



UNIVERSITÀ DEGLI STUDI DI CAGLIARI

Dipartimento di Fisica

RELAZIONE DI TIROCINIO CURRICULARE

Prototipo di un Autoencoder per la classificazione delle Regioni Attive solari

Svolto presso: Osservatorio Astronomico di Cagliari, Istituto Nazionale di Astrofisica



Tirocinante:

Emilia Serra

Matr. 60/60/66039

Tutor Aziendale:

Dott. Marco Marongiu

Tutor Accademico:

Prof. Alessandro Riggio

Anno Accademico 2025/2026

Sommario: In questo progetto di tirocinio è stato creato un prototipo di Autoencoder per la classificazione delle Regioni Attive nelle mappe solari. Il modello, basato su Reti Neurali Convolutionali, è stato addestrato su mappe solari acquisite tramite i radiotelescopi dell'INAF nell'ambito del progetto SunDish. Per valutare l'accuratezza del prototipo nell'individuazione delle coordinate delle Regioni Attive, i risultati ottenuti dall'Autoencoder sono stati messi a confronto con le classificazioni ufficiali fornite dai cataloghi solari ufficiali inclusi dentro il tool JHelioviewer, calcolandone la distanza euclidea. L'analisi comparativa ha rivelato che, nell'identificazione delle coordinate delle Regioni Attive, il prototipo mostra discrepanze che talvolta superano le dimensioni del beam del ricevitore. Trattandosi di un prototipo, questo lavoro si è rivelato fondamentale per individuare i necessari margini di miglioramento del modello e l'ottimizzazione dei dati di input.

1 Introduzione

Il Sole è una stella con un ciclo di attività della durata di 11 anni [1] che emette lungo tutto lo spettro elettromagnetico, dalle onde radio fino ai raggi gamma [2]. Alle frequenze radio, il Sole è dominato dall'emissione di Sole quieto¹, che ne costituisce lo sfondo pressoché uniforme sull'intera superficie; a questa componente di base si sovrappongono strutture localizzate a maggiore brillantezza, come ad esempio le Regioni Attive. Essendo la materia solare un plasma di ioni ed elettroni liberi, la sua trasparenza e l'osservazione alle radiofrequenze sono regolate dalla frequenza di plasma. Quest'ultima rappresenta la frequenza di oscillazione naturale del mezzo e dipende unicamente dalla densità elettronica locale, crescendo man mano che si scende in profondità verso l'interno della stella. Quando si osserva il Sole quieto in banda radio in condizioni di equilibrio termico locale, la radiazione viene prodotta dal meccanismo di emissione free-free, dove gli elettroni liberi del plasma solare emettono onde radio dopo essere stati deviati dagli ioni. Tuttavia, il plasma non è sempre trasparente: la possibilità che un'onda radio lo attraversi dipende dal confronto diretto tra la frequenza dell'onda usata per l'osservazione e la frequenza di plasma del mezzo. Se la frequenza di osservazione è più bassa di quella di plasma, gli elettroni riescono a muoversi abbastanza velocemente da schermare l'onda elettromagnetica, rendendo il mezzo otticamente spesso, cioè del tutto opaco e impenetrabile come un muro. Al contrario, se viene utilizzata una frequenza di osservazione più alta rispetto alla frequenza locale del plasma, l'onda oscilla troppo rapidamente perché gli elettroni riescano a bloccarla; in questo scenario, il plasma diventa otticamente sottile, comportandosi come un vetro trasparente che si lascia attraversare senza opporre resistenza. Questo significa che per ogni frequenza radio esiste una quota limite precisa, chiamata raggio limite, al di sotto della quale il plasma è troppo denso e opaco per far propagare l'onda verso l'esterno, mentre al di sopra di essa il plasma è otticamente sottile e permette alla radiazione di viaggiare indisturbata fino ai radiotelescopi. Dunque, è possibile scegliere quale strato dell'atmosfera solare visualizzare cambiando la frequenza di osservazione [3, 4].

Lo studio del Sole nell'ambito del radio è possibile grazie al progetto SunDish (PI: A. Pellizzoni), nato da una collaborazione tra l'Istituto Nazionale di Astrofisica (INAF) e l'Agenzia Spaziale Italiana (ASI)[5, 6], e ha come obiettivo quello di sondare gli strati più profondi dell'atmosfera solare nella banda radio. Questo progetto permette di osservare il Sole nella banda K, 18-26 GHz, dove è possibile osservare principalmente la cromosfera e la corona [7].

La cromosfera è uno strato dell'atmosfera solare situato sopra la fotosfera, caratterizzato da emissioni energetiche specifiche nell'ultravioletto e nelle onde radio, spesso legate all'attività magnetica. Il suo nome deriva dalle parole greche colore (cromo) e palla (sfera), e alludono al bordo colorato che si può osservare nel contorno della Luna durante un'eclissi solare. La radiazione che si osserva $\lambda = 656,3\text{nm}$, ovvero la riga H_α della serie di Balmer, è dovuta principalmente all'emissione dell'idrogeno nella transizione dallo stato $n=3$ allo stato $n=2$. La cromosfera può essere osservata anche in assenza di eclissi, utilizzando un filtro centrato in H_α . Questo strato dell'atmosfera solare è caratterizzato da una temperatura che ha un minimo di circa 4.000 K a 500 km sopra la superficie visibile, fino a un aumento della temperatura a 8.000 K a un'altezza di circa 2.000 km [8]. All'interno della cromosfera, le Regioni Attive mostrano le loro fenomenologie più interessanti. Queste sono caratterizzate da un campo magnetico locale molto intenso e possono immagazzinare l'energia necessaria a innescare i brillamenti solari e le emissioni di massa coronale [9]. Le Regioni

¹Con il termine Sole quieto ci si riferisce al Sole in una fase di minima attività o, più nello specifico, alle aree della sua superficie e della sua atmosfera che non sono influenzate da fenomeni magnetici violenti e/o rapidamente variabili.

Attive hanno origine in una zona di transizione chiamata tachocline: un sottile strato interno situato tra la zona radiativa, più interna e caratterizzata da una rotazione uniforme, e la zona convettiva, più esterna e caratterizzata da una rotazione differenziale. Questa zona di transizione si trova al di sotto della fotosfera, la regione del Sole da cui emerge la maggior parte della radiazione visibile. È però nella cromosfera che le Regioni Attive mostrano i loro effetti più interessanti, in quanto appaiono come le zone più brillanti e intense di tutto il disco solare; tuttavia, durante eventi transitori e violenti come i Brillamenti o le Espulsioni di massa coronale, questi ultimi prendono il sopravvento in termini di intensità.

La dinamica intrinseca che governa la transizione tra l'emissione uniforme del Sole quieto e la formazione di aree più brillanti, come le Regioni Attive, rappresenta una sfida per l'identificazione automatica di queste strutture. Con l'obiettivo di superare tali limitazioni attraverso gli strumenti offerti dallo stato dell'arte informatico, questo progetto di tirocinio ha visto la realizzazione di un prototipo di classificatore di Regioni Attive, basato su tecniche di Intelligenza Artificiale (AI) strutturate su reti neurali. Questo prototipo ha fatto uso di mappe solari acquisite dal 2018 a oggi. L'analisi automatizzata dei dati radio è possibile grazie a SUNDARA, un codice sofisticato e completamente autoconsistente [9] in grado di identificare diversi eventi all'interno del disco solare (come le Regioni Attive) e misurarne i flussi e gli indici spettrali. Il prototipo sviluppato durante questo tirocinio è parte di un processo di potenziamento di SUNDARA, attualmente in corso, che verte all'implementazione di tecniche AI per la classificazione di eventi solari all'interno del disco solare.

