Expert Systems with Applications 40 (2013) 7151–7158

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Image analysis of histological features in molar pregnancies

Patison Palee [a], Bernadette Sharp [b,*], Len Noriega [b], N.J. Sebire [c], Craig Platt [d]

[a] Faculty of Computing, Engineering and Technology, Staffordshire University, Beaconside, Stafford ST18 0AD, UK
[b] School of Computing, Staffordshire University, Beaconside, Stafford ST18 0AD, UK
[c] GOSHCC Professor of Paediatric and Developmental Pathology, Great Ormond Street Hospital/Institute of Child Health (ICH/UCL), GOSH BRC Theme Lead (Diagnostics and Imaging), UK
[d] Department of Cellular Pathology, Level 9, Bristol Royal Infirmary, Marlborough Street, Bristol BS2 8HW, UK

ARTICLE INFO

Keywords:
Fuzzy c-mean clustering
Anomaly detection
Image analysis of molar pregnancy stained slides

ABSTRACT

Molar pregnancy (also known as hydatidiform mole, hydatid mole, gestational trophoblastic disease) represents forms of abnormal conception caused by defective fertilisation resulting in excess expression of paternal genes in placental tissue. There are two forms of hydatidiform mole: complete (diploid androgenetic) and partial (paternal triploid), the distinction between which is important for determining appropriate prognosis and management of patients. Both complete and partial hydatidiform moles are associated with increased risk of development of malignant gestational trophoblastic tumours, the risk being much greater for complete hydatidiform moles. Whilst in most cases the diagnosis of these moles can be reliably achieved on morphological histological assessment, these represent a continuing diagnostic problem for histopathologists since in early pregnancy complete hydatidiform moles, partial hydatidiform moles and non-molar hydropic miscarriages may be difficult to distinguish.

In this paper, we propose a computational image analysis approach guided by the knowledge of expert pathologists in identifying essential distinguishing morphological criteria. The approach, which combines Fuzzy C-Means clustering with hue, saturation and value colour space, shows promising results as it is able to classify successfully the villi into appropriate regions, namely trophoblast and stroma, and extract areas of blood. However, because of the marked variations in size, shape and outline of the villi, and trophoblast proliferation, both within and between cases, the analysis shows that there is no single criteria which can reliably classify these products of conception and a combination of criteria is required.

1. Introduction

Hydatidiform mole, also known as molar pregnancy or gestational trophoblastic disease, results from the over-expression of paternal genes and the proliferation of tissue which should develop into the placenta during pregnancy; instead these tissues develop abnormally into an abnormal placental mass. Hydatidiform mole is common in women under age 15 or over 40 and affects around one in 1,000 pregnancies in the UK (Hancock, Newlands, Berkowitz, & Cole, 2003); a higher rate of hydatidiform mole is reported in some countries in Asia and Africa (Gul, Yilmaztürk, & Erden, 1997; Khaskheli, Khushk, Baloch, & Shah, 2007).

Hydatidiform mole can be categorised into complete (CHM) and partial (PHM) mole according to their genetic origin and morphological features. CHM are diploid androgenetic and lack normal fetal blood vessels, the villi have an abnormal budding architecture and show trophoblast proliferation., whereas PHM are paternal triploid, have some normal villi mixed with abnormally shaped, irregular villi and only focal abnormal trophoblastic proliferation (Sebire, 2010). The morphological characteristics of CHM and PHM are different from normal placental villi.

Tissues, obtained via spontaneous or surgical uterine evacuation are routinely examined by histopathologists to diagnosis the disease. The tissues are processed, embedded and then sectioned into thin 2–5 micrometer sections and then stained, usually with haematoxylin and eosin, to enhance the contrast of the nucleus and cytoplasm within cells which allows for microscopic examination. The distinction between CHM and PHM is important for determining the appropriate counselling and treatment of patients since it may be associated with persistent gestational trophoblastic disease including invasive mole, choriocarcinoma and placental site trophoblastic tumours. Sebire, Fisher, and Rees (2003), Sumithran, Cheah, Susil, and Looi (1996) and Howat et al. (1993) explain that the diagnosis of these moles continues to be a problem for many and experienced histopathologists because in early pregnancy CHM and PHM may be difficult to morphologically distinguish from other abnormal pregnancy products. Further studies by Landolsi et al. (2009) and

* Corresponding author. Tel.: +44 1785353585.
E-mail addresses: P.Palee@staffs.ac.uk (P. Palee), b.sharp@staffs.ac.uk (B. Sharp), l.a.noriega@staffs.ac.uk (L. Noriega), neil.sebire@gosh.nhs.uk (N.J. Sebire), Craig.Platt@UHBristol.nhs.uk (C. Platt).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.06.034

Kim, Park, Hong, Kwon, and Robboy (2009) confirm the challenges of histopathological diagnosis of molar pregnancy and conclude that even experienced pathologists cannot always easily distinguish CHM, PHM and non-molar hydropic abortion (HA) at early gestational ages. There are several critical areas that can lead to diagnostic error, namely the diagnosis of early complete mole as partial mole, the over-diagnosis of hydatidiform mole in tubal pregnancy and the diagnosis of placental site non-villus trophoblast as placental site trophoblastic tumour or choriocarcinoma (Wells, 2007). Paul et al. (2010) explain that morphological analysis is inadequate to make a confident diagnoses in many cases and that the histological features of complete mole at an early gestation are frequently confused with partial mole, hydropic miscarriage or non-molar chromosomal abnormalities. Finally the limited sampling material available to histopathologists adds to the diagnostic challenge.

