Università di Perugia
Dipartimento di Matematica e Informatica

Corso di Laurea in Informatica

A Deep Learning approach for Time Series Imputation on Photovoltaic data

Relatori
Valentina Poggioni
Enrico Bellocchio
Alessandro Devo

Laureando
Nicolò Vescera

Anno Accademico 2022-2023

Contents

Abstract

La crescente necessità dell'adozione di strumenti in grado di produrre energia pulita da risorse rinnovabili e sostenibili ha portato ad una vasta generazione e raccolta di dati di produzione energetica provenienti, specialmente, da pannelli fotovoltaici installati in tutto il mondo. Tuttavia, questi dati spesso presentano lacune e mancanze dovute a vari fattori, come guasti temporanei, condizioni meteorologiche avverse o malfunzionamenti dei sensori e degli strumenti di raccolta dati. L'accurata imputazione di queste lacune è cruciale per garantire l'affidabilità delle analisi e delle previsioni basate su questi dati. Questa tesi si propone di affrontare il problema dell'imputazione di serie temporali di dati provenienti da pannelli fotovoltaici utilizzando tecniche avanzate di deep learning. In particolare, viene proposto un modello di deep learning basato su reti neurali fully connected (FCNN) e reti neurali ricorrenti (RNN) progettato per catturare le relazioni temporali complesse tra l'energia totale prodotta (target feature) e i vari componenti dell'impianto. I modelli sono stati allenati su un dataset formato di dati reali provenienti da un impianto fotovoltaico della potenza di $1MW$.

Chapter 1

Background

1.1 Problem Definition

La seguente tesi si pone l'obbiettivo di risolvere il problema dell'imputazione di serie temporali di dati provenienti da impianti fotovoltaici. Nello specifico, accade molto spesso che gli strumenti di acquisizione dati di un impianto falliscano momentaneamente causando un periodo di tempo, più o meno lungo, dove la curva dell'energia totale prodotta è assente. Per cercare di colmare questo "buco" non basta un semplice calcolo dato che vanno tenuti in considerazione vari fattori come l'irraggiamento solare, temperatura ambientale, presenza di nuvole, pioggia, ecc. Formalmente possiamo definire il problema come segue:

Definition 1.1.1. Siano dati:

1. Un insieme di serie temporali di dati rappresentanti un impianto fotovoltaico, $S = \{S_1, S_2, ..., S_N\}$, dove N rappresenta il numero di serie temporali disponibili;

2. Una serie temporale target $t \in S$ che rappresenta l'Energia Totale Prodotta;

3. Ogni serie temporale S_i è composta da coppie ordinate (t_i, v_i), dove t_i è un timestamp che rappresenta il momento in cui è stato registrato il valore v_i.

L'obiettivo del problema dell'imputazione è stimare i valori mancanti o danneggiati nella serie temporale t.

1.2 Photovoltaic Implant

Solar photovoltaic (PV) energy systems are made up of different components. Each component has a specific role. The type of component in the system depends on the type of system and the purpose. For example, a simple PV-direct system is composed of a solar module or array (two or more modules wired together) and the load (energy-using device) it powers. A solar energy system produces direct current (DC). This is electricity which travels in one direction. The loads in a simple PV system also operate on direct current (DC). A stand-alone system with energy storage (a battery) will have more components than a PV- direct system.

1.2.1 Solar Module

The majority of solar modules available on the market and used for residential and commercial solar systems are silicon- crystalline. These modules consist of multiple strings of solar cells, wired in series (positive to negative), and are mounted in an aluminum frame. The size or area of the cell determines the amount of amperage. The larger the cell, the higher the amperage.

