Journal of Biological Sciences and Medicine
Available online at www.jbscim.com

ISSN: 2455-5266

Research Article **Open Access**

Image segmentation and association analysis of Endothelin-1 and CD-31 expression in tobacco smokers' placenta using automated Nearest Neighbour and Genetic Algorithm

Banajit Bastia[1], Kapil Kumar[2], Shashi Nandar Kumar[1], Pradeep Kumar Naik[4], Nihar Ranjan Behera[3] and Arun Kumar Jain[1, *]

[1]Environmental Toxicology & Electron Microscope Lab, ICMR-National Institute of Pathology, Safdarjung Hospital Campus, New Delhi-110029, India.
[2]Department of Zoology, Meerut College, Meerut (C.C.S University), U.P- 250001, India.
[3]Department of Obstetrics & Gynaecology, M.K.C.G Medical College & Hospital, Berhampur, Odisha-760004, India.
[4]Department of Biotechnology and Bioinformatics, Sambalpur University, Burla, Sambalpur, Odisha. India.

*Corresponding Author: drakjain@gmail.com

ARTICLE INFO

Article History:
Received
29 January 2018

Accepted
15 March 2018

Available online
31 March 2018

Key words:
CD-31; Endothelin-1; Genetic Algorithm; Nearest Neighbour; Placenta; Tobacco smoker

ABSTRACT

Expression of the protein markers, Endothelin-1 (ET-1) and CD-31 in human placenta is reported to be associated with the amount of exposure to tobacco smoke. We have established through immunohistochemical evaluation that the expression level of ET-1 protein is consistently very high among the women who directly smoke tobacco (active smokers) compared to the other group of women who inhale environmental tobacco smoke or ETS (passive smokers) and non-smokers. To determine the relative expression of both the proteins in the immunohistochemistry images from both the group of women, we have utilized combination of genetic algorithm and nearest neighbor (GA-NN model). Based on the extracted features from the image the GA-NN model efficiently detected the differential expression of the marker proteins and categorized them as active smokers or passive smokers. The computational model was found to be very robust (predictive accuracy was 90.76 ± 2.50 %, Matthews Correlation Coefficient (MCC) of 0.7125, Q_{pred} = 82.85 ± 4.246%, sensitivity = 89.24 ± 2.45% and specificity = 83.43 ± 2.38%. The rate/intensity of expression for ET-1 and CD-31 was more for active smokers as compared to passive smokers. From these results, it could be concluded that computer aided diagnosis based on the immunohistochemical (IHC) images of tissue biopsies from different diseases can support the clinicians in framing their opinion about the comprehensive diagnosis of patient's ailment and its cure, especially in country like ours which lacks experienced histopathologists.

Introduction

Endothelin-1 (ET-1) and CD-31 expression level in the placenta depends on the exposure to tobacco smoke, either through active or passive smoking or both (Bastia et al. 2017). It has been seen that in case of women who smoke tobacco very frequently (Active Smokers), the rate of expression of these two proteins is high; whereas in the second category of women who do not

smoke tobacco, but they inhale the smoke from other external sources, such as, environmental tobacco smoke (Passive Smokers), the rate of expression is low (Mastrogiannis et al. 1991; Nova et al. 1991; Schiff et al. 1992; Kozuka et al. 1989; Wolff et al. 1993). The higher expression of these two proteins (ET-1 and CD-31) leads to the placental abnormalities that may in turn lead to birth failure or major pregnancy associated birth defects, such as, Low birth weight (LBW), intrauterine growth restriction (IUGR), preterm delivery, fetal growth restriction (FGR), sudden infant death syndrome (SIDS), developmental defects, abruptio placenta and increased risks of stillbirth (Sabra et al. 2017). Hence, the study of these two proteins and their expression is very important for proper identification of the affected mother and fetus pairs and thereafter their appropriate protection and salvage from the adverse toxic effects of exposure to tobacco smoke. Traditionally, these types of studies were typically done with the immunohistochemistry (IHC) technique (Lösch and Kainz 1996; Nuovo 2006; Huang et al. 2013; Luiza et al. 2014); the inference was generally derived based on the visualization of the images by an experienced pathologist. Since this is a manual method sometimes the diagnosis is erroneous and wrongly classifies the images. In this regard, it was essential to construct an automatic tool using a reproducible algorithm for prediction, classification and proper diagnosis based on the obtained IHC images. Hence in this study, an attempt has been made to develop a tool using Image Processing and Genetic Algorithm to study the rate of expression of both the proteins in the placenta of smoking mothers.

