European Journal of Molecular & Clinical Medicine

ISSN 2515-8260 Volume 07, Issue 10, 2020

A Study Of Preprocessing Techniques And Features For Ovarian Cancer Using Ultrasound Images

Ms.ArathiBoyanapalli[1], Dr.Shanthini. A[2]

[1]*Research Scholar, Department of Computer Science and Engineering, SRMIST, Kattankulathur, Chennai*
[2]*Associate Professor, Department of Information Technology, SRMIST, Kattankulathur,Chennai*

Abstract: Ovarian Cancer is the third leading cancer among women in India. The early detection-rate of ovarian cancer is very low [1]. Transvaginal ultrasound is the most common screening test to detect the presence of tumors but adnexal masses are very common in patients, the challenging part is to discriminate whether the masses are benign or malignant. This distinction is very essential for optimal surgical management, but reliable pre-surgical differentiation is sometimes difficult using clinical features, ultrasound examination, or tumor markers alone[2]. Recent trends in medical imaging facilitate the detection of most cancers at a very initial stage. Still, an ovarian cancer diagnosis is not accurate. The patient has to undergo painful practices such as biopsies or surgeries, even with benign nodules. Ultrasound images with deep learning techniques in ovarian cysts help in diagnosis whether the cyst is benign or malignant at a very early stage without any surgeries. This method not only cuts the medical expenses of the patient but also reduces the mental stress of the patients.

Keywords: Ovarian Cancer detection, Deep Learning, Ultrasound Images, Neural Networks

1. INTRODUCTION

Ovarian cancer is a category of cancer to effects the ovaries. The female reproductive system consists of two ovaries. Ovarian cancer is often detected when it is spread to the pelvis and abdomen, it is very hard to treat at this stagesovarian cancer can be successfully treated when the disease is confined to the ovary. The patient has to undergo surgery and chemotherapy to treat ovarian cancer. Ovarian cancer can be classified as epithelialtumors, Stromal tumors, and Germ cell tumors. Epithelial tumors begin in the thin layer of tissue that covers the ovaries, 90% of ovarian cancers are epithelial tumors. Stromal tumors begin in the ovarian tissue that contains hormone-producing cells, 7% of ovariantumors are stromal. Germ cell tumors begin in egg-producing cells. This is rare cancer which generally occurs in younger women [3]

ISSN 2515-8260 Volume 07, Issue 10, 2020

Ultrasound Imaging is the most convenient and secure method for the diagnosis of internal organs. Many imaging tools can be used to scan internal organs such as Magnetic Resonance Image (MRI), Computerized Tomography (CT), but ultrasound images can be easily portable. Ultrasound images are less expensive, safe, and very useful in diagnosing the infections of the internal organs[4][5]. Ovarian Cancer is one of the most frequently occurring cancers in women, according to statistics nearly 93% of the patients can be survived if cancer can be detected at an early stage, but only 20% of the cases can be diagnosed due to the availability of non-accurate methods [6]. Tumors are common in patients but it is very challenging to differentiate whether the tumor is benign or malignant. The patient has to undergo painful practices such as biopsies or surgeries that cause mental stress to the patient. The introduction of Deep Leaning in medical imaging has made a great impact on health care and other sectors too. Radiology has seen significant developments with Deep Learning which provides better judgment while lowering healthcare costs [7].

1.1 Relevant Terms

Some of the frequently used architectures for image processing are discussed below.

Convolution Neural Network:convolutional neural network also called CNN/ ConvNet. It is a Deep Learning algorithm essentially images are given as input and the algorithm can assign importance to different objects in the image. CNN requires very little pre-processing as compared to other classification algorithms.While in primitive methods filters are hand-engineered, with enough training, ConvNets can learn these filters/characteristics[8].Feature extraction and classification are the major parts of the convolutional neural network. Input Layer, Convolutional Layer, and Pooling Layer are the layers of feature extraction, Fully connected nodes and Output Layer are the layers of classification. The output can be either binary o multiclass classification.

