



Methods for Automatic Cyst Detection and Classification in Ultrasound Images of the Female Genitalia Using Image Processing

Ahmad Ali AlZubi^{1*}, Hannoud Al Moghrabi², Mallak Ahmad AlZubi³, Sufian Ahmad AlZubi³

¹Computer Science Department, King Saud University, Riyadh, Saudi Arabia.

²Frances Payne Bolton School of Nursing, Case Western Reserve University, USA

³Faculty of Medicine, Jordan University of Science and Technology, Jordan

*Corresponding author: Ahmad Ali AlZubi, Computer Science Department, King Saud University, Riyadh, Saudi Arabia, Email: aalzubi@ksu.edu.sa

Submitted: 17 February 2023; Accepted: 20 March 2023; Published: 01 April 2023

ABSTRACT

Ovarian cysts are a condition that affects female reproductive organs. Experts are able to detect ovarian cysts, which is a disorder that affects a woman's uterus, by examining the cyst's size and characteristics on an ultrasound device. Because the manual interpretation of ultrasound examination data for ovarian cyst size generally produces erroneous results, a tool is necessary to assess the size of the cyst and identify the characteristics of the cyst based on the papillary growth in the cyst. The method proposed here involves taking an ultrasound picture as its input and then running a pre-processing phase to eliminate noise before going on to a segmentation stage using the watershed approach. The last step of the process involves the extraction of individual features from the image. The findings of the segmentation are then utilised for feature extraction, namely, to identify cysts and papillary and their diameters using contour analysis using the bounding box approach.

Keywords: *Ovarian Cysts, Papillary growth, Watershed segmentation, Pre-Processing, Automatic cyst detection*

INTRODUCTION

The cervix and vagina are also part of the female reproductive system together with the uterus, ovaries, and fallopian tubes. The system generates female hormones that regulate the reproductive cycle. The uterus, which has the form of a pear, is a pelvic organ that sits between the rectum and the bladder. Two ovaries, one on each side of the uterus, reside lower in the pelvis than the upper abdomen. They play a crucial role in reproduction by producing both the female gamete and female sex hormones. The pituitary gland secretes luteinizing hormone and follicle

stimulating hormone in response to a signal from the brain. On the day of ovulation, these hormones prompt the follicle to release the egg. In the event of an unsuccessful fertilisation, the uterine lining, blood, and egg are all expelled during the menstrual period. When trying to conceive, if fertilisation does not occur up to one year, it may be categorised as a case of infertility. The infertility has become an important cause of concern in present days and seems to be increasing both in males and females. There could be number of reasons for this rise in infertility cases especially in females including

the changed lifestyle (smoking and alcohol consumption), dietary insufficiency (eating disorders), hormonal disorders, altered structures of reproductive organs (fallopian tube blockage or damage) as well as some underlying health disorders (tumors, infections etc.)

Sonography, or ultrasound, has greatly influenced how doctors diagnose and treat infertility. To "image" or visualise structures in the female pelvis, ultrasound equipment are a valuable addition to the gynaecologist's diagnostic arsenal. s, the most common problem to be diagnosed is associated with ovulation. This is why ovaries are the most common reproductive organ to be checked up for infertile women using ultrasound. This is the first step in evaluating a woman who is unable to conceive where the ovarian status is determined and development of her follicles is tracked.

The development of a big ovulatory cyst, caused by the expansion of a dominant follicle beyond a preovulatory diameter, is a possible cause of infertility. Imaging of the ovary is used to assess its health (normal, polycystic, or multicystic), detect any abnormalities (cysts, dermoids, endometriomas, tumours, etc.), track the development of follicles during ovulation monitoring, and look for signs of ovulation and the formation and function of the corpus luteum. With the use of an ovulation scan, the doctor can pinpoint the exact moment when egg becomes fertile. Most infertility treatments start with this step since ovulation is at the centre of the process. Follicle growth may be seen on the ultrasound screen as a black bubble, and this is monitored on a daily basis. Both using edges to estimate ovarian borders and using thresholding as a segmentation approach provide suboptimal results. Because of enhancements made to the approach in employing active contours, the quality of identified follicles has increased dramatically (Zhang, Huang & Liu, 2019).

