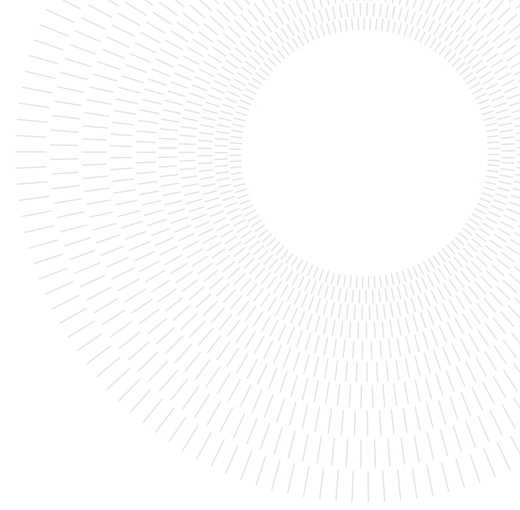**POLITECNICO**

MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Enhanced Convolutional Neural Network for solar radiation nowcasting: All-Sky camera infrared images embedded with exogenous parameters

Tesi di Laurea Magistrale in
Energy Engineering - Ingegneria Energetica

Paolo Cusa, 964768

Advisor:
Prof. Emanuele Giovanni
Carlo Ogliari

Co-advisor:
Alfredo Nespoli

Academic year:
2021-2022

Abstract: The need for an accurate prediction of solar radiation is becoming crucial for both the stability of grid connected and efficient management of stand-alone networks due to the ever-increasing diffusion of electricity production from renewable energy sources. This thesis focuses on nowcasting, a very-short term forecast which ranges from a few minutes to a maximum of half an hour. Specifically, three forecast horizons equal to 5, 10 and 15 minutes ahead have been taken into account. The proposed methodology is based on the development of a Convolutional Neural Network (CNN) which elaborates sequences of images from an infrared all sky imager to detect clouds' features for solar solar radiation forecast. The case study exploits about six months of data, which coincides with 33,686 images filtered by rainy days and hours characterized by a Sun elevation angle lower than 20°. Furthermore, Global Horizontal Irradiance (GHI) measurements and the Ineichen and Perez clear sky model are used as input for the neural network. Images and irradiance data sets have been collected at the SolarTech^{LAB} of the Politecnico di Milano. Moreover, an innovative technique in which exogenous data are inserted as information directly in the images is proposed. The results show that the CNN outperforms the predictions provided by a weather broadcaster company which has exploited the same data. The simple CNN has forecast skill (FS) scores of 0.09% and 0.41% over persistence for 10 and 15 minutes ahead respectively. The Enhanced CNN (ECNN) got positive FS of 5.05%, 9.57% and 9.07% moving from the shortest to the longest forecast horizon. A deeper analysis of the results reveals that the proposed models are more accurate than persistence when high fluctuations of irradiance are registered. In fact, the CNN models get a forecast skill always higher than 19% for all the forecast horizons. The main problems, possible solutions and future works are also discussed.

Introduction

The ever-increasing energy demand leads to a larger spread of power generation in which renewable sources play a key role. Green-technologies are being pushed forward by stronger policies and ambitious climate targets

announced for COP26 along with the current global energy crisis, both of which contribute to the acceleration towards a clean energy transition. This trend is highlighted by the worldwide annual renewable capacity additions, which increased by 6% in 2021 reaching almost 295 gigawatts of power installed [1]. This growth was expected to further increase by over 8% in 2022 despite the continuation of pandemic-driven supply chain challenges and record-level prices for raw materials.
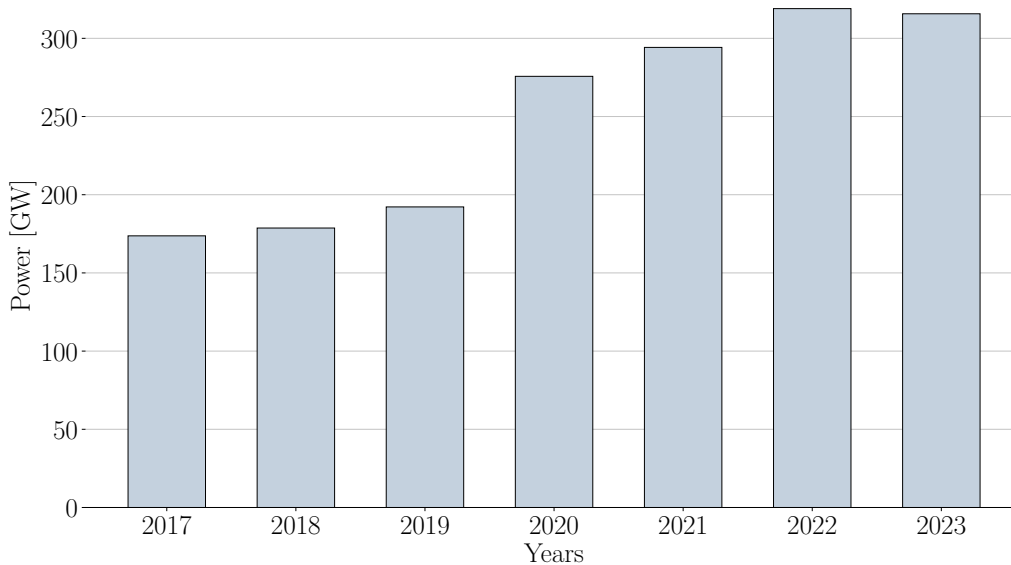

Figure 1: Net renewable capacity additions, 2017-2023 [1].

Solar photovoltaic (PV) is expected to account for 60% of the total growth [2] thanks to its multiple applications, such as residential, commercial and industrial usage; either via grid-connection or stand-alone networks. Thus, solar energy is emerging as one of the most promising source, which will take a major share in energy production in the near future. However, its integration into existing or future energy supply structures is becoming the major challenge due to its nature, which is dynamic and dependent on both geographic locations than by sudden climate change [3].

The need for an accurate solar radiation forecast

The intermittent and unpredictable nature of the solar source, which can induce a power loss of up to 80% in a few minutes, could generate relevant problems in balancing between the power generation and the load demand, i.e. voltage fluctuations, system stability, reliability, reactive power compensation and frequency response [4]. Many approaches have been proposed to mitigate the aforementioned problems: increased storage capabilities, increase the accuracy of resource and load forecasting, demand response, etc. [5][6]. These solutions, singularly or combined in the same system, contribute to integrating higher levels of variable *Renewable Energy Sources* (RES). Among the listed solutions, power generation forecasting is becoming an indispensable tool for the energy network. A study made by H. Beltran et al. [7] proposed an optimized energy management strategy for photovoltaic power plants with energy storage based on the estimation of the daily solar energy. The authors demonstrate the possibility to reduce storage capacity ratings without excessive production deviations, permitting a solar power plant to participate in the intra-day energy market without incurring penalties. Rodríguez et al. emphasize the benefits of source forecasting even for distributed power networks as microgrids [8]. Specifically, the forecasting of the energy produced allows better optimization and control of the elements involved (i.e. distributed generations based on RES, diesel generators, storage devices), reducing the losses and improving the system's stability. Hence, production forecasts constitute a good instrument to ensure a smooth power supply and real-time supply and demand balance, becoming even more essential in the absence of a local energy storage system. But the benefits of solar forecasting extend beyond grid operators: highly precise forecasts allow adapting to changing weather conditions while helping to minimize disruptions and the overall cost of operations, as they enable optimized hardware usage and control schemes.

Finally, the necessity of accurate solar forecasting states as a significant way towards the integration of larger shares of variable renewable energy sources, managing solar intermittency and ensuring power continuity.

1.   Forecast horizons and classification techniques

The forecast horizon is "the future time period for output forecasting or the time duration between the actual and effective time of prediction" [9]. Its temporal resolution can vary from a few minutes to hundreds of hours depending on the set goal. Nowadays, researchers generally refer to four horizon categories [10]: very-short term, short-term, medium-term and long-term forecasting as listed in Table 1.

	Forecast horizon	Applications
Very-short term forecasting	1 minute - 30 minutes	It facilitates electricity marketing, power smoothing processes, monitoring of real-time electricity dispatch and PV storage control.
Short-term forecasting	30 minutes - 6 hours	It manages load dispatch and plant operations.
Medium-term forecasting	6 hours - 1 day	It is useful for maintenance scheduling.
Long-term forecasting	1 day - 4 weeks	It is helpful for long-term power generation, transmission, distribution and for taking into account seasonal trends.

Table 1: Main forecast horizon categories

1.1.   State-of-the-art

As a general rule, the longer the forecast horizon is, the lower the accuracy is [3]. As regards PV, the incorrectness for high prediction time is attributed to the stochastic nature of clouds, which has a significant impact on solar radiation and becomes hardly predictable with time dilatation [11]. Therefore, most studies in literature move forward short time forecasts, combining or improving the previously mentioned consolidated techniques.

A. Catalina et al. made a comparison between three different models to nowcast hourly PV energy production for horizons up to six hours over Peninsular Spain and the islands of Majorca and Tenerife [12]. The first model uses a *Numerical Weather Prediction* (NWP) forecast with atmospheric equations, the second one adopts *satellite-based measures* plus clear sky estimates and the final one combines together the previous two techniques. Three years of data were collected and used by a machine learning technique, specifically the Gaussian Support Vector Regression model. The forecasting method was evaluated using Mean Absolute Error (MAE), whose value strongly varies with the forecast horizon and the zone of interest since the observation of a wider area smooths spikes and drops typical of solar radiation. The conclusions state the superiority of the third model, which outperforms the results achieved by the approaches which adopt numerical weather prediction or satellite images singularly.

A. Nespoli et al. developed a model to detect clouds and their interceptions with sunrays directed toward a specific geographic location. Satellite images from ten days of 2019 were analyzed, classifying each pixel into 19 categories of clouds. Such data and Sun coordinates (zenith and azimuth angles) were combined with meteorological data sampled by a weather station located on the roof of a building of the Politecnico di Milano to determine the clearness index. This parameter was chosen as an indicator to distinguish four classes of weather and it constitutes the object of the forecasting. This information represents the input data set, which was evaluated to make a classification forecast fifteen minutes in advance by three different machine learning techniques, among which Random Forest got the best accuracy of 84.2% [13].

Works based on satellite images guarantee good accuracy in prediction since they focus on large spatial and temporal resolutions, which smooth sudden solar radiation changes. However, their disadvantages are the difficulties to detect the Cloud Based Height (CBH) - a fundamental parameter to define cloud types and their thickness - and the high time requested to receive and process the information, which complicates the possibilities to make a very-short term forecast. As a result, the application of an *All Sky Imager* (ASI) was investigated in several studies. All-sky images reduce process time and are widely spread to perform cloud classification.

A. Kazantzidis et al. used a digital camera with a fish-eye lens and a hemispheric dome for automatic detection of the clouds. The percentage cloud coverage was determined by applying thresholds to each pixel of the images based on red, green and blue (RGB) colors. Cloudy pixels were then classified into seven categories by a k-

Nearest-Neighbor method, which achieved an average accuracy of 87.9% [14]. This methodology represents the first step for nowcasting based on ASI.

T. Schmidt et al. applied a cloud detection method based on color thresholding to classify seven different categories of sky conditions with the aim of developing a multi-step model to forecast the Global Horizontal Irradiance (GHI) up to 25 minutes ahead. Historical data of solar irradiance were provided by pyranometers and CBH by a ceilometer. The Sun position, along with clear sky irradiance, was computed by a theoretical model for ray tracing and subsequent cloud shadow mapping. Cloud movement was determined by applying an optical flow algorithm. The results vary according to forecast horizon and sky conditions, revealing a Root-Mean-Square Error (RMSE) which ranges from $70W/m^2$ at clear sky to a maximum of $250W/m^2$ in presence of cumulus clouds. Despite the numerous model steps made, the sky imager forecast was not able to outperform the *persistence method* under all cloud conditions. [15].

To date, physical methods are limited by their accuracy in predictions, but a specific type of deep learning model, the *Convolutional Neural Network* (CNN), outperforms traditional methodologies. Its advantage is the ability to process images by simulating animal visual cortex, automatically and adaptively learning spatial features of the pixels.

C. Shi et al. exploited CNN models for ground-based cloud classification. Two public data sets were used to train and validate the ANN. The first one is characterized by five different typologies of sky, the second one is divided into seven categories of clouds. Two different configurations for the CNN were defined to compare a fast and simple model with a more complex one. Both outperform conventional handcrafted features. The faster model got a best accuracy of 98.20% and 89.90% respectively for the first and second data sets. The more complex one had a best accuracy percentage, respectively, of 98.20% and 90.32% [16]. A similar study was made by J. Zhang et al., who built a data set containing 11 categories of clouds, including contrails - a type of cloud generated by airplanes - for the first time. The CNN takes as inputs a series of fixed-size red-green-blue cloud images and provides a sequence of label predictions as output. The network is optimized by subtracting the mean RGB value of each pixel to augment the contrast and to improve training speed and accuracy. The results show good accuracy for such specific classification, ranging from 86% to 90% for all the categories [17].

Logothetis et al. made a benchmarking of solar irradiance nowcast performance up to twenty minutes ahead derived from all-sky imagers. Five different methodologies were evaluated and two of them can be compared to assess the impact on cloud classification by means of a CNN. In fact, the first one exploits physical and empirical techniques to detect CBH, cloud coverage and motion from sky images whereas the other one repeats the same processing steps, but combining a CNN to detect the CBH. The results show that both models outperform the persistence method, but the ensemble model is always more accurate than the physical one for each of the five different sky conditions considered. The RMSE obtained by the model which makes use of the CNN ranges from 18.1 W/m^2 under clear sky to 151.1 W/m^2 for cloudy conditions [18].

In solar forecasting convolutional neural networks are becoming a promising technology for cloud classification, but their ability on extracting features from images makes them also relevant for directly predicting solar irradiance.

Y. Sun et al. used a CNN to correlate power output of two PV panels with simultaneous all-sky images. The Authors made a sensitivity analysis to get the most suitable CNN architecture for their specific study and the results demonstrated that images contain significant information for solar prediction. Despite meteorological data and dynamic features not being exploited by the neural network, the resulting normalized Root-Mean-Square Error (nRMSE) is equal to 28% [19].

D. Pothineni et al. presented an image-based approach to anticipate the irradiance state both five and ten minutes ahead of time by means of a CNN. The future irradiance state ("occluded", "clear") is defined considering a threshold of 80% on the clear sky irradiance computed using the Perez-Ineichen clear sky model [20]. Two large datasets of images taken in Italy and Switzerland were used to validate the model, getting an accuracy of 92.93% and 91.34% respectively. Those values demonstrate that the network is not influenced by local site-specific properties [21].

CNNs are not limited to single images, as they can also process sequences of images, to detect features both in spatial and temporal dimensions. They are called *3D-CNN* and a study made by X. Zhao et al. is based on this approach. Multiple consecutive all-sky images were analyzed by the convolutional neural network, which predicts the clearness sky index among five categories of sky conditions. Subsequently, the extracted features, along with historical irradiance measurements, were used by two forecasting models (linear autoregressive and nonlinear multilayer perceptron) to predict GHI 10-30 minutes ahead. The results showed a mean accuracy in the clearness index classification of 67.0% and a best value of nRMSE equal to 30.82% in the regression computed with the nonlinear multilayer perceptron model [22].

In the last years, a specific ASI that works in the infrared spectrum, the *Thermal Infrared All Sky Imager*, is becoming popular in the forecasting context due to its capabilities in providing cloud properties and atmospheric data [23] [24]. Furthermore, *InfraRed* (IR) images show a better contrast between clouds and clear sky. In fact, visible cameras display a significant Sun flare whereas the Sun disk affects very low infrared photos [25].

A study by Niccolai and Nespoli makes use of the advantages provided by the thermal infrared ASI to detect

the Sun position, which is a useful parameter for solar prediction algorithms. The researchers compared three techniques based on theoretical solar angle, image processing and neural network. The solar angle method revealed problems in accuracy, specifically it got a wrong identification of 82.9%. An identification was evaluated incorrectly whether the identified position of the Sun was more than three pixels away from the manually detected one. Image processing technique instead was not always able to detect the Sun due to cloud coverage, halving the available number of sample with respect to the other strategies. Finally, neural network overcame all drawbacks of the previous techniques. It memorizes and learns from historical data, making possible the processing of all images, even in presence of clouds. Furthermore, it eliminates the deformation errors introduced by the camera, getting almost a perfect accuracy (wrong identification of 0.01%) [26].

An application of the thermal infrared ASI in forecasting field is present in the work of Liandrat et al. The study aims to optimize a hybrid PV-Diesel system and evaluate the effects of 10 minutes solar prediction compared to a scenario without forecast. A fixed load of 3MW, a nominal PV power plant of 1MW and a series of 15 diesel generators with a power of 0.2MW each were considered. The features of the images are extracted via two steps: pixels thresholding to define cloud/clear sky and optical flow algorithm to detect cloud motion. Four months of data derived from images and irradiance measurements are used by a random forest ANN for forecasting. The authors demonstrated the benefits of solar prediction with respect to the "no-forecast" scenario. In fact, the load balance error (difference between load and production) decreases for an equal number of blackouts in the two cases, resulting in lower fuel consumption up to 17% [27].

