

UNIVERSITA' DEGLI STUDI DI VERONA



Corso di Laurea in Bioinformatica

Elaborato finale

**CLASSIFICAZIONE DI FENOMENI
SISMICI VULCANICI CON TECNICHE
DI TRANSFER LEARNING**

Relatore

Manuele BICEGO

Candidato

Luca Antonelli

Marcello Cilluffo

Anno Accademico 2018 – 2019

INDICE

SOMMARIO	3
CAPITOLO 1 – INTRODUZIONE	4
<i>1.1 Descrizione dei fenomeni sismici</i>	<i>4</i>
<i>1.2 Le classi dei terremoti vulcanici</i>	<i>5</i>
<i>1.3 Utilizzo della pattern recognition nei fenomeni sismici vulcanici</i>	<i>5</i>
CAPITOLO 2 – BACKGROUND.....	8
<i>2.1 Rappresentazione: FFT, spettrogramma, spettrogramma medio, PCA</i>	<i>8</i>
<i>2.2 Il Transfer Learning</i>	<i>9</i>
<i>2.2.1 Il transfer learning negativo</i>	<i>10</i>
<i>2.3 Classificatori utilizzati</i>	<i>11</i>
<i>2.4 Validazione di un classificatore.....</i>	<i>11</i>
<i>2.4.1 Validazione utilizzata nello studio</i>	<i>12</i>
CAPITOLO 3 – METODOLOGIA PROPOSTA	13
<i>3.1 Domain adaptation utilizzata nella tecnica di mapping</i>	<i>13</i>
<i>3.2 Pipeline creata per la realizzazione delle metodologie proposte</i>	<i>15</i>
CAPITOLO 4 – RISULTATI SPERIMENTALI.....	16
<i>4.1 Analisi del dataset</i>	<i>16</i>
<i>4.2 Esperimenti effettuati.....</i>	<i>16</i>
<i>4.2.1 Esperimento 1.....</i>	<i>17</i>
<i>4.2.2 Esperimento 2.....</i>	<i>20</i>
<i>4.2.3 Esperimento 3.....</i>	<i>21</i>
<i>4.3 Commenti</i>	<i>21</i>
CAPITOLO 5 – CONCLUSIONI	23
BIBLIOGRAFIA	24

SOMMARIO

Questo lavoro si colloca nel contesto della Pattern Recognition, un campo di ricerca volto alla classificazione di oggetti, o gruppi di oggetti, in determinate classi di appartenenza.

Essendo un processo automatizzato, la pattern recognition trova diverse applicazioni come nel caso dell'analisi e identificazione di fenomeni sismici vulcanici.

Per l'ottimizzazione di questi processi ci si avvale di una particolare ed avanzata tecnica di learning chiamata Transfer Learning che permette di sfruttare la conoscenza acquisita precedentemente per la risoluzione di nuovi problemi correlati ad esso.

Gli esperimenti riportati sono stati effettuati su dati di fenomeni sismici vulcanici provenienti dal vulcano Nevado del Ruiz situato in Colombia utilizzando diversi classificatori e analizzando gli errori ottenuti.

CAPITOLO 1 – INTRODUZIONE

In questo capitolo verrà introdotto il contesto applicativo ovvero l'insieme di concetti e nozioni utili alla comprensione degli argomenti in esame. Verrà poi approfondita l'analisi dei segnali sismici e la loro distinzione nelle diverse classi ed infine verrà analizzata la pipeline di pattern recognition applicata in questo contesto.

1.1 Descrizione dei fenomeni sismici

Lo scaturimento di un terremoto è causato dalla propagazione di onde sismiche, particolari onde elastiche che si diffondono in direzioni diverse sulla base delle proprietà elastiche delle rocce. Esse si differenziano per modo, velocità e direzione di propagazione, infatti possono essere distinte in due principali tipologie di onde sismiche: onde di corpo, o di volume, e onde superficiali.

Le prime possono essere ulteriormente distinte in onde longitudinali e onde trasversali. Le onde longitudinali sono anche chiamate onde primarie (onde P) e si propagano all'interno del corpo roccioso deformandone il volume. Essendo le più veloci sono le prime ad essere rilevate in superficie per mezzo dei sismografi. Vengono chiamate onde longitudinali in quanto provocano, nel materiale attraversato, oscillazioni parallele alla direzione di propagazione. Sono inoltre in grado di propagarsi attraverso un materiale liquido come ad esempio il magma. Le onde trasversali vengono anche chiamate onde secondarie (onde S) in quanto sono più lente delle onde P e quindi raggiungono la superficie dopo queste ultime. Contrariamente alle onde P, le onde S provocano nel materiale attraversato delle oscillazioni dal basso verso l'alto e viceversa in maniera perpendicolare alla direzione di propagazione. Inoltre, esse non sono in grado di diffondersi nei materiali liquidi. Le oscillazioni più attenuate che susseguono i picchi delle onde P ed S sono chiamate coda. Dal momento che però la coda delle onde P è nascosta dalle onde S, solitamente con l'espressione coda ci si riferisce alla coda delle onde S (Fig. 1).

Le onde superficiali, al contrario delle onde di corpo, si originano quando queste ultime interagiscono con la superficie terrestre e si estendono su di essa.

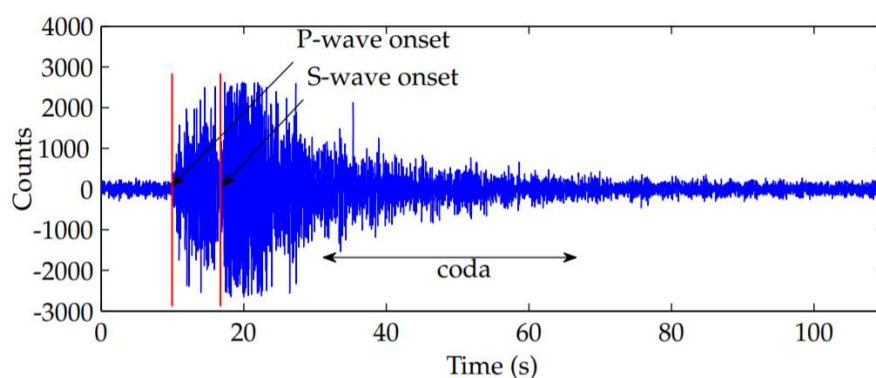


Fig. 1 Parte di un segnale sismico [1]

1.2 Le classi dei terremoti vulcanici

In generale i fenomeni sismici vulcanici vengono classificati in diverse classi in virtù della durata e della modalità d'origine. Tali classi sono:

- *Volcano tectonic* (VT): questi fenomeni sono associati a rotture nel sottosuolo terrestre; il sismografo rileva dei chiari picchi costituiti dalle onde P ed S ed una breve coda che può durare dai 7 ai 15 secondi;
- *Long Period* (LP): questi eventi sismici vengono generati a causa di cambiamenti di pressione nei canali contenenti magma e fluidi idrotermali; in questo caso il sismografo non rileva dei chiari e nitidi picchi di arrivo del segnale. Inoltre, questi eventi sono caratterizzati da una mancanza di onde S e presentano una coda che può durare fino a 1.5 minuti;
- *Tremors* (TR): i tremori sono molto simili agli eventi LP in quanto sono generati dagli stessi fenomeni che provocano i precedenti, con la differenza che presentano oscillazioni che possono durare da minuti a giorni o addirittura mesi; non c'è quindi una differenza significativa con gli eventi LP se non per quanto riguarda la durata;
- *Hybrid* (HB): questi fenomeni sono una combinazione di eventi LP e VT [1].