In a minority (15% of CHM and 0.5% of PHM), HM can develop into a persistent disease such as choriocarcinoma, a malignant form of gestational trophoblastic disease hence it is of critical importance to identify and distinguish hydatidiform moles from non-molar specimens and the development of new methods that help differentiate these diagnoses in doubtful cases could be important. The aim of this research project is therefore to develop computational methods to support histopathological evaluation for classification of HM. In this paper we describe our computational approach to the analysis of HM images combining expert heuristics and their tacit knowledge with image processing and artificial intelligence techniques.

The paper introduces the challenges associated with the diagnosis of HM. Section 2 reviews key approaches developed recently in the study of image analysis of various type of cancer. Section 3 explains our approach to analyse HM images in order to detect differences which could help distinguish CHM and PHM from non-molar villi. Sections 4 and 5 describe our experiments and our findings.

2. Related work

A review of the literature on the study of computational image analysis of cancer indicates a high level of activity into the study of breast, lung, skin, cervical and prostate cancers but research into its image analysis of HM appears unexplored. Most techniques applied to image processing and segmentation include statistical methods as well as neural networks, fuzzy logic and cluster analysis. A detailed review of these methods and application is found in Beutel, Kundel, and Metter (2000) and Bankman (2000). In medical image processing these conventional image processing techniques are not sufficient to distinguish between tissues due to poor resolution or contrast, high data variability and noise. Several approaches have been developed ranging from applying semantic indexing of histopathology images for breast cancer grading (Tutac et al., 2008), developing a mammographic ontology to describe features of abnormalities in order to improve the performance of CAD systems (Qi, Denton, & Zwiggelaar, 2006), and combining conventional image processing techniques with artificial intelligence techniques to improve medical image segmentation (Chen, Chung, Chen, Tsai, & Chang, 1998; Dhawan & Arata, 1991; Dhawan, Wolf, Rabinovitz, & Poulos, 1990; Stansfield, 1986).

The traditional approach to cancer image analysis consists of the three following major steps.

(i) **Image pre-processing.** The purpose of this step is to remove unwanted objects (such as noise) and improve the quality of an image, using thresholding to separate the bright regions from the dark regions (Crisan, Dobrescu, & Planinsic, 2007), or to select suspicious regions to distinguish between the breast tissue and background in mammographic images (Lee et al., 2008; Yuan & Shi, 2004). Other statistical methods include a Gaussian filter applied to colour images for smoothing Fine Needle Aspiration Cytology (FNAC) images (Niwas, Palanisamy, & Sujathan, 2010a), Discrete Wavelet Transforms (DWT) to eliminate low frequency image components in a digital mammogram (Lahmiri & Boukadoum, 2011), background masking based on measuring entropy to separate the background from cells (Kazmar, Smid, Fuchs, Luber, & Mattes, 2010) and median filtering to remove all irrelevant data in images for better classification (Bai & Qian, 2008). To improve image intensity distribution histogram equalisation is explored by Raman, Sumari, and Raj (2010), Naghdy, Ross, Todd, and Norachmawati (2010), and Han, Lee, and Choi (2007). Other image enhancement methods include using DWT on digital mammogram images (Hamdi, Auhmani, & Hassani, 2008), gradient, mean contrast, discrete DWT and Gaussian filters (Tiu, Jong, & Hsieh, 2008), anisotropic diffusion to enhance contrast of CT and microscopy (Cui et al., 2010; Linguraru et al., 2009), and a set of morphological operators on a grayscale cyto-image of cervical cancer (Allwin, Kenny, & Manian, 2010).

(ii) **Image segmentation** is defined as a partitioning of an image into meaningful regions for specific tasks. Some researchers adopt either a region-based approach based on either statistical approaches or on machine learning algorithms, or boundary-based approaches. Simple region-based segmentation thresholding is widely used to segment Region of Interest (ROI) in medical applications, such as brain cancer regions from MRI images (Naghdy et al., 2010), micro-calcifications from digital mammograms (Hamdi et al., 2008) and pathological images of prostate cancer (Xu, Xing, Huang, & Wang, 2008). The machine learning approach employs Hopfield Neural Networks (HNN) and a Fuzzy C-Mean (FCM) clustering algorithm to identify lung cancer on sputum colour images (Taher & Sammouda, 2011), the K-means algorithm to separate the nuclei and cytoplasm from the background (Naghdy et al., 2010), the K-means algorithm with Otsu's algorithm to separate the abnormal nuclei regions from all nuclei regions in cancer microscopic images (Cui et al., 2010). The boundary-based approach is the method based on finding out the border of the objects. For example, Crisan et al. (2007) apply the automatic contour trace algorithm to digital mammograms to find out the boundary of lesions from the images, whereas Naik, Madabhushi, Tomaszeweski, and Feldman (2007) combine the contour-based approach with a level set algorithm initialised by a user to enhance the performance of prostate tissue image segmentation. Doukas & Maglogiannis, 2007 apply active contour techniques (e.g. snake) to microscopic images for cell death (apoptosis) segmentation, and Parolin, Herzer, and Jung (2010) use an edge map that utilises DWT to guide the gradient vector flow (GVF) snakes to segment the lesion in dermatological images.

(iii) **Feature extraction and classification**. The feature extraction focuses on extracting the relevant features that can be used to classify the objects. Two major classification approaches are employed. The statistical approach includes Bayesian decision rules (Waheed, Moffitt, Chaudry, Young, & Wang, 2007), *k*-nearest neighbour classifiers (Niwas, Palanisamy, & Sujathan, 2010b) and *k*-means classification (Qi & Head, 2001; Zhang, Shi, Gao, & Li, 2007); the main challenge is the assumption that the samples or features are independent. However, for some applications, the samples or features depend on each other (Demir & Yener, 2005), so the machine learning approach is adopted; a wide range of artificial intelligence techniques are applied namely Naive