However, it is difficult to automatically determine appropriate settings for snakes. Semi-automated outer follicular wall segmentation utilizes the watershed segmentation methodology, which involves regular human tracing of the inner border of all follicles. Some tiny, neighbouring follicles were combined by

watershed segmentation applied to smoothed image data.

For this reason, the field of binary mathematical morphology was used to carve these regions apart. Follicle segmentation uses cellular automata and cellular neural networks . While the results are encouraging, the challenge of determining the necessary criteria for follicular segmentation is an evident downside of these two approaches.

Ultrasound pictures often include a lot of speckle noise, making it hard to determine object boundaries and resulting in subpar segmentation. As part of the pre-processing step, Roth et al. (2019) used an edge-based technique for segmentation and a speckle reduction method based on the contourlet transform for denoising medical ultrasound pictures.

LITERATURE REVIEW

Ovarian cancer (OC) is the sixth greatest cause of cancer mortality among women. Symptoms of OC, such as bloating, pelvic discomfort, weight gain, and expansion of the belly, are often ignored by women who have the disorder (World Health Organization, Cancer Fact Sheet, 2019). Unfortunately, by the time it is diagnosed, the cancer has usually progressed to other organs, making treatment very challenging, and sometimes futile. Over the course of a woman's life, the ovary undergoes major anatomical and functional changes that affect the reproductive system. Women who have gone through menopause and anyone with a family history of OC are at a greater risk of getting the disorder.

Finding an OC in its infancy is challenging. Most women who have a gynaecological tumour eventually die from OC (American Cancer Society, cancer-facts-and-figures-2019). Multiple imaging modalities and serum markers have been utilised in the research to improve the chance of early OC diagnosis (Torre et al., 2018; Chornokur et al., 2013; El-Nabawy et al., 2018). Although there has been promising progress in the use of biomarkers to diagnose ovarian cancer, this approach is not without its downsides, including missed detections, lengthy testing procedures, and the need for specialists.

One of the most widely used biomarkers for ovarian tumour detection is serum carbohydrate antigen 125 (CA125). More than 80% of advanced-stage female patients with OC had elevated blood levels of CA-125 [6]. Ultrasound imaging, magnetic resonance imaging (MRI), and positron emission tomographic (PET) are among of the imaging modalities used to diagnose and characterise human OC tumours. Classification accuracy cannot be guaranteed using machine learning techniques such as conventional support vector machine (SVM), random forest, ensemble SVM, logistic regression, or boosting (Wu et al., 2018). An early diagnosis of ovarian cancer may be possible with the use of a biomarker and a machine learning algorithm. Researchers have supervised machine learning techniques that use manually derived features to identify photos as malignant or benign which have proved useful for early diagnosis of ovarian cancer (Wu et al., 2018; Shibusawa et al., 2016).

Support vector machines were enhanced with textural and pathological cues to classify thyroid nodules, as shown by Chen et al. (2010). Chang et al. (2010) employed SVM on ultrasound images to identify graves illness. In order to automatically classify ovarian cancers from ultrasound pictures, Martínez-Más et al. (2019) compared many popular Machine Learning (ML) algorithms, including KNN, Linear Discriminant (LD), Support Vector Machine (SVM), and Extreme Learning Machine (ELM). Logistic regression and other machine learning methods were also used to make OC forecasts by the authors (Mingyang et al., 2010).

Regular feature extraction approaches need computationally intensive processes that are created by hand. The ensuing high dimensionality, hefty workloads, inefficiency, and low classification rates are all the outcome of this (Chang et al., 2010; Mingyang et al., 2010). As an added complication, knowing features inside and out is necessary for extracting the best characteristics when collecting data. Fortunately, deep learning is here to help machine learning get beyond its limitations so that it can be used to handle enormous datasets (Mingyang et al., 2010).

The ability of deep learning algorithms to automatically learn aspects of raw data is a major benefit of these methods.

Classifying OCs using a 15-neuron ANN model was done by Rahman et al. (2020). Many studies have cautioned that, while using Deep Learning algorithms, the system must not depend on human feature computation (Ma et al., 2016; Costa et al., 2018).

Many scientists are interested in applying deep learning algorithms to analyse medical photos. Recent developments in deep learning algorithms have a wide range of medical applications (Krizhevsky et al., 2012). These include cancer prediction, tumour cell segmentation, disease detection, and many more. The Deep Convolutional Neural Network (DCNN) approach was presented by Roth et al. (2014) for developing a lymph node identification system.

