OG-Miner: an Intelligent Health Tool For Achieving Millennium Development Goals (MDGs) in m-Health Environments*

M. Jamal Afridi and Muddassar Farooq
Next Generation Intelligent Networks Research Center (nexGIN RC)
National University of Computer & Emerging Sciences (NUCES)
Islamabad, Pakistan
Email: {jamal.afridi, muddassar.farooq}@nexginrc.org

***Abstract*—The latest statistics of WHO show that approximately $500,000$ women die worldwide every year – the majority of them residing in developing countries – due to pregnancy related complications. The situation is so grave that UN has set a target of reducing Maternal Mortality Rate (MMR) by 75% till the year 2015 in its millennium development goals (MDGs). Therefore, the current focus of health care researchers is to advocate the use of e-health technology in developing countries that have the capability: (1) to remotely monitor patients in their homes by semiskilled health professionals, and (2) to use data mining techniques to raise alarms about high risk patients. In this paper, we develop an intelligent health tool – Obstetrics and Gynaecology (OG) OG-Miner – that presents a novel combination of data mining techniques for accurate and effective classification of high risk pregnant women. The scheme classifies four major risk factors of mortality – hypertension, hemorrhage, septicemia and obstructed labor – in a reliable, autonomous and accurate fashion. We have collected a real world data of more than 1200 patients from tertiary care hospitals and rural areas. Our tool achieves more than 98% accuracy on the collected OG dataset. Moreover, our evaluations of OG-Miner on eight other medical datasets show that its learning paradigm can be generalized to other domains as well. Last but not least we are using OG-Miner as an integral component of a health value chain in our m-health project to autonomously filter a significant number of low risk patients in rural areas; as a result, only high risk patients are referred to specialized gynecologist in tertiary care hospitals. As a consequence, the reduced workload enables them to provide quality care to the patients.**

I. INTRODUCTION

The latest statistics of WHO show that approximately $500,000$ women die worldwide every year – the majority of them residing in developing countries – due to pregnancy related complications [1]. The situation is so grave that UN has set a target of reducing Maternal Mortality Rate (MMR) by 75% till the year 2015 in its millennium development goals (MDGs) [1]. The major causes that contribute to MMR – especially in resource constrained African and Asian countries – are Hemorrhage, Hypertension, Obstructed Labor and Septicemia. It is disappointing to note that these causes of mother mortality could have been easily removed through effective and efficient preventive countermeasures. But, this demands the need of timely antenatal visits to a trained medical expert. Unfortunately, in remote rural areas, a specialized medical expert is generally not available and thus the people suffer. Also, in the urban areas – due to very low doctor to population ratio – the medical experts remain overwhelmed with the workload; therefore, it becomes a challenging task for them to provide quality care to the patients.

To meet the challenge of saving lives of mothers and new born, we provide an integrated e-health solution – using existing mobile IT infrastructure and intelligent health tools – that provides antenatal services at the door step of women residing in rural areas of our country[1]. We build the capacity of semiskilled lady health workers (LHWs) to integrate them into the work-flow of our project (see Figure 1). The basic idea is that a LHW electronically enters her antenatal visit reports on the mobile phones and then sends it to a server, where it is stored in an Electronic Medical Record (EMR) system, housed in a tertiary care hospital. The server then invokes Clinical Decision Support System (CDSS) which is based on an intelligent health tool (OG-Miner). The tool automatically analyzes the patient's electronic record and classifies her as either normal or high risk. The motivation of using intelligent health tool is to assist semiskilled LHWs in accurate and timely referrals. As a result, it acts as a filter and reduces the workload of medical consultants who now only mange high risk patients. Our tool is of great value for developing countries where doctor to population ratio is very low. The results of our experiments show that our tool achieves more than 98% accuracy[2] on an OG medical dataset of high risk patients. OG-Miner (with such a high accuracy) will become an important component of IT based health environments; as a result, it will play a critical role in achieving MDGs through

* This work is supported by National ICT R & D Fund, Ministry of IT, Government of Pakistan.

[1] see http://rpms.nexginrc.org for an overview of our project

[2] Throughout this text, the terms *detection accuracy* and *Area Under ROC Curve (AUC)* are used interchangeably. The AUC ($0 \leq \text{AUC} \leq 1$) is used as a yardstick to determine the detection accuracy from ROC curve. In this paper we report AUC in %

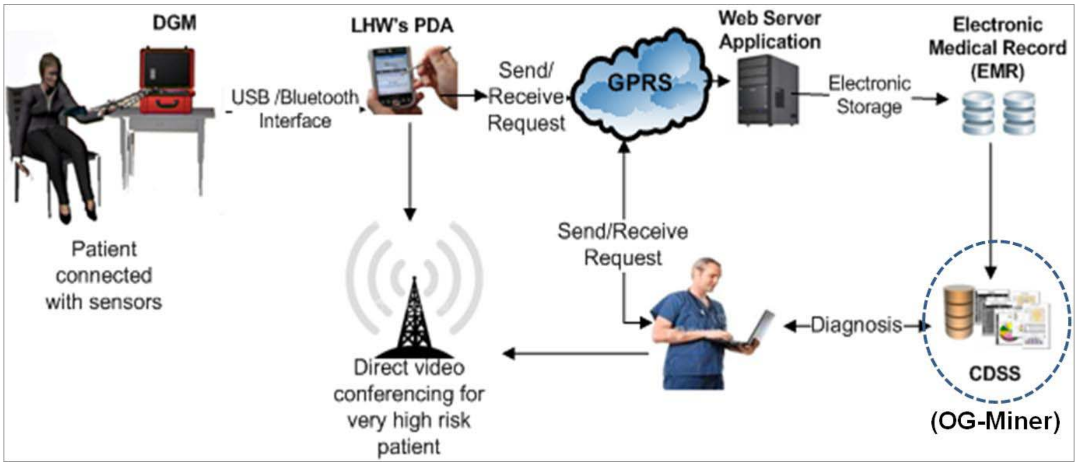


Fig. 1. Work flow of information in our project of Remote Patient Monitoring System

the reduction of MMR in developing countries.