Materials and Methods

Sample Collection

Placentas (n=200) were collected immediately after normal singleton vaginal delivery from healthy mothers at term who were visiting the Antenatal Clinic (ANC) and/or admitted in Labor Room of Dept. of Obstetrics & Gynecology, M.K.C.G Medical College and Hospital, Berhampur, Odisha, India and had been previously enrolled for this study. The participants were classified chiefly into two groups (n=100 each), Active smokers and Passive smokers based on the tobacco exposure before and during pregnancy. Active Smokers were those pregnant women who directly smoked 15 cigarettes or 30 'bidis' or more per day for at least 3 months before and during pregnancy while Passive Smokers were mothers who were passively exposed to tobacco smoke during pregnancy as her husband/close relatives were chain smokers (who smoke 30 bidis/15 cigarettes or more /day).

Subjects were matched for maternal age, parity and gestational age. Those with history of any debilitating or infectious diseases were excluded from our study. For IHC studies, the placental tissues were collected from the fetal side of placenta after fixed by perfusion technique in 10% neutral buffered formalin fixative. All placentas were examined for conventional gross and light microscopic features as per the guidelines for examination of the placenta developed by the College of American Pathologists (Langston et al. 1997). The study was approved by the Scientific Ethical Committee of M.K.C.G Medical College and Hospital, Berhampur (Odisha) and samples were collected according to the "Ethical guidelines for biomedical research on human participants" issued be Indian Council of Medical Research (ICMR).

Tissue Processing

Five mm^2 (approx) thick placental tissue sections were immersion fixed in 10% buffered formalin after delivery or expulsion of placenta. The fixed placental tissues were processed for routine paraffin embedding. Five micron (5μm) thick sections were cut in a rotary microtome and mounted on poly-L- lysine coated clean glass slides.

Immunohistochemistry (IHC)

The slides selected for IHC study were kept in the oven at 60°C for 20 minutes. Tissue sections were then de-paraffinized and rehydrated. Quenching of the endogenous peroxidase activity was done by immersing the de-paraffinized sections in 3% H_2O_2 solution (in Methanol) for 30 minutes. Sections were rinsed in Phosphate Buffered Saline (PBS, pH-7.2) and then exposed to microwaves for10 minutes in Antigen Retrieval Solution (8.2 mM sodium citrate, 1.8mM Citric acid-pH-6.0, containing 0.01% Triton X-100). After subsequent washing with PBS, the sections were treated with blocking serum for 45 minutes to block the non-specific binding. Followed by incubation in a humidified chamber (at 37ºC) with primary anti-Endothelin-1 monoclonal antibody (mAb) [Oncogene Research Products, Boston, MA, USA] at a dilution of 1:250 (tested over a range of 1:100 to 1:1000), diluted in PBS containing 1% BSA and 0.2% Tween-20 for overnight at 4-8°C. Negative controls were accomplished by omitting primary antibody while rest of the procedure remained the same.

The sections were then washed with PBS containing 0.2% Tween-20 and immuno-staining was attained using an Avidin-Biotin Complex (ABC) developer kit (Vectastain Elite ABC Kit, Vector Laboratories; Burlingame, CA, USA). The complex was detected using 3, 3'-diaminobenzidine (DAB) as a substrate. Sections were counterstained with Mayer's hematoxylin for 2 minutes, washed and mounted (Hsu et al. 1981). In each experiment, an additional serial section was stained with anti-Human CD-31 mAb (pre-diluted concentration of 200μg/ml; Sigma, St. Louis, MO, USA) that served as a positive control and permitted a comparison between the distribution of ET-1 and that of the endothelial cell marker, CD-31. Olympus upright microscope (BX51) fitted with digital camera (Olympus, E-520) was used for capturing microphotographs.

