

Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Heuristic neural network approach in histological sections detection of hydatidiform mole

Patison Palee
Bernadette Sharp
Leonard Noriega
Neil Sebire
Craig Platt

SPIE.

Patison Palee, Bernadette Sharp, Leonard Noriega, Neil Sebire, Craig Platt, "Heuristic neural network approach in histological sections detection of hydatidiform mole," *J. Med. Imag.* **6**(4), 044501 (2019), doi: 10.1117/1.JMI.6.4.044501.

Heuristic neural network approach in histological sections detection of hydatidiform mole

Patison Palee,^a Bernadette Sharp,^b Leonard Noriega,^{c,*} Neil Sebire,^d and Craig Platt^e

^aChiang Mai University, Chiang Mai, Thailand

^bStaffordshire University, Department of Computer Science, College Lane, Stoke on Trent, Staffordshire, England

^cHochschule Furtwangen University, Fakultät von Wirtschaftsinformatik, Furtwangen, Germany

^dUniversity College London, Great Ormond Street Institute of Child Health, Professor of Paediatric and Developmental Pathology, Population, Policy and Practice Department, London, United Kingdom

^eConsultant Perinatal Pathologist, University Hospitals Bristol NHS Foundation Trust, Bristol, United Kingdom

Abstract. A heuristic-based, multineural network (MNN) image analysis as a solution to the problematical diagnosis of hydatidiform mole (HM) is presented. HM presents as tumors in placental cell structures, many of which exhibit premalignant phenotypes (choriocarcinoma and other conditions). HM is commonly found in women under age 17 or over 35 and can be partial HM or complete HM. Appropriate treatment is determined by correct categorization into PHM or CHM, a difficult task even for expert pathologists. Image analysis combined with pattern recognition techniques has been applied to the problem, based on 15 or 17 image features. The use of limited data for training and validation set was optimized using a k -fold validation technique allowing performance measurement of different MNN configurations. The MNN technique performed better than human experts at the categorization for both the 15- and 17-feature data, promising greater diagnostic consistency, and further improvements with the availability of larger datasets. © 2019 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.6.4.044501]

Keywords: multineural network; image analysis; hydatidiform mole; molar pregnancy; diagnosis.

Paper 19066R received Mar. 13, 2019; accepted for publication Oct. 7, 2019; published online Nov. 5, 2019.

1 Introduction

Molar pregnancy occurs with the formation of hydatidiform moles (HM) in their partial (PHM) or complete (CHM) form on the placenta and are abnormal products of conception. With CHM, there is no fetal development, and with PHM, there is only partial development, resulting in malformations and an unviable fetus. Therefore, the distinction between PHM and CHM is important in determining the management of the condition (with CHM having a greater malignant potential). Although moles can be identified from placental cells using clinical, ultrasonographic, gross morphological, histological and genetic criteria, and given each type has a different prognosis, the final diagnosis must be confirmed by pathologists.^{1,2} Many experienced pathologists have confirmed the difficulties in distinguishing these moles.³ Typically, misclassification is due to the absence of unambiguous morphological criteria to differentiate HM (PHM or CHM) and normal placenta, and some atypical cases are not easy to detect using morphological criteria alone. Misclassifications can also occur if no recent tissue samples are available and also if abundant tissue of maternal origin is present.⁴ Figure 1 shows an example of nonmolar villi, PHM, and CHM. Distinguishing these moles on the basis of morphological features alone is both challenging and is affected by interobserver variability.³

The identification of cell morphology information can be extracted from digital images of the microscope slides used by pathologists in the diagnosis of HM. In this paper, expert pathologist knowledge has been modeled using morphological image processing techniques to produce an artificially

intelligent system to classify digital images of placental cells into normal, PHM, or CHM categories, with the aim of producing a system either capable of making reliable diagnoses or assisting pathologists in diagnosis. Rendering the diagnostic problem as a pattern recognition task uses a standard methodology involving three stages: image preprocessing, which removes irrelevant information and enhances relevant information; image segmentation which partitions the image into regions of interest (relevant cell structures); the extraction and identification of relevant features (numerical measurements); and finally, classification. This paper focuses on the classification stage, as the earlier steps have been reported by Palee et al.⁵

The remainder of the paper is arranged as follows. Section 2 reviews the recent study of various types of cancer image analysis. The proposed heuristic-based multineural network (MNN) to classify anomalous villi of normal, PHM, and CHM images is explained in Sec. 3. Section 4 describe the experiments, which are followed by a discussion and conclusion.

2 Related Work

Pattern classification is the process of categorizing or classifying the objects of interest, and the inherent difficulties involved depend on the properties of the data used to make the classification decisions. Statistically, the data should be “well-behaved” (minimum dimensionality, minimum correlation, and rich in statistical information content). The approaches to categorization are all statistical in nature, and some are based on machine learning (ML) approaches, where properties of the data are “learned” from a “training set” and used to make the categorization decisions. The traditional statistical classification approaches include Bayesian decision rule,⁶ k -nearest

*Address all correspondence to Leonard Noriega, E-mail: l.a.noriega@uclmail.net

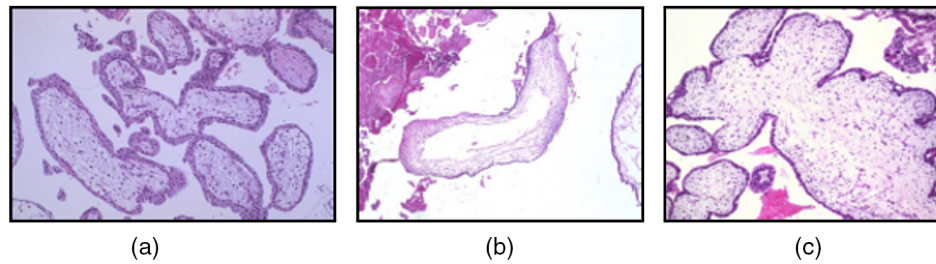


Fig. 1 Hydatidiform moles: (a) nonmolar villi, (b) PHM villi, and (c) CHM villi.

neighbor (kNN),⁷ k-means,^{6,7} and Gaussian mixture model⁶ based systems. ML uses fuzzy set,⁸ artificial neural network (ANN),^{7,9} support vector machine (SVM),⁹ case-based reasoning,¹⁰ and extreme learning machine (ELM)¹¹ algorithms.