DCNN's ability to extract useful features from pictures serves a variety of detection, recognition, and retrieval needs. By using a non-linear network, these deep learning techniques produce a multilayer neural network for feature extraction and data reading. The combination of both low- and high-level features allows for fast learning of the dataset's important properties and the construction of an in-depth representation of incoming data. It has been possible to categorise OC from pathological images using the DCNN-based technique developed by Spanhol et al. (2016). Li et al. (2016) used the DCNN method to categorise pulmonary modules.

According to the research, normal AI algorithms are crucial in detecting ovarian cancer, but they still can't compete with a human pathologist's level of accuracy. In recent years, deep learning methods have shown promising results in medical image analysis applications such as detecting thyroid nodules, analysing breast cancer, and detecting lung cancer. While deep learning has not been extensively studied in the context of ovarian cancer, this may change as new tools become available. Pre-trained deep convolutional neural network (DCNN) models like the one used by Wu M. and colleagues for OC classification from histopathological images

could not achieve an accuracy of more than 78% as they reported.

ImageNet, a training dataset utilised by cutting-edge models like AlexNet and VGG-Net, has 14 million images over 1000 classifications. AlexNet came out in 2012, VGG-Net and GoogleNet in 2013, ResNet in 2015, MobileNet and DenseNet in 2016, and VGG-Net and GoogleNet in 2017 were the last two releases in this series. These are the basis of most pretrained CNN architectures. This paper introduces a novel architecture for utilising deep learning to predict ovarian cancer using histopathology pictures. For the purpose of algorithm training, we only utilise histology images.

OBJECTIVE OF THE RESEARCH

The study's overarching goal is to create a Computer-Aided Process (CAS) for automated follicle identification and ovarian image classification into Normal, Cyst, and PCOS categories (figure 1). Steps in the CAS includes:

Noise reduction in ovarian imaging by pre-processing

To create innovative methods for follicle image segmentation using ultrasonography of the ovary.

To attempt classification of ovarian kinds after they've been analysed and categorised, with a focus on the follicle characteristics.

METHOD

The creation of a Computer Assisted System (CAS) for cyst and PCOS identification in ultrasound ovarian pictures is aimed at delivering helpful data to radiologists.

Detailed steps of the proposed CAS are shown in Figure 2 (CAS). Ultrasound pictures of the ovaries are sent into the CAS computer. Ultrasound pictures of the ovaries are first pre-processed using a variety of filters, including the

Lee, Kuan, Frost, Gaussian, Wiener, Median, Hybrid median, Modified hybrid median, and Fuzzy filters. Metrics for measuring effectiveness are used to assess the filters' efficiency. According to the findings, the hybridization between the Modified hybrid median and the Fuzzy filters yields improved denoising outcomes. To categorise the ovarian pictures as Normal, Cyst, or PCOS, geometrical characteristics are derived from the segmented follicles to identify real follicles. To better categorise ovarian pictures, SVM classifier has been utilised. For this reason, the suggested CAS may use a hybrid fuzzy filter for pre-processing, the MIWO approach for segmentation, and an SVM classifier for ovarian type classification. The use of this CAS in radiology and gynaecology has great promise.

Pre-Processing Stage

Since ultrasound pictures are always noisy owing to the technique of image capture itself (for example, the head of both the ultrasound device is indeed not moist enough), denoising that input image is part of the pre-processing step. Disturbing noise, and speckle noise in particular, is a common problem. Thus, we resort to denoising medical ultrasound pictures using the more effective speckle reduction approach based here on contourlet transform. In this approach, we first conduct a transform similar to a wavelet in order to identify edges, and then we apply a development practitioner's transform in order to identify contour segments. So, to get sparse expansions for common pictures with contours, the contourlet transform uses a double filter bank strategy. To further improve the brightness of the despeckled picture, histogram equalisation is then done. Since the suggested segmentation approach is effective on maximum energy valued objects, we also apply a negative modification to the distribution equalised picture.