Methodology for Image segmentation, Extraction of features and Classification

We have used image region analysis (IRA) program to select the spatial region (labeled as brown color) of the immunohistochemical image representing the presence of ET-1 and CD31 proteins. From this region, a series of pixels were extracted using image segmentation method, measured their RGB intensity and were used to correlate the overall expression of the marker proteins. The intensity was calculated and written to an XML file. The intensity of the area represented the expression level of the marker proteins. For each image type of known class, such as Endothelin-1 Active smoker, Endothelin-1 Passive smoker, CD31 Active smoker and CD31 Passive smoker, the intensity of the brown punctate was calculated and written to an XML file. We have collected 1200 images out of 200 patients; out of which 600 were from active smokers group (100 pregnant women) and 600 were from passive smokers (100 pregnant women). The images from each category were grouped into a training set (70% of the total images) and a test set (30% of the total images). The training set images were used to train the GA-NN model for automated detection and classification into different groups, whereas the test set was used to test the predictive accuracy of the computational model (Schmitt et al. 1998; Schmitt 2001; Schmitt 2004; Cha and Tappert 2009; Whitley 1994; Akbari and Ziarati 2011).

Five-fold cross-validation

Five-fold cross-validation (5-fold CV) technique had been used for training and testing the GA-NN model, in which the dataset was randomly divided into five subsets, each containing equal number of active smokers and passive smokers. Each set was a balanced set that consist of 50 percent of active smokers and 50 percent passive smokers. The data set had been divided into training and testing set. The training set consisted of five subsets. The GA-NN algorithm was validated for minimum error on testing set to calculate the performance measure for each fold of validation. This had been done five times to test for each subset. The final prediction results were averaged over five testing sets.

Performance Measures

The prediction results of GA-NN model developed was evaluated using the following statistical measures as shown in Table 1 (Naik et al. 2007).

Results

Endothelial cells (EC) were identified and confirmed by immunostaining with endothelial cell marker, anti-human CD-31 mAb. Fine granular and dark brown deposits of ET-1 (achieved due to DAB substrate & chromogen reaction) were observed in the ECs lining the capillaries of tertiary villi of full-term passive smokers' placenta (Fig1). Immunostaining was limited to ECs only and was not discernible in syncytiotrophoblasts or other cells of placenta. The distribution pattern of ET-1 was similar to the immunostaining of endothelial cell marker, anti-human CD-31 (Fig2). This confirmed that the expression and localization of ET-1 was only in the endothelial cells of placental villi.

The distribution pattern of ET-1 immunoreactivity was also investigated in the placenta of active smokers and observed to be more intense as compared to passive smokers. In addition to the ET-1 immunostaining in the ECs lining the capillaries, focal immunostaining was observed in ECs lining small to medium sized blood vessels (Fig3). When compared with the distribution pattern of CD-31 mAb in the serial sections of smoker's placenta, the distribution of ET-1 immunostaining followed similar pattern as in case of CD-31 (Fig4).

This computational model used the RGB color annotations for measuring the intensity and classification of the images into active and passive smokers' category. A series of homogeneous pixels with identical color intensity of RGB were screened out using GA from each respective region of interest and was used for the prediction of the level of expression of both ET-1 and CD-31 marker proteins. The intensity of RGB was distinctly different for both active and passive smokers (Table 3 & 4), which was sufficient to classify both types of images. This information was used to automatically classify the given IHC image to active endothelin-1, passive endothelin-1, active CD-31 and passive CD-31 based on neighboring network. The predictability for the classification of the images in case of Endothelin-1 protein was found to be quite accurate (90.76 ± 2.50 %).

Table 1: Evaluation of the predictability of GA-NN algorithm by several performance measures.

Accuracy of the prediction method	$Q_{ACC} = \frac{P+N}{T}$ Where T=(P+N+O+U)
Matthews Correlation Coefficient (MCC)	$\frac{(P \times N) - (O \times U)}{\sqrt{(P+U) \times (P+O) \times (N+U) \times (N+O)}}$
Sensitivity	$Q_{sens} = \frac{P}{P+U}$
Specificity	$Q_{spec} = \frac{N}{N+O}$
Probability of correct prediction (Q_{Pred})	$Q_{pred} = \frac{P}{P+O} \times 100$

Where, *P* and *N* refer to correctly predicted enzymes and non-enzymes, and *O* and *U* refer to over and under predictions, respectively.