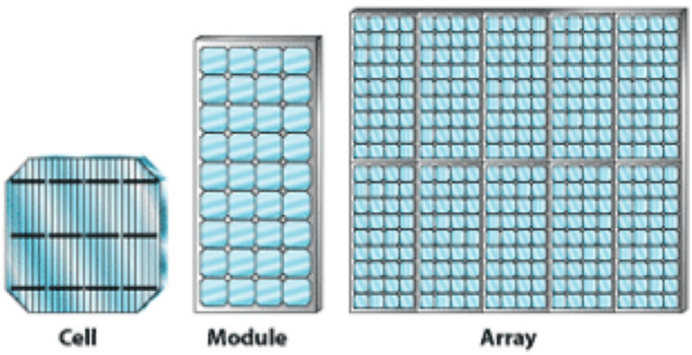

Figure 1.1: The solar cell is the basic component. Cells wired together and mounted in a frame compose a solar module. Several modules wired together form an array.

1.2.2 Solar Array

The solar array is made up of multiple PV modules wired together. Connecting the negative wire of one module to the positive wire of a second module is the beginning of a series string. Wiring modules in series results in the voltage of each of the two modules is added together. A series string represents the summed voltages of each individual module. The negative cable of one module is connected to the positive cable of the next module. In a large system, multiple strings are assembled and the non-connected ends are connected to homerun leads which are landed at the terminals of an enclosure located near the array. The goal is to wire modules in series to build voltage.

1.2.3 Junction Box

A PV system array with multiple strings of modules will have a positive lead and a negative lead on the end of each string. The positive leads will be connected to individual fuses and the negative leads will be connected to a negative busbar in an enclosure. This is called the source circuit. The junction box serves to "combine" multiple series strings into one parallel circuit. For example, an array with three strings of 10 modules wired in series would produce 300 volts (10 modules x 30 volts) per string and 4 amps per string. When the leads are landed in the combiner box, the circuit would produce 300 volts at 12 amps (3 strings x 4 amps/string). Once the circuits are combined, leaving the box it is referred to as the output circuit.

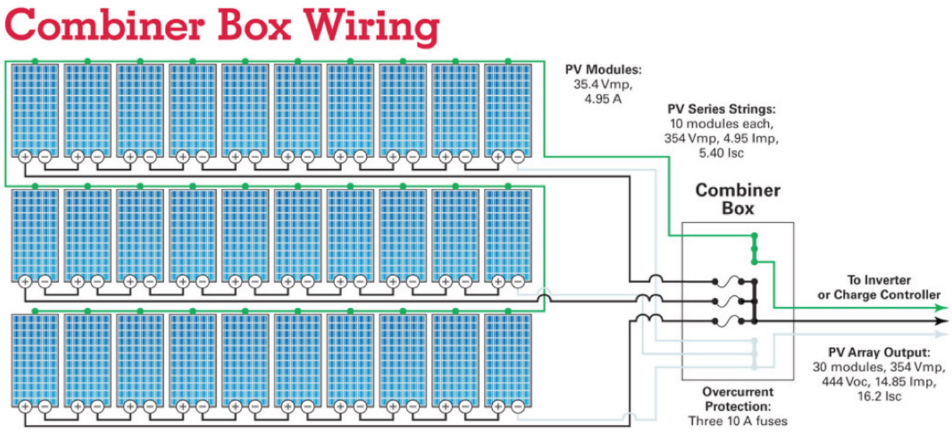

Figure 1.2: This figure represent an output circuit made of 3 string, each one hosts 10 solar modules.

1.2.4 Inverter

Energy from an array or a battery bank is direct current (DC). This will provide for DC loads such a lights, fans, pumps, motors, and some specialty equipment. However, if the energy is to be used to power loads that operate on alternating current (AC), as what is found in a residence, the current needs to be converted. The inverter changes DC energy to AC energy. Inverters are available in many different sizes for various-sized loads. A string inverter is used to convert DC power from a solar array to AC power and can be connected to an AC distribution power panel (service panel) in a residence or facility.

1.2.5 System Metering

Several tools are available to help the solar user to monitor their system. On stand-alone or off-grid PV systems, the battery meter is used to measure the energy coming in and going out of the battery bank. Charging and discharging of batteries, and proper functioning of the charging system is important to alert the user to incomplete charging, battery decline, or possible system shutdown. System monitoring with web-based tools and apps allow the solar user to see system activity using a cell phone or tablet from a location away from their system.

