. In addition, there appears to be no correlation between TL approaches, the environment/case or network of the original ANN, and the improvement in performance.

Future works for this investigation, which focused on developing an effective forecasting of building in-door air temperature, can include to investigate potential applications of the same methodology in comparable environments. Based on its structural characteristics, this investigation was conducted on a building that can be either public or private. Future research could investigate the applicability of this method to industrial structures, whose structural characteristics and environmental factors vary significantly. Future research could examine its applicability within the context of Smart Buildings. Thus, Demand/Response control strategies can be evaluated within the context of a Smart Building. According to Yu in [148], digital age efficiency requirements are one of the many challenges power grids have faced in recent years, and an increasing number of AI and big data-driven solutions are emerging to advance smart-grid development. Thus, the proposed solution can make a significant contribution to the ongoing endeavor to reduce energy consumption, which is a topic of great interest in contemporary society.

During the investigation aimed at developing an innovative solution for the accurate forecasting of PV power generation, a novel method for forecasting PV power generation with ANNs is presented when only a limited quantity of actual data is available. Initially, feature selection is used to investigate various meteorological features and their potential impact on enhancing the prediction of PV power generation. Then, a simulator that accurately replicates an actual PV installation is used to construct an artificial, but accurate and realistic dataset of PV power generation that is large enough to effectively train and test various ANNs. The ANN models trained and evaluated on the simulated dataset are then applied to a portion of the real, but limited, dataset of the real power generated by the real PV installation upon which the simulator is based in order to evaluate their prediction performance against real data. Different TL techniques are then used to fine-tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, and their ability to improve the prediction performance of PV power generation against the same real data is investigated. The methodology has been tested and validated on a real-life PV installation located on the rooftop of a building of a university campus in Turin, Italy.

Different feature selection methodologies are leveraged to analyze various meteorological data features, to determine which have the greatest impact on the precision of data prediction forecasts. GHI, humidity, temperature, dew point, UV index, solar duration, and time of day have the greatest influence on the prediction of power. The PV power generation simulator presented in citeRefwork:21, which accurately simulates actual PV installations, is then used to generate an artificial, but accurate and realistic dataset of PV power generation large enough to train and test various ANNs. Along with the previously selected meteorological features, the initial training and testing of two distinct ANNs is completed using this artificial but accurate and realistic dataset: 1D-CNN and LSTM.

The ANN models trained and evaluated on the simulated dataset are then applied to a portion of the real, but limited, dataset of the real power generated by the real PV installation upon which the simulator is based in order to evaluate their prediction performance against real data. Different TL techniques are then used to fine-tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, and their ability to improve the prediction performance of PV power generation against the same real data is investigated. On the rooftop of a structure on our university campus in Turin, Italy, an actual PV installation was utilized to test and validate the methodology.

The most performing architecture for the LSTM ANN includes three recurrent layers, with the first two containing 100 units with hyperbolic activation functions, the third containing 24 units with tanh activation functions, and the output layer containing 100 units with linear activation functions. TL also enhances the efficacy of this LSTM architecture, especially the third TL model which includes retraining of all layers. Short-term (up to 30 minutes) and long-term (up to 120 minutes) improvements are substantial and moderate, respectively.

The optimal architecture for 1D-CNN consists of two 1-dimensional convolution layers with filter sizes of 170 and 2, a hyperbolic tangent activation function, followed by a flatten and dense layer. Using a hyperbolic tangent, a completely connected layer with 100 units is inserted at the conclusion of the network. The output layer was comprised of sixteen outputs, 500 epochs, and 200 batches. When only the second layer of this 1D-CNN architecture is retrained using the second TL model, performance can be improved using TL. This enhancement is however minimal and only temporary (up to 30 minutes).

The outcomes demonstrate that 1D-CNN and LSTM ANNs can effectively predict PV power generation even when trained on simulated data. In addition, the implementation of TL techniques has improved the efficacy of PV power forecasting in both the short-term (15-30 minutes for 1D-CNN) and the medium-term (up to 2 hours for LSTM).

Future works for this investigation, which aimed at developing an innovative solution for the accurate forecasting of PV power generation, can include investigating the applicability of the developed methodology with other ANNs, such as Transformers and Hybrid Models. The suggestion to pursue further research in this field with the Transformer ANN, an ANN from the Recurrent Neural Network (RNN) family, come from understand the characteristics and limitations of the other RNNs. As explained by [168], first of all, RNNs are slow to train: they take input sequentially one by one, which doesn't use up GPUs very well, which are designed for parallel computation. Furthermore, they are not so capable of remembering old connections from long sequences, since long sequences lead to vanishing gradient or the problem of long-term dependencies. The vanishing gradient problem is partially, but not completely, solved with the development of LSTM ANN: LSTM is a special kind of RNN, specially made for solving vanishing gradient problems. They are capable of learning long-term dependencies, since their default behaviour is remembering information for long periods of time: their neuron structure allows the network to retain memory for a longer period of time thus improving the vanishing gradient problem, but not terribly well. Furthermore, like the simple RNN, LSTM ANN are also very slow to train, perhaps even slower. So the RNN family has historically dealt with two main issues: the vanishing gradient problem (difficulty to learn long-term dependencies, or to remember information for long periods of time, partially solved with LSTM), and slow training.

As explained by [168], the Transformer ANN is a novel RNN architecture that aims to solve sequence-to-sequence tasks while handling long-range dependencies with ease. It was first proposed in the paper [169], and can be considered the current state-of-the-art technique in the field of Natural Language Processing (NLP), although it does perform well also in a variety of other applications including time series forecasting, as presented by [170]. As presented by [169], the Transformer model relies on the parallel multi-head attention mechanism and requires less training time than previous RNN architectures, such as LSTM, since, as explained by [168], in a Transformer model the input sequence can be passed parallelly so that GPU can be used effectively and the speed of training can be increased. The multi-headed attention layer also allows the Transformer model to easily overcome the vanishing gradient problem, which is an issue with traditional RNN architectures (partially, but not completely, solved by LSTM).

However, although the Transformer ANN presents many interesting characteristics which makes it a suitable subject for this investigation, it does present a major drawback. The traditional form of Transformer ANN is not capable of receiving as input a multivariate dataset, such as the one created and described in Section

5.3.1, where the meteorological features are integrated with the PV power generation data. This drawback of the Transformer ANN can however be solved with a hybrid model ANN. For example, a hybrid model based on architectural integration, more specifically an attention-based LSTM, similar to those presented by [171] and [172]. This model architecture allows to effectively combine two different powerful ANN models: the RNN, or more specifically the LSTM, and the Attention model, which is a key feature of the Transformer ANN. This allows the hybrid model to leverage many of the strengths of the Transformer ANN previously described, together with the power of the multivariate dataset input.

As explained in Section 3.2, the techniques belonging to the fields of AI and ML, more specifically to the domain of ANNs, present one of their greatest challenges in requiring a substantial amount of data to make sure they are effectively trained and produce acceptable accuracy. [98], for example, recommend selecting a data length of at least three years. However, availability of large enough datasets of real data, complete, accurate and reliable is still hard to achieve, and this difficulty has increased even more during the Covid-19 emergency, since access to industrial sites was restricted. Consequently, the research focused on developing an innovative solution for effective forecasting using ANN models even when faced with limited real data availability, by leveraging accurate simulators and TL.