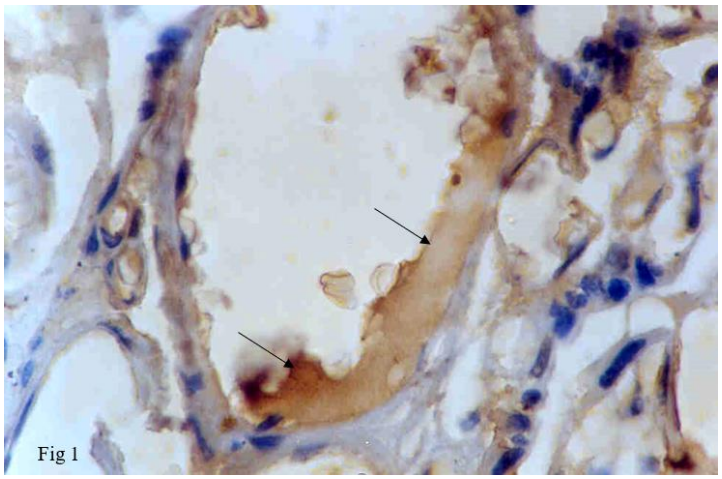

Fig 1. Endothelin-1 (ET-1) expression (arrow) as brown color punctate observed lining the Endothelial cells (EC) of placental chorionic villi. (Passive smoker, 100x)

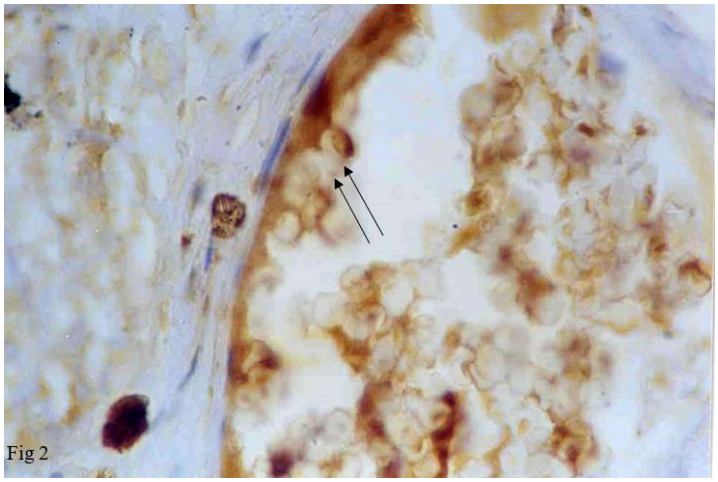

Fig 2. CD-31 immunoreactivity (double arrow) in Endothelial Cells (EC) of placental capillaries is localized to similar regions of ET-1 localization as in Fig1. (Passive smoker, 100x)

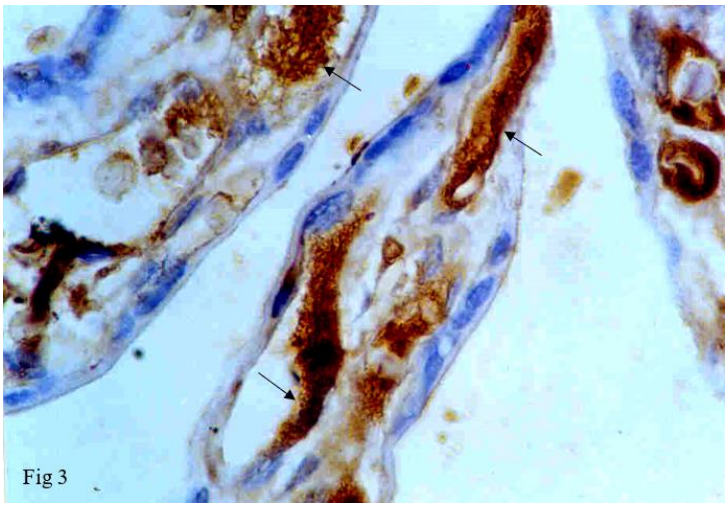

Fig 3. Endothelin-1(single arrow) staining depicted in endothelial cells (EC) of placental villi showing intense staining. (Active smoker, 100X)