La struttura della relazione è organizzata come segue. Nel Paragrafo 2 è presente una breve descrizione dei dati e dei radiotelescopi dell'INAF utilizzati. Nel Paragrafo 3 vengono illustrate l'architettura dell'Autoencoder e le tecniche utilizzate per l'analisi. I risultati sono presentati e discussi nel Paragrafo 4, con un'analisi comparativa tra i risultati ottenuti dalla rete neurale e quelli calcolati manualmente. Infine, nel Paragrafo 5 sono riportate le conclusioni.

2 Dati

Le mappe solari sono state acquisite con la tecnica di osservazione *single – dish* in banda K, che esplora le frequenze radio da 18 *GHz* a 26 *GHz*, utilizzando i radiotelescopi dell'INAF [7]:

- Medicina 'Gavril Grueff' Radio Telescope (32 m di diametro)
- Sardina Radio Telescope (SRT; 64 m di diametro)

Le osservazioni solari con il Noto Radio Telescope, situato nel sud della Sicilia, non sono ancora state effettuate in quanto lo strumento è in fase di potenziamento, in particolar modo per la modalità solare.

Le osservazioni con il Grueff Radio Telescope, situato a Medicina, vicino a Bologna, si svolgono in media una volta alla settimana da febbraio 2018. Sono presenti diversi ricevitori che coprono l'intervallo 1,3–26,5 GHz. Il Sole è stato osservato con il dual-feed in banda K, utilizzando principalmente 18,3 e 25,8 GHz come frequenze centrali, le cui corrispondenti dimensioni del fascio (beam) sono rispettivamente di 2,1 e 1,5 arcominuti [3]. Un esempio di mappe acquisite con il Medicina Radio Telescope è mostrato in Fig. 1.

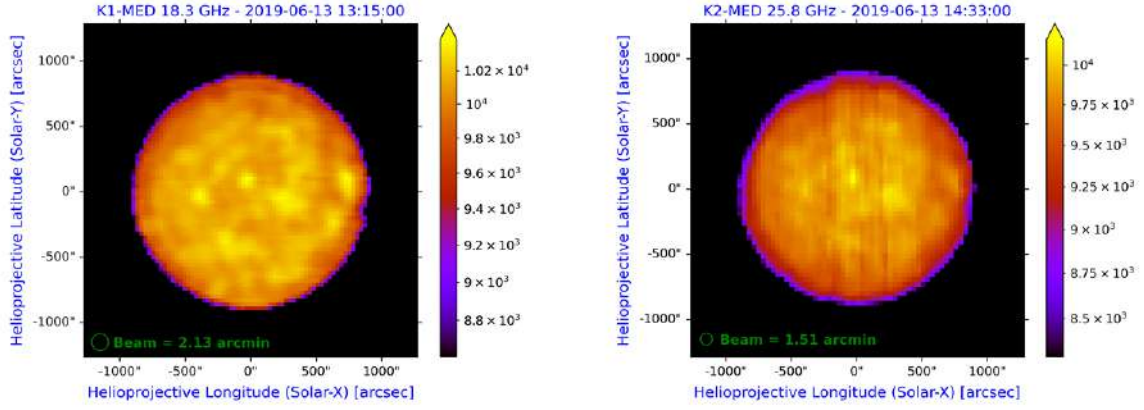


Figura 1: Immagine del disco solare a 18,3 GHz (sinistra) e 25,8 GHz (destra), acquisite con il Grueff Radio Telescope il 13 giugno 2019, processate con SUNDARA [9]. La barra di colore indica la temperatura di brillantezza della mappa del Sole in Kelvin. In entrambe le mappe, il cerchio verde in basso a sinistra indica il beam del ricevitore alle frequenze osservate [7].

Le osservazioni con il Sardinia Radio Telescope (SRT), situato a circa 50 km a nord di Cagliari, si svolgono in media una volta al mese (Fig. 2). Le frequenze prese in esame vanno da 18,8 a 24,7 GHz, corrispondenti a una dimensione del fascio (beam) rispettivamente di 1,0 e 0,8 arcminuti [3], e sono state osservate attraverso un ricevitore a sette feed a doppia polarizzazione [10]. Questo ricevitore è stato adattato per le osservazioni solari [5].

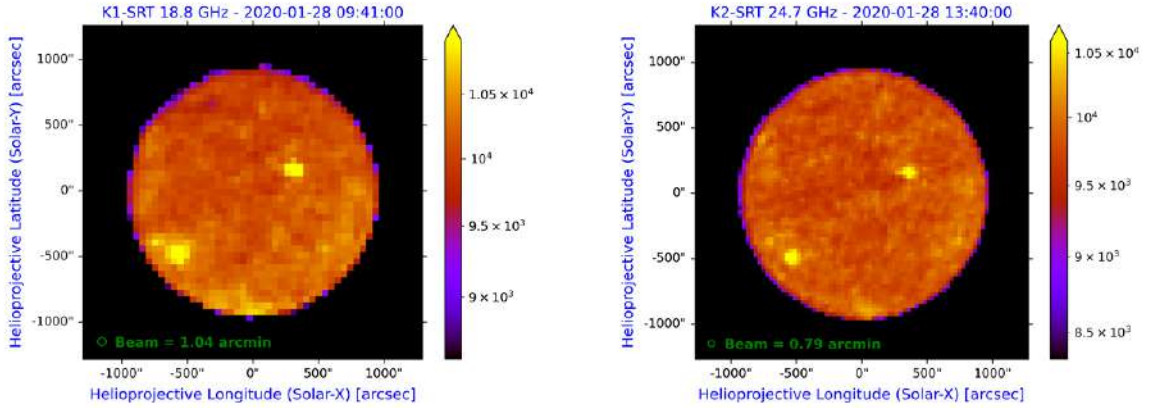


Figura 2: Immagine del disco solare a 18,8 GHz (sinistra) e a 24,7 GHz (destra), acquisite con SRT il 28 gennaio 2020, processate con SUNDARA [9]. Per ulteriori dettagli si veda la descrizione della Fig. 1.

Le mappe solari sono state acquisite utilizzando la scansione on-the-fly (OTF) [10]. L'antenna esegue passaggi consecutivi avanti e indietro lungo strisce parallele, muovendosi nel sistema di coordinate equatoriali celesti Ascensione Retta (AR) o Declinazione (DEC). Durante questo movimento continuo, il sistema acquisisce i dati a intervalli temporali brevi (solitamente 20 ms) e regolari. Infine, i flussi di dati vengono combinati e ri-campionati su una griglia regolare per ricostruire l'immagine finale [11].

La scansione è controllata tramite il sistema di controllo dell'antenna denominato DISCOS [12]. Queste mappe coprono un'area di $1,3^\circ \times 1,3^\circ$ nel cielo a Medicina (Fig. 1) e $1,5^\circ \times 1,5^\circ$ a SRT (Fig. 2) [3].

Le immagini solari sono state estratte da file FITS², che al loro interno non includono solo la matrice delle temperature di brillantezza³ che compongono l'immagine solare, ma anche una struttura più complessa che

²Tipico formato di dati astronomici, un acronimo che sta per Flexible Image Transport System (Sistema Flessibile di Trasporto Immagini)

³La temperatura di brillantezza è la temperatura che un corpo nero dovrebbe avere per emettere la stessa intensità di radiazione (brillantezza) osservata da una sorgente reale a una determinata frequenza ν [13].

permette di avere diverse informazioni per ogni singola immagine. Un file FITS è composto da una o più unità chiamate HDU (Header Data Unit). Ogni unità contiene due parti: l'Header (etichetta), un blocco di testo che descrive l'oggetto, il telescopio, il tempo di esposizione e le coordinate celesti; e i dati effettivi per ricostruire l'immagine.

