

# FHDO Biomedical Computer Science Group at Medical Classification Task of ImageCLEF 2015

Obioma Pelka and Christoph M. Friedrich

Department of Computer Science  
University of Applied Sciences and Arts Dortmund (FHDO)  
Emil-Figge-Strasse 42, 44227 Dortmund, Germany  
obioma.pelka@googlemail.com and christoph.friedrich@fh-dortmund.de  
<http://www.inf.fh-dortmund.de>

**Abstract.** This paper presents the modelling approaches performed by the FHDO Biomedical Computer Science Group for the compound figure detection and subfigure classification tasks at ImageCLEF 2015 medical classification. This is the first participation of the group at an accepted lab of the Cross Language Evaluation Forum. For image visual representation, various state-of-the-art visual features such as Bag-of-Keypoints computed with dense SIFT descriptors and the new Border Profile presented in this work, were adopted. Textual representation was obtained by vector quantisation on Bag-of-Words codebook generated using attribute importance derived from  $\chi^2$ -test and the Characteristic Delimiters feature presented in this paper. To reduce feature dimension and noise, the principal component analysis was computed separately for all features. Various multiple feature fusion were adopted to supplement visual image information with their corresponding textual information. Random forest models with 100 to 500 deep trees grown by resampling, a multi class linear kernel SVM with  $C = 0.05$  and a late fusion of the two classifiers were used for classification prediction. Six and Eight runs of submission categories: Visual, Textual and Mixed were submitted for the compound figure detection task and subfigure classification task, respectively.

**Keywords:** bag-of-keypoints, bag-of-words, compound figure detection, modality classification, medical imaging, image border profile, principal component analysis, random forest, support vector machine

## 1 Introduction

This paper describes the modelling methods and experiments performed by the FHDO Biomedical Computer Science Group (BCSG) at the ImageCLEF 2015 medical classification. This is the first participation of the BCSG, a research group from the University of Applied Sciences and Arts Dortmund, at the cross-language image retrieval track ImageCLEF [28] of the Cross Language Evaluation Forum (CLEF)<sup>1</sup>.

<sup>1</sup> <http://www.clef-initiative.eu/>

The ImageCLEF 2015 medical classification task consists of four subtasks: compound figure detection, multi-label classification, figure separation and sub-figure classification of which the BCSG participated in two subtasks [14]. The remaining of this paper is organised as follows: In section 2 for the sub-task compound figure detection, various image representations extracted are presented and the model classifier setup as well as submitted runs and their corresponding results are described. Modelling approach, submitted runs and results for the subfigure classification task are elaborated in section 3. Finally, conclusions are drawn in section 4.

## 2 Compound Figure Detection

### 2.1 Task Definition

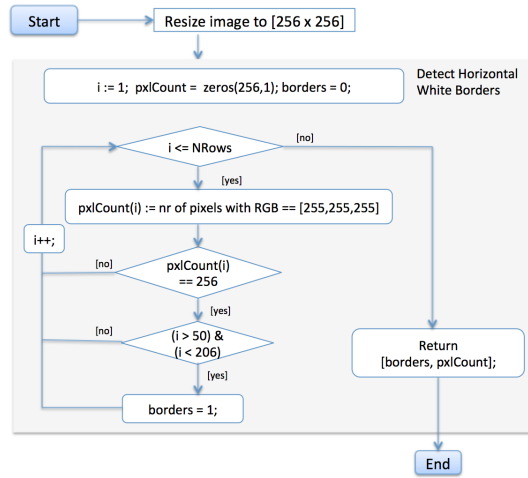
Several figures found in biomedical literature consist of several subfigures. To obtain efficient image retrieval on a given search, it is necessary that these figures are separated and not considered as single figures. The first step in achieving this goal is to detect these compound figures. The detailed task definition is presented in [14].

### 2.2 Visual Features

For the visual image representation, a combination of high level and low level features was pursued. This is an important step in order to have 'whole-image' and 'detail' representation of an image. The Bag-of-Keypoints feature and the new Border Profile feature specifically adapted for this task were used for the visual image representation. The feature definitions and extraction procedures are described in the following subsections.