We follow a two step engineering approach to develop our tool. In the first step, we collect real world OG data of 1200 patients from two tertiary care hospitals and rural areas. We visualize and analyze the dataset to find trends and patterns that help us in doing requirements engineering of an accurate and reliable back end classifier. We describe collection of OG data in Section III and analysis of its features' set in Section IV. In the second step, we first intuitively argue about a possible architecture of OG-Miner that could meet our requirements. We then follow a series of pilot studies to select a best features' selection scheme and a back end classifier. This step is discussed from Section V to Section VI. In Section VII, we study the ability of OG-Miner to generalize its knowledge to other medical domains. (We specially focus on 8 challenging datasets of UCI). Finally, we provide the feedback – collected through live interaction with LHWs and medical experts – about our intelligent health tool once it is deployed in a real world m-health environment. Subsequently, we conclude the paper with an outlook to our future research.

II. Related Work

To the best of our knowledge, no system like OG-Miner exists that uses data-mining techniques to determine the risk factors, which could lead to the death of a pregnant woman, on the basis of the knowledge extracted from her electronic health record. It is important to have such a system because of the reasons mentioned in the previous section. Pregnancy related complications show a specific pattern which otherwise might be considered normal. For example a blood pressure of $130/90$ is an alarm of high risk for a pregnant women but for other women it might be a routine increase. Similarly, Uterus needs large blood flow in pregnant women and vaginal bleeding (in case of abnormality) can result in hemorrhage. To conclude, we should have a dataset for pregnancy related complications because the methodology of risk assessment for pregnant women is totally different compared with the normal ones.

The work reported in [2],[3],[4],[5] does apply data mining to hypertension datasets but the dataset is not focusing on pregnant women. The only known study on pregnant women is detecting pre-eclampsia – a complication resulting because of hypertension – by identifying metabolic patterns in patients' plasma [6]. Their dataset is specific to this problem and hence cannot be used for our purpose. Goodwin et al. [7] conducted research on $19,970$ patients of pregnant women with an aim to predict – considering demographic factors – pre-term births. The idea of using an expert system for the risk assessment of pre-term births is the focus of Woolery et al. [8]. To conclude, OG-Miner is the first tool that targets the high risk factors of maternal mortality with the use of data mining and machine learning techniques.

The next section presents the Step 1 of the OG-Miner's development process by explaining the collection and development of OG dataset.

III. Dataset of High Risk Patients

As mentioned before, no dataset of high risk pregnant women exists; therefore, we need to collect such a dataset before we can utilize data mining techniques. We signed Memorandum of Understanding (MoU) with two tertiary hospitals that mandates us to develop a web based Electronic Medical Record (EMR) system for their OG departments and they provide us with medical records after removing the identity of patients in order to ensure privacy of the patients [3]. The EMR has been developed by keeping OG consultants in a feedback loop. We also trained doctors in the senior year of residency to enter records using our EMR. Since Jan 2009, our EMR is deployed in the inpatient wards of OG departments in both hospitals. During the one year period, both hospitals admitted more than 1200 pregnant women. The aim is to develop datasets that can be used by an effective classifier for detecting the high risk anomalies in OG domain. In order to make the paper self contained, we briefly define each anomaly:

[3]We have publicly released the datasets that could be downloaded from http://www.nexginrc.org

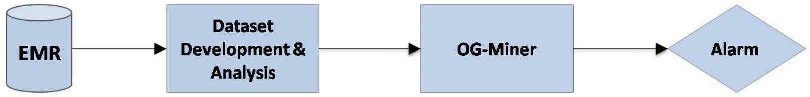


Fig. 2. A Two step Engineering Approach for Developing OG-Miner

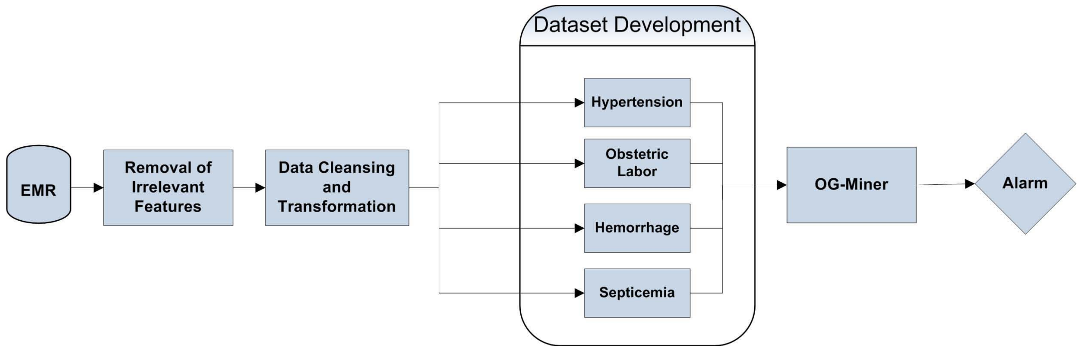


Fig. 3. Top Level Architecture of OG-Miner

TABLE I
MAIN BRANCHES OF THE SELECTED FEATURES

Main Information Areas of the Selected Features				
Patient Information	Routine Investigation	Present Pregnancy	General Examination	Contraception History
Vaginal Examination	Previous Pregnancies	Per Speculum Inspection	Conditional Follow-Up	Abdominal Examination
Specific Investigation	Presenting Complaints History	Gynecological History	Drug History	Family History

- **Hypertension (HYP).** This is because of an increase in the blood pressure, which not only hinders the growth of the baby but also puts a mother's life at risk as well.
- **Hemorrhage (HMG).** This is because of excessive blood loss from the body. Vaginal bleeding is one of the main reasons for hemorrhage in pregnancy.
- **Obstructed Labor (Ob-L).** This is an abnormality that arises during the process of labor.
- **Septicemia (SEPT).** This is an infection which is caused by different types of bacteria that contaminates a patient's blood.

Now we follow a three step methodology to develop OG dataset that can be readily used by our intelligent tool. The steps are: (1) removing least relevant features with the help of gynecologists, (2) applying data cleansing and transformation techniques, and (3) conversion into ARFF format to make it compliant with Wakaito Environment for Knowledge Acquisition (WEKA) tool [9]. The remainder of this section describes each of these steps separately.

Removing Least Relevant Features. Our EMR contains more than 600 attributes. In order to reduce the dimensionality of the features' space, we analyzed raw health records with the help of gynecologists and manually removed 395 attributes. As a result, we are now left with 205 relevant features that consultants use to detect anomalies in pregnant women.