1.3 Utilizzo della pattern recognition nei fenomeni sismici vulcanici

Nel contesto analizzato la pattern recognition è un sistema molto utile per classificare automaticamente ed analizzare i dati di fenomeni sismici vulcanici.

Secondo una definizione storica, la pattern recognition è “il processo che prende in input dati grezzi (*raw*) ed effettua un'azione sulla base della categoria dei dati” [2]. In generale con pattern recognition si intende un processo volto ad analizzare i dati ricevuti in ingresso al fine

di categorizzarli in varie classi o gruppi, sulla base di determinate e specifiche caratteristiche di interesse estrapolate dai dati.

Il punto di partenza di un sistema di pattern recognition consiste nell'acquisizione dei dati e delle informazioni iniziali. Nello specifico, la collezione di dati sismici viene effettuata seguendo delle procedure automatiche per permettere ed assicurare un'acquisizione continua ed un controllo costante dei fenomeni vulcanici.

I tre passaggi che permettono la collezione automatica dei dati sui fenomeni sismici vulcanici prevedono: la misurazione per mezzo di specifici dispositivi in grado di captare le oscillazioni del terreno e convertirle in segnali misurabili, l'acquisizione dei dati con successiva conversione del segnale da analogico a digitale ed infine la trasmissione dei dati per mezzo di sistemi radio [1].

Vengono utilizzati in primis dei sensori per la rilevazione degli eventi sismici in formato analogico e successivamente la loro conversione avviene attraverso l'elaborazione dei dati in segnali digitali. La fase successiva consiste nella rappresentazione dei dati che permette, attraverso l'estrazione di apposite features e sfruttando fondamentali criteri di similarità/dissimilarità, la classificazione per mezzo di approcci statistici o di machine learning dei fenomeni osservati (Fig. 2).

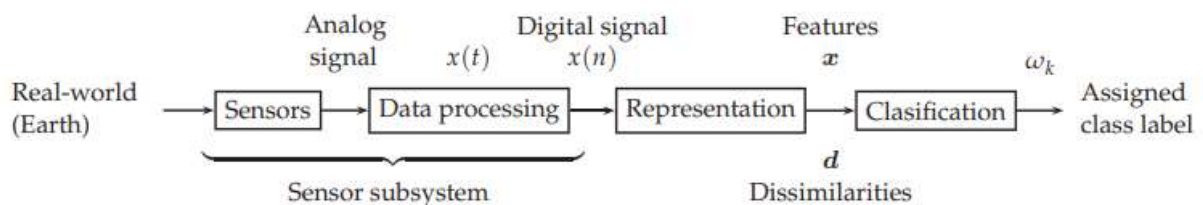


Fig. 2 Sistema applicativo di pattern recognition [1]

Nello studio compiuto è stata analizzata ed utilizzata una particolare tecnica avanzata di pattern recognition denominata “Transfer Learning” che permette di sfruttare e adattare la conoscenza acquisita da un task precedente (source task) per un nuovo task (target task).

In alcuni contesti applicativi il procedimento di collezione dei dati per l'addestramento di un classificatore può essere molto lungo e costoso. Di conseguenza può essere vantaggioso sfruttare la conoscenza estratta da dati acquisiti precedentemente in altri contesti simili; le tecniche di transfer learning mirano esattamente a questo, permettendo di ottimizzare tempi e costi [3].

Nello studio proposto è stata infatti simulata, attraverso tecniche di transfer learning, la creazione di una nuova antenna evitando il passaggio di acquisizione dei dati, ma sfruttando le informazioni appositamente elaborate ricavate da un'altra antenna.

Questa procedura può essere attuata ogni qualvolta che si registrano problematiche relative all'acquisizione dati come la costruzione o la rottura di un'antenna o la perdita di dati improvvisa.

CAPITOLO 2 – BACKGROUND

In questo capitolo verranno analizzate le informazioni di background per la comprensione delle tecniche applicate. Esse comprendono la rappresentazione dei dati, il transfer learning, i classificatori utilizzati e il fondamentale protocollo di validazione.

2.1 Rappresentazione: FFT, spettrogramma, spettrogramma medio, PCA

La rappresentazione è uno dei punti focali di un sistema di pattern recognition, quindi da una buona rappresentazione si è in grado di estrapolare delle features corrette e di conseguenza ottenere un classificatore performante.

I dati inizialmente estratti sono segnali descritti nel dominio temporale, ma per poterli analizzare affondo e manipolarli è necessario operare un cambio di rappresentazione. Ci si avvale quindi della Trasformata di Fourier che permette di scomporre un'onda raffigurata nel dominio temporale nelle sue sottocomponenti e visualizzarle nel dominio delle frequenze denominato anche dominio di Fourier. Il segnale ottenuto viene chiamato spettro della trasformata e la sua rappresentazione in funzione del tempo è detta spettrogramma. L'illustrazione grafica di uno spettrogramma è realizzata posizionando sull'asse delle ascisse il tempo mentre sull'asse delle ordinate la frequenza. Nella colormap utilizzata nella Fig. 3 i colori freddi corrispondono a valori del segnale bassi, mentre i colori caldi corrispondono a valori del segnale alti.

Gli spettri analizzati in questo studio risultavano essere di lunghezza variabile, di conseguenza per poter attuare un confronto ed operare con essi è stato necessario ricavare lo spettrogramma medio di ogni segnale ottenendo in questo modo dei segnali della stessa lunghezza che possono essere quindi paragonati e studiati.

Una tecnica che permette la standardizzazione dei dati con l'obiettivo di visualizzare e rappresentare i dati stessi è la Principal Component Analysis (PCA). Questa procedura consente di trasformare i dati attraverso un approccio lineare non supervisionato, applicando una riduzione della dimensionalità dello spazio assumendo che le informazioni più importanti siano codificate dalla varianza. La PCA è la miglior tecnica che permette di mantenere la maggior coerenza possibile con i dati originali.

