

Classificatore BDTB

Applicazione su variabili di RMN

V. Marè - 13/04/2017

Il Classificatore

Boosted Decisional Tree with Bagging (BDTB):

- Tipo di classificatore: Albero binario decisionale
- Numero di alberi = 400
- *Bagging*: Ogni albero viene cresciuto su un campione i cui elementi sono scelti casualmente – con rimpiazzo – dall'insieme degli oggetti
- Dimensione minima nodo: 2,5% della dimensione del campione
- Criterio decisionale = CrossEntropy: $i(t) = \sum_{j=1}^c P_j * \log(P_j)$, in cui t è il nodo in

esame, $P_j = \frac{\text{n° oggetti della classe } j \text{ che hano raggiunto il nodo } t}{\text{numero di oggetti totali}}$ è la probabilità della classe j di

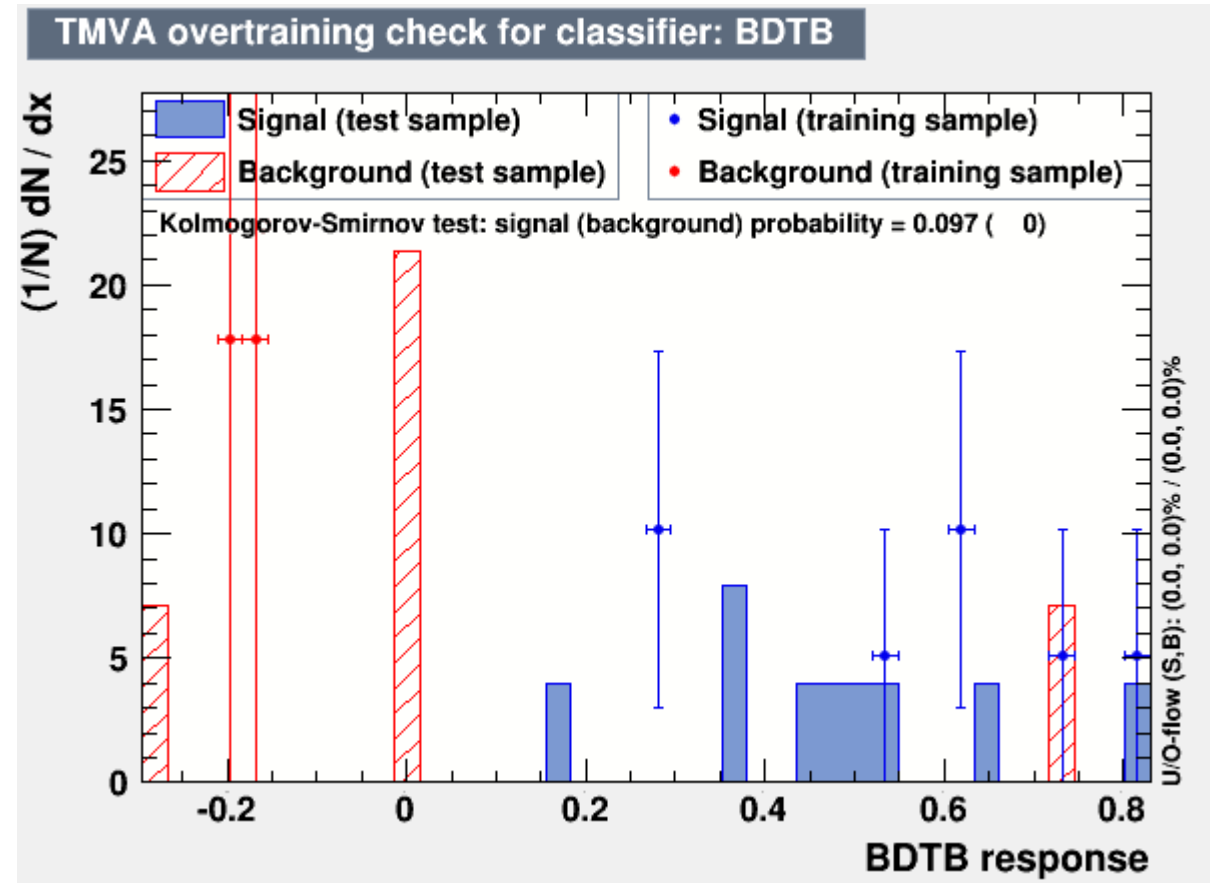
trovarsi nel nodo t e c è il numero totale di classi. i(t) vale 0 per nodi puri e log (c) nel caso di distribuzione uniforme degli oggetti nelle varie classi