Data cleansing and Transformation. This step is very important for development of real world medical datasets. The raw data collected from EMR contained a number of fields that require text entry by gynecologists. It became a challenge to extract useful information from these fields. For example, in "delivery mode" attributes different doctors used different abbreviations that meant the same thing – Simple Vaginal Delivery (SVD) and Normal Vaginal Delivery (NVD). Sometimes the gynecologists even did not enter correct units in the numerical fields. In order to have a meaningful dataset, we applied data cleansing and transformation cycle (see Figure 3).

Dataset Development and conversion in ARFF Format. Once we have the meaningful attributes, we need to develop separate datasets for hypertension, hemorrhage, obstructed labor and septicemia on the basis of short listed 205 attributes. To get a better insight about each risk factor of maternal mortality, we made separate datasets for each risk factor. Table II shows the basic statistics of these datasets.

In the next section we discuss another important activity of Step 1: analyze the trends and patterns in obtained datasets.

TABLE II
BASIC PROPERTIES OF OUR DATASETS

Measures	HYP	Ob-L	HMG	SEPT
Total instances	2734	2738	2299	2328
Diseased instances	107	83	34	26
total attributes	205	205	205	205

IV. FEATURES OF OG DATA

OG datasets are based upon a diverse set of features that covers different areas as shown in Table I. The analysis of the distributions of the remaining 205 features in OG data shows that most of the features have low potential for discriminating high risk patients from the normal ones. Figure 4 shows the distribution of blood flow values for normal and hemorrhage patients. It is clear from the figure that all 34 hemorrhage patients do not have a “Heavy” blood flow. However, we also see that more than two thousand normal women also have the same attribute value. As a result, it is difficult to classify a hemorrhage patient on such a diffused features’ set. Similarly, in Figure 6, we see that in 26 out of 34 hemorrhage cases, the pregnant women belong to the illiterate category. But again, this information could not create a discriminating potential for hemorrhage patients as we see that more than 800 normal pregnant women are also illiterate. On a similar note, Figure 5 shows the distribution of blood group values for hypertensive and normal patients. As expected, we could not correlate hypertension with the blood group. Figures 7 and 8 further elaborate that OG dataset still contain many ineffective features hindering accurate classification. Many of these features are even redundant and therefore accurate classification on such datasets is a challenge.

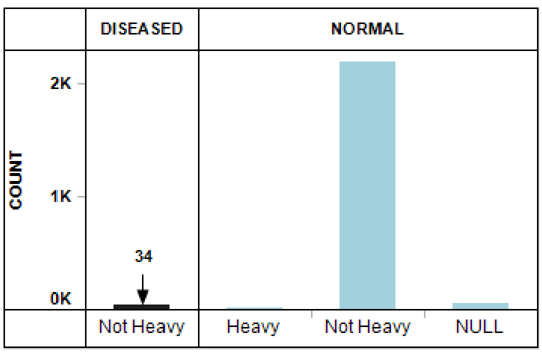


Fig. 4. Distribution for Blood Flow

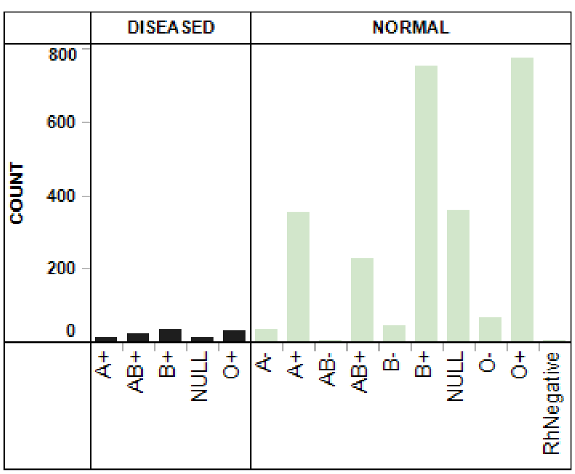


Fig. 5. Distribution for Blood Group Values

It is evident that in order to develop accurate classification model, we should focus on our system’s ability to select relevant features. This is possible if we extract effective OG subsets for each risk factor and then remove redundant and ineffective features from the dataset. Therefore, it is imperative to incorporate an effective feature selection scheme in our design. In the following section, we start with the Step 2 and will explain the design process of our classification tool.

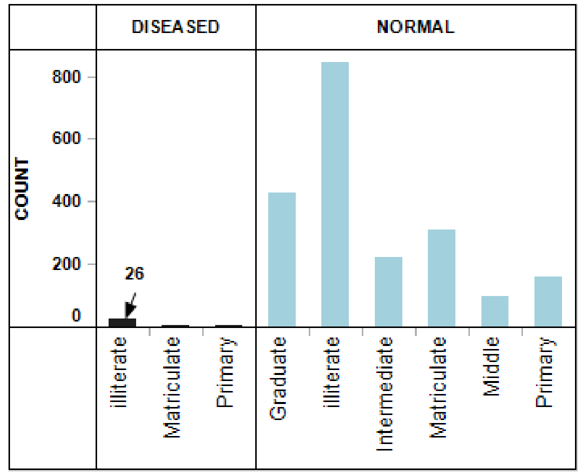


Fig. 6. Distribution for the wife education level

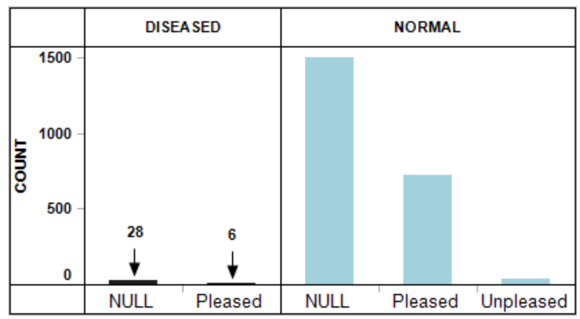


Fig. 7. Distribution for pregnancy reaction

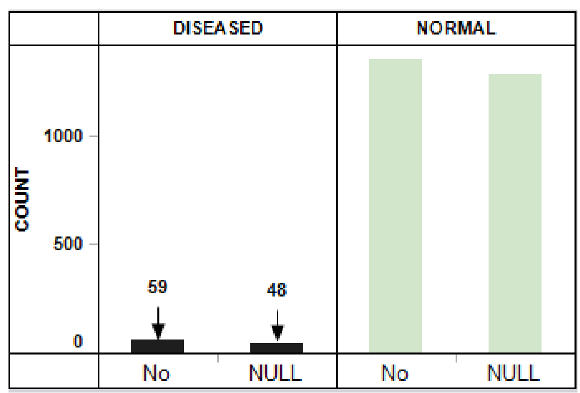


Fig. 8. Distribution for the history of Heart surgery

V. Designing the Architecture of OG-Miner

In this section, we first do a qualitative comparison and analysis of our OG data to intuitively contemplate on an expected architecture of OG-Miner. This work is mainly based on novel insights of the OG data and will be useful in exploring the design space. To finalize this architecture, we adopt an experimental approach that investigates different options. Consequently the architecture of OG-Miner is finalized.