1.3 Multi Perceptron

The multilayer perceptron is the most known and most frequently used type of neural network. On most occasions, the signals are transmitted within the network in one direction: from input to output. There is no loop, the output of each neuron does not affect the neuron itself. This architecture is called feed- forward. Layers which are not directly connected to the environment are called hidden. There are also feed-back networks, which can transmit impulses in both directions, due to reaction connections in the network. These types of networks are very powerful and can be extremely complicated. Introduction of several layers was determined by the need to increase the complexity of decision regions. A perceptron with a single layer and one input generates decision regions under the form of semi planes. By adding another layer, each neuron acts as a standard perceptron for the outputs of the neurons in the anterior layer, thus the output of the network can estimate convex decision regions, resulting from the intersection of the semi planes generated by the neurons. The power of the multilayer perceptron comes precisely from non-linear activation functions. Almost any non-linear function can be used for this purpose, except for polynomial functions. Currently, the functions most commonly used today are the single-pole (or logistic) sigmoid.

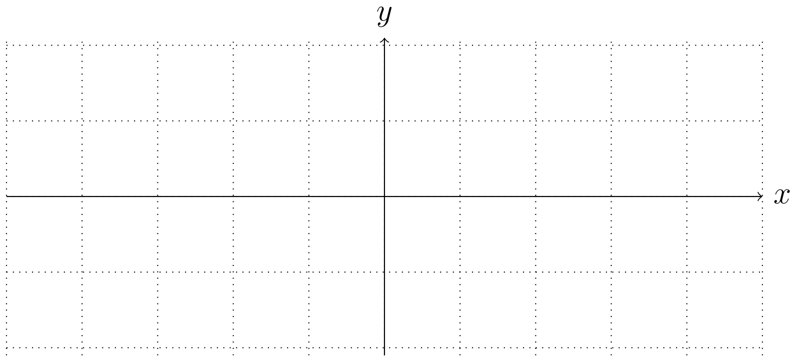

Figure 1.3: Sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$

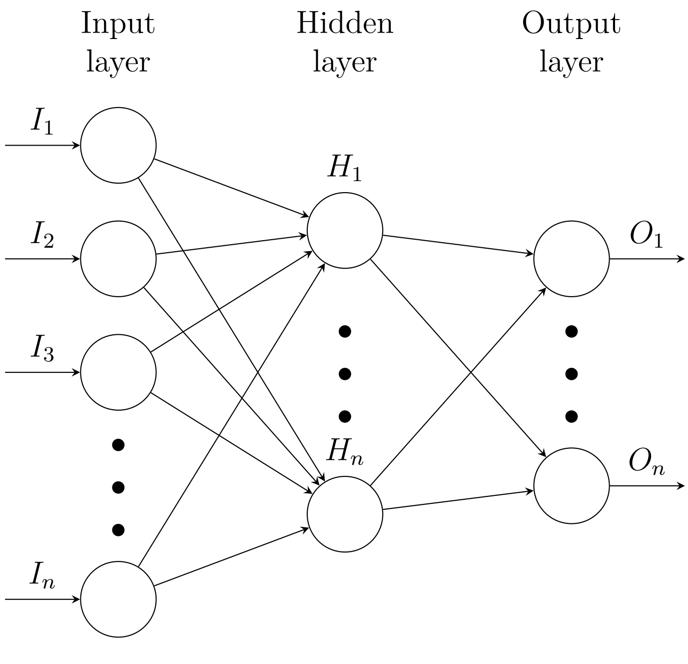

Figure 1.4: Multi Layer Perceprton architecture.

1.4 Recurrent Neural Network

Recurrent Neural Networks (RNNs) are a type of neural network architecture which is mainly used to detect patterns in a sequence of data. Such data can be handwriting, genomes, text or numerical time series which are often produced in industry settings (e.g. stock markets or sensors). However, they are also applicable to images if these get respectively decomposed into a series of patches and treated as a sequence. On a higher level, RNNs find applications in Language Modelling and Generating Text, Speech Recognition, Generating Image Descriptions or Video Tagging. What differentiates Recurrent Neural Networks from Feedforward Neural Networks also known as Multi-Layer Perceptrons (MLPs) is how information gets passed through the network. While Feedforward Networks pass information through the network without cycles, the RNN has cycles and transmits information back into itself. This enables them to extend the functionality of Feedforward Networks to also take into account previous inputs $X_{0:t-1}$ and not only the current input X_t. This difference is visualised on a high level in Figure 1.5. Note, that here the option of having multiple hidden layers is aggregated to one Hidden Layer block H. This block can obviously be extended to multiple hidden layers.