Il file pazienti.root

- Variabile ypT2: discrimina tra segnale (=0) e fondo (>0)
- All'interno della root-upla ci sono, oltre a ypT2, 3 macro gruppi di variabili:
 1. Texture: sono 6 (media, sd, entropy, mpp, skewness, entropy) x 5 (livelli di filtro) x 3 (pre, durante e post CR) = 90
 2. T2 Intensity: sono 8 (Vol_tot, mediana, skewness, media, dev.std, skewness, kurtosis, Normf) x 3 = 24
 3. RMN Diffusione: sono 2 (ADC, DWIB1000) x 3 = 6
 4. RMN Perfusion: sono 4 (media, sd, min, max) x 13 parametri di perfusione (TME, AUC, WI, WO, VE, KT, KEP, VP, PEAK, CHI2, SER, PE, CW0) x 2 (pre e post CR) = 108

ROOT 5 vs ROOT 6

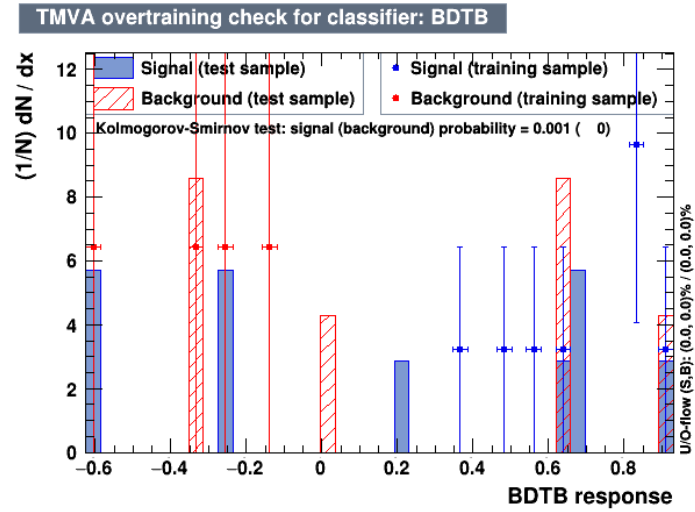
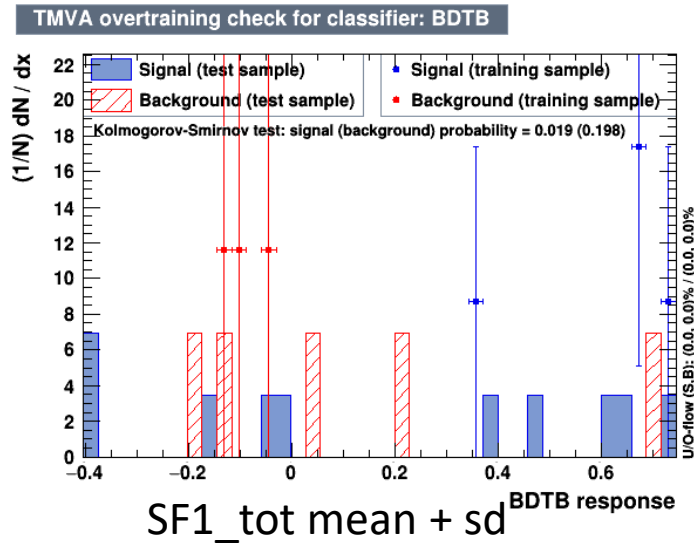
- Alcune classi di TMVA sono cambiate tra ROOT 5 e ROOT 6 (p.e. factory e dataloader) -> necessità di adattare il precedente codice alla nuova versione.
- Stessi risultati del codice precedente (di R. Paramatti)