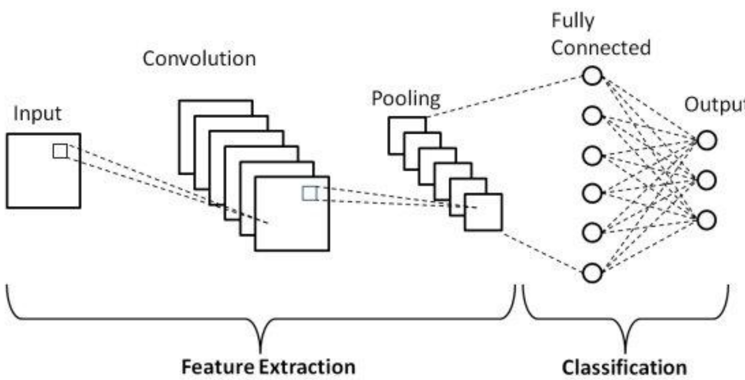

Figure 1.1[9] CNN architecture

VGGNet*:* A CNN architecture proposed by Karen Simonyan and the Andrew of the Zisserman university of Oxford in 2014[10]. ImageNet dataset is used to pre-train the CNN model. The input is an RGB image with 224 * 224 pixels. Pre-processing Layer subtracts the

ISSN 2515-8260 Volume 07, Issue 10, 2020

mean image values from pixel values of 0-255. This pre-processing is adapted for the whole ImageNet dataset. The images after pre-processing are passed through a set of convolution layers. VGGNet has two types depending on the number of convolutional layers. They are VGG16 and VGG19. VGG16 consists of 13 convolutional layers and 3 fully connected layers. VGG19 consists of 16 convolutional layers and 3 fully connected layers, both the variants have 5 max pool layers. VGG16 and VGG19 can predict 1000 labels, last fully connected layer consists of the softmax layer, used for classification.

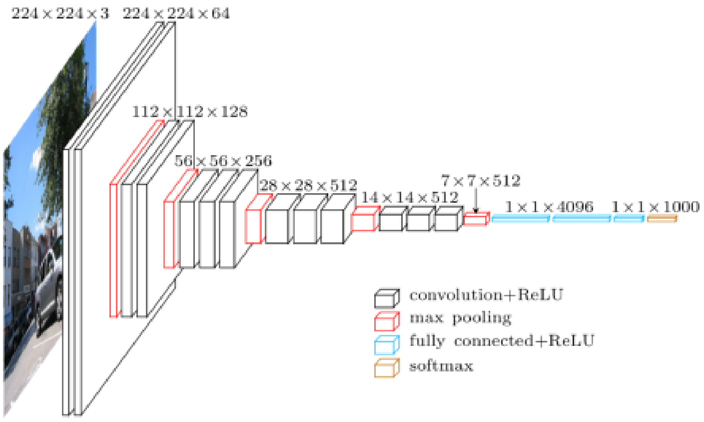

Figure 1.2 VGG16 architecture

AlexNet: AlexNet is apowerfulmodelthat achieves great accuracies on very challenging datasets. AlexNet is one of the most preferred architecture for object detection tasks and very useful in the applications of computer vision and artificial intelligence. Some new approaches incorporated are Relu Nonlinearity, Multiple GPUs, Overlapping Pooling [11].

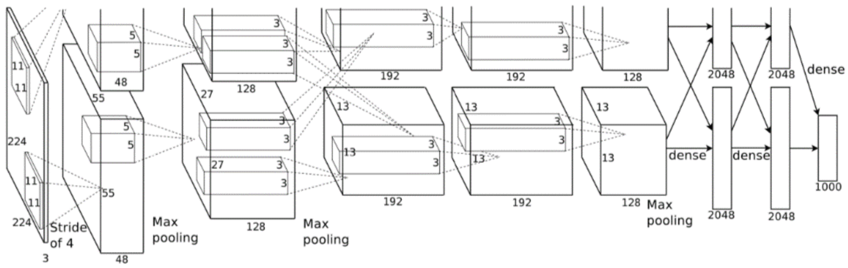

Figure 1.3 AlexNet architecture[12]

GoogLeNet:GoogLeNet is evidenced to be a suitable classifier for images. GoogLeNet is the ILSVRC classification champion in 2014. GoogLeNet has of 27-layers that consist of an inception layer which is a combination of 1x1, 3x3, and 5x5 convolutional layers, whose output is combined into a single output and given as input to the next stage.[13]