1.2. Forecasting techniques

The bibliographic review resumes some of the main methodologies used to forecast solar radiation. Going more into detail, predictive methods can be classified into four different groups: *naïve methods*, *physical methods*, *statistical techniques* and *ensemble methods* [3][4]. Figure 2 illustrates the previous-cited techniques, which are briefly discussed in the next paragraphs to provide a clearer overview of each of them.
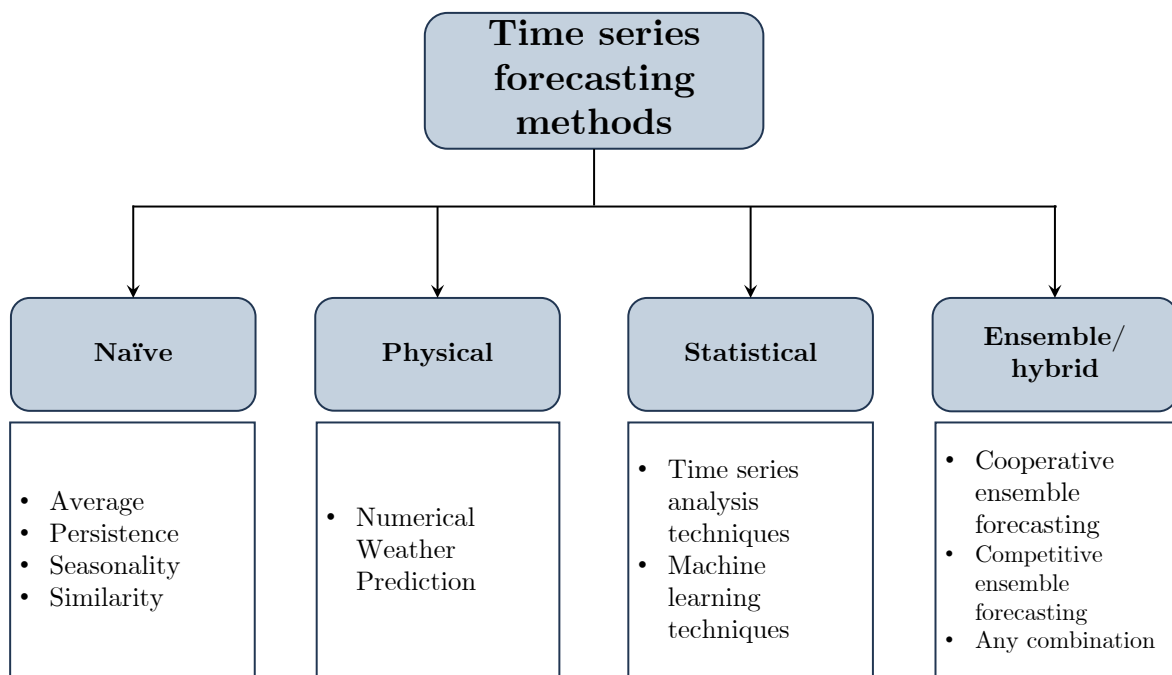

Figure 2: Time series forecasting methods

Naïve methods

Naïve methods are often used as benchmark methods due to their simplicity, which leads both to low computational and time costs. Furthermore, they are surprisingly effective and considered sufficiently accurate for many applications. Among them, average, persistence, seasonality and similarity methods are the most popular.

Average method The *average method* is based on the concept of predicting all future values equal to the average of the historical data [28]. Considering the notation for historical data as $y_1, \ldots, y_t$, the prediction is:

$$\hat{y}_{t+h|t} = \overline{y} = (y_1 + \cdots + y_t)/t \tag{1}$$

where t is the instant of time and h is the forecast horizon.

Persistence method *Persistence* is used to benchmark more complex forecasting techniques. In particular, it is useful as a baseline model for understanding whether other methods can outperform it.
Persistence forecast is based on the concept that today equals tomorrow, meaning that every forecast is expected to be identical to the current conditions [28]. Formally, it is written as:

$$\hat{y}_{t+h|t} = y_t \tag{2}$$

Prediction based on the last recorded observation yields a fair accuracy in the short-term with respect to the uncertainty of weather forecasts [29], but on the other hand a not negligible drawback is implicit. In fact, it is not able to take into account the weather variability, such as clouds' motion, rainfalls, etc. at the forecast window.

Seasonality A similar method, the *seasonal naïve method*, is useful for highly seasonal data [28]. Each forecast is set to be equal to the last observed value from the same period of the interval considered. Mathematically, it results:

$$\hat{y}_{t+h|t} = y_{t+h-m(k+1)} \tag{3}$$

where m is the seasonal period and k is the integer part of the ratio *(h-1)/m*, which represents the number of times a complete time-interval (day, week, month, year, etc.) is passed in the forecast window prior the instant $t+h$.

Similarity The *similarity method* is based on the investigation of the past to forecast the future. Patterns of historical data which match the current conditions are exploited for forecasting since they can be more informative than the entire past behaviour of the variables of interest [30]. In particular, analogy among several boundary conditions in the historical data (i.e. irradiance, temperature, wind velocity) can be analysed to assess the target by a *similarity-weighted averaging* [31] [32].
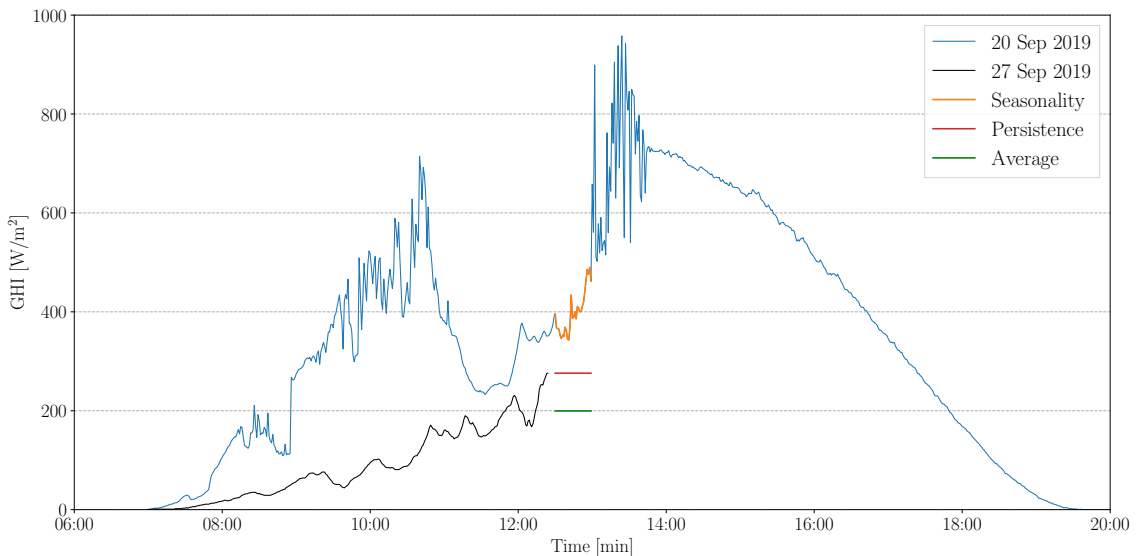

Figure 3: Examples of naïve methods with forecast horizons from 5 to 35 minutes

Figure 3 shows the forecasts of the naïve methods, with the exception of the similarity one due to its higher complexity. At 12:25 on Monday 27 September 2019, the forecast methods make solar radiation predictions from 12:30 for the next half hour. So, forecast horizons range from 5 to 35 minutes.
Specifically, the average forecast method considers the average of the last hour; the naïve method hypothesizes the last value measured; the seasonal naïve predicts the same trend as the previous Monday.
Generally, *naïve methods* are useful for very short time horizon predictions because their accuracy strongly decreases with the dilatation of the forecasting times, becoming unreliable for predictions higher than one hour.

Physical methods

Physical methods study the interactions between solar radiations and the atmosphere. Depending on the object of the study, they can be retrieved through image-based techniques or with information on the atmospheric properties achieved from measurements made by local weather stations.

Numerical Weather Prediction (NWP)   *NWP* models generally combine local meteorological data with atmosphere model equations to determine future states and make predictions for long-term forecasting. In particular, it is highly adopted for prediction up to 15 days ahead. Commonly, NWP tools are categorized in *global* and *mesoscale* in relation to their spatial scale. Global technique acts on a worldwide scale whereas mesoscale is limited to a specific region of interest (i.e. countries or continents). Given the huge areas involved, these models are characterized by a coarse resolution which hinders accurate computation for small spatial scales. The solution for investigating shorter spatial and temporal scales through this method is the use of images, which are differentiated between *ground-based* and *satellite*.
Ground-based systems, also known as All Sky Imagers (ASI), consist of a digital camera that provides a hemi-spherical image of the sky. They are usually applied for cloud classification, cloud-based height measurements and cloud motion determination. They are used mainly to provide information for very-short term forecasting. The focus on a such high temporal resolution is twofold: field of view and clouds transformation limit the skill of this method for wider horizons, while it also fills a gap among other data sources which do not have the capability to forecast for a time lower than few minutes [33].
Satellite images are used instead for the classification of clouds and the definition of their patterns but, unlike ground-based systems, they are appropriate for a broader forecast horizon due to larger observed areas and slower update time, which lead them to be more appropriate for short term forecasting.

Statistical approaches

Statistical approaches are used to reconstruct the relations between meteorological time series and real-time measured data. The more accurate and larger the data set, the more reliable the results are. Two main methods belong to this category: *linear models or time series-based techniques* and *non-linear models or machine learning techniques*.

Time series analysis techniques   Time series provides statistical information to depict linear relation-ships among variables [34]. Data are recorded and collected over time on a regular basis, such as minutely, hourly, daily or monthly. Their advantage is a lower computational burden than that of non-linear models.

Machine learning techniques   Learning methods are based on the ability of artificial intelligence to learn from historical data and to further hone its predictive abilities via training runs. An *Artificial Neural Network* (ANN) mimics the information processing mechanism of the human brain and it is able to approximate nonlinear functions with high fidelity and accuracy because of important properties as self-learning, adaptivity and fault tolerance [35]. Its high computational time is justified by better forecasting in the short term [36].
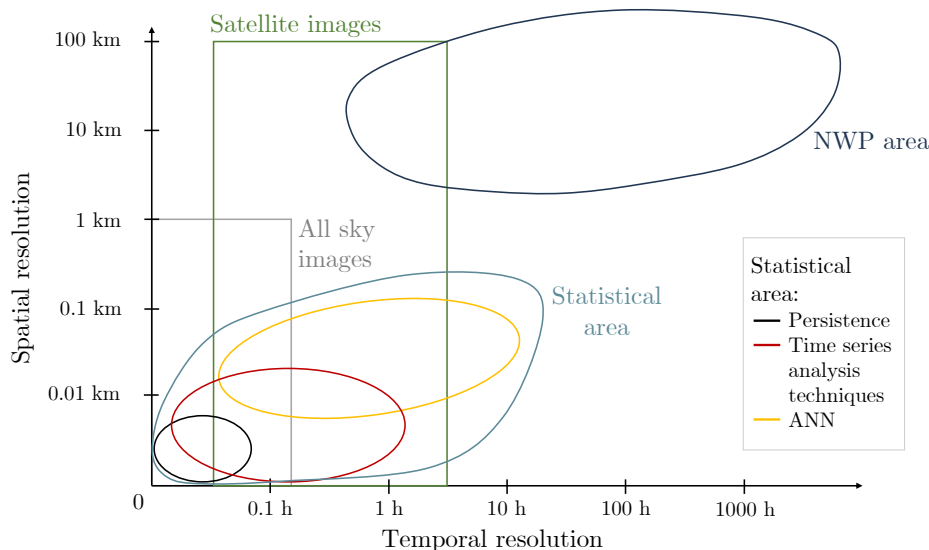

Figure 4: Model classification based on temporal and spatial resolutions

Ensemble\hybrid methods

Ensemble\hybrid methods combine individual methods with the aim to reduce their weaknesses and drawbacks, improving the robustness and accuracy of forecasting at the cost of a higher computational burden [37].
Any combination of single methods which operate in different ways from the cooperative or competitive methods is considered part of the hybrid methods.

Cooperative ensemble forecasting   Cooperative ensemble forecasting is a technique which operates by dividing the final task into a few sub-tasks. Thus, smaller problems are solved individually and the forecast is based on the aggregation of the decisions taken by each predictor.

Competitive ensemble forecasting   Competitive ensemble forecasting exploits different models in which each predictor makes the forecast independently. Diversity in forecasting techniques, data sets, or parameters is the strength of this method since their results are collected and averaged to determine the final decision.

2.   Motivations and research questions

The previous sections highlight the need for accurate solar forecasting and the numerous benefits which it could bring. Among the different forecast horizons, very-short term, also known as *nowcasting*, plays a crucial and essential role to facilitate the access of variable RES into electrical networks. The aim of this thesis is to provide an innovative forecasting technique for the immediate future based on the promising technologies that are developing nowadays: the Convolutional Neural Networks and the Thermal Infrared All Sky Imager.

This goal is achieved by proposing a methodology which aims to solve two typical drawbacks related to machine learning and solar forecasting: the high computational burden and the inaccuracy of traditional methods with respect to the naïve methods, in particular persistence. The former derives from the huge amount of data needed by Artificial Neural Networks for processing and training, which is further incremented in the case of CNN where information is extracted from images. The latter is caused by the good accuracy of the persistence method for such short timescales. In fact, considering the daylight hours, in a few instants strong variations in solar irradiation could be observed. This means that statistically the solar radiation changes gradually, resulting in a small overall persistence error on a minutely time basis.

With that said, a new analysis linked to the performance of the forecasting methods is introduced for the first time. Specifically, an evaluation of the errors based on the specific instants in which consistent variations of solar irradiance occur was performed. Those moments are indeed crucial for a plant owner who needs to manage power production in the near future.

Lastly, a new methodology to provide exogenous data to the CNN without changing its architecture is proposed. In particular, the level of solar irradiance is passed to the neural network directly from the image, modifying those pixels that are not part of the sky. Thus, a specific process was developed to codify solar radiation in pixels language.

To summarize, this work aims to answer the following questions:
- **How much solar forecasting based on CNN and ICI is accurate?**
    - Which are the main drawbacks of the technique propososed?
    - How does the CNN perform in those moments of high irradiance fluctuations?

3. Framework

Machine Learning (ML) is a subfield of computer science and it is classified as an Artificial Intelligence (AI) method (Fig. 5). It is the science of programming computers so that they can *learn from data*.

In this section, the main machine learning techniques are described, as well as their applications and advantages. Then, the evolution of artificial intelligence and how it works is introduced. Lastly, the most adopted technique to train AI is described and a deeper focus on convolutional neural network is proposed.

Artificial Intelligence
A program that can sense, reason, act and adapt.

Machine Learning
Algorithm whose performance improve as they are exposed to more data over time.

Deep Learning
Subset of machine learning in which multi-layered neural networks learn from vast amount of data.
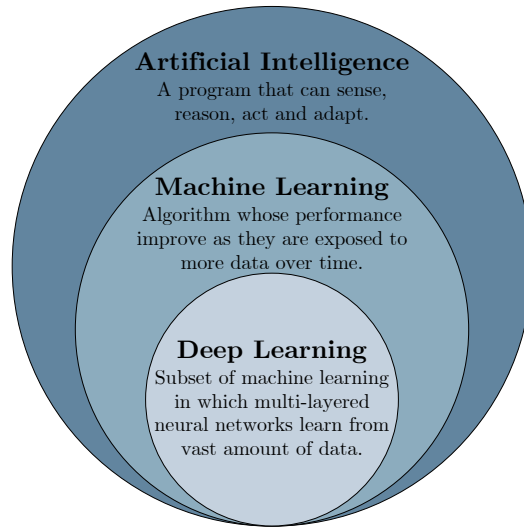
Figure 5: Deep learning family

3.1. Machine learning

Machine learning systems can be classified in different ways, but generally, they are differentiated based on whether or not they are trained with human supervision. Hence, the following categories are defined:

- Supervised Learning;
- Unsupervised Learning;
- Semi-supervised Learning;
- Reinforcement Learning (RL).

A brief description of their characteristics is summarized in Table 2.

Machine learning is adopted in many applications and represents a new frontier in *data analysis* due to its efficiency in terms of results and time with respect to the human brain. More in detail, it can be a solution to the following problems:

- Situations for which existing solutions require a lot of fine-tuning or long lists of rules;
- Cases where the size of the problem is extremely large and exceeds human reasoning abilities;
- Problems with a high level of complexity for which traditional approaches do not yield any good solutions;
- Cases characterized by fluctuating environments, where data updates over time.

These advantages are then reinforced by *Artificial Neural Networks*, which have become the pivot of Deep Learning: the deepest part of AI. Intelligent machines with multi-layered neural networks add the following strengths to the potential of machine learning:

- **Robustness**: the optimized features are recognized in an automated fashion related to the task under study;
- **Generalization**: Different data types or different applications can use the same deep learning technique;
- **Scalability**: the architecture, with layers and nodes, can be implemented in accordance with the specific application.

3.1.1 Arificial neurons, the perceptron and the multilayer perceptron

The first artificial model was proposed by McCulloch and Pitts to reproduce a biological neuron. It was based on binary (on/off) signals, in which the *artificial neuron* takes one or more inputs and provides a single output. In particular, the neuron activates only when determined logical conditions (i.e. not, and, or) on the inputs are met. It represented the pillar for ANN since even with such a simple structure, it is possible to create any desired logical proposition.
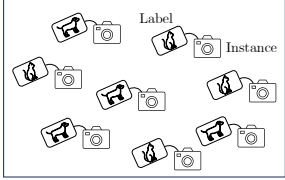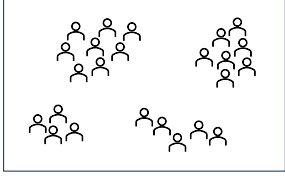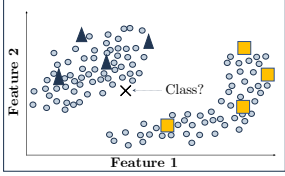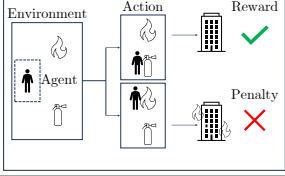
| Task	Description	Graphical representation
Supervised Learning	In *supervised learning* each data point contains features and an associated label. The goal of the algorithms is to learn a function that maps feature vectors (inputs) to labels (output) [38].	
Unsupervised learning	In *unsupervised learning* the training data is unlabeled, forcing the machine to learn and discover correlations and trends by itself.	
Semi-supervised Learning	It combines a small amount of labeled data with a large amount of unlabeled data during training. This approach alleviates the burden of obtaining hand-labeled data sets, which can be costly or impractical.	
Reinforcement Learning	*RL* concerned with how an *agent* ought to take actions in an environment in order to maximize its *rewards*. The agent gets a *reward* or a *penalty* if it did a good or bad action.	

Table 2: Machine learning techniques
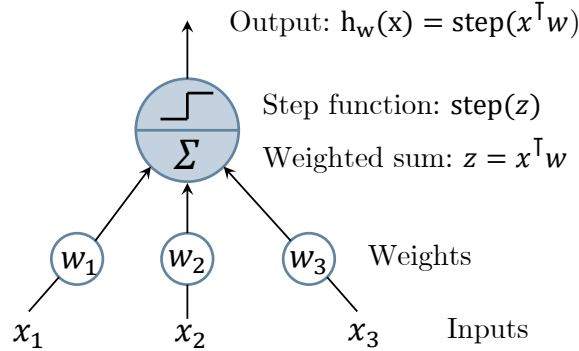
Figure 6: The Threshold Logic Unit. Image reproduced from [39].