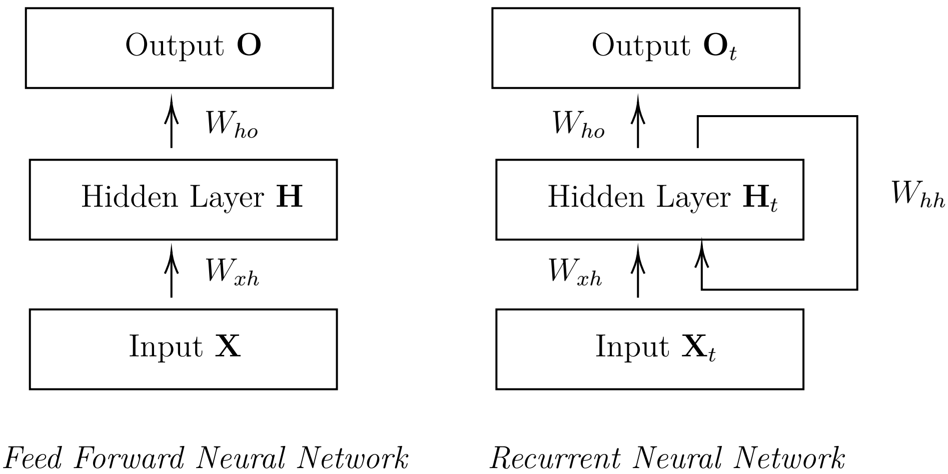

Figure 1.5: Visualization of difference between MLPs and RNN.

1.4.1 Gated Recurrent Unit

Gated Recurrent Units (GRU) are an advanced variation of RRNs (Recurrent Neural Network). GRUs use update gate and reset get for solving a standard RNNs vanishing gradient issue. These are essentially 2 vectors that decide the type of information to be passed towards the output. What makes these vectors special is that programmers can train them to store information, especially from long ago. They do all of this by utilizing its gated units which help solve vanishing/exploding gradient problems often found in traditional recurrent neural networks. These gates are helpful for controlling any information to be maintained or discarded for each step. It is also worth keeping in mind that gated recurrent units make use of reset and update gates.

- **Update Gates Function**: The main function of the update gate is to determine the ideal amount of earlier info that is important for the future.

One of the main reasons why this function is so important is that the model can copy every single past detail to eliminate fading gradient issue.

- **Reset Gate's Function**: A major reason why reset gate is vital because it determines how much information should be ignored. It would be fair to compare reset gate to LSTMs forget gate because it tends to classify unrelated data, followed by getting the model to ignore and proceed without it.