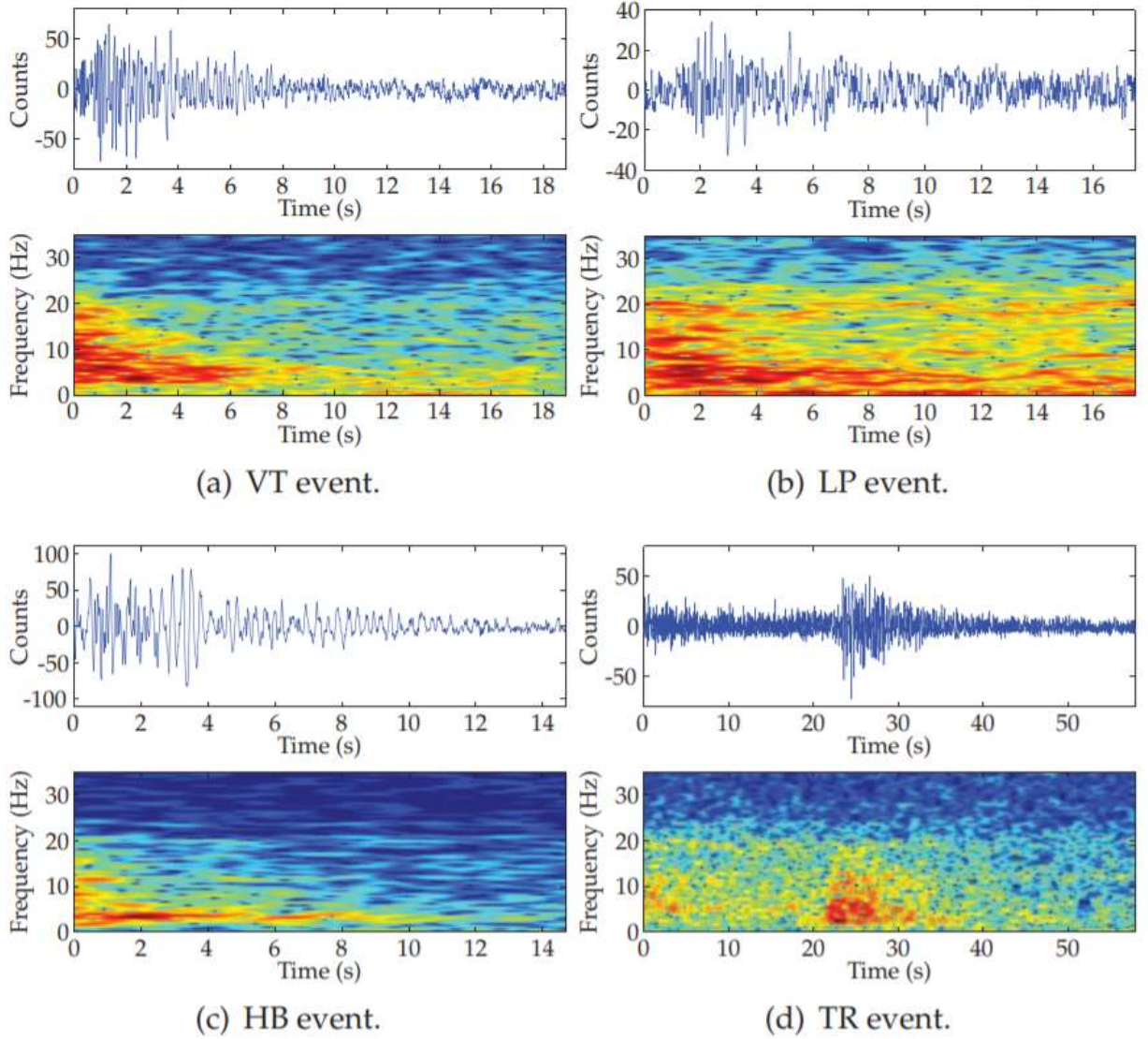


Fig. 3 Esempi di segnali sismici e relativi spettrogrammi rilevati dal vulcano Nevado del Ruiz [1]

2.2 Il Transfer Learning

Le tecniche di transfer learning possono essere suddivise in tre diverse categorie in base al contesto dei domini e dei task di source e target. Il dominio è composto dal contesto in cui si collocano i dati di interesse (in questo studio il dominio è rappresentato da segnali sismici vulcanici), mentre il task, sfruttando i dati del dominio, si pone di trovare una soluzione al problema in esame (la classificazione di tali segnali in questo caso). Nel transfer learning può essere vantaggioso sfruttare le informazioni provenienti dal dominio dei dati di source ed applicarle al dominio dei dati di target. Source rappresenta la sorgente dei dati ovvero i dati di partenza i quali, dopo un processing attraverso particolari tecniche, produrranno i dati di target.

Le tre categorie in cui il transfer learning si suddivide sono: induttivo, trasduttivo e non supervisionato.

Nel transfer learning induttivo si analizzano gli stessi domini di source e target, mentre i task sono differenti ma correlati tra loro. Dal momento che le labels di entrambi i domini sono disponibili, esso si limita a trasferire la conoscenza acquisita dal task di source a quello di target (previo processing dei dati). D'altra parte, nel caso in cui le labels del dominio di target siano disponibili e quelle relative al dominio di source non lo siano, le informazioni e i dati del dominio di source non possono essere sfruttati direttamente nel dominio di target.

Nel transfer learning trasduttivo invece si analizzano diversi domini di source e target ma relazionati tra loro, mentre i tasks di source e target considerati sono gli stessi. Inoltre, in questo caso, le labels di source sono disponibili mentre le labels di target non lo sono.

Nel transfer learning non supervisionato infine entrambi i domini ed entrambi i tasks sono diversi ma relazionati tra di loro. Inoltre, in questo caso, sia le labels del dominio di source che quelle del dominio di target non sono disponibili; per risolvere questo tipo di problematiche, dunque, possono essere utilizzate delle tecniche di clustering con l'obiettivo di creare dei gruppi (cluster) sulla base di un preciso criterio di similarità [3].

Nello studio proposto i tasks di source e target sono i medesimi in quanto ci si propone di classificare fenomeni sismici vulcanici registrati da due antenne differenti situate in posizioni differenti dello stesso vulcano. D'altra parte, però, i domini di source e target sono diversi ma relazionati in quanto in entrambi i casi si tratta di segnali sismici acquisiti da stazioni, che però, essendo dislocate in posizioni differenti del vulcano, collezionano dati diversi per lo stesso tipo di fenomeno. Il segnale registrato dipende infatti dall'evento (lo stesso) e anche dal mezzo fisico attraversato per raggiungere le stazioni; di conseguenza saranno registrati due diversi segnali che corrispondono allo stesso fenomeno. Il problema in questione si riflette quindi nel transfer learning trasduttivo che prevede delle tecniche di domain adaptation per il suo studio.

2.2.1 Il transfer learning negativo

La relazione che intercorre tra il task di source e il task di target influisce profondamente sull'efficacia del metodo di transfer learning: più essa è forte e più il metodo di transfer learning produrrà un risultato ottimale. Se d'altra parte non si verifica una relazione solida tra task di source e task di target o se essa non è sfruttata appieno dal metodo di transfer learning, i risultati non solo non rispecchiano le aspettative, ma potrebbero inficiare ancor di più il risultato.

Il metodo di transfer learning ideale è un metodo in grado di ottimizzare il transfer positivo tra task correlati, denominato così in quanto le tecniche utilizzate permettono il conseguimento

dei risultati desiderati, e simultaneamente in grado di minimizzare e/o evitare il transfer negativo che indica il fallimento del metodo di transfer learning in quanto i risultati ottenuti non rispecchiano le aspettative quando i tasks non sono strettamente correlati. Conciliare questi due concetti non è facilmente realizzabile. Infatti, possono essere utilizzate delle misure precauzionali come lo sviluppo di un algoritmo “safe” che non produce necessariamente risultati ottimali ma in grado di evitare il transfer negativo; se invece si vogliono ottenere dei risultati altamente performanti si può utilizzare un approccio più diretto con il rischio però di ottenere l’effetto opposto sfociando quindi nel transfer learning negativo [4].