**Border Profile:** A highly distinguishing feature characterising a compound figure is the existence of a separating border. These borders are usually of white or black color. Hence the first visual feature computed is to detect the presence of such horizontal and vertical black and white color profiles for all images. For comprehension, a white or black border is present when all pixels of a row or column have  $RGB_{value} = [255, 255, 255]$  or  $RGB_{value} = [0, 0, 0]$  respectively. To detect this presence, the functions listed in Table 1 were implemented and their respective results were concatenated to obtain the complete feature vector. Fig. 1 depicts a flowchart containing the steps computed for detection of white horizontal borders.

To visually demonstrate the outcomes of the functions in Table 1, compound figures separated with white as well as black borders were selected. The compound figure in Fig. 2 displays the central nervous system and skeletal involvement by breast cancer of a rat and was adapted from [25].

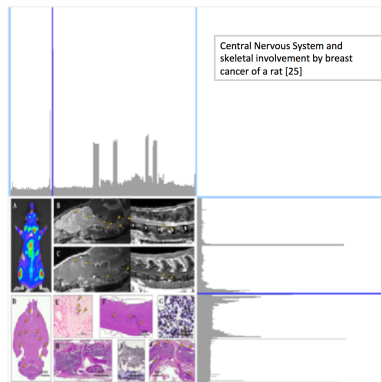


**Table 1.** Functions implemented for image border profile feature extraction

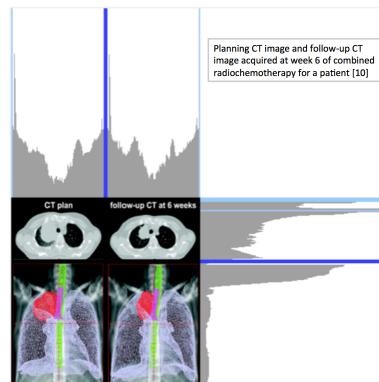
| Functions Implemented         |
|-------------------------------|
| detectWhiteHorizontalBorder() |
| detectWhiteVerticalBorder()   |
| detectBlackHorizontalBorder() |
| detectBlackVerticalBorder()   |

**Fig. 1.** *detectWhiteHorizontalBorder()* function, computed for detection of white horizontal border profiles

The horizontal and vertical bars adjoining the resized [256 x 256] figure show number of white pixels present in the rows and columns respectively. Considering that not all existing borders actually separate the existing subfigures, the next step is to detect and eliminate such frame borders. The cut-off threshold used was [1:50] and [206:256], i.e only borders located in the rows and columns [51:205] are treated as separating borders. The light blue bars in Fig. 2 and 3 show frame borders while the dark blue bars displays detected separating borders.



**Fig. 2.** Detected horizontal and vertical white separating borders as dark blue bars and frame borders as light blue bars



**Fig. 3.** Detected horizontal and vertical black separating borders as dark blue bars and frame borders as light blue bars

Compound figures can also be separated using borders with colors other than white. Figure 3 display the detection of horizontal and vertical black separating borders. The compound figure adapted from [10], shows a planning CT image and its corresponding follow-up CT image acquired at week 6 of combined radiochemotherapy of a patient. The same cut-off threshold outlined above was used.

**Bag-of-Keypoints:** For whole-image classification tasks, the bag of feature approach has achieved high accuracy results [29] and [18]. The motivation to this idea comes from bag-of-word approach used for text categorisation. Limitations of invariance present in [19] was eliminated in the comprehensively evaluated approach presented in [7], which has now become a common state of the art approach for image classification. They proposed a method called Bag-of-Keypoints (BoK) which is based on vector quantization of affine invariant descriptors of image patches. Apart from the invariance to affine transformation, another advantage that comes with this method is the simplicity.

The task here to tackle being a whole-image classification task, the Bag-of-Keypoints approach was adopted as a visual image representation. The functions used for this approach are from the VLFEAT library [27]. As visual descriptors, dense SIFT descriptors applied at several resolutions were uniformly extracted with an interval grid of 4 pixels using the VL-PHOW function. To speed up computational time,  $k$ -means clustering with approximated nearest neighbours (ANN) [15] was computed on these randomly chosen descriptors using the VL-KMEANS function, to partition the observations into  $k$  clusters so that the within-cluster sum of squares is minimised [12].