In 1957 Frank Rosenblatt invented one of the simplest ANN architectures, called *perceptron*. It is based on the *Threshold Logic Unit (TLU)*, a slightly different type of artificial neuron with respect to the one devised by McCulloch and Pitts because its inputs and outputs are not impulse signals anymore, but numbers; and each input connection is associated with a weight. The TLU computes the weighted sum between its inputs and their associated weights ($z = w_1x_1 + w_2x_2 + \cdots + w_nx_n = x^\intercal w$), then it applies a *step function* to determine the output.

The most common step function adopted by the TLU is the *Heaviside step function*.

$$H(z) = \begin{cases} 0, & \text{if } z < 0 \\ 1, & \text{if } z \geq 0 \end{cases} \tag{4}$$

The architecture of the perceptron is constituted by a single layer of TLUs, connected to all the inputs. Hence, it is a *fully connected layer*, also called *dense layer*, since it is linked to all neurons of the previous layer. Inputs are fed in the *input layer*, which is built with particular *input neurons* that output the input without changing. Furthermore, a special type of neuron, the *bias neuron*, is added. It always outputs 1 and it works as an extra bias feature.
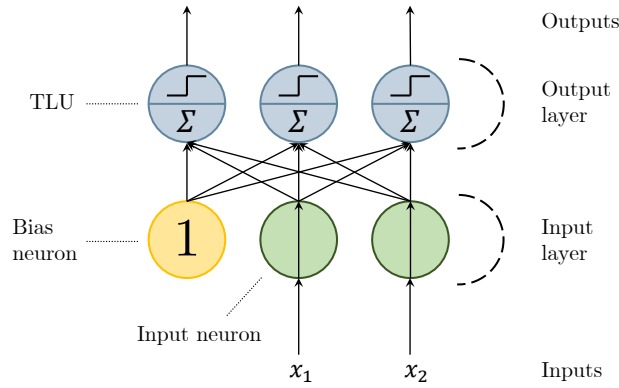
Figure 7: The architecture of perceptron. Image reproduced from [39].

The output of a dense layer of an AI is mathematically defined in Equation 5.

$$h_{W,b}(X) = \phi(XW + b) \tag{5}$$

Where:
- **X** represents the matrix of inputs.
- **W** stands for the weight matrix, which contains all the connection weights except for the bias neuron.
- **b** is the bias vector, which memorizes the connection weights between the bias neuron and the artificial neurons.
- ϕ is defined as the *activation function* (for TLUs it is a step function).

The evolution of the perceptron is the Multilayer Perceptron (MLP). An MLP recovers indeed the same structure of the perceptron but adds one or more layers of TLUs, called *hidden layers*, between the input and the output layers. It is the most similar concept to today's ANN.
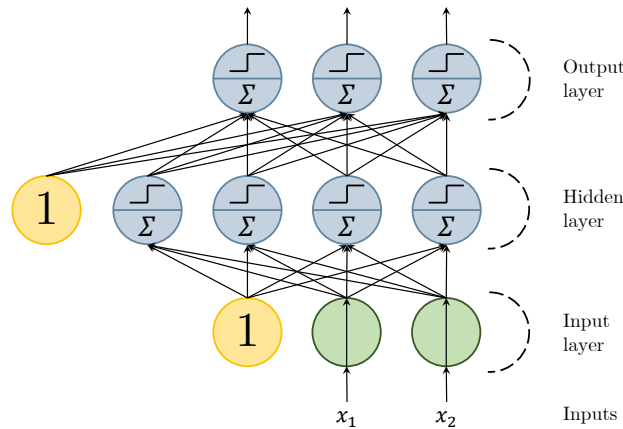

Figure 8: Example of a Multilayer Perceptron architecture

3.1.2 The Backpropagation

The MPLs have trained following the *backpropagation* algorithm. This technique is able to compute the gradient of the network's error in just two steps: one forward and one backward. Going into more detail, it first makes a prediction (forward pass) and measures the error, then it goes through each layer in reverse to find out how each connection weight and each bias term affect the error. Once it has these gradients, it performs a regular Gradient Descent step, and the whole process is repeated until the network converges to the solution.

This algorithm needs a key change to the MLP's architecture to work properly. The step function is indeed constituted by flat segments, so there is no gradient to work with. The solution was to modify the step function with other logistic functions characterized by a nonzero derivative (especially for positive values), allowing Gradient Descent to make some progress at every step.
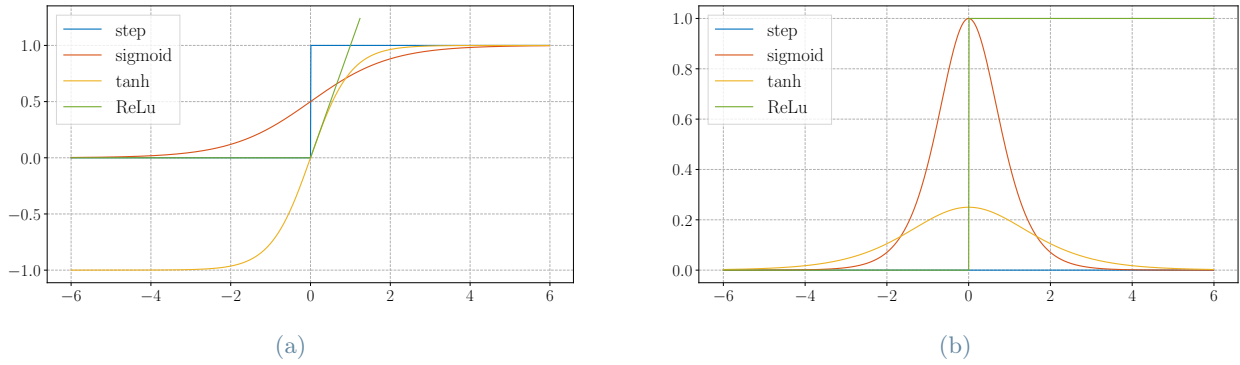
Figure 9: Activation functions (a) and their derivatives (b)

3.2. Convolutional neural networks

To understand how a CNN operates and how its structure is conceived, it is necessary to look at the functioning of the brain's visual cortex. Here, each neuron focuses on local fields, meaning that a single neuron sees a limited part of the whole picture. The "*local receptive fields*" can overlap between the neurons and together provide the entire visual field. Another characteristic of the visual cortex is the different capability of each neuron to react to specific features (i.e. horizontal or vertical lines), which gives the brain the ability to detect all kinds of patterns in any area of the visual field.

These analyses lead to the development of the *convolutional neural network* of the present day.

3.2.1   Overall architecture

The CNNs are built with three types of layers: *convolutional layers*, *pooling layers* and *fully connected layers* (already described in 3.1.1).

The basic functionality of a CNN can be divided into four key areas:

1. The **input layer** manages the data provided to the CNN.
2. The **convolutional layer** computes the output of a neuron through the scalar product between its weight and the data contained in the area of the input under its receptive field.
3. The **pooling layer** filters the information reducing the number of parameters through the downsampling along the spatial dimension of the given input.
4. The **fully-connected layer** attempts to learn a non-linear function and produce class scores for classification and regression purposes.

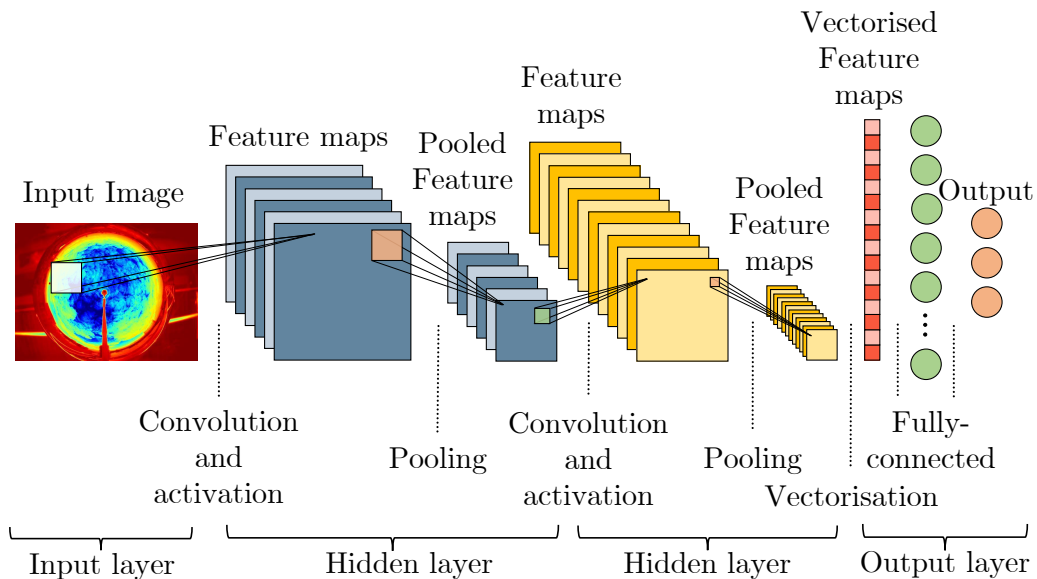Figure 10 shows the main layers and elements involved in a CNN.

Figure 10: General architecture of a convolutional neural network

3.2.2 Inputs

The most common inputs for convolutional neural networks are images. Although they appear similar to our eyes, there are different types of images which contain different amounts of information. Going into more detail, they are characterized by a spatial dimension, which is composed of the *height*, *width* and *depth* (or *channels*). The first one represents the number of pixel rows in the image (or the number of pixels in each column of the image); the second one stands for the number of pixel columns in the picture (or the number of pixels in each row of the picture) and the third one is the number of components used to represent each pixel.

Depending on the value of the depth, several categories of images are distinguished. Most images belong to the *grayscale* or *RGB* (which stands for Red, Green and Blue colours) types. A grey image is composed of just one channel with pixels in the greyscale, which go from black to white shades; the RGB images instead are composed of three channels, one per *colour channel*.

Their level of data is represented by the *bit depth*, which in most of the cases is equal to 8. It means that an 8-bit grayscale image holds 256 (2^8) colour values (since a bit can assume a value of 0 or 1), whereas an RGB picture has 8 bits for each group of colours, so it offers 16.7 million (2^{24}) colour variations.

Considering a single-channel image, its information stored in a computer is represented in Figure 11, where each pixel is linked to a value of intensity which goes from 0 to 255 (exactly 256 total shades).
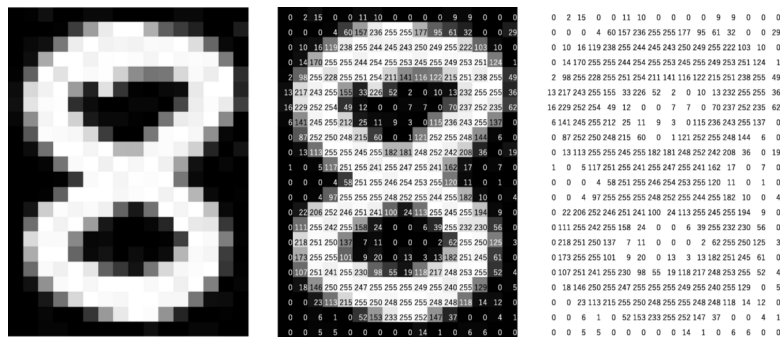

Figure 11: Images showing pixels and their associated intensity values

3.2.3 Convolutional layers

A convolutional layer is the main building block of a CNN [40]. Here, neurons are not connected to every single pixel of the input image (differently from the dense layer), but only to pixels in their local receptive fields. Usually, convolutional layers are set following a hierarchical structure: the higher layer is connected only to neurons of the lower layer which belongs to its local receptive field (see Fig. 12). It allows the network to capture the low-level features (e.g., lines) of images in the initial hidden layer, then assemble them and extract the high-level features (e.g., shapes, specific objects) in the later layers [41].
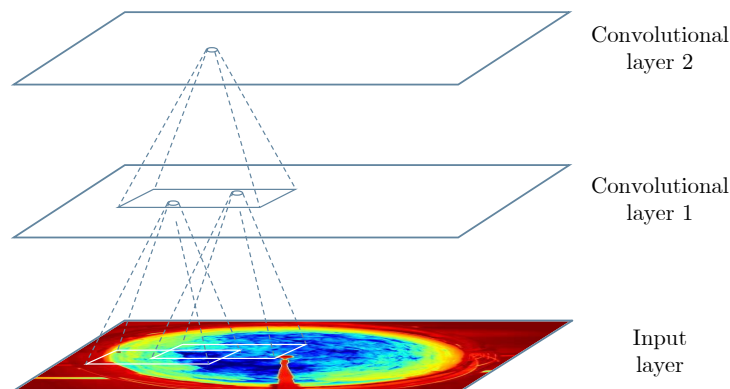

Figure 12: Hierarchical structure of convolutional layers

The neurons between one hidden layer and another can interact in different ways and provide different outputs according to the settings of two terms: the *padding* and the *stride*. The former refers to the number of pixels added to an image. Its main application is the so-called *zero padding*, which consists in adding zeros around the inputs to allow the next layer to have the same height and width as the previous layer: see figure 13.
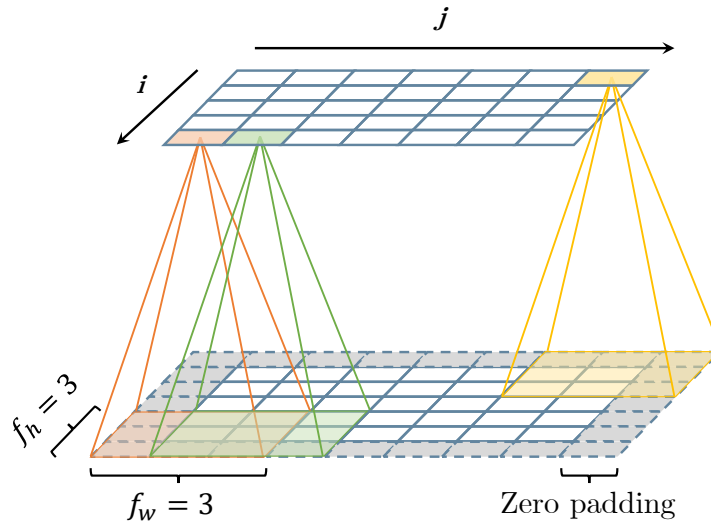
Figure 13: Zero padding: f_h and f_w are the height and width of the neuron's receptive field respectively

The latter is a parameter that refers to the amount of movement of the neural network's filter. It is fundamental, first of all, because it determines the region under each neuron's inspection, therefore whether or not their receptive fields overlap. The lower the stride, the higher the observed levels of detail. Despite this benefit, a higher computational complexity is implicit, since the dimension of the next layer is not reduced, as in the case of a higher number of stride.
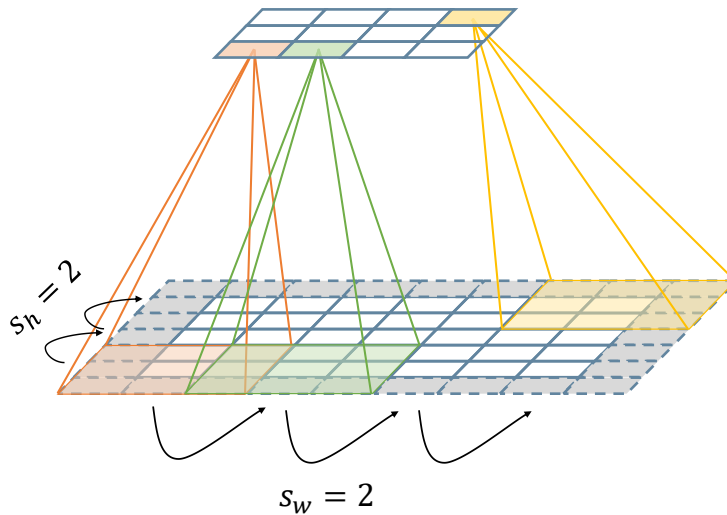

Figure 14: Illustration of the stride. The s_h and the s_w are the movements made in the height and the width direction of the receptive field

In the convolution process, the receptive field of each neuron works as a filter which is applied to the input so that it results *activated*. The parameters of the filter, also known as *kernel*, are the *weights* that the ANN learns throughout the training to detect features.

In case all the neurons in a layer use the same filter, it results in the creation of a map of activations called *feature map*, which indicates the locations and strength of a detected feature.

The activation is made through the dot product between every element of the filter and the input. Figure 15 shows an example of the convolution process. The first entry of the activation map (marked in green in Figure 15) is calculated by convolving the filter with the portion marked blue in the input image. The activation map is generated by repeating this process for every element of the input image.

Since a convolutional layer is characterized by many filters (more specifically the number is decided by who designs the CNN) and outputs a feature map for each filter, the result is a volume generated by stacking the activation maps of every filter along the depth dimension. In other words, every pixel in each feature map is associated to a neuron, and all neurons within a given feature map share the same parameters (which are
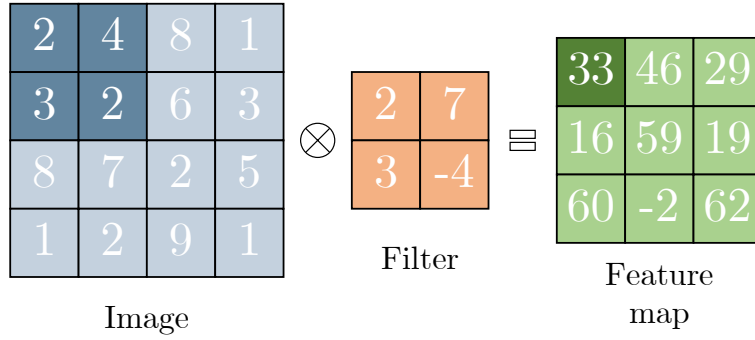
different among the feature maps).