3 Autoencoder e descrizione del codice

Come già evidenziato nel Paragrafo 1, per migliorare ulteriormente la classificazione delle Regioni Attive e altri eventi del disco solare, il gruppo di ricerca di Radioastronomia Solare dell'Osservatorio Astronomico di Cagliari sta aggiornando SUNDARA tramite l'implementazione di un nuovo prototipo di IA, basato su Reti Neurali Convoluzionali (CNN), all'interno del sistema di riconoscimento dei pattern già in uso. Una CNN è un modello di Deep Learning strutturato per estrarre caratteristiche visive dalle immagini. Nel presente lavoro, questa tecnologia è stata applicata all'interno di una specifica architettura di tipo Autoencoder.

3.1 Autoencoder

L'architettura di un Autoencoder si compone di due parti fondamentali: l'Encoder, che comprime un'immagine in uno spazio latente, e il Decoder, che la ricostruisce cercando di generare un output specifico. In questo progetto, tale struttura è stata impiegata per compiere una segmentazione d'immagine supervisionata. Sebbene l'Autoencoder nasca originariamente come modello di apprendimento non supervisionato, l'approccio qui utilizzato mappa i dati basandosi sulle posizioni delle Regioni Attive di un catalogo esterno (ottenuto tramite osservazioni in EUV); per questo motivo, l'addestramento può essere considerato un ibrido tra l'apprendimento supervisionato e quello non supervisionato.

La scelta di questa architettura è guidata dalla sua capacità di effettuare una classificazione a livello di singolo pixel, operando di fatto una segmentazione semantica dell'immagine. Invece di limitarsi a una ricostruzione fedele dell'input, il modello viene addestrato per localizzare e mappare specifici eventi solari, come le Regioni Attive e i Buchi Coronali. A differenza dei tradizionali algoritmi di classificazione d'immagine — i quali associano un'etichetta globale all'intero frame perdendo la risoluzione spaziale — questo approccio permette di preservare l'informazione morfologica e di posizione di ciascun fenomeno.

Computazionalmente, l'Autoencoder è stato implementato con il seguente codice.

```
1 @tf.keras.utils.register_keras_serializable()
2 class SourceDetect(Model):
3
4     def __init__(self, input_dimensions, output_dimensions):
5         super(SourceDetect, self).__init__()
6         self.input_dimensions = input_dimensions
7         self.output_dimensions = output_dimensions
8         self.encoder = tf.keras.Sequential([
9             layers.Input(shape=input_dimensions),
10
11             layers.Conv2D(filters=64, kernel_size=(3, 3), activation='relu', padding='same',
12               strides=2),
13             layers.MaxPool2D(pool_size=(2, 2)),
14             layers.Conv2D(filters=32, kernel_size=(3, 3), activation='relu', padding='same',
15               strides=2),
16             layers.MaxPool2D(pool_size=(2, 2)),
17             layers.Conv2D(filters=16, kernel_size=(3, 3), activation='relu', padding='same',
18               strides=2),
19
20             layers.Conv2D(filters=8, kernel_size=(3, 3), activation='relu', padding='same',
21               strides=2),
22             layers.Flatten()
23         ])
24         self.n_pixels = np.prod(self.output_dimensions)
25
26         self.decoder = tf.keras.Sequential([
27             layers.Dense(latent_dim, activation='relu'),
28             layers.Dense(int(self.n_pixels // 4), activation='relu'),
```

```

25 layers.Dense(int(self.n_pixels // 2), activation='relu'),
26 layers.Dense(int(self.n_pixels), activation='sigmoid'),
27 layers.Reshape(target_shape=self.output_dimensions)])

```

Listing 1: Codice contenente il modello di rete neurale utilizzato in questo progetto di tirocinio.

Nelle prime due righe viene creato il modello personalizzato, chiamato '*SourceDetect*', utilizzando una libreria tipica del Machine Learning (*TensorFlow*); tale modello viene quindi salvato su disco in modo che il codice sia in grado di ricaricarlo all'occorrenza. Successivamente, il modello viene inizializzato specificando le dimensioni delle immagini in input e in output. Nella riga 11 inizia il processo di Encoder, dove le immagini passano attraverso una serie di "imbuti" chiamati *Conv2D* (livelli convoluzionali) e *MaxPool2D* che scansionano l'immagine per estrarre le caratteristiche visive più importanti e ne riducono la risoluzione geometrica. Alla fine di questo processo, nella riga 18, *layers.Flatten()* trasforma l'immagine compressa contenente le caratteristiche geometriche tridimensionali in un unico vettore monodimensionale. Nella riga 22 inizia il processo di Decoder, che fa l'esatto opposto dell'Encoder: prende il vettore compresso e cerca di ricostruire l'immagine finale. Utilizza i *layers.Dense* per aumentare progressivamente la dimensione, prima una frazione dei pixel totali ($//4$), poi la metà ($//2$), fino ad arrivare al numero esatto di pixel totali *self.n_pixels*. Lo strato finale della rete, implementato tramite il livello *layers.Dense*, utilizza una funzione di attivazione sigmoidea (*sigmoid*). Questa scelta permette di schiacciare i valori di output nell'intervallo continuo $[0, 1]$, interpretando l'attivazione di ciascun pixel in termini di probabilità di appartenenza alla classe target. Infine, nella riga 27, *layers.Reshape* prende il vettore piatto finale e lo rimodella per portarlo a una matrice con le dimensioni geometriche iniziali dell'immagine, incluse nel parametro *output dimensions*.

3.2 Descrizione del codice

Il presente progetto di tirocinio ha visto il riadattamento di una pipeline in linguaggio Python, in corso di sviluppo da Marco Marongiu, per il monitoraggio e l'analisi di molteplici fenomeni all'interno del disco solare (quali Buchi Coronali, Brillamenti e Macchie Solari). Nello specifico, il flusso di lavoro è stato rimodulato e ottimizzato per concentrare l'analisi su un'unica classe target: le Regioni Attive. Il codice è diviso in tre parti principali, descritte in dettaglio nei seguenti paragrafi.

3.2.1 Acquisizione e filtraggio dei dati

Nella prima parte del codice vengono preparati i dati per l'addestramento (training, Paragrafo 3.2.2). Per prima cosa, è stata definita una funzione in grado di filtrare le frequenze specifiche da analizzare. Sono stati importati i file FITS e sono state create due liste differenti: una contenente gli Headers e l'altra le matrici dei dati. Successivamente, a ogni mappa è stato associato il catalogo della mappatura ufficiale delle Regioni Attive fornito dalla NOAA (National Oceanic and Atmospheric Administration), l'agenzia federale statunitense che si occupa del meteo terrestre e dello spazio⁴. Infine, prima di procedere con la fase di training vero e proprio, sono state escluse le mappe solari che presentavano anomalie visive (come, ad esempio, strisciate o artefatti).

Il dataset finale è formato da 193 mappe di Medicina, acquisite dal 2018 a oggi a una frequenza compresa tra i 18,0 GHz e i 18,9 GHz.