2.1 Traditional Statistical Classification

This approach is followed widely in cancer diagnosis using many image types. Waheed et al. used multiclass Bayesian decision rules to differentiate between benign and malignant renal cell carcinoma on pathological images.¹² Qi and Head applied the k-means classifier to identify breast cancer regions from thermal infrared images,¹³ and Hamdi et al. applied used Fisher's linear discriminant to categorize the normal and abnormal microcalcifications on digital mammograms.¹⁴ In addition, the quadratic discriminant analysis method was used to classify the features of colon cancer in histological images, resulting in better classification than those based on techniques using Fisher's score.¹⁵ Meng et al. applied a classifier based on similarity measurements (city block metric) to compute similarity between data values for ovarian cancer samples.¹⁶ This method was selected because the data values extracted from samples are binary values. Zheng et al. developed a breast cancer classification application by combining k-means classification with a multilayer perceptron (MLP) neural network using the backpropagation algorithm to train and classify a dataset of benign and malignant regions in ultrasound images.¹⁷ A kNN classifier was applied by Niwas et al.¹⁸ to classify cancerous and noncancerous cells on cytological images and by Petroudi et al.¹⁹ for breast cancer classification. Li et al. used a classification method based on Gaussian mixture models to classify breast tumor mammographic images.²⁰

The requirement for the statistical features to be well-behaved is often an assumption in much of the statistical work, but many of the diagnostic features often correlate (i.e., they are not statistically independent).²¹ To find the best classifier for breast cancer diagnosis, Filipczuk et al. compared the classification performance of statistical based kNN classifier with ML-based classifiers: naive Bayes (NB), decision tree, and SVMs showed that these performed better than the kNN approach.²²

Other classifiers include the use of NB to classify cervical cancer cells from colposcopy images,²³ case-based reasoning to classify breast cancer cells in digital mammograms,²⁴ ELM algorithm to classify the malignant masses in digital mammograms,²⁵ and particle swarm optimization to classify anomalous areas in digital mammograms of breast cancer²⁶ and analyze skin cancer diagnosis based on dermoscopic images for feature optimization.²⁷ Yaguchi et al. apply linear discriminant analysis (LDA) to classify biopsy images of stomach cancer²⁸ while Torrent et al. proposed the GentleBoost algorithm based on the concept that the sum of weak classifiers can deliver a

strong classifier to improve the robustness of their system.²⁹ Wei et al. compared the performance of classifiers, such as SVM, kernel Fisher discriminant (KFD), relevance vector machine (RVM), and committee machines based on the adaptive boosting algorithm.³⁰ They showed that SVM, KFD, and RVM yielded the best performance for microcalcification classification in digital mammograms. George et al. compared four ML-based classifiers, namely, MLP, probabilistic neural network (PNN), learning vector quantization, and SVM for a breast cancer cytological image classification application, and showed that the two best classifiers of this image type are PNN and SVM.³¹ Shahare and Giri³² have recently carried out a comparative study on SVM with different kernels for breast cancer detection and reported that SVM outperformed ANN.

2.2 Machine Learning Classification Approaches

ML classification uses a training set to learn how to classify data. The training set associates a class with each element in the set so that the performance of an ML classification can be assessed during the learning process. An ML system is said to have learned when its errors in classifying the training set reach a minimum (ideally zero). Having learned the ML system is then expected to "generalizes" or "recall," correctly classify data not included in the training set.²¹ The SVM is a common supervised ML approach applied to cancer applications. It was used by Lahmiri and Boukadoum,³³ and Hassanien and Kim³⁴ to classify cancerous breast cells, by Allwin et al.³⁵ to study cervical cancer and by Han et al.³⁶ and Naik et al.³⁷ in prostate cancer. Xu et al.³⁸ and Doyle et al.^{39,40} have also applied SVM to classify histology image slides while Yuan et al.⁴¹ have used SVM based on texture only to detect skin cancer with acceptable results, though they recommend adding more features to improve the classification. Deepa and Bharathi⁴² and Kim et al.⁴³ use SVM to distinguish between benign and malignant regions in digital mammogram images, whereas Torheim et al.⁴⁴ applied SVMs to classify MR images of cervical cancer.

To solve the unclassifiable region problems in SVM and to improve its classification performance, Bai and Qian⁴⁵ proposed fuzzy SVMs to classify the different noise levels of digital mammograms of breast cancer. The experiments yielded better results and decreased the computational cost. Furthermore, Alolfe et al.⁴⁶ used four classifiers (SVM, kNN, neural network, and fuzzy classifiers) with microcalcification data from digital mammography. The classification results of the four classifiers were unsatisfactory due to the small number of benign and malignant samples used in training and testing. To improve the classification, Alolfe et al.⁴⁷ combined SVM with LDA. Kounelakis et al.⁴⁸ however, showed that the application of SVM, embedded with the Relief-F filter method under a recursive feature elimination manner, provided greater accuracy

compared to Relief-F on Weka software in classifying statistical and biological features of brain gliomas. To extract significant statistical and morphological features from pulmonary nodules in computed tomography images of the lung and to determine whether the nodule is cancerous or healthy. Farahani et al.⁴⁹ employed an ensemble of three classifiers comprising MLP, kNN, and SVM.

ANNs are much used in cancer applications. Niwas et al.⁵⁰ classified images of breast tissue samples with them, and Karnan and Gandhi⁵¹ used them to study digital mammograms. Moradi et al.⁵² have applied MLP neural networks based on the back-propagation algorithm to classify cancerous and noncancerous regions, while Mini⁵³ applied PNNs to classify breast cancer in digital mammograms. To improve the performance of the classification results of ANNs, Neofytou et al.⁵⁴ combined PNN with SVM while Zheng et al.¹⁷ applied a MLP network and the k-means classification method to classify ultrasound images of tumors. Naghdy et al.⁵⁵ also applied ANN and neurofuzzy classifiers to a real-time brain cancer classification application. Tiu et al.⁵⁶ combined a self-organizing map neural network with a fuzzy criterion classifier to classify microcalcification on digital mammograms, whereas Elizabeth et al.⁵⁷ applied a radial basis function neural network (RBFNN) to classify lung cancer tomography images. The experimental results show that the proposed method yields 94.44% accuracy. Wan et al.⁵⁸ developed a multiple level feature set, including pixel, object, and semantic level features derived from convolutional neural networks (CNNs) to distinguish low, intermediate, and high grades of breast cancer in digitized histopathology images and achieved an overall accuracy of 69%. Recently, Anaraki et al.⁵⁹ combined CNNs and genetic algorithm to classify different grades of Glioma using magnetic resonance imaging. Sajjad et al.⁶⁰ developed a deep learning framework to segment and classify brain tumor into four different grades using a fine-tuned CNN model.