As presented in Figure 3.1, in all applications a small, real, but limited dataset of real data is available. For each application a simulator, accurately modelled to replicate the real environment, was found and leveraged to generate more data. These simulators, which accurately model the case-study environments, are used to create an artificial, but accurate and realistic, dataset large enough to effectively train and test different ANN models. It is worth noting that, for both case studies, the real and simulated environments are accurately coincident, since the identified simulators have been developed to accurately replicate precisely the two environments which were chosen for the two different applications. The accuracy of such simulators are presented by [102] and [103], and by [104], as described in Sections 4.2.2 and 5.2. Therefore, although using simulated data to train the ANN models clearly introduces some form of bias and limitation in the accuracy of the models' predictions, the use of such accurate simulators, which have been developed to accurately replicate precisely the two environments which were chosen for the two different applications, and whose accuracy is presented by [102] and [103], and by [104], significantly reduces this limitation, allowing to create an artificial, but accurate and realistic, dataset large enough to accurately train the different ANN models.

Furthermore, in order to evaluate the prediction performance of these models with the highest possible level of accuracy, while they are trained and tested on the artificial, but accurate and realistic, dataset, they are then exploited only a portion of the real, but limited, dataset of real data. Similarly, also the different TL techniques which are used to tune the ANN models are applied only on the remaining portion of the real dataset. By using only the real dataset to exploit the ANN models and to apply the TL techniques, any bias and limitations coming from using a simulated, artificial dataset to train the model are intercepted when the model's prediction accuracy is evaluated.

Regardless of the potential bias and limitations coming from the use of simulators, the topic of potential bias in forecasting outcomes, together with other wider ethical implications such as data privacy and security, is particularly critical considering the potential applicability of the topics of this Doctoral research. As discussed in Section 1.4.2, one of the potential applications of the topics of this Doctoral research is their integration in smart grids. The incorporation of similar innovations in power grids, turning them into smart grids, can improve the administration and stability of power grids by allowing them to better coordinate power demand and supply, with power demand and consumption being adjusted to align with the available power demand, and vice versa, as presented by [38]. However, if one considers some of the design principles of I4.0 discussed in Section 1.2.2, such as decentralization and real-time capability, the consequences of inaccurate forecasting outcomes, both in terms of power demand and supply, can be mildly disturbing at best, catastrophic at worst. By underestimating the power demand or overestimating the power supply, for example, the power grid could be brought to produce less energy, shutting down key power production facilities which require long start-up times, and then crash due to excessive power demand compared to supply. On the other hand, by overestimating power demand or underestimating power supply, the power grid could be brought to generating more power than needed, depleting precious resources and increasing costs for the community. Similar consequences could happen at the industrial site level. These issues can be the result of inaccurate forecasting outcomes, for example due to bias in the developed model, but can also be a consequence of false data being fed to the models. The topic of data security remains therefore extremely important, as presented in Section 1.2.3, where cybersecurity is presented as one of the key enabling technologies for I4.0, as presented by [1] and [2].

Finally, as far as data privacy is concerned, the type of applications on which this Doctoral investigation focused (indoor air temperature prediction and PV power generation), and the type of data involved, do not raise particular concerns in terms of data privacy, especially considering potential applications in an industrial reality. First of all, the type of data to be collected for the described methodology does not fall under the "sensible data" category: it does not include information which can be used to identify any sensible information of the organization or its people, and it does not provide any particular insight on any core industrial processes which could be the subject of industrial espionage. Furthermore, the data on PV power prediction can be easily estimated once the characteristics of the PV installation have been determined, which can be easily achieved since it is an external installation. As far as the data on indoor air temperature is concerned, if used in residential applications it could provide indirect information on the habits and presence of the inhabitants of the residential structure and therefore be used for unethical purposes, but in the case of an industrial application they hardly pose any threat. The only possible threat could come, maybe, from the indoor air temperature data collected from a business environment where the indoor air temperature does not depend on the industrial process, but on the presence of employees (such as an office building). In this case, like for residential buildings, the data could provide indirect information on the habits and presence of the employees and therefore be used for unethical purposes. However, the actual risk associated with this possibility can be considered very low.

The results of this Doctoral investigation also succeed in highlighting the significance of the research in addressing real-world challenges related to energy management in the industrial world, and the effective application of I4.0 solutions. Such results aim to effectively demonstrate the link between academic research and industrial applicability, showcasing potential areas of improvement for manufacturing organizations in energy management (and potentially other applications too) through the application of IIoT and Big Data & Analytics. This Doctoral investigation therefore also contributed to bridging the gap between theory and practice.

As highlighted in Section 1.2 and [7], the successful development and implementation of innovations in the industrial reality requires a tighter collaboration between academia and business, to direct academic research towards industrial needs and speed up its industrial application. As highlighted in Section 1.2.1, when it comes to I4.0, the academic world thrives with research highlighting how I4.0 promises to transform manufacturing and produce a paradigm change in industrial processes, product creation, and customer experiences through the confluence of physical and digital technologies. However, as emphasized by [8], because the I4.0 has only just been formed and its environment is still evolving, providing an univocal definition for "Industry 4.0" is problematic. For example, as discussed in Section 1.2.3, literature presents a range of technologies, varying from four [15] to thirteen [8], which can be identified as the drivers of the ongoing I4.0 transformation, as they merge to form an innovative ecosystem that serves as a powerful mean to increase the productivity of manufacturing systems and create a new production paradigm. The absence of a single, stand-alone enabling component is a continuous contrast between the Fourth Industrial Revolution and its preceding ones. On the other hand, companies need clear guidelines when executing innovation roadmaps, so this ambiguity in I4.0 content and, especially, implementation guidelines represent a barrier for rapid and successful implementation in the industrial world. As presented in Section 1.2.3, after recognizing and understanding the benefits of transforming their country's industrial footprint through I4.0 innovations, governments have worked to provide both incentives and, especially, some guiding framework to support their country's companies to navigate the I4.0 revolution. In Italy for example, as presented by [2], the Piano Nazionale Industria 4.0 outlines the essential enabling technologies for the country's I4.0 transformation. This government intervention in providing not only policies and incentives, but also an operating framework, clearly highlights the difficulty of academic research to permeate into industrial applications quickly and efficiently.

As presented in Section 1.3, in order to strengthen the link between academic research and industrial applicability, this Doctoral investigation focuses on industrial topics which are considered critical for their impact on manufacturing performance (cost, quality, delivery), taking into consideration also aspects such as readiness, cost, robustness, reliability and flexibility during the research. Energy management was chosen after a data-driven prioritization process which highlighted it as the most critical topic for manufacturing performance, amongst those which could be impacted by the I4.0 features IIoT and AI, presented in 1.2.1, design principles of Interoperability, Real-time Capability, and Modularity, presented in 1.2.2, and key enabling technologies Simulation Technologies, IIoT and Big Data & Analytics, presented in 1.2.3. The topics of this Doctoral investigation therefore aimed to leverage I4.0 solutions for Smart Energy applications in the manufacturing environment, with Section 1.4.1 further highlighting why energy represents such an important competitiveness factor in industrial manufacturing companies, particularly those with energy-intensive activities. Section 1.2.4 therefore presents the Smart Factory concept, describing I4.0 applications in the industrial reality, while Section 1.4.2 introduces the topic of smart energy systems for smart energy management, for I4.0 solutions applied to the energy sphere.

This Doctoral investigation has worked to provide a clearer guideline for industrial companies on how to understand the fundamental characteristics of I4.0 innovation, and on how to leverage them in a specific area, energy management, in order to increase their competitiveness. The purpose is therefore not only to provide innovative research in the smart energy management field, but also to bridge the gap between theory and practice.