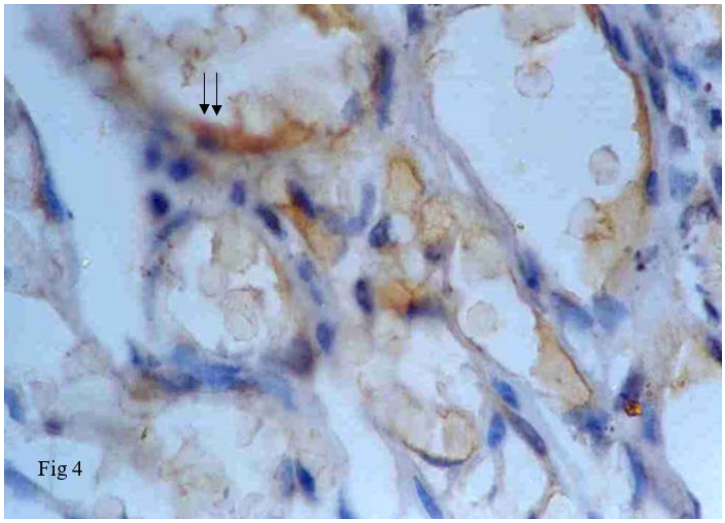

Fig 4. CD-31 immunostaining (double arrow) in endothelial cell (EC) indicating similar distribution pattern as ET-1 positivity. (Active Smoker, 100x)

The prediction results are presented in Table 2. The GA-NN algorithm had achieved an MCC of 0.7125. The other performance measures were: Qpred = 82.85 ± 4.246%, sensitivity = 89.24 ± 2.45% and specificity = 83.43 ± 2.38%. The vast majority of the predictions have been contained within 0.0 to 0.1 for the passive smokers and 0.9 to 1.0 for the active smokers in the case of endothelin-1 marker. Similarly, the prediction accuracy for the classification of images into passive smokers and active smokers was 88.94% based on CD-31 marker. The other performance measures are included in Table 2.

Table 2: Results of predictive measures of GA-NN for the classification of images

5-fold CV	Accuracy	Specificity	Sensitivity	MCC	Q_{Pred}	Prediction range (Active smokers)	Prediction range (Passive smokers)
(a) Classification of the images based on Endothelin-1 marker							
C1	0.9473	0.8420	0.9231	0.8042	87.35	0.943-1.00	0.00-0.312
C2	0.8929	0.8641	0.8843	0.7939	85.32	0.928-1.00	0.00-0.482
C3	0.9162	0.8394	0.8656	0.6364	79.18	0.914-1.00	0.00-0.263
C4	0.8843	0.8266	0.9125	0.6390	84.88	0.929-1.00	0.00-0.474
C5	0.8974	0.7992	0.8765	0.6891	77.54	0.892-1.00	0.00-0.415
Mean	0.9076± 0.0250	0.8343± 0.02379	0.8924± 0.0245	0.7125± 0.0818	82.85± 4.246		
(b) Classification of the images based on CD-31 marker							
C1	0.9260	0.8651	0.8148	0.7648	87.46	0.949-0.973	0.210-0.205
C2	0.8738	0.9128	0.9128	0.8260	88.63	0.942-0.928	0.321-0.606
C3	0.8384	0.8282	0.9203	0.8138	91.38	0.961-0.956	0.166-0.421
C4	0.8967	0.8948	0.8712	0.8345	92.63	0.954-0.923	0.329-0.533
C5	0.9123	0.8745	0.8821	0.8891	86.54	0.913-0.943	0.305-0.483
Mean	0.8894± 0.0345	0.8751± 0.0321	0.8802± 0.0419	0.8256± 0.0446	89.33± 2.591		

Table 3: Extraction of pixel intensity (red, green and blue) of Endothelin-1 expression in the placenta of tobacco smoking pregnant women and their Image segmentation.