2. RELATED WORKS:

In recent years there are several studies for automated early detection and classification of ovarian cancer. In a study [14]convolutional neural networks, a fully connected residual network and a U-Net with the binary and multiclass approach. The models are trained with five different types of ultrasound data which varied from beam-formed radio frequency to brightness mode. The best results are provided by B-mode for ovary and follicles segmentation on both the models. The multiclass classification was beneficial as it provided the spatial relationship between follicles and ovary. C. Wu et. al.[15] used VGGNet, DenseNet, ResNet, GoogleNet models on 988 ultrasound images for the classification of ovarian cancer and best performance using GoogleNet model, The accuracy provided by GoogleNet model is 92.50%.M. Ahmad et. al.[16] used a new fuzzy histogram equalization technique on the fuzzy normalized histogram of an image. The effectiveness of the algorithm is justified over MRI images, CT scan images, and Ultrasound images and attained 20% to 35% enhancement in the identification of ovarian cancer as compared to traditional image processing solutions.U. Rajendra Acharya *et al*. [17]used Probabilistic Neural Networks (PNN) classifiersfor classifying images into benign and malignant. The Gabor transform parameters and entropies are first extracted from the ultrasound images and then trained to a PNN model. The model with 1300 benign images and 1300 malignant images provides the best performance using the Genetic Algorithm and evaluated with 10-fold cross-validation with an accuracy of 99.8%, the sensitivity of 99.2 %, and specificity of 99.6 % with an σ of 0.264.S. Khazendar *et al.*[18] used transvaginal 2D B mode ultrasound images are pre-processed and enhanced. Local Binary Pattern Histograms are extracted from each image. An SVM model is used to train the images using stratified cross-validation with 100 random samples selected for each round repeating the process 15 times. This model achieved 77% accuracy with Local Binary Pattern operator. D. Timmerman *et al.*[19] proposed a method for whichdata is collected from 9 European centers in which 800 patients have benign tumors and 266 has malignant tumors, 12 variables are considered for the logistic regression model and achieved an accuracy of 95%.

3. COMPARISON:

Reference	Dataset	Sample size	Features	classification and accuracy	Highlights
[20]	Used a combination of open source and generated dataset. **Open Source:** • 428 ultrasound images of 560 x 360 pixels • 357- malignant samples • 71- benign samples **Realtime dataset:** • 1400 ultrasound images • 277 samples of ovary	**Total Samples:** Training: 2745 Validating: 549 Testing: 549 **Malignant Samples:** Training: 2340 Validating: 459 Testing: 450 **Benign Samples:** Training: 452 Validating: 101 Testing: 99	**Low-Level texture features:** The Low-level features that are extracted from the images are Uniform Local Binary Pattern (ULBP) **high-level deep features:** GoogLeNet is finely tuned to normalize ULBP features and combined with high-level features.	• VGNet - 97.12 % accuracy • AlexNet- 98.29 % • FCNN - 93.08 % • CNN - 92.06 % • GoogleNet - 96.68 % ● Random Classifier - 99.15%	1) Data Expansion Method for augmentation of data 2) The deep features are extracted and combined with texture features. 3) 10-fold cross-validation is used

	cancer • 299 samples with a benign cyst • normal ovary ultrasound images				
[21]	The real-time dataset generated in the Institute of Tokyo Women's Medical University Medical Center East was used under the approval of the Institutional Review Board (IRB).	**Sample Size:** 53 - ovarian cancers, 23 - borderline malignant tumors 126 - benign ovarian tumors.	1. Age (yr) 2. Gravity 3. Parity 4. Menopause 5. Endometriosis 6. BMI 7. WBC 8. Hb 9. Platelet (×103/μl) 10. Albumin (g/dl) 11. CRP (mg/l) 12. CA125 (U/ml) 13. CA19-9 (U/ml) 14. CEA (ng/ml) 15. Size (cm) 16.	**Support vector machine-** 62.4 % accuracy **Random Forest** - 78.1% accuracy **Naive Bayes** - 62.2% accuracy **Logistic Regression-** 67.1 % accuracy **XGBoost** - 80.4 % accuracy	**Best Predictive features:** albumin, ascites, and CRP **Best Regression coefficient:** CA125, CA19-9, and CRP **Best Features:** platelet, albumin, and CA125 were the top 3 hits in the feature importance of random forest

ISSN 2515-8260 Volume 07, Issue 10, 2020

			Ascites		
[22]	Generated	The Dataset is divided into 100,200,300,400 and 500 images taken during different stages of pregnancy	LRC is used for feature map extraction, segmentation, and classification	Logistic Regression classifier - Precision - 96.5, Recall-99.1	The ML-CNN is a method that uses a combination of the machine learning algorithm and convolution neural network for the diagnosis of an ovarian tumor during pregnancy. The proposed model is integrated with the IoMT platform.
[23]	Generated	240 Ultrasound Images with normal and cyst. Training :160 Validation :80		VGG16 - 92.11%	1) Imagenet dataset is used for pre-training the model. 2) VGG-16 model is used and the modification is done on the last four layers of the neural network.
[24]	Generated in the hospital of Queen Charlotte's and Chelsea London	100 images 44 malignant	The machine learning algorithm proposed automatically extracts the 9 quantitative texture-based feature vectors of different vectors. The filters used are Gabor	Support Vector Machine (SVM): 93%	The proposed method uses the IoTA platform for the study of ovarian cancer using Transvaginal and Transabdominal static 2D B-mode images.