A. Initiating the Design.

The domain of Obstetrics and Gynecology presents an imbalanced and high dimensional hierarchical data. The aim of designing the architecture is to find the best combination of data mining techniques for accurate classification of OG risk factors. OG datasets, even after manual feature removal, still contains 205 attributes (a number of them are still not useful in classification as shown in the previous section). So, it is relevant that our OG-Miner tool should: (1) use a feature selection scheme that automatically extracts the most relevant subset of features from the data and removes the redundant features, and (2) an effective and accurate *classification scheme* that gives reliable and consistent results on OG datasets.

Feature Selection Scheme. The job of feature selection is to remove least relevant and redundant features. In order to find the most relevant features, it is important to find the predictive ability of each feature and rank them. As a result, only top ranking features are selected and used in the data subset. Chi-square and entropy based measures – Information Gain, Gain Ratio – determine predictive ability of features but they do not guarantee that redundant features will not coexist. For medical datasets, it is imperative to have a features' set that covers the complete knowledge of the dataset. Therefore, we also evaluate correlation based feature selection schemes that have the ability to remove redundant features. Moreover, they also select features with high predictive ability. The final scheme will be selected after empirical evaluations.

Expected Classification Scheme. A classification algorithm relates and finds the similarity of a test instance to the previously known instances or to the knowledge learned on the previous known data. In case of imbalanced data like OG datasets, the knowledge learned by the most classifiers shows a bias towards the major class which affects the overall accuracy. We compare the imbalanced ratio of our dataset to the well known UCI medical datsets. Using the definition in [10], we have computed imbalance ratios of UCI medical datasets. The top 3 imbalance datasets are: (1) Hyper-thyroid (Hp-T) dataset with an imbalance ratio of (28.81), (2) Hypo-thyroid (Ho-T) with an imbalance ratio of 9.99, and (3) Ann-thyroid(An-T) with an imbalance ratio of 8.37. In comparison, the imbalance ratio of SEPT, HMG, Ob-L and HYP datasets in our OG dataset are 44.27, 33.1, 16 and 12.3 respectively (see Figure 9). It is logical to have imbalanced OG datasets because the number of patients suffering from high risk factors are significantly smaller compared with normal pregnant women who visit hospitals for antenatal examinations.

The algorithms that use only local data distribution in the decision process are more relevant in the case of imbalance datasets. Instance based learning *k*-NN[4][11] is a well known technique that maps an instance into an N-dimensional space, where 'N' is the total number of attributes. IB*k* considers – depending on its location in space – K nearest neighbors of this test instance for voting. The neighbors might belong to different classes and each neighbor casts its vote in favor of its own class. The voting process is weighted and its weight depends on the distance of a neighbor from the test instance. After the weighted voting, the class of a test instance is finally decided. Therefore, we conclude that IB*k* classifies an instance on the basis of local information about the neighbors of an instance without looking at the global distribution of attributes. Due to this behavior, it is expected to provide good results on the imbalanced OG datasets. The algorithm will be unaware of the global imbalance distribution and after mapping the test instance it will focus only on the location based similarity with its neighbors.

Such individual algorithms, however, might fail to correctly classify a test instance that is located in a region where the class noise is significant[12]. Class noise measures the

[4]Throughout this text, the terms IB*k* and *k*-NN are used interchangeably.