A decision tree classifier has also been used by Peng et al.⁶¹ to classify cervical nuclei on microscope images. They claim that their proposed method yields a promising accuracy (97.8%). To improve the classification of cervical nuclei on microscope images, Peng et al.⁶² used C4.5 decision tree as a classifier and *F*-score as a feature selection algorithm to minimize the computational cost.

Although the literature review reveals a high level of research activity, most applications tend to be limited to the study of breast, lung, skin, cervical and prostate cancers. The study of HMs or molar pregnancy cancer is confined to the management and care of patients (Gul et al.,⁶³ Khaskheli et al.,⁶⁴ and Sebire⁶⁵). The literature review also shows that knowledge and expert heuristics are still unexplored, yet such expertise has the potential of improving the accuracy in classifying cancerous cells.

3 Integrated Heuristic Approach to Analyze and Classify Molar Pregnancy Slides

Our research project has developed an MNN configuration, which integrates experts' heuristic-based approaches to classify images taken from stained slides of HM. The aim of this MNN configuration is to capture expert knowledge used in identifying the most critical features for distinguishing nonmolar placenta from PHM and CHM and PHM from CHM. In the development of this approach, there have been challenges. Many expert strategies are based on expert "tacit knowledge" and are difficult to communicate. Also, experts differ in their approaches so it becomes difficult to reach a common consensus. The percentage

of correct classification of villi on the basis of morphology can vary from 55% to 75% for an individual pathologist and may reach up to 75% if the consensus view of several pathologists is used (Vang et al.³). To simulate expert pathologists' heuristics, we have adopted a task analysis approach and divided each task into subtasks, which are further explored as a set of experiments, in order to manage complexity and rank the information-bearing features. The pathologists' approach to the detection of anomalies begins by first separating normal villi from non-normal villi, then focusing on the non-normal villi and separate PHM from CHM. This approach is reflected in our MNN configuration.

3.1 Feature Selection

The initial test data for the experiment consisted of 939 images of villi are divided into 294 normal, 455 PHM, and 190 CHM, in which each image contained intact and complete villi with clear contours. This enabled the identification of features within the image, which consisted of numerical descriptors of specific morphological anomalies associated with the trophoblast, stroma, and red blood cells of PHM and CHM (Fig. 2). Numerical descriptors of other important characteristics such as the shape and the perimeter of the villi, the villous scalloping aspect, trophoblastic hyperplasia, and the ratio between the stroma and villi were also used. In total, 15 numerical features were extracted and applied to the MNN configuration. A summary of these features and associated anomalies is given in Table 1.

3.2 Experimental Study Based on 15 Features

In the first experiment, the MNN (consisting of NN1, NN2, and NN3) is configured so that it closely models the process used by pathologists in determining the diagnosis discussed above, and illustrated in Fig. 3. NN1 performs the initial categorization into two groups: "possible-normal" and "abnormal." Abnormal villi are trained by NN2 to distinguish between PHM and CHM. Possible-normal villi are further trained by NN3 to classify normal villi from PHM. The primary settings of NN1, NN2, and NN3 consist of 15 input nodes representing the 15 features, 1 hidden layer, a set of hidden nodes varying from 1 to 15, and 2 output nodes.

The neural network training method used here is the back-propagation multilayer perceptron (MLP) model,⁷ which can be seen as a "matrix-fitting" approach to solving the problem. The initial matrices consist of random weights, whose values are changed and over the training process so as to minimize an error function. The error function is a measure of the difference

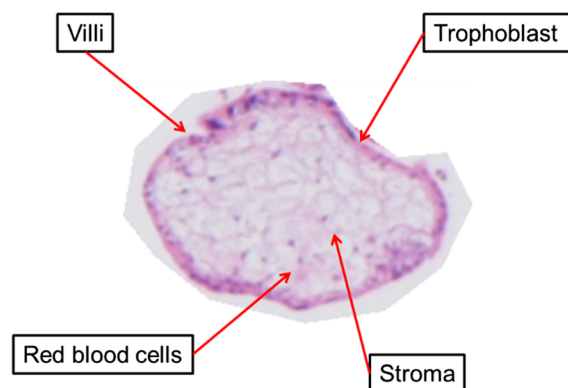


Fig. 2 Villous features.

Table 1 Elicited morphological features and associated anomalies.

| Morphological features given by expert pathologists | Proposed features | Types of anomalies |
|---|--|-----------------------|
| RBC factor: density and distribution within a given villi | Percentage of red blood cells inside stroma regions | Point anomaly |
| Villi: shape and size | Villi size | Morphological anomaly |
| | Number of villi boundary corner points | Morphological anomaly |
| | Different area between villi's bounding box and villi area | Morphological anomaly |
| | Major axis | Morphological anomaly |
| | Minor axis | Morphological anomaly |
| | Elongation ratio | Morphological anomaly |
| | Ratio between number of villi boundary corner points and all pixels belonging to villi perimeter | Morphological anomaly |
| | The notion of four quadrants | Morphological anomaly |
| Stroma's morphological characteristics | Percentage of stroma regions inside villi | Contextual anomaly |
| Stroma's textural characteristics | Percentage of edge inside stroma regions | Contextual anomaly |
| | Variance of gray scale of stroma regions | Point anomaly |
| Trophoblast thickness and proliferation | Percentage of trophoblast inside villi | Contextual anomaly |
| | Trophoblast skeleton per trophoblast perimeter ratio | Density anomaly |
| | Trophoblast analysis | Morphological anomaly |

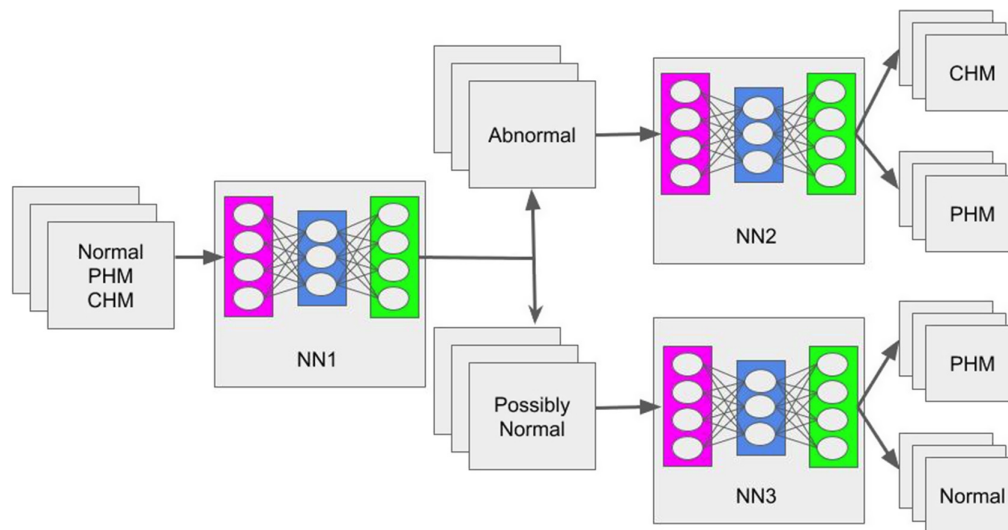


Fig. 3 MNN architecture.