Energy monitoring and analytics are two examples of smart energy management opportunities for industries to better manage power demand and supply: [35] and [36] discuss the benefits of efficiently managing electricity usage in smart industries and smart cities through the installation or integration of efficient smart grids, in which innovative methods and tools such as data analytics and ML techniques are integrated with sensors and remote controls, while [31] and [33] present the opportunities coming from renewable energy integration, the on-site installation of renewable energy generation systems, such as solar panels, to offset their reliance on the grid, lower energy costs, and reduce environmental impact.

For such applications to function effectively, however, it has become increasingly essential to forecast power demand and supply with varying time horizons, especially considering the variable nature of renewable energy sources, which poses a challenge to the resilience of power grids. When looking at innovative solutions for forecasting, with the support of Big Data & Analytics, one enters the ML realm. As presented by [39], ML methods are gaining popularity in the forecasting field and, of the different ML methods, those based on ANNs present substantial improvements in forecasting modelling compared to benchmarks. The Doctoral investigation therefore focused on developing innovative solutions for the effective forecasting of both power demand and supply, leveraging ML methods and particularly those based on ANNs. Furthermore, it has worked to provide a clearer guideline for industrial companies on how to understand the fundamental characteristics of I4.0 innovation, and on how to leverage them in a specific area, energy management, in order to increase their competitiveness. The purpose is therefore not only to provide innovative research in the smart energy management field, but also to bridge the gap between theory and practice.

Finally, as discussed in Section 1.3, it is important that when proposing innovation in the I4.0 realm these are set into a clear industrial transformation strategy. Section 1.2.5 aims to give a precise definition of the terms "Digitization," "Digitalization," and "Digital Transformation" and, as presented by [1], places them as sequential steps that can be followed to varying degrees to innovate traditional production to a fully integrated and digitalized sector. However, once again academic research does not offer a clear definition and framework of these important innovation milestones: while a definition for "Digitization" and "Digitalization" can be found in Gartner's Information Technology Glossary [18], "Digital Transformation" still does not have a clear definition in literature. The topics presented in this Doctoral investigation clearly fall under the "Digitalization" phase but cannot yet be considered part of a greater "Digital transformation" unless they can be framed in a clear strategic program driven by industrial companies, aiming to leverage digitalization to create new processes dealing with people's competences, mindsets and business culture in order to transform the organization and maximize the new opportunities provided by the digital approach. This concept is further underlines in Section 1.3, where it is clearly discussed that in order to bring true transformational value I4.0 solutions must be understood and integrated in a wider strategic transformation program. They must be considered a catalyst towards improved performance, not the final purpose. Lessons learned from past projects and applications have demonstrated that unless I4.0 solutions are developed within a clear continuous improvement framework (such as Lean Production, TPM, WCM, ...), rather than increasing the competitiveness of the industrial system they can have the opposite effect, for example by digitalizing its losses. On the other hand, if applied on top of a clear continuous improvement framework, I4.0 can help fine-tune improvements by identifying losses which were previously impossible to identify, or by attacking losses which were previously impossible to attack. For example, [5] and [22] highlight how, in terms of performance, industrial continuous improvement programs can generate savings up to 10% a year, onto which I4.0 and Digitalization can add a further 2%.

The results of this Doctoral investigation therefore further research and confirm the applicability and potential benefits of innovative applications for smart energy management involving the I4.0 features IIoT and AI, key enabling technologies Simulation Technologies, IIoT and Big Data & Analytics, design principles of Interoperability, Real-time Capability, and Modularity, all to develop innovative solutions for the effective forecasting of both power demand and supply, leveraging ML methods and particularly those based on ANNs. However, to successfully bridge the gap between academic research and industrial applicability, these innovations must be understood and integrated in a wider strategic transformation program. They must be considered a catalyst towards improved performance, not the final purpose.

The results of this Doctoral investigation, and the ability of its proposed models to successfully predict forecasting outcomes accurately, can benefit industrial companies not only from the economic and competitiveness point of view, but also in terms of environmental sustainability. As discussed in Section 1.3, according to [23], the environmental problems directly related to energy production and consumption include air pollution, climate change, water pollution, thermal pollution, and solid waste disposal. For example, the emission of air pollutants from fossil fuel combustion is the major cause of urban air pollution, and burning fossil fuels is also the main contributor to the emission of greenhouse gases. As presented by [24], energy derived from fossil fuels contributes significantly to global climate change, accounting for more than 75% of global greenhouse gas emissions and approximately 90% of all carbon dioxide emissions. According to [25], due to their high energy density, fossil fuels are the primary energy source worldwide; however, fossil fuel combustion produces greenhouse gases; approximately 35% of greenhouse gases are emitted by existing power plants. [26], on the other hand, presents that China's coal-fired power plants emit 42% of nitrous oxides and 38% of sulphur dioxides, for a total of 40% of the heat-trapping greenhouse gases, thereby increasing global temperature. As presented by [24], over 300 natural disasters were caused by climate change in 2018, affecting more than 68 million people and causing approximately $131.7 billion in economic losses, with storms, wildfires, floods, and droughts accounting for 93%. In order to reduce the environmental impact of their energy needs, companies and societies as a whole can follow the following paths: (i) reducing energy demand by reducing and/or optimizing energy consumption; (ii) better synchronize energy demand and supply, to avoid energy waste and thus energy over-production; and (iii) migrating to more sustainable energy sources.

For the first two points, as presented in Section 1.4.2, smart energy management provides businesses with numerous opportunities to better manage power demand and supply. [35] and [36], for example, discuss the benefits of efficiently managing electricity usage in smart industries and smart cities through the installation or integration of efficient smart grids, in which innovative methods and tools such as data analytics and ML techniques are integrated with sensors and remote controls, while [31] and [33] present the opportunities coming from renewable energy integration, the on-site installation of renewable energy generation systems, such as solar panels, to offset their reliance on the grid, lower energy costs, and reduce environmental impact. As presented by [173], the application of smart energy management solutions to better manage power demand and supply can reduce peak load and energy expenditures by 7% and 10% respectively, thus reducing primary fossil energy usage by up to 14.5%, and therefore reducing energy expenses up to 55%. For the third point, as discussed by [24], alternative energy from renewable sources must be utilized to decarbonize the energy sector. From the supply point of view, sources of alternative energy generation, such as solar energy sites, can be installed to balance energy needs. Renewable energy integration, according to [31] and [33], occurs when industries install on-site renewable energy generation systems, such as solar panels or wind turbines, to offset their reliance on the grid, lower energy costs, and reduce environmental impact. Within this framework, the increasing significance of renewable energy sources such as PV energy is evident. For example, as presented by [174], environmental analysis showed that the greenhouse gas emissions of renewable energy production such as PV energy can be up to 90% lower than those of traditional fossil fuel systems. However, as discussed in Section 5.1, PV energy falls under the category of variable renewable energy sources (VRE) due to its fluctuating power output derived from solar energy. PV power's variable character poses a challenge to its use as a reliable energy source in power systems, whose stability is highly dependent on the equilibrium between energy generation and consumption. So for such applications to function properly, it has become increasingly essential to be able to effectively forecast power demand and supply with varying horizons. When looking at innovative solutions for forecasting, with the support of Big Data & Analytics, one enters the ML realm. As presented by [39], ML methods are gaining popularity in the forecasting field and, of the different ML methods, those based on ANNs present substantial improvements in forecasting modelling compared to benchmarks, also for accurate forecasting of PV power generation, as discussed by [47], [48], [49], [50], and [149]. Therefore the results of this Doctoral investigation, and the ability of its proposed models to successfully predict forecasting outcomes accurately, can benefit industrial companies not only from the economic and competitiveness point of view, but also in terms of environmental sustainability.