Color Tone	Active Smoker			Passive Smoker		
	Max.	Min.	Avg.	Max.	Min.	Avg.
RED	154	21	87.5	143	31	87
GREEN	138	24	81	152	27	89.5
BLUE	80	34	57	144	31	87.5

Table 4: Extraction of pixel intensity (red, green and blue) of CD-31 expression in the placenta of tobacco smoking pregnant women and their Image segmentation.

Color Tone	Active Smoker			Passive Smoker		
	Max.	Min.	Avg.	Max.	Min.	Avg.
RED	126	19	72.5	142	30	86
GREEN	154	16	85	139	31	85
BLUE	132	14	73	116	42	79

Discussion

The developed GA-NN computational model efficiently detected presence and absence of ET-1 and CD-31 proteins in the IHC images of placenta. It also effectively measured the differential expression of both the proteins from the active and passive smoking mothers (Akbari and Ziarati 2011; Luiza et al. 2014). The computational model simply used the RGB color annotations for measuring the intensity and classification of the images into active and passive smokers' category. A series of homogeneous pixels with identical color intensity of RGB were screened out using GA from each respective region of interest and was used for the prediction of the level of expression of both ET-1 and CD-31 marker proteins (Schmitt et al. 1998; Schmitt 2001 and 2004). Clearly the level of expression of both the proteins was determined to be different for the active smokers' and passive smokers' classes (Luiza et al. 2014; Bastia et al. 2017). The intensity of RGB was distinctly different for both active and passive smokers which were sufficient to classify both types of images. This information was used to automatically classify the given IHC image to active endothelin-1, passive endothelin-1, active CD-31 and passive CD-31 based on neighboring network (Naik et al. 2007; Cha SH and Tappert CC 2009). The predictability for the classification of the images was found to be very accurate (90.76 %) for classification of images based on Endothelin-1 protein. The GA-NN algorithm had achieved an MCC of 0.7125 for this protein and the other performance measures were: Qpred = 82.85 ± 4.246%, sensitivity = 89.24 ± 2.45% and specificity = 83.43 ± 2.38%. The vast majority of the predictions were observed to be contained within 0.0 to 0.1 for the passive smokers and 0.9 to 1.0 for the active smokers in the case of endothelin-1 marker. This illustrated that 0.1 and 0.9 cut-off values provide very adequate separation of the two classes of images (Cha SH and Tappert CC 2009). Similarly, the prediction accuracy for the classification images into passive smokers and active smokers is determined to be 88.94% based on CD-31 marker. The rate/intensity of expression for ET-1 and CD-31 was more for active smokers as compared to passive smokers. It had revealed a wide difference in the range of intensity between both the classes of images and hence it was possible to classify the user given image into its corresponding class (Schmitt et al. 1988; Lösch A and Kainz C 1996).

Conclusion

The results of the present work demonstrated that the image derived features with GA-NN model appear to be a very fast image identification mechanism providing good results, comparable to some of the current efforts in the literature. The predictability and the classification of the images were found to be very accurate (90.76 ± 2.50 %). The prediction results for both the proteins, namely, Endothelin-1 and CD-31 were also observed to be quite accurate in case of the two groups, active and passive smoking mothers. The tool has achieved an MCC of 0.7125 (Endothelin-1). This tool had demonstrated a wide difference in the range of intensity between both the classes of images and hence it was possible to classify the input images into its corresponding class. The rate/intensity of expression of both the proteins was more for active smokers in comparison to passive smokers. Thus, it could be concluded that computer aided diagnosis of the immunohistochemical (IHC) images of tissue biopsies from different diseases will help the clinicians in classifying and grading the diseases according to their severity and in turn take appropriate therapeutic measures after proper diagnosis, in areas lacking experienced histopathologists.

Conflict of Interest

The authors declare that there are no conflicts of interest in this study.

Acknowledgments

This study received financial support from Indian Council of Medical Research (ICMR), New Delhi in the form of project grants. The authors gratefully acknowledge the assistance of the staff at the ICMR-National Institute of Pathology, Safdarjung Hospital Campus, New Delhi.