A maximum of 20 iterations was defined to allow the  $k$ -means algorithm converge. The cluster centres were initialised using random data points. As  $k = 12000$ , a codebook containing 12,000 keypoints was generated and was further optimised by adapting a kd-tree with metric distance  $L_2$  for quick nearest neighbour lookup using VL-KDTREEBUILD function.

### 2.3 Textual Features

Text representations for all images was derived from their figure caption. All figures in the ImageCLEF collection originate from biomedical literature published in PubMed Central<sup>2</sup>. The original figure caption and journal title are extracted from the provided XML files of this task.

**Bag-of-Words:** The Bag-of-Words (BoW) approach [24] is one of the common methods used for text classification. The basic concept here is to extract features by counting the frequency or presence of words in the text to be classified. These words have to be defined first in a dictionary or codebook. To generate the needed dictionary, all words from the captions of all images in the distributed

---

<sup>2</sup> <http://www.ncbi.nlm.nih.gov/pmc/>

collection were extracted. Several text processing procedures such as removal of stop-words and stemming using PorterStemmer [23] were induced to obtain a positive effect on computational time performance. The occurrence (%) for all words in both classes was computed. Words with less than 85% difference between the two classes were eliminated to further reduce the dictionary size. For the BoW representation two dictionaries were created:

- **Dictionary<sub>1</sub> (D<sub>1</sub>):** 455 words obtained with porter stemming, removal of stop-words and word occurrence.
- **Dictionary<sub>2</sub> (D<sub>2</sub>):** 3906 words obtained with removal of stop-words and word occurrence.

The benefit of  $\chi^2$ -test and Information Gain have been investigated, but not further used since no relevant advantage was detected during feature selection.

**Charateristic Delimiters:** When captions of compound figures are written, it is most likely that existing subfigures are addressed using some delimiter. Depending on the certainty that a figure can only be called 'compound figure' when it contains at least two subfigures, the presence of two delimiters was determined.

To achieve this task, a set of possible double delimiters characterising compound figures was computed. This step was manually done by analysing the captions of compound figures from the training set and selecting words with very high occurrence. Such words that appear often and hence significantly characterise the presence of subfigures are referred in this work as 'Characteristic Delimiters'. A sub-collection of delimiters used are listed in Table 2.

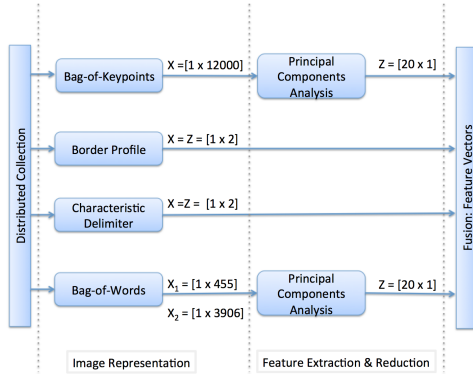
If existence of a delimiter pair is detected in the caption of an image, the figure is textually represented by assigning the value [1, 1] and otherwise [0, 0] to the feature vector.

**Table 2.** Delimiters characterising captions of compound figures

| First subfigure | Second subfigure | First subfigure | Second subfigure |
|-----------------|------------------|-----------------|------------------|
| A.              | B.               | 1.              | 2.               |
| (A).            | (B).             | (1).            | (2).             |
| A).             | B).              | 1).             | 2).              |
| Lower           | Upper            | Left            | Right            |
| i.              | ii.              | I).             | II).             |

## 2.4 Classifier Setup

A fusion of all textual and visual representations will result to a feature vector with 15910 columns. To model an efficient and effective classifier, the feature dimensions and noise is reduced using the principal component analysis [8]. The principal component analysis is separately computed on each feature vector group as shown in Fig. 4. Subsequently, the best number of principal components needed to describe the feature were estimated by model selection.



