Automatic Classification of Toco-Signals: An Approach Towards Comprehensive Monitoring of Labour Progress

R. Jyothi
Dept. of Biomedical Engineering
M.I.T., Manipal University
Manipal, Karnataka-576104, India.

Rijul Saurabh Soans
Dept. of Biomedical Engineering
M.I.T., Manipal University
Manipal, Karnataka-576104, India.

Parvati V. Bhat
Dept. of Obstetrics and Gynecology
M.M.M.C., Manipal University
Manipal, Karnataka-576104, India.

***Abstract*— Childbirth is an intricate process which is marked by an increased cervical dilation rate caused due to steady increments in the frequency and strength of uterine contractions. The contractions may be characterized by its strength, duration and frequency (count) – which are monitored through Tocography. However, the procedure is prone to subjectivity and an automated approach for the classification of the contractions is needed. In this paper, we use three different Weighted K-Nearest Neighbor classifiers and Decision Trees to classify the contractions into three types: Mild, Moderate and Strong. Further, we note the fact that our training data consists of fewer samples of Contractions as compared to those of Non-contractions – resulting in "Class Imbalance". Hence, we use the Synthetic Minority Oversampling Technique (SMOTE) in conjunction with the K-NN classifier and Decision Trees to alleviate the problems of the same.**

The ground truth for Tocography signals was established by a doctor having an experience of 36 years in Obstetrics and Gynaecology. The annotations are in three categories: Mild (33 samples), Moderate (64 samples) and Strong (96 samples), amounting to a total of 193 contractions whereas the number of Non-contraction samples was 1217. Decision Trees using SMOTE performed the best with accuracies of 95%, 98.25% and 100% for the aforementioned categories, respectively. The sensitivities achieved for the same are 96.67%, 96.52% and 100% whereas the specificities amount to 93.33%, 100% and 100%, respectively. Our method may be used to monitor the labour progress efficiently.

Keywords*— *Uterine contractions, Subjectivity, K-Nearest Neighbour, Decision Trees, SMOTE.

I. Introduction

The uterus undergoes changes right from the early weeks of pregnancy until the birth of the child - it transforms from being in a quiescent state to a powerful contracting muscle. In the early stages of pregnancy, the uterus contracts irregularly; however, as the labour progresses, the intensity and the frequency of the contractions gradually increase. Finally, they become strong enough to completely dilate the cervix in order to expel the baby [1]. Generally, doctors categorize these contractions into three groups: Mild, Moderate and Strong. In particular, the frequency (counts) of the contractions is of great importance - its gradual rise is caused due to Oxytocin (a hormone produced by the Pituitary gland). Sometimes, sufficient amount of Oxytocin is not stimulated by the pregnant woman, leading to sub-optimal levels of contraction intensities and frequency. This type of labour is called as *Prolonged Labour* wherein, the doctors usually "*augment*" the labour using synthetic Oxytocin [2]. However, over dosage of the same can cause high-frequency uterine contractions (*Uterine Hyperstimulation*) and possibly rupture the uterus as well as cause foetal distress. Hence, synthetic Oxytocin must be given with proper attention and care to avoid such unwanted complications.

Presently, external Tocograph is used to monitor the strength, duration and frequency of uterine contractions. The device consists of a pressure sensitive transducer that picks up mechanical contractions from the uterus. It is held onto the fundus of the uterus and is kept in place with the help of a belt. The contractions cause a change in the resistance of the transducer, which is subsequently converted to an electric signal and is displayed on the graph with the X-axis as Time (sec) and the Y-axis as Amplitude (mmHg). A typical Tocography signal is shown in Fig. 1.

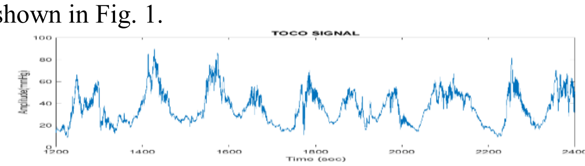

Fig. 1. Tocography signal

However, Tocography has a few shortcomings - subjectivity in the interpretation of Tocography signal [3] and the presence of Baseline wandering (the movement of the Base/ X- axis either upward or downward) [4].