Finally, following the success of the common methodology developed during this Doctoral investigation, it would be interesting to further pursue the topics of this Doctoral investigation in other areas of the industrial and manufacturing world where the methodology and ANN models can be applied, beyond power demand and supply forecasting. One possible are could be predictive maintenance, which has traditionally been a very vibrant area for forecasting and prediction, as demonstrated for example by the research carried out by [175], [176] and [177]. However, as discussed in Section 1.3, it is important that when proposing innovation in the I4.0 realm these are set into a clear industrial transformation strategy, because in order to bring true transformational value I4.0 solutions must be understood and integrated in a wider strategic transformation program. They must be considered a catalyst towards improved performance, not the final purpose. Lessons learned from past projects and applications have demonstrated that unless I4.0 solutions are developed within a clear continuous improvement framework (such as Lean Production, TPM, WCM, . . .), rather than increasing the competitiveness of the industrial system they can have the opposite effect, for example by digitalizing its losses. On the other hand, if applied on top of a clear continuous improvement framework, I4.0 can help fine-tune improvements by identifying losses which were previously impossible to identify, or by attacking losses which were previously impossible to attack. For example, [5] and [22] highlight how, in terms of performance, industrial continuous improvement programs can generate savings up to 10% a year, onto which I4.0 and Digitalization can add a further 2%. This is particularly important especially in the maintenance sphere, since strong and extensive literature on maintenance theory, equipment efficiency improvement, and maintenance cost optimization already exists, and the academic research has widely been integrated with long and mature industrial applications in the maintenance area. The application of I4.0 solutions can therefore be studied to further enhance existing maintenance systems, but cannot replace them. They must be considered a catalyst towards improved performance, not the final purpose.

References

[1] Alessandro Merlino. Sustainable approach for investment in industrial digitalization. Master's thesis, Politecnico di Torino, 2023.

[2] Governo Italiano MISE Ministero dello Sviluppo Economico. Piano nazionale industria 4.0, 2016. [Online; accessed 15-September-2019].

[3] M.E. Porter. *The Competitive Advantage: Creating and Sustaining Superior Performance*. NY: Free Press, 1985.

[4] Detlef Zühlke. Smartfactory – from vision to reality in factory technologies. *IFAC Proceedings Volumes*, 41(2):14101–14108, 2008. 17th IFAC World Congress.

[5] F. Canuto. World class manufacturing evolution and innovation - fca (italy), 2017. [XXVIII Riunione Scientifica Annuale, RSA AiIG 2017. Bari, 19-20 October 2017. Document online, accessed 10-July-2020].

[6] Alessandro Aliberti. *Machine learning techniques to forecast non-linear trends in smart environments*. PhD thesis, Politecnico di Torino, 2020.

[7] Confindustria. Dove va l'industria italiana - rapporto 2019, 2019. [Online; accessed 15-September-2020].

[8] Giovanna Culot, Guido Nassimbeni, Guido Orzes, and Marco Sartor. Behind the definition of industry 4.0: Analysis and open questions. *International Journal of Production Economics*, 226:107617, 01 2020.

[9] H. Kagermann, J. Helbig, W. Wahlster, and A. Hellinger. *Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry ; Final Report of the Industrie 4.0 Working Group*. Forschungsunion, 2013.

[10] Henning Kagermann. *Change Through Digitization—Value Creation in the Age of Industry 4.0*, pages 23–45. Springer, 12 2015.

[11] EESC European Economic and Social Committee. For a european industrial renaissance, Apr 2014. [Online; accessed 15-September-2020].

[12] Inanc Kabasakal, Fatma Demircan Keskin, Keti Ventura, and Haluk Soyuer. From mass customization to product personalization in automotive industry: potentials of industry 4.0. *Pressacademia*, 4:244–250, 04 2017.

[13] Jay Lee, Behrad Bagheri, and Hung-An Kao. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters*, 3:18–23, 2015.

[14] D. Uckelmann, M. Harrison, and F. Michahelles. *Architecting the Internet of Things*. Springer Berlin Heidelberg, 2011.

[15] Mario Hermann, Tobias Pentek, and Boris Otto. Design principles for industrie 4.0 scenarios: A literature review, 01 2015.

[16] Barbara Bigliardi, Eleonora Bottani, and Giorgia Casella. Enabling technologies, application areas and impact of industry 4.0: a bibliographic analysis. *Procedia Manufacturing*, 42:322–326, 01 2020.

[17] Arianna Martinelli, Andrea Mina, and Massimo Moggi. The enabling technologies of industry 4.0: examining the seeds of the fourth industrial revolution, 01 2021.

[18] Gartner. Information technology glossary, 2023.

[19] European Commission. Digital maturity tool and the innovation radar, May 2021. [Online; accessed 15-September-2022].

[20] Michael E. Porter and James E. Heppelmann. How smart, connected products are transforming competition, Nov 2014.

[21] T. Stock and G. Seliger. Opportunities of sustainable manufacturing in industry 4.0. *Procedia CIRP*, 40:536–541, 2016.

[22] Antonella Petrillo, Fabio De Felice, and Federico Zomparelli. Performance measurement for world-class manufacturing: a model for the italian automotive industry. *Total Quality Management & Business Excellence*, 30(7-8):908–935, 2019.

[23] EEA European Environment Agency. Transformer neural network: Step-by-step breakdown of the beast. Accessed on: 2023-08-27.

[24] Ahmed I. Osman, Lin Chen, Mingyu Yang, Goodluck Msigwa, Mohamed Farghali, Samer Fawzy, David W. Rooney, and Pow-Seng Yap. Cost, environmental impact, and resilience of renewable energy under a changing climate: a review. *Environmental Chemistry Letters*, 21(2):741–764, Apr 2023.

[25] Ala'a H. Al-Muhtaseb, Ahmed I. Osman, Paskalis Sahaya Murphin Kumar, Farrukh Jamil, Lamya Al-Haj, Abdulrahman Al Nabhani, Htet Htet Kyaw, Myo Tay Zar Myint, Neha Mehta, and David W. Rooney. Circular economy approach of enhanced bifunctional catalytic system of cao/ceo2 for biodiesel production from waste loquat seed oil with life cycle assessment study. *Energy Conversion and Management*, 236:114040, 2021.

[26] Ying Yang, Pietro Elia Campana, and Jinyue Yan. Potential of unsubsidized distributed solar pv to replace coal-fired power plants, and profits classification in chinese cities. *Renewable and Sustainable Energy Reviews*, 131:109967, 2020.

[27] EIA U.S. Energy Information Administration. Natural gas explained - factors affecting natural gas prices, Oct 2022. [Online; accessed 15-December-2022].

[28] IEA International Energy Agency. World energy outlook 2020, Oct 2020. [Online; accessed 10-December-2020].

[29] Daron Acemoglu, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr. Innovation, reallocation, and growth. *American Economic Review*, 108(11):3450–91, November 2018.

[30] Carlos Herce, Enrico Biele, Chiara Martini, Marcello Salvio, and Claudia Toro. Impact of energy monitoring and management systems on the implementation and planning of energy performance improved actions: An empirical analysis based on energy audits in italy. *Energies*, 14(16), 2021.