2.3 Classificatori utilizzati

Nello studio proposto sono stati utilizzati vari classificatori per testare le tecniche di transfer learning analizzate in questo studio. Essi sono:

- K-Nearest Neighbor: è un classificatore non parametrico. In particolare il suo funzionamento prevede la scelta di un valore intero K e si costruisce una regione R che contiene i K oggetti del training set più vicini a quello da classificare secondo una certa misura di distanza; la classe che risulta essere la più frequente all’interno di R sarà quella assegnata all’oggetto da classificare. Il Nearest Neighbor è la variante del K-NN in cui $K=1$.
- Nearest Mean Classifier: è un classificatore molto semplice basato sull’assunzione che ogni classe sia rappresentata dalla media degli oggetti che le appartengono. L’oggetto da classificare viene assegnato alla classe con distanza valore-media minore.
- Support Vector Machine: è un classificatore discriminativo molto utilizzato che si basa su concetti geometrici. Nella sua versione binaria, mira a trovare la retta, o iperpiano, in grado di massimizzare la distanza tra le classi. Nella sua versione più flessibile viene utilizzato il kernel trick per avere classi non lineari. Il classificatore che utilizza questa tecnica è chiamato Radial Basis SVM.

2.4 Validazione di un classificatore

La validazione è una fase di fondamentale importanza per un sistema di pattern recognition, in quanto permette di valutare l’efficacia di un sistema di classificazione. Lo scopo principale della validazione è misurare ed esaminare quanto un classificatore sia in grado di generalizzare, ovvero quanto esso è in grado di classificare correttamente anche oggetti non presenti nel training set. È necessario avere quindi due insiemi separati, uno per costruire il modello e uno

per testarlo. Per fare ciò, tipicamente, si dividono i dati presenti nel dataset in due diversi sottoinsiemi: il training set che contiene i dati con cui viene addestrato il classificatore ed il testing set che viene usato per testare il classificatore. Questa metodologia è definita cross-validation, e prevede vari approcci per la suddivisione del dataset.

2.4.1 Validazione utilizzata nello studio

Nello studio proposto ci si è avvalsi di una procedura di validazione particolare disegnata ad-hoc per il caso specifico, che permette la simulazione del seguente scenario (Fig. 4). È stato diviso l'intero dataset in due parti pressoché uguali sulla base di una determinata data temporale. I valori appartenenti al dataset di source sono rimasti invariati, mentre quelli relativi al dataset di target sono stati ulteriormente suddivisi con la necessità di creare un testing set per la validazione. In particolare, la prima parte è stata suddivisa ripetutamente nel 10%, 20%, 30%, 40% e 50% utilizzando ogni volta ciascuna porzione singolarmente come training set per addestrare il classificatore, mentre la seconda, dovendo effettuare ripetuti esperimenti con dati differenti, è stata utilizzata come testing set invariato per conservare una correlazione tra i risultati ed essere in grado di confrontarli tra di loro.

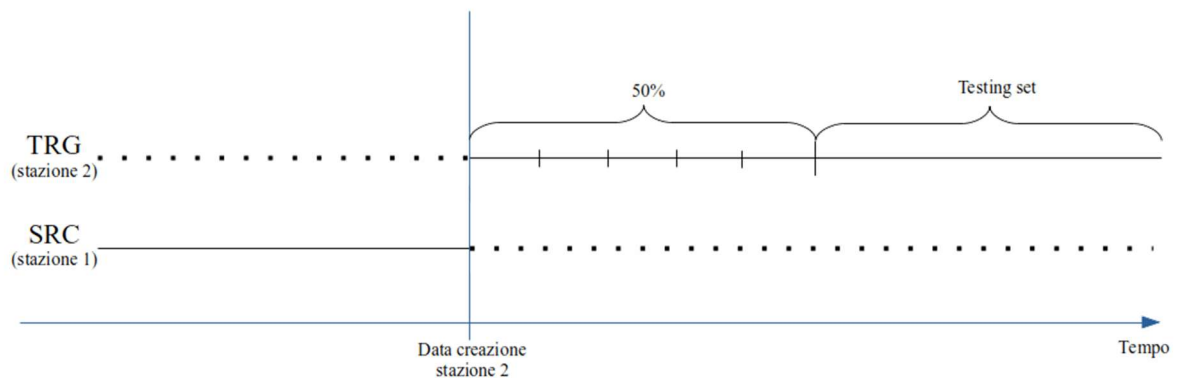


Fig. 4 Schema di suddivisione del dataset

CAPITOLO 3 – METODOLOGIA PROPOSTA

In questo studio preliminare sono state studiate alcune tecniche di transfer learning basilari. È importante osservare che nella letteratura non ci sono ancora studi di questo tipo che mirano ad investigare il transfer learning nel contesto dei segnali sismici vulcanici.

In particolare sono state applicate quattro diverse tecniche diverse:

- SRC-ONLY: è una tecnica che utilizza esclusivamente i dati relativi al dominio di source ignorando quelli relativi al dominio di target. Non prevede una manipolazione dei dati, ma si limita ad utilizzare il dominio di source per addestrare il classificatore e quello di target per testare il classificatore.
- TRG-ONLY: è una tecnica di classificazione standard che prevede solamente l'utilizzo dei dati del dominio di target sia per addestrare che per testare il classificatore.
- ALL: è una tecnica che prevede la combinazione dei due domini in esame, in particolare comporta l'utilizzo di entrambi i training set di source e target per addestrare il classificatore mantenendo però invariato, come negli altri casi, i dati di test. Un potenziale problema di questa tecnica è dato dalla probabile diversa dimensione dei dataset. Infatti, tipicamente i dati del dominio di source sono di gran lunga in maggior numero rispetto a quelli del dominio di target, di conseguenza i dati di target potrebbero addirittura essere ininfluenti per l'addestramento del classificatore.
- MAPPING: è una tecnica che prevede l'utilizzo di una particolare metodologia chiamata domain adaptation attraverso la quale viene realizzata una nuova rappresentazione dei dati mediante un aumento delle features con modalità differenti in entrambi i domini. Tra le tecniche proposte, il mapping, apportando una manipolazione dei dati, è la metodologia che dovrebbe fornire le prestazioni migliori [5]. Per questo motivo viene spiegata in maniera più approfondita successivamente.

3.1 Domain adaptation utilizzata nella tecnica di mapping

La domain adaptation utilizzata in questo studio ha l'obiettivo di prendere ogni feature nel problema originale e produrre tre versioni di esso: una versione generale, una versione source-specifica e una versione target-specifica. I dati di source nella nuova rappresentazione conterranno solamente le versioni generale e source-specifica. I dati di target nella nuova rappresentazione conterranno solamente le versioni generale e target-specifica [5].

