AUTOMATED DETECTION OF INDIVIDUAL MICRO-CALCIFICATIONS FROM MAMMOGRAMS USING A MULTI-STAGE CASCADE APPROACH

Zhi Lu†  Gustavo Carneiro†  Neeraj Dhungel‡  Andrew P. Bradley* *

†Australian Centre for Visual Technologies, The University of Adelaide  ‡Phenomics and Bioinformatics Research Centre, University of South Australia  *School of ITEE, The University of Queensland

ABSTRACT

In mammography, the efficacy of computer-aided detection methods depends, in part, on the robust localisation of micro-calculifications (μCs). Currently, the most effective methods are based on three steps: 1) detection of individual μC candidates, 2) clustering of individual μC candidates, and 3) classification of μC clusters. Where the second step is motivated both to reduce the number of false positive detections from the first step and on the evidence that malignancy depends on a relatively large number of μC detections within a certain area. In this paper, we propose a novel approach to μC detection, consisting of the detection and classification of individual μC candidates, using shape and appearance features, using a cascade of boosting classifiers. The final step in our approach then clusters the remaining individual μC candidates. The main advantage of this approach lies in its ability to reject a significant number of false positive μC candidates compared to previously proposed methods. Specifically, on the INbreast dataset, we show that our approach has a true positive rate (TPR) for individual μCs of 40% at one false positive per image (FPI) and a TPR of 80% at 10 FPI. These results are significantly more accurate than the current state-of-the-art, which has a TPR of less than 1% at one FPI and a TPR of 10% at 10 FPI. Our results are competitive with the state of the art at the subsequent stage of detecting clusters of μCs.

Index Terms— Micro-calcification, Mammogram, Cascade of Boosting Classifiers

1. INTRODUCTION

Breast cancer is the most diagnosed cancer amongst women worldwide, with 23% of all diagnosed cancers [1]. Breast screening programs aim to detect breast cancer at its early stages, when treatment is generally more effective [2]. These programs are usually based on the analysis of mammograms, where one of the main goals is the detection of micro-calculifications (μC) given that almost half of all breast cancers are associated with μC [3]. As shown in Fig. 1, μCs are represented by tiny calcium deposits that are displayed as small white spots on a mammogram. Their automated localisation by computer-aided detection (CADe) methods has the potential to streamline mammogram analysis and reduce the inter-user variance of μCs [4].

The current state-of-the-art methods for the automated detection of μCs consists of the following standard pipeline: S.1) detection of individual μC candidates, S.2) clustering of individual μC candidates based on their geometric distribution, and S.3) classification of those μC clusters. The fact that step S.1 above usually produces a large number of false positive individual μCs combined with the evidence that calculation malignancy is correlated with clusters of μCs [3] has motivated the inclusion of S.2 and S.3. These last two steps are able to eliminate a large number of isolated false positive μC detections, but they often fail to reject individual false positive μCs within clusters, which can potentially bias the analysis of a mammogram. In this paper, we propose a novel pipeline comprising the following steps: P.1) detection of individual μC candidates, P.2) classification of individual μC candidates, and P.3) clustering of individual μC candidates based on their geometric distribution. Compared to the standard pipeline, steps P.1 and S.1 are the same, step P.2 is new, step P.3 is the same as step S.2, and step S.3 has been removed. We have two goals with our new approach: 1) a significant reduction of the number of false positive individual μC detections, especially within true positive clusters, and 2) competitive detection rate of μC clusters. A quantitative analysis of our approach is performed using the publicly available INBreast dataset [5], where the main results obtained show that our method achieves a true positive detection of individual μCs (TPR) of 40% at 1 false positive detection per image.
the availability of a training set in step P.2. Each step is explained below, where we assume

tures [13]; and P.3) clustering of the individual
dates using pixel-based cascade of boosting classifiers [12]. Given that this new process is effective for remov-
ing false positive detections [6]. However, we observed that within a cluster, false positive detections were still prevalent. This motivated us to propose a new pipeline with the intro-
duction of a classification step between standard steps S.1 and S.2 that filters out individual \( \mu C \) candidates using their shape and appearance features with a cascade of boosting classi-
fiers [12]. Given that this new process is effective for removing false positive \( \mu C \) detections, we no longer require step S.3 (cluster classification).