References

Akbari R, Ziarati K (2011) A multilevel evolutionary algorithm for optimizing numerical functions. Int J Indust Engg Compu 2(2):419–430

Bastia B, Kumar K, Kumar SN, Behera NR, Jain AK (2017) Expression of Endothelin-1 in Human Placenta of Active Smokers: An Immunohistochemical Study. International Journal of Research Studies in Biosciences (IJRSB) 5(12):34-40. http://dx.doi.org/10.20431/2349-0365.0512005

Cha SH, Tappert CC (2009) A Genetic Algorithm for Constructing Compact Binary Decision Trees. J Pattn Recog Res 4(1):1–13

Hsu SM, Raine L, Fanger H (1981) Use of avidin-biotin-peroxidase complex (ABC) in immunoperoxidase techniques: a comparison between ABC and unlabeled antibody (PAP) procedures. J Histochem Cytochem 29(4):577-580

Huang F, Zheng W, Liang Q, Yin T (2013) Diagnosis and treatment of placental site trophoblastic tumor. International J Clin and Experimental Pathology 6(7):1448-1451

Kozuka M, Ito T, Hirose S, Takahashi K, Hagiwara H (1989) Endothelin induces two types of contractions of rat uterus: phasic contractions by way of voltage-dependent calcium channels and developing contractions through a second type of calcium channels. Biochem Biophys Res Commun 159:317-323

Langston C, Kaplan C, Macpherson T, Manci E, Peevy K, Clark B, Murtagh C, Cox S, Glenn G (1997) Practice guideline for examination of the placenta: developed by the Placental Pathology Practice Guideline Development Task Force of the College ofAmerican Pathologists. Arch Pathol Lab Med 121(5):449-476

Lösch A, Kainz C (1996) Immunohisto- chemistry in the diagnosis of the gestational trophoblastic disease. Acta Obstetricia et Gynecologica Scandinavica 75:753–756. doi: 10.3109/00016349609065741

Luiza JW, Taylor SE, Gao FF, Edwards RP (2014) Placental site trophoblastic tumor: Immunohistochemistry algorithm key to diagnosis and review of literature. Gynecologic Oncology Case Reports 7:13-15. doi:10.1016/j.gynor.2013.11.001

Mastrogiannis DS, O'Brien WF, Krammer J, Benoit R (1991) Potential role of endothelin-1 in normal and hypertensive Pregnancies. Am J Obstet Gynecol 165:1711-1716

Naik PK, Mishra VS, Gupta M, Jaiswal K (2007) Prediction of enzymes and non-enzymes from protein sequences based on sequence derived features and PSSM matrix using artificial neural network. Bioinformation 2(3):107-112

Nova A, Sibai BM, Barton JR, Mercer BM, Mitchell MD (1991) Maternal plasma levels of endothelin is increased in pre-eclampsia. Am J Obstet Gynecol 165:724-727

Nuovo G (2006) The utility of immuno-histochemistry and in situ hybridization in placental pathology. Arch Pathol Lab Med 130(7):979-983

Sabra S, Gratacos E, Gomez Roig MD (2017) Smoking-Induced Changes in the Maternal Immune, Endocrine, and Metabolic Pathways and Their Impact on Fetal Growth: A Topical Review. Fetal Diagn Ther 41(4):241-250

Schiff E, Ben-Baruch G, Peleg E, Rosenthal T, Alcalay M, Devir M, Mashiach S (1992) Immuno-reactive circulating endothelin-1 in normal and hypertensive pregnancies. Am J Obstet Gynecol 166:624-628

Schmitt LM (2001) Theory of Genetic Algorithms. Theor Compu Sci 259:1–61

Schmitt LM (2004) Theory of Genetic Algorithms II: models for genetic operators over the string-tensor representation of populations and convergence to global optima for arbitrary fitness function under scaling. Theor Compu Sci 310:181–231

Schmitt LM, Nehaniv CL, Fujii RH (1998) Linear analysis of genetic algorithms. Theor Compu Sci 208:111–148

Whitley D (1994) A genetic algorithm tutorial. Stats and Compu 4(2):65–85

Wolff K, Nisell H, Modin A, Lundberg JM, Lunell NO, Lindblom B (1993) Contractile effect of endothelin 1 and endothelin 3 on myometrium and small intra-myometrial arteries of pregnant women at term. Gynecol Obstet Invest 36:166-171