Figure 15: Convolution process

Now, it is possible to mathematically define the computation made by a neuron to generate its output:

$$z_{i,j,k} = b_k + \sum_{u=0}^{f_h-1} \sum_{v=0}^{f_w-1} \sum_{k'=0}^{f_{n'}-1} x_{i',j',k'} \cdot w_{u,v,k',k} \qquad \text{with} \begin{cases} i' = i \times s_h + u \\ j' = j \times s_w + v \end{cases} \tag{6}$$

where:
- $z_{i,\,j,\,k}$ is the output of the neuron located in row i, column j in feature map k of the convolutional layer l;
- s_w and s_h are the horizontal and vertical strides, f_w and f_h are the width and height of the receptive field, and $f_{n'}$ is the number of feature maps in the previous layer (l-1);
- $x_{i',\,j',\,k'}$ is the output of the neuron located in layer $l-1$;
- b_k is the bias term for feature map k in layer l;
- $w_{u,\,v,\,k',\,k}$ is the connection weight between any neuron in feature map k of the layer l and its input located at row u, column v (relative to the neuron's receptive field), and feature map k'.

3.2.4 Pooling layers

The goal of pooling layers is to gradually reduce the dimensionality of the representation, thus reducing the number of parameters, the memory usage and the computational burden [42].

The operation of neurons in a pooling layer is similar to the one described for convolutional layers: their size, stride and padding have to be set in the same way. However, a pooling neuron has no weights because its task is to unify the input through an aggregation function, such as the maximum or mean.

Figure 16 shows the result of the pooling operation for the max and mean respectively. The example characterized by a kernel size of 2×2 and a stride of 2 determines a 25% reduction of the original activation map, highlighting the efficiency in dropping the parameters in excess.
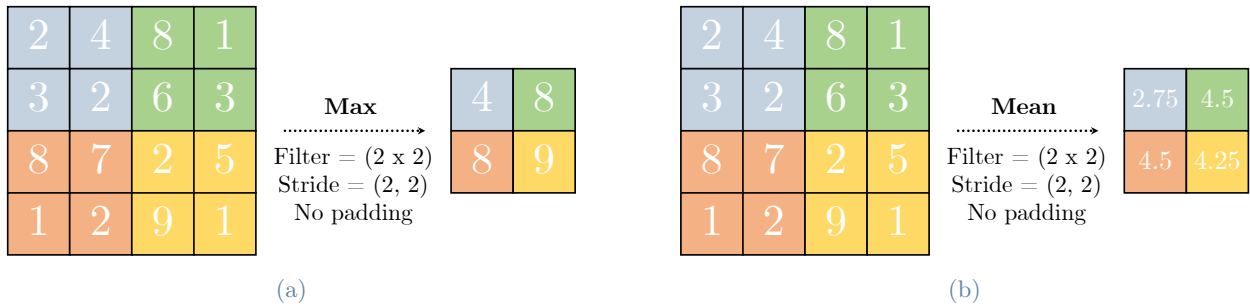

Figure 16: Comparison between the max-pooling (a) and average pooling (b)

4. Proposed methodology

Before providing a specific description of the adopted method, the definition of an important forecasting characteristic is now introduced to better understand the work of this thesis. Apart from the forecast horizon, another time-related parameter is identified to characterize the temporal nature of a forecast: the *forecast update rate*, which indicates how often forecasts are generated.

This research develops a deep learning technique based on a CNN which concatenates a certain number of consecutive images to create a sequence; in which three forecast horizons (5, 10 and 15 minutes) are analysed with an update rate of 1 minute.

Figure 17 illustrates the proposed methodology, which will be presented step by step in the next paragraphs.
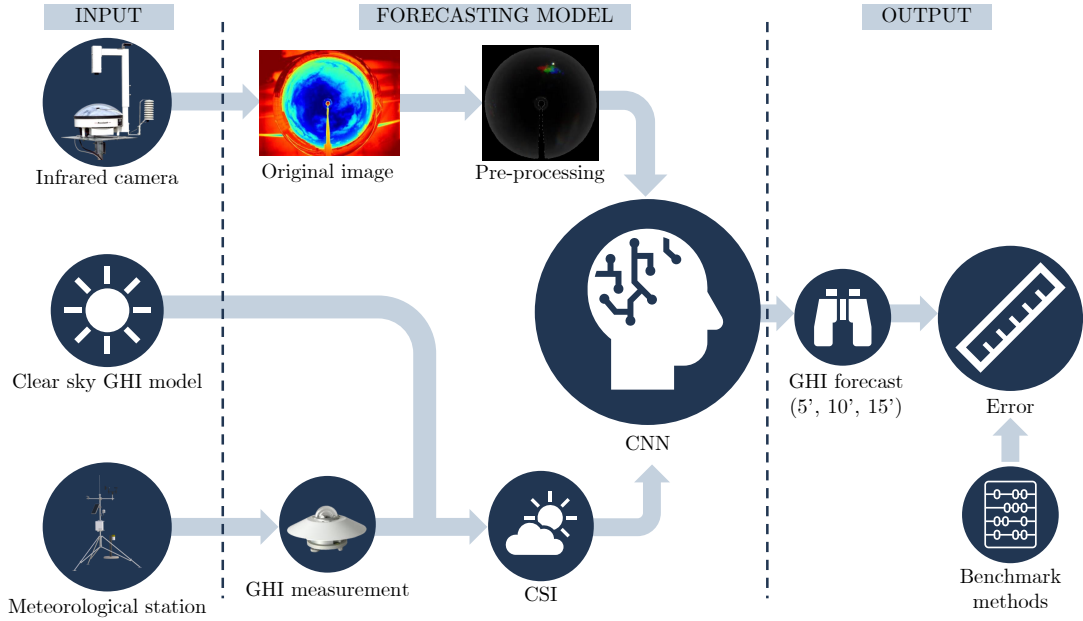

Figure 17: Scheme of the proposed methodology

4.1. Problem formulation

The proposed model can be represented as a non-linear mapping function that correlates the known inputs to the unknown output: the future global horizontal irradiance.

Hence, the problem is formulated as:

$$y_{\Delta t} = F_{\Delta t_{back}}(X, W) \tag{7}$$

where $y_{\Delta t} \in \mathbb{R}^{N \times 1}$ is the future GHI; $X \in \mathbb{R}^{N \times W \times H \times C}$ is the sequence of images; N, W, H, C are the sample number, the image width and the image height in terms of pixel, the number of channels of the image (i.e. red, green, blue), respectively; $\Delta t \in \{5, 10, 15\}$ is the forecast horizon in terms of minutes; W is the trainable parameters matrix of the CNN; $F_{\Delta t_{back}}(\cdot)$ is the proposed CNN model whose inputs change and are trained accordingly with the *historical timestep* (Δt_{back}). $\Delta t_{back} \in \{2, 4, 8\}$ represents the time interval in terms of minutes between the images used to build the sequence (see 4.3).

4.2. Pre-preprocessing

The available image data set was processed with the aim of obtaining better results both in terms of performance and execution time.

The first step was the *"cleaning phase"*, in which not exploitable images depicting rainfall are removed (Figure 18a). This process was made through two filtering modes: firstly, a visual check was carried out to identify the rainy days, then images taken at times when the rain gauge measures values higher than 0 mm/h were excluded from the dataset.

Precipitation is not the only phenomenon that can affect the forecast with infrared images. The thermal infrared all sky imager has a problem in case of low Sun elevation angle due to the increase of the atmospheric thickness (see 5.2.1) as depicted in Figure 18b. Moreover, at Sunrise and Sunset, the region of the sky around the Sun

is limited due to the presence of land. Hence, in this area it is particularly hard to detect clouds and their movements.

This is the reason why even the images and the time samples characterized by a Sun elevation angle lower than 20° were not considered by the forecasting model.

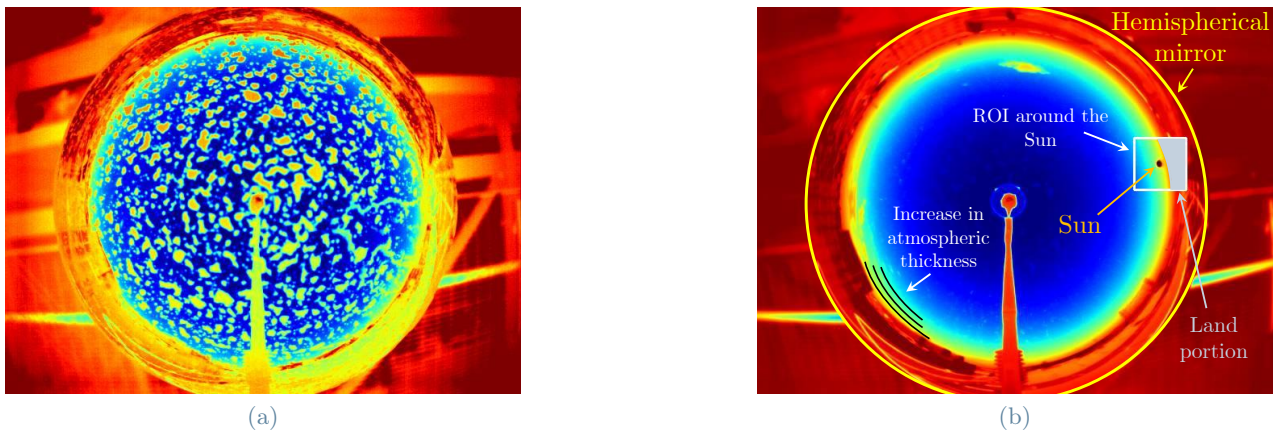<div align="center">(a)</div>

<div align="center">(b)</div>

Figure 18: a) shows the rain effect on infrared image. b) highlights both the colour change near the horizon due to the increase in atmospheric thickness and the limited portion of the sky in the Region Of Interest (ROI) around the Sun at a low Sun elevation angle.

The second step is the *conversion phase*. Here, the goal is to convert the images such that they pass from RGB format (three channels) to greyscale format (one channel). This operation aims to accelerate the training process since the amount of image data is reduced by two-thirds. In practice, it consists in making the inverse process of the one realized at the moment of the shot. In that case, a colormap was applied to a greyscale picture with a bit depth equal to 8, generating an RGB image with 256 shades.

Python libraries have several functions that convert images from RGB to grayscale through the computing of a weighted sum of the red, green and blue pixels. However, they do not ascertain that different RGB values can provide the same result. This problem was solved by using the *jet*[1] function implemented in MATLAB® [43]. This function returns a colormap from blue to red with the selected number of colors. Since the pictures taken by the camera have an 8 bits depth, the original shades in the images were 256. Usually, RGB pictures are 24-bit (8 bits per channel), however in this case the original image was in greyscale and during post-processing, the computer applied a colormap to allow the human eye to see the information more clearly.

The transform operation consists in comparing each RGB pixel value of the original image with the 256 colors generated by the jet function. The color of the jet function that has the least difference from the RGB pixel was selected, and its index represents the grey intensity value for the monochrome image (see Figure 19).
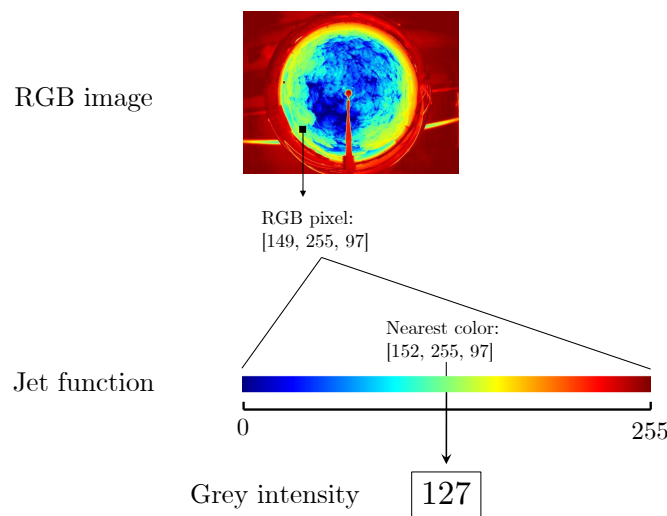

Figure 19: Illustration of the image transformation process from RGB to greyscale

---

[1]Matlab jet function: www.mathworks.com/jet. Last accessed: March 10, 2023.

The color of the RGB pixel in the image and the closest one generated by the jet function may not match perfectly. This is caused by the jpeg2000 format, which uses a number of algorithms to compress information as much as possible, but which could lead to the formation of different colors than the originals.

The second phase is concluded through the application of a mask to remove the part that is not object of interest. In particular, the region of the ground that surrounds the hemispherical mirror is darkened together with the buildings, the gasometer as well as the mast of the camera which all cover part of the sky. Furthermore, even the horizon where the air mass increases (region characterized by a color similar to one of the clouds at clear sky conditions) is obscured to avoid confusion during the training of the CNN.

The second phase steps are depicted in Figure 20.

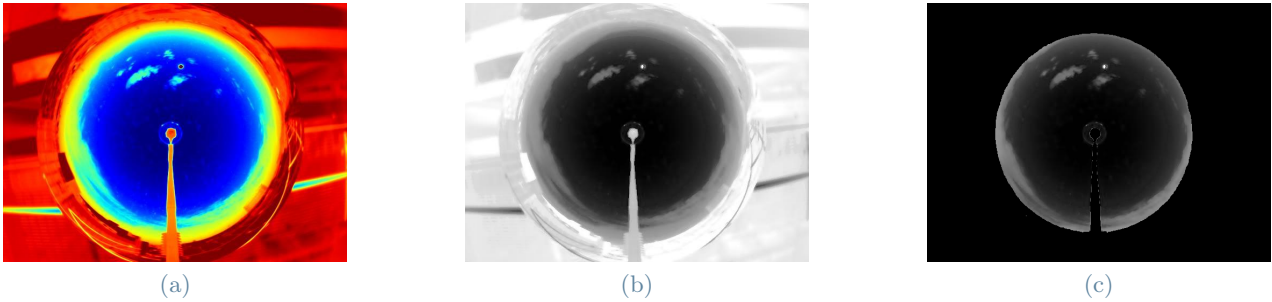(a)   (b)   (c)

Figure 20: a) shows the infrared image. b) is the equivalent of the RGB image in the greyscale. c) shows the mask application.

Once the background was removed and only the sky remains as the analyzable part of the image, the last pre-processing phase was carried out: the *cropping phase* (Figure 21).

Here, a square area of 384×384 pixels size was cropped from the original picture and *scaling* was performed to reduce the image size by two-thirds. In this way, a final image with 128×128 resolution was generated to reduce the data that has to be elaborated by the CNN.
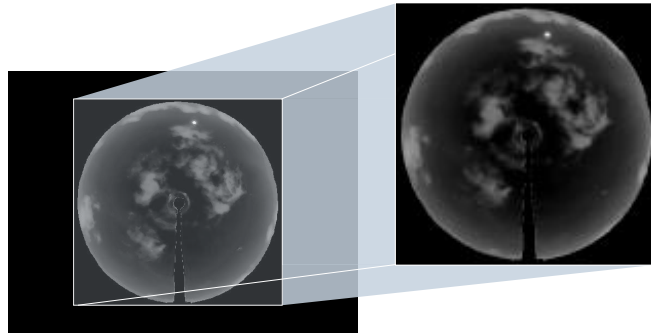

Figure 21: Cropped image

4.3.   Training

20/09/2019 13:20   20/09/2019 13:21   20/09/2019 13:22   20/09/2019 13:23   20/09/2019 13:24
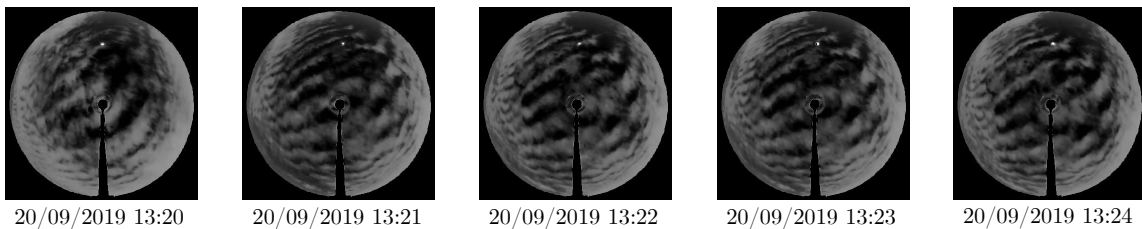
Figure 22: Dynamic motion of clouds in a 5-minute temporal span

Figure 22 shows a 5 minutes temporal window and highlights how the IR camera allows extracting clouds thermal dynamics and motion. To derive features from both temporal and spatial dimensions, consecutive IR images are fed into the CNN.

Image sequences are generated by concatenating them along the channel dimension. Hence, suppose that k images are concatenated as the input of the forecasting model, the input has the size of $H \times W \times (k \cdot C)$.

The nowcasting task is to forecast the ramp events in the near future. Thus, sequences with last images taken before the prediction were concatenated. This operation is formally described in:

$$I_{input} = (I_{t-(l-1)\Delta t_{back}} \oplus I_{t-(l-2)\Delta t_{back}} \oplus \cdots \oplus I_t) \tag{8}$$

In the Equation 8, the symbol $\oplus$ indicates the concatenation between images (I), t is the actual time, l is the sequence length and Δt_{back} is the time interval between images in the sequence.

In this study, a sequence of *three* images was evaluated since it constitutes a good compromise between a limited number of information provided by the concatenation of just two images and a higher computational cost derived from longer sequences.

The result can be visualized as an RGB image, in which each channel is constituted by a greyscale picture taken at a specific time. The clouds movement is extracted by the color shades generated by the images concatenation, which coincide with red, green and blue when clouds do not overlap as shown in Figure 23.
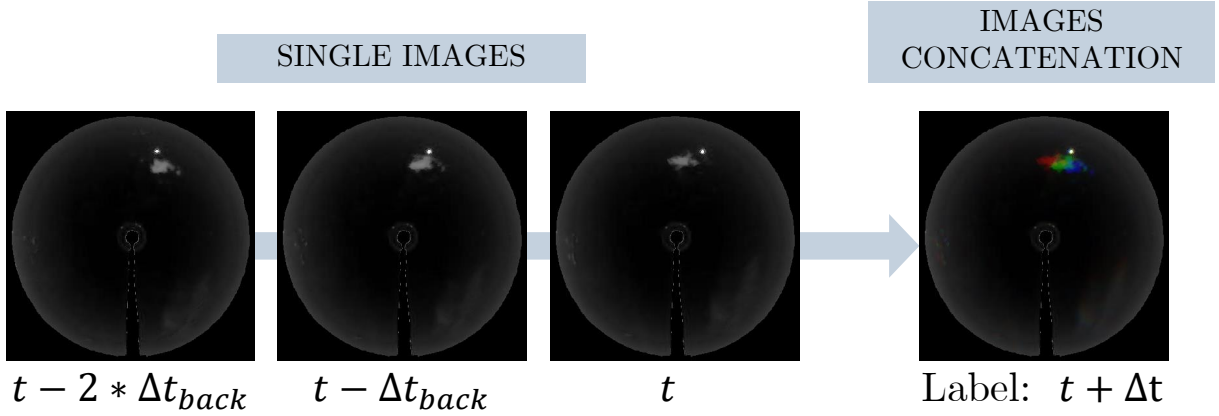

Figure 23: Three images concatenation

During training, the image sequences were passed multiple times through the CNN. Each iteration on the whole dataset corresponds to an *epoch*, at the end of which network parameters (the weights) are updated.

