

## Using Decision Trees for the Classification of Pregnancy Outcomes

Peer Waseem Ul Aslam, Sanjay Kumar , Muheet Ahmad Butt

**Abstract— Background and Objective:** Early pregnancy hormones play an important role in gestational outcome. Most abortions (miscarriages) occur in the first trimester, which require follow-up procedures and specific measures for the next pregnancy. Early detection of a pregnancy and its outcome enables us to assess optimal management for another healthy pregnancy.

**Methods:** We used decision trees to classify pregnancy hormones, which will detect the final outcome. Based on particular levels of hormones and a patient's previous history, we can handle the situation in advance. The hormone, human chorionic gonadotropin (HCG), has also been used with decision trees; it is a widely-used model for similar datasets.

**Results:** The groups into which a pregnancy might fall, based on HCG levels, is defined by Decision Trees. This data mining technique would support the concept of 'outcome generation,' which is based on previous data. For example, historic data from two hospitals could be used to classify various patients and their pregnancy's eventual outcome.

**Conclusions:** This multi-level classification should help with the diagnosis of many conditions, e.g., a normal pregnancy, an ectopic pregnancy, a molar pregnancy (hydatidiform mole), an intrauterine pregnancy, choriocarcinoma, and Down syndrome

**Index Terms—** Decision Tress, Human chorionic gonadotropin, Hormones, Expectant Management

### I. INTRODUCTION

Decision Tree is a technique used for classification, it is very simple and easily understandable. DTC involves some steps which are very important for the classification. The first step is to create the model, this involves knowing the class prior to the classification. Then we apply the decision tree model to another dataset so that we predict which class they belong. In other words, it is a simple classification process for the new dataset. Decision trees are used for various purposes, such as categorizing cells as malignant or benign based on the results, and classifying students based on their grades. They have better accuracy when compared to other models such as statistical, artificial neural networks and genetic models. The Decision tree model classification has high accuracy even when the size of the dataset increases.

Pregnancy Hormone:

During pregnancy, a hormone is produced which is known as human chorionic gonadotrophin (HCG). It is secreted by

trophoblastic cells of placenta, which nourishes the ovum soon after the conception and implantation in the uterine cavity. The levels of the HCG can be detected by a blood test after 11 days of conception. Soon after 12-14 days, these levels can be detected by a urine test. Typically, the HCG level doubles after every 42-72 hours. The level will reach its peak in the first 9-12 weeks of the pregnancy and decline gradually towards the last weeks of the pregnancy. During the course of pregnancy, the HCG level rises. A low level of HCG may result in a normal pregnancy. The Scale suggests that a lower level of HCG (<5 mIU/ml) is considered to be negative for pregnancy and a higher value (25 mIU/ml) is positive for pregnancy. The HCG hormone level is measured by a unit known as Milli-international units per milliliter (mIU/ml). We have to take multiple HCG values for most accurate diagnoses. The HCG levels at different trimester give a more accurate assessment of its being either eventful or uneventful. We will us the decision tree approach to classifying the HCG levels so that we can predict the outcome of the pregnancy. This prediction can also lead us to the diagnosis of choriocarcinoma, miscalculation of pregnancy dating, possible miscarriage or obligated ovum, ectopic pregnancy, molar pregnancy and multiple pregnancies. In males if this test is positive it can predict testicular cancer. The range of HCG for males is 0.5 mIU/ml. Combined with alpha-fetoprotein Beta HCG is an excellent marker for the monitoring of germ cell tumors.

### II. EXISTING WORK

The study shows a majority of pregnancies of unknown location are abnormal; many resolve spontaneously when managed expectantly. This paper suggests that the use of this logistic regression model in clinical decision making is likely to improve the management of pregnancies of unknown locations [S. Banerjee et al. 1999]. The study concluded that HCG level predicts the pregnancy outcome based on some variables like Age etc. Serum levels of HCG in samples taken 16 days after ovulation are powerful predictors of good or poor pregnancy outcomes [Gillian Homan et al. 1999]. In this Paper they have used three approaches- synchronization tree construction, partitioned tree construction, hybrid approach. The Processor performance has been taken into the consideration. The author suggested that the hybrid approach yields better performance on processors [Anurag Srivastava et al. 1999]. This study suggested that HCG has more predictive value than other biochemical markers for assessing pregnancy outcome (Honan et al. 300). They have included the assay of HCG 12 days after embryo transfer in their IVF protocol [P. Poikkeus et al. 2002]. This paper describes the design and implementation of MRDTL-2 an algorithm for learning decision tree classifiers from relational databases, which are based on the framework for multi-relational data mining originally [Anna Atramentov et al. 2003]. This Study can predict which pregnancies of unknown locations become failing PULs, IUPs and most importantly EPs based