In this paper, an objective method of classification is presented wherein the uterine contractions are classified into three types: Mild, Moderate and Strong. Further, the frequency of each type of contraction can be found by counting them. The algorithm may be used by medical professionals to avoid over dosage of Oxytocin and the associated ramifications. The paper is organized as follows. Section 2 highlights the previous work done in Tocography and identifies gaps in the literature. Section 3 consists of details of our algorithm. Subsequently in Section 4, we discuss important results and list out important contributions in Section 5.

II. Literature Review

Cardiotocograph (CTG) is a device which is used to measure the foetal heart rate (Cardio) and uterine activity

978-1-5090-1746-1/16/$31.00 ©2016 IEEE

(Tocography). The uterine activity is commonly measured to detect foetal distress. Specifically, if the foetal heart rate decreases (termed as *Deceleration*) simultaneously with uterine activity, then it is concluded that there is a “Foetal Distress”. Therefore, most of the work done on CTG is on the automated detection of uterine contractions [5, 6, and 7]. However, other works are on the feature extraction from the foetal heart rate signal [8]. Some of the features include morphological features, time and frequency domain features, statistical features obtained from Heart Rate Variability, Wavelet-based and non-linear features.

Cardiotocography is also corrupted by high frequency noise. [9] have concentrated on removing the same from CTG signal using Empirical Mode Decomposition. But the algorithm is very slow due to the sifting process and consequently, we have not used the technique for noise correction in the present work. On the other hand, [10] have used Tocography to classify uterine contractions into True and False labour (one that does not lead to the delivery of the baby). Recently, we came up with a novel algorithm that uses Tocography signals to explicitly classify the contractions into “Mild”, “Moderate” and “Strong” using a K-NN classifier [11]. However, we found that the method had low accuracies and specificities.

In this paper, we look to improve on our previous work [11] specifically, in terms of the classifier’s accuracy, sensitivity and specificity. We try to achieve the same through the use of Weighted K-NN classifier and Decision Trees. Further, we acknowledge the fact that the samples in our dataset are not equally represented by each of the classes and use a suitable technique to overcome the possible effects of the same. In particular, we make use of the SMOTE technique and associate it with Decision Trees and K-NN classifiers. Also, we prove the efficacy of the interpolation technique which was used to correct baseline wandering in [11].

III. METHODOLOGY

The methodology is as follows:

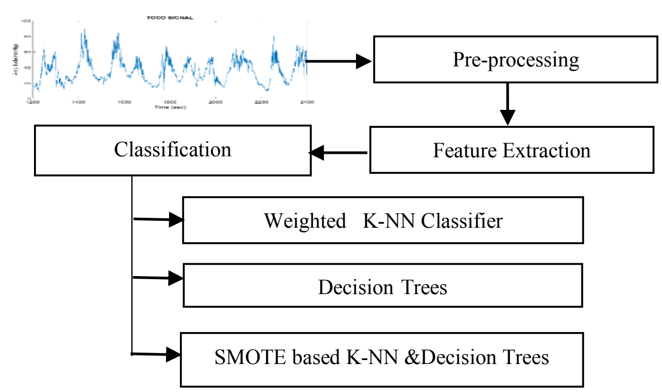

Fig. 2. Methodology

A. Pre-processing and Feature Extraction

Tocography signal is corrupted by high and low frequency noise. The high-frequency noise is due to the stress induced during labour and is removed by implementing a Moving Average filter, as in [11]. The low frequency noise (also known as baseline wandering), shifts the X-axis either upward or downward and makes the interpretation of the contractions difficult. The noise was removed by an interpolation technique, the details of which can be found in [11].

We implement a FIR high pass filter to remove the low frequency noise and compare it with results of the interpolation technique mentioned in [11]. The cut off frequency of the high pass filter was determined by taking the Power Spectral Density (Welch’s method) of the low frequency noise. The aforementioned noise is obtained by subtracting the baseline wandered signal from the baseline corrected signal by following the steps in [11]. The order of the filter was increased empirically until the desired signal was obtained.

Subsequently, eight features are extracted from the pre-processed signal. They are listed as: Intensity, Duration, Area and Power of each contraction, Time to Rise, Time to fall, Mean Duration Time and Standard Deviation Time. The mathematical equations of each feature can be found in our previous work [11].