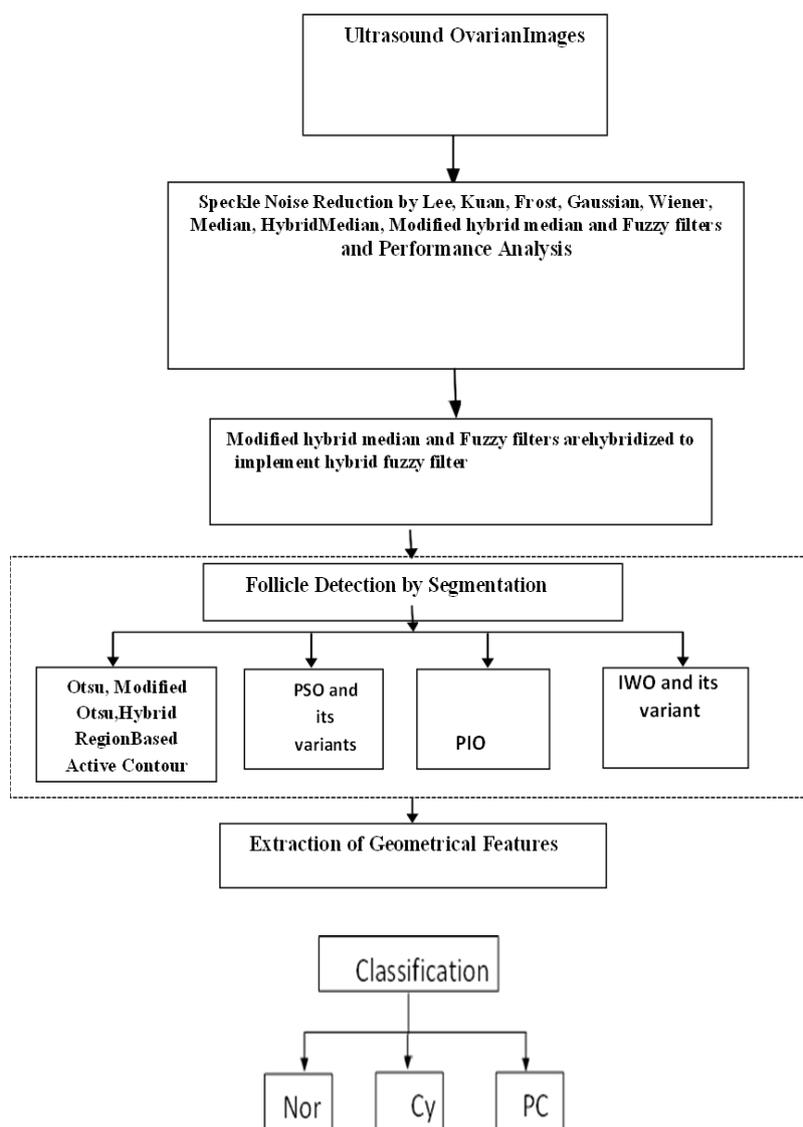


FIGURE 1: CAS overall system

Segmentation Stage

The picture after the negative transformation is used as the input image for the segmentation procedure. We employed the active contour technique without borders for segmentation. When the active contour approach is applied to a picture, it produces a final product that includes segmented sections. The geographical area around and including eliminating the boundaries. Morphological erosion gets rid of any noise-induced false areas. Areas smaller than the threshold T are filtered out of the segmented picture. After that, we gave names to the areas and plug any gaps we find inside them.

Feature Extraction

The ovarian follicles are compact structures with an oval shape that resembles a circle or ellipse. These follicles can be identified by their seven geometric features: the ratio R of major axis length to minor axis length; the compactness C_p ; the circularity C_r ; the tortuosity T_r ; the extent E ; and the extent and $C =$ a rotational constant that defines a spherical symmetry (C_x, C_y). One important follicular region description is the ratio of follicle major axis length to minor axis length. Ultrasound pictures show the follicles to be loosely packed, spherical entities with circular projections.

Follicle Detection

With SVM, you may learn the hyper planes that divide a high-dimensional feature space in a computationally efficient manner. The SVM's outstanding generalisation performance without the need for specialised training data has made it a viable solution for a wide variety of practical situations must supplement with earlier information. As a result, SVM is widely recognised as a powerful method for pattern classification, image analysis, and bioinformatics. For the SVM model, a priori-chosen non-linear mapping is all that's needed to transform the input vectors into the high-dimensional feature space. To achieve this goal, we use the structural risk reduction principle to create a separating hyperplane in this space with the aim of reducing the maximum possible generalisation error. The SVM outperforms other classifiers even when just a few training examples are available, and its performance is not reliant on the size of the feature space. It is applicable to both one-class and n-class classification issues.