between the answer produced by the network, and the answers given in the training set expressed as a function of the network weights. The technique is based on gradient descent (similar to Newton's method) and, since its discovery in 1985, has been proven effective in problems, where there is no clear solution (such as the diagnosis of molar pregnancy). The error function to be minimized is assumed to have a global minimum—a setting of the matrix values in which the error cannot go any

lower, however, the behavior of the error function is entirely unpredictable and so there is a high likelihood (almost a certainty in most difficult problems) of local minima existing, which can confound the training process (as the gradient at these points is zero), implying no change is possible. The effect of local minima is managed by setting parameters within the network and regularly randomizing the order in which training data is presented.

In this experiment, a k -fold cross-validation⁶⁶ was applied to provide data for results evaluation. The precise amount of data required to train an MLP is dependent on the number of weights in the network. As the neural network technique is innately statistical, the number of training samples should be a substantial multiple of the number of weights,⁹ however, there is no agreement as to what this multiple should be. The use of the k -fold cross-validation technique is therefore designed to function in a manner similar (but not identical) to bootstrapping in situations where data availability is limited to make the best use of the resource. In any supervised ML algorithm, such as a neural network, there is a requirement that the training data and the test data conform to the same distribution as the original data. The available training data is divided into k equal sized segments, where membership of the segments is taken at random from the available data. The algorithm for training and validating the data is as follows (illustrated in Fig. 4).

1. Divide the available data into K segments, labeled $i = 1, \dots, K$.
2. Train the neural network K times, as follows.
3. Remove one data segment to serve as a validation set.

4. Train the network with the remaining $K - 1$ segments.
5. Record the performance statistics, and select the next validation segment. Repeat from 2.

This technique serves to validate the number of hidden nodes in the neural network, as K repetitions of the training enable the determination of the consistency of the performance over the K iterations, as neural networks are known to be sensitive to starting conditions and local minima. In cases where sufficient data exist so that 10% of samples are likely to provide a similar distribution to the training set, K is usually set to 10. In cases where fewer data are available, a smaller number is selected.

This is an established, informed estimation method to determine the consistency and performance of supervised ML algorithms (such as MNNs) operating under different training conditions. The composition of each partition should reflect the statistical distribution of the categories in the whole dataset (i.e., normal, CHM, and PHM).⁶⁶ The repetition of the training and validation process with different datasets yields insights (after analysis), as to the applicability of the technique to the problem, given the recognized sensitivities of neural networks to starting conditions and local minima (see above).

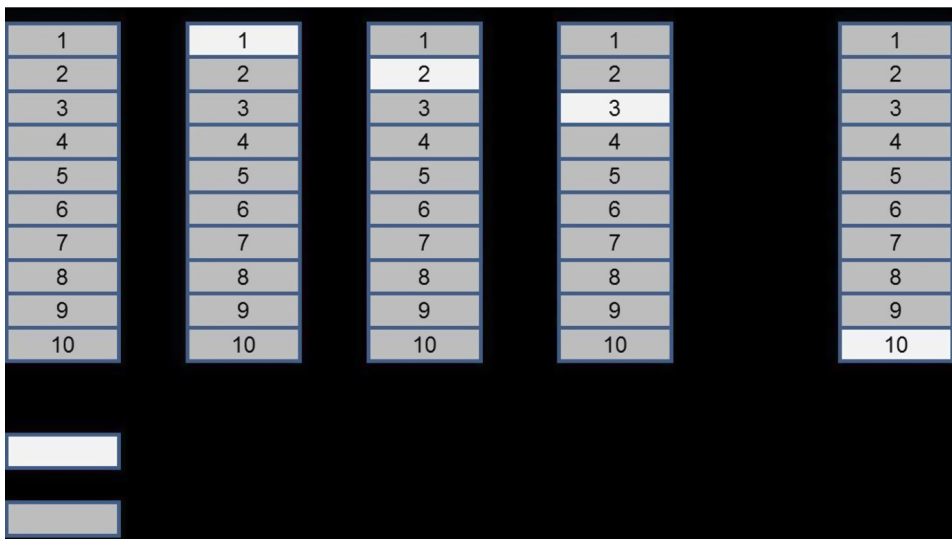


Fig. 4 Ten-fold cross validation.

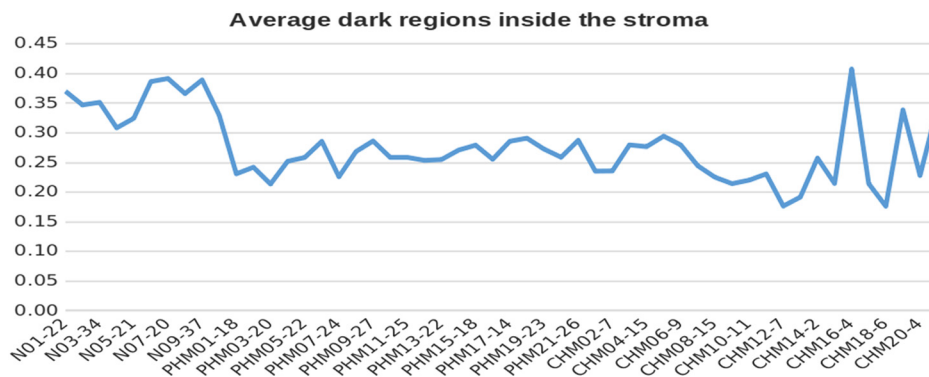


Fig. 5 Dark regions inside the stroma (e.g., N101-22 denotes normal placenta slide 1 consisting of 22 villi).