**Fig. 4.** Original and truncated vector size for extracted visual and textual representations after dimension and noise reduction with principal components analysis

**Table 3.** Adjustment in accuracy on classification model Run4 by omitting a feature in comparison to when all features are used. The evaluation was done ex-post

| Feature                  | Loss of Eval. Accuracy (%) | Loss of Dev. Accuracy (% ± sd) |
|--------------------------|----------------------------|--------------------------------|
| BoK                      | -4.50                      | -2.63(±0.54)                   |
| Border Profile           | -2.65                      | -2.69(±0.58)                   |
| Characteristic Delimiter | -0.94                      | -1.62(±0.63)                   |
| BoW                      | -1.81                      | -1.09(±0.63)                   |

The feature vector Border Profile and Characteristic Delimiter have both 2 columns and hence do not need any dimension reduction. Different combinations of the derived principal components are concatenated to obtain the final feature vector used for training the classifier. These combinations are the various runs submitted for evaluation. Table 3 lists the effects on prediction accuracy when certain features are left out during the feature fusion stage. In this ex-post analysis, the contribution (%) for each feature was computed by applying the classifier model of Run4 on the evaluation set and on 10 sampled learning and validation sets. It can be seen that all features contribute positively.

The distributed collection was split into 10 different learning and validation sets using the bootstrap algorithm [9]. For category prediction, a random forest (RF) classifier [2] using the FITENSEMBLE function from the MATLAB software package [22] was modelled. The list below is an excerpt of several parameters used to tune the classifier model.

- Number of Trees = 200
- Number of Leaf Size = [0.04, 0.06, 0.3]
- Split Criterion = Exact
- Ensemble grown = By resampling

## 2.5 Submitted Runs

In this section, the six compound figure detection runs submitted by the Biomedical Computer Science Group for evaluation are presented.

- **task1\_run1\_mixed\_stemDict:** A combination of BoW with Dictionary<sub>1</sub> textual features and BoK visual features was used to train the classifier.

- **task1\_run2\_mixed\_sparse1**: Visual features: Border profile and Characteristic Delimiter combined with textual features derived from the BoW Dictionary<sub>1</sub>.
- **task1\_run3\_mixed\_sparse2**: Same as **run2** without the BoW textual representation.
- **task1\_run4\_mixed\_bestComb**: Fusion of all features described. BoW features extracted using Dictionary<sub>2</sub>.
- **task1\_run5\_visual\_sparseSift**: This random forest model classifier is trained only with the visual features: Bag-of-Keypoints and Border Profile.
- **task1\_run6\_text\_sparseDict**: Model was trained only with the textual features BoW with Dictionary<sub>1</sub> and Characteristic Delimiter.

## 2.6 Results

Six runs (four Mixed, one Visual and one Textual) were submitted for evaluation. Table displays the official evaluation accuracy and retrieval type for each run. The fourth column displays the standard mean accuracy and standard deviation achieved on 10 sampled learning and validation set derived using the bootstrap algorithm.

**Table 4.** Official evaluation accuracy of submitted runs showing retrieval type with standard mean accuracy and corresponding standard deviation achieved on 10 sample validation sets using the same modelling approach

| Run ID                        | Retrieval Type | Evaluation Accuracy (%) | Development Accuracy (% ± sd) |
|-------------------------------|----------------|-------------------------|-------------------------------|
| task1_run1_mixed_stemDict     | Mixed          | 83.88                   | 85.64 (±0.44)                 |
| task1_run2_mixed_sparse1      | Mixed          | <b>85.39</b>            | 85.71 (±0.54)                 |
| task1_run3_mixed_sparse2      | Mixed          | 80.07                   | 86.24 (±0.54)                 |
| task1_run4_mixed_bestComb     | Mixed          | 73.32                   | <b>90.11</b> (±0.36)          |
| task1_run5_visual_sparseSift  | Visual         | 72.51                   | 85.69 (±0.48)                 |
| task1_run6_textual_sparseDict | Textual        | 78.34                   | 85.17 (±0.45)                 |

## 3 Subfigure Classification

### 3.1 Task Definition

Clinicians have implied on the importance of the modality of an image in several user-studies. The usage of modality information significantly increases the retrieval efficiency, thus image modality has become an essential and relevant factor regarding medical information retrieval [13]. The subfigure classification subtask aims to evaluate approaches that automatically predict the modality of

medical images from biomedical Journals. For further task definition, refer to [14].