Classification Of Ovarian Images

The retrieved characteristics of the follicles have been used in a Support Vector Machine to classify a picture of an ovary as Normal, Cyst, or PCOS (SVM). With the SVM, the number of

dimensions is irrelevant even with little data, it outperforms other classifiers in the feature space and may provide impressive results.

Data Augmentation

Adding fresh data is a critical step in deep learning. Because of the large quantity of data required for DCNN, data augmentation is often required; acquiring a large number of pictures is not always possible. Ultimately, it contributes to expanding the database and introducing more ambiguity. Having too little data to train on is another common source of overfitting [7, 15]. This picture modification or enhancement is accomplished in a variety of ways, such as via zooming, tilting, and accentuating certain features. We flipped the original picture horizontally and vertically, increased the brightness, zoomed in to capture more details, and rotated the original image by 90 degrees. After the images were improved, an additional 24,742 were acquired, bringing the total to roughly 50 times the number of shots in the original dataset. All RGB pictures were scanned as JPG files, scaled to the identical size of 227x227 pixels. Methods for improving the quantity of training data are shown in a number of forms in Figure 2. Two models were developed, one for the standard dataset and another for the improved image set, which was around 100 times bigger.

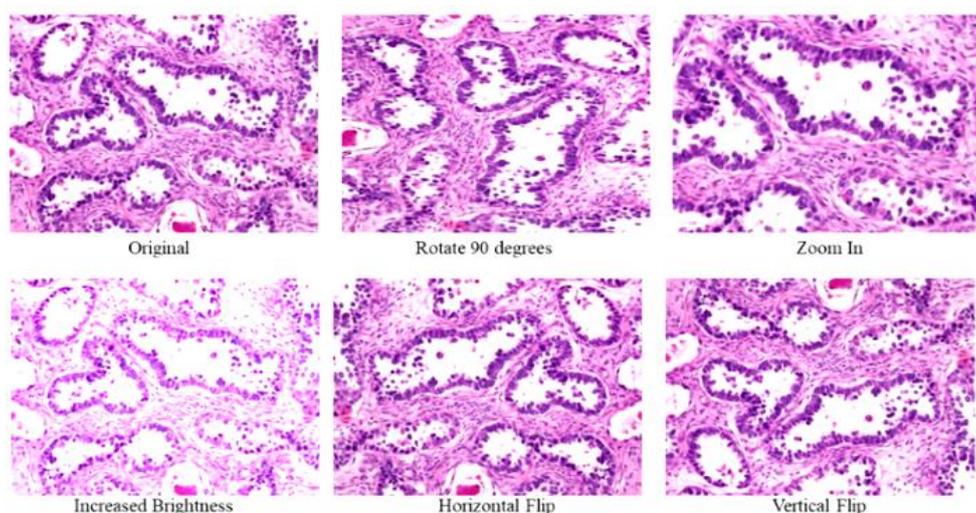


FIGURE 2: Augmented histopathological images

Results Of Proposed Model: Kk-Net

An original and modified dataset was used for DCNN training and evaluation. Tenfold cross-validation was used to test its categorization accuracy. According to Table 1, each class's original and enhanced photos are provided. Each

class has roughly 5000 photos after augmentation. Figure 2 depicts the classification model's training accuracy, training loss, validation accuracy, and validation loss; all plotted against epoch. Results of KK-Nets measurements are shown in the third table

TABLE 1: Number of images of each class

Class	Original Images	Augmented Images
Serous	200	5850
Mucinous	300	5950
Endometrioid	70	4353
Clear Cell	90	5852
Non-Cancerous	125	4689
Total	785	26694

TABLE 2: Shows the classification accuracy of the classes in scope with the original and augmented models.