B. Classification: Weighted K-NN and Decision Trees

We choose non-parametric methods of classification as they do not require any statistical assumptions on the distribution of the extracted features. The task involves a multi-class classification and consequently, the One-vs.-all classification technique is implemented wherein, a binary classifier is trained for each type of contraction. This translates to three binary classifiers in each classification method. Moreover, we split the data into a set of Training data (70%) and Testing data (30%).

First, we implement the Weighted K-NN and the Decision Tree classifiers and store their results to compare with subsequent SMOTE-based methods. Each of them is explained in detail in the following sections.

1. WEIGHTED K-NN

Typically, in the K-Nearest Neighbor (K-NN) classification technique, all the neighbors are given equal weights. However, in the Weighted K-NN technique, the nearest neighbor is given more weight using an appropriate weighting function. In this paper, three weighting functions are implemented. They are:

$$Inverse\ Function = \frac{1}{D(testvector,trainvector)} \tag{1}$$

$$Squared\ Inverse\ Function(SID) = \frac{1}{D(testvector,trainvector)^2} \tag{2}$$

$$Exponential\ Function = e^{-|D(testvector,trainvector)|} \tag{3}$$

where *D (test vector, train vector)* is the Euclidean distance between the training data vectors and the test data vectors. From Eq. 1, 2 and 3, it is evident that the weights increase as the distance between the training vectors and the test vectors decrease.

The training phase of the Weighted K-NN classifier involves storing all the 8 features of the training data as a feature vector and the corresponding class labels. In the test phase, the following steps are implemented:

1. Find the Euclidean distance between the test data vector and every training data vector.
2. Find the K-Nearest Neighbors and their corresponding class labels. The value of K is found by 10-fold technique [12].
3. Multiply the labels found in step 2 with the weights assigned and sum them to get the class label of the test vector.

2. Decision Trees

Decision trees are non-parametric classification methods structured in the form of a tree. They are constructed in a top-down fashion by means of finding the best feature vector using a *Gain* measure i.e. a feature is selected when it has the maximum *Gain* as compared to other features [12]. The training data is split based on the best feature and the process is repeated until a stopping criterion is reached. The measure that we employ to obtain the decision criterion is the *Gini's Diversity Index*. It is defined as follows:

$$Gini\ Index(n) = 1 - \sum_{i=0}^{c-1} p(i/n)^2 \quad (4)$$

where:
$p\ (i/n)$ is the number of examples belonging to class i in a particular node n.

C. SMOTE (Synthetic Minority Oversampling Technique) based Classification: K-NN and Decision Trees

Typically, classifiers are biased towards the *Majority class* i.e. the class which is represented by a far greater number of samples as compared to those in other classes [13]. Our dataset is a perfect example of such a situation – commonly termed as a "Class Imbalance" problem. In our case, the experienced doctor identified 193 Contraction samples and 1057 Non-contraction samples. The Contraction samples include 33 Mild, 64 Moderate and 96 Strong samples; the Non-Contractions generally arise due to uterine muscle twitch [14]. Our task is to perform Multi-class classification and consequently, we use the One vs. All approach i.e. Mild vs Non-Mild, Moderate vs Non-Moderate & Strong vs Non-Strong. Note that, the 'All' category consists of Non-Contraction samples as well. As one can easily see, the ratios of 'One' to 'All' classes are 33/1217, 64/1186 and 96/1154 for the Mild, Moderate and Strong binary classifiers, respectively.

We undersampled the training data points while training the weighted K-NN and the Decision Tree classifiers, so as to avoid the bias of the classifiers towards the 'All' classes. The majority class (typically, the 'All' class) was undersampled until a balance was reached between the data points in the 'One' and the 'All' class for Mild, Moderate and Strong contraction categories. However, due to a significant amount of under sampling, the classifier may miss many negative (majority class) examples thereby, reducing the performance of the classifier. One approach to avoid losing too much of information is to oversample the positive examples by simply replicating them; however, the classifiers tend to overfit the data [15].