Some image categories were represented by few annotated examples, thus the expansion of the original collection was strived in order to counteract the imbalanced dataset. Additional datasets created are described below:

1. **DataSet<sub>1</sub> (DS<sub>1</sub>):** The original training collection distributed for the sub-figure classification task in ImageCLEF2015 Medical Classification.
2. **DataSet<sub>2</sub> (DS<sub>2</sub>):** Additive to DataSet<sub>1</sub>, the complete collection distributed in ImageCLEF2013 AMIA<sup>3</sup> Medical Task. The collection contains over 300,000 images from over 45,000 biomedical research articles of the PubMed Central Repository hosted by the U.S. National Library of Medicine.
3. **DataSet<sub>3</sub> (DS<sub>3</sub>):** Additive to DataSet<sub>1</sub>, the collection distributed for the Modality Classification ImageCLEF2013 AMIA medical subtask. This is a sub-collection of DataSet<sub>2</sub> and contains figures annotated into 31 categories. Figures belonging to the compound figure 'COMP' category were eliminated to attain the same categories in DataSet<sub>1</sub>.
4. **DataSet<sub>4</sub> (DS<sub>4</sub>):** The sub-collection for the Modality Classification task in ImageCLEF 2013 AMIA medical task without the 'COMP' category.

### 3.2 Visual Features

Over the years, various techniques for medical imaging have been developed. Each having not only its advantages and disadvantages, but also different acquiring technique. Hence various feature extracting methods are needed to apprehend the possible characteristics of medical images [5]. In addition, images have to be completely represented, i.e. 'whole image' and 'detail' representation. This can be acquired by extracting global and local features. These features: BAF, Gabor, JCD, Tamura, PHOG were extracted using functions from the LIRE: Lucene Image Retrieval library [21].

- **Bag-of-Keypoints:** Visual image representation using the Bag-of-Keypoints approach described in the subsection 2.2. With the distinction, that three different datasets were used to create various codebooks accordingly.
- **BAF:** The global features (brightness, clipping, contrast, hueCount, saturation, complexity, skew and energy) represented as a 8-dimensional vector.
- **CEDD:** Low-level feature CEDD (Color and Edge Directivity Descriptor) [4] incorporating color and textures information were extracted and represented as a 144-dimensional vector.
- **FCH:** The Fuzzy Color Histogram considers through fuzzy-set membership function the color similarity of each pixel's color to all histogram bins and is represented as a 10-dimensional vector using the fuzzy linking method [11],[17].
- **Gabor:** A 60-dimensional vector was used to represent texture features based on Gabor functions.

<sup>3</sup> <http://www.imageclef.org/2013/medical>



- **JCD:** The Joint Composite Descriptor (JCD) is a combination of two Compact Composite Descriptors: Color and Edge Directivity Descriptor (CEDD) and Fuzzy Color Texture Histogram (FCTH) [4]. The feature made up of merging the texture areas of CEDD and FCTH was represented as a 168-dimensional vector.
- **Tamura:** The Tamura features consisting of six basis textural feature: coarseness, contrast, directionality, line-likeness, regularity and roughness, were represented as a 18-dimensional vector [26].
- **PHOG:** The Pyramid of Histograms of Oriented Gradients (PHOG) feature proposed in [1] represents an image by its local shape and the spatial layout of the shape. A 630-dimensional vector was used for feature representation.

### 3.3 Textual Features

Similar to the compound figure detection task, textual representation for the figures was adopted using their corresponding captions.

**Bag-of-Words:** The process of textual representation executed is complementary to the process for the compound figure detection task described in subsection 2.3. With an adjustment in dictionary generation and word selection method. The figures distributed for the subfigure classification task are subfigures extracted from compound figures, hence their corresponding captions actually describe compound figure and not the single subfigures. Considering that multipane figures consist of subfigures not only from the same category but also from multiple categories, using the original captions to represent the subfigures will not lead to a valuable characterisation.