Manuscript received Feb 04, 2017

Peer Waseem Ul Aslam, Department of Computer Science & Engineering, Jaipur National University, Jaipur (Rajasthan), India

Sanjay Kumar , Department of Computer Science & Engineering, Jaipur National University, Jaipur (Rajasthan), India

Muheet Ahmad Butt, Department of Computer Science, University of Kashmir, Kashmir (J&K), India

on the patients HCG ratio alone. It significantly outperforms current diagnostics criteria for the prediction of EPs. The variables used in this study are HCG levels on a 48-hour basis, progesterone level, transvaginal scans, Demographical, hormonal and ultra-sonographic characteristics of women with a PUL. They have actually studied the women's which had already been diagnosed with the pregnancy of unknown location. a) PULs: Pregnancies of Unknown Location b) IUPs: Intrauterine Pregnancy c) EPs: Ectopic Pregnancy [G. Condous et al. 2004]. Decision tree analysis may be used as a guide to estimate the probability of successful expectant management in individual cases. Initial serum beta-HCG level was the best predictor of the outcome of expectant management. These difference enabled a construction of a four-level decision tree to estimate the likelihood of successful expectant management. The aim of this study was to establish a combination of parameters that could be used to identify ectopic pregnancies undergoing spontaneous resolution and thus improve the selection of patients for expectant management. They have created a database, a statistical analysis on SPSS. The likelihood of the final outcome can also be easily conveyed to the patient [J. Elson et al. 2004]. The study concluded that the single value of HCG is not a powerful predictor of an ectopic pregnancy [G. Condous et al. 2005]. In this paper, the author used some of the data mining techniques. They have studied the rule induction, decision tree, artificial neural network (ANN). These techniques are applied to healthcare data especially patient data. They have classified the healthcare data based on diseases. Some of the diseases which they have studied with the data mining's are Diabetes Mellitus and Diabetes Insipidus. The previous data has been applied to train the data mining techniques so that they can predict accurately. Some of the variables used in this study is Age, Sex, Duration, Mass, Sugar Level and Microalbuminuria [Harleen Kaur et al. 2006]. This paper suggests only predictive model for miscarriage had a clinically relevant predictive value of 28%. Results together do not justify closer surveillance of chromosomally normal pregnancies with PAPP-A or f beta HCG levels below the fifth percentile [R. van Ravenswaaij et al. 2010]. In this study the author has classified preeclampsia. They have developed a software which identifies the preeclampsia in early stages. The model behind this software package is Bayesian Network. They have created three screens a) Input – for inputting the data like: Diabetes, Age, Family, Previous history of hypertension. B) The blood pressure: is monitored by the Bluetooth machine which is then synced with the android phone c) automatic analysis of urine test strips using the camera phone. After all these details have been entered the model then predicts whether preeclampsia would happen and what would be the level [Marc Spaanderman et al. 2010]. This paper gives us an idea of how to detect the breast cancer. They have applied Single Decision Tree, Boosted Decision Tree, Decision Tree Forest (SDT, BDT, and DTF). Early Diagnosis of breast cancer is an important role in reducing the mortality and improving the prognosis of this disease. The Performance of the proposed structure is evaluated in terms of sensitivity, specificity, accuracy and ROC. They have concluded that Decision Tree Forest provides satisfying results for classification [Ahmad Taher et al. 2012]. This paper gives us an idea how hospitals & clinicians can use their own

discharge data to create decision trees that produce highly transparent clinically relevant decision rules. The Decision tree helped the hospitals to predict the Readmission Rate of Patients [John P. Hilbert et al. 2014].