3. METHODOLOGY

The proposed methodology consists of a initial pre-processing step based on quantum noise equalisation [6], which is followed by three steps: P.1) detection of individual \( \mu C \) candidates using pixel-based cascade of boosting classifiers [12] and Haar like features [6]; P.2) classification of the individual \( \mu C \) detections (from step P.1) using a region-based cascade of boosting classifiers [12] with appearance and shape features [13]; and P.3) clustering of the individual \( \mu C \)s detected in step P.2. Each step is explained below, where we assume the availability of a training set \( D = \{ (x_i, M_i) \}_{i=1}^N \), where \( x : \Omega \rightarrow \mathbb{R} \) denotes the mammogram (\( \Omega \) represents the image lattice) and \( M = \{ y_j \}_{j=1}^S \) represents the set of \( \mu C \) annotations for image \( i \) with \( y : \Omega \rightarrow \{0,1\} \) (i.e., each \( \mu C \) annotation is a binary map, where pixels of \( y_{i,j} \) labelled with 1 denote part of the \( j^{th} \) \( \mu C \) of \( i^{th} \) image).

Pre-processing: Our pre-processing is based on quantum noise equalisation proposed by Bria et al. [6], where the source of noise fluctuations in full-field digital mammograms (FFDM) can be described by a Poisson distribution with a standard deviation that can be estimated from the image.

Step P.1: Detection of individual \( \mu C \) candidates: This first step consists of a pixel-based classifier [6], represented by \( H(q) \), which estimates the likelihood that the pixel \( q \in \Omega \) represents part of a \( \mu C \) given the information extracted from a sub-window of size \( M \times M \) around the pixel. This classifier is represented by a cascade of boosting classifiers, where a pixel \( q \) is accepted to be part of a \( \mu C \) if it is positively classified by all stages of the cascade. In this cascade classifier [14], the detection rate \( D \) and false positive rate \( F \) of a cascade with \( S \) stages are computed with \( D = \prod_{s=1}^S d_s \) and \( F = \prod_{s=1}^S f_s \), where \( d_s \) and \( f_s \) represent the detection and false positive rates of stage \( s \). Therefore, if \( d_{s=0.99} \) and \( f_{s=0.3} \) and \( S = 5 \), then \( D = 0.951 \) and \( F = 0.002 \).
The training of the classifier at each cascade stage $s \in \{1, \ldots, S\}$, denoted by $H_s(p)$, uses a set of positive samples $\mathcal{P}_s$ and negative samples $\mathcal{N}_s$, where each sample $x_{M \times M}(q) \in \mathcal{P}_s$ consists of a sub-image of $x$ of size $M \times M$ centred at position $q$, such that one of the $J$ $\mu$Cs annotations contains $y(q) = 1$ (a negative sample is similarly defined with $y(q) = 0$). The main issue in training such classifier is the fact that $|\mathcal{N}_s| >> |\mathcal{P}_s|$, and this is solved by under-sampling the negative set, such that the proportion $|\mathcal{N}_s|/|\mathcal{P}_s|$ is constant over the training of each cascade stage. The classifier utilised in this work is the RUSBoost \cite{12}, which is designed to deal with such class imbalance with this under-sampling procedure. Finally, the feature set used is the Haar-like features \cite{14}, which are efficiently computed using integral image \cite{14} (note that we use a set of 1,697 features instead of the original 14,709 features from \cite{6} as this smaller set is faster to train and we did not notice a significant difference in the results). The final part of this step consists of finding the connected components of the pixel-based classification to form the $\mu$C candidates, where connected components that have width and length larger than 1 mm are removed because they represent macro-calculations that are not to be processed further \cite{6}. The step P.1 is defined by:

$$\{\tilde{y}_k, d_k\}_{k=1}^K = f_1(x, \theta_1),$$

where $\tilde{y}_k$ denotes a binary map of the $k^{th}$ $\mu$C candidate, $\theta_1$ is the classifier parameter set, and $d_k \in \mathbb{R}^4$ represents the top-left and bottom-right corner coordinates of the bounding box of this detection.