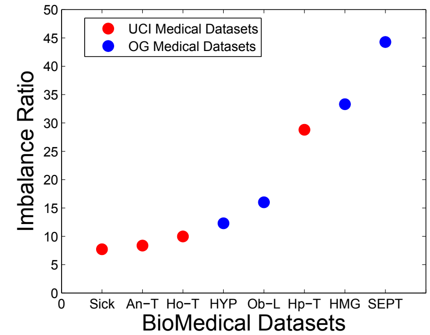


Fig. 9. Imbalance Ratio for the high risk factors

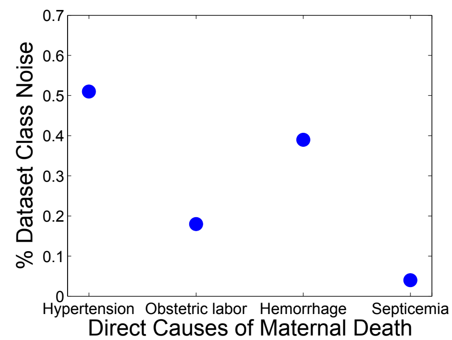


Fig. 10. Data class Noise for the high risk factors

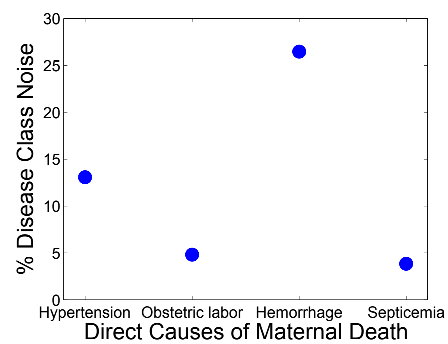


Fig. 11. Disease Class Noise for the high risk factors

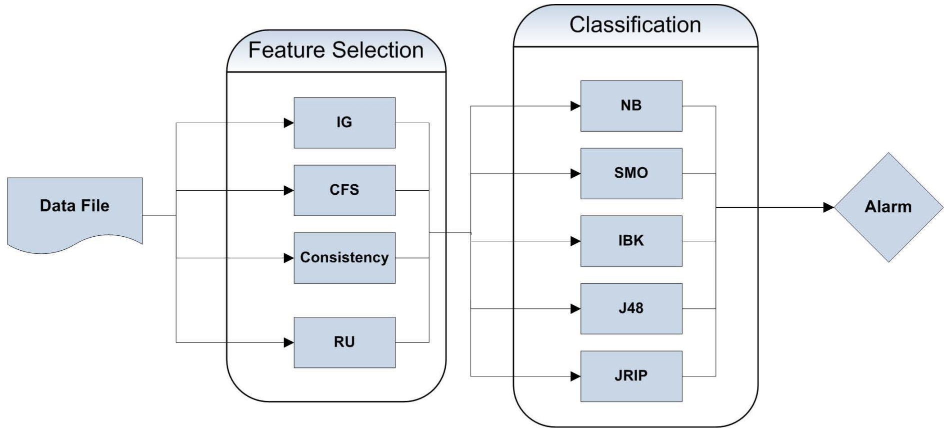


Fig. 12. Experimental framework of OG-miner

number of instances that belong to different classes but are located at the common region in space. (Class noise is directly proportional to the area of confusion region.) The accuracy of a classifier decreases with an increase in the value of class noise for a dataset. We plot class noise for OG datasets in Figure 10. We can see that HYP dataset has the maximum class noise of 0.51%; while SEPT has the smallest class noise of 0.04%. Unlike imbalance ratio, class noise values are comparable with the class noise of UCI medical datasets.

We argue that for two class medical datasets that have high imbalance ratio, it is relatively more important to identify disease instances (compared with normal ones) that confuse a classifier. We now define "disease class noise" as the ratio of misclassified disease instances to the total number of disease instances. Figure 11 shows that hemorrhage has the highest diseased class noise. Note that the hemorrhage dataset also has the second highest imbalance ratio as well; therefore, it becomes a challenge to correctly classify this dataset. As mentioned before, hemorrhage is the most prevalent cause of maternal death; therefore, it is important to accurately detect it. In this case, IB*k* classifier might fail because it cannot accurately classify a test instance that belongs to the confusion region. Also in all other OG datasets, when a test instance belongs to the confusion region, there will always be a chance of its misclassification. Rather than completely depending upon the instance based learning (IB*k*), the classification scheme will also need the use of another classification approach in order to view the test instance from multiple dimensions in its decision making process. Therefore, we believe that a classification scheme that combines the use of instance based learning with some other accurate classification method, will produce more reliable results. To conclude, our intelligent tool is expected to use correlation based feature selection scheme and a hybrid classifier combining multiple classification approaches. The following section explains the used experimental framework for finalizing our OG-Miner tool.

B. *Finalizing Architecture of OG-Miner*

We use empirical based design exploration approach to evolve architecture of OG-Miner. The idea is to evaluate representative candidates from each design paradigm, evaluate them on OG dataset and select the best candidate for our tool. The framework of our experimental study is shown in Figure 12.

1) Feature Selection Module: The objective of our pilot studies in this module is to evaluate merits/demerits of correlation based feature selection schemes – CFS – with other well known techniques. In our study, we use four well known feature selection schemes – Information Gain (IG), Correlation based Feature Subset Evaluator (CFS), Consistency based Feature subset evaluator, and Remove useless (RU) – to select the best scheme for our OG-Miner tool. Remember, CFS and consistency based feature subset selection use a search method for finding subsets. To remove the bias of search method we used genetic search for both.

2) Classification Module: The objective of experiments here is to find the effectiveness of the use of IB*k* classifier for OG domain and investigate the hybrid architecture of our classifier. To remove any bias in our study, we use a well known diverse group of classifiers containing probabilistic (Naive bayes), functional (Sequential Minimal Optimization), Instance based Learner(IB*k*), Decision Tree (J48) and Rule Learning (Jrip) algorithms. An interested reader can see the details of these paradigms in [13]. Now an experimental comparison of these techniques is presented in the following section.

VI. Experiments, Results and Discussions

We initially perform 120 experiments using a total of six feature selection schemes as shown in Table III. After this step, we get six reduced subsets from each anomaly dataset in OG. (Remember our OG dataset consists of four risk factors – hemorrhage, hypertension, obstructed labor and septicemia.) Each of these subsets are then given to above-mentioned five classifiers. We have used the standard 10 fold cross validation process in all these experiments. This process allows the classifiers to use 9 folds of data for training and 1 unused fold, containing previously unknown data, for testing. This is repeated 10 times for every combination. In all experiments we use percentage average AUC of class values to define accuracy. The results of our experiments are tabulated in Table III. Now, we use this experimental study to finalize the feature selection and classification module of OG-Miner.

A. Finalizing the Classification Module

To compare the performance of IB*k* classifier and hence to finalize the architecture of this module, it is imperative to evaluate the accuracy of classifiers on each OG dataset. The accuracy of a classifier for the all the subsets (obtained from 6 feature selection approaches) of each risk factor are tabulated in table IV.

TABLE IV
PERFORMANCE OF CLASSIFIERS FOR EACH OG-RISK FACTOR

Classifiers	Risk Factors				
	HYP	Ob-L	HMG	SEPT	Mean
NB	93.9	97.8	**94.6**	96.7	**95.75**
SMO	92.6	97.2	79.9	98.2	91.975
IBK	**97.7**	**98.6**	88.4	**98.3**	**95.75**
J48	84.6	95.7	72.6	96.6	87.375
JRIP	89.5	95.9	83.6	96.3	91.325

This study supports our hypothesis in section V-A about the use of IB*k* on imbalanced OG datasets. IB*k* achieves the best accuracy on three OG datasets. However, its accuracy on Hemorrhage dataset is not the best because of its sensitivity to high disease class noise. Naive bayes on the other hand uses a probabilistic approach and produces the best accuracy on hemorrhage data.

Probabilistic techniques such as Naive bayes use the global model of data to predict the class of an instance. The mean accuracy of Naive bayes is equal to that of IB*k* classifier in Table IV. This justifies the use of Naive bayes in combination with IB*k*. To develop a hybrid classifier, we perform Voting at the meta-level that uses the average probabilities as a combination rule. Now it is necessary to compare the experimental results of this hybrid classification technique with the two best classifiers of OG datasets (namely Naive Bayes and IB*k*). We present the accuracy of *Voting* approach on four OG datasets in Table V. It is clear from the table that the hybrid classifier improves the average accuracy by 2% on all datasets. So, for a dataset with 2000 patients, 2% increase would mean the correct classification of 40 patients and for larger data the significance of this 2% increase becomes even more prominent.