Come sopraccitato, la manipolazione dei dati apportata prevede una modificazione delle features tramite due tipi di mapping, uno per i dati di source e uno per i dati di target ottenendo una nuova rappresentazione per ciascuno dei due domini. Il dominio di source prodotto sarà quindi:

$$\Phi^s(x) = \langle x, x, 0 \rangle$$

mentre il dominio di target risulterà essere:

$$\Phi^t(x) = \langle x, 0, x \rangle$$

dove: x rappresenta il vettore delle features e 0 costituisce un vettore di n zeri dove n corrisponde alla cardinalità di x .

In entrambe le casistiche la prima parte della nuova rappresentazione delle features simboleggia la versione generale, ovvero dati comuni che non sono dominio-specifici. La seconda parte della nuova rappresentazione raffigura la versione source-specifica in cui i dati sono incompatibili con il dominio di target e sono caratteristici e unici del dominio di source. Infine, la terza posizione, la versione target-specifica, è l'opposto della precedente infatti contiene i dati che sono incompatibili con il dominio di source e sono caratteristici e unici del dominio di target [5] (Fig. 5).

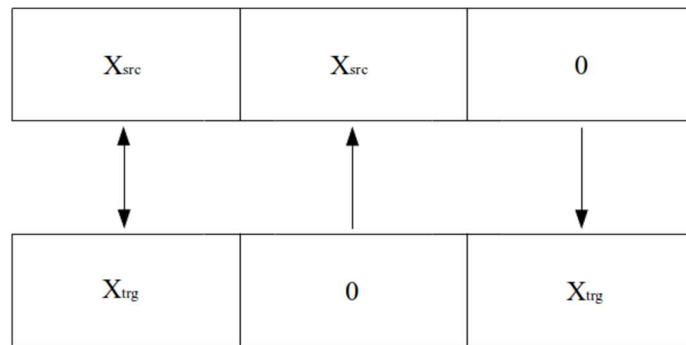


Fig. 5 Domain adaptation utilizzata nello studio

La tecnica di MAPPING è costruita attraverso la manipolazione dati sopraccitata, la quale sfrutta la nuova rappresentazione di entrambi i domini di source e target per la creazione del classificatore ponendo in tal modo le basi per un transfer in grado di migliorare le prestazioni di questa tecnica.

3.2 Pipeline creata per la realizzazione delle metodologie proposte

Per lo sviluppo dell'algoritmo proposto è stato utilizzato il linguaggio di programmazione MATLAB ed un particolare toolbox per l'implementazione di applicazioni di pattern recognition chiamato PRTools. Sono stati creati tre metodi: Main, DividiClassi e SplitData.

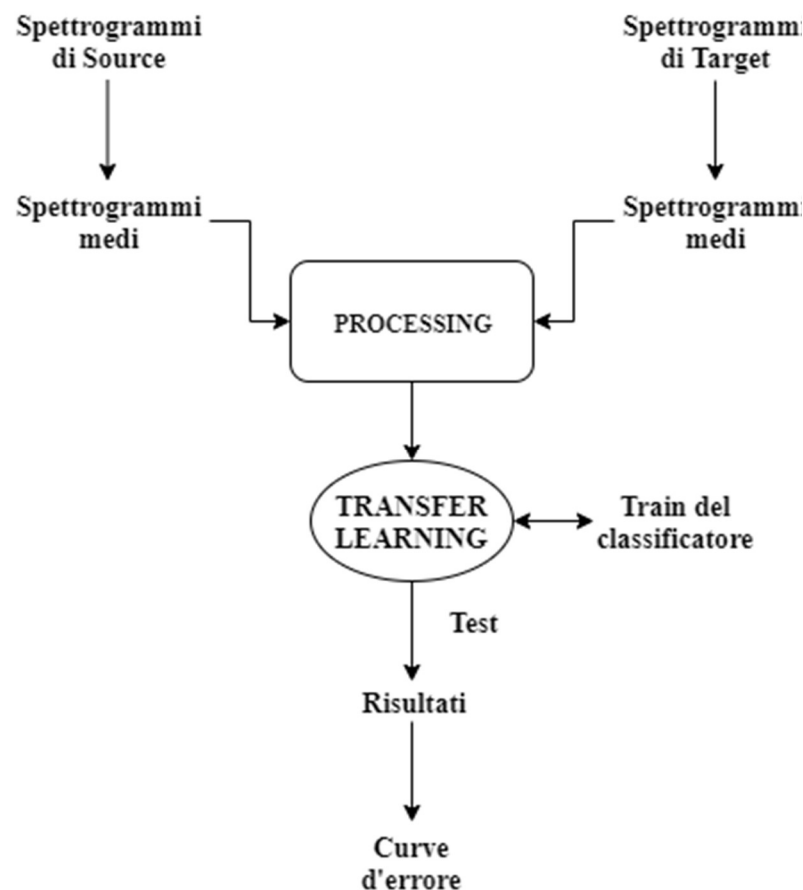


Fig. 6 Pipeline proposta

Il punto di partenza dell'algoritmo prevede la creazione degli spettrogrammi medi per ogni segnale sismico proveniente da ciascuna delle due antenne di source e di target. Segue una fase di processing nella quale viene effettuata la suddivisione del dataset come sopraccitato. Ottenuti i dati elaborati, sono state applicate le quattro metodologie proposte addestrando i diversi classificatori ed ottenendo, dopo una fase di testing, risultati analizzabili che vengono poi rappresentati sotto forma di curve d'errore (Fig. 6).

CAPITOLO 4 – RISULTATI SPERIMENTALI

Lo studio proposto è basato su dati riguardanti segnali sismici vulcanici contenuti in un dataset fornito dal Depto. de Informática y Computación dell'Universidad Nacional de Colombia Sede Manizales in collaborazione con l'Osservatorio Vulcanologico y Sismológico de Manizales. I risultati ricavati sono stati ottenuti dalla versione del dataset aggiornata a fine maggio 2019. Nel contesto applicativo considerato sono state utilizzate due particolari antenne di rilevazione (sensori) con un'attività di monitoraggio ed archiviazione dei dati decennale, ubicate in due zone differenti del vulcano Nevado del Ruiz. Per testare le tecniche di transfer learning, i dati relativi alla prima antenna (BIS - source) sono stati mantenuti integralmente, mentre la seconda antenna (OLL - target) è stata considerata come relativamente nuova utilizzando i dati di pochi mesi di monitoraggio.

4.1 Analisi del dataset

Il dataset in questione presenta 1078 serie temporali che raffigurano segnali sismici vulcanici ciascuno dei quali associato ad una specifica classe di appartenenza e alla data di registrazione di tale evento. Le classi relative ai segnali in esame sono cinque e sono rappresentate dagli eventi vulcano tectonic (VT), gli eventi long period (LP), i tremori (TR), gli eventi ibridi (HB) ed infine gli eventi Tornillo (TO). La cardinalità delle classi è differente ed i segnali sono così distribuiti: sono presenti 153 elementi per la classe VT, 333 elementi appartenenti alla classe LP, 240 elementi rappresentanti la classe TR, 393 elementi costituenti la classe HB ed infine 104 elementi appartenenti alla classe TO. Come sopraccitato, i segnali presenti all'interno del dataset riguardano fenomeni sismici vulcanici registrati dalle due antenne BIS e OLL.