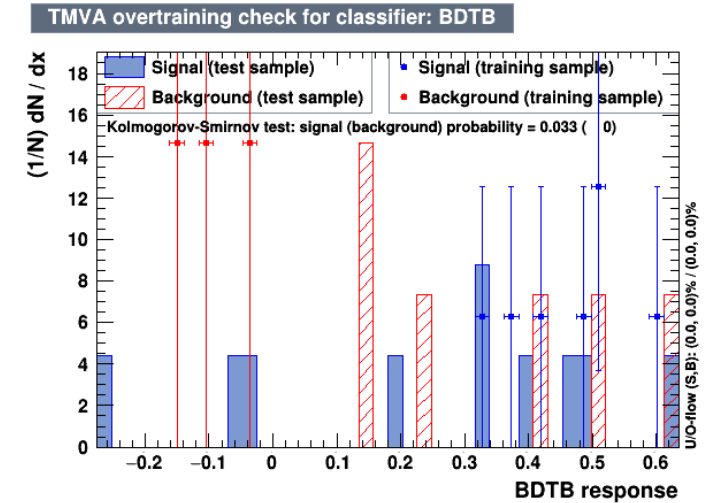
Variabili:

Vol_tot_pre – SF0_sd_int – SF0_skewness_pre –
SF1_skewness_post – SF1_kurtosis_post – SF1_5_post –
SF1_8_kurtosis_post

Texture



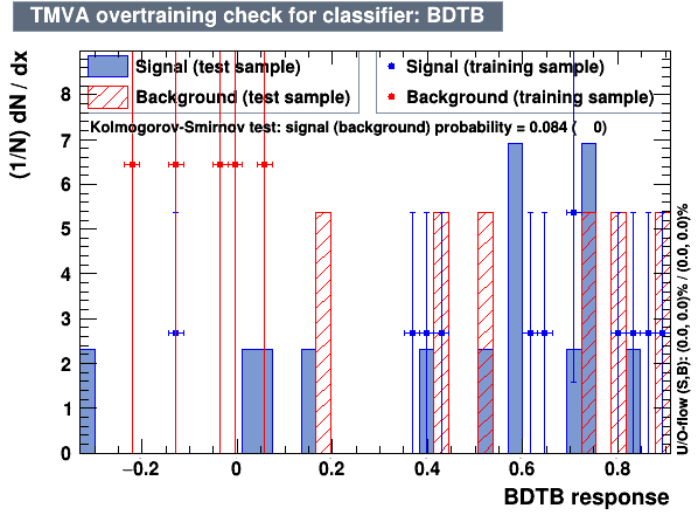
Combinazione dei parametri più discriminanti per ogni livello di filtro (pre int e post)



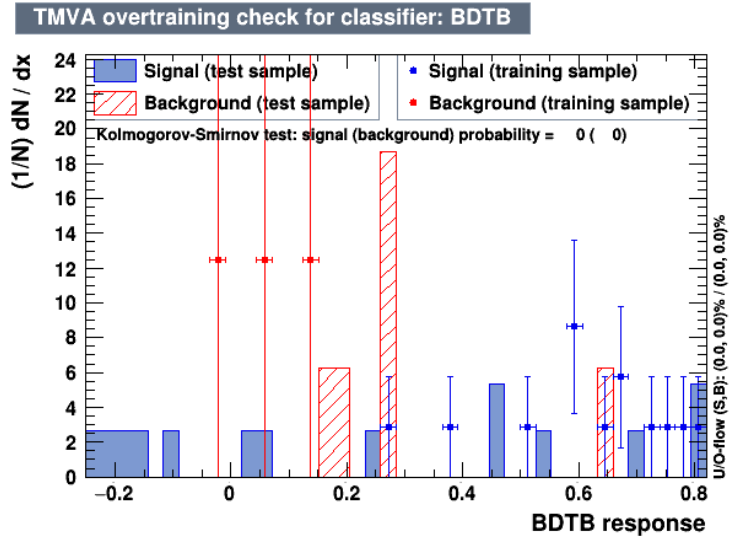
Variabili SF0_pre (senza filtro)