To overcome this limitation, the dictionary was built using the DataSet<sub>4</sub>. The figures in this dataset do not originate from multipane figures and thus have characteristic captions that can be mapped to the 30 subfigure categories. All words from all captions were retrieved, removal of stop-words and stemming was done in the text preprocessing stage. To develop a dictionary containing relevant words for each category, vector quantisation on all figures and the  $\chi^2$ -test [6] was computed on the derived matrix. With this step, attribute importance for all words was effectuated.

A dictionary with 438 words was finally obtained by selecting words with attribute importance over a fixed cutoff threshold. The captions of the subfigures were trimmed to relevance using the characteristic delimiters presented in subsection 2.3 before vector quantisation on the generated dictionary was performed.

### 3.4 Classifier Setup

Contrary to the compound figure detection task, not only a random forest classifier model was used. A multiclass linear kernel SVM from the libSVM library [3]

was modelled to compare prediction accuracies between the two classifier models, as it has been a popular approach in former ImageCLEF medical challenges [13]. The cost parameter used was  $C = 0.05$ . The random forest model was tuned with the same parameters mentioned in subsection 2.4. Ten samples of learning and validation sets were obtained using the bootstrap algorithm [9].

To reduce computational time, feature dimension and noise reduction was achieved using the principal component analysis. All features beside the BAF features were reduced using this method. Table 5 presents the original and truncated vector size after computing the principal component analysis on each feature. The contribution of a feature to prediction performance is an important attribute that assists efficient feature selection. To obtain each feature contribution, the difference between the accuracy when all features are combined and the accuracy when a certain feature is omitted was calculated and displayed in the fourth column of Table 5. The feature contribution analysis was done ex-post. The prediction accuracy used for this analysis was computed by applying the classifier model Run1 on the original evaluation set.

**Table 5.** Descriptors with original and truncated feature vector size and feature contribution (%) evaluated on classifier model Run1. The evaluation done ex-post

| Descriptor       | Original Vector Size | Reduced Vector Size | Loss of Prediction Accuracy (%) |
|------------------|----------------------|---------------------|---------------------------------|
| Bag-of-Keypoints | 12000                | 25                  | -2.99                           |
| Bag-of-Words     | 438                  | 40                  | -6.42                           |
| BAF              | 8                    | 8                   | -4.06                           |
| CEDD             | 144                  | 5                   | -0.49                           |
| FCH              | 10                   | 3                   | -0.67                           |
| Gabor            | 60                   | 0                   | 00.00                           |
| JCD              | 168                  | 5                   | -0.43                           |
| Tamura           | 18                   | 2                   | -0.76                           |
| PHOG             | 630                  | 2                   | +0.27                           |

Drawing conclusions using Table 5, it can be seen that omitting most of the extracted features leads to a negative effect on prediction performance. The representations BoK, BoW and BAF have the most contributions. In contrary, the omission of PHOG feature has a positive effect on the prediction performance and hence increases the evaluation accuracy with +0.27%. The principal components computed from Gabor image representation did not improve the prediction accuracy and was omitted from the final fused feature vector used for classification.

### 3.5 Submitted Runs

The BCSG submitted eight (six Mixed, one Textual and one Visual) runs for evaluation. The several fusion approaches defining the submitted runs are displayed in Table 6. In addition for each set, the prediction performance obtained on 10 sampled learning and validation set using the same modelling approach is listed.