**Step P.2: Classification of individual $\mu$C detections with shape and appearance features**

The contribution of this paper consists of this individual $\mu$C classification step, where we extract a large set of shape and appearance features \cite{13} from each $\mu$C candidate in \cite{1}, and use a second cascade of RUSBoost \cite{12} classifiers to further eliminate false positive $\mu$C detections. These features are extracted with:

$$z = g(x, d, y).$$

A set of 11 shape features are calculated from $y$ in \cite{2}, which describe the following geometric information: area, perimeter, ratio of perimeter to area, rectangularity, circularity, and etc. Another set of 27 appearance features in \cite{2} are calculated from the sub-image of $x$ limited by the bounding box $d$, consisting of information (energy, correlation, entropy, inertia, and etc.) extracted from the spatial grey level dependence (SGLD) matrix \cite{13} \cite{15}. In addition, we compute the 1,697 Haar-like features of step P.1 and the local binary pattern (LBP) \cite{16} from the sub-image $x$ limited by the bounding box $d$. The step P.2 is defined by:

$$\{\tilde{y}_l, d_l\}_{l=1}^{L} = f_2(x, \{\tilde{y}_k, d_k\}_{k=1}^K, \theta_2),$$

which selects a subset of the detections from step P.1, with $L \leq K$, where $\theta_2$ is the parameter set of the classifier.

**Step P.3: Clustering of individual $\mu$C detections**

The clustering of the $\mu$C detections $\{\tilde{y}_l, d_l\}_{l=1}^{L}$ from step P.2 is based on the following algorithm \cite{6}: 1) construction of a weighted graph formed by nodes represented by the centroid of the detected $\mu$Cs, and edges that connect nodes that are closer than 10 mm; and 2) estimation of clusters from the connected components of this graph, where clusters with fewer than 3 $\mu$Cs are rejected. The step P.3 is defined by:

$$C = f_3(x, \{\tilde{y}_l, d_l\}_{l=1}^{L}, \theta_3),$$

where $\theta_3$ is the parameter set of the classifier, and $C$ represents the set of clusters, where each element of this set is formed by a graph computed from a subset of $\{\tilde{y}_l, d_l\}_{l=1}^{L}$ from step P.2.

4. MATERIALS AND METHODS

The experiments use the INBreast dataset \cite{5}, which contains 115 cases with 410 images, where 19 cases have no findings, 68 cases have benign findings and 28 cases have malignant findings (note that findings include $\mu$Cs and masses), where 6,880 individual $\mu$Cs have been identified by two radiologists. The experiments are performed using this dataset for the following reasons: it is a public domain (allowing direct comparison with other methods) full-field digital dataset where the individual manual $\mu$C annotations are both precise and reliable. In order to evaluate the detection of $\mu$C clusters, we produce the annotation of $\mu$C clusters using step P.3 of Sec. \cite{5} from the individual $\mu$C manual annotations. We perform a quantitative evaluation of the individual $\mu$C detection and cluster of $\mu$Cs detection by randomly dividing the 115 INBreast cases into five cross-validation folds with 60%
In this paper we propose a new C detection pipeline that introduces a step that effectively filters out individual false positive C detections using shape and appearance features in a cascade of boosting classifiers. We empirically show that our method displays a significantly more effective detection of the individual as well as clusters of Cs.

5. DISCUSSION AND CONCLUSIONS

The results from the Fig. 3 show that our approach is significantly more effective at the detection of individual Cs compared to the baseline [6]. It is also interesting to note from Fig. 4 and Fig. 5 that our approach is competitive with the baseline in terms of cluster detection and case-based performance (note that the results in those figures agree with the published results by Bria et al. [6], even though we use a different dataset). This apparent discrepancy in results is explained by the large number of individual false positive Cs detections that are preserved within true positive clusters of Cs detections by the baseline approach [6]. Our method is able to eliminate a significant number of these false positives and thus provide a more reliable result on which to perform further assessment of the mammogram. Finally, Fig. 6 shows that our proposed methodology is robust to normal mammograms, while Fig. 6b-d displays visually accurate detection of the individual as well as clusters of Cs.
6. REFERENCES

[1] Ahmedin Jemal, Rebecca Siegel, Elizabeth Ward, Yongping Hao, Jiaquan Xu, Taylor Murray, and Michael J Thun, “Cancer statistics, 2008,” *CA: a cancer journal for clinicians*, vol. 58, no. 2, pp. 71–96, 2008.