[31] Christos Timplalexis, Georgios-Fotios Angelis, Stylianos Zikos, Stelios Krinidis, Dimosthenis Ioannidis, and Dimitrios Tzovaras. *A comprehensive review on industrial demand response strategies and applications*, pages 1–23. The Institution of Engineering and Technology, 06 2022.

[32] Imane Worighi, Abdelilah Maach, Abdelhakim Hafid, Omar Hegazy, and Joeri Van Mierlo. Integrating renewable energy in smart grid system: Architecture, virtualization and analysis. *Sustainable Energy, Grids and Networks*, 18:100226, 2019.

[33] Nicholas Preston, Azadeh Maroufmashat, Hassan Riaz, Sami Barbouti, Ushnik Mukherjee, Peter Tang, Javan Wang, Ehsan Haghi, Ali Elkamel, and Michael Fowler. How can the integration of renewable energy and power-to-gas benefit industrial facilities? from techno-economic, policy, and environmental assessment. *International Journal of Hydrogen Energy*, 45(51):26559–26573, 2020.

[34] IEA International Energy Agency. Technology roadmap - energy-efficient buildings - heating and cooling equipment, Jun 2011. [Online; accessed 10-December-2020].

[35] A Bonati, G De Luca, S Fabozzi, N Massarotti, and L Vanoli. The integration of exergy criterion in energy planning analysis for 100% renewable system. *Energy*, 174:749–767, 2019.

[36] Maedeh Ghorbanian, Sarineh Hacopian Dolatabadi, and Pierluigi Siano. Big data issues in smart grids: A survey. *IEEE Systems Journal*, 13(4):4158–4168, 2019.

[37] Lars T Berger and Krzysztof Iniewski. *Smart grid applications, communications, and security*. John Wiley & Sons, 2012.

[38] Pierluigi Siano. Demand response and smart grids—a survey. *Renewable and sustainable energy reviews*, 30:461–478, 2014.

[39] Casper Solheim Bojer. Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities. *International Journal of Forecasting*, 38(4):1555–1561, 2022. Special Issue: M5 competition.

[40] European Parliament. Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings, 2010. [Online; accessed 12-October-2020].

[41] Zakia Afroz, GM Shafiullah, Tania Urmee, and Gary Higgins. Modeling techniques used in building hvac control systems: A review. *Renewable and Sustainable Energy Reviews*, 83:64–84, 2018.

[42] Xingchao Zhou. Thermal and energy performance of a solar-driven desiccant cooling system using an internally cooled desiccant wheel in various climate conditions. *Applied Thermal Engineering*, 185:116077, 2021.

[43] Ali Deihimi and Hemen Showkati. Application of echo state networks in short-term electric load forecasting. *Energy*, 39(1):27–340, 2012.

[44] IEA International Energy Agency. World energy outlook 2022. Technical report, IEA - International Energy Agency, 2022.

[45] Hugo Chandler. Empowering variable renewables-options for flexible electricity systems, 2008.

[46] Semich Impram, Secil Varbak Nese, and Bülent Oral. Challenges of renewable energy penetration on power system flexibility: A survey. *Energy Strategy Reviews*, 31:100539, 2020.

[47] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.J. Martinez de Pison, and F. Antonanzas-Torres. Review of photovoltaic power forecasting. *Solar Energy*, 136:78–111, 2016.

[48] Fei Wang, Zengqiang Mi, Shi Su, and Hongshan Zhao. Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. *Energies*, 5:1355–1370, 12 2012.

[49] H.Z. Wang, G.B. Wang, G.Q. Li, J.C. Peng, and Y.T. Liu. Deep belief network based deterministic and probabilistic wind speed forecasting approach. *Applied Energy*, 182:80–93, 2016.

[50] Hugo T.C. Pedro and Carlos F.M. Coimbra. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86(7):2017–2028, 2012.

[51] Andreas Weigend. *Time Series Prediction: Forecasting The Future And Understanding The Past*. Routledge, 05 2018.

[52] P.J. Brockwell and R.A. Davis. *Introduction to Time Series and Forecasting*. Springer Texts in Statistics. Springer New York, 2006.

[53] Rob J. Hyndman. Yule-Walker Estimates For Continuous-Time Autoregressive Models. *Journal of Time Series Analysis*, 14(3):281–296, May 1993.

[54] Donald Specht. A general regression neural network. *IEEE Transactions on Neural Networks*, 2:568–578, 11 1991.

[55] Mark Deakin and Husam Al Waer. From intelligent to smart cities. *Intelligent Buildings International*, 3(3):140–152, 2011.

[56] Pierluigi Siano. Demand response and smart grids—a survey. *Renewable and Sustainable Energy Reviews*, 30:461–478, 2014.

[57] Kelly E. Parmenter Clark W. Gellings. *Demand-Side Management*, chapter 15. CRC Press, 2015.

[58] Miroslav Kubat. *Artificial Neural Networks*, pages 91–111. Springer International Publishing, 01 2017.

[59] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.

[60] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.

[61] Amanpreet Singh, Narina Thakur, and Aakanksha Sharma. A review of supervised machine learning algorithms. *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 1310–1315, 2016.

[62] Alboukadel Kassambara. *Practical guide to cluster analysis in R: Unsupervised machine learning*, volume 1. Sthda, 2017.

[63] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[64] José María Cavanillas, Edward Curry, and Wolfgang Wahlster. The big data value opportunity. *New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in Europe*, pages 3–11, 2016.

[65] Omid Ameri Sianaki, Ashkan Yousefi, Azadeh Rajabian Tabesh, and Mehregan Mahdavi. Machine learning applications: The past and current research trend in diverse industries. *Inventions*, 4(1):8, 2019.

[66] Nils J Nilsson. Artificial intelligence: A modern approach. *Prentice Hall*, 1996.

[67] Martin T Hagan, Howard B Demuth, and Mark Beale. *Neural network design*. PWS Publishing Co., 1997.

[68] Frederic B Fitch. Warren s. mcculloch and walter pitts. a logical calculus of the ideas immanent in nervous activity. bulletin of mathematical biophysics, vol. 5 (1943), pp. 115–133. *The Journal of Symbolic Logic*, 9(2):49–50, 1944.

[69] Donald O Hebb. The organization of behavior; a neuropsycholocigal theory. *A Wiley Book in Clinical Psychology.*, pages 62–78, 1949.

[70] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology press, 2005.

[71] BWAC Farley and W do Clark. Simulation of self-organizing systems by digital computer. *Transactions of the IRE Professional Group on Information Theory*, 4(4):76–84, 1954.

[72] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[73] Marvin Minsky and Seymour Papert. An introduction to computational geometry. *Cambridge tiass., HIT*, 479:480, 1969.

[74] Paul Werbos. Beyond regression: New tools for prediction and analysis in the behavioral sciences. *PhD thesis, Committee on Applied Mathematics, Harvard University, Cambridge, MA*, 1974.

[75] Douglas C Montgomery, Cheryl L Jennings, and Murat Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.

[76] Danilo Mandic and Jonathon Chambers. *Recurrent neural networks for prediction: learning algorithms, architectures and stability*. Wiley, 2001.

[77] Tomas Mikolov, Armand Joulin, Sumit Chopra, Michael Mathieu, and Marc'Aurelio Ranzato. Learning longer memory in recurrent neural networks. *arXiv preprint arXiv:1412.7753*, 2014.