Table 2 A confusion matrix.⁶⁷

| | | Actual | |
|-----------|----------|----------|----------|
| | | Positive | Negative |
| Predicted | Positive | TP | FP |
| | Negative | FN | TN |

Table 3 Comparative performance of NN with different numbers of hidden nodes with 15 features.

| Number of hidden nodes | Average accuracy | | Standard deviation | |
|------------------------|------------------|----------------|--------------------|----------------|
| | Training set | Validation set | Training set | Validation set |
| 1 | 92.7 | 90.1 | 0.3 | 0.7 |
| 2 | 92.1 | 89.4 | 1.6 | 1.7 |
| 3 | 92.7 | 90.3 | 0.1 | 1.3 |
| 4 | 92.9 | 90.0 | 0.6 | 1.2 |
| 5 | 92.9 | 90.2 | 0.5 | 1.2 |
| 6 | 93.0 | 90.3 | 0.7 | 0.5 |
| 7 | 93.5 | 90.2 | 0.7 | 1.1 |
| 8 | 93.9 | 90.1 | 0.4 | 1.4 |
| 9 | 93.9 | 89.3 | 0.3 | 1.0 |
| 10 | 94.5 | 90.8 | 0.4 | 1.0 |
| 11 | 94.3 | 90.0 | 0.3 | 0.8 |
| 12 | 94.4 | 90.1 | 0.3 | 0.4 |
| 13 | 94.6 | 90.8 | 0.3 | 0.4 |
| 14 | 94.6 | 91.0 | 0.2 | 1.3 |
| 15 | 94.6 | 89.6 | 0.6 | 1.6 |

The results of the *k*-fold cross-validation are then used with a confusion matrix (Table 2) to compare predicted results against actual diagnoses. Additionally, recall and precision determined as a function of true/false and positives/negatives are used and defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP},$$

$$\text{Recall} = \frac{TP}{TP + FN},$$

where TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

This experiment was repeated with several configurations of MNN with results shown in Table 3, with the best results achieved with six hidden nodes and 93.1% average accuracy for NN1, three hidden nodes with 89.0% average accuracy for NN2,

Table 4 Average accuracy of MNN based on 15 features.

| Sampling set | Average accuracy of validation set | | | Overall accuracy |
|--------------------|------------------------------------|---------|---------|------------------|
| | NN1 = 6 | NN2 = 3 | NN3 = 5 | |
| 1 | 93.3 | 88.5 | 92.2 | 80.4 |
| 2 | 93.2 | 89.1 | 91.4 | 81.4 |
| 3 | 93.2 | 89.4 | 90.6 | 80.4 |
| 4 | 93.5 | 89.0 | 90.9 | 81.3 |
| 5 | 93.6 | 88.8 | 91.4 | 80.2 |
| 6 | 92.8 | 89.2 | 91.2 | 81.1 |
| 7 | 92.4 | 88.9 | 91.2 | 80.5 |
| 8 | 93.3 | 88.1 | 92.1 | 83.2 |
| 9 | 93.3 | 89.2 | 90.8 | 82.4 |
| 10 | 92.6 | 89.7 | 92.1 | 81.4 |
| Average | 93.1 | 89.0 | 91.4 | 81.2 |
| Standard deviation | 0.4 | 0.5 | 0.6 | 1.0 |

Table 5 Precision and recall based on 15 features.

| MNN | Precision | | | Recall | | | Average accuracy |
|--------------|-----------|------|------|--------|------|------|------------------|
| | Normal | PHM | CHM | Normal | PHM | CHM | |
| Experiment 1 | 82.7 | 81.3 | 80.8 | 83.0 | 87.1 | 64.3 | 81.2 |

and five hidden nodes with 91.4% average accuracy for NN3 (Table 4).

The precision achieved for normal, PHM, and CHM villi is 82.7%, 81.3%, and 80.8%, respectively, whereas the recall for normal and PHM villi is 83.0% and 87.1%, respectively, which are higher than the CHM villi (Table 5).

3.3 Experimental Study Based on 17 Features

Further consultation with expert diagnosticians led to the identification of two additional characteristics: dark regions within the stroma and the trophoblast. The dark regions inside the stroma are high in normal placental and low in PHM and CHM villi samples (Fig. 5). In nonmolar villi, the structure of the stroma is net-like, whereas, in PHM, it is scalloped and dentate, and in CHM, it is enlarged, irregular, and karyorrhectic. Within the initial HM dataset, these characteristics were observed: with “dark” regions within the trophoblast being more pronounced in normal placentas than in PHM and with great variation among the CHM villi samples (Fig. 6). Numerical features based on image analysis were identified, which represented these morphological features and which were then used in the MNN in order to classify the image data.”

These two additional features are added to the previous 15 features and used to train the MNN. In this experiment, the

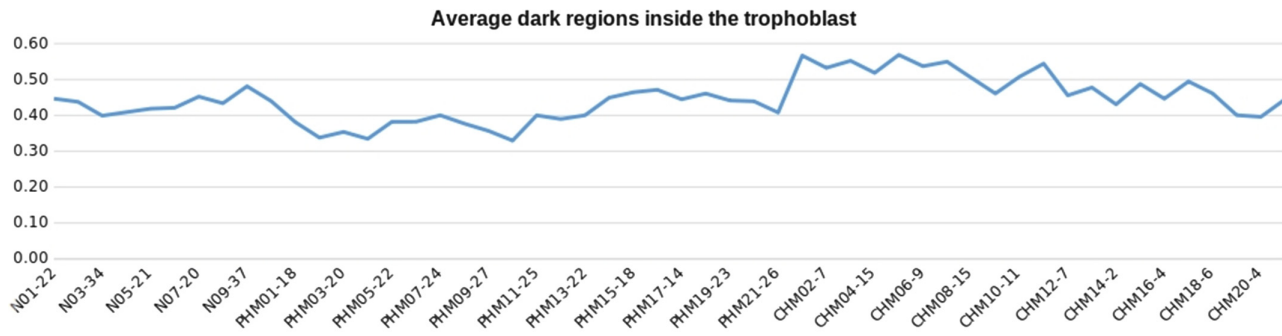


Fig. 6 Dark regions inside the trophoblast.