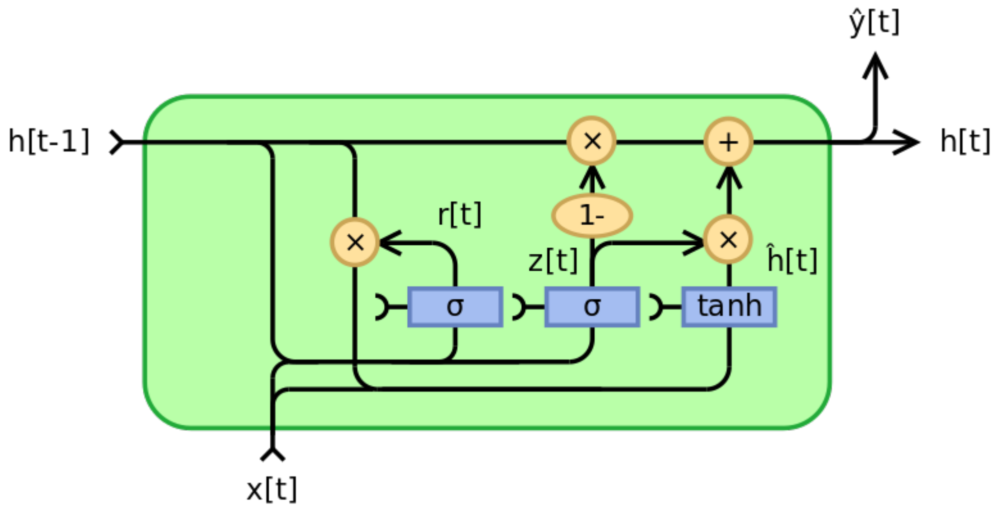

Figure 1.6: Gated Recurrent Unit (GRU) architecture.

1.5 Transformer

Recurrent neural networks, long short-term memory and gated recurrent neural networks in particular, have been firmly established as state of the art approaches in sequence modeling and transduction problems such as language modeling and machine translation. Recurrent models typically factor computation along the symbol positions of the input and output sequences. Aligning the positions to steps in computation time, they generate a sequence of hidden states h_t, as a function of the previous hidden state h_{t-1} and the input for position t. This inherently sequential nature precludes parallelization within training examples, which becomes critical at longer sequence lengths, as memory constraints limit batching across examples. Attention mechanisms have become an integral part of compelling sequence modeling and transduction models in various tasks, allowing modeling of dependencies without regard to their distance in the input or output sequences. The Transformer allows for significantly more parallelization.

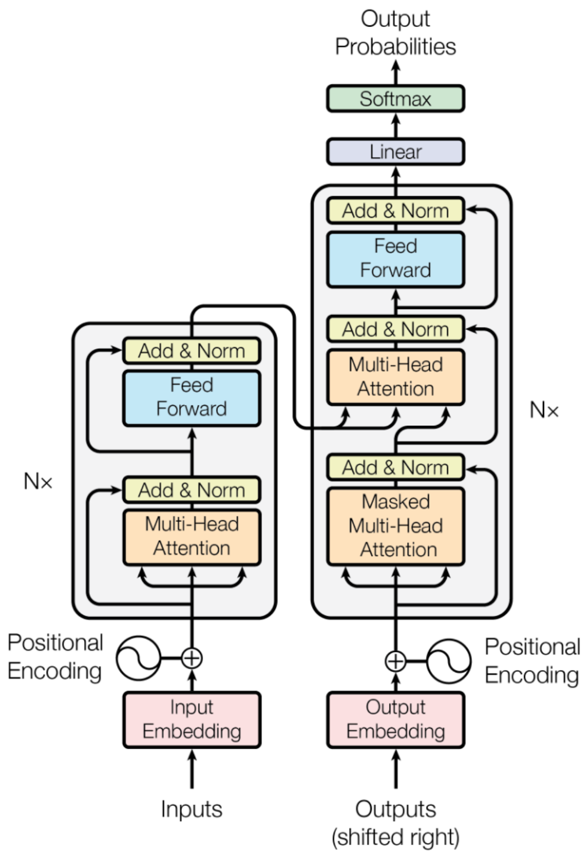

Figure 1.7: This figure shows the Transformers model architecture.

Most competitive neural sequence transduction models have an encoder-decoder structure. Here, the encoder maps an input sequence of symbol representations $(x_1, \ldots, x_n)$ to a sequence of continuous representations $z = (z_1, \ldots, z_n)$. Given z, the decoder then generates an output sequence $(y_1, \ldots, y_m)$ of symbols one element at a time. At each step the model is auto-regressive, consuming the previously generated symbols as additional input when generating the next. The Transformer follows this overall architecture using stacked self-attention and point-wise, fully connected layers for both the encoder and decoder, shown in the left and right halves of Figure 1.7, respectively.

- *Encoder*: The encoder is composed of a stack of $N = 6$ identical layers. Each layer has two sub-layers. The first is a multi-head self-attention mechanism, and the second is a simple, position- wise fully connected feed-forward network. Residual connection are employed around each of the two sub-layers, followed by layer normalization. That is, the output of each sub-layer is LayerNorm(x+Sublayer(x)), where Sublayer(x) is the function implemented by the sub-layer itself. To facilitate these residual connections, all sub-layers in the model, as well as the embedding layers, produce outputs of dimension $dmodel = 512$.