Basso potere discriminante in ogni caso

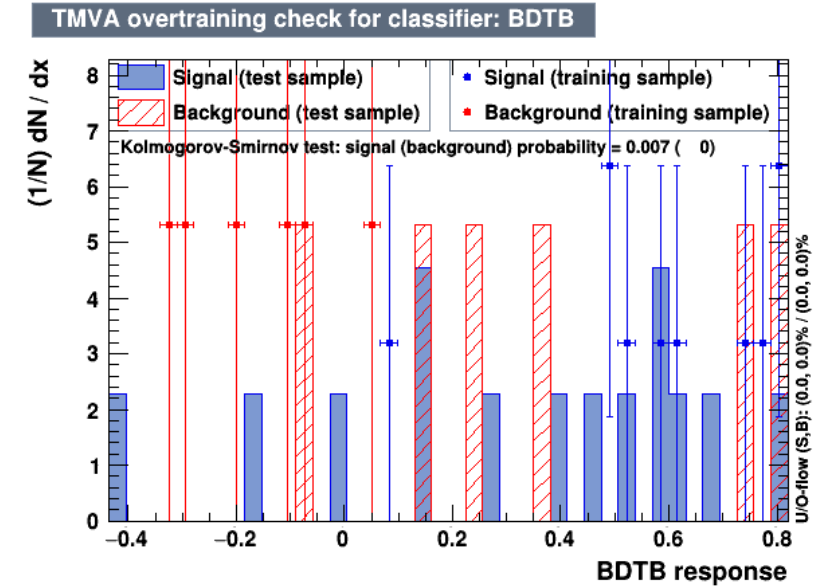
T2 Intensity



Durante CR



Pre CR

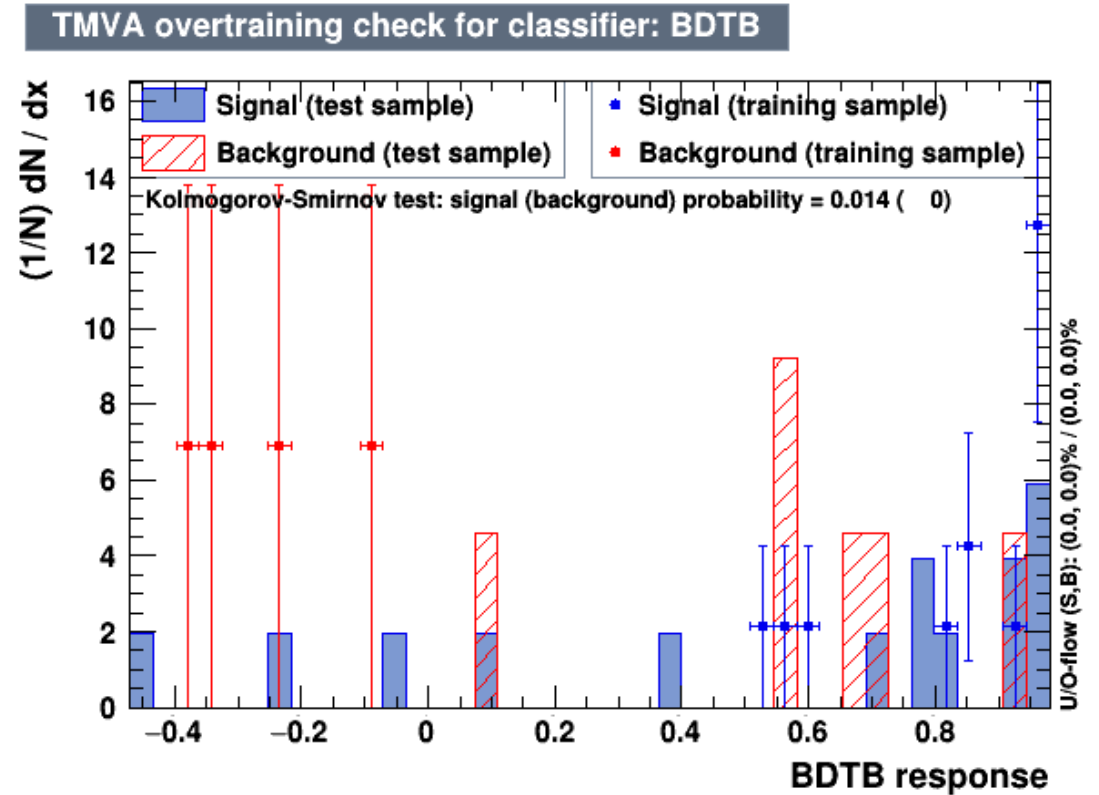


Post CR

Basso potere discriminante in ogni caso

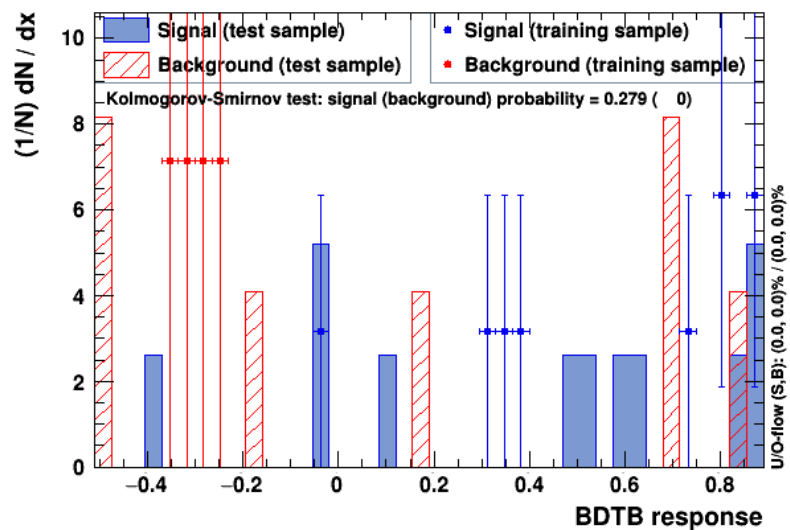
Diffusione

- Parametri pre, durante e post CR
- Basso potere discriminante

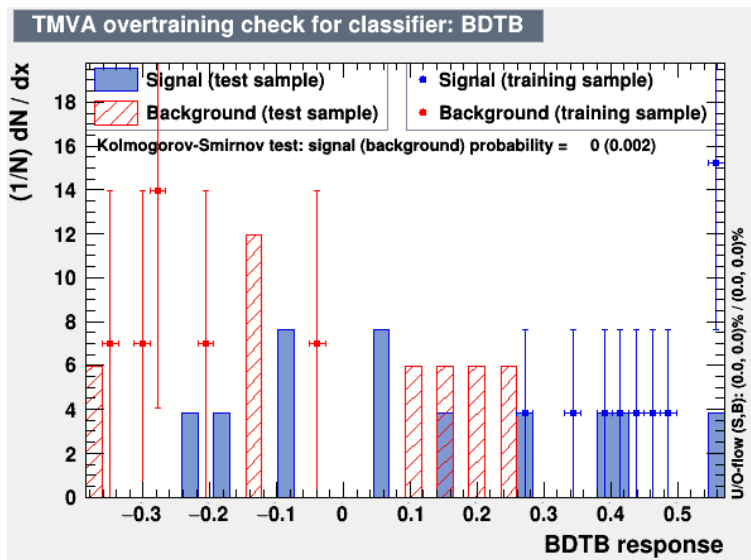


Perfusione

TMVA overtraining check for classifier: BDTB



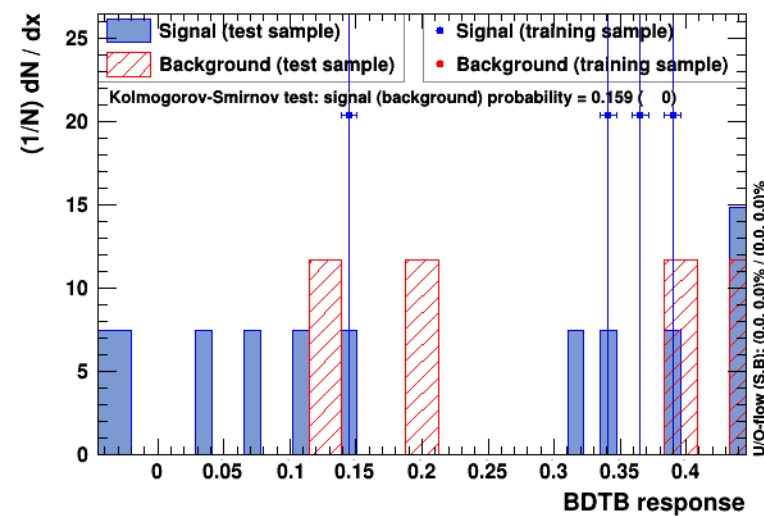