4.2 Esperimenti effettuati

Nello studio condotto sono stati effettuati numerosi esperimenti utilizzando diverse versioni del dataset e diversi classificatori. L'obiettivo comune di tutti gli esperimenti è valutare le tecniche di transfer learning su dati relativi ad eventi sismici vulcanici stimando l'efficacia delle quattro metodologie proposte.

Per ciascuna metodologia sono stati annotati in apposite tabelle gli errori dei classificatori utilizzando in modo incrementale il training set partendo inizialmente dal 10% (prima colonna di valori) arrivando fino al 50% (ultima colonna).

4.2.1 Esperimento 1

Nel primo esperimento effettuato, sono stati utilizzati tutti i segnali presenti nel dataset applicando le quattro diverse tecniche di transfer learning ed analizzandone i risultati ottenuti. Queste metodologie sono state valutate attraverso l'utilizzo dei classificatori Nearest Neighbor, K-Nearest Neighbor, Nearest Mean Classifier, Support Vector Machine e Radial Basis Support Vector Machine il quale, attraverso una cross-validation del dataset, stima in modo automatico i parametri da applicare in fase di addestramento del classificatore.

I risultati sono presentati nella tabella sottostante (Tab. 1).

		10%	20%	30%	40%	50%
KNN	SRC-ONLY	0.7279	0.7279	0.7279	0.7279	0.7279
	TRG-ONLY	0.4596	0.4154	0.4044	0.3824	0.4154
	ALL	0.5110	0.5000	0.4926	0.4559	0.4706
	MAPPING	0.4696	0.4301	0.4228	0.4118	0.4412
1-NN	SRC-ONLY	0.6765	0.6765	0.6765	0.6765	0.6765
	TRG-ONLY	0.4596	0.4154	0.4596	0.4044	0.4118
	ALL	0.4853	0.4522	0.4485	0.4338	0.4449
	MAPPING	0.4596	0.4154	0.4596	0.4044	0.4118
NMC	SRC-ONLY	0.7500	0.7500	0.7500	0.7500	0.7500
	TRG-ONLY	0.4853	0.4632	0.4632	0.4081	0.3566
	ALL	0.7684	0.7756	0.7904	0.7794	0.7647
	MAPPING	0.8566	0.8566	0.9118	0.9044	0.9871
SVM	SRC-ONLY	0.7978	0.7978	0.7978	0.7978	0.7978
	TRG-ONLY	0.4522	0.4963	0.5331	0.4963	0.4669
	ALL	0.6728	0.6397	0.5993	0.6026	0.5993
	MAPPING	0.6213	0.6140	0.5000	0.4522	0.4890
RBSVM	SRC-ONLY	0.7721	0.7721	0.7721	0.7721	0.7721
	TRG-ONLY	0.4265	0.4154	0.3750	0.3529	0.3566
	ALL	0.4743	0.4596	0.4632	0.4669	0.4779
	MAPPING	0.4037	0.4301	0.4044	0.3493	0.3750

Tab. 1 Risultati esperimento 1

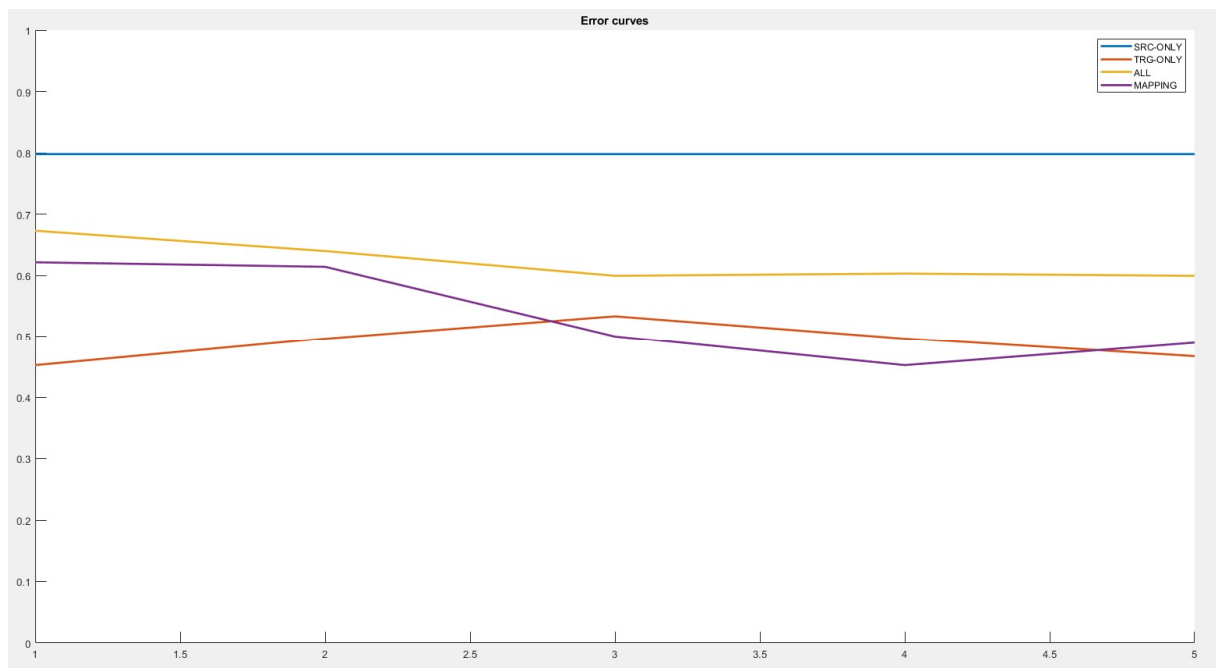


Fig. 7 Risultati esperimento 1 con classificatore SVM

Come si può notare dall'esempio riportato, utilizzando come classificatore SVM (Fig. 7) gli errori di classificazione ottenuti sono eccessivamente elevati; pertanto si è deciso di analizzare e studiare in modo approfondito il dataset in esame.

Attraverso una PCA, è stata applicata una riduzione del numero delle features da uno spazio a 65 dimensioni ad uno spazio a tre dimensioni per permettere la rappresentazione dei dati registrati da ciascuna delle due antenne (BIS e OLL) per ognuna delle cinque classi. L'obiettivo era volto a capire se una delle cinque classi conduce il classificatore a riportare degli errori così elevati.