- *Decoder*: The decoder is also composed of a stack of $N = 6$ identical layers. In addition to the two sub-layers in each encoder layer, the decoder inserts a third sub-layer, which performs multi-head attention over the output of the encoder stack. Similar to the encoder, there are residual connections around each of the sub-layers, followed by layer normalization. The self-attention sub-layer is modified in the decoder stack to prevent positions from attending to subsequent positions. This masking, combined with fact that the output embeddings are offset by one position, ensures that the predictions for position i can depend only on the known outputs at positions less than i.

- *Attention*: An attention function can be described as mapping a query and a set of key-value pairs to an output, where the query, keys, values, and output are all vectors. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key.

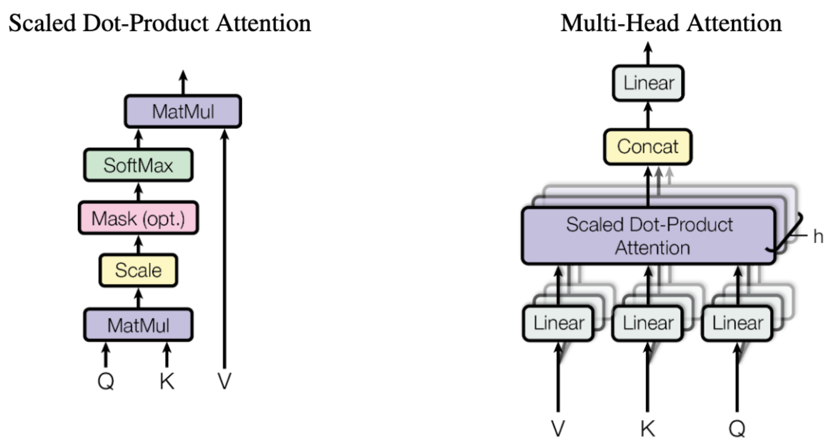

Figure 1.8: Scaled Dot-Product Attention (left). Multi-Head Attention consists of several attention layers running in parallel (right).

1.6 Dataset

Il dataset a nostra disposizione descrive un periodo di quasi due anni (dal 02/02/2022 al 16/06/2023) ed è proveniente da un impianto fotovoltaico da 978 kW situato nella provincia di Bari. È composto da 3 inverter, 27 quadri di campo, 1 contatore, 1 solarimetro, 2 protezioni interfaccia e 1 dispositivo “impianto” nel quale vengono immagazzinati i dati provenienti da Solargis. Il dataset è organizzato in file, uno per tipologia di dispositivo, che rappresentano la singola giornata e i dati sono aggregati a 5 minuti ed ogni riga riporta il riferimento al dispositivo di appartenenza (`deviceName` e `deviceId`).

File Name
`2022_02_02_Rofilo_NP00003174_inverter.csv`
`2022_02_02_Rofilo_NP00003174_meteorology.csv`
`2022_02_02_Rofilo_NP00003174_meter.csv`
`2022_02_02_Rofilo_NP00003174_other.csv`
`2022_02_02_Rofilo_NP00003174_plantDevice.csv`
`2022_02_02_Rofilo_NP00003174_stringbox.csv`
`2022_02_03_Rofilo_NP00003174_inverter.csv`
`2022_02_03_Rofilo_NP00003174_meteorology.csv`
`2022_02_03_Rofilo_NP00003174_meter.csv`
`2022_02_03_Rofilo_NP00003174_other.csv`
`2022_02_03_Rofilo_NP00003174_plantDevice.csv`
`2022_02_03_Rofilo_NP00003174_stringbox.csv`
...