[78] Georg Dorffner. Neural networks for time series processing. *Neural network world*, 6(4):447–468, 1996.

[79] Jerome Connor, R.Douglas Martin, and L.E. Atlas. Recurrent neural networks and robust time series prediction. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 5:240–54, 02 1994.

[80] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, 5:157–66, 02 1994.

[81] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. *Neural computation*, 12(10):2451–2471, 2000.

[82] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[83] Christopher Olah. Understanding lstm networks, 2015.

[84] David H Hubel and Torsten N Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243, 1968.

[85] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[86] Steve Lawrence, C Giles, Ah Tsoi, and Andrew Back. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8:98–113, 02 1997.

[87] Hokuto Kagaya, Kiyoharu Aizawa, and Makoto Ogawa. Food detection and recognition using convolutional neural network. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 1085–1088, 2014.

[88] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, 1, 04 2014.

[89] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J Inman. 1d convolutional neural networks and applications: A survey. *Mechanical systems and signal processing*, 151:107398, 2021.

[90] Magnus Nørgaard, Ole Ravn, Niels Poulsen, and Hansen LK. *Neural Networks for Modelling and Control of Dynamic Systems: A Practitioner's Handbook*. Springer, 01 2000.

[91] Benjamin Schrauwen, David Verstraeten, and Jan Van Campenhout. An overview of reservoir computing: theory, applications and implementations. In *Proceedings of the 15th european symposium on artificial neural networks. p. 471-482 2007*, pages 471–482, 2007.

[92] Girish Chandrashekar and Ferat Sahin. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28, 2014.

[93] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[94] Geoffrey F Miller, Peter M Todd, and Shailesh U Hegde. Designing neural networks using genetic algorithms. In *ICGA*, volume 89, pages 379–384, 1989.

[95] Song Han, Jeff Pool, John Tran, and William Dally. Learning both weights and connections for efficient neural network. *Advances in neural information processing systems*, 28, 2015.

[96] Douglas M Kline. Methods for multi-step time series forecasting neural networks. In *Neural networks in business forecasting*, pages 226–250. IGI Global, 2004.

[97] Donna Xu, Yaxin Shi, Ivor W Tsang, Yew-Soon Ong, Chen Gong, and Xiaobo Shen. Survey on multi-output learning. *IEEE transactions on neural networks and learning systems*, 31(7):2409–2429, 2019.

[98] Kejun Wang, Xiaoxia Qi, and Hongda Liu. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Applied Energy*, 251:113315, 2019.

[99] Hasan Tercan, Alexandro Guajardo, Julian Heinisch, Thomas Thiele, Christian Hopmann, and Tobias Meisen. Transfer-learning: Bridging the gap between real and simulation data for machine learning in injection molding. *Procedia CIRP*, 72:185–190, 01 2018.

[100] Yujiao Chen, Zheming Tong, Yang Zheng, Holly Samuelson, and Leslie Norford. Transfer learning with deep neural networks for model predictive control of heating, cooling, and natural ventilation in smart buildings. *Journal of Cleaner Production*, 254:119866, 12 2020.

[101] Christian A Gueymard. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. *Renewable and Sustainable Energy Reviews*, 39:1024–1034, 2014.

[102] Lorenzo Bottaccioli, Alessandro Aliberti, Francesca Ugliotti, Edoardo Patti, Anna Osello, Enrico Macii, and Andrea Acquaviva. Building energy modelling and monitoring by integration of iot devices and building information models. In *2017 IEEE 41st annual computer software and applications conference (COMPSAC)*, volume 1, pages 914–922. IEEE, 2017.

[103] Matteo Del Giudice, Anna Osello, Arianna Fonsati, Daniela De Luca, Alberto Musetti, Francesca Marchi, Edoardo Patti, Francesco Brundu, and Andrea Acquaviva. District data management, modelling and visualization via interoperability. *Building simulation. San Francisco*, 2017. [DOI not available. Article available at the following link: https://iris.polito.it/handle/11583/2680251. Accessed 07-September-2020.].

[104] Lorenzo Bottaccioli, Edoardo Patti, Enrico Macii, and Andrea Acquaviva. Gis-based software infrastructure to model pv generation in fine-grained spatio-temporal domain. *IEEE Systems Journal*, 12(3):2832–2841, 2017.

[105] Edoardo Patti, Andrea Acquaviva, Marco Jahn, Ferry Pramudianto, Riccardo Tomasi, Damien Rabourdin, Joseph Virgone, and Enrico Macii. Event-driven user-centric middleware for energy-efficient buildings and public spaces. *IEEE Systems Journal*, 10(3):1137–1146, 2016.

[106] Pierluigi Siano. Demand response and smart grids—a survey. *Renewable and Sustainable Energy Reviews*, 30:461–478, 2014.

[107] Jochen L Cremer, Marco Pau, Ferdinanda Ponci, and Antonello Monti. Optimal scheduling of heat pumps for power peak shaving and customers thermal comfort. In *SMARTGREENS*, pages 23–34, 2017.

[108] Michal Veselỳ and Wim Zeiler. Personalized conditioning and its impact on thermal comfort and energy performance–a review. *Renewable and Sustainable Energy Reviews*, 34:401–408, 2014.

[109] Francesco Gavino Brundu, Edoardo Patti, Anna Osello, Matteo Del Giudice, Niccolò Rapetti, Alexandr Krylovskiy, Marco Jahn, Vittorio Verda, Elisa Guelpa, Laura Rietto, and Andrea Acquaviva. Iot software infrastructure for energy management and simulation in smart cities. *IEEE Transactions on Industrial Informatics*, 13(2):832–840, 2017.

[110] Gianni Bianchini, Marco Casini, Antonio Vicino, and Donato Zarrilli. Demand-response in building heating systems: A model predictive control approach. *Applied Energy*, 168:159–170, 04 2016.

[111] Alessandro Aliberti, Lorenzo Bottaccioli, Giansalvo Cirrincione, Enrico Macii, Andrea Acquaviva, and Edoardo Patti. Forecasting short-term solar radiation for photovoltaic energy predictions. In *SMARTGREENS*, pages 44–53, 2018.

[112] Joseph Andrew Clarke and JLM Hensen. Integrated building performance simulation: Progress, prospects and requirements. *Building and Environment*, 91:294–306, 2015.

[113] Gonçalo Marques and Rui Pitarma. A cost-effective air quality supervision solution for enhanced living environments through the internet of things. *Electronics*, 8(2):170, 2019.

[114] Hao Gao, Christian Koch, and Yupeng Wu. Building information modelling based building energy modelling: A review. *Applied Energy*, 238:320–343, 2019.

[115] Drury B Crawley, Linda K Lawrie, Frederick C Winkelmann, Walter F Buhl, Y Joe Huang, Curtis O Pedersen, Richard K Strand, Richard J Liesen, Daniel E Fisher, Michael J Witte, and Jason Glazer. Energyplus: creating a

new-generation building energy simulation program. *Energy and buildings*, 33(4):319–331, 2001.

[116] SA Klein, WA Beckman, JW Mitchell, JA Duffie, NA Duffie, TL Freeman, JC Mitchell, JE Braun, B Evans, J Kummer, et al. Trnsys 16: A transient system simulation program: mathematical reference. *Trnsys*, 5:389–396, 2007. [DOI not available. Material available on numerous online sources, including the following link: http://web.mit.edu/parmstr/Public/Documentation/05-MathematicalReference.pdf. Accessed 03-October-2020.].