### III. OUR CONTRIBUTION

The decision tree helps us to determine the outcome of Pregnancy. The HCG level in non-pregnant women is always < 5.0 mIU/ml, < 9.6 mIU/ml in post-menopausal women and 0-5 in men. The HCG level in pregnant women becomes detectable or raises soon after conception (9-11 days) after LH surge even when she has not missed her periods, so HCG becomes the first hormone that detects pregnancy at earliest. There are different levels of HCG at different trimesters of pregnancy that can help us to diagnosis different diseases. In the first trimester, the HCG level rises drastically maintaining the doubling time that is 48-72 hours. It usually raises up to 18-20 Weeks of pregnancy. If Beta HCG level decreases and USG showing empty uterine cavity rises the probability of ectopic pregnancy. In this condition, the patient may complain of severe abdominal pain with a history of (h/o) missed periods or may be in a state of shock. That is (decreased BP or non-recordable BP + tachycardia + cold clammy skin). The second situation is when Beta HCG level rises up to 1000000 mIU/ml indicating exuberant trophoblastic growth increases the susceptibility of molar pregnancy or hydatidiform mole although the USG is the criterion standard for identifying both complete and partial mole (incomplete mole). Complete mole in pregnancy with edematous placental villi, hyperplasia of trophoblastic and lack of fetal vessels and fetal red cells are present within the mesenchyme of villi. The treatment for molar pregnancy is suction and evacuation (S&E). Beta HCG level shows a plateau or rise in patients being observed after diagnosis of a hydatidiform mole or after suction and evacuation confirms the diagnosis of GTN (Gestational Trophoblastic Neoplasia). It is a collective term for Gestational Trophoblastic diseases. That includes hydatidiform mole, choriocarcinoma, invasive mole (Chorioadenoma destruens), and placental site Trophoblastic Tumor (PSTT) staging of GNT (by the official international federation of Gynecology and obstetrics)

Stage:

1. Confined to the uterus.
2. Limited to genital organs
3. Lung metastasis
4. Other metastasis

Patients with nonmetastatis GTN - Rx (Treatment) is single Chemotherapy (methotrexate). In this condition, the HCG levels are monitored every week. Six weeks of maintenance chemotherapy is administered after normal HCG levels, after 3-4 normal S.HCG levels, the levels are observed once per month for 1 year. If HCG level raises again the methotrexate is changed to Actinomycin d.

In the second trimester, the HCG plays an important role in diagnosing Down's syndrome. The increase in HCG level + decreased serum alpha + Fetoprotein + unconjugated serum estradiol (UE 3) usually at (15-18 weeks) confirms the diagnosis of Down's syndrome

( ↑HCG + ↓S. AFP + ↓VE3 = Down's syndrome)

The ↑S.HCG + ↓S.ALFP + ↓VE3 + serum inhibin forms the Quad test which is usually done at 18-20 weeks to rule out any chromosomal abnormality in the fetus.

The major placental hormone HCG is also well known for its important role in third trimester of pregnancy. The complication of third trimester known as preeclampsia (mild and severe) can be easily detected and the outcome of the pregnancy can be assumed as well. The eclampsia is a complication of pregnancy in which the patient develops HTN (hypertension) greater than 140/90 mmHg at least twice and is associated with proteinuria (greater than 300 ml/dl). It is noticed that patients having preeclampsia have high HCG level then normotensive pregnancy patients. The HCG level in patients having severe preeclampsia is much higher than the patients having mild preeclampsia. The mild preeclampsia patients are managed conservatively with antihypertensive therapy whereas severe preeclampsia patients have much chance of going into eclampsia if associated with convulsions in which medical termination of pregnancy becomes the only choice.

The main action of HCG hormone towards the end of third trimester is that helps in the onset of labor. The HCG level decreases towards the end of pregnancy that allows the progesterone and estrogen to act on muscles of uterus thus directly helps in contraction of uterus.