TME pre e post CR



Wash In (WI) pre e post CR

Basso potere discriminante in ogni caso

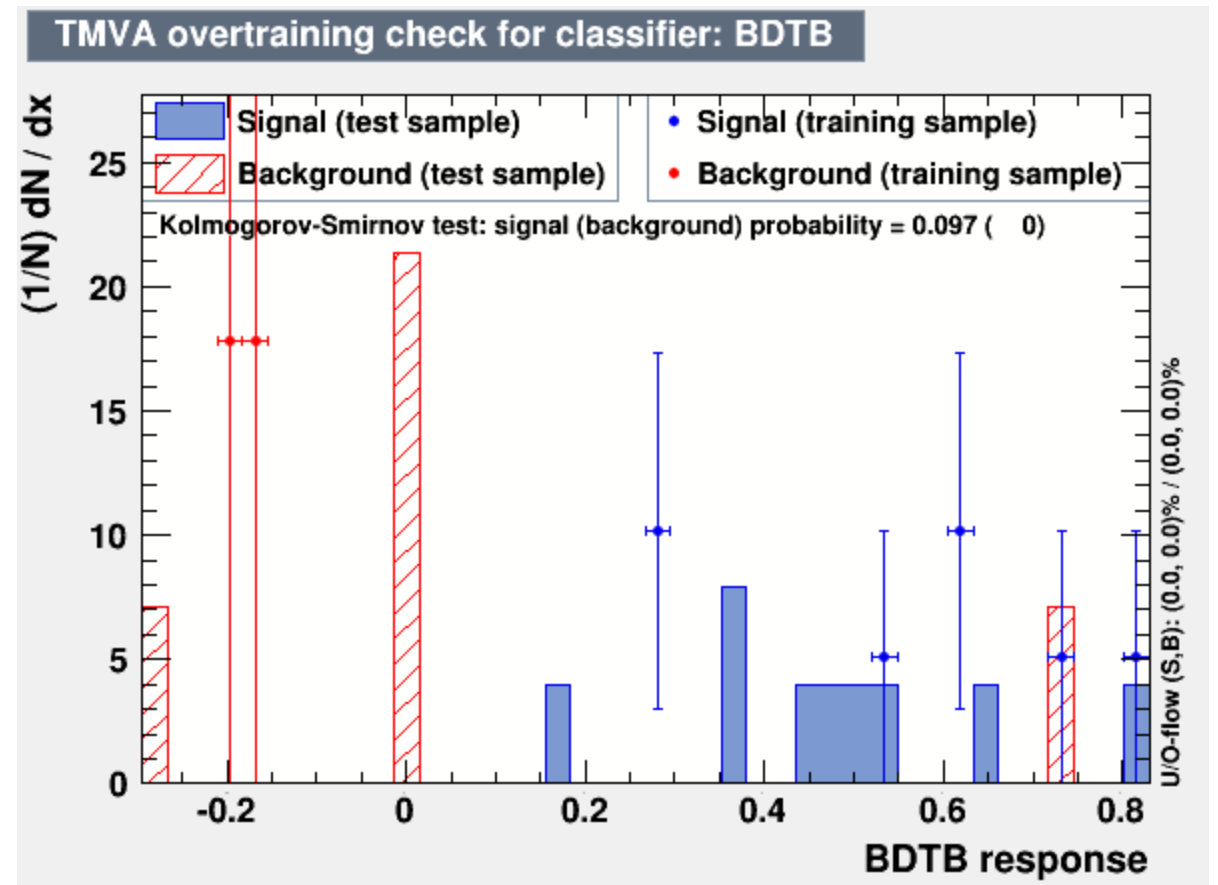
TMVA overtraining check for classifier: BDTB



Wash Out (WO) pre e post CR

Considerazioni

- Configurazione «di prova»:
 - Realizzata da R. Paramatti
 - sottoinsieme di variabili scelte in base al p-value
 - presenta una buona discriminazione -> possibile miglioramento miscelando variabili di macrogruppi diversi
- Bassa statistica -> possibile miglioramento con l'introduzione del Random Forest (in teoria più stabile in questo regime del BDTB)



Variabili:

Vol_tot_pre – SF0_sd_int – SF0_skewness_pre –
SF1_skewness_post – SF1_kurtosis_post – SF1_5_post –
SF1_8_kurtosis_post

To Do List

- Implementazione del Random Forest in ROOT 6
- Determinazione delle (eventuali) variabili discriminanti
- Studio ad alta statistica (MonteCarlo su pazienti fake e/o di LHC)
- Verifica stabilità nel limite di bassa statistica

Grazie dell'attenzione!