3.2.2 Training

L'addestramento dell'Autoencoder è stato eseguito per un totale di 500 epoche⁵. In linea con le pratiche standard del Machine Learning, per valutarne le prestazioni il dataset è stato suddiviso in due parti: l'80% è stato utilizzato per addestrare la Rete Neurale, mentre il restante 20% (set di validazione o convalida) è stato riservato come test set per valutare la capacità di generalizzazione del modello su dati non visti in precedenza. Le mappe, acquisite con SRT ad altre frequenze in banda K, sono state impiegate per la

⁴<https://www.spaceweather.gov/>

⁵Nel machine learning, un'epoca definisce un ciclo computazionale completo in cui l'algoritmo di ottimizzazione esegue un passaggio sull'intero dataset di addestramento, campionando ogni istanza per l'aggiornamento dei parametri del modello [14].

validazione esterna del modello su dati non noti. Si è scelto di testare questo specifico approccio in quanto configurato come un prototipo preliminare, sviluppato appositamente per le finalità di questo tirocinio. Durante questo processo, la libreria *Keras* permette di monitorare l'evoluzione di due metriche fondamentali, calcolate simultaneamente sia sui dati di training sia su quelli di validazione:

- **Loss Function (Funzione di Perdita):** quantifica l'errore commesso dal modello, misurando quanto l'immagine ricostruita dal decoder si discosti dalla mappa originale delle Regioni Attive della NOAA; l'obiettivo dell'algoritmo è ridurla il più possibile.
- **Accuracy (Accuratezza):** misura la precisione della ricostruzione a livello di singolo pixel. Indica la percentuale di pixel che l'Autoencoder ha classificato e posizionato correttamente, distinguendo accuratamente le strutture delle Regioni Attive dal Sole quieto.

Il confronto diretto tra l'andamento di queste metriche sul set di addestramento e su quello di validazione è cruciale. Analizzare come variano le curve di Training Loss/Accuracy rispetto alle rispettive controparti di Validation Loss/Accuracy permette di capire se l'Autoencoder sta imparando correttamente, oppure se sta riscontrando problemi di underfitting (apprendimento insufficiente) o di overfitting (fenomeno per cui il modello memorizza i dati di addestramento anziché apprendere le caratteristiche generali, compromettendo la capacità di generalizzare su nuove immagini).

L'andamento di queste funzioni nel corso delle 500 epoche è illustrato in Figura 3, dove si osserva come la Loss decresca rapidamente tendendo a zero, mentre l'Accuracy converga progressivamente verso il valore 1, per poi stabilizzarsi entrambe nelle epoche successive. Questo andamento standard è una prima chiara evidenza del corretto apprendimento del modello. Un'ulteriore e fondamentale conferma è data dal comportamento delle curve di validazione (Validation Loss e Validation Accuracy): il fatto che seguano da vicino l'andamento delle rispettive curve di training, senza mostrare divergenze o inversioni di tendenza, dimostra che l'Autoencoder non è incorso in fenomeni di overfitting e possiede un'ottima capacità di generalizzazione.

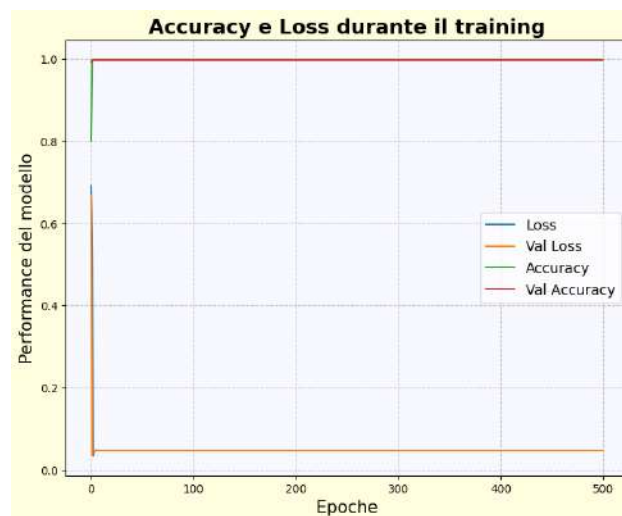


Figura 3: Andamento delle funzioni che mostrano la prestazione dell'addestramento.

3.2.3 Analyser

Il flusso di lavoro del sistema di analisi (analyser) si articola in tre fasi principali. In primo luogo, è stata sviluppata una funzione specifica che riceve in input le mappe solari radio grezze in formato FITS e ne esegue il pre-processing, che comprende la pulizia del rumore e il condizionamento matematico dei dati, al fine di renderle idonee all'analisi da parte dell'Autoencoder. Successivamente, sono state generate le mappe di confidenza, ovvero rappresentazioni grafiche create a partire dalle predizioni dell'Autoencoder. Queste mappe consentono di rilevare automaticamente le strutture solari di interesse – nello specifico, le Regioni Attive – evidenziando visivamente la loro esatta localizzazione sul disco solare. Infine, l'intera pipeline è

stata testata analizzando 13 nuove mappe radio. I risultati qualitativi e quantitativi ottenuti da questa elaborazione saranno discussi nel dettaglio nel Paragrafo 4.

4 Risultati

In questo paragrafo vengono riportati i risultati ottenuti, selezionando le 5 migliori mappe. Ogni mappa è presentata secondo questo schema:

- confronto visivo tra la mappa solare originale e la mappa delle Regioni Attive identificate dal modello;
- mappa di confidenza associata alle predizioni;
- tabella comparativa contenente: le coordinate delle Regioni Attive stimate dall'Autoencoder, la probabilità di presenza di una Regione Attiva nell'area d'interesse, le coordinate di riferimento ufficiali (estratte tramite il software di imaging **SAOImage DS9**) presenti su *JHelioviewer* [15] e, infine, la distanza angolare tra le coordinate stimate e quelle ufficiali.

Le coordinate delle Regioni Attive e gli assi nelle immagini sono espresse in arcosecondi. L'incertezza associata alle coordinate stimate dall'Autoencoder è associata alla dimensione di un singolo pixel (pari a ± 42 arcsec), mentre l'errore associato alle coordinate di riferimento estratte con DS9 è definito dal beam del ricevitore utilizzato per l'acquisizione della mappa solare. Quest'ultimo è calcolato mediante la nota relazione:

$$beam = 1,22 \frac{\lambda}{D} \quad (1)$$

dove 1,22 è un coefficiente dovuto alla struttura del beam pattern del ricevitore, λ (m) è la lunghezza d'onda a cui viene fatta l'osservazione e D (m) è il diametro della parabola. L'unità di misura associata al beam è l'arcosecondo.

Infine, l'errore associato alla distanza tra le coordinate delle Regioni Attive ricostruite e le coordinate ufficiali è stato calcolato propagando l'errore.

Ogni mappa solare è stata nominata secondo questo criterio: nome del radiotelescopio, la data, l'orario (UTC) e la frequenza con cui è stata acquisita.