[117] U.S. Department of Energy's (DOE) Building Technologies Office (BTO). Energyplus, Sep 2020. [Online; accessed 15-September-2020].

[118] U.S. Department of Energy's (DOE) Building Technologies Office (BTO). Energyplus: Energy simulation software for buildings, Sep 2020. [Online; accessed 15-September-2020].

[119] A Castell and C Solé. Design of latent heat storage systems using phase change materials (pcms). In *Advances in Thermal Energy Storage Systems*, pages 285–305. Elsevier, 2015.

[120] Christoph F Reinhart and Carlos Cerezo Davila. Urban building energy modeling–a review of a nascent field. *Building and Environment*, 97:196–202, 2016.

[121] Marco Massano, Enrico Macii, Edoardo Patti, Andrea Acquaviva, and Lorenzo Bottaccioli. A grey-box model based on unscented kalman filter to estimate thermal dynamics in buildings. In *2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, pages 1–6. IEEE, 2019.

[122] Peder Bacher and Henrik Madsen. Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings*, 43:1511–1522, 07 2011.

[123] Duo Li, Sheldon X-D Tan, Eduardo H Pacheco, and Murli Tirumala. Parameterized architecture-level dynamic thermal models for multicore microprocessors. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 15(2):1–226, 2010.

[124] Thom J Eguia, Sheldon X-D Tan, Ruijing Shen, Duo Li, Eduardo H Pacheco, Murli Tirumala, and Lingli Wang. General parameterized thermal modeling for high-performance microprocessor design. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 20(2):211–224, 2012.

[125] Antonio E Ruano, Eduardo M Crispim, Eusébio ZE Conceiçao, and Ma Manuela JR Lúcio. Prediction of building's temperature using neural networks models. *Energy and Buildings*, 38(6):682–694, 2006.

[126] P.M. Ferreira, A.E. Ruano, S. Silva, and E.Z.E. Conceição. Neural networks based predictive control for thermal comfort and energy savings in public buildings. *Energy and Buildings*, 55:238–251, 2012.

[127] Fernando Mateo, Juan José Carrasco, Abderrahim Sellami, Mónica Millán-Giraldo, Manuel Domínguez, and Emilio Olivas. Machine learning methods to forecast temperature in buildings. *Expert Systems with Applications*, 40:1061––1068, 03 2013.

[128] Alessandro Aliberti, Francesca Maria Ugliotti, Lorenzo Bottaccioli, Giansalvo Cirrincione, Anna Osello, Enrico Macii, Edoardo Patti, and Andrea Acquaviva. Indoor air-temperature forecast for energy-efficient management in smart buildings. In *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, pages 1–6. IEEE, 2018.

[129] Chengliang Xu, Huanxin Chen, Jiangyu Wang, Yabin Guo, and Yue Yuan. Improving prediction performance for indoor temperature in public buildings based on a novel deep learning method. *Building and Environment*, 148:128–135, 2019.

[130] E. Kamel, A. Javan-Khoshkholgh, N. Abumahfouz, S. Huang, X. Huang, A. Farajidavar, and Y. Qiu. A case study of using multi-functional sensors to predict the indoor air temperature in classrooms. *ASHRAE Transactions*, 126:3–11, 2020. [DOI not available. Article available at the following link: https://www.thefreelibrary.com/A+Case+Study+of+Using+Multi-Functional+Sensors+to+Predict+the+Indoor...-a0627513870. Accessed 27-August-2020.].

[131] Alessandro Aliberti, Lorenzo Bottaccioli, Enrico Macii, Santa Di Cataldo, Andrea Acquaviva, and Edoardo Patti. A non-linear autoregressive model for indoor air-temperature predictions in smart buildings. *Electronics*, 8(9):979, 2019.

[132] Jenny Cifuentes Quintero, Geovanny Marulanda, Antonio Bello, and Javier Reneses. Air temperature forecasting using machine learning techniques: A review. *Energies*, 13:4215, 08 2020.

[133] Sadi Alawadi, David Mera, Manuel Fernández-Delgado, Fahed Alkhabbas, Carl Magnus Olsson, and Paul Davidsson. A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings. *Energy Systems*, pages 1–17, 2020.

[134] Zhanhong Jiang and Young M. Lee. Deep transfer learning for thermal dynamics modeling in smart buildings. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 2033–2037. IEEE, 2019.

[135] Nan Gao, Wei Shao, Mohammad Saiedur Rahaman, Jun Zhai, Klaus David, and Flora D. Salim. Transfer learning for thermal comfort prediction in multiple cities. *Building and Environment*, 195:107725, 2021.

[136] Shichao Xu, Yixuan Wang, Yanzhi Wang, Zheng O'Neill, and Qi Zhu. One for many: Transfer learning for building hvac control. In *Proceedings of the 7th ACM international conference on systems for energy-efficient buildings, cities, and transportation*, pages 230–239, 2020.

[137] Zhipeng Deng and Qingyan Chen. Reinforcement learning of occupant behavior model for cross-building transfer learning to various hvac control systems. *Energy and Buildings*, 238:110860, 2021.

[138] Poul O Fanger. *Thermal comfort. Analysis and applications in environmental engineering*. Copenhagen: Danish Technical Press., 1970.

[139] M.S. Owen, Refrigerating American Society of Heating, and Air-Conditioning Engineers. *2009 ASHRAE Handbook: Fundamentals*. 2009 Ashrae Handbook - Fundamentals. American Society of Heating, Refrigeration and Air-Conditioning Engineers, 2009.

[140] B.W. Olesen. International standards and the ergonomics of the thermal environment. *Applied ergonomics*, 26:293–302, 09 1995.

[141] Michelangelo Grosso, Salvatore Rinaudo, Edoardo Patti, and Andrea Acquaviva. An energy-autonomous wireless sensor network development platform. In *2018 13th International Conference on Design Technology of Integrated Systems In Nanoscale Era (DTIS)*, pages 1–6, 2018.

[142] ANSI/ASHRAE Standard 55. Thermal environmental conditions for human occupancy, 2013. [ASHRAE 55-2013 standard available for purchase on this website. Purchased in March 2015.].

[143] Charu C Aggarwal. *Neural networks and deep learning*, volume 10. Springer, 2018.

[144] Douglas M Kline. Methods for multi-step time series forecasting neural networks. In *Neural networks in business forecasting*, pages 226–250. IGI Global, 2004.

[145] Siamak Mehrkanoon. Deep shared representation learning for weather elements forecasting. *Knowledge-Based Systems*, 179:120–128, 06 2019.

[146] Deepak Soekhoe, Peter Putten, and Aske Plaat. On the impact of data set size in transfer learning using deep neural networks. In *International symposium on intelligent data analysis*, pages 50–60, 10 2016.

[147] Ricardo Ribani and Maurício Marengoni. A survey of transfer learning for convolutional neural networks. In *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*, pages 47–57. IEEE, 10 2019.

[148] Yixin Yu. Editorial for the special issue on smart grid and energy internet, 2020.

[149] Alfredo Nespoli, Emanuele Ogliari, Sonia Leva, Alessandro Massi Pavan, Adel Mellit, Vanni Lughi, and Alberto Dolara. Day-ahead photovoltaic forecasting: A comparison of the most effective techniques. *Energies*, 12(9):1621, 2019.

[150] Marco Pierro, Matteo De Felice, Enrico Maggioni, David Moser, Alessandro Perotto, Francesco Spada, and Cristina Cornaro. Data-driven upscaling methods for regional photovoltaic power estimation and forecast using satellite and numerical weather prediction data. *Solar Energy*, 158:1026–1038, 2017.

