24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

EXTREME LEARNING MACHINE AND DECISION TREE BASED ENSEMBLE TECHNIQUES FOR DIAGNOSIS OF ROLLING ELEMENT BEARING DEFECTS

Nitin Upadhyay, Pavan Kumar Kankar

PDPM Indian Institute of Information Technology Design and Manufacturing, Jabalpur, India

Email: nitin.upadhyay7@gmail.com

Bearings are the most critical component used in all kind of rotating machines. Due to the long continuous run, the fatigue stresses are generated which result, the defects developing over the surfaces of the bearing. For smoothly running the operation, it is important to diagnose these defects before the severe damages occur. This study is based on the diagnosis of rolling element bearing surface defects with the help of new machine learning algorithms like extreme learning machine, rotation forest, random forest, and random subspace. An experiment has been conducted to obtain the vibration signals for the analysis of various surface defects such as Inner race defect, outer race defect, and ball defect. For further processing of these signals ten time-domain features are extracted from the experimental data. Results show that Extreme learning machine is more efficient than the other discussed ensemble techniques for the classification of rolling element bearing defects.

Keywords: Bearing defects, Feature extraction, Extreme learning machine, Ensemble techniques, Vibration signal analysis

1. Introduction

Rolling element bearing is one of the very essential and critical mechanical components used in almost all kind of rotating machines from large industrial system to the small handheld devices. Defects in these bearings will lead to the sudden breakdown of the system. It is important to identify these defects in advance. Defects are mainly classified into two categories the first is localized defects i.e. spall, pits, dents etc. and second is distributed defects which is spread over the surface of bearing example of distributed defects are waviness on inner race, waviness on outer race and off-sized rolling element. Condition monitoring techniques will help to diagnose these defects. There are several condition monitoring techniques such as vibration analysis, acoustic emission, stator current analysis, oil lubrication analysis, etc. out of these vibration analysis is most widely used and accurate prediction method used for diagnosis of the rolling element bearing defects.

For effective analysis of the vibration signals several studies have been carried out by using signal processing and machine learning techniques. A process of fault diagnosis of rolling element bearing by using intrinsic mode decomposition (IMF) envelope spectrum and support vector machine (SVM) proposed by Yang et al.[1]. In their work, the authors have considered the defect on inner and outer race of the bearing and collected the vibration signals; features have been extracted by using IMF envelope analysis and these features with known output used as test and training set for SVM. Lei et al. [2] have developed a new artificial intelligent methodology i.e. adaptive neuron-fuzzy inference system, improve distance evaluation and EMD for the bearing defect diagnosis.

Recently several authors have employed the artificial intelligent techniques such as artificial neural (ANN), SVM, EMD, etc. for the fault diagnosis of the rolling element bearing combined with wavelet and feature selection techniques [3-10]. Prabhakar et al. [3] used discrete wavelet transform to detect races defects of a ball bearing. In their work, the authors collected the vibration response of rolling element bearing having single and multiple point defects on the inner and outer race of the bearing. Paya et al. [4] used ANN for fault diagnosis of the ball bearing. The authors have used wavelet transform as a pre-processor. Purushotham et al. [5] proposed an integrated wavelet analysis and hidden Markov model (HMM) base model to identify and classify the multiple faults on the ball bearing and achieve 99% classification efficiency. Vyas and Satishkumar [6] proposed neural network based technique for fault identification of rotating machinery, in their work multilayer network and back propagation learning algorithm have been used and authors observed overall success rate is up to 90%.

Kankar et al. [7,8] used machine learning techniques such as SVM, learning vector quantization (LVQ) and ANN for fault diagnosis of the rolling element bearing. The authors observed that the performance of SVM was better than the ANN and LVQ. Most recently, Kavathekar et al. [9] and Sharma et al. [10] presented ensemble techniques i.e. rotation forest and random forest etc. and observed that rotational forest is more efficient than other methods.

In this paper, the machine learning techniques such as extreme learning machining (ELM), and decision tree based ensemble technique such as rotational forest, random forest (RF), and random subspace (RS) have been used for the classification of bearing defects. An experiment was conducted to extract the vibration response of healthy as well as defective bearing. Ten time domain features have been extracted from the vibration response of each case. These extracted features are used for training and testing set with known output of ELM and ensemble techniques used for the classification of bearing defects.