**Table 6.** Breakdown of feature fusion approach applied for the subfigure classification submitted runs

| Run ID: task4_run  | Submission Type Category | Classifier Model | DataSet used for Training | Bag-of-Keypoints Codebook Build | Bag-of-Words Codebook Build | Vector Size Feature Fusion | Official Evaluation Accuracy (%) | Development Mean Accuracy (% $\pm$ sd) |
|--------------------|--------------------------|------------------|---------------------------|---------------------------------|-----------------------------|----------------------------|----------------------------------|----------------------------------------|
| 1.combination      | Mixed                    | Random Forest    | DS <sub>1</sub>           | DS <sub>3</sub>                 | DS <sub>4</sub>             | 90                         | 66.48                            | <b>91.27</b><br>( $\pm 0.40$ )         |
| 2.visual           | Visual                   | Random Forest    | DS <sub>1</sub>           | DS <sub>2</sub>                 | DS <sub>4</sub>             | 50                         | 60.91                            | 87.58<br>( $\pm 0.61$ )                |
| 3.textual          | Textual                  | Random Forest    | DS <sub>1</sub>           | DS <sub>1</sub>                 | DS <sub>4</sub>             | 40                         | 60.91                            | 79.73<br>( $\pm 0.87$ )                |
| 4.clean_rf         | Mixed                    | Random Forest    | DS <sub>1</sub>           | DS <sub>1</sub>                 | DS <sub>4</sub>             | 90                         | 67.24                            | 91.15<br>( $\pm 0.40$ )                |
| 5.train_20152013   | Mixed                    | Random Forest    | DS <sub>3</sub>           | DS <sub>3</sub>                 | DS <sub>4</sub>             | 90                         | <b>67.60</b>                     | 82.92<br>( $\pm 0.59$ )                |
| 6.clean_libnorm    | Mixed                    | LibSVM           | DS <sub>1</sub>           | DS <sub>3</sub>                 | DS <sub>4</sub>             | 90                         | 64.34                            | 88.95<br>( $\pm 0.84$ )                |
| 7.clean_comb.librf | Mixed                    | LibSVM RF        | DS <sub>1</sub>           | DS <sub>1/3</sub>               | DS <sub>4</sub>             | 90                         | 65.99                            | 88.85<br>( $\pm 0.76$ )                |
| 8.clean_short_rf   | Mixed                    | Random Forest    | DS <sub>1</sub>           | DS <sub>3</sub>                 | DS <sub>4</sub>             | 90                         | 66.44                            | 88.62<br>( $\pm 0.64$ )                |

### 3.6 Results

The BCSG submitted runs in all submission categories: Visual, Textual and Mixed. Most of the submitted runs belong to the submission category 'Mixed' which is a combination of textual and visual representation. This decision was made because not only were better accuracies obtained during development, but also evaluation results presented by other ImageCLEF participant groups in the previous years tasks have proven to be better when the 'Mixed' submission category is induced [13],[16]. Figure 5 depicts the achieved performance of all submitted runs for the subfigure classification task. Runs belonging to the Biomedical Computer Science Group are represented in as colored bars and the gray bars represent submissions of other participants.



## 4 Conclusions

Various classification prediction approaches based on multiple feature fusion and combination of classifier models were explored for the ImageCLEF 2015 medical classification task. Negative differences in the prediction performance were observed when the Bag-of-Keypoints representation was computed using SIFT [20] instead of dense SIFT descriptors, feature vectors weren't normalised and single precision format was used rather than double precision format to define floating-points numbers. The discrepancy between prediction performance on the evaluation set and on the sampled learning and validation sets is assumed to be an overfitting problem. Supplementing visual image representation with corresponding textual representation proved to be a beneficial strategy regarding classification accuracy. Omitting any of the described features apart from the PHOG feature, results to a negative decrease on the official evaluation accuracy. The proposed Border Profile image representation could be further enhanced by implementing additional functions to detect border profile of colors other than black and white.

## References

1. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM International Conference on Image and Video Retrieval. pp. 401–408. CIVR '07, ACM, New York, NY, USA (2007)
2. Breiman, L.: Random forests. *Mach. Learn.* 45(1), 5–32 (2001)
3. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2, 27:1–27:27 (2011)
4. Chatzichristofis, S.A., Boutalis, Y.S.: Compact Composite Descriptors for Content Based Image Retrieval: Basics, Concepts, Tools. VDM Verlag, Saarbrücken, Germany (2011)
5. Chen, C.h.: Computer vision in medical imaging. World Scientific (2013)
6. Cochran, W.G.: The  $\chi^2$  test of goodness of fit. *Ann. Math. Statist.* 23(3), 315–345 (1952)
7. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: In Workshop on Statistical Learning in Computer Vision, ECCV. pp. 1–22 (2004)
8. Dunteman, G.H.: Principal Components Analysis. Sage University paper. Quantitative applications in the social sciences, Sage publications, Newbury Park, London, New Delhi (1989)
9. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. Chapman & Hall, New York (1993)
10. Guckenberger, M., Baier, K., Richter, A., Wilbert, J., Flentje, M.: Evolution of surface-based deformable image registration for adaptive radiotherapy of non-small cell lung cancer (NSCLC). *Radiation Oncology* 4(68), 2169–2178 (2009)
11. Han, J., Ma, K.K.: Fuzzy color histogram and its use in color image retrieval. *IEEE Transactions on Image Processing* 11(8), 944–952 (2002)
12. Hartigan, J.A., Wong, M.A.: A k-means clustering algorithm. *JSTOR: Applied Statistics* 28(1), 100–108 (1979)