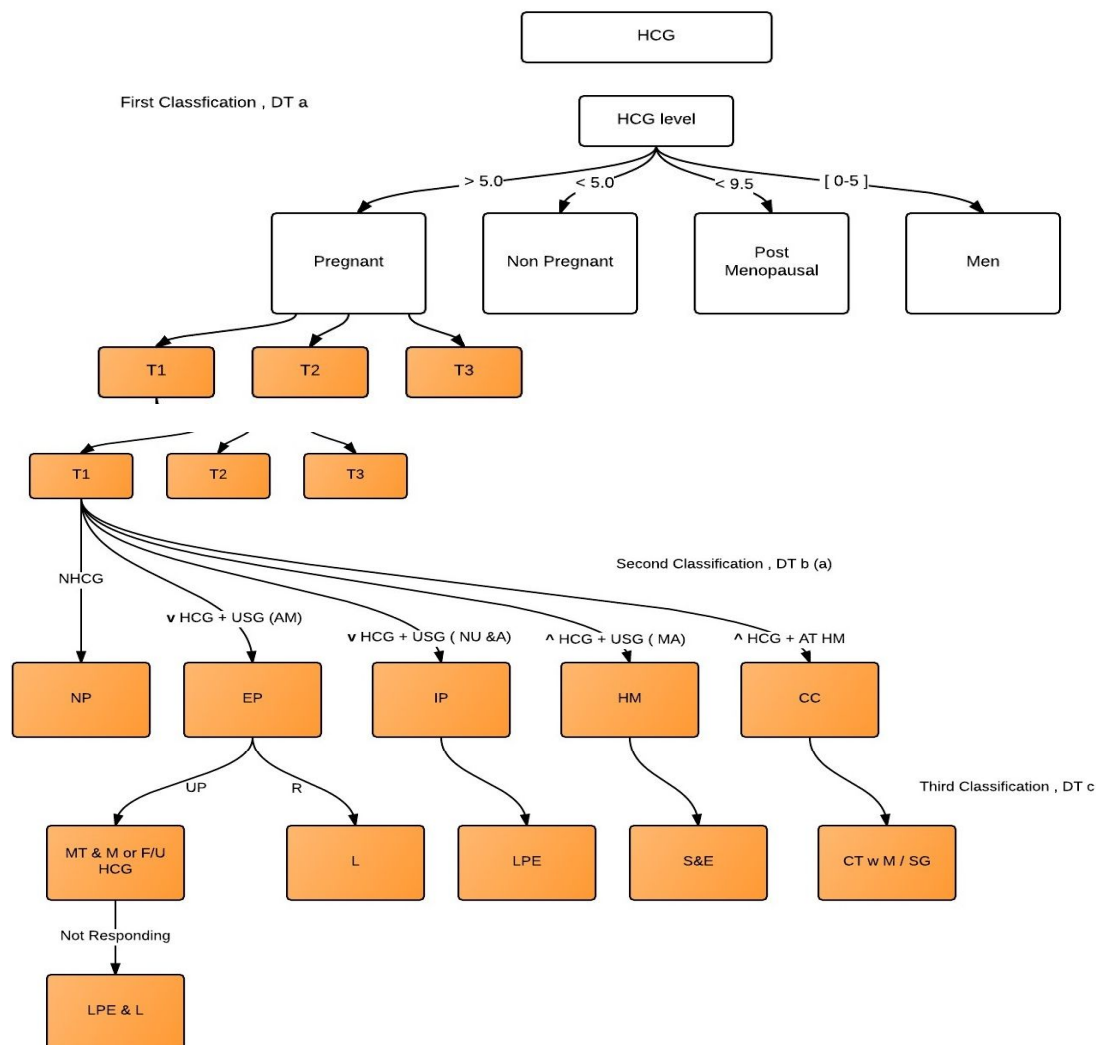
weeks since LMP	mIU/mL
3	5 – 50
4	5 – 426
5	18 – 7,340
6	1,080 – 56,500
7 – 8	7,650 – 229,000
9 – 12	25,700 – 288,000
13 – 16	13,300 – 254,000
17 – 24	4,060 – 165,400
25 – 40	3,640 – 117,000
Non-pregnant females	<5.0
Postmenopausal females	<9.5