[151] Changsong Chen, Shanxu Duan, Tao Cai, and Bangyin Liu. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Solar energy*, 85(11):2856–2870, 2011.

[152] Maria Grazia De Giorgi, Paolo Maria Congedo, and Maria Malvoni. Photovoltaic power forecasting using statistical methods: impact of weather data. *IET Science, Measurement & Technology*, 8(3):90–97, 2014.

[153] Utpal Kumar Das, Kok Soon Tey, Mehdi Seyedmahmoudian, Mohd Yamani Idna Idris, Saad Mekhilef, Ben Horan, and Alex Stojcevski. Svr-based model to forecast pv power generation under different weather conditions. *Energies*, 10(7):876, 2017.

[154] A Mellit, A Massi Pavan, and V Lughi. Deep learning neural networks for short-term photovoltaic power forecasting. *Renewable Energy*, 172:276–288, 2021.

[155] Vidisha De, Tiong Teck Teo, Wai Lok Woo, and T Logenthiran. Photovoltaic power forecasting using lstm on limited dataset. In *2018 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, pages 710–715. IEEE, 2018.

[156] Donghun Lee and Kwanho Kim. Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information. *Energies*, 12(2):215, 2019.

[157] Marco Castangia, Alessandro Aliberti, Lorenzo Bottaccioli, Enrico Macii, and Edoardo Patti. A compound of feature selection techniques to improve solar radiation forecasting. *Expert Systems with Applications*, 178:114979, 2021.

[158] Siyu Zhou, Lin Zhou, Mingxuan Mao, and Xinze Xi. Transfer learning for photovoltaic power forecasting with long short-term memory neural network. In *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 125–132. IEEE, 2020.

[159] Xing Luo, Dongxiao Zhang, and Xu Zhu. Combining transfer learning and constrained long short-term memory for power generation forecasting of newly-constructed photovoltaic plants. *Renewable Energy*, 185:1062–1077, 2022.

[160] Zhichao Yu, Zichao Yang, Fiedler Frank, Hong Rao, and William Wie Song. *Power Generation Prediction of Residential Photovoltaic Equipment Based on Online Transfer Learning Model- A Case Study of a Residential Solar Power System*, page 58–65. Association for Computing Machinery, 2021.

[161] The Dark Sky Company LLC. Dark sky api. Accessed on: 2019-06-01.

[162] William Holmgren, Clifford Hansen, and Mark Mikofski. pvlib python: a python package for modeling solar energy systems. *Journal of Open Source Software*, 3:884, 09 2018.

[163] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[164] Ian H Witten, Eibe Frank, Mark A Hall, CJ Pal, and MINING DATA. Practical machine learning tools and techniques. In *DATA MINING*, volume 2, page 4, 2005.

[165] Christian A Gueymard. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. *Renewable and Sustainable Energy Reviews*, 39:1024–1034, 2014.

[166] Deepak Soekhoe, Peter Putten, and Aske Plaat. On the impact of data set size in transfer learning using deep neural networks. In *International symposium on intelligent data analysis*, pages 50–60, 10 2016.

[167] Ricardo Ribani and Maurício Marengoni. A survey of transfer learning for convolutional neural networks. In *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*, pages 47–57. IEEE, 10 2019.

[168] Utkarsh Ankit. Transformer neural network: Step-by-step breakdown of the beast. Accessed on: 2023-08-23.

[169] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[170] R Farsani, Ehsan Pazouki, and Jecei Jecei. A transformer self-attention model for time series forecasting. *Journal of Electrical and Computer Engineering Innovations*, 9:1–10, 01 2021.

[171] Majid Moradi Aliabadi, Hajar Emami, Ming Dong, and Yinlun Huang. Attention-based recurrent neural network for multistep-ahead prediction of process performance. *Computers and Chemical Engineering*, 140:106931, 2020.

[172] Yao Qin, Dongjin Song, Haifeng Cheng, Wei Cheng, Guofei Jiang, and Garrison Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. 04 2017.

[173] Mutiu Shola Bakare, Abubakar Abdulkarim, Mohammad Zeeshan, and Aliyu Nuhu Shuaibu. A comprehensive overview on demand side energy management towards smart grids: challenges, solutions, and future direction. *Energy Informatics*, 6(1):4, Mar 2023.

[174] Prashant Malik, Mamta Awasthi, and Sunanda Sinha. Techno-economic and environmental analysis of biomass-based hybrid energy systems: A case study of a western himalayan state in india. *Sustainable Energy Technologies and Assessments*, 45:101189, 2021.

[175] R Naren and J Subhashini. Comparison of deep learning models for predictive maintenance. *IOP Conference Series: Materials Science and Engineering*, 912(2):022029, aug 2020.

[176] Mounia Achouch, Mariya Dimitrova, Khaled Ziane, Sasan Sattarpanah Karganroudi, Rizck Dhouib, Hussein Ibrahim, and Mehdi Adda. On predictive maintenance in industry 4.0: Overview, models, and challenges. *Applied Sciences*, 12(16), 2022.

[177] Muhammad Sohaib, Shiza Mushtaq, and Jia Uddin. *Deep Learning for Data-Driven Predictive Maintenance*, pages 71–95. 06 2021.

.1 List of Publications

Andrea Bellagarda has been author/coauthor in:

- 1 accepted paper in peer review international journals;
- 2 accepted chapters in peer review international books;
- 2 accepted papers in peer review international conferences.

Peer reviewed International Journals:

J.1 Andrea Bellagarda, Silvia Cesari, Alessandro Aliberti, Francesca Ugliotti, Lorenzo Bottaccioli, Enrico Macii, Edoardo Patti, "Effectiveness of neural networks and transfer learning for indoor air-temperature forecasting." Automation in Construction, Volume 140, 2022, 104314, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2022.104314.

Peer reviewed Books:

B.1 Patti, Edoardo & Brundu, Francesco & Bellagarda, Andrea & Bottaccioli, Lorenzo & Rapetti, Niccolò & Vittorio, Verda & Elisa, Guelpa & Rietto, Laura & Macii, Enrico & Acquaviva, Andrea & Krylovskiy, Alexandr & Jahn, Marco. (2022). Combining BIM, GIS, and IoT to Foster Energy Management and Simulation in Smart Cities. 10.4018/978-1-6684-7548-5.ch016.

B.2 Cerquitelli, Tania & Nikolakis, Nikolaos & Morra, Lia & Bellagarda, Andrea & Orlando, Matteo & Salokangas, Riku & Saarela, Olli & Hietala, Jani & Kaarmila, Petri & Macii, Enrico. (2021). Data-Driven Predictive Maintenance: A Methodology Primer. 10.1007/978-981-16-2940-2-3.

Peer reviewed International Conferences:

IC.1 A. Bellagarda, E. Patti, E. Macii and L. Bottaccioli, "Human daily activity behavioural clustering from Time Use Survey," 2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE), Turin, Italy, 2020, pp. 1-6, doi: 10.23919/AEITAUTOMOTIVE50086.2020.9307408.

IC.2 F. M. Solinas, A. Bellagarda, E. Macii, E. Patti and L. Bottaccioli, "An Hybrid Model-Free Reinforcement Learning Approach for HVAC Control," 2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Bari, Italy, 2021, pp. 1-6, doi: 10.1109/EEEIC/ICPSEurope 51590.2021.9584805.