The effect of the training and validation error is function of the number of epochs, since after an initial decrease of both types of error, a minimum in the validation error is reached, whereas training error continues to decrease. The increase in the generalization error is related to the machine learning issue of overfitting and variance error for models with high expressive power [44]. For this reason, early stopping technique was used to stop the model training at the minimum of the error curve on the validation subset.

The maximum number of epochs was set to 70. Lastly, results were obtained evaluating the trained model on the test set.
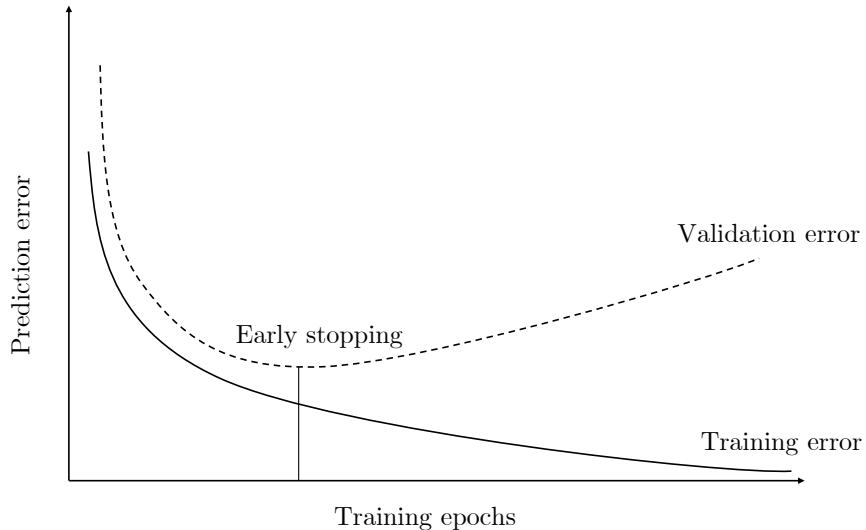

Figure 24: Learning curve

4.3.1 Clear sky model and clear sky index

The dynamic motion of the clouds extrapolated from the sequences of images is correlated with the global horizontal irradiance through the *labels*. They are associated with each image and constitute the *solution* for the CNN during the training phase. In particular, the model does not directly predict the GHI, but the GHI normalized by the *Clear Sky Global Horizontal Irradiance* (GHI_{cs}). This last term was computed through the Ineichen and Perez clear sky model which was simulated in the pvlib Python package [45]. The ratio between clear sky and actual GHI permits to operate with a limited target parameter between 0 and 1, except for particular conditions where it could be greater than 1 due to solar radiation multiple reflections between the ground and direct beam.

The resulting parameter is called *Clear Sky Index* (CSI) and it is a dimensionless parameter which directly informs on the weather conditions: the greater the attenuation of the radiation by the clouds, the smaller the CSI.

$$CSI = \frac{GHI}{GHI_{cs}} \tag{9}$$

In addition to the usefulness of the information provided by the CSI, two other important reasons justify its calculation: firstly, normalized data are processed more easily by neural networks, and secondly, it is a suitable parameter for identifying irradiance ranges and defining classes for classification tasks.

4.3.2 Regression phase

Each sequence of images has to be associated with a label (the target) which was defined taking into account the forecast horizon. In fact, the target corresponds to the CSI measured at the prediction time, which is the time of the last image in chronological order of the sequence plus the forecast horizon. Considering a sequence composed of three images taken at t, t-Δt_{back} and t-$2\Delta t_{back}$, the CSI associated with that sequence is the one measured at t+Δt. In terms of time, considering a prediction made at 9:04 with a forecast horizon of 5 minutes and a historical timestep equal to 2, the CSI at 9:09 is the forecasting target of the sequence generated by the concatenation of the images taken at 9:00, 9:02 and 9:04.

Table 3 resumes inputs, outputs and other fundamental information adopted for this process.

Input\Output	Variable	Dimension	Samples	Type
Input	Images	128×128	33,686	double
Output	CSI	1×1	33,686	double

Table 3: Details about data exploited in the regression phase

4.3.3 Enhanced Convolutional Neural Network (ECNN)

The ECNN is a CNN to which exogenous data information is provided in the form of images.

This model exploits a further step in image manipulation in order to improve the CNN based forecasting.

Moreover, the ECNN was developed in order to maintain the same structure of the neural network and avoid the adoption of more complex forecasting methods which could handle different types of data: both IR pictures from ASI (images) and ambient parameters (real numbers). Exogenous data were codified into pixels and placed horizontally in the upper left corner of the image to not affect the sky region. Specifically, the data of interest selected was the solar radiation due to its strong relation with photos of the sky. For each grayscale picture, the corresponding GHI was codified as described in:

$$\begin{cases} w_{pixels}[0, 0 : n_{w_{pixels}}] = 255.0 \\ l_{pixel}[0, n_{w_{pixels}} + 1] = GHI\%255.0 \end{cases} \tag{10}$$

Where:

$$n_{w_{pixels}} = int(GHI/255.0) \tag{11}$$

Equation 11 defines the number of white pixels ($n_{w_{pixels}}$). Equation 10 informs about the value for the white pixels (w_{pixels}) and the last pixel (l_{pixel}) of the raw. The modulus operator (%) returns the remainder of the division and the square brackets indicate the raws and columns of the image matrix.

At first, number of white pixels was defined computing the integer value obtained by the division between irradiance and 255, the maximum value that a pixel can assume. That value corresponds to the white color

and was assigned to the white pixels. Then, the remainder of the division was the value assigned to the pixel located after the white pixels raw. In this way, the overall sum of the modified pixels is exactly the GHI, as shown in Figure 25.

This method ensures the presence of the three solar radiation measures in the three image sequence. Thus, the ECNN can identify in a more accurate way the impact of clouds and memorize the irradiance trends.
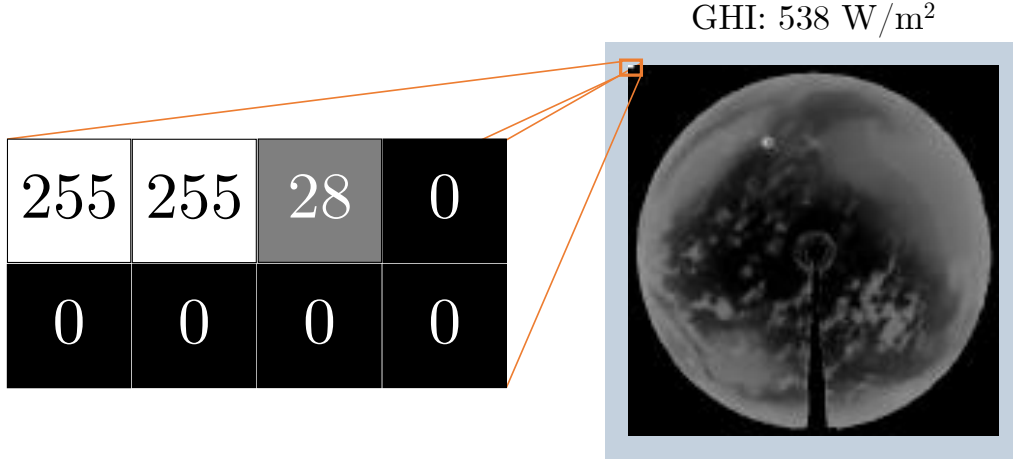

Figure 25: Illustration of the values associated with the pixels located on the left upper corner of the image for a GHI value of 538 W/m^2

4.4. Banchmarks

Two models were used as a baseline for this work. The first one is the dataset of forecasts computed in 2020 by a meteorological provider, which exploited, as the model proposed in this thesis, the infrared images and the GHI measurements collected by the meteorological station at the SolarTech$^{\text{LAB}}$. The second one is the persistence method, which basically assumes a constant CSI within the forecast horizon, resulting in a time-shifting of the GHI curve.

4.5. Evaluation metrics

The aim of this thesis is to predict the future GHI at different forecast horizons. The accuracy of the forecasting and the relative error can be evaluated through several indexes, but the main statistical indicators differ depending on the goals. In particular, the task of this study is *regression*, which aims to predict a *target* numeric value from a set of *features* called *predictors*.

4.5.1 Regression metrics

In this work, the three most used indexes in forecasting literature are selected. Specifically, they are the *Root Mean Squared Error* (RMSE), the *Mean Bias error* (MBE) and the *Forecast Skill* (FS). These metrics are necessary because they reflect the average ability of the model for predicting sudden fluctuations [46].

The RMSE measures the global error during the entire prediction period. Its formula is reported:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(y_{\text{pred}} - y_{\text{meas}}\right)^2} \tag{12}$$

Where N is the number of samples, y_{pred} is the predicted GHI value and y_{meas} is the measured GHI value.

The Mean Bias Error (MBE) indicates whether the model overestimates or underestimates the predicted irradiance.

$$MBE = \frac{1}{N} \sum_{i=1}^{N} (y_{\text{pred}} - y_{\text{meas}}) \tag{13}$$

The FS is a metric which indicates the forecast improvement against the reference naïve persistence model.

$$FS = 1 - \frac{RMSE}{RMSE_p} \tag{14}$$

Where $RMSE$ and $RMSE_p$ are the root mean squared error obtained by the forecasting model and by the persistence respectively.

In this thesis, the results of the forecasting model will be expressed according to the cited metrics. The error and the indexes will be calculated on the different time spans of the forecast.

5. Case study

In this section data sets are introduced as well as and the instrumentation adopted to collect them and the chosen architecture for the neural network and its loss function.

5.1. Datasets

The dataset is collected from a thermal infrared all sky imager and a meteorological station both located in Milan, at the SolarTech$^{\text{LAB}}$ (latitude: 45.502921°N; longitude: 9.156564°E, elevation: 140 a.s.l.), placed on the roof of the Energy department building of the Politecnico di Milano.

The two types of data, infrared all sky images and numerical meteorological measurements, were collected in the period between 18$^{\text{th}}$ September 2019 and 9$^{\text{th}}$ April 2020 considering only the daylight hours, with an interruption period of recording between 23$^{\text{th}}$ December 2019 and 12$^{\text{th}}$ February 2020.

Images were taken with a resolution of one minute and their total number is equal to 116,891. After filtering out rainy days and images where the Sun's elevation angle is too low, the remaining images used by the neural network are 33,686. They were then shuffled by day in order to generalize the learning stage of CNN. In fact, if the neural network first observes a period of sunshine and then a period of overcast sky, it learns to identify only the first type of weather condition. Images were then concatenated to build the sequences and split into three subsets: *training*, *validation* and *test*; whose size is 64%, 16% and 20% respectively.

Meteorological measurements were also collected to guide the network training and to build the benchmark models. In particular, the meteorological station performs ambient conditions measurements every ten seconds, hence they were averaged to 1 minute to keep consistency with sky images.

Images dataset	Number
Total	116,891
Rainy	41,160
Low Sun elevation	42,045
Processed	**33,686**

Table 4: Images dataset specifications

5.2. Thermal Infrared All Sky Imager
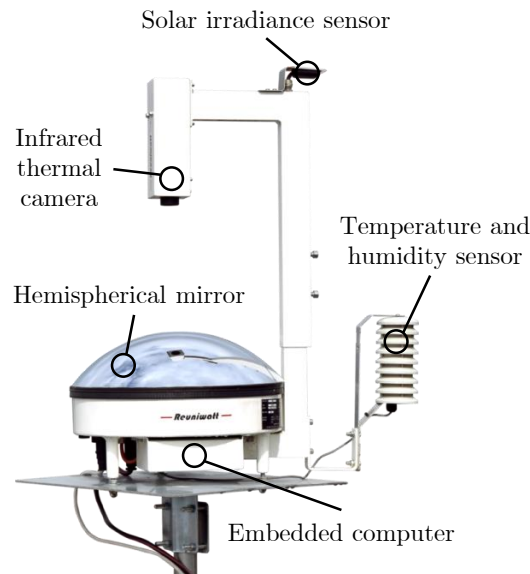
Figure 26: Thermal infrared all-sky imager

Operating conditions	Details
Operation	Day and night with identical accuracy
Field of view	360°x180°
Resolution	640x480 pixels
Format	8bit-JPEG2000
Environmental conditions	Range and rating
Operating temperature	-20°C to +60°C
IP	IP 65
Power and communications	Details
Power consumption	85 W
Power supply	110-220 VAC, 50-60 Hz
Communications	TCP/IP Ethernet 100 Mbit/s
Controller	Embedded PC
Autonomy	Uninterrupted power supply
Mechanical specifications	Details
Dimensions	350x615x800mm
Weight	10 kg
Sensors	Object of measure
Irradiance reference cell	Global Horizontal Irradiance (GHI)
Environmental parameters	Ambient temperature and relative humidity

Table 5: Technical specifications

The Thermal Infrared All Sky Imager selected for the installation is the Sky InSight[TM] developed by Reuniwatt [2]. It includes a long-wavelength infrared camera that takes a picture with a 640 by 480 spatial resolution every 60 seconds. It is mounted on a mast to film a hemispherical mirror pointing at the sky, which provides a 180° field of view. The chromed coating of the mirror surface ensures an optimal reflection of LWIR radiation. A solar reference cell measures the Global Horizontal Irradiance, whereas temperature and humidity are recorded by sensors integrated under a naturally ventilated shield. Furthermore, a computer is embedded in the housing for data logging and transmission. The housing dimensions are 350 mm in diameter and 800 mm in height. The equipment weighs 10 kg. It is designed to be dust-tight and to resist liquid ingress.

Figure 26 shows a picture of the camera and Table 5 lists its main technical specifications.

5.2.1 Cloud observation in thermal infrared spectrum

Infrared radiation has broad prospects for providing valuable cloud and atmospheric properties [47]. The effect of the atmosphere on infrared radiation varies with the spectral band: the region which ranges from 8 to 13 μm is referred to as the Long Wave InfraRed (LWIR) thermal atmospheric window due to its low emission and low absorption. This wavelengths range is appropriate for detecting cloud with high accuracy, since here the spectral emission of sky is very sensitive to water vapor presence, facilitating the identification of the type of clouds, and consequently of the meteorological conditions.

Figure 27 shows the spectral emission in the atmospheric window for three different sky states, highlighting the role of clouds as infrared emitters. The red curve represents the typical conditions of a clear sky, the blue one instead refers to the emission of thick clouds, which is significantly higher. It is evident how the radiation received on the ground varies with the water vapor contents, going from small values in the case of cloudless weather to higher ones when clouds appear. Despite a strong contrast between sunshine and overcast weather,

---

[2] https://reuniwatt.com/en/247-all-sky-observation-sky-insight/. Last accessed: April 10, 2023.

the thinner the clouds are, the lower the distinction from the cloud-free condition. An example is represented by the green curve, which is the case of a sky with cirrus clouds.
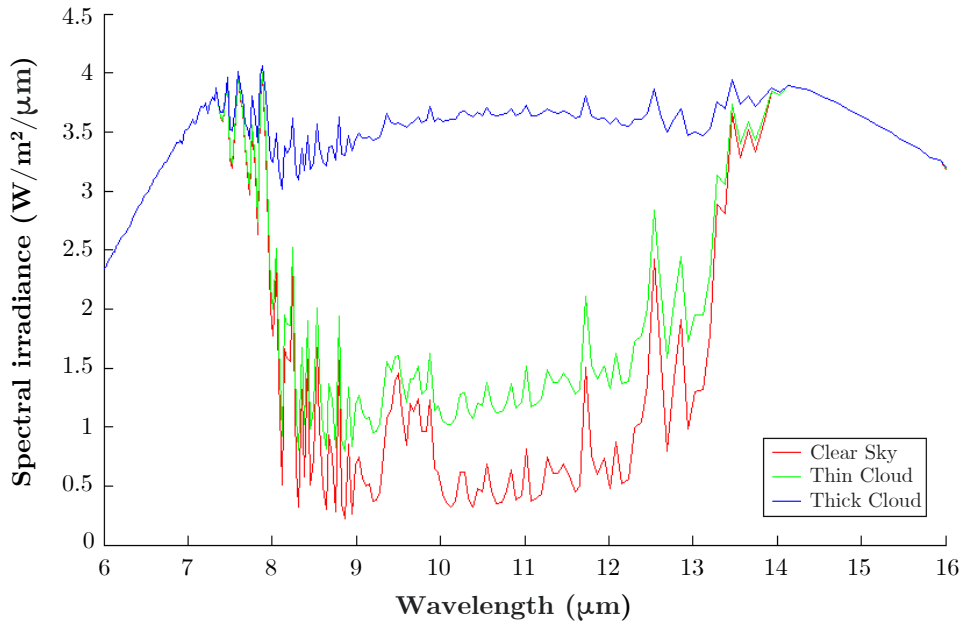

Figure 27: Spectral emission of clear sky, thin cloud, and thick cloud in the LWIR band. Image acquired from [25]

This highlights the need to evaluate carefully the atmospheric conditions to detect in a precise way the water vapor contents. A thermal infrared imager is developed exactly with this aim since it can directly detect cloud emission rather than rely on reflected and scattered sunlight as in the case of visible - or near-infrared - observation systems. Thus, cloud detection performance does not depend on the illumination of the cloud cover by the Sun. The differences between the two types of camera are shown in Figure 28

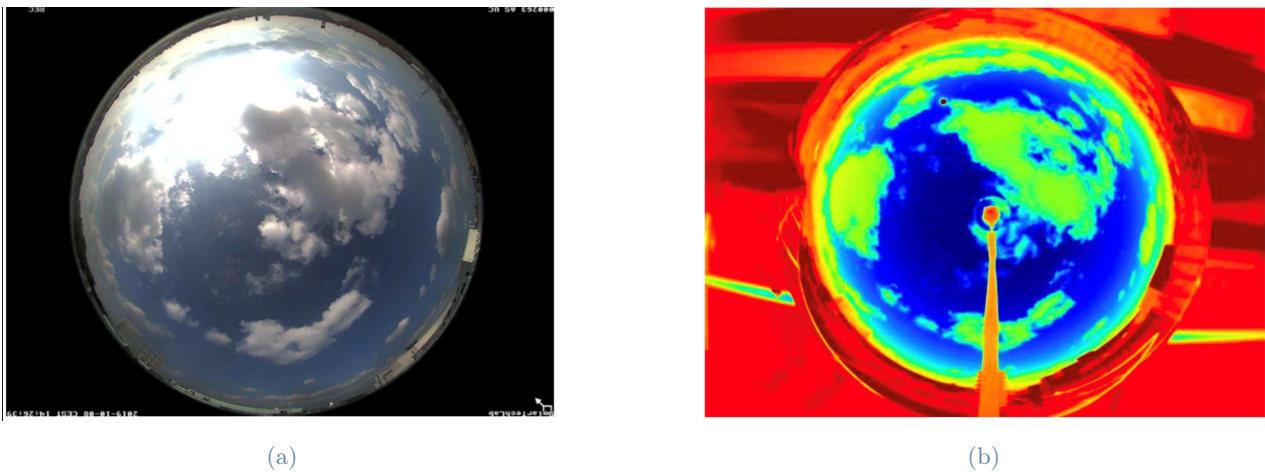(a)   (b)

Figure 28: Images taken in visible spectrum (a) and infrared spectrum (b)

These features determine the main advantages of operating in the infrared channel. They are summarized in the following list:

- No Sun glare effect in Sun-region;
- Better overall cloud features.