Figure 1. HCG levels

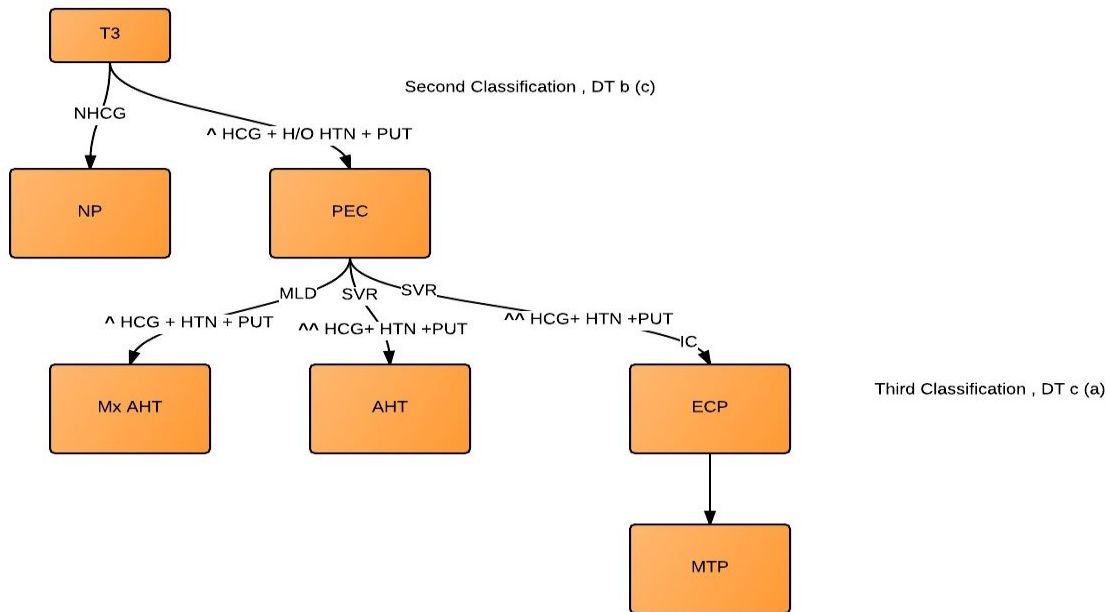
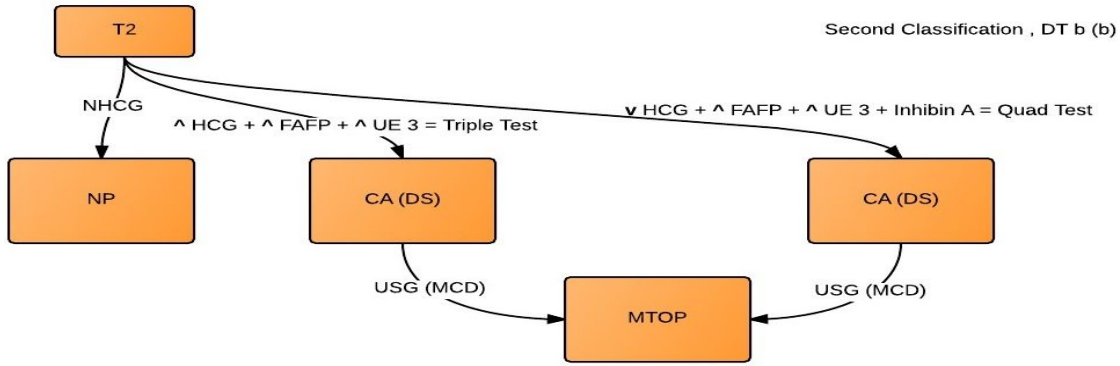
LMP: Last Menstrual Period

There are different levels of HCG at different trimester of pregnancy that can help us to diagnosis different conditions [Figure 1].

**Classification by Decision Tree**



## Using Decision Trees for the Classification of Pregnancy Outcomes



### Abbreviations used in the Decision tree.

**T1 : First Trimester**  
**T2 : Second Trimester**  
**T3 : Third Trimester**

**HCG : Human Chorionic Gonadotropin**  
**v HCG : Decrease in HCG**  
**^ HCG : Increase in HCG**  
**USG : Ultrasonography**  
**AM : Adnexal Mass**  
**NU & A : Empty Uterus and Adnexal**  
**MA : Molar Appearance**  
**S&E : Suction and Evacuation**

**NP : Normal Pregnancy**  
**EP : Ectopic Pregnancy**  
**IP : IntraAbdominal Pregnancy**  
**HM: Hydatidiform Mole**  
**CC: Choriocarcinoma**

**CT w M / SG : Chemotherapy with Methotrexate / Surgery**  
**LPE : Laparoscopic Evaluation**  
**LPE & L : Laparoscopic Evaluation & Laparotomy**  
**UP : Unruptured**  
**R: Ruptured**  
**MT & M or F/U : Methotrexate Therapy and Monitoring / Follow Up**  
**FAFP : Serum Alpha FetoProtein**  
**UE3 : Unconjugated E3**  
**CA (DS) : Chromosomal Anomalies ( Down's Syndrome )**  
**MTOP : Medical Termination of Pregnancy**  
**MCD : Multiple Conjunctival Defects**  
**PEC : Preeclampsia**  
**PUT: Proteinuria**  
**HTN: Hypertension**  
**AHT : Anti Hypertensive Therapy**  
**ECP: Eclampsia**  
**MTP : Medical Termination of Pregnancy**

Conditions and Values.