2. Machine learning techniques

Machine learning techniques are an approach to create the program from the data. If the data have both input value and output value, it is known a supervised learning. In other case, when only input parameters are known with the unknown output and the learning job is to gain some understanding of the method that produced the data, this kind of learning approach are said to be unsupervised. In this present work, the four machine learning techniques have used, namely, ELM, rotation forest, RF, and RS for the classification of rolling element bearings defect.

2.1 Extreme learning machine

Extreme learning machine is used to overcome the challenging problem faced by the back propagation (BP) learning algorithm.

ELM is mainly developed for the single layer feed-forward neural network (SFNN) and then extended to the general SFNN where the neuron need not to be neuron alike [11-13]. In ELM, the hidden layer needs not to be tuned. The ELM output function for the generalized SFNNs is written as:

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta \tag{1}$$

where, $\beta = [\beta_1, \dots\dots\dots\dots.., \beta_N]^T$ represent the output weight matrix between the hidden layers of L nodes and output nodes. $h(x) = [h_1(x), \dots\dots\dots\dots., h_L(x)]$ is the output matrix of the hidden layer in respect of the input x. Generally, this is a process to map the data from d dimensional input space to hidden layer feature space of L dimensional (ELM features space) H, *h(x)* is actually a feature mapping. for the real application the value of $h_i(x)$ can be:

$$h_i(x) = G(a_i, b_i, x)\,,\; a_i \in R^d \text{ and } b_i \in R \tag{2}$$

where, $G(a_i, b_i, x)$ is a nonlinear piecewise continuous function which satisfy the ELM universal approximation statement [12].

From the learning perspective, unlike classic learning algorithm, ELM main focus is to achieve smallest training error as well as smallest weight output norms [13].

$$Minimize : \| \beta \|_p^{\sigma_1} + C \| H\beta - T \|_q^{\sigma_2} \tag{3}$$

where, $\sigma_1 > 0, \sigma_2 > 0$ and $p, q = 0\ , \ \frac{1}{2}, 1, 2, \ldots\ldots, +\infty$ and H is the hidden layer output matrix:

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdots & \cdots & \cdots & h_L(x_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_1(x_N) & \cdots & \cdots & \cdots & h_L(x_n) \end{bmatrix} \tag{4}$$

and training data matrix T can be written as:

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & \cdots & \cdots & t_{1m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ t_{N1} & \cdots & \cdots & \cdots & t_{Nm} \end{bmatrix} \tag{5}$$

Some basic learning principles of ELM are as:

Principle 1: The hidden neurons of SFNN are randomly generated and independent of training samples as well as its learning environments.

Principle 2: In view of generalization performance and system stability, the output weight norm of generalized SFNN need to be small with some optimization constrain.

Principle 3: From the optimization perspective the SFNN output node should be unbiased (or set bias zero)

So as to satisfy the second learning principle, the minimal norm least square method have been used for basic execution of ELM (when $C = \infty$) in place of standard optimization techniques:

$$\beta = H^{\dagger} T \tag{6}$$

where, $H^{\dagger} = Moore - penrose\ generalized\ inverse\ of\ matrix\ H$, different methods have been used to obtained the value of $H^{\dagger}$ i.e. singular value decomposition (SVD), orthogonal projection method and projection method. The output vector β can be estimated by other iterative methods.

2.2 Rotation forest

Rotation forest is a new ensemble technique which is derived to overcome the issue with classifiers i.e. decision trees, random forest, random subspace, boosting, Adaboost etc., firstly introduce by Rodríguez et al. [14]. The primary objective of the rotation forest ensemble is to construct accurate and diverse classifiers. Steps involve in rotation forest ensemble are as follow:

Let $A = [a_1, \ldots \ldots a_n]^T$ is represent a n feature data set, which having n features, and A represent data point containing training objects in the form of $N \times n$ data matrix. Let B is the vector with class labels, $B = [b_1 \ldots \ldots b_n]$. Where, b_j can take value from the class labels $\beta = \{\beta_1 \ldots \ldots \beta_c\}$. Taking, $P_1 \ldots \ldots P_L$ is L base classifiers in the ensemble technique, the value of L must be decided in advance. To build the classifier training set. Following steps have been carried out:

Step-1 Split the feature vector S into K feature subset. The feature subset may be disjoint or intersecting. Disjoint feature subset have been selected to increase the diversity of the classifier. For the simplicity, assume that K is a factor of n, so, each feature subset having $Q = n/k$ features.