A better method of alleviating the bias and preventing overfit is SMOTE – first introduced by [15]. The method presents a simple way of oversampling and operates in the feature space rather than in the data space. One can oversample the data in the *Minority class* (under-represented class) from 100% upto 500%. For this purpose, the number of nearest neighbors to be considered for each point in the feature space is fixed at 5. Depending upon the amount of oversampling required, suitable numbers of neighbors (i.e. 1 for 100%, 2 for 200% and so on) are chosen from the aforementioned fixed 5 nearest neighbors. For example, if the minority class is to be oversampled by 300% then, 3 neighbors are randomly chosen from the 5 nearest neighbors of each point in the class and subsequently, 3 new *Synthetic samples* are generated. The synthetic sample is generated in the following way:

i. The difference between the feature vector and its nearest neighbor is taken.
ii. The obtained difference is multiplied by a random number between 0 and 1 and added with the feature vector.

In this paper, we have oversampled the Minority class (i.e. the 'One' classes corresponding to Mild, Moderate and Strong) by 500%, respectively. The technique results in 5 new synthetic samples for every feature point in the feature space. Hence, the total number of Mild, Moderate and Strong contraction samples obtained after implementing the SMOTE technique is 198 (33+500% of 33), 384 (64+500% of 64) and 576 (96+500% of 96), respectively. Note that, there is still a possibility of the classifier being biased towards the 'All' classes as the number of Non-contraction samples alone amounts to 1217. Therefore, in order to achieve a balance between the 'One' and the 'All' classes, we randomly undersample the negative (majority) class until the 'One' and the 'All' classes are balanced evenly. Although undersampling is done again as in the preceding section, more samples are currently included in the training set thereby, making the classification robust. Table 1 shows the total number of samples in the 'All' classes (negative examples) chosen for training the classifier before and after applying SMOTE technique.

Table I. No. of negative examples included in the training set before & after SMOTE

Category	Before SMOTE	After SMOTE
'All' class for Mild	23	138
'All' class for Moderate	44	269
'All' class for Strong	67	403

Subsequently, the generated data is given as input to the Decision Tree classifier and the standard K-NN classifier.

IV. Results

A. Data Collection and Annotation

The Tocography data was obtained from the CTU-UHB Intrapartum Cardiotocography Database of Physionet [16]. The sampling frequency of the signal is 4 Hz and 20 minute segments were taken as input for the classification. The ground truth was established by a doctor having around 36 years of experience in Obstetrics and Gynecology. Based on the expertise, the doctor was asked to mark the contractions as Mild, Moderate or Strong. After the annotation, 193 contraction samples consisting of 33 Mild, 64 Moderate and 96 Strong contractions were obtained. Moreover, 1057 Non–Contraction samples were also identified.

B. Pre-Processing

The Input signal and the pre- processed signal are shown below:

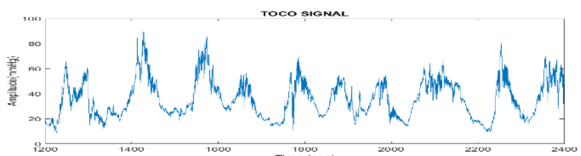

Fig. 3. Input Signal

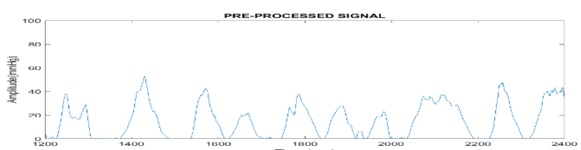

Fig. 4. Pre-processed signal

A peak frequency of 0.0031 Hz was found after taking the PSD of low frequency noise. Fig 5 shows the Power Spectrum Density (PSD) of the low frequency noise.

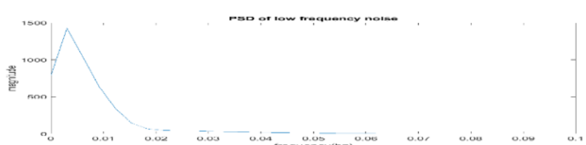

Fig. 5. PSD of the low frequency noise

A Kaiser Window FIR filter was implemented with the obtained cut off frequency of 0.0031 Hz and an order of 300. Fig. 6. shows the baseline corrected signal using a FIR High pass filter.