4.1 SUN SRT 2025-02-12 09:22:51 18,8 GHz

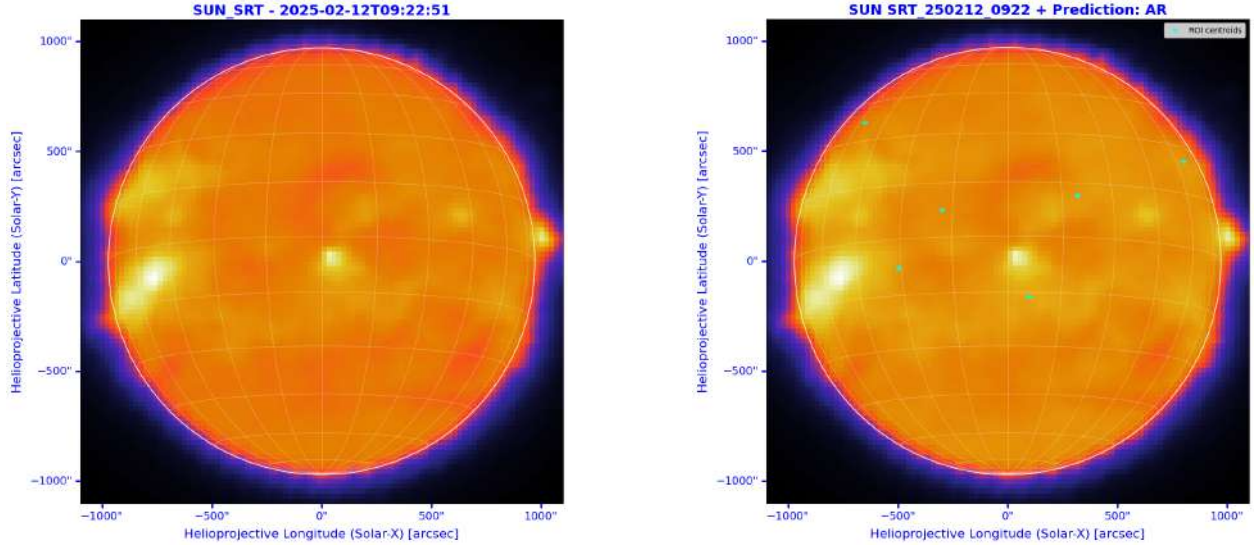


Figura 4: Confronto visivo tra la mappa solare originale (sinistra) e la mappa delle Regioni Attive identificate attraverso l'Autoencoder (destra) relativo a SUN SRT 2025-02-12 09:22:51 18,8 GHz.

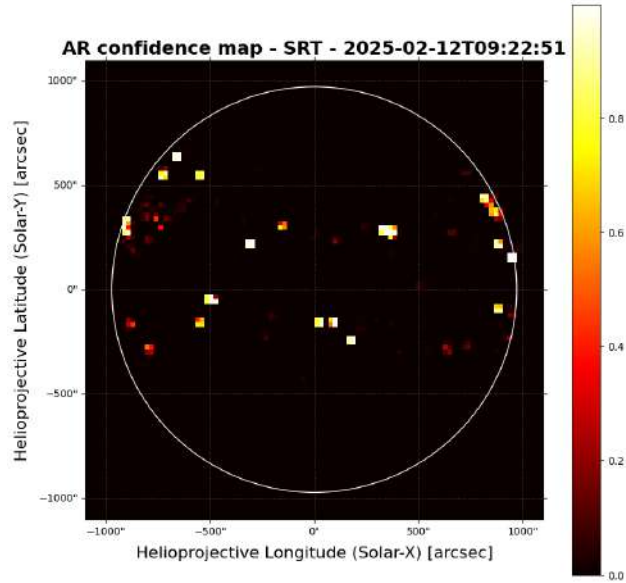


Figura 5: Mappa di confidenza relativa a SUN SRT 2025-02-12 09:22:51 18,8 GHz

X (arcsec)	Y (arcsec)	Probabilità	x DS9 (arcsec)	y DS9 (arcsec)	Distanza (arcsec)
$+319 \pm 42$	$+297 \pm 42$	99,4 %	$+280 \pm 60$	$+280 \pm 60$	90 ± 73
$+99 \pm 42$	-165 ± 42	99,2 %	$+110 \pm 60$	$+10 \pm 60$	160 ± 73
-649 ± 42	$+627 \pm 42$	99,1 %	-520 ± 60	$+450 \pm 60$	220 ± 73
-495 ± 42	-33 ± 42	99,1 %	-500 ± 60	-180 ± 60	140 ± 73
-297 ± 42	$+231 \pm 42$	99,1 %	-150 ± 60	$+80 \pm 60$	210 ± 73
$+803 \pm 42$	$+451 \pm 42$	94,1 %	$+880 \pm 60$	$+320 \pm 60$	150 ± 73

Tabella 1: Confronto tra le coordinate stimate dall'Autoencoder e quelle rilevate tramite DS9 relativo a SUN SRT 2025-02-12 09:22:51 18,8 GHz.

4.2 SUN SRT 2025-02-12 11:16:54 24,7 GHz

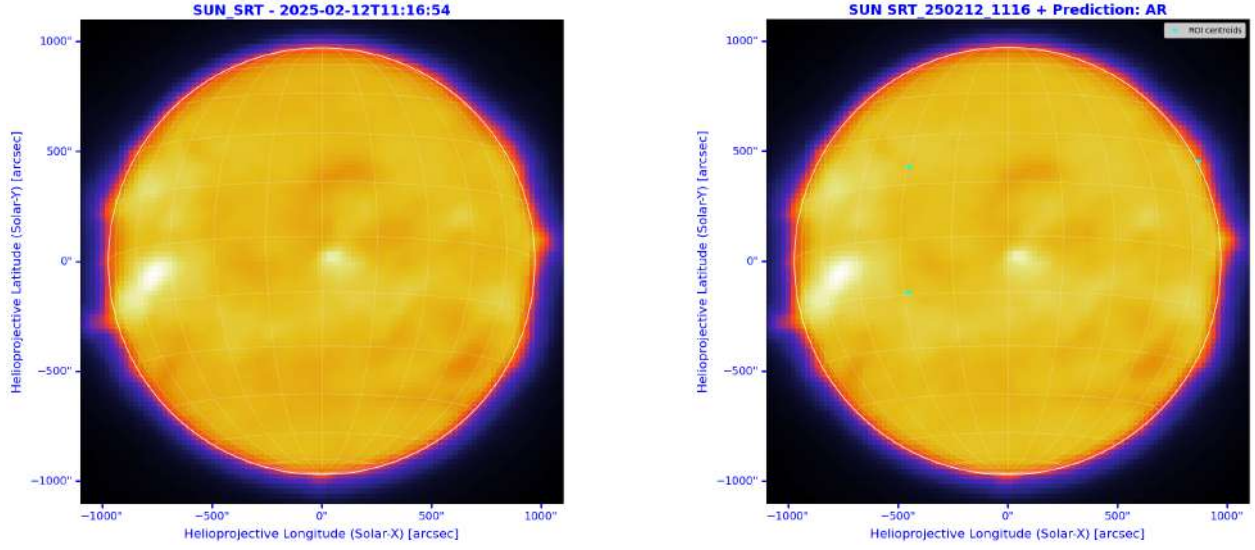


Figura 6: Confronto visivo tra la mappa solare originale (sinistra) e la mappa delle Regioni Attive identificate attraverso l'Autoencoder (destra) relativo a SUN SRT 2025-02-12 11:16:54 24,7 GHz.

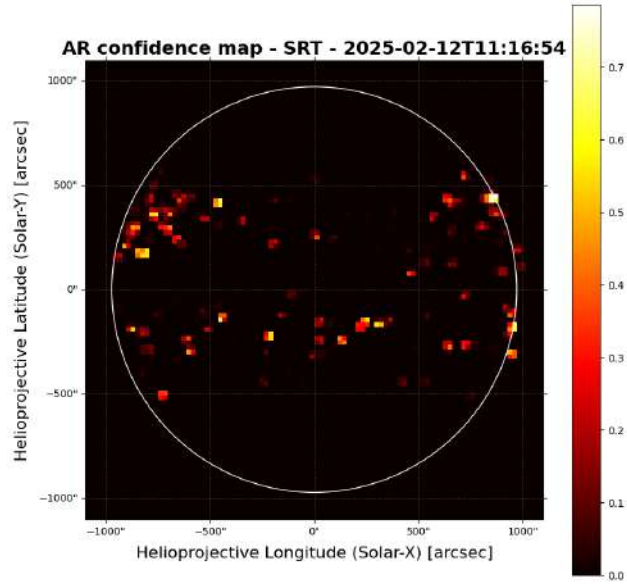


Figura 7: Mappa di confidenza relativa a SUN SRT 2025-02-12 11:16:54 24,7 GHz.