Table 6 Comparative performance of NN with different numbers of hidden nodes with 17 features.

| Number of hidden nodes | Average accuracy | | Standard deviation | |
|------------------------|------------------|----------------|--------------------|----------------|
| | Training set | Validation set | Training set | Validation set |
| 1 | 70.4 | 69.5 | 1.6 | 1.8 |
| 2 | 85.4 | 82.9 | 2.6 | 2.7 |
| 3 | 87.1 | 84.3 | 1.3 | 1.2 |
| 4 | 87.8 | 84.3 | 0.3 | 0.6 |
| 5 | 88.1 | 85.1 | 0.4 | 0.6 |
| 6 | 88.3 | 84.4 | 0.3 | 0.6 |
| 7 | 88.6 | 84.2 | 0.2 | 0.6 |
| 8 | 88.7 | 84.9 | 0.4 | 0.9 |
| 9 | 88.9 | 85.2 | 0.6 | 0.5 |
| 10 | 89.3 | 85.1 | 0.6 | 0.6 |
| 11 | 89.9 | 84.7 | 0.6 | 0.7 |
| 12 | 89.8 | 85.3 | 0.4 | 0.8 |
| 13 | 89.6 | 84.7 | 0.4 | 0.7 |
| 14 | 90.2 | 85.2 | 0.5 | 0.6 |
| 15 | 90.2 | 85.0 | 0.5 | 0.9 |
| 16 | 90.4 | 85.6 | 0.4 | 0.7 |
| 17 | 90.7 | 85.6 | 0.3 | 0.7 |

number of input nodes is configured to 17 to represent the new features and the number of hidden nodes ranges from 1 to 17 nodes (Table 6). The results of the MNNs show that the best number of hidden nodes for NN1, NN2, and NN3 hidden nodes is 5 with 96.7% average accuracy, 8 with 92.8% average accuracy, and 14 with 94.9% average accuracy, respectively. The repeated 10-fold cross-validation is reapplied and shows that the average accuracy is 86.1% (Table 7).

Furthermore, the precision of the second experiment has increased from 82.7% to 89.5% for the normal villi class, from

Table 7 Average accuracy of MNN based on 17 features.

| Sampling set | Average accuracy of validation set | | | Overall accuracy |
|--------------------|------------------------------------|---------|----------|------------------|
| | NN1 = 5 | NN2 = 8 | NN3 = 14 | |
| 1 | 96.9 | 92.4 | 94.9 | 85.0 |
| 2 | 96.5 | 92.7 | 95.2 | 86.7 |
| 3 | 96.5 | 93.0 | 94.7 | 85.5 |
| 4 | 96.6 | 92.7 | 94.1 | 86.8 |
| 5 | 96.9 | 93.2 | 94.9 | 85.9 |
| 6 | 97.1 | 93.5 | 94.9 | 86.1 |
| 7 | 96.7 | 92.8 | 94.6 | 86.4 |
| 8 | 96.0 | 92.8 | 95.4 | 85.9 |
| 9 | 96.6 | 92.6 | 95.0 | 85.7 |
| 10 | 96.8 | 92.5 | 95.4 | 87.0 |
| Average | 96.7 | 92.8 | 94.9 | 86.1 |
| Standard deviation | 0.3 | 0.3 | 0.4 | 0.6 |

81.3% to 84.7% for PHM villi class, and from 80.8% to 86% for CHM villi class. Similarly, the recall has improved from 83.0% to 87.7% for the normal villi class, from 87.1% to 90.3% for the PHM villi class, and from 64.3% to 73.5% for the CHM villi class (Table 5).

4 Discussion and Concluding Remarks

In terms of accuracy, the performance of the MNN in the task compares favorably with the accuracy reportedly achieved by diagnosticians.³ Figures do not appear to be available for the precision and recall of clinicians in order to make a more detailed comparison. This remains as further work.

The role of the system as a whole is under consideration, so as to determine whether it represents a complete, self-contained diagnostic tool, or as a means of obtaining objective evidence to assist a pathologist in making a diagnosis (i.e., decision support). In either case, the risks of misdiagnosis need to be factored into the system, based on the confusion matrix statistics of false positives/negatives.

The experimental studies have identified 17 critical features, which can be used to distinguish between PHM and CHM. As the difficulties faced by pathologists in achieving correct classification on the basis of morphology vary significantly, our MNN configuration is able to achieve over 80% in terms of precision and recall for both PHM and CHM. The study has also helped elicit tacit critical features employed by expert pathologists. While the experts suggest that the most important features are the size and shape of villi, our experimental studies have shown that texture features are more important. One major difficulty is the availability of well-characterized samples to improve the training of the MNN, which is currently limited to a total of 939 villi, and a larger dataset is expected to improve understanding of the diagnostic problem as well as the issues of training and classification.

Further work will require the acquisition of more cell images, better to characterize the statistical behavior of the data and to explore different architectures of the MNN for better performance. For example, it is proposed to enhance the developed heuristic MNN configuration by integrating fuzzy criteria into the hidden layers to address the issue of the fuzziness of villi characteristics. The current research is based on normal placental, PHM and CHM villi; other types of abnormal villi, such as persistent gestational trophoblastic and choriocarcinoma, are to be investigated to support pathologists in their complex and challenging classification of HM.

Disclosures

The authors are aware of no conflicts of interest.

Ethical Approval

The PhD was subject to ethical scrutiny by the Research Ethics Committee at Staffordshire University. The identities of the patients whose data were used were anonymized, and remained unknown to the principal researchers on the project. There were no issues with the consent of the patients concerned.