Table 1.1: Estratto di alcuni file del dataset.

timestamp	serial	...	TotalEnergy(kWh)	Frequency(Hz)	deviceid
01/02/2023 00:05	INV01	...	431324.36	49.88	83204
01/02/2023 00:10	INV01	...	431324.36	49.88	83204
01/02/2023 00:15	INV01	...	431324.36	49.88	83204
...	...	...	...	...	...
01/02/2023 12:00	INV01	...	431324.36	49.88	83204
01/02/2023 12:05	INV01	...	431324.36	49.88	83204
01/02/2023 12:10	INV01	...	431324.36	49.88	83204
...	...	...	...	...	...
01/02/2023 23:45	INV01	...	431324.36	49.88	83204
01/02/2023 23:50	INV01	...	431324.36	49.88	83204
01/02/2023 23:55	INV01	...	431324.36	49.88	83204

Table 1.2: Contenuto di un file che rappresenta un inverter

1.6.1 Inverter

Un inverter in un impianto fotovoltaico è un componente fondamentale che svolge una funzione vitale: converte l'energia elettrica prodotta dai pannelli solari, che è in corrente continua (DC), in energia elettrica utilizzabile in corrente alternata (AC). Questa conversione è cruciale perché la maggior parte delle apparecchiature elettriche domestiche e delle reti di distribuzione elettrica utilizzano corrente alternata per funzionare. Regolano la tensione e la frequenza della corrente alternata prodotta per garantire che siano conformi agli standard di rete elettrica locale. Questo è importante per evitare danni agli apparecchi elettrici e per garantire l'interoperabilità con la rete elettrica. Sono spesso dotati di tecnologie avanzate come il "Maximum Power Point Tracking" (MPPT) che ottimizzano costantemente la produzione di energia solare. Questo significa che l'inverter regola la tensione in ingresso dai pannelli solari per ottenere la massima potenza possibile in base alle condizioni di luce solare in tempo reale. Le feature che caratterizzano il nostro impianto sono riassunte dalla seguente tabella:

Name	Unit Symbol	Description
CommunicationCode	-	COMMUNICATION CODE
Failure 3	-	Allarme Attivo
Failure 4	-	Allarme di isolamento
CurrentDC	A	Corrente di campo fotovoltaico
CurrentAC	A	Corrente di Rete
CurrentAC Phase1	A	Corrente RMS di Linea Fase R
CurrentAC Phase2	A	Corrente RMS di Linea Fase S
CurrentAC Phase3	A	Corrente RMS di Linea Fase T
TotalEnergy	kWh	Energia Attiva Erogata
Frequency	Hz	Frequenza di Rete
PowerAC Phase1	kW	PA di Linea Fase R
PowerAC Phase2	kW	PA di Linea Fase S
PowerAC Phase3	kW	PA di Linea Fase T
PowerAC	kW	PA Erogata
PowerDC	kW	Potenza di campo fotovoltaico
Status	-	Stato Inverter
Failure	-	Stato PLL per Aggancio Rete
Failure 2	-	Stato Rete 1
Failure 1	-	Stato Rete 2
InternalTemperature	C	Temp. CPU
HeatSinkTemperature	C	Temp. IGBT
VoltageDC	V	Tensione di campo fotovoltaico
VoltageAC	V	Tensione di Rete
VoltageAC Phase1	V	Tensione RMS di Linea Fase R
VoltageAC Phase2	V	Tensione RMS di Linea Fase S
VoltageAC Phase3	V	Tensione RMS di Linea Fase T

Table 1.3: Lista di tutte le features degli inverter, con descrizione.

1.6.2 StringBox

Le stringbox in un impianto fotovoltaico sono contenitori elettrici progettati per ospitare e proteggere una serie (o stringa) di pannelli solari collegati in serie. Questi dispositivi svolgono diverse funzioni importanti:

- Protezione: Le stringbox includono dispositivi di protezione come interruttori automatici o fusibili che prevengono cortocircuiti e sovraccarichi nel circuito fotovoltaico.

- Connettori: Forniscono connettori sicuri per collegare i cavi provenienti dai pannelli solari alla stringa di cavi principale dell'impianto.
- Monitoraggio: Alcune stringbox sono dotate di sistemi di monitoraggio che consentono di rilevare prestazioni o guasti dei pannelli solari all'interno della stringa.
- Isolamento: Possono anche includere dispositivi di isolamento che consentono di interrompere l'alimentazione elettrica verso la stringa di pannelli solari per scopi di manutenzione o sicurezza.