X (arcsec)	Y (arcsec)	Probabilità	x DS9 (arcsec)	y DS9 (arcsec)	Distanza (arcsec)
$+869 \pm 42$	$+451 \pm 42$	78,6 %	$+890 \pm 50$	$+120 \pm 50$	330 ± 65
-451 ± 42	-143 ± 42	66,9 %	-480 ± 50	-210 ± 50	70 ± 65
-451 ± 42	$+429 \pm 42$	64,3 %	-440 ± 50	$+450 \pm 50$	30 ± 65

Tabella 2: Confronto tra le coordinate stimate dall'Autoencoder e quelle rilevate tramite DS9 relativo a SUN SRT 2025-02-12 11:16:54 24,7 GHz.

4.3 SUN SRT 2025-08-19 11:44:14 24,7 GHz

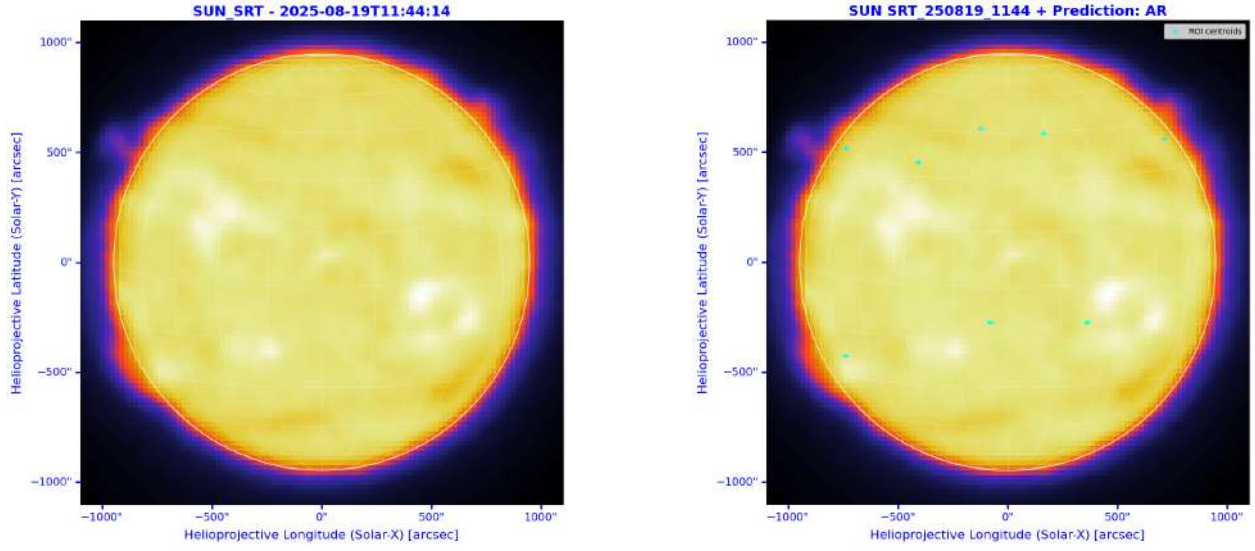


Figura 8: Confronto visivo tra la mappa solare originale (sinistra) e la mappa delle Regioni Attive identificate attraverso l'Autoencoder (destra) relativo a SUN SRT 2025-08-19 11:44:14 24,7 GHz.

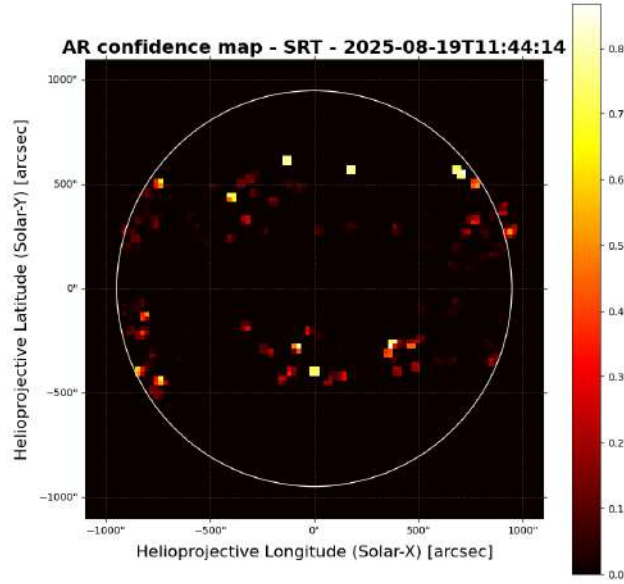


Figura 9: Mappa di confidenza relativa a SUN SRT 2025-08-19 11:44:14 24,7 GHz.

X (arcsec)	Y (arcsec)	Probabilità	x DS9 (arcsec)	y DS9 (arcsec)	Distanza (arcsec)
+715 ± 42	+561 ± 42	86,8 %	+760 ± 50	+450 ± 50	110 ± 65
+363 ± 42	-275 ± 42	83,0 %	+330 ± 50	-280 ± 50	30 ± 65
+165 ± 42	+583 ± 42	79,3 %	-4 ± 50	+370 ± 50	270 ± 65
-121 ± 42	+605 ± 42	78,8 %	-4 ± 50	+370 ± 50	260 ± 65
-77 ± 42	-275 ± 42	78,7 %	-250 ± 50	-350 ± 50	180 ± 65
-407 ± 42	+451 ± 42	72,6 %	-340 ± 50	+410 ± 50	40 ± 65
-737 ± 42	-429 ± 42	69,4 %	-710 ± 50	-400 ± 50	40 ± 65
-737 ± 42	+517 ± 42	64,3 %	-690 ± 50	+450 ± 50	80 ± 65

Tabella 3: Confronto tra le coordinate stimate dall'Autoencoder e quelle rilevate tramite DS9 relativo a SUN SRT 2025-08-19 11:44:14 24,7 GHz.

4.4 SUN SRT 2025-08-20 07:04:23 18,8 GHz

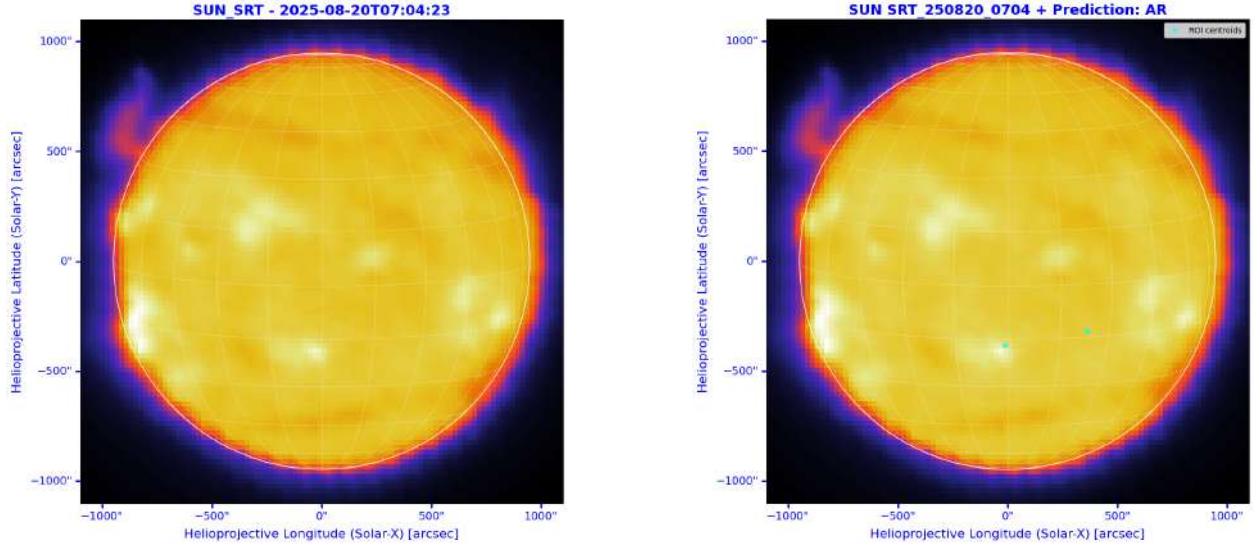


Figura 10: Confronto visivo tra la mappa solare originale (sinistra) e la mappa delle Regioni Attive identificate attraverso l'Autoencoder (destra) relativo a SUN SRT 2025-08-20 07:04:23 18,8 GHz.