Despite the multiple benefits that an infrared sky imager can bring, the observation in the infrared spectrum determines also two significant drawbacks. In particular, at a low angular elevation angle (typically less than 20° above the horizon) the atmospheric thickness increases. This means a higher emission of the atmospheric components (in particular water vapor), which reduces the ability of the thermal infrared sky imager to detect

clouds, especially the thinner ones.

As a result, on a sunny day, the radiance at the center of the infrared image is the smallest and gradually increases from the center to the boundary. On the contrary, on a cloudy day, even if there are clouds at the zenith, the radiation can be less than that of the cloudless area at the border of the image.

Another drawback is related to overcast conditions. The clouds cover the entire sky making appear to the infrared cam the sky dome with approximately the same temperature everywhere, which means that there is no possibility to distinguish sky and clouds as in the case of partly cloudy weather. Furthermore, the temperature of the clouds is very low due to their high thickness. The result is an image characterized by the same color shades of a picture taken at clear sky (see Figure 29).
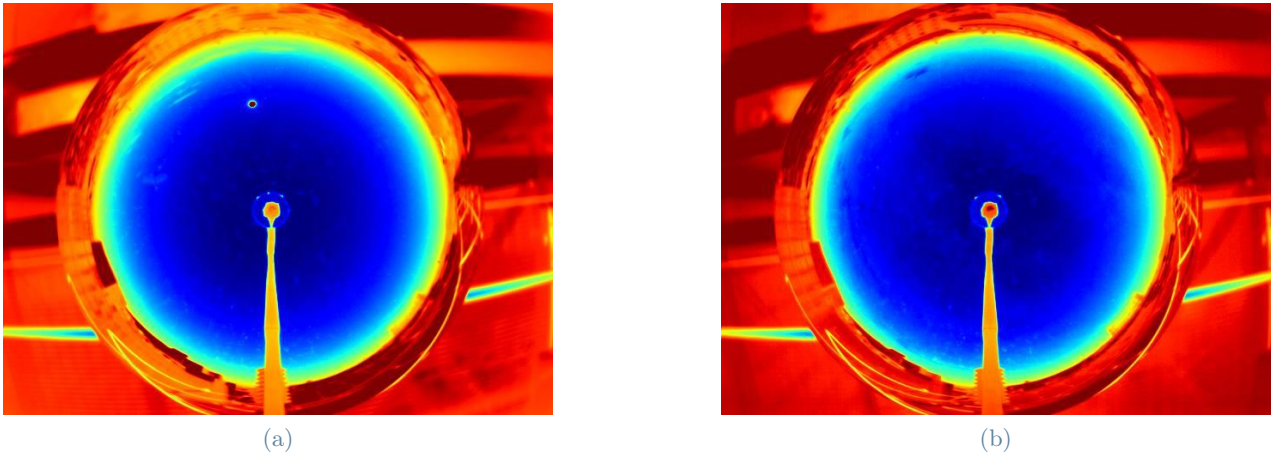

(a)   (b)

Figure 29: Similarity between a) taken under clear sky (2019-09-25 12:46:00; GHI = 648W/m^2) and b) taken under overcast condition (2019-09-23 10:55:00; GHI = 115W/m^2)

5.3.   Meteorological station

The environmental conditions were collected with a weather station equipped with solar irradiance sensors, temperature and humidity sensors, a wind speed and direction sensor and a rain collector. Solar irradiance was measured with two secondary standard pyranometers, which evaluate the global irradiance on the horizontal and 30° tilted planes. Moreover, the diffuse irradiance was collected with a pyranometer shadowed by an occultation band. The measuring equipment is shown in Figure 30 and the characteristics of each instrument are reported in Table 6.

The meteorological station performed ambient conditions measurements every ten seconds. The average, maximum, minimum and standard deviation of the values measured by the sensors were calculated every minute and stored in a database.
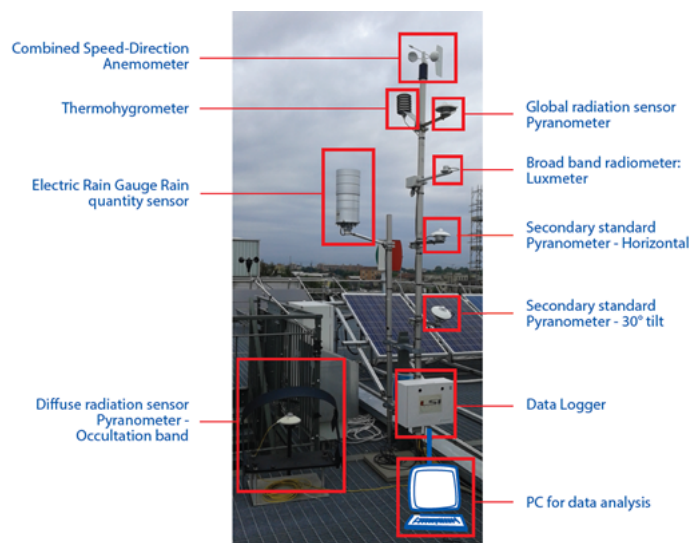

Figure 30: Meteorological station and its instrumentation

Irradiance sensors	Pyranometer (LSI, DPA252)
Standard	Secondary standard ISO 9060
Measurements range	< 2000 W/m^2
Spectral range	300-3000 nm
Total achievable daily uncertainty	$< 2\%$
Directional response	$<\pm5.4$ W/m^2
Thermal drift	$< 2\%$
Temperature and humidity sensor	Thermo-hygrometer (LSI, DMA 875)
Temperature sensor	Pt100 1/3 B (DIN EN 60751)
Measurements range	[-30°C, +70°C]
Uncertainty	0.2 °C (at 0 °C)
Resolution	0.04 °C
Response time (T90)	3 min: with filter;
	20 s: without filter(air speed 0.2 m/s)

Table 6: Solar irradiance and temperature sensor characteristics

5.4. Convolutional neural network architecture

The architecture of the neural network adopted in this study is inspired by the deep convolutional neural network developed by the Visual Geometry Group (VGG) [48]. This structure is well-researched and widely used due to its good performance at image classification and low complexity compared to the other most popular convolutional architectures.

The neural network is 2-dimensional, hence the images are processed in two directions: height and width.

Figure 31 shows the layers involved in detail. In particular, there are five Feature Learning Blocks (FLBs) to detect all the features in the images. Each FLB consists of two or three convolutional layers and a max- pooling layer. The number of filters in the first FLB is 64, but they are doubled with each next FLB to extract the most latent information.

After the convolution part, there are two fully-connected layers. The first one has 256 neurons and the second has 1 neuron. This last neuron is responsible for providing the value predicted. The rectified linear activation function was used in convolutional layers, whereas the linear activation function was used in fully-connected layers. Lastly, a 0.2 dropout was put before each fully-connected layer to avoid overfitting [49].

The model was trained by using stochastic gradient descent (SGD) with an initial learning rate of 0.0001 and mini-batch size of 128 examples. The loss function adopted was the *Mean Absolute Error*.
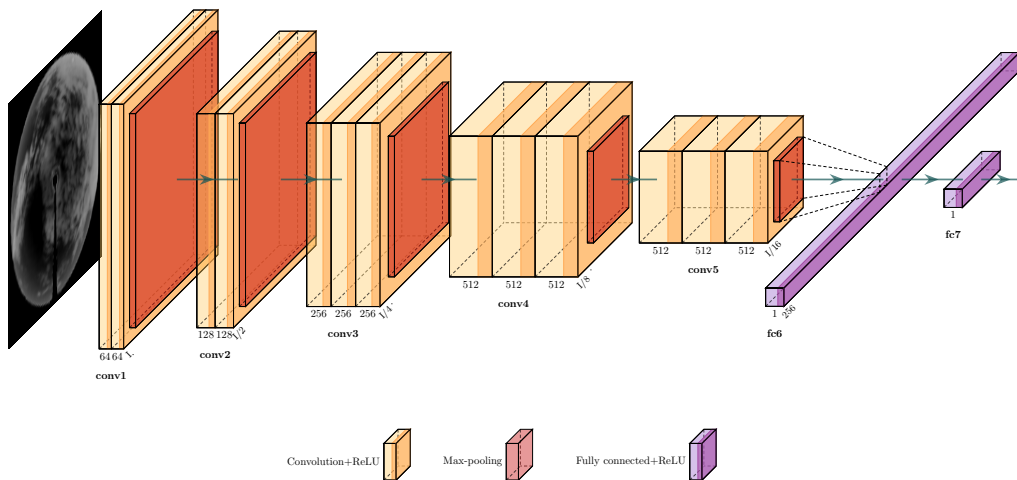

Figure 31: CNN architecture

6.   Results

In this section, the forecasting metrics obtained by the weather broadcaster provider, persistence and CNN methods are compared for the 17 testing days.

Table 7 summarizes the Forecast Skill (FS) as functions of the forecast horizons and the historical timestep. Specifically, three different historical timesteps which follow the trend of the powers of two (2^1, 2^2, 2^3) were considered for each forecast horizon.

The results show how the simple convolutional neural network model is able to match the persistence method except for predictions 5 minutes ahead. This is due to the high accuracy of persistence at those timescales. The best result was obtained by the ECNN method with a historical timestep of 4 minutes for a forecast horizon of 10 minutes, getting a FS of 9.57%. The FS trend shows the importance of increasing past observation for longer forecast horizons: the lower the historical timesteps the higher the forecast skill for 5-minute predictions; on the other hand, for 15-minute forecasts the higher the historical timesteps, the higher the FS. This trend could be explained by the importance that clouds have depending on the forecast horizon. In fact, looking at the moments near the prediction time allows for identifying sudden irradiance changes in the next few minutes: the convolutional neural network studies all the clouds near the Sun and evaluates the effects that could generate from there in a few moments. On the other hand, increasing the historical timestep makes the CNN model focus on the motion of the main clouds, which have a higher impact compared to the thinner ones, which can dissolve during the observation time and not be relevant for the forecast.

FS [%]	Forecast horizon [min]	Historical timestep [min]	Weather broadcaster	Persistence	CNN	ECNN
		2	-40.74	-	**-17.97**	**5.05**
	5	4	-40.62	-	-28.3	-0.70
		8	-40.40	-	-44.18	-11.82
		2	-17.82	-	-7.40	5.62
	10	4	-17.82	-	-32.44	**9.57**
		8	-16.07	-	**0.09**	3.89
		2	-4.24	-	0.36	5.18
	15	4	-4.15	-	-22.53	8.70
		8	-3.72	-	**0.41**	**9.07**

Table 7: FS results

Table 8 shows the Root Mean Squared Error (RMSE) and Mean Bias Error (MBE) for the simulations which scored the best FS. All the methods are characterized by an increasing trend of RMSE for longer forecast horizons due to the higher difficulty in forecasting going farther in time.

MBE is quite constant for the weather broadcaster predictions. In its model all the values are negative, meaning that the forecasts tend to underestimate future solar radiation. On the contrary, MBE of the two CNN methods is always positive. Lastly, persistence is characterized by a limited MBE due to the compensation between morning and afternoon derived by the GHI curve shifting (see figure 33).

RMSE [W/m^2]	Forecast horizon [min]	Weather broadcaster	Persistence	CNN	ECNN
	5	90.57	63.47	74.88	60.26
	10	86.99	76.66\77.54	77.46	69.32
	15	91.46	88.18	87.82	80.18
MBE [W/m^2]	Forecast horizon [min]	Weather broadcaster	Persistence	CNN	ECNN
	5	-8.30	-0.40	20.70	3.90
	10	-6.59	-1.12\-1.58	13.48	9.42
	15	-8.59	-2.59	1.05	16.47

Table 8: RMSE and MBE results for the simulations with the best FS

For the 10-minute forecast horizon both the persistence errors related to the historical timestep of 4 and 8 were listed. This is because the two CNN methods have the optimal FS for different historical timesteps for that forecast horizon. Specifically, the persistence forecast RMSE varies with the historical timestep for the same forecast horizon due to data alignment: the higher the historical timestep, the lower the available dataset.

Appendix A contains in more detail all the results obtained by each simulation.

Figure 32 shows the FS comparison among the different benchmarks and methods considered. In particular, the plot refers to the cases which got the best forecast skill for each benchmark\method: weather broadcaster predictions, CNN and ECNN (persistence is the reference).

Figure 32 depicts the ECNN ability to always outperform persistence, reaching its maximum value for a forecast horizon of 10 minutes. On the contrary, weather broadcaster predictions are always worst than persistence. Lastly, the CNN is characterized by an increasing trend which permits to do better than persistence for forecast horizons longer than 5 minutes.
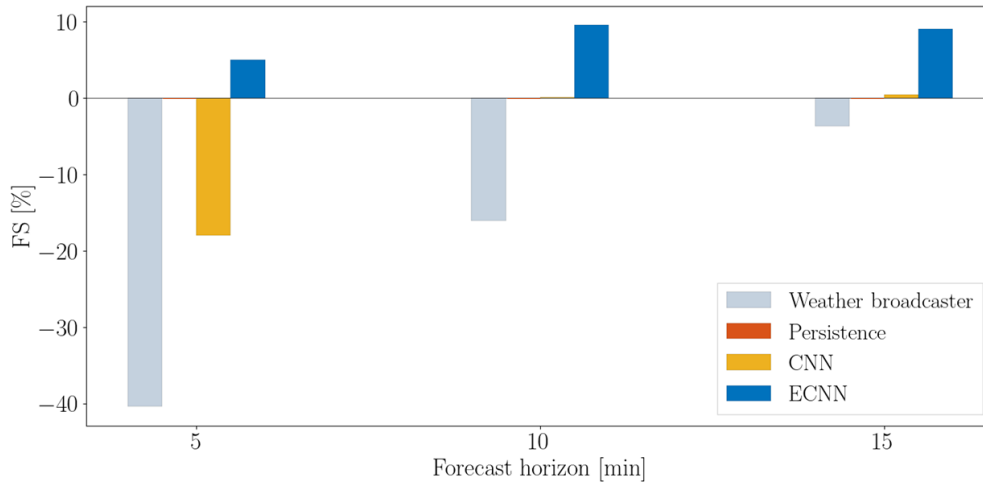

Figure 32: FS comparison

To better understand the results, predictions of 4 days, which are part of the 17 test days, are analyzed. More in detail, different weather conditions were selected to make an overview of the neural network behaviour with the highest performance (forecast horizon: 10 min; timestep: 4 min). Moreover, the persistence model, which has the best results among the benchmark methods, was added to the plots for comparison.
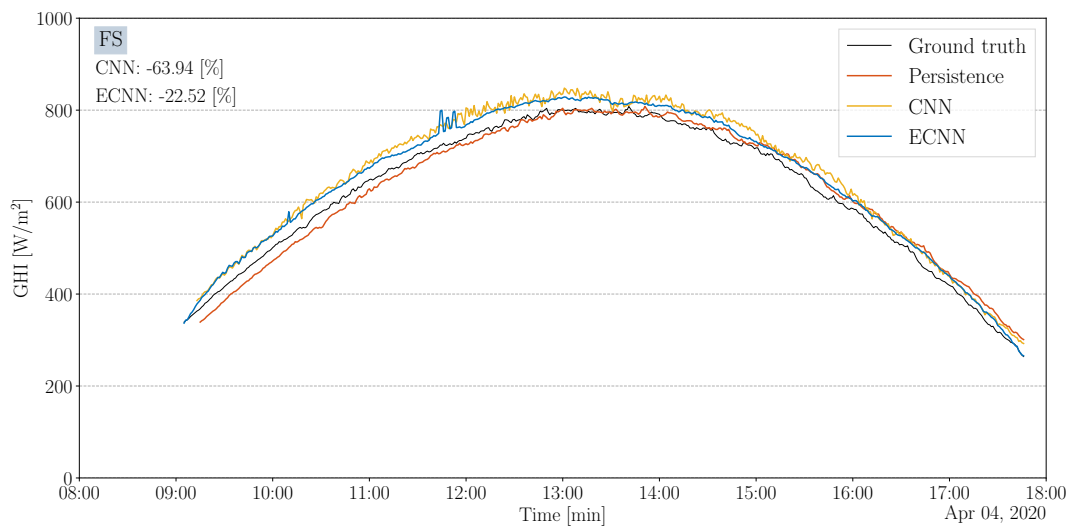