Class	Original Images	Augmented Images
Clear Cell	78%	89%
Endometrioid	87%	94.%
Mucinous	75%	89%
Non-Cancerous	82%	82%
Serous	88%	92%

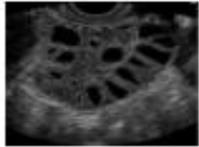
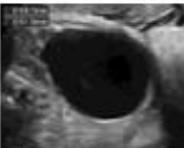
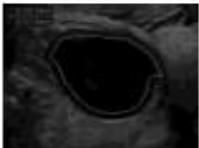
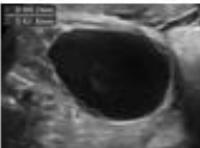
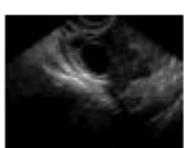
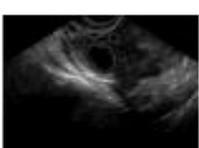
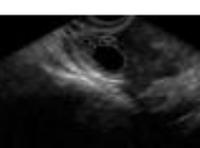
Original Image of ovary	Ovary with detected follicles (Proposed method)	Manual follicle detection by expert	Ovary classification
			Polycystic ovary
			Cystic ovary
			Normal ovary with 2 antral follicles and 1 dominant follicle.

FIG 3: Sample results for the proposed ovarian classification method

CONCLUSION

In this study, we provided a unique approach to classify ovaries from ultrasound pictures. The active contour without edges approach is used for pre-processing the ultrasound picture of the ovary, and the contourlet transform is used for follicle segmentation. SVM classification is used for follicle detection. After follicle detection, the ovary is further categorised based on two factors: follicle number (N) and follicle size (S). To identify whether an ovary is healthy, cystic, or polycystic, an SVM classifier is used. The experimental findings corroborate the effectiveness of the suggested strategy by showing excellent agreement with the manual identification of the ovarian classes by medical specialists. Both the support vector machine and the fuzzy classifier are perfect at distinguishing ovarian tissue from other types. Because of this, the suggested technique provides a solid foundation for the continuous automated categorization of ovaries during the whole female reproductive cycle. It aids in the study of ovarian morphology and greatly enhances the quality of diagnosis and therapy for infertile individuals.

ACKNOWLEDGEMENT

The authors would like to thank the editors and reviewers for their review and recommendations and also to extend their thanks to King Saud University for funding this work through the Researchers Supporting Project (RSP2023R395), King Saud University, Riyadh, Saudi Arabia.

Funding Statement

This work was supported by the Researchers Supporting Project (RSP2023R395), King Saud University, Riyadh, Saudi Arabia.

Authors' contributions

All authors contributed toward data analysis, drafting and revising the paper and agreed to be responsible for all the aspects of this work.

Declaration of Conflicts of Interests

Authors declare that they have no conflict of interest.

Data Availability Statement

The database generated and /or analysed during the current study are not publicly available due to privacy, but are available from the corresponding author on reasonable request.

Declarations

Author(s) declare that all works are original and this manuscript has not been published in any other journal.

REFERENCES

1. American Cancer Society., Cancer facts, and figures. US: American Cancer Society (2019) Available: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf> 2019.
2. Chang, C. Y., Liu, H. Y., Tseng, C. H., & Shih, S. R. (2010). Computer-aided diagnosis for thyroid graves' disease in ultrasound images. *Biomedical Engineering: Applications, Basis and Communications*, 22(2), 91-99. <https://doi.org/10.4015/S1016237210001815>
3. Chen, S. J., Chang, C. Y., Chang, K. Y., Tzeng, J. E., Chen, Y. T., Lin, C. W., Hsu, W. C., & Wei, C. K. (2010). Classification of the thyroid nodules based on characteristic sonographic textural feature and correlated histopathology using hierarchical support vector machines. *Ultrasound in medicine & biology*, 36(12), 2018-2026. <https://doi.org/10.1016/j.ultrasmedbio.2010.08.019>
4. Chornokur, G., Amankwah, E. K., Schildkraut, J. M., and Phelan, C. M. (2013). Global ovarian cancer health disparities. *Gynecologic Oncology*, 129(1), 258-264. <https://doi.org/10.1016/j.ygyno.2012.12.016>
5. Costa, A. C., Oliveira, H. C. R., Catani, J. H., de Barros, N., Melo, C. F. E. & Vieira, M. A. C. (2018) "Data augmentation for detection of architectural distortion in digital mammography using deep learning approach," arXiv:1807.03167. [Online] Available: <http://arxiv.org/abs/1807.03167>