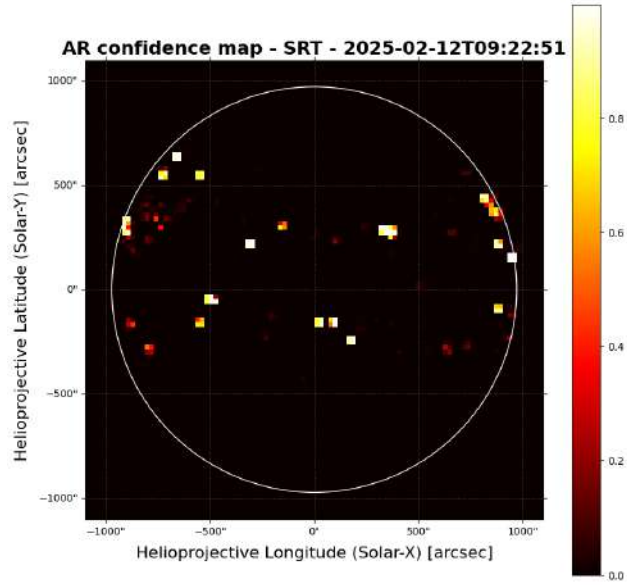


Figura 11: Mappa di confidenza relativa a SUN SRT 2025-08-20 07:04:23 18,8 GHz.

X (arcsec)	Y (arcsec)	Probabilità	x DS9 (arcsec)	y DS9 (arcsec)	Distanza (arcsec)
-11 ± 42	-385 ± 42	69,1 %	-30 ± 60	-420 ± 60	40 ± 73
$+363 \pm 42$	-385 ± 42	60,7 %	$+640 \pm 60$	-320 ± 60	290 ± 73

Tabella 4: Confronto tra le coordinate stimate dall'Autoencoder e quelle rilevate tramite DS9 relativo a SUN SRT 2025-08-20 07:04:23 18,8 GHz.

4.5 SUN MED 2026-04-26 10:27:21 25,5 GHz

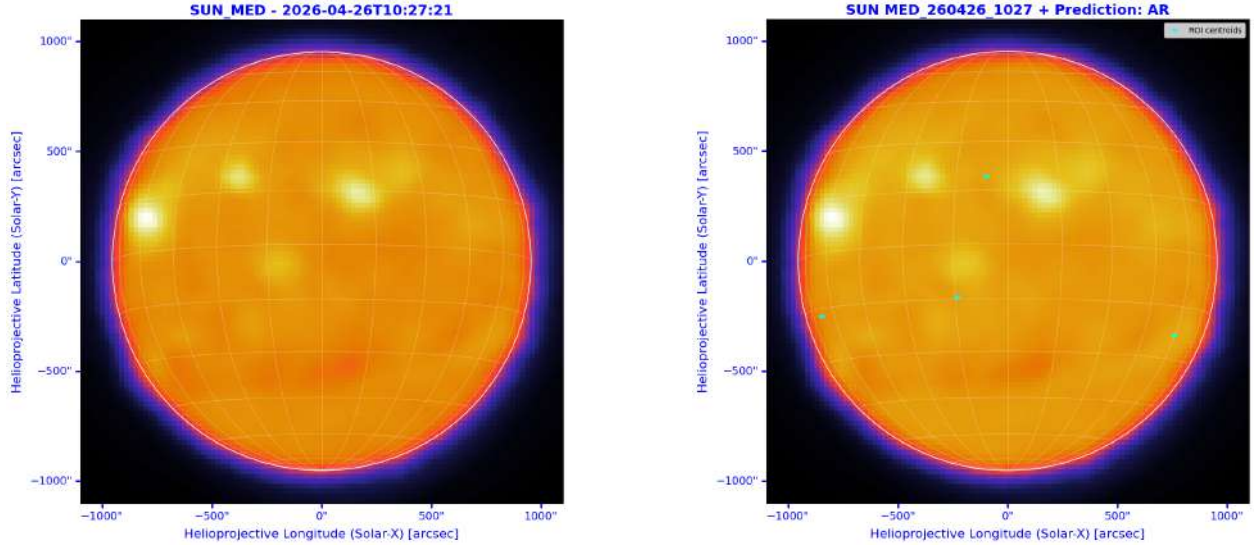


Figura 12: Confronto visivo tra la mappa solare originale (sinistra) e la mappa delle Regioni Attive identificate attraverso l'Autoencoder (destra) relativo a SUN MED 2026-04-26 10:27:21 25,5 GHz.

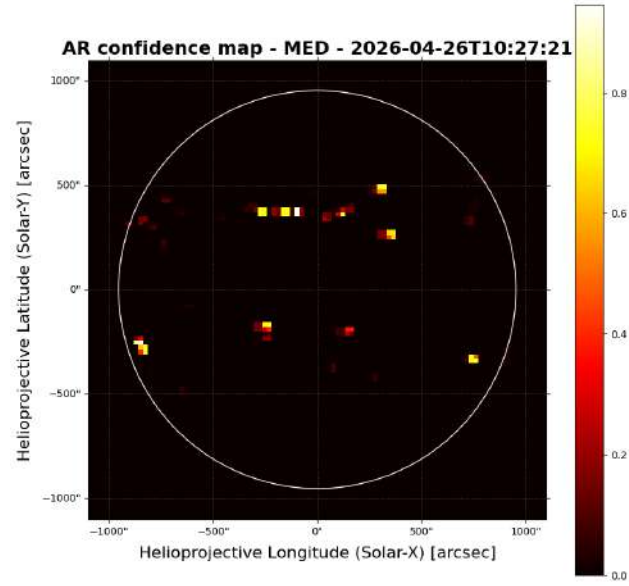


Figura 13: Mappa di confidenza relativa a SUN MED 2026-04-26 10:27:21 25,5 GHz.

X (arcsec)	Y (arcsec)	Probabilità	x DS9 (arcsec)	y DS9 (arcsec)	Distanza (arcsec)
-847 ± 42	-253 ± 42	94,7 %	-840 ± 90	$+50 \pm 90$	310 ± 99
-99 ± 42	$+385 \pm 42$	94,1 %	-30 ± 90	$+460 \pm 90$	100 ± 99
$+759 \pm 42$	-341 ± 42	77,6 %	$+840 \pm 90$	$+180 \pm 90$	530 ± 99
-231 ± 42	-165 ± 42	73,3 %	-400 ± 90	-30 ± 90	220 ± 99

Tabella 5: Confronto tra le coordinate stimate dall'Autoencoder e quelle rilevate tramite DS9 relativo a SUN MED 2026-04-26 10:27:21 25,5 GHz.

4.6 Discussione dei risultati

Dall'analisi comparativa tra le coordinate stimate dall'Autoencoder e quelle di riferimento estratte tramite **SAOImage DS9** (riportate nelle Tabelle 1, 2, 3, 4, 5), si osserva che il prototipo mostra discrepanze variabili nella localizzazione delle Regioni Attive. In particolare, confrontando le distanze calcolate con il beam del ricevitore, se in alcuni casi la discrepanza rientra nell'incertezza sperimentale complessiva, in altri si registrano scostamenti significativi che superano la dimensione del beam.

