UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Instituto Universitario de Microelectrónica Aplicada

Master of Science in Telecommunication Technologies

Master Thesis

Semi-Supervised Classification of Hyperspectral Images for Brain Tumours detection

Patricia Beltran Alonso

Gustavo Marrero Callico, Samuel Ortega Sarmiento y Beatriz Martinez Vega

September 2021

Abstract

In the medical field, **hyperspectral** (HS) images have represented a technological breakthrough due to their non-invasive nature and because they provide useful information for the **diagnosis** of diseases. In many practical classification applications the number of available **labelled samples** is **limited**, and the amount of **unlabelled samples** is **large**. It is interesting to develop **algorithms** able to exploit both labelled and unlabelled samples in the classification process to obtain high-performance classifiers**. Semi-Supervised Learning** (SSL) is a **powerful tool** to generate learning models when the number of **labelled samples** is **low**. This project describes different methodologies of the design of semi-supervised algorithm for brain tumour detection. For the evaluation of these designs, the Support Vector Machines (SVM) and the Random Forest (RF) classifiers were employed.

Methodology

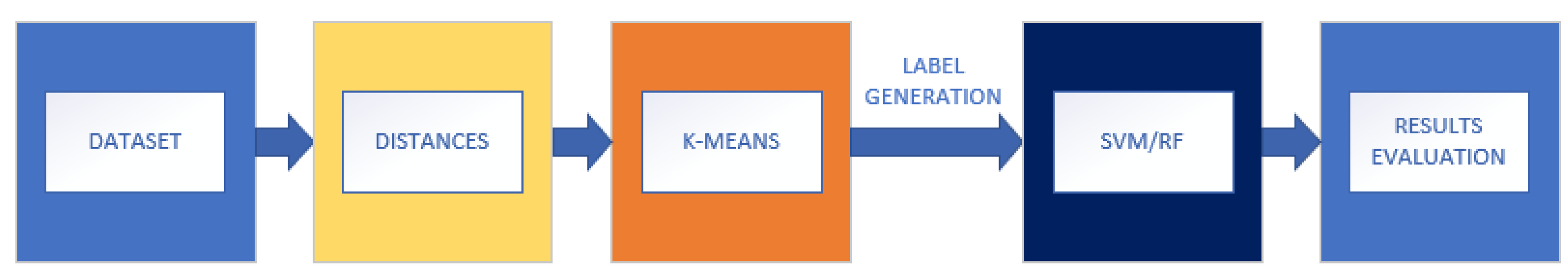
Proposed methodology

The motivation of this work is to **simulate a realistic case** in the operating room, where there is a previously labelled database and the new acquired data of the patient who is going to receive the intervention.

This **HS database** is composed by **26 HS cubes** belonging to a total of **16 different patients** diagnosed with **Glioblastoma primary brain tumour**.

The objective is to **include** this **current patient data** in the database with which to train the supervised classifiers. The methodology proposed to develop the semi-supervised classification of HS images of brain tumours is:

- To evaluated the **most suitable distance** metric.
- To evaluated which **value of the k parameter** best fits our database.
- To **generated** the **labels** of the new patient.



The new patient data labelled with this methodology and the dataset of the previous patients are fed into the classifiers, in order to train them, generate a model, and finally evaluate its performance.

The supervised classifiers selected for this work are the **Support Vector Machines** (SVM) and **Random Forests** (RF).

Results

Four case studies:

- **SVM** (k-means parameters: **cosine** distance**, k=10**)
  - Without condition (all clusters are used to generated the labels of the new patients).
  - With condition (only the cluster with a **high presence** of a certain class (**more than 60%**) are used.
- **RF** (k-means parameters: **cosine** distance, **k=15**)
  - Without condition (all clusters)
  - With condition (only three clusters).

SVM and RF algorithms were trained without semi-supervised vision in order to compare if using semi-supervised classification improves the results.

		Sensitivity				Specificity				
	OA	Normal	Tumour	Blood Vessel	Background	Normal	Tumour	Blood Vessel	Background	**Kappa**
Supervised process										
SVM	**78.77%**	93.00%	28.03%	87.44%	95.39%	81.96%	98.09%	93.32%	93.94%	-
RF	**76.99%**	97.04%	9.91%	89.79%	91.73%	76.92%	99.88%	95.67%	91.09%	**0.67**
Semi-supervised process										
SVM without Condition	**45.57%**	45.18%	1.90%	46.34%	64.69%	64.53%	99.62%	86.34%	57.10%	**0.25**
SVM with condition	**44.70%**	48.05%	1.43%	33.26%	87.28%	74.95%	99.30%	97.64%	46,.6%	**0.28**
RF without Condition	**46.56%**	44.80%	0.07%	33.96%	98.77%	96.55%	99.92%	97.67%	27.48%	**0.27**
RF (evaluating with three clusters)	**45.89%**	64.20%	0.03%	63.87%	70.99%	62.45%	100.00%	61.55%	77.32%	**0.19**

If focusing on the **semi-supervised process**, perhaps the **RF** evaluation approach **with three clusters** gives the **best sensitivity** results for all class types **except tumour**. The proposed processing method may not be adequate to improve the results. The semi-supervised algorithm proposal worsens the classification results compared to the non-semi-supervised.

Conclusions

It is considered that the image used in the semi-supervised **to automatically label** it and thus increase the database with which the model is generated, must be an image that does **not include any tumour pixels**. In this way we can ensure that when the automatic labelled is generated, there will be **no mislabelled tumour pixels.**

If we **improve the balance** of **specificity** and **sensitivity** of the rest of the classes, we will also be able to improve it for the tumour class.

IUMA
Universidad de Las Palmas de Gran Canaria
Campus Universitario de Tafira
Las Palmas de Gran Canaria, SPAIN
t: +34 928 451 150
+34 928 451 286
f: +34 928 451 083
e: iuma@iuma.ulpgc.es **w:** www.iuma.ulpgc.es

Poster TFM IUMA versión 3 – 12 febrero 2018