Condition	MSAFP	UE3	HCG
Tube Defect	Increase	N	N
Trisomy 21	Decrease	Decrease	Increase
Trisomy 18	Decrease	Decrease	Decrease
Molar Pregnancy	Decrease	Decrease	Very High
Multiple Gestation	Increase	N	Increase
Fetal Death (Still Birth)	Increase	Decrease	Decrease

#### CONCLUSION & FUTURE SCOPE

We have included three types of classifications in this decision tree, the first classification is based on the HCG level. This classification will be defined on the location, age and previous history of the patient. We can also predict the outcome of the pregnancy-specific to a region based on the previous data. An analysis of different location with respect to the pregnancy will help us understand the problem of that region- it might include poor health conditions, self-medications etc. The second classification is based on the outcome of the pregnancy. The high and low of the HCG level plus other tests can easily let us predict the outcome of the pregnancy. This would enable us to create follow-up procedure to maintain the health and well-being of these patients. This classification will also lead us to detect many cancers related to pregnancy and others as well- so managing the prediction of these diseases would lead to a successful model. In the last classification, we have defined a pathway to manage these conditions which might occur during the pregnancy. The future scope of this study is actually creating a broader predictive model which will encompass different scenarios plus a larger dataset could be examined. We can apply different predictive models to achieve better results. HCG in males also plays an important role to secrete testosterone so creating a predictive model for cancers detection in males.

#### REFERENCES

[1] Elson, J., Taylor, A., Banerjee, S., Salim, R., Hillaby, K., & Jurkovic, D. (2004). Expectant management of tubal ectopic pregnancy: prediction of successful outcome using decision tree analysis. *Ultrasound in Obstetrics & Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 23(6), 552–556. <http://doi.org/10.1002/uog.1061>

[2] Hilbert, J. P., Zasadil, S., Keyser, D. J., & Peele, P. B. (2014). Using Decision Trees to Manage Hospital Readmission Risk for Acute Myocardial Infarction, Heart Failure, and Pneumonia. *Applied Health Economics and Health Policy*, 12(6), 573–585. <http://doi.org/10.1007/s40258-014-0124-7>

[3] Tong, W., Hong, H., Fang, H., Xie, Q., & Perkins, R. (2003). Decision forest: Combining the predictions of multiple independent decision tree models. In *Journal of*

*Chemical Information and Computer Sciences* (Vol. 43, pp. 525–531). <http://doi.org/10.1021/ci020058s>

[4] Gehrke, J., Ramakrishnan, R., & Ganti, V. (1998). RainForest—a framework for fast decision tree construction of large datasets. 24rd International Conference on Very Large Data Bases, 416–427. Retrieved from <http://www.springerlink.com/index/G652567687177764.pdf>

[5] Azar, A. T., & El-Metwally, S. M. (2013). Decision tree classifiers for automated medical diagnosis. *Neural Computing and Applications*, 23(7-8), 2387–2403. <http://doi.org/10.1007/s00521-012-1196-7>

[6] Srivastava, a., Han, E.-H. S. H. E.-H. S., Singh, V., & Kumar, V. (1998). Parallel formulations of decision-tree classification algorithms. *Proceedings. 1998 International Conference on Parallel Processing (Cat. No.98EX205)*, 24, 1–24. <http://doi.org/10.1109/ICPP.1998.708491>

[7] Atramentov, A., Leiva, H., & Honavar, V. (2003). *Inductive Logic Programming*. (T. Horváth & A. Yamamoto, Eds.) *Lecture notes in computer science* (Vol. 2835). Berlin, Heidelberg: Springer Berlin Heidelberg. <http://doi.org/10.1007/b13700>

[8] Schumacher, A., Heinze, K., Witte, J., Poloski, E., Linzke, N., Woidacki, K., & Zenclussen, A. C. (2013). Human chorionic gonadotropin as a central regulator of pregnancy immune tolerance. *Journal of Immunology* (Baltimore, Md. : 1950), 190, 2650–8. <http://doi.org/10.4049/jimmunol.1202698>

[9] Seeber, B. E., & Barnhart, K. T. (2006). Suspected ectopic pregnancy. *Obstetrics and Gynecology*, 107(2 Pt 1), 399–413. <http://doi.org/10.1097/01.AOG.0000198632.15229.be>

[10] Banerjee, S., Aslam, N., Zosmer, N., Woelfer, B., & Jurkovic, D. (1999). The expectant management of women with early pregnancy of unknown location. *Ultrasound in Obstetrics & Gynecology : The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 14(4), 231–6. <http://doi.org/10.1046/j.1469-0705.1999.14040231.x>