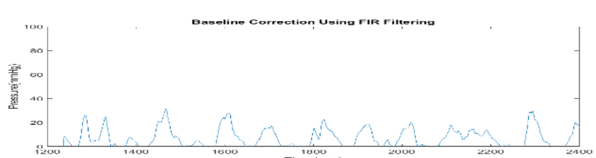

Fig. 6. Baseline correction using FIR High Pass Filter

Comparing Figs 4 & 6, a significant delay is observed due to which a contraction is lost. The delay can be calculated using the formula:

$$Delay = \frac{N-1}{2fs}\, seconds \qquad (5)$$

Where N is the order of the filter and fs is the sampling frequency. The delay is almost equivalent to 37.375 seconds. An effort to strike a balance between the order and delay of the signal was made; however, we found out that such an operation would result in poor baseline correction, thereby affecting subsequent steps. The interpolation technique developed in [11] does not have a delay and moreover, the order required to correct the baseline wander is very less. Therefore, the same interpolation technique is implemented in this paper.

C. Classification: Weighted K-NN and Decision Trees

We use 10-fold technique to find the correct number of neighbors in case of the weighted K-NN classifiers. The numbers of neighbors obtained for Mild, Moderate and Strong categories are found to be 3, 3 and 7, respectively.

The classifiers were evaluated using Accuracy, Sensitivity and Specificity parameters. Table 2, 3 and 4 show the performance of the Inverse distance, SID and Exponential Weighted K-NN classifiers.

Table II. Performance of the Inverse distance weighted K-NN classifier

Type	Accuracy (%)	Sensitivity (%)	Specificity (%)
Mild	90.91	90.91	90.91
Moderate	77.5	75	80
Strong	83.93	92.86	75

Table III. Performance of the SID weighted K-NN classifier

Type	Accuracy (%)	Sensitivity (%)	Specificity (%)
Mild	90.91	90.91	90.91
Moderate	75	70	80
Strong	82.14	89.29	75

Table IV. Performance of the Exponential weighted K-NN classifier

Type	Accuracy (%)	Sensitivity (%)	Specificity (%)
Mild	90.91	90.91	90.91
Moderate	77.5	75	80
Strong	85.71	92.86	78.57

Moreover, we used Decision Trees with Gini's Diversity Index as a measure of impurity. The index was chosen by its performance based on the 10-fold technique. Table 5 shows the performance of Decision Trees.

Table V. Performance of Decision Trees

Type	Accuracy (%)	Sensitivity (%)	Specificity (%)
Mild	81.82	100	63.64
Moderate	92.5	100	85
Strong	89.29	92.86	85.71

As seen from Tables 2, 3, 4 & 5, Decision trees have higher accuracies for Moderate and Strong categories. Moreover, we found that Decision Trees used only Intensity, Mean duration time and Area features for classification in all the three categories. Therefore, there is a considerable reduction in the features of the dataset.

D. Classification: SMOTE based K-NN and SMOTE based Decision Trees

Tables 6 and 7 show the performance of K-NN classifier and Decision Trees after oversampling the minority class using the SMOTE technique. The number of neighbors chosen for K-NN classifier after oversampling is 3 for all the class types.

Table VI. Performance of the K-NN classifier using the SMOTE technique

Type	Accuracy (%)	Sensitivity (%)	Specificity (%)
Mild	92.5	100	85
Moderate	99.13	99.13	99.12
Strong	100	100	100

Table VII. Performance of Decision Trees using the SMOTE technique

Type	Accuracy (%)	Sensitivity (%)	Specificity (%)
Mild	95	96.67	93.33
Moderate	98.25	96.52	100
Strong	100	100	100

As seen from the Tables 6 and 7, the performance of the K-NN classifier and Decision Trees improved after oversampling. Further, it was seen that Decision Trees used only 3, 5 and 4 features for classifying Mild, Moderate and Strong categories, respectively – resulting in feature reduction.

E. Discussion

No results pertaining to classification (except [11]) of uterine contractions using Tocography (Mild, Moderate and Strong), was found in the literature for comparison. While [10] classified the labour into True or False, they did not explicitly classify the contractions. Table 8 reproduces the results of classification obtained in [11] and shows the efficacy of our new algorithm. A comparison of Tables 7 and 8 shows the better performance of Decision tree using SMOTE over K-NN classifier in [11]. Moreover, the fact that the former technique performed the best among the classifiers in this work suggests that the issue of 'Class Imbalance' has been addressed.