[2] Edward A Sickles, “Breast cancer screening outcomes in women ages 40-49: clinical experience with service screening using modern mammography,” *Journal of the National Cancer Institute. Monographs*, no. 22, pp. 99–104, 1996.

[3] Ulrich Bick, “Mammography: How to interpret microcalcifications,” in *Diseases of the Abdomen and Pelvis 2014–2017*, pp. 313–318. Springer, 2014.

[4] Maryellen L Giger, Nico Karssemeijer, and Julia A Schnabel, “Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer,” *Annual review of biomedical engineering*, vol. 15, pp. 327–357, 2013.

[5] Inês C Moreira, Igor Amaral, Inês Domingues, António Cardoso, Maria João Cardoso, and Jaime S Cardoso, “Inbreast: toward a full-field digital mammographic database,” *Academic radiology*, vol. 19, no. 2, pp. 236–248, 2012.

[6] Alessandro Bria, Nico Karssemeijer, and Francesco Tortorella, “Learning from unbalanced data: a cascade-based approach for detecting clustered microcalcifications,” *Medical image analysis*, vol. 18, no. 2, pp. 241–252, 2014.

[7] Seung Yeon Shin, Soochahn Lee, Il Dong Yun, Ho Yub Jung, Yong Seok Heo, Sun Mi Kim, and Kyoung Mu Lee, “A novel cascade classifier for automatic microcalcification detection,” *PloS one*, vol. 10, no. 12, pp. e0143725, 2015.

[8] Athanasios Papadopoulos, Dimitrios I Fotiadis, and Lena Costaridou, “Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques,” *Computers in biology and medicine*, vol. 38, no. 10, pp. 1045–1055, 2008.

[9] Issam El-Naqa, Yongyi Yang, Miles N Wernick, Nikolas P Galatsanos, and Robert M Nishikawa, “A support vector machine approach for detection of microcalcifications,” *IEEE transactions on medical imaging*, vol. 21, no. 12, pp. 1552–1563, 2002.

[10] Liyang Wei, Yongyi Yang, Robert M Nishikawa, Miles N Wernick, and Alexandra Edwards, “Relevance vector machine for automatic detection of clustered microcalcifications,” *IEEE transactions on medical imaging*, vol. 24, no. 10, pp. 1278–1285, 2005.

[11] Arnau Oliver, Albert Torrent, Xavier Lladó, Meritxell Tortajada, Lidia Tortajada, Melcior Sentís, Jordi Freixenet, and Reyer Zwiggelaar, “Automatic microcalcification and cluster detection for digital and digitised mammograms,” *Knowledge-Based Systems*, vol. 28, pp. 68–75, 2012.

[12] Chris Seiffert, Taghi M Khoshgoftaar, Jason Van Hulse, and Amri Napolitano, “Rusboost: improving classification performance when training data is skewed,” in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.

[13] C Varela, S Timp, and N Karssemeijer, “Use of border information in the classification of mammographic masses,” *Physics in Medicine and Biology*, vol. 51, no. 2, pp. 425, 2006.

[14] Paul Viola and Michael Jones, “Rapid object detection using a boosted cascade of simple features,” in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. IEEE, 2001, vol. 1, pp. I–511.

[15] Neeraj Dhungel, Gustavo Carneiro, and Andrew P Bradley, “Automated mass detection in mammograms using cascaded deep learning and random forests,” in *Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*. IEEE, 2015, pp. 1–8.

[16] Timo Ojala, Matti Pietikainen, and David Harwood, “Performance evaluation of texture measures with classification based on kullback discrimination of distributions,” in *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision &amp; Image Processing., Proceedings of the 12th IAPR International Conference on*. IEEE, 1994, vol. 1, pp. 582–585.

[17] Hao Jing, Yongyi Yang, and Robert M Nishikawa, “Detection of clustered microcalcifications using spatial point process modeling,” *Physics in medicine and biology*, vol. 56, no. 1, pp. 1, 2010.