			Filter, Fractal Dimension, Histogram of Gradient, Local Binary Pattern (LBP-256 Bins), 7-moments, - Uniform LBP (ULBP-59), Statistics Histogram, Skewness, Kurtosis.		
[25]	Generated	Total Samples: 469 (non-suspicious: 238, suspicious: 231)	The relief-f method is used for feature selection from the extracted non-linear features and 39 features are selected using the relief-f method from 796 features.	KNN (K=5) - 67.16% Random Forest: 63.75 , FRNN - 71.86 , Fuzzy Forest - 80.60	1) 10- fold cross-validation 2) The separate datasets are created with 200, 225,250,275,300,325,350,375, 400.425,450, and 469 samples and balanced to keep the benign and malicious instances the same in every dataset.

ISSN 2515-8260 Volume 07, Issue 10, 2020

4. DISCUSSION

Ovarian cancer ranked 5th in cancer deaths. Early detection of ovarian cancer helps in curing the disease. A cyst in the ovary can be benign or malignant, to find the cyst to be malignant or benign the patient has to undergo surgery or biopsy. This leads to unnecessary medical expenses and trauma to the patients. Detecting ovarian cancer through ultrasound images can help to decrease deaths due to ovarian cancer and the medical expenses of the patient. Deep Learning techniques can help in the detection of ovarian cancer in ultrasound images. This research can be further extended to MRI and CT scan images too, but ultrasound machines are less expensive and easily available in all the hospitals. The different pre-processing techniques and deep learning algorithms can improve the accuracy of ovarian cancer detection.

5. REFERENCES

[1] "Special Section: Ovarian Cancer." Accessed: Oct. 22, 2020. [Online]. Available: http://surveillance.cancer.gov/devcan/.

[2] M. Grigore, R. Popovici, D. Gafitanu, L. Himiniuc, M. Murarasu, and R. Micu, "Logistic models and artificial intelligence in the sonographic assessment of adnexal masses – a systematic review of the literature," pp. 1–7, 2020, doi: 10.11152/mu-2583.

[3] "Ovarian cancer - Symptoms and causes - Mayo Clinic." https://www.mayoclinic.org/diseases-conditions/ovarian-cancer/symptoms-causes/syc-20375941 (accessed Oct. 22, 2020).

[4] M. Nagayama, Y. Watanabe, A. Okumura, Y. Amoh, S. Nakashita, and Y. Dodo, "Fast MR imaging in obstetrics," *Radiographics*, vol. 22, no. 3, pp. 563–580, May 2002, doi: 10.1148/radiographics.22.3.g02ma03563.

[5] C. B. Benson and P. M. Doubilet, "The history of imaging in obstetrics," *Radiology*, vol. 273, no. 2. Radiological Society of North America Inc., pp. S92–S110, Nov. 01, 2014, doi: 10.1148/radiol.14140238.

[6] "(1) (PDF) MAX POOLING TECHNIQUE TO DETECT AND CLASSIFY MEDICAL IMAGE FOR OVARIAN CANCER DIAGNOSIS." https://www.researchgate.net/publication/340495718_MAX_POOLING_TECHNIQUE_TO_DETECT_AND_CLASSIFY_MEDICAL_IMAGE_FOR_OVARIAN_CANCER_DIAGNOSIS (accessed Oct. 22, 2020).

[7] L. Saba *et al.*, "The present and future of deep learning in radiology," *European Journal of Radiology*, vol. 114. Elsevier Ireland Ltd, pp. 14–24, May 01, 2019, doi: 10.1016/j.ejrad.2019.02.038.

[8] "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way | by Sumit Saha | Towards Data Science." https://towardsdatascience.com/a-comprehensive-guide-

to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 (accessed Oct. 28, 2020).