13. García Seco de Herrera, A., Kalpathy-Cramer, J., Demner Fushman, D., Antani, S., Müller, H.: Overview of the ImageCLEF 2013 medical tasks. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum) (2013)
14. García Seco de Herrera, A., Müller, H., Bromuri, S.: Overview of the ImageCLEF 2015 medical classification task. In: Working Notes of CLEF 2015 (Cross Language Evaluation Forum) (2015)
15. Indyk, P., Motwani, R.: Approximate nearest neighbors: Towards removing the curse of dimensionality. In: Proceedings of the 30th Annual ACM Symposium on Theory of Computing. pp. 604–613. STOC '98, ACM, New York, NY, USA (1998)
16. Kalpathy-Cramer, J., García Seco de Herrera, A., Demner-Fushman, D., Antani, S., Bedrick, S., Müller, H.: Evaluating performance of biomedical image retrieval systems— an overview of the medical image retrieval task at ImageCLEF 2004–2014. *Computerized Medical Imaging and Graphics* (2014)
17. Konstantinidis, K., Gasteratos, A., Andreadis, I.: Image retrieval based on fuzzy color histogram processing. *Optics Communications* 248(4–6), 375 – 386 (2005)
18. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2. pp. 2169–2178. CVPR '06 (2006)
19. Li, S.Z., Zhu, L., Zhang, Z., Blake, A., Zhang, H., Shum, H.: Statistical learning of multi-view face detection. In: In Proceedings of the 7th European Conference on Computer Vision. pp. 67–81 (2002)
20. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 91–110 (2004)
21. Lux, M., Chatzichristofis, S.A.: Lire: Lucene image retrieval – an extensible java cbir library. In: El-Saddik, A., Vuong, S., Griwodz, C., Bimbo, A.D., Candan, K.S., Jaimes, A. (eds.) *ACM Multimedia*. pp. 1085–1088. ACM (2008)
22. MATLAB: version 8.5.0.197613 (R2015a). The MathWorks Inc., Natick, Massachusetts (2015)
23. Porter, M.: An algorithm for suffix stripping. *Program-electronic Library and Information Systems* 14, 130–137 (1980)
24. Salton, G., McGill, M.J.: Introduction to modern information retrieval. McGraw-Hill computer science series, McGraw-Hill, New York (1983)
25. Song, H.T., Jordan, E.K., Lewis, B.K., Liu, W., Ganjei, J., Klaunberg, B., Despres, D., Palmieri, D., Frank, J.A.: Rat model of metastatic breast cancer monitored by MRI at 3 tesla and bioluminescence imaging with histological correlation. *Journal of Translational Medicine* 7 (2009)
26. Tamura, H., Mori, S., Yamawaki, T.: Texture features corresponding to visual perception. *IEEE Transactions on System, Man and Cybernetic* 6 (1978)
27. Vedaldi, A., Fulkerson, B.: Vlfeat: An open and portable library of computer vision algorithms. In: Proceedings of the International Conference on Multimedia. pp. 1469–1472. MM '10, ACM (2010)
28. Villegas, M., Müller, H., Gilbert, A., Piras, L., Wang, J., Mikolajczyk, K., de Herrera, A.G.S., Bromuri, S., Amin, M.A., Mohammed, M.K., Acar, B., Uskudarli, S., Marvasti, N.B., Aldana, J.F., del Mar Roldán García, M.: General Overview of ImageCLEF at the CLEF 2015 Labs. *Lecture Notes in Computer Science*, Springer International Publishing (2015)
29. Zhang, H., Berg, A.C., Maire, M., Malik, J.: Svm-knn: Discriminative nearest neighbor classification for visual category recognition. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2. pp. 2126–2136. CVPR '06 (2006)