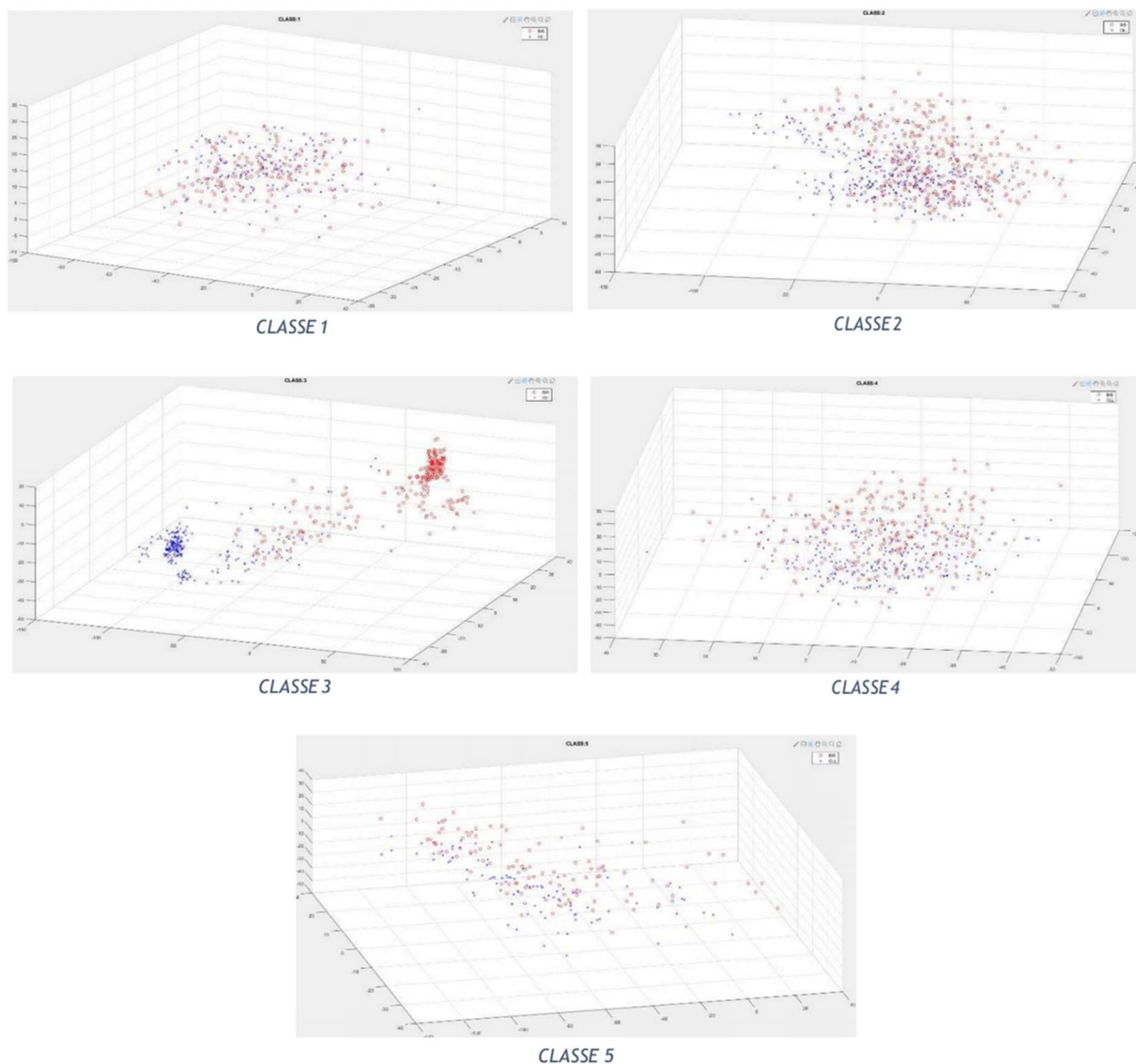


Fig. 8 Visualizzazione dei dati delle cinque classi

Come visibile nell'esempio soprastante (Fig. 8) i dati acquisiti dalle due antenne, colorati in modo differente (rosso per i dati dell'antenna BIS e blu per i dati dell'antenna OLL), mostrano che nella classe 1 (VT) i dati sono in forte relazione tra di loro e sono quindi facilmente confrontabili, mentre nella classe 3 (TR) essi hanno valori completamente differenti e si può quindi dedurre che le cattive prestazioni del classificatore dipendono da questa classe.

Allo stesso modo, la classe 2 (LP) ha un comportamento simile alla classe 1 con valori paragonabili, mentre la classe 5 (TO) presenta dati molto sparsi e pressoché separati. Un ragionamento a parte deve essere effettuato per la quarta classe che sembra mostrare valori confrontabili ma, rappresentando i fenomeni ibridi ovvero la combinazione degli eventi VT ed LP, potrebbe provocare un forte calo delle prestazioni in atto di classificazione.

4.2.2 Esperimento 2

Alla luce di quanto osservato, è stato deciso di creare il nuovo dataset in cui sono presenti solamente le classi VT ed LP eliminando le classi TR, HB e TO. Di seguito vengono riportati i risultati ottenuti (Tab. 2).

		10%	20%	30%	40%	50%
KNN	SRC-ONLY	0.0813	0.0813	0.0813	0.0813	0.0813
	TRG-ONLY	0.5447	0.3496	0.3171	0.2439	0.1789
	ALL	0.0894	0.0894	0.0976	0.0976	0.0976
	MAPPING	0.5447	0.3496	0.3171	0.2439	0.1789
1-NN	SRC-ONLY	0.0813	0.0813	0.0813	0.0813	0.0813
	TRG-ONLY	0.5447	0.3496	0.3171	0.2439	0.1789
	ALL	0.0894	0.0894	0.0976	0.0976	0.0976
	MAPPING	0.5447	0.3496	0.3171	0.2439	0.1789
NMC	SRC-ONLY	0.3821	0.3821	0.3821	0.3821	0.3821
	TRG-ONLY	0.4390	0.3252	0.3089	0.1870	0.1382
	ALL	0.3821	0.3821	0.3821	0.3821	0.3821
	MAPPING	0.6829	0.6829	0.6829	0.6829	0.6829
SVM	SRC-ONLY	0.2358	0.2358	0.2358	0.2358	0.2358
	TRG-ONLY	0.4715	0.3577	0.3171	0.0650	0.0732
	ALL	0.2439	0.1626	0.0976	0.0732	0.1220
	MAPPING	0.2358	0.2033	0.1626	0.0488	0.0569
RBSVM	SRC-ONLY	0.1220	0.1220	0.1220	0.1220	0.1220
	TRG-ONLY	0.3415	0.1951	0.1707	0.1134	0.1057
	ALL	0.1626	0.0894	0.0976	0.0732	0.0894
	MAPPING	0.3821	0.1870	0.2114	0.0813	0.1134