[9] V. H. Phung and E. J. Rhee, “A High-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets,” *Appl. Sci.*, vol. 9, no. 21, Nov. 2019, doi: 10.3390/app9214500.

[10] Karen Simonyan∗ & Andrew Zisserman+, “VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION Karen,” *Am. J. Heal. Pharm.*, vol. 75, no. 6, pp. 398–406, 2018.

[11] T. F. Gonzalez, “Handbook of approximation algorithms and metaheuristics,” *Handb. Approx. Algorithms Metaheuristics*, pp. 1–1432, 2007, doi: 10.1201/9781420010749.

[12] “AlexNet: The Architecture that Challenged CNNs | by Jerry Wei | Towards Data Science.” https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951 (accessed Nov. 06, 2020).

[13] C. Szegedy *et al.*, “Going Deeper with Convolutions.”

[14] C. C. B, “for Gynaecological Ultrasound Image Segmentation : A Radio-Frequency vs B-mode Comparison,” vol. 1, pp. 295–306, doi: 10.1007/978-3-030-27272-2.

[15] C. Wu, Y. Wang, and F. Wang, *Deep learning for ovarian tumor classification with ultrasound images*, vol. 11166 LNCS. Springer International Publishing, 2018.

[16] M. Ahmad, N. Zaman, and M. Al-Amin, “An experimental research in health informatics for enhancing ovarian cancer identification in ovarian imaging analysis using fuzzy histogram equalization,” *J. Med. Imaging Heal. Informatics*, vol. 7, no. 6, pp. 1385–1390, Oct. 2017, doi: 10.1166/jmihi.2017.2218.

[17] U. Rajendra Acharya *et al.*, “Evolutionary algorithm-based classifier parameter tuning for automatic ovarian cancer tissue characterization and classification,” *Ovarian Neoplasm Imaging*, pp. 425–440, 2013, doi: 10.1007/978-1-4614-8633-6_27.

[18] S. Khazendar *et al.*, “Automated characterisation of ultrasound images of ovarian tumours: the diagnostic accuracy of a support vector machine and image processing with a local binary pattern operator.,” *Facts, views Vis. ObGyn*, vol. 7, no. 1, pp. 7–15, 2015, Accessed: Oct. 28, 2020. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/25897367.

[19] D. Timmerman *et al.*, “Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: A multicenter study by the International Ovarian Tumor Analysis Group,” *J. Clin. Oncol.*, vol. 23, no. 34, pp. 8794–8801, 2005, doi: 10.1200/JCO.2005.01.7632.

[20] L. Zhang, J. Huang, and L. Liu, “Improved Deep Learning Network Based in combination with Cost-sensitive Learning for Early Detection of Ovarian Cancer in Color Ultrasound Detecting System,” *J. Med. Syst.*, vol. 43, no. 8, 2019, doi: 10.1007/s10916-019-1356-8.

[21] M. Akazawa and K. Hashimoto, “Artificial intelligence in ovarian cancer diagnosis,” *Anticancer Res.*, vol. 40, no. 8, pp. 4795–4800, 2020, doi: 10.21873/anticanres.14482.

[22] Z. Zhang and Y. Han, “Detection of Ovarian Tumors in Obstetric Ultrasound Imaging Using Logistic Regression Classifier with an Advanced Machine Learning Approach,” *IEEE Access*, vol. 8, pp. 44999–45008, 2020, doi: 10.1109/ACCESS.2020.2977962.

[23] S. Srivastava, P. Kumar, V. Chaudhry, and A. Singh, “Detection of Ovarian Cyst in Ultrasound Images Using Fine-Tuned VGG-16 Deep Learning Network,” *SN Comput.*

ISSN 2515-8260 Volume 07, Issue 10, 2020

Sci., vol. 1, no. 2, pp. 1–8, 2020, doi: 10.1007/s42979-020-0109-6.

[24] D. Al-Karawi *et al.*, "OC04.04: A machine-learning algorithm to distinguish benign and malignant adnexal tumours from ultrasound images," *Ultrasound Obstet. Gynecol.*, vol. 54, no. S1, pp. 9–10, 2019, doi: 10.1002/uog.20445.

[25] U. R. Acharya *et al.*, "Use of Nonlinear Features for Automated Characterization of Suspicious Ovarian Tumors Using Ultrasound Images in Fuzzy Forest Framework," *Int. J. Fuzzy Syst.*, vol. 20, no. 4, pp. 1385–1402, 2018, doi: 10.1007/s40815-018-0456-9.