Queste discrepanze non necessariamente indicano un fallimento del modello, ma possono essere attribuite a diversi fattori sistematici e fisici:

1. **Estensione spaziale delle Regioni Attive:** le Regioni Attive non sono sorgenti puntiformi, ma strutture altamente estese e morfologicamente irregolari. Se da un lato l'analisi manuale tramite **DS9** tende a identificare il centro geometrico o il picco di massima emissione visiva, l'Autoencoder stima il baricentro basandosi sulla distribuzione di probabilità dei pixel attivati, portando a definizioni intrinsecamente diverse dello stesso centroide.
2. **Effetti di proiezione al bordo del disco solare:** da queste prime analisi, le discrepanze maggiori sembrano verificarsi in prossimità del bordo del disco solare (il cui raggio è di circa 960 arcsec), dove gli effetti di proiezione geometrica schiacciano le strutture, rendendo la determinazione delle coordinate estremamente complessa sia per l'algoritmo sia per l'analisi manuale.
3. **Problemi di astrometria nelle mappe solari:** durante la fase di training sono emerse problematiche relative all'astrometria delle mappe. L'individuazione di tale anomalia in corso d'opera non costituisce un limite del lavoro, ma bensì un valore aggiunto: progetti di questo genere, infatti, hanno il merito di far emergere indicazioni preziose per ottimizzare e innalzare lo standard qualitativo delle osservazioni radio solari. Il processo di correzione astrometrica è attualmente in fase di sviluppo e si prevede possa incrementare in modo significativo le prestazioni future dell'Autoencoder.
4. **Natura prototipale del modello:** essendo un prototipo sviluppato per un contesto specifico, l'analisi risente inevitabilmente del dataset particolarmente ridotto, composto da meno di 200 mappe solari, utilizzato in fase di addestramento. Solitamente le tecniche di Machine Learning richiedono una mole ingente di dati (almeno migliaia) per questa fase; l'uso di campioni limitati espone il modello al rischio di overfitting. Inoltre, l'efficacia del modello dipende strettamente dall'accuratezza delle maschere target, motivo per cui la risoluzione dei problemi astrometrici sopracitati è fondamentale. Una tecnica per ovviare al numero di mappe solari disponibili è quella del Data Augmentation, che permette di espandere artificialmente il dataset applicando trasformazioni geometriche ai dati esistenti. Ad esempio, ruotando una singola mappa solare con intervalli di 10° , è possibile generare 36 immagini distinte da un unico input originario. Ulteriori variazioni includeranno operazioni di flipping (riflessione speculare) rispetto agli assi cartesiani.

5 Conclusioni

Il presente lavoro di tirocinio ha riguardato la caratterizzazione e la valutazione di un prototipo basato su architetture di Autoencoder, finalizzato all'identificazione e alla localizzazione automatica delle Regioni Attive su mappe radio solari nell'ambito del progetto **SunDish**. Il confronto sistematico tra le coordinate stimate dal modello e quelle di riferimento, estratte manualmente tramite il software **SAOImage DS9**, ha evidenziato discrepanze talvolta superiori alla dimensione del beam del ricevitore. L'analisi sistematica di queste discrepanze non sminuisce l'efficacia del modello, ma consente di ricondurle a precise cause fisiche, geometriche e sistematiche.

In primo luogo, sono emersi fattori legati alla natura stessa delle sorgenti e alla geometria di osservazione, quali l'estensione morfologica irregolare delle Regioni Attive, che rende intrinsecamente complessa la definizione univoca di un centroide, e i forti effetti di proiezione al bordo del disco solare. In secondo luogo, sono state isolate due limitazioni tecniche cruciali: la presenza di anomalie astrometriche nelle mappe originarie e la ridotta dimensione del dataset di addestramento (inferiore a 200 mappe).

Proprio l'identificazione delle problematiche astrometriche rappresenta uno dei principali valori aggiunti di questo lavoro, avendo avviato lo sviluppo di una procedura di correzione sistematica attualmente in fase di implementazione, che consentirà di incrementare in modo significativo la qualità dei dati di input.

Per superare i limiti evidenziati e condurre il modello verso una maturità operativa, gli sviluppi futuri si concentreranno su:

- **Risoluzione delle anomalie astrometriche** e ottimizzazione delle maschere target per l'addestramento.
- **Espansione del dataset** tramite tecniche di *Data Augmentation* (rotazioni e riflessioni speculari delle mappe esistenti), riducendo drasticamente il rischio di overfitting e migliorando la capacità di generalizzazione dell'Autoencoder.

In conclusione, sebbene il prototipo necessiti di ulteriori fasi di affinamento e un addestramento su scala più ampia prima di un impiego sul campo, la metodologia strutturata e l'analisi critica condotte in questa sede pongono solide e promettenti basi per lo sviluppo di sistemi di monitoraggio dell'attività solare completamente automatizzati.

References

- [1] Wolfgang Gleissberg. "Ascent and descent in the eighty-year cycles of solar activity." In: *J. Brit. Astron. Ass.*, 76: 265-8 (June 1966). (1965).
- [2] Markus J Aschwanden. *Physics of the solar corona: an introduction with problems and solutions*. Springer, 2006.
- [3] Marco Marongiu et al. "Study of solar brightness profiles in the 18–26 GHz frequency range with INAF radio telescopes-I. Solar radius". In: *Astronomy & Astrophysics* 684 (2024), A122.
- [4] Christian Vocks et al. "LOFAR observations of the quiet solar corona". In: *Astronomy & Astrophysics* 614 (2018), A54.
- [5] A Pellizzoni et al. "Solar observations with single-dish INAF radio telescopes: continuum imaging in the 18–26 GHz range". In: *Solar Physics* 297.7 (2022), p. 86.
- [6] Christina Plainaki et al. "Current state and perspectives of Space Weather science in Italy". In: *Journal of Space Weather and Space Climate* 10 (2020), p. 6.
- [7] Marco Marongiu et al. "Study of solar brightness profiles in the 18–26 GHz frequency range with INAF radio telescopes-II. Evidence of coronal emission". In: *Astronomy & Astrophysics* 684 (2024), A123.
- [8] Mats Carlsson, Bart De Pontieu, and Viggo H Hansteen. "New view of the solar chromosphere". In: *Annual Review of Astronomy and Astrophysics* 57.1 (2019), pp. 189–226.
- [9] Marco Marongiu et al. *A Python approach for solar data analysis: SUNDARA (SUNDish Active Region Analyser)*. Tech. rep. Preliminary Development, Tech. Rep. 81, 2021.
- [10] Isabella Prandoni et al. "The Sardinia radio telescope-from a technological project to a radio observatory". In: *Astronomy & Astrophysics* 608 (2017), A40.
- [11] JG Mangum, DTn Emerson, and EW Greisen. "The on the fly imaging technique". In: *Astronomy & Astrophysics* 474.2 (2007), pp. 679–687.
- [12] Marco Bartolini et al. "DISCOS Project Status and Evolution Towards Continuous Integration". In: *Astronomical Data Analysis Software and Systems XXVI* 521 (2019), p. 193.
- [13] Costas E Alissandrakis, TS Bastian, and Roman Brajša. "The quiet sun at mm wavelengths as seen by ALMA". In: *Frontiers in Astronomy and Space Sciences* 9 (2022), p. 981320.
- [14] Léon Bottou. "Large-scale machine learning with stochastic gradient descent". In: *Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics Paris France, August 22-27, 2010 Keynote, Invited and Contributed Papers*. Springer. 2010, pp. 177–186.
- [15] D Müller et al. "JHelioviewer-Time-dependent 3D visualisation of solar and heliospheric data". In: *Astronomy & Astrophysics* 606 (2017), A10.