TABLE V
COMPARISON OF HYBRID CLASSIFIER FOR EACH RISK FACTOR

Classifier	**Risk Factors**				
	HYP	Ob-L	HMG	SEPT	Mean
Hybrid	97.7	98.6	95.6	98.2	**97.525**
NB	93.9	97.8	94.5	96.7	95.725
IB*k*	97.7	98.6	88.4	98.3	95.75

B. Finalizing the Feature Selection Module

In this section we present a comparative study to evaluate the performance of CFS compared with other feature selection schemes using the below criteria:

Comparison of mean Accuracy of classifiers. Each feature selection scheme reduces the number of features in the features' set and then produces subsets for four anomaly types – hemorrhage, hypertension, obstructed labor, and septicemia. The subsets are then given as input to each classifier. Each data subset produced by the feature selection scheme has a different classification potential because each scheme selects different attributes. We tabulate mean accuracy of all classifiers using different selection schemes in Table VI. The table clearly shows that IG_{100} and CFS provide the best accuracies.

TABLE VI
MEAN %AUC OF CLASSIFIERS FOR FEATURE SELECTION SCHEMES

Schemes	**Mean AUC of classifiers**
IG_{100}	93.2
IG_{50}	92.34
IG_{25}	91.98
Remove useless	92.28
CFS	93.1
consistency	91.54

Fitness with the hybrid Classifier. It is imperative to determine the fitness of feature selection scheme with our hybrid classifier. The results are tabulated in Table VII. It is clear that CFS produces the best accuracy of 98.62 with our hybrid classifier.

TABLE VII
ACCURACY OF HYBRID CLASSIFIER WITH DIFFERENT FEATURE SELECTION SCHEMES

SCHEME	RISK FACTORS				Mean
	HYP	Ob-L	HMG	SEPT	Accuracy
IG_{100}	97.8	98.6	95.7	97.3	97.35
IG_{50}	98.2	98.7	95	98.4	97.58
IG_{25}	97.9	99	94.5	98.7	97.53
CFS	98.3	98.8	98.7	98.7	**98.62**
Consistency	97.5	98.9	93.4	98.2	97
Remove Useless	97.7	97.8	95.9	98.2	97.4

We see in Table VII that CFS provides the best fitness with our hybrid classification scheme. Moreover, the results in Table VI also advocate the use of CFS. Moreover, unlike Information Gain, CFS also considers the correlation between the features along with their predictive ability. A well known disadvantage of an IG based features subset is that it might miss a relevant feature because of its low value of Information Gain. On the other hand, it might select redundant features which might not be able to cover the complete knowledge in the dataset.

Towards the end of this section, we would now present the final architecture of our intelligent decision support system tool in Figure 13. OG-Miner is a novel health tool because it combines the use of CFS based feature selection technique with a hybrid classifier that uses the voting technique to combine Naive Bayes classifier with an instance based learning IB*k*. Table VII shows that OG-Miner achieves the best mean

TABLE III
PERFORMANCE OF CLASSIFIERS ON THE SUBSETS OF DIFFERENT FEATURE SELECTION SCHEMES

Algorithms	Obs and Gyne Datasets				Mean accuracy
	Hypertension	**Obs-labor**	**Hemorrhage**	**septicemia**	
selected Top 100 features based on Information Gain (IG_{100})					
NB	96.2	97.8	97.9	96.2	**97.02±0.95**
SMO	93.3	97.5	85.2	98.1	93.52±5.95
IBK	97.5	98.8	88	97.3	95.40±4.98
J48	81.9	97.7	73.8	96.6	87.50±11.63
JRIPPER	90.4	97.1	86.2	96.3	92.50±5.15
selected Top 50 features based on Information Gain (IG_{50})					
NB	97.1	97.1	97.8	97.7	**97.42±0.377**
SMO	93.8	97.5	85.2	98.1	93.65±5.94
IBK	97.2	97.2	90.1	98.4	95.72±3.79
J48	88.1	88.1	70.6	96.6	85.85±10.93
JRIPPER	89.7	89.7	80.4	96.3	89.02±6.54
selected Top 25 features based on Information Gain (IG_{25})					
NB	97.5	98.7	97.2	97.8	**97.8±0.65**
SMO	93.3	97.5	50	98.1	84.72±23.25
IBK	97.6	98.8	89.1	98.8	96.10±4.68
J48	86.7	96.3	73.7	96.6	88.32±10.78
JRIPPER	89.7	97.1	88.8	96.3	92.98±4.33
Correlation based Feature Selection					
NB	94.5	97.9	95.1	96.3	95.95±1.5
SMO	94.3	97.5	86.6	98.1	94.12±5.29
IBK	98.4	98.9	89	98.8	**96.28±4.85**
J48	89.8	97.5	67.4	96.6	87.82±14.04
JRIPPER	88.8	97.1	83.2	96.3	91.35±6.60
Consistency based Feature selection					
NB	89.9	97.9	86.5	96.8	92.78±5.48
SMO	87.2	95.8	80.8	98.1	90.48±7.97
IBK	98.7	99.1	85.9	98.9	**95.65±6.50**
J48	80.9	96.8	81.2	96.5	88.85±9.00
JRIPPER	86.4	96.5	80.6	96.3	89.95±7.81
Remove Useless					
NB	87.9	97.6	91.8	96.2	93.38±4.41
SMO	92.8	97.5	85.2	98.8	93.58±6.15
IBK	98.3	98.8	89.8	98.2	**96.28±4.32**
J48	81.3	96.7	70.9	96.6	86.38±12.60
JRIPPER	91.2	97	82.8	96.3	91.82±6.55
Mean Accuracy	91.68±5.22	96.98±2.35	83.69±10.11	**97.27±0.98**	