Table VIII. Performance of Weighted K-NN [11]

Type	Accuracy (%)	Sensitivity (%)	Specificity (%)
Mild	90.91	90.91	90.91
Moderate	85	90	80
Strong	85.71	100	71

V. CONCLUSION AND FUTURE WORK

A scheme to classify uterine contractions into Mild, Moderate and Strong categories, has been presented. The contraction signal is pre-processed with a moving average filter and an interpolation technique. Further, eight features were extracted and subjected to three different weighted K-NN classifiers and Decision Trees. Moreover, the issue of class imbalance was recognized and consequently, SMOTE-based K-NN classifiers and SMOTE-based Decision Trees were employed to reduce classifier bias. Our classifiers perform better than those in [11] and may be used by healthcare professionals to automatically monitor the labour progress efficiently. Future work involves including an additional doctor to agree on the ground truth and a comprehensive hardware implementation of our algorithm.

VI. REFERENCES

[1] N. Hatfield, "Introductory maternity and paediatric nursing," Wolters Kluwer Health, China, 3rd Edition, 2013.

[2] N. J. Klossner, "Introductory Maternity Nursing: Lippincott Williams & Wilkins", USA, 1st Edition, 2006.

[3] J. Bernardes, A. Costa-Pereira, D. Ayres-de-Campos, H. Van Geijn, and L.Pereira-Leite, "Evaluation of interobserver agreement of Cardiotocograms," International Journal of Gynaecology & Obstetrics, vol. 57, pp. 33-37, 1997

[4] J. A. Marques, P. C. Cortez, J. P. Madeiro, and F. S. Schlindwein, "Computerized Cardiotocography analysis system based on Hilbert Transform," Expert System with Applications, vol. 40, pp. 7159-7658, 2013.

[5] J. A. Marques, P. C. Cortez, J. P. Madeiro, and F. S. Schlindwein, "Computerized Cardiotocography Analysis System Based on Hilbert Transform," Expert System with Applications, vol. 40, pp. 7159-7658, 2013.

[6] S. Cazares, L. Tarassenko, L. Impe, M. Moulden, and C. Redman, "Automated Identification of Abnormal Cardiotocograms Using Neural Network Visualization Techniques," Engineering in Medicine and Biology Society, 2001. *Proc.23rd Annual International Conference of the IEEE*, vol. 2, pp. 1629-1632, 2001.

[7] M. Jeżewski, R. Czabański, J. Wróbel, and K. Horoba, "Analysis of Extracted Cardiotocographic Signal Features to Improve Automated Prediction of Fetal Outcome," Biocybernetics and Biomedical Engineering, vol. 30, pp. 29-47, 2010.

[8] V. Chud´aˇcek, J. Spilka, M. G. Huptych, G, and L. Lhotsk´a, "Linear and Non-Linear Features for Intrapartum Cardiotocography Evaluation," Computing in Cardiology, vol. 64, pp. 999-1002, 2010.

[9] B. Krupa, M. Ali, and E. Zahedi, "The Application of Empirical Mode Decomposition for the Enhancement of Cardiotocograph Signals," Physiological Measurement, vol. 30, pp. 729-743, 2009.

[10] S. Hiwale, V. Pallavi, F. Celine, and R. S. Singh, "Automated Uterine Contraction Analysis for Detecting False Labour Contractions," presented at the First AMA IEEE Medical Technology Conference on Individualized Healthcare, 2010.

[11] R.Jyothi, S. Hiwale, P. Bhat "Classification of Labour Contractions using K-NN Classifier", *Proc.IEEE International Conference on Systems in Medicine and Biology*, IIT Kharagpur, in press, 2016.

[12] L. Breiman, J. Friedman, C. Stone, and R. Olshen, "Classification and Regression Trees, " Taylor & Francis, 1994.

[13] E. Alpaydin, "Introduction to Machine Learning", MIT Press, USA, 2nd Edition, pp.163-209, 2014

[14] I. V. Gestel, M.M.I. Jland, H.J.Hoogland and J.L.H.Evers, "Endometrial wave-like activity in the non-pregnant uterus", Human Reproduction Update, Vol.9, No.2, pp. 131-138, 2003

[15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.

[16] V. Chudáček, J. Spilka, M. Burša, P. Janků, and L. Hruban, "Open access intrapartum CTG database," BMC Pregnancy and Childbirth, vol. 14, pp. 14-16, 2014.