6. El-Nabawy, A., El-Bendary, N., Belal, N. (2018). Epithelial Ovarian Cancer Stage Subtype Classification using Clinical and Gene Expression Integrative Approach. *Procedia Computer Science*, 131, 23-30. <https://doi.org/10.1016/j.procs.2018.04.181>
7. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, Nevada, 1097-1105.
8. Li, W., Cao, P., Zhao, D., & Wang, J. (2016). Pulmonary nodule classification with deep convolutional neural networks on computed tomography images. *Computational and Mathematical Methods in Medicine*, 6215085. <https://doi.org/10.1155/2016/6215085>
9. Lu, M., Fan, Z., Xu, B., Chen, L., Zheng, X., Li, J., Znati, T., Mi, Q., & Jiang, J. (2020). Using machine learning to predict ovarian cancer. *International journal of medical informatics*, 141, 104195. <https://doi.org/10.1016/j.ijmedinf.2020.104195>
10. Ma, S., Sigal, L. & Sclaroff, S. 2016. Learning Activity Progression in LSTMs for Activity Detection and Early Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 1942-1950, <https://doi.org/10.1109/CVPR.2016.214>
11. Martínez-Más J, Bueno-Crespo A, Khazendar S, Remezal-Solano M, Martínez-Cendán JP, Jassim S, Du H, Al Assam H, Bourne T, Timmerman D, "Evaluation of machine learning methods with Fourier Transform features for classifying ovarian tumors based on ultrasound images," *PLOS ONE*, Available: 2019. <https://doi.org/10.1371/journal.pone.0219388>
12. Menchón-Lara, R-M., Sancho-Gómez, J-L., Bueno-Crespo, A. (2016). Early-stage atherosclerosis detection using deep learning over carotid ultrasound images. *Applied Soft Computing*, 49, 616-628. <https://doi.org/10.1016/j.asoc.2016.08.055>
13. Rahman, M.A., Muniyandi, R.C., Islam, K.T., & Rahman, M.M. (2019). Ovarian Cancer Classification Accuracy Analysis Using 15-Neuron Artificial Neural Networks Model. 2019 *IEEE Student Conference on Research and Development (SCORED)*, 33-38. <https://doi.org/10.1109/SCORED.2019.8896332>
14. Roth, H. R., Lu, L., Seff, A., Cherry, K. M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., & Summers, R. M. (2014). A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 17(Pt 1), 520-527. https://doi.org/10.1007/978-3-319-10404-1_65
15. Shibusawa, M., Nakayama, R., Okanami, Y., Kashikura, Y., Imai, N., Nakamura, T., Kimura, H., Yamashita, M., Hanamura, N., & Ogawa, T. (2016). The usefulness of a computer-aided diagnosis scheme for improving the performance of clinicians to diagnose non-mass lesions on breast ultrasonographic images. *Journal of medical ultrasonics (2001)*, 43(3), 387-394. <https://doi.org/10.1007/s10396-016-0718-9>
16. Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). Breast cancer histopathological image classification using Convolutional Neural Networks. *International Joint Conference on Neural Networks (IJCNN)*, Vancouver, 2560-2567. <https://doi.org/10.1109/IJCNN.2016.7727519>
17. Torre, L. A., Trabert, B., DeSantis, C. E., Miller, K. D., Samimi, G., Runowicz, C. D., Gaudet, M. M., Jemal, A. and Siegel, R. L. (2018.). *Ovarian cancer statistics, 2018. CA: A Cancer Journal for Clinicians*. 68(4), 284-296. <https://doi.org/10.3322/caac.21456>
18. World Health Organization (WHO), *Cancer Fact Sheet*, 2019. Available: <https://www.who.int/news-room/fact-sheets/detail/cancer>
19. Wu, M., Yan, C., Liu, H., & Liu, Q. (2018). Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks. *Bioscience reports*, 38(3), BSR20180289. <https://doi.org/10.1042/BSR20180289>
20. Zhang, L., Huang, J., & Liu, L., "Improved deep learning network based in combination with cost-sensitive learning for early detection of ovarian cancer in color ultrasound detecting system," *Journal of Medical Systems*, Vol. 43, no. 8, pp.243 - 251, 2019. <https://doi.org/10.1007/s10916-019-1356-8>