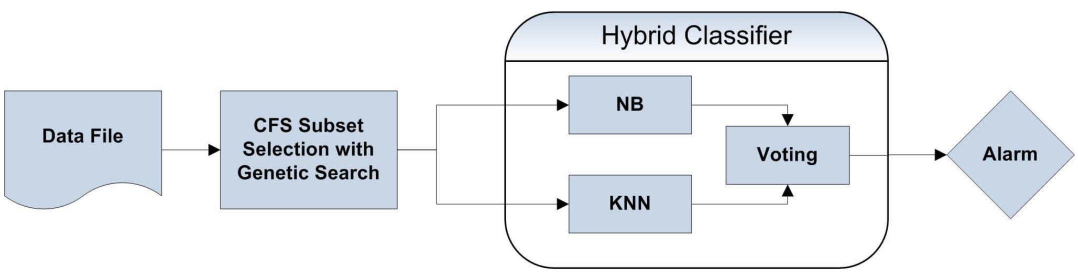


Fig. 13. Final Architecture of OG-miner

accuracy of 98.62% on OG datasets. In order to analyze its ability to generalize learning, we now evaluate OG-Miner on other medical datasets.

VII. EVALUATION OF OG-MINER IN OTHER MEDICAL DOMAINS

OG-Miner can generalize its knowledge extraction process to other medical datasets. To substantiate it, we have also evaluated it on 8 medical datasets obtained from UCI repository. In a previous study, extensive evaluations were done to determine the best classifier for each UCI dataset [4]. We select the best classifiers from this study for each dataset and compare the accuracy of OG-Miner with it. The results are tabulated in Table VIII.

From Table VIII, we can see that six out of eight times, OG-miner outperforms the previous best schemes for the medical dataset. This shows the effectiveness of OG-Miner tool for these medical datasets. Therefore, we can conclude that our tool easily generalizes its knowledge extraction process to a new medical dataset with unknown statistics. (The statement is based on the experiments in Table VIII). The following section presents the field evaluations of our intelligent health tool in m-health environments of our mobile IT project.

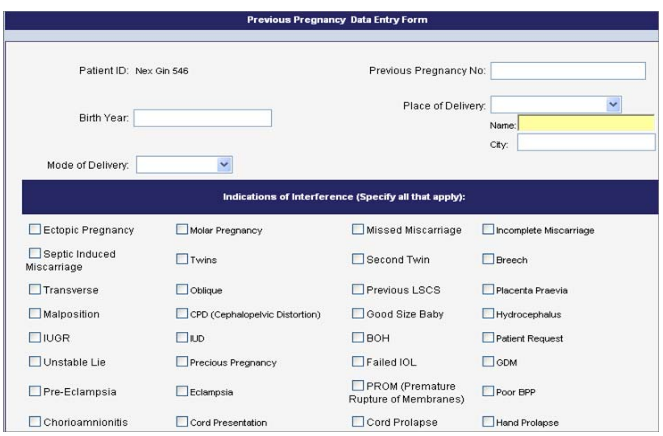
TABLE VIII
EVALUATION OF OG-MINER ON OTHER MEDICAL DATASETS

Datasets	Classification schemes		
	OG-miner	%AUC & name of Best in[4]	
Ann Thyroid	**99.8**	99.59	Bagging J48
Dermatology	**99.9**	99.75	NB
Cleveland heart	**79.2**	79	Stalking best three
Echo Cardiogram	**91.9**	83.5	Voting best three
Pima Indian Diabetes	81.8	**82.9**	NB
Horse colic	**91.7**	89.9	Voting best three
Breast cancer diag.	**99.3**	99.2	Bagging MLP
Breast cancer Prog.	72.8	**73.7**	Bagging NB

VIII. FIELD EVALUATIONS OF OG-MINER IN M-HEALTH ENVIRONMENTS

In this section, we evaluate the efficacy and effectiveness of our tool – in light of the users's feedback along three dimensions: (1) evaluation of the data obtained from rural areas, (2) an analysis of the features selected by our tool, and (3) feedback from consultant gynecologists on the effectiveness of the tool.

Evaluation of OG-Miner on the Field Data. Kindly recall that we have collected 2200 records of 1200 patients from two tertiary care hospitals. These records are used to not only evolve the architecture of OG-Miner but also train and test it. This data, however, is entered by medical experts using a custom tool. (see Figure 14 for the snapshot of the tool).


Previous Pregnancy Data Entry Form

Patient ID: Nex Gin 546 | Previous Pregnancy No:

Birth Year: | Place of Delivery: | Name: | City:

Mode of Delivery:

Indications of Interference (Specify all that apply):

Ectopic Pregnancy	Molar Pregnancy	Missed Miscarriage	Incomplete Miscarriage
Septic Induced Miscarriage	Twins	Second Twin	Breech
Transverse	Oblique	Previous LSCS	Placenta Praevia
Malposition	CPD (Cephalopelvic Distortion)	Good Size Baby	Hydrocephalus
IUGR	IUD	BOH	Patient Request
Unstable Lie	Precious Pregnancy	Failed IOL	GDM
Pre-Eclampsia	Eclampsia	PROM (Premature Rupture of Membranes)	Poor BPP
Chorioamnionitis	Cord Presentation	Cord Prolapse	Hand Prolapse

Fig. 14. Snapshots of the data entry tool

Now once we have completely deployed our intelligent tool in an m-health environment (as depicted in Figure 1), it becomes imperative to test it on the patient's data that is entered by semiskilled paramedic staff (Lady Health Workers (LHW)) on a PDA in the rural areas. LHWs use our customized sensor box (see Figure 16) and our PDA application (see Figure 15) for collecting the patients' vitals and entering the patients' records respectively. In this environment, the primary use of OG-Miner is to act as a filter on the records of the patients; as a result, alerts of only high risk patients are sent on the smart phone of medical experts. As a consequence, not only their workload is significantly reduced but also they are able to provide quality care to the patients. The system is deployed in a rural region (serving 1000 households) for more than 6 months now, and we have received 487 records of 295 patients in our EMR.

Fig. 15. LHW's PDA and its training simulator

To evaluate the performance of OG-Miner on this data, we compare its classification with the decisions of a specialized Gynecologist. A comparison of results (in the form of a confusion matrix) is tabulated in Table IX. The bold entries along the diagonal shows the correct classification done by our intelligent tool (taking the physician as the gold standard). The results show that only nine patients are misclassified by our system.