[11] Homan, G., Brown, S., Moran, J., Homan, S., & Kerin, J. (2000). Human chorionic gonadotropin as a predictor of outcome in assisted reproductive technology pregnancies. *Fertility and Sterility*, 73(2), 270–274. [http://doi.org/10.1016/S0015-0282\(99\)00512-9](http://doi.org/10.1016/S0015-0282(99)00512-9)

[12] Esim, E., Turan, C., Unal, O., Dansuk, R., & Cengizglu, B. (2003). Diagnosis of premature rupture of membranes by identification of ??-HCG in vaginal washing fluid. *European Journal of Obstetrics Gynecology and Reproductive Biology*, 107(1), 37–40. [http://doi.org/10.1016/S0301-2115\(02\)00277-4](http://doi.org/10.1016/S0301-2115(02)00277-4)

[13] Poikkeus, P., Hiilesmaa, V., & Tiitinen, A. (2002). Serum HCG 12 days after embryo transfer in predicting pregnancy outcome. *Human Reproduction* (Oxford, England), 17(7), 1901–5. <http://doi.org/10.1093/humrep/17.7.1901>

[14] Tul, N., Pušenjak, S., Osredkar, J., Spencer, K., & Novak-Antolič, Ž. (2003). Predicting complications of pregnancy with first-trimester maternal serum free-βhCG, PAPP-A and inhibin-A. *Prenatal Diagnosis*, 23(12), 990–996. <http://doi.org/10.1002/pd.735>

[15] Zhu, Z., Wang, S., Xi, Q., Bai, J., Yu, X., & Liu, J. (2011). Logistic regression model for predicting stone-free rate after minimally invasive percutaneous nephrolithotomy. *Urology*, 78(1), 32–36. <http://doi.org/10.1016/j.urology.2010.10.034>

[16] Condous, G., Kirk, E., Lu, C., Van Huffel, S., Gevaert, O., De Moor, B., ... Bourne, T. (2005). Diagnostic accuracy of varying discriminatory zones for the prediction of ectopic pregnancy in women with a pregnancy of

- unknown location. *Ultrasound in Obstetrics and Gynecology*, 26(7), 770–775. <http://doi.org/10.1002/uog.2636>
- [17] Goetzinger, K. R., Singla, A., Gerkowicz, S., Dicke, J. M., Gray, D. L., & Odibo, A. O. (2009). The efficiency of first-trimester serum analytes and maternal characteristics in predicting fetal growth disorders. *American Journal of Obstetrics and Gynecology*, 201(4). <http://doi.org/10.1016/j.ajog.2009.07.016>
- [18] Van Ravenswaaij, R., Tesselaar-Van Der Goot, M., De Wolf, S., Van Leeuwen-Spruijt, M., Visser, G. H. A., & Schielen, P. C. J. I. (2011). First-trimester serum PAPP-A and  $\beta$ -hCG concentrations and other maternal characteristics to establish logistic regression-based predictive rules for adverse pregnancy outcome. *Prenatal Diagnosis*, 31(1), 50–57. <http://doi.org/10.1002/pd.2610>
- [19] Narayanan, R., Honbo, D., Memik, G., Choudhary, A., & Zambreno, J. (2007). An FPGA implementation of decision tree classification. In *Proceedings -Design, Automation and Test in Europe, DATE* (pp. 189–194). <http://doi.org/10.1109/DATE.2007.364589>
- [20] Kaur, H., & Krishan, S. (2006). Empirical study on applications of data mining techniques in healthcare. *Journal of Computer Science*, 2(2), 194–200. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.8981&rep=rep1&type=pdf>
- [21] Kirk, E., Papageorghiou, A. T., Condous, G., Tan, L., Bora, S., & Bourne, T. (2007). The diagnostic effectiveness of an initial transvaginal scan in detecting ectopic pregnancy. *Human Reproduction*, 22(11), 2824–2828. <http://doi.org/10.1093/humrep/dem283>
- [22] Chen, X., Wen, S. W., Fleming, N., Demissie, K., Rhoads, G. G., & Walker, M. (2007). Teenage pregnancy and adverse birth outcomes: a large population based retrospective cohort study. *International Journal of Epidemiology*, 36(2), 368–373. <http://doi.org/10.1093/ije/dyl284>