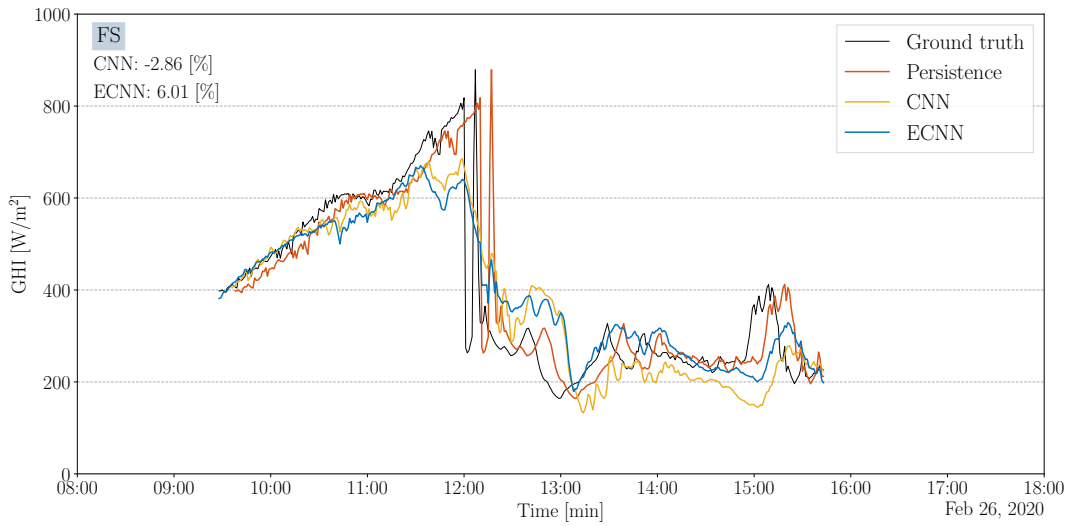
Figure 33: 4 April 2020 - clear sky conditions

Figure 34: 26 February 2020 - sunshine during the morning and partly cloudy in the afternoon
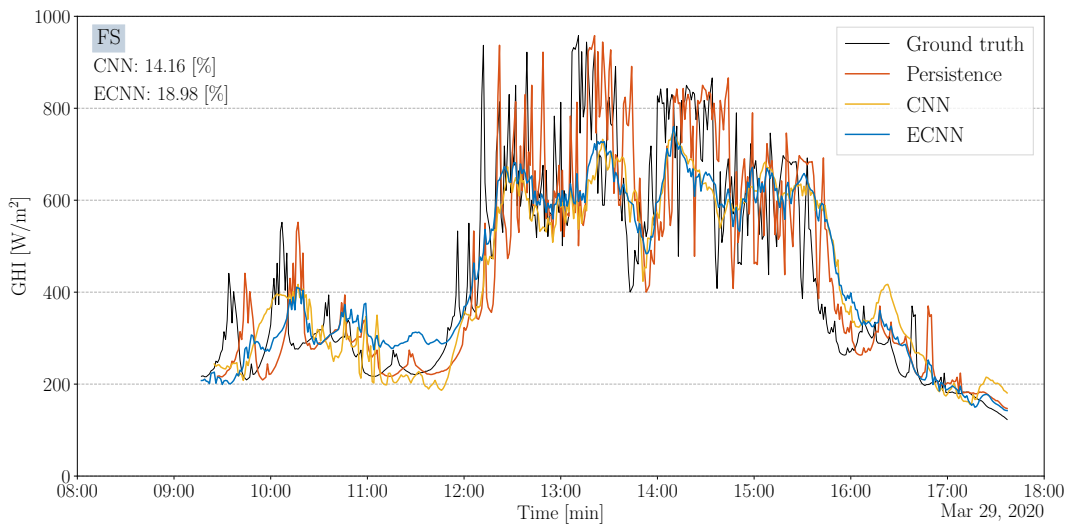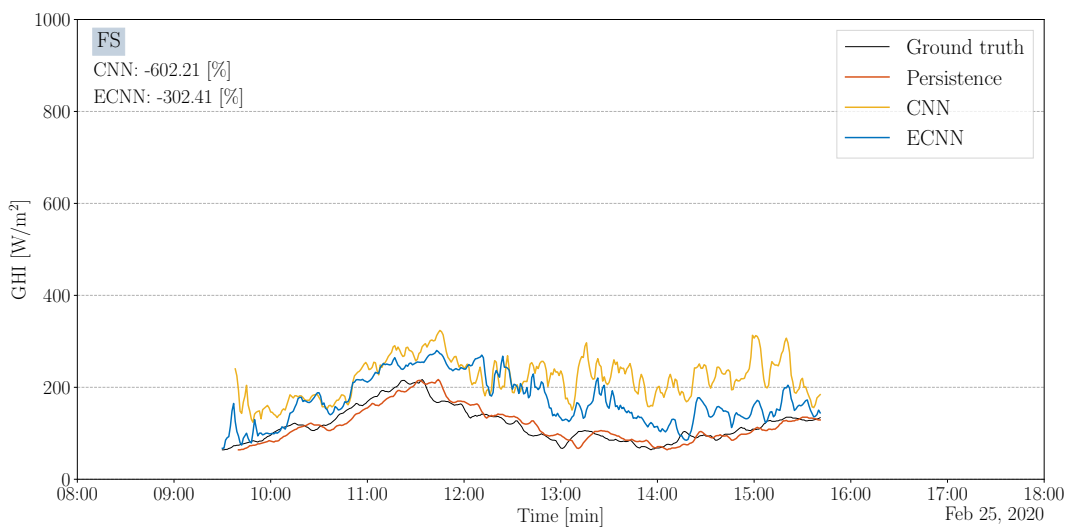

Figure 35: 29 March 2020 - partly cloudy

Figure 36: 25 February 2020 - overcast

31

Figure 33 shows the differences between persistence, CNN and ECNN methods at clear sky conditions. The three forecasting series are characterized by a similar trend and the CNN methods follow the real curve with quite a good accuracy. However, they are not able to outperform persistence due to the very low variation of solar radiation in 10 minutes, as expressed by the FS: -63.94 W/m^2 for CNN versus -22.52 W/m^2 for the ECNN. However, a difference between persistence and the two forecasting methods can be detected: persistence is quite balanced in terms of bias since it tends to underestimate the irradiance during the morning and to overestimate it in the afternoon. This is a key aspect of persistence, which repeats every time there are clear conditions. On the other hand, CNN and ECNN have a positive bias which is maintained throughout the whole day. This effect is caused by the difficulty of the neural network in forecasting very precise values of irradiance evaluating exclusively images, despite in the case of the ECNN the measured values being inserted in the picture through the modified pixels.

Figure 34, on the other hand, highlights a significant difference between the two methods. In particular, the convolutional neural network tends to behave like a very smart persistence model, avoiding large errors at the expense of missing peaks and having regular time delays. The smoothing is evident by looking at the significant drop that occurs between 12:00 and 12:30, where the blue and the yellow curves gradually decrease whereas the ground truth registers a sudden decrease. This behaviour is even more evident in Figure 35, in which all the ramps are missed.

Despite these drawbacks, the ECNN is able to mitigate the errors. Specifically, it performs better than persistence for both days represented in Figures 34 and 35, obtaining positive FS of 6.01% and 18.98%. Furthermore, the two figures show the better skill of the ECNN to be nearer to the ground truth, despite it being characterized by almost the same trend as the CNN. This characteristic makes visible the beneficial effects of providing exogenous data through the images.

Lastly, Figure 36 shows the main problem of the proposed methods: the overcast days. Specifically, looking at the above-cited figure, both the CNN performs badly for irradiance values lower than 200 W/m^2. These weather conditions are hard to evaluate by the CNN for two reasons: first, the impossibility of the CNN to know the Sun position, which means that the artificial neural network does not know the region it should focus on; second, cloudy days appear quite similar to clear sky in infrared images (see 5.2.1), making them difficult to distinguish.

This last aspect is the main cause of the high positive bias in most of the analyses reported in Table 8. In most cases, the best results were obtained for the lowest MBE since the CNN found values of weights which were able to better detect those conditions. An example is constituted by the MBEs obtained for the 10-minute forecast horizon but with 4 and 8 historical timesteps, as depicted in Figure 37. Specifically, a MBE of 52.60 W/m^2 was got for the historical timestep of 4 minutes, while the resulting MBE for the 8-minute historical timestep was 63.42 W/m^2.
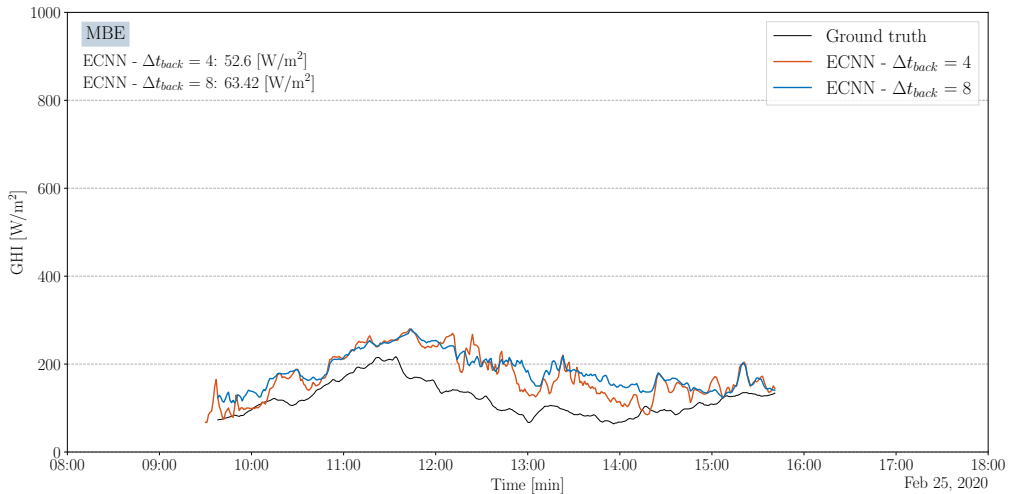

Figure 37: Comparison of the MBEs for different historical timesteps

6.1. Analysis on high irradiance fluctuation samples

In literature, the results provided by the forecasting model are usually carried out on the basis of different meteorological conditions. Larger errors are found on cloudy and partially cloudy days due to the greater variation of GHI compared to clear skies. This section proposes a different approach for the analysis of the results, looking directly at those instants characterized by high fluctuations in solar irradiance. For this reason,

Probability Density Functions (PDF) of the difference between the actual GHI variation and the one object of prediction were computed for each forecast horizon. Hence, three datasets with the difference between the GHI at 5, 10 and 15 minutes ahead and the instantaneous one were obtained for each CNN method. Then, the mean and standard deviation were calculated. Standard deviation was used to define the right number of bins in the histogram as told by one of the most recognized bin size rules: Scott's Rule [50]. Its formula is reported in:

$$b_w = 3.49\sigma n^{-1/3} \tag{15}$$

Where b_w is the bin width, σ is the standard deviation and n is the number of elements within the set.

Once the range was defined, the frequencies of each bin in the histograms were computed as well as two other important parameters for a distribution curve: the skewness and the kurtosis. The former is a measure of asymmetry, the latter instead shows whether the data are heavy-tailed or light-tailed relative to a normal distribution. In all cases, curves cannot be identified as normal since the kurtosis is always higher than three, which means that the probability density curve is higher and thinner with respect to a Gaussian. On the other hand, the mean and the skewness are close to zero, which means that fluctuations are equally distributed between negative and positive. The curves reported in Figures 38, 39 and 40 show how most of the irradiance variations are small. However, the high fluctuations are fundamental for both electric grid balancing and microgrid management. The proposed analysis focuses precisely on this aspect, evaluating the performance of the best CNN models in terms of FS against the persistence method under those conditions that rarely occur, but which are significant. In particular, according to the bin width of each dataset, at least 80% of the most probable data[3] were removed and not considered by the error metrics. This means that only variations approximately[4] lower than -80 W/m^2 and higher than +60 W/m^2 were considered. Figures 38, 39 and 40 depict in light blue the region which was not evaluated, whereas red tails represent the analysed data.

The results show a significant improvement of the CNN model in these conditions, reaching forecast skill values at least of 19.91%, demonstrating the effectiveness of the proposed models for the most crucial and significant periods.

Forecast horizon [min]	FS CNN [%]	FS ECNN [%]
5	19.91	20.20
10	24.19	22.25
15	20.53	25.47

Table 9: Forecast skill for the best models in high irradiance fluctuation conditions
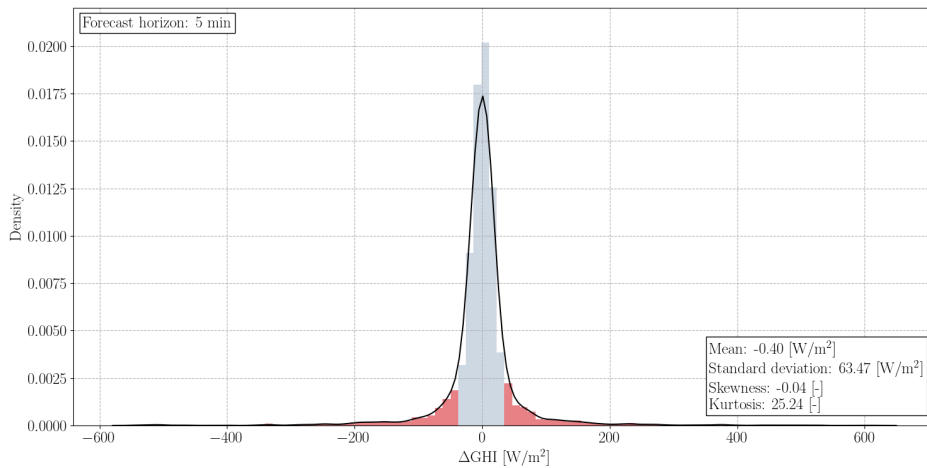

Figure 38: Probability density function for the forecast horizon equal to 5 minutes

---

[3]Bin size is finite, so it is not possible to consider exactly the middle 80%. Specifically, the percentages considered are 82.46%, 81.56% and 80.55% for forecast horizons of 5, 10 and 15 minutes respectively.

[4]Specific values of fluctuation depend by the forecast horizon dataset considered, the larger values among them were taken to be conservative.
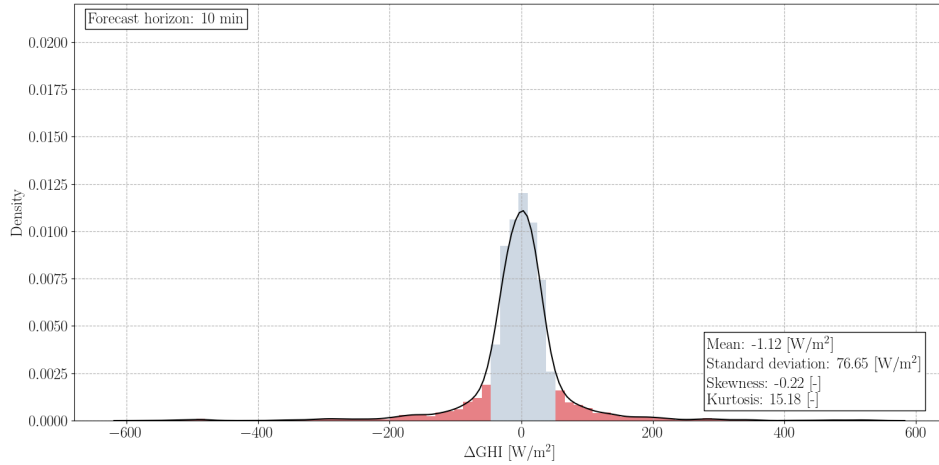
Figure 39: Probability density function for the forecast horizon equal to 10 minutes
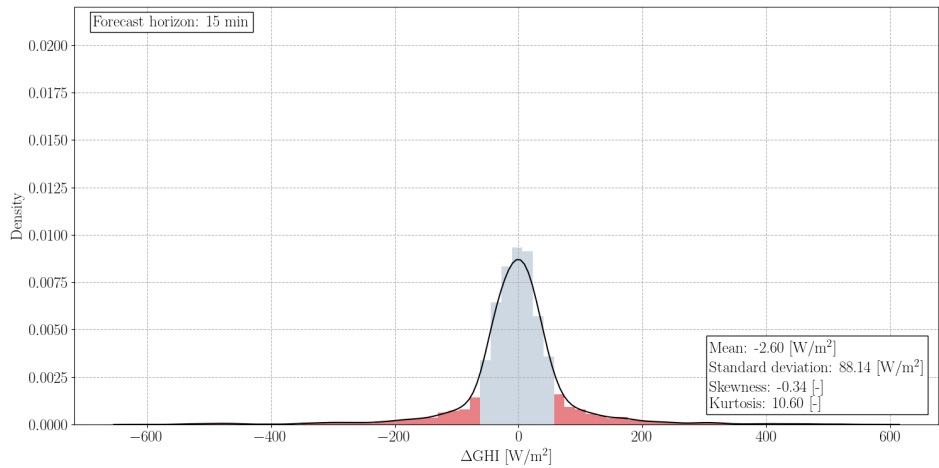

Figure 40: Probability density function for the forecast horizon equal to 15 minutes

This study demonstrated the improvement that neural networks can bring to the forecasting field. However, this technique is still at an early stage of its development. Results highlight that persistence is still the most accurate method in very short term (5') with respect to the basic CNN. A solution to improve the forecasting skill is to provide more information to the CNN at the expense of incrementing the computational burden. In particular, two different sequences of images belonging to further and closer times can be fed into the neural network, so that it can detect both the dynamics of the clouds in the moments just before the prediction, as well as the clouds at the border of the image that could cover the Sun in the next minutes. Furthermore, the combination between the images and exogenous data can help the artificial model. It is shown by the results of the ECNN, which reveals a significant improvement due to the better skill in fluctuations day provided by the GHI data. In fact, the enhanced method detects the information related to the variation of irradiance among the images in the sequence. However, it does not understand that they indicate the exact value of irradiance. It is evident by looking at the overcast days, where ECNN is still not able to detect solar radiation with high precision. Thus, it could be helpful to design an improved method in which data codified in pixels can be more easily understandable by the ECNN. Moreover, despite the root means squared error and the forecast skill are better than the persistence ones and further increased in the case with high fluctuations of irradiance, the GHI curves reveal the tendency of both the CNN methods to operate conservatively, favouring a smoothing trend in the predictions with respect to forecast ramps. This last aspect suggests the need of additional metrics to evaluate forecastings, which measure the forecast delay or the ability to predict peaks and drops [51]. Lastly, training on a larger dataset could improve the accuracy of the CNN methods since they can memorize and detect more features.