1.6.3 JunctionBox

Name	**Unit Symbol**	**Description**
CommunicationCode	-	COMMUNICATION CODE
Failure	-	Allarme Stringhe
CurrentString1	A	Corrente I1
CurrentString2	A	Corrente I2
CurrentString3	A	Corrente I3
CurrentString4	A	Corrente I4
CurrentString5	A	Corrente I5
CurrentString6	A	Corrente I6
CurrentString7	A	Corrente I7
AverageStringCurrent	A	Corrente Media
Irradiance	W/m^2	Irraggiamento moduli
Failure 1	-	Stringhe Aperte
Failure 2	-	Stringhe Non Perform.
EnvironmentTemperature	C	Temperatura ambiente
ModuleTemperature	C	Temperatura moduli
InternalTemperature	C	Temperatura Scheda

Table 1.4: Lista di tutte le features deglle JunctionBox, con descrizione.

1.6.4 Solargis

Solargis è una società specializzata nella fornitura di dati e servizi di previsione solare per impianti fotovoltaici e progetti legati all'energia solare. Il loro principale obiettivo è fornire informazioni precise e affidabili sull'irradiazione solare e sulle condizioni meteorologiche solari in qualsiasi parte del mondo. Questi dati sono fondamentali per la progettazione, l'ottimizzazione e la gestione degli impianti fotovoltaici. Solargis raccoglie e fornisce dati dettagliati sull'irradiazione solare globale, diretta e diffusa in ogni posizione geografica. Questi dati consentono agli sviluppatori di impianti fotovoltaici di valutare la quantità di energia solare disponibile in una determinata area, il che è fondamentale per dimensionare correttamente l'impianto e calcolare le previsioni di produzione.

timestamp	...	**SolargisGHI(W/m2)**	**SolargisGTI(W/m2)**
2022-08-01 11:40:00	...	896	978
2022-08-01 11:45:00	...	896	978
2022-08-01 11:50:00	...	896	978
2022-08-01 11:55:00	...	914	1001
2022-08-01 12:00:00	...	914	1001
2022-08-01 12:05:00	...	914	1001
2022-08-01 12:10:00	...	928	1019
2022-08-01 12:15:00	...	928	1019
2022-08-01 12:20:00	...	928	1019

Table 1.5: Dati di Solargis a nostra disposizione, provenienti dal file `2022_08_01_Rofilo_NP00003174_plantDevice.csv`

Solar radiation takes a long journey until it reaches Earths surface. So when modelling solar radiation, various interactions of extra-terrestrial solar radiation with the Earths atmosphere, surface and objects are to be taken into account. The component that is neither reflected nor scattered, and which directly reaches the surface, is called direct radiation; this is the component that produces shadows. The component that is scattered by the atmosphere, and which reaches the ground is called diffuse radiation. The small part of the radiation reflected by the surface and reaching an inclined plane is called the reflected radiation. These three components together create global radiation.

In solar energy applications, the following parameters are commonly used in practice:

- Direct Normal Irradiation/Irradiance (DNI) is the component that is involved in thermal (concentrating solar power, CSP) and photovoltaic concentration technology (concentrated photovoltaic, CPV).
- Global Horizontal Irradiation/Irradiance (GHI) is the sum of direct and diffuse radiation received on a horizontal plane. GHI is a reference radiation for the comparison of climatic zones; it is also essential parameter for calculation of radiation on a tilted plane.

- Global Tilted Irradiation/Irradiance (GTI), or total radiation received on a surface with defined tilt and azimuth, fixed or sun-tracking. This is the sum of the scattered radiation, direct and reflected. It is a reference for photovoltaic (PV) applications, and can be occasionally affected by shadow.

Name	**Unit Symbol**	**Description**
SolargisGHI	W/m^2	Solargis Global Horizontal Irradiation
SolargisGTI	W/m^2	Solargis Global Tilted Irradiation

Table 1.6: Lista di tutte le features di Solargis, con descrizione.