TABLE IX
A COMPARISON WITH THE DECISIONS OF A MEDICAL EXPERT

Medical Expert	**Classification Results of OG-miner**				
	Normal	**HYP**	**OB-L**	**HMG**	**SEPT**
Normal	**428**	5	0	0	1
HYP	1	**17**	0	0	0
OB-L	0	0	**9**	0	0
HMG	2	0	0	**7**	0
SEPT	0	0	0	0	**11**

To further elaborate, the system raised 6 false alarms on the smart phone of the gynecologist and has missed only 3 high risk patients. Remember false alarm are not a significant problem, but missing a high risk patient is. Therefore, we are investigating on how to reduce the false negatives to nearly 0%. It is interesting to see that no patient for obstructed Labor and Septicemia is misclassified by the system. On a positive note, the system successfully filtered 428 normal patients. The results – though obtained from a pilot study of 1000 households – are intriguing enough to show the effectiveness of our OG-Miner tool. To conclude, such intelligent health care tools have a significant benefit in resource constrained developing countries where the doctor to population ratio is very low.

Evaluation of Selected Feature by Medical Experts. In order to validate the effectiveness of our feature selection

Fig. 16. LHW's Sensor box for measuring vitals

scheme, we present our features' set to the specialized Gynecologists. The medical experts are excited to know that most of the features – selected by our tool – are also used by them in the decision making process. We show some relevant features in Table X. This further establishes the effectiveness of OG-Miner in selecting relevant features from electronic records of patients.

TABLE X
MEDICAL EVALUATION OF SELECTED FEATURE SETS

OG- risk Factor	Important Features
Hemorrhage	PV bleeding Gestational age
Hypertension	Blood pressure Edema Pulse Rate Renal function Creatinine
Septicemia	TLC count Renal Function Urine Respiratory Rate
Obstructed labor	Os-Vaginal examination Blood pressure Respiratory rate Consistency-vaginal examination

Feedback from Gynecologists. The Gynecologists involved in in our project are inclined towards integrating OG-Miner not only in our project but also using it as a standalone tool in the decision making process at outdoor patients at their hospitals. They believe that the tool will be of help in providing them a number of options in the beginning of diagnosis process for detecting the high risk factors of maternal mortality; as a result, they do not prematurely converge at a wrong diagnosis. Moreover, they have also suggested to augment OG-Miner with a rule based learning scheme that will empower them to audit, enhance and amend the diagnosis rules used by the tool. In near future, we will complement OG-Miner with a a rule learning classifier that enables a medical expert to edit/refine diagnosis rules in the rule base. As a result, the acceptability of such a transparent and interactive tool among medical experts will be significantly enhanced.

IX. CONCLUSION AND FUTURE WORK

The major contributions of the paper are: (1) an OG dataset with more than 2200 records of 1200 pregnant women that contains mortality patients because of hemorrhage, hypertension, obstructed labor and septicemia, (2) OG-Miner framework identifies high risk factors in an autonomous, accurate and reliable fashion for our OG dataset, and (3) OG-Miner framework is shown to provide consistent accuracies even on other UCI medical datasets. Remember that accuracy of OG-Miner on four OG datasets is above 98%. OG-Miner is now an integral component of our m-health solution and it provides significant support to the consultant gynecologists by filtering more than 90% of normal pregnant women in peripheral rural areas. As a result, record of only 10% high risk patients are referred to the experts. Consequently, they can provide quality care to the most deserving patients. The preliminary outcome of using our intelligent tool is that high risk patients are accurately identified and treated early in the pregnancy; therefore, MMR is expected to be significantly reduced leading to achieving the MDGs.

In near future, we will incorporate a rule learning classifier in our tool to generate diagnosis rules. The rules will not only provide a better insight about the classification process but they can be also evaluated, and enhanced/refined by the medical experts.

REFERENCES

[1] U. Fund, "State of the Worlds Children 2009. Maternal and Newborn Health," *New York: UNICEF*, vol. 60, 2008.

[2] C. Su and C. Yang, "Feature selection for the SVM: An application to hypertension diagnosis," *Expert Systems with Applications*, vol. 34, no. 1, pp. 754–763, 2008.

[3] Y. Chae, S. Ho, K. Cho, D. Lee, and S. Ji, "Data mining approach to policy analysis in a health insurance domain," *International journal of medical informatics*, vol. 62, no. 2-3, pp. 103–111, 2001.

[4] A. Tanwani, J. Afridi, M. Shafiq, and M. Farooq, "Guidelines to Select Machine Learning Scheme for Classification of Biomedical Datasets," in *Proceedings of the 7th European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*. Springer, 2009, pp. 128–139.

[5] M. Wozniak, "Boosted Decision Trees for Diagnosis Type of Hypertension," *Entropy*, vol. 1.

[6] L. Kenny, W. Dunn, D. Ellis, J. Myers, P. Baker, and D. Kell, "Novel biomarkers for pre-eclampsia detected using metabolomics and machine learning," *Metabolomics*, vol. 1, no. 3, pp. 227–234, 2005.

[7] L. Goodwin, M. Iannacchione, W. Hammond, P. Crockett, S. Maher, and K. Schlitz, "Data mining methods find demographic predictors of preterm birth," *Nursing research*, vol. 50, no. 6, p. 340, 2001.

[8] L. Woolery and J. Grzymala-Busse, "Machine learning for an expert system to predict preterm birth risk," *Journal of the American Medical Informatics Association*, vol. 1, no. 6, p. 439, 1994.

[9] I. Witten, E. Frank, L. Trigg, M. Hall, G. Holmes, and S. Cunningham, "Weka: Practical machine learning tools and techniques with Java implementations."

[10] A. Tanwani and M. Farooq, "Performance evaluation of evolutionary algorithms in classification of biomedical datasets," in *Proceedings of the 11th annual conference companion on Genetic and evolutionary computation conference*. ACM, 2009, pp. 2617–2624.

[11] D. Aha, D. Kibler, and M. Albert, "Instance-based learning algorithms," *Machine learning*, vol. 6, no. 1, pp. 37–66, 1991.

[12] A. Tanwani and M. Farooq, "The Role of Biomedical Dataset in Classification," *Artificial Intelligence in Medicine*, pp. 370–374, 2009.

[13] I. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Pub, 2005.