6.2. Computational load

This study's results were obtained using an NVIDIA® T4 GPU on Google Colaboratory. The forecasting model was implemented using the `Keras` library with `Tensorflow` backend in `Python 3.10.0`.

Forecast horizon [min]	Historical timestep [min]	CNN Execution time [h:m:s]	ECNN Execution time [h:m:s]
	2	02:01:18	01:53:20
5	4	01:26:41	01:35:12
	8	03:41:25	01:47:13
	2	01:51:44	01:58:24
10	4	01:08:05	02:02:37
	8	03:32:45	01:12:32
	2	01:32:58	01:21:31
15	4	01:10:07	01:13:45
	8	01:33:04	01:05:27

Table 10: Execution times

The training time needed to perform the simulations is shown in Table 10. The average time for both the simple and the ECNN methods is less than two hours. This is a very competitive time for computations which exploit artificial intelligence, especially for calculations that involve convolutional neural networks, which can take one day or more.

Moreover, the two outliers in the simple CNN methods, characterized by a simulation time higher than three hours, were caused by the variability in the speed of the connection to the online services of Google Colaboratory.

Conclusions

The research of this thesis develops a convolutional neural network (CNN) to make very-short term forecasting of the solar radiation on the horizontal plane. The inputs are constituted by a dataset of images of the sky taken with a field of view of 360°×180°, measurements of the global horizontal irradiance collected by a meteorological station located at the Department of Energy of Politecnico di Milano (latitude: 45.502921°N; longitude: 9.156564°E, elevation: 140 a.s.l.) and the Ineichen and Perez clear sky model. In particular, the images adopted for the study were taken in the infrared spectrum to evaluate their advantages, especially for their capabilities in providing cloud properties and atmospheric data. The CNN method stacked three greyscale images with a 128×128 resolution with the aim to detect clouds motion and dynamics. More in detail, a sensitivity analysis on the concatenation of the images considering different time intervals was carried out. The input dataset refers to about 6 months, made by 33,686 images which were analysed. Furthermore, an innovative technique exploiting exogenous data inserted as additional information directly in the pixels of the available images is here proposed: the Enhanced Convolutional Neural Network (ECNN). The proposed models were compared with two baselines: a dataset of predictions provided by a weather broadcaster provider and the forecasts computed through the persistence method. Moreover, both the CNN based models outperform the predictions of the meteorological company for the three forecast horizons considered (5, 10 and 15 minutes). On the contrary, the simple CNN method is not able to make better than persistence for predictions 5 minutes ahead. The ECNN method outperforms all the considered models in all the forecasting time horizons, besides it is more accurate than persistence: positive forecast skill scores of 5.04%, 9.57% and 9.07% were got for 5, 10 and 15 minutes ahead respectively. All the results are negatively affected by the bad performance of the model on overcast days since infrared images appear similar to clear sky under those conditions. Lastly, an analysis of those periods characterized by high fluctuations of irradiance was carried out since they are the most crucial events for both grid and microgrid management despite they being less frequent. Thus, probability density functions were computed and just variations which fall outside the 80% of the most probable events were taken into account by the error metrics. Results show a significant improvement in the proposed methods with respect to persistence, which reach forecast skill score always higher than 19.91%.

References

[1] IEA. Renewable energy market update 2022, 2022.

[2] IEA. Renewables 2021 - analysis and forecast to 2026, 2021.

[3] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif. A review and evaluation of the state-of-the-art in pv solar power forecasting: Techniques and optimization, 5 2020.

[4] Sobrina Sobri, Sam Koohi-Kamali, and Nasrudin Abd Rahim. Solar photovoltaic generation forecasting methods: A review, 1 2018.

[5] Amanpreet Kaur, Lukas Nonnenmacher, Hugo T.C. Pedro, and Carlos F.M. Coimbra. Benefits of solar forecasting for energy imbalance markets. *Renewable Energy*, 86:819–830, 2016.

[6] Simon R. Sinsel, Rhea L. Riemke, and Volker H. Hoffmann. Challenges and solution technologies for the integration of variable renewable energy sources—a review, 1 2020.

[7] Hector Beltran, Emilio Pérez, Néstor Aparicio, and Pedro Rodriguez. Daily solar energy estimation for minimizing energy storage requirements in pv power plants. *IEEE Transactions on Sustainable Energy*, 4:474–481, 2013.

[8] Fermín Rodríguez, Alice Fleetwood, Ainhoa Galarza, and Luis Fontán. Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. *Renewable energy*, 126:855–864, 2018.

[9] Utpal Kumar Das, Kok Soon Tey, Mehdi Seyedmahmoudian, Saad Mekhilef, Moh Yamani Idna Idris, Willem Van Deventer, Bend Horan, and Alex Stojcevski. Forecasting of photovoltaic power generation and model optimization: A review, 1 2018.

[10] Muhammad Qamar Raza, Mithulananthan Nadarajah, and Chandima Ekanayake. On recent advances in pv output power forecast, 10 2016.

[11] Chi Wai Chow, Serge Belongie, and Jan Kleissl. Cloud motion and stability estimation for intra-hour solar forecasting. *Solar Energy*, 115:645–655, 6 2015.

[12] Alejandro Catalina, Carlos M. Alaiz, and Jose R. Dorronsoro. Combining numerical weather predictions and satellite data for pv energy nowcasting. *IEEE Transactions on Sustainable Energy*, 11:1930–1937, 7 2020.

[13] Alfredo Nespoli, Alessandro Niccolai, Emanuele Ogliari, Giovanni Perego, Elena Collino, and Dario Ronzio. Machine learning techniques for solar irradiation nowcasting: Cloud type classification forecast through satellite data and imagery. *Applied Energy*, 305, 1 2022.

[14] A. Kazantzidis, P. Tzoumanikas, A. F. Bais, S. Fotopoulos, and G. Economou. Cloud detection and classification with the use of whole-sky ground-based images. *Atmospheric Research*, 113:80–88, 9 2012.

[15] Thomas Schmidt, John Kalisch, Elke Lorenz, and Detlev Heinemann. Evaluating the spatiooral performance of sky-imager-based solar irradiance analysis and forecasts. *Atmospheric Chemistry and Physics*, 16:3399–3412, 3 2016.

[16] Cunzhao Shi, Chunheng Wang, Yu Wang, and Baihua Xiao. Deep convolutional activations-based features for ground-based cloud classification. *IEEE Geoscience and Remote Sensing Letters*, 14:816–820, 6 2017.

[17] Jinglin Zhang, Pu Liu, Feng Zhang, and Qianqian Song. Cloudnet: Ground-based cloud classification with deep convolutional neural network. *Geophysical Research Letters*, 45:8665–8672, 8 2018.

[18] Stavros Andreas Logothetis, Vasileios Salamalikis, Stefan Wilbert, Jan Remund, Luis F. Zarzalejo, Yu Xie, Bijan Nouri, Evangelos Ntavelis, Julien Nou, Niels Hendrikx, Lennard Visser, Manajit Sengupta, Mário Pó, Remi Chauvin, Stephane Grieu, Niklas Blum, Wilfried van Sark, and Andreas Kazantzidis. Benchmarking of solar irradiance nowcast performance derived from all-sky imagers. *Renewable Energy*, 199:246–261, 11 2022.

[19] Yuchi Sun, Gergely Szucs, and Adam R. Brandt. Solar pv output prediction from video streams using convolutional neural networks. *Energy and Environmental Science*, 11:1811–1818, 7 2018.

[20] Richard Perez, Pierre Ineichen, Kathy Moore, Marek Kmiecik, Cyril Chain, Ray George, and Frank Vignola. A new operational model for satellite-derived irradiances: Description and validation, 2002.

[21] Dinesh Pothineni, Martin R Oswald, Jan Poland, and Marc Pollefeys. Kloudnet: Deep learning for sky image analysis and irradiance forecasting, 2018.

[22] Xin Zhao, Haikun Wei, Hai Wang, Tingting Zhu, and Kanjian Zhang. 3d-cnn-based feature extraction of ground-based cloud images for direct normal irradiance prediction. *Solar Energy*, 181:510–518, 3 2019.

[23] Yiren Wang, Dong Liu, Wanyi Xie, Ming Yang, Zhenyu Gao, Xinfeng Ling, Yong Huang, Congcong Li, Yong Liu, and Yingwei Xia. Day and night clouds detection using a thermal-infrared all-sky-view camera. *Remote Sensing*, 13, 5 2021.

[24] Brentha Thurairajah and Joseph A. Shaw. Cloud statistics measured with the infrared cloud imager (ici). *IEEE Transactions on Geoscience and Remote Sensing*, 43:2000–2007, 9 2005.

[25] Clément Bertin, Sylvain Cros, Laurent Saint-Antonin, and Nicolas Schmutz. Prediction of optical communication link availability: real-time observation of cloud patterns using a ground-based thermal infrared camera. volume 9641, page 96410A. SPIE, 10 2015.

[26] Alessandro Niccolai and Alfredo Nespoli. Sun position identification in sky images for nowcasting application. *Forecasting*, 2:488–504, 11 2020.

[27] Olivier Liandrat, Antonin Braun, Etienne Buessler, Marion Lafuma, Sylvain Cros, André Gomez, and Etienne Boudreault. Sky-imager forecasting for improved management of a hybrid photovoltaic-diesel system, 2018.

[28] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

[29] Shreya Dutta, Yanling Li, Aditya Venkataraman, Luis M. Costa, Tianxiang Jiang, Robert Plana, Philippe Tordjman, Fook Hoong Choo, Chek Fok Foo, and Hans B. Puttgen. Load and renewable energy forecasting for a microgrid using persistence technique. volume 143, pages 617–622. Elsevier Ltd, 2017.

[30] Yiannis Dendramis, George Kapetanios, and Massimiliano Giuseppe Marcellino. A similarity-based approach for macroeconomic forecasting. *Available at SSRN 3330082*, 2019.

[31] Itzhak Gilboa, Offer Lieberman, and David Schmeidler. A similarity-based approach to prediction. *Journal of Econometrics*, 162(1):124–131, 2011.

[32] Qingqing Mu, Yonggang Wu, Xiaoqiang Pan, Liangyi Huang, and Xian Li. Short-term load forecasting using improved similar days method. In *2010 Asia-Pacific Power and Energy Engineering Conference*, pages 1–4. IEEE, 2010.

[33] R Tawn and J Browell. A review of very short-term wind and solar power forecasting. *Renewable and Sustainable Energy Reviews*, 153:111758, 2022.

[34] Bella Espinar, José-Luis Aznarte, Robin Girard, Alfred Mbairadjim Moussa, and Georges Kariniotakis. Photovoltaic forecasting: A state of the art. In *5th European PV-hybrid and mini-grid conference*, pages Pages–250. OTTI-Ostbayerisches Technologie-Transfer-Institut, 2010.

[35] Oludare Isaac Abiodun, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat AbdElatif Mohamed, and Humaira Arshad. State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4:938, 2018.

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

[37] Javier Antonanzas, Natalia Osorio, Rodrigo Escobar, Ruben Urraca, Francisco J Martinez-de Pison, and Fernando Antonanzas-Torres. Review of photovoltaic power forecasting. *Solar energy*, 136:78–111, 2016.

[38] Stuart J Russell. *Artificial intelligence a modern approach*. Pearson Education, Inc., 2010.

[39] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.", 2019.

[40] Sakib Mostafa and Fang-Xiang Wu. Diagnosis of autism spectrum disorder with convolutional autoencoder and structural mri images. In *Neural Engineering Techniques for Autism Spectrum Disorder*, pages 23–38. Elsevier, 2021.

[41] Manassés Ribeiro, André Eugênio Lazzaretti, and Heitor Silvério Lopes. A study of deep convolutional auto-encoders for anomaly detection in videos. *Pattern Recognition Letters*, 105:13–22, 2018.

[42] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[43] The MathWorks Inc. Matlab version: 9.13.0 (r2022b), 2022.

[44] Claudia Perlich. Learning curves in machine learning., 2010.

[45] William F Holmgren, Clifford W Hansen, and Mark A Mikofski. pvlib python: A python package for modeling solar energy systems. *Journal of Open Source Software*, 3(29):884, 2018.

[46] Ruiyuan Zhang, Hui Ma, Tapan Kumar Saha, and Xiaofang Zhou. Photovoltaic nowcasting with bi-level spatio-temporal analysis incorporating sky images. *IEEE Transactions on Sustainable Energy*, 12(3):1766–1776, 2021.

[47] Paul W Nugent, Joseph A Shaw, and Sabino Piazzolla. Infrared cloud imaging in support of earth-space optical communication. *Optics express*, 17(10):7862–7872, 2009.

[48] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[49] Cong Feng, Jie Zhang, Wenqi Zhang, and Bri-Mathias Hodge. Convolutional neural networks for intra-hour solar forecasting based on sky image sequences. *Applied Energy*, 310:118438, 2022.

[50] Philip A Legg, Paul L Rosin, David Marshall, and James E Morgan. Improving accuracy and efficiency of mutual information for multi-modal retinal image registration using adaptive probability density estimation. *Computerized Medical Imaging and Graphics*, 37(7-8):597–606, 2013.

[51] Quentin Paletta, Guillaume Arbod, and Joan Lasenby. Benchmarking of deep learning irradiance forecasting models from sky images–an in-depth analysis. *Solar Energy*, 224:855–867, 2021.

A. Appendix A

RMSE [W/m²]	Forecast horizon [min]	Historical timestep [min]	Weather broadcaster	Persistence[*]	CNN	ECNN
		2	89.33	63.47	**74.88**	**60.26**
	5	4	89.76	63.83	81.90	64.28
		8	**90.57**	64.51	93.01	72.14
		2	89.89	76.29	81.94	72.00
	10	4	90.32	76.66	101.53	**69.32**
		8	**86.99**	77.54	**77.46**	74.52
		2	90.40	86.72	86.41	82.23
	15	4	90.84	87.22	106.87	79.63
		8	**91.46**	88.18	**87.82**	**80.18**

MBE [W/m²]	Forecast horizon [min]	Historical timestep [min]	Weather broadcaster	Persistence	CNN	ECNN
		2	-8.32	-0.40	**20.70**	**3.90**
	5	4	-8.26	-0.51	10.43	17.94
		8	**-8.30**	-0.74	25.24	22.27
		2	-8.47	-0.84	20.90	15.57
	10	4	-8.43	-1.12	18.36	**9.42**
		8	**-6.59**	-1.58	**13.48**	17.52
		2	-8.44	-1.41	14.01	25.53
	15	4	-8.40	-1.74	26.72	4.34
		8	**-8.59**	-2.59	**1.05**	**16.47**

FS [%]	Forecast horizon [min]	Historical timestep [min]	Weather broadcaster	Persistence	CNN	ECNN
		2	-40.74	-	**-17.97**	**5.05**
	5	4	-40.62	-	-28.3	-0.70
		8	**-40.40**	-	-44.18	-11.82
		2	-17.82	-	-7.40	5.62
	10	4	-17.82	-	-32.44	**9.57**
		8	**-16.07**	-	**0.09**	3.89
		2	-4.24	-	0.36	5.18
	15	4	-4.15	-	-22.53	8.70
		8	**-3.72**	-	**0.41**	**9.07**

Table 11: RMSE, MBE and FS results. Values in bold refer to the simulation which got the best forecast skill (persistence is not considered since it constitutes the benchmark).

[*]The persistence forecast RMSE varies with the historical timestep for the same forecast horizon due to data alignment: the higher the historical timestep, the lower the available dataset.

Abstract in lingua italiana

Ad oggi, la necessità di una previsione accurata della radiazione solare sta assumendo sempre più importanza a seguito dello sviluppo esponenziale della generazione di potenza da fonti rinnovabili, in particolare quella da fotovoltaico. Questa tesi studia la previsione dell'irradianza solare a brevissimo termine, la quale si pone in un lasso di tempo compreso fra pochi minuti e mezz'ora. In particolare, questa tesi ha analizzato orizzonti di previsione di 5, 10 e 15 minuti. Il metodo utilizzato si fonda sullo sviluppo di una rete neurale convoluzionale in grado di elaborare sequenze di immagini per rilevare il movimento delle nuvole. Per effettuare questa ricerca sono stati sfruttati circa 6 mesi di immagini, le quali sono risultate in totale 33,686 dopo essere state opportanemente filtrate e processate. I valori di irradianza globale sul piano orizzontale e quelli equivalenti in caso di cielo sereno sono stati ottenuti rispettivamente dalla stazione meteorologica situata al SolarTech^{LAB} del Politecnico di Milano e dal modello di cielo sereno sviluppato da Ineichen e Perez. Essi sono entrambi stati utilizzati per guidare la rete neurale durante l'allenamento. Inoltre, è stata ideata una tecnica innovativa a cui si forniscono alla rete neurale convoluzionale informazioni relative a parametri esogeni sotto forma di immagine. I risultati ottenuti hanno mostrato che il modello proposto supera le previsioni fornite dall'azienda meteorologica che ha sfruttato gli stessi dati usati da questo studio. Tuttavia, la rete neurale convoluzionale non è in grado di fare meglio della persistenza per l'orizzonte temporale di 5 minuti. Le previsioni a 10 e 15 minuti invece hanno ottenuto dei risultati superiori al metodo della persistenza, in particolare con "forecast skills" (FS) pari a 0.09% e 0.41%. Mentre la rete convoluzionale in grado di elaborare dati esogeni riesce a superare la persistenza per tutti gli orizzonti temporali in analisi, ottenendo rispettivamente FS del 5.05%, 9.57% e 9.07% per i 5', 10' e 15'. Inoltre, un'analisi specifica dei risultati nei casi di alta variazione dell'irradianza solare ha mostrato come il metodo proposto riesca a battere significativamente la persistenza, ottenendo valori di forecast skill maggiore del 19% in tutti i casi. Infine, sono stati discussi i principali problemi riscontrati, insieme alle loro possibili soluzioni e agli sviluppi futuri.

Parole chiave: intelligenza artificiale, previsione solare, fotovoltaico, immagini infrarossi

Acknowledgements