References

- P. Vassilakos, G. Riotton, and T. Kajii, "Hydatidiform mole: two entities. A morphologic and cytogenetic study with some clinical consideration," *Am. J. Obstet. Gynecol.* **127**(2), 167–170 (1977).
- P. Petignat et al., "Is genetic analysis useful in the routine management of hydatidiform mole?," *Hum. Reprod.* **18**(2), 243–249 (2003).
- R. Vang et al., "Diagnostic reproducibility of hydatidiform moles: ancillary techniques (p57 immunohistochemistry and molecular genotyping) improve morphologic diagnosis," *Am. J. Surg. Pathol.* **36**(3), 443–453 (2012).
- K. A. Bell et al., "Molecular genetic testing from paraffin-embedded tissue distinguishes nonmolar hydropic abortion from hydatidiform mole," *Mol. Diagn.* **4**(1), 11–19 (1999).
- P. Palee et al., "Image analysis of histological features in molar pregnancies," *Expert Syst. Appl.* **40**(17), 7151–7158 (2013).
- R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., Wiley, New York (Bayesian Decision Rules) (2001).
- R. Schalkov, *Pattern Recognition: Statistical, Structural and Neural Approaches*, Wiley, London (1992).
- M. Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, 2nd ed., Pearson, Edinburgh (2004).
- S. Haykin, *Neural Networks and Learning Machines*, 3rd ed., Pearson, Upper Saddle River, New Jersey (2008).
- G. F. Luger, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, 6th ed., Pearson, London (2009).
- G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing* **70**, 489–501 (2006).
- S. Waheed et al., "Computer aided histopathological classification of cancer subtypes," in *Proc. 7th IEEE Int. Conf. Bioinf. and Bioeng.*, pp. 503–508 (2007).
- H. Qi and J. F. Head, "Asymmetry analysis using automatic segmentation and classification for breast cancer detection in thermograms," in *Proc. 23rd Annu. Int. Conf. IEEE Eng. Med. and Biol. Soc.*, Vol. 3, pp. 2866–2869 (2001).
- N. Hamdi, K. Auhmani, and M. M. Hassani, "Design of a high-accuracy classifier based on fisher discriminant analysis: application to computer-aided diagnosis of microcalcifications," in *Proc. Int. Conf. Comput. Sci. and Appl.*, pp. 267–273 (2008).
- A. N. Esgiar and P. K. Chakravorty, "Fractal based classification of colon cancer tissue images," in *Proc. 9th Int. Symp. Signal Process. and Appl.*, pp. 1–4 (2007).
- H. Meng et al., "Feature extraction and analysis of ovarian cancer proteomic mass spectra," in *Proc. 2nd Int. Conf. Bioinf. and Biomed. Eng.*, pp. 668–671 (2008).
- K. Zheng et al., "Recognition of breast ultrasound images using a hybrid method," in *Proc. IEEE/ICME Int. Conf. Complex Med. Eng.*, pp. 640–643 (2007).
- S. I. Niwas, P. Palanisamy, and K. Sujathan, "Wavelet based feature extraction method for breast cancer cytology images," in *Proc. Ind. Electron. Appl.*, pp. 686–690 (2010).
- S. Petroudi et al., "Investigation of AM-FM methods for mammographic breast density classification," in *Proc. IEEE 13th Int. Conf. Bioinf. and Bioeng.*, pp. 1–4 (2013).
- Z. Li et al., "A new histogram-based breast cancer image classifier using Gaussian mixture model," in *Proc. ACM Res. Appl. Comput. Symp.*, pp. 143–147 (2012).
- C. Demir and B. Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey," Rensselaer Polytechnic Institute, Department of Computer Science (2005).
- P. Filipczuk et al., "Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies," *IEEE Trans. Med. Imaging* **32**(12), 2169–2178 (2013).
- H. G. Acosta-Mesa et al., "Cervical cancer detection using colposcopic images: a temporal approach," in *Proc. Sixth Mex. Int. Conf. Comput. Sci.*, pp. 158–164 (2005).
- V. Raman, P. Sumari, and G. D. P. Raj, "Performance based CBR mass detection in mammograms," in *Int. Conf. Commun. Control and Comput. Technol.*, pp. 565–568 (2010).
- G. Vani, R. Savitha, and N. Sundararajan, "Classification of abnormalities in digitized mammograms using extreme learning machine," *Proc. 11th Int. Conf. Control Autom. Rob. and Vision*, pp. 2114–2117 (2010).
- W. C. Lin, S. C. Hsu, and A. C. Cheng, "Mass detection in digital mammograms system based on PSO algorithm," in *Proc. IEEE Int. Symp. Comput., Consum. and Control*, pp. 662–668 (2014).
- T. Y. Tan et al., "Intelligent skin cancer detection using enhanced particle swarm optimization," *Knowledge-Based Syst.* **158**, 118–135 (2018).
- A. Yaguchi et al., "Cancer detection from biopsy images using probabilistic and discriminative features," in *Proc. IEEE Int. Conf. Image Process.*, pp. 1609–1612 (2011).
- A. Torrent et al., "A supervised micro-calcification detection approach in digitised mammograms," in *17th IEEE Int. Conf. Proc. Image Process.*, pp. 4345–4348 (2010).
- L. Wei et al., "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Trans. Med. Imaging* **24**, 371–380 (2005).
- Y. M. George et al., "Remote computer-aided breast cancer detection and diagnosis system based on cytological images," *IEEE Syst. J.* **8**(3), 949–964 (2014).
- P. D. Shahare and R. N. Giri, "Comparative analysis of artificial neural network and support vector machine classification for breast cancer detection," *Int. Res. J. Eng. Technol.* **2**, 2114–2119 (2015).
- S. Lahmiri and M. Boukadoum, "Hybrid discrete wavelet transform and Gabor filter banks processing for mammogram features extraction," in *Proc. New Circuits and Syst. Conf.*, pp. 53–56 (2011).
- A. E. Hassanien and T.-H. Kim, "Breast cancer MRI diagnosis approach using support vector machine and pulse coupled neural networks," *J. Appl. Logic* **10**(4), 277–284 (2012).