Tab. 2 Risultati esperimento 2

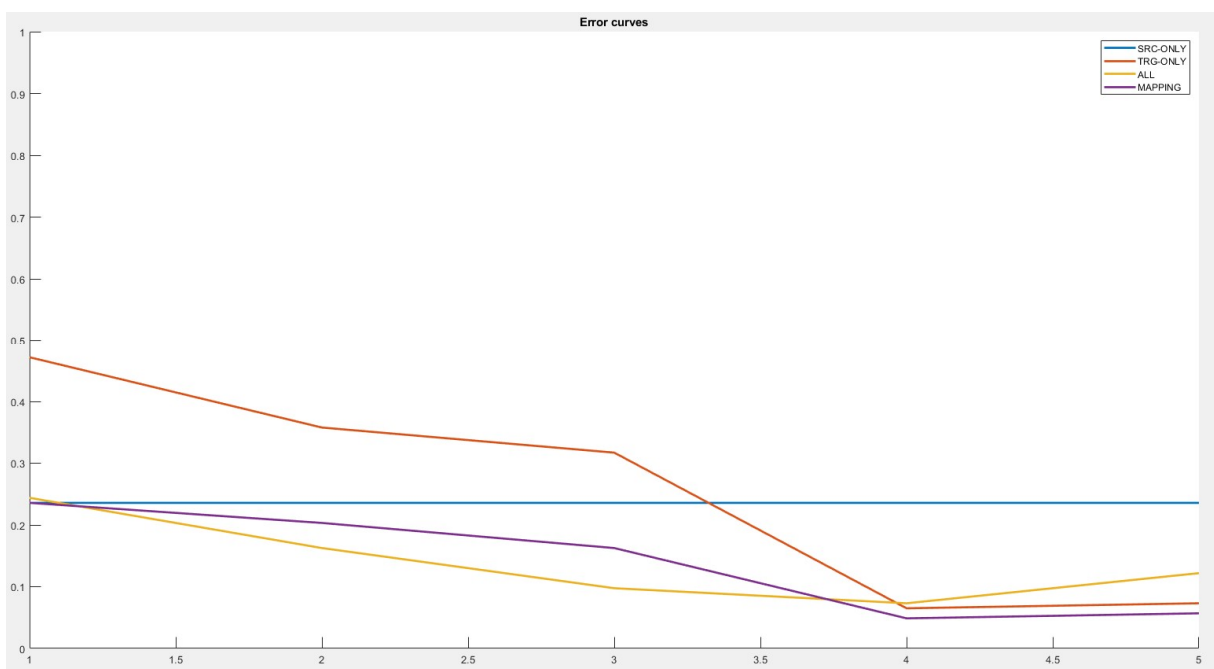


Fig. 9 Risultati esperimento 2 con classificatore SVM

Confrontando i valori ottenuti dai primi due esperimenti, si può notare come il secondo fornisca un risultato migliore in quanto gli errori calano drasticamente ma soprattutto si può osservare come le tecniche di transfer learning abbiano un andamento congruo alle aspettative. Si può notare, inoltre, che il metodo MAPPING (viola) attraverso la nuova rappresentazione dei dati secondo la domain adaptation diminuisce l'errore migliorando le prestazioni del metodo ALL (giallo) (Fig. 9).

4.2.3 Esperimento 3

L'ultimo esperimento è mirato allo studio della seguente ipotesi: dopo particolari eventi sismici, la conformazione terrestre viene alterata modificando di conseguenza la rilevazione degli eventi successivi e causando imprecisioni a livello di classificazione. Basandosi su questa supposizione, risulterebbe più opportuno utilizzare solamente i dati acquisiti più recentemente. Questo è stato ripetuto sia nelle condizioni del primo esperimento tenendo in considerazione cioè l'intero dataset, che nelle condizioni del secondo esperimento utilizzando quindi solamente le classi VT ed LP. In entrambe le casistiche sono state ripetute le classificazioni sfruttando sia gli ultimi 100 segnali di ogni classe considerata che gli ultimi segnali con un numero pari alla dimensione della classe più piccola.

I risultati prodotti non sono stati in grado però di migliorare quanto ottenuto precedentemente nell'esperimento 2.

4.3 Commenti

In conclusione gli esperimenti finali effettuati si possono dividere in due gruppi: quelli effettuati utilizzando tutte le classi e quelli effettuati utilizzando solamente le prime due classi. In ognuno dei due gruppi sono state ripetute varie classificazioni: sono stati utilizzati tutti i valori nell'arco temporale, gli ultimi 100 valori più recenti ed infine gli ultimi N valori dove N corrisponde alla cardinalità della classe più piccola (104 utilizzando tutte le classi, 153 utilizzando solamente le prime due classi).

Dai risultati ottenuti è possibile notare, come atteso, che gli errori della metodologia SRC-ONLY rimangono invariati nelle diverse classificazioni, in quanto il training set, costituito da tutti i valori della stazione BIS antecedenti alla data di divisione, rimane immutato. Di conseguenza sono stati studiati a fondo gli errori generati dall'applicazione delle altre tre metodologie.

Deve essere inoltre riportato come molte classificazioni dell'esperimento 3 abbiano prodotto risultati con errori dello 0%. La causa di questo fenomeno è legata all'overtraining, un episodio in cui un classificatore invece di apprendere dai dati, li memorizza rendendo impossibile la generalizzazione.

Il problema affrontato in questo studio si è rivelato essere molto classificatore-dipendente, in quanto sono stati testati svariati classificatori e per ciascuno di essi sono stati ottenuti risultati molto diversi; Tra i molteplici esperimenti effettuati, il miglior risultato si è rivelato essere quello ottenuto nell'esperimento 2 in particolare utilizzando come classificatore la Support Vector Machine.

I risultati ottenuti non si sono dimostrati ottimali in quanto le tecniche applicate risultano essere semplici. Bisognerebbe infatti analizzare più approfonditamente il problema ed adottare delle tecniche più complesse e specifiche create ad-hoc.

CAPITOLO 5 – CONCLUSIONI

In questo studio sono state analizzate delle tecniche di transfer learning applicate alla classificazione di segnali relativi a fenomeni sismici vulcanici.

È stata ipotizzata la creazione di una nuova antenna ubicata sul vulcano Nevado del Ruiz situato in Colombia. Dal momento che, in questo caso, la procedura di acquisizione dei segnali e di addestramento del classificatore richiedono uno storico elevato, sono state testate quattro tipologie di transfer learning analizzando la buona riuscita (o meno) del trasferimento della conoscenza dall'antenna precedentemente installata a quella costruita successivamente.

Per studiare i metodi di transfer learning, sono stati presi in considerazione cinque diversi classificatori con caratteristiche e funzionamenti altrettanto differenti.

Dagli esperimenti realizzati, solo uno è risultato essere coerente con le aspettative dimostrando che il problema di classificazione analizzato è complicato e necessita di un ulteriore studio specifico ed approfondito.

BIBLIOGRAFIA

- [1] Mauricio Orozco-Alzate, Carolina Acosta-Muñoz and John Makario Londoño-Bonilla, *The Automated Identification of Volcanic Earthquakes: Concepts, Application and Challenges, Earthquake Research and Analysis - Seismology, Seismotectonic and Earthquake Geology*, InTech, 2012, pp. 377-402
- [2] Richard O. Duda, Peter E. Hart, David G. Stork, *Pattern Classification*, Wiley-Interscience, 2001
- [3] Sinno Jialin Pan, Qiang Yang, *A Survey on Transfer Learning*, IEEE Transaction On Knowledge And Data Engineering, 2009, Vol. 22, No. 10, pp. 1345-1359
- [4] Lisa Torrey, Jude Shavlik, *Transfer Learning, Handbook of Research on Machine Learning Applications*, IGI Global, 2009, pp. 242-264
- [5] Hal Daumè III, *Frustratingly Easy Domain Adaptation*, *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 2007, pp. 256–263