35. S. Allwin, S. P. K. Kenny, and V. Manian, "Classification of stages of malignancies using textron signatures of a cervical cyto image," in *IEEE Int. Conf. Comput. Intell. and Comput. Res.*, pp. 1–4 (2010).
36. S. M. Han, H. J. Lee, and J. Y. Choi, "Prostate cancer detection using texture and clinical features in ultrasound image," in *Proc. Int. Conf. Inf. Acquis.*, pp. 547–552 (2007).
37. S. Naik et al., "A quantitative exploration of efficacy of gland morphology in prostate cancer grading," in *Proc. IEEE 33rd Annu. Northeast Bioeng. Conf.*, pp. 58–59 (2007).
38. X. Xu et al., "On the classification of prostate pathological images based on Gleason score," in *Proc. Intell. Inf. Technol. Appl. Workshops*, pp. 605–608 (2008).
39. S. Doyle et al., "Automated grading of prostate cancer using architectural and textural image features," in *Proc. 4th IEEE Int. Symp. Biomed. Imaging: From Nano to Macro*, pp. 1284–1287 (2007).
40. S. Doyle et al., "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," in *5th IEEE Int. Symp. Biomed. Imaging: From Nano to Macro*, pp. 496–499 (2008).
41. X. Yuan et al., "SVM-based texture classification and application to early melanoma detection," in *Int. Conf. IEEE Eng. Med. and Biol. Soc.*, pp. 4775–4778 (2006).
42. S. Deepa and V. S. Bharathi, "An efficient digital mammogram image classification using DTCWT and SVM," in *Proc. Second Int. Conf. Comput. Sci., Eng. and Inf. Technol.*, pp. 288–293 (2012).
43. H. I. Kim et al., "SVM-based Harris corner detection for breast mammogram image normal/abnormal classification," in *Proc. Res. Adapt. and Convergent Syst.*, pp. 187–191 (2013).
44. T. Torheim et al., "Classification of dynamic contrast enhanced MR images of cervical cancers using texture analysis and support vector machines," *IEEE Trans. Med. Imaging* **33**(8), 1648–1656 (2014).
45. X.-L. Bai and X. Qian, "Medical image classification based on fuzzy support vector machines," in *Proc. Int. Conf. Intell. Comput. Technol. and Autom.*, Vol. 2, pp. 145–149 (2008).
46. M. A. Alolfe et al., "Computer-aided diagnostic system based on wavelet analysis for microcalcification detection in digital mammograms," in *Proc. Int. Conf. Biomed. Eng.*, pp. 1–5 (2008).
47. M. A. Alolfe et al., "Computer aided diagnosis in digital mammography using combined support vector machine and linear discriminant analysis classification," in *Proc. 16th IEEE Int. Conf. Image Process.*, pp. 2609–2612 (2009).
48. M. Kounelakis et al., "On the relevance of glycolysis process on brain gliomas," *IEEE J. Biomed. Health Inf.* **17**(1), 128–135 (2013).
49. F. V. Farahani, A. Ahmadi, and M. H. F. Zarandi, "Hybrid intelligent approach for diagnosis of the lung nodule from CT images using spatial kernelized fuzzy c-means and ensemble learning," *Math. Comput. Simul.* **149**, 48–68 (2018).
50. S. I. Niwas, P. Palanisamy, and K. Sujathan, "Complex wavelet as nucleus descriptors for automated cancer cytology classifier system using ANN," in *Proc. IEEE Int. Conf. Comput. Intell. and Comput. Res.*, pp. 1–5 (2010).
51. M. Karan and K. R. Gandhi, "Diagnose breast cancer through mammograms, using image processing techniques and optimization techniques," in *Proc. IEEE Int. Conf. Comput. Intell. and Comput. Res.*, pp. 1–4 (2010).
52. M. Moradi et al., "Detection of prostate cancer from RF ultrasound echo signals using fractal analysis," in *Proc. 28th Annu. Int. Conf. IEEE Eng. Med. and Biol. Soc.*, pp. 2400–2403 (2006).
53. M. G. Mini, "Neural network based classification of digitized mammograms," in *Proc. Second Kuwait Conf. e-Serv. and e-Syst.*, pp. 1–5 (2011).
54. M. S. Neofytou et al., "Color multiscale texture classification of hysteroscopy images of the endometrium," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. and Biol. Soc.*, pp. 1226–1229 (2008).
55. G. Naghdy et al., "Classification cervical cancer using histology images," in *Proc. Int. Conf. Comput. Eng. and Appl.*, Vol. 1, pp. 515–519 (2010).
56. C. M. Tiu, T. L. Jong, and C. W. Hsieh, "Self organizing map neural network with fuzzy screening for micro-calcifications detection on mammograms," in *IEEE Conf. Soft Comput. Ind. Appl.*, pp. 421–425 (2008).
57. D. S. Elizabeth et al., "Computer-aided diagnosis of lung cancer based on analysis of the significant slice of chest computed tomography image," *IET Image Process.* **6**(6), 697–705 (2012).
58. T. Wan et al., "Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features," *Neurocomputing* **229**, 34–44 (2017).
59. A. K. Anaraki, M. Ayati, and F. Kazemi, "Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms," *Biocybern. Biomed. Eng.* **39**(1), 63–74 (2019).
60. M. Sajjad et al., "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *J. Comput. Sci.* **30**, 174–182 (2019).
61. Y. Peng et al., "Detection of nuclei clusters from cervical cancer microscopic imagery using C4.5," in *Proc. 2nd Int. Conf. Comput. Eng. and Technol.*, pp. 593–597 (2010).
62. Y. Peng et al., "Clustering nuclei using machine learning techniques," in *Proc. IEEE/ICME Int. Conf. Complex Med. Eng.*, Vol. 3, pp. 52–57 (2010).
63. T. Gul, A. Yilmaztürk, and A. C. Erden, "A review of trophoblastic diseases at the Medical School of Dicle University," *Eur. J. Obstet. Gynecol. Reprod. Biol.* **74**, 37–40 (1997).
64. M. Khaskheli et al., "Gestational trophoblastic disease: experience at a tertiary care hospital of Sindh," *J. Coll. Phys. Surg. Pak.* **17**, 81–83 (2007).
65. N. J. Sebire, "Histopathological diagnosis of hydatidiform mole: contemporary features and clinical implications," *Fetal Pediatr. Pathol.* **29**(1), 1–16 (2010).
66. P. Refaailzadeh, L. Tang, and H. Liu, "Cross-validation," in *Encyclopaedia of Database Systems*, L. Liu and M. T. Ozsu, Eds., pp. 532–538, Springer, New York (2009).
67. A. P. Dhawan, *Medical Image Analysis*, Wiley-IEEE Press, London (2003).

Patison Palee is currently a lecturer in Chiang Mai University in Thailand and defended his PhD in image analysis at Staffordshire University in 2016.

Bernadette Sharp is a professor emerita of applied artificial intelligence at Staffordshire University, United Kingdom. She received her BSc degree in computer mathematics in 1973, her MPhil in statistical forecasting in 1979, and her PhD in automatic abstracting in 1989. She is a fellow and chartered IT professional fellow of the British Computer Society. Her research focuses on the application of artificial intelligence and knowledge discovery to industrial, educational, and medical organizations. She has published over 130 research papers in journals, and conferences, and chapters in books. She is the coeditor of 10 books. She is the chair of the International Workshop on Natural Language Processing and Cognitive Science (NLPCS), which she founded in 2002.

Leonardo Noriega is a professor of legal economic analysis and data science at the Faculty of Business Informatics at the Hochschule Furtwangen University. Previously, he has been a private consultant and senior lecturer at a number of universities in the United Kingdom. He achieved his PhD in machine vision from Nottingham Trent University in 1998 and has researched in many related disciplines, including color physics, statistical, data science and signal processing.

Neil Sebire is a professor of paediatric and developmental pathology at the Institute of Child Health (UCL), consultant pathologist at GOS Hospital for Children, the Regional Trophoblastic Disease Unit at Charing Cross Hospital, London. A fellow of the Royal College of Pathologists, and a member of numerous national expert committees, he contributes to the Royal College of Obstetricians and Gynaecologists guidelines for management of gestational trophoblastic disease and has over 1000 peer-reviewed publications.

Craig Platt is a consultant perinatal pathologist at University Hospital's Bristol NHS Foundation Trust. His research interests include the pathology of the placenta, and perinatal deaths and malformations. Academic degrees: MBChB FRCPath.