Step-2 $S_{i,j}$ is the j^{th} subset of feature for the training classifier P_i, for each subset a non-empty subset of classes are randomly selected and later draw a bootstrap sample size which is *75%* of the data count.

Step-3 Apply PCA on Q feature set of $S_{i,j}$ and selected subset of A. Save the principal component coefficients, $m_{i,j}^{(1)}, \ldots \ldots \ldots m_{i,j}^{(Qj)}$ each of size $Q \times 1$. Note that there is a chance that some of the Eigen values are zero. So that, the number of vectors obtaining after applying the PCA, may not have Q vectors i.e. $Q_j < Q$. The PCA have only applied on each feature subset instead of whole data set to avoid the alike feature coefficient if the same feature set have been selected for distinct classifiers.

Step-4 Finally arrange the principal component coefficient in the form of distributed “rotation matrix” F_i.

$$F_i = \begin{bmatrix} m_{i,j}^{(1)}, \dots\dots\dots, m_{i,j}^{(Q_i)} & \dots\dots & [0] \\ \vdots & \ddots & \vdots \\ \vdots & & \vdots \\ [0] & \dots & m_{i,j}^{(1)}, \dots\dots\dots, m_{i,j}^{(Q_j)} \end{bmatrix}_{(n \times \sum_j Q_j)} \quad (7)$$

The dimensionality of the rotation matrix is $n \times \sum_j Q_j$. The training data set for the classifier P_i is computed by rearranging the column of the rotation matrix, so that they are corresponding to the original feature set and the rearrange rotation matrix is denoted by F_i^a of size $N \times n$. The training set for the classifier D_i is XR_i^m

2.3 Random forest

Random forest is an efficient and very popular ensemble method, based on the idea of model aggregation, for both regression and classification problems, it is firstly introduce by Breiman [15]. Let the learning set $L = \{(X_1, Y_1), \dots\dots.., (X_n, Y_n)\}$, where X_n is an input data and Y_n is an output predictor for the classification, a classifier t is a mapping $t: R^P \to Y$, and for the regression, suppose that $Y = S(x) + \varepsilon$. Where, $E[\varepsilon|x] = 0$ and S is regression function. RF is a model formation approach providing estimators of either the Bayes classifier, which is the mapping minimizing the classification error $P(y \neq \pm(x))$, of the regression function.

The basic RF method is the combination of many decision trees built by using various bootstrap samples taking from the L learning samples. The following Steps are involve in the RF method:

1. Build of n tree bootstrap sample of the original learning data set L.
2. For the each bootstrap sampling, grow a decision tree. When the decision tree is growing, at the each node, n samples are randomly selected from the N samples.
3. Basically, $n << N$. It is recommending that start with $n = \sqrt{N}$ or $[log_2(N) + 1]$ and then repeatedly increasing and decreasing the values of n until the minimum error of the out bag data (OOB) are obtained. At each node, only one variable which give the best split has used out of the n samples.

2.4 Random subspace

Random subspace is an ensemble technique, similar to the random forest, used for the classification and regression, firstly introduce by Ho [16]. In the RS method we modify the training data, but the modification is only made in feature space. Let us assume training data point $X_i(i = 1, \dots..n)$ within the training data set $X = (X_1, X_2, \dots..X_n)$ having n -dimensional vector $X_i = (x_{i1}, x_{i2}, \dots\dots x_{in})$. In the RS method feature subset which is containing r features $(r < n)$ are randomly selected from the data set X. Later, the n dimension feature subset is divided into r dimension random subspace. Therefore, the new modified training set $X^b = (X_1^b, X_2^b, \dots..X_n^b)$ having r-dimensional training objects $X_i^b = (X_{i1}^b, X_{i2}^b, \dots\dots..X_{r2}^b)$ $i = 1 \dots..n$, where r components $X_{i.j}^b$ $j = 1, \dots.r$ are randomly selected from m components $X_{i.j}$ $j = 1, \dots.m$ of the training sample X_i. Then the RS classifier have been built and combined by simple majority voting. The organization of the random subspace is as follow:

1. Repeat for $b = 1, \dots\dots\dots.B$;

(a) From original m-dimensional feature space X, random subspace of r dimension selected

(b) Build a classifier $C_b(x)$ in the X^b (with the decision boundary $C_b(x) = 0$)

2. Combined the classifier $C_b(x)$, $b = 1, \dots..B$ by simple majority voting to a final decision rule.

$$\beta(x) = \underset{y=\{-1,1\}}{\arg\max} \sum_b \delta_{\mathrm{sgn}}(\mathrm{C_b(x)})_{\mathrm{y}} \quad (8)$$

where, $\delta_{i,j}$ is Kronecker symbol and $y \in \{-1,1\}$ class label of the classifier.

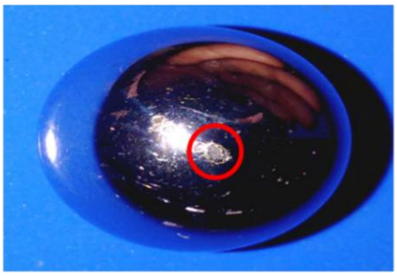
3. Experimental Setup and Features Extraction

To reduce the plant downtime and full utilization of the machine production capacity, it is necessary to predict the degradation of the machine component (Bearing in present study) before they cross the failure threshold. An experimental setup constructed to record the vibration response of the healthy and defective bearing. An artificial defect produced on the inner race, outer race and rolling element with the help of the electric discharge machining (EDM) as shown in the Fig.1. The experimental setup have been consist of a 3HP induction motor whose maximum rotating speed is 1980 *rpm* and the speed is controlled by variable frequency drive. The one end of shaft is connected with motor with the help of flexible coupling and another end supported by test bearing. The constant load of 1764 *N* has applied to the test bearing. The schematic diagram of the experimental setup is shown in Fig.2.

Vibration responses of the each bearing have been recorded with the help of uniaxial accelerometer whose sensitivity is 100 mv/g. The data have been recorded by varying the rotating speed from 500 to 1500 *rpm* with an interval of 100 *rpm*. For the fast data acquisition, and record vibration data DEWESOFT data acquisition system (DAQ). Table 1 represents the detailed dimension of the bearing use to record the vibration signal.

(a)

(b)



(c)

Figure 1 Bearing defects (a) Inner race defect (b) Outer race defect (c) Ball defect

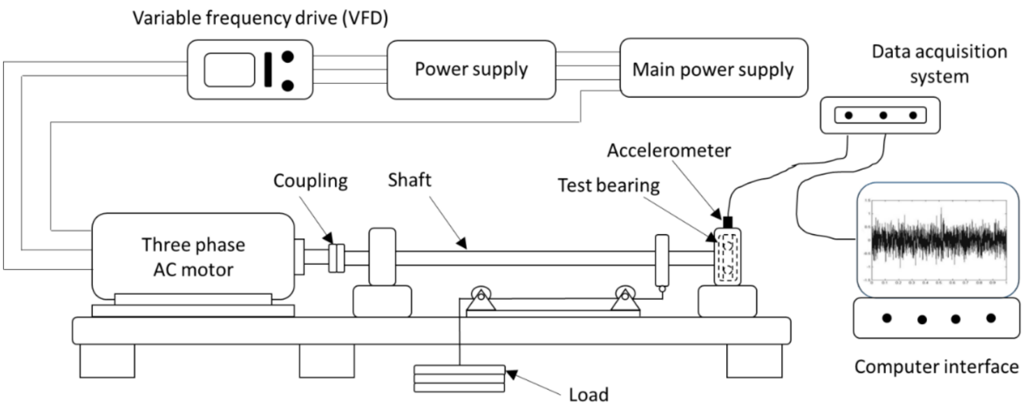


Figure 2 schematic of experimental setup of rotor bearing system

Table 1 Detail specification of the bearing

Parameters	Values
Bearing number	NBC 6307
Inner Race Diameter (D_i)	*35mm*
Outer Race Diameter (D_0)	*80mm*
Ball diameter (d_b)	*11.25mm*
Width (B)	*21mm*
Number of balls (N_b)	*8*
Contact angle (φ)	*0^o*
Radial internal clearance (γ)	*20µm*

4. Results and Discussion

In this section, training and testing data are classified by using ELM, rotation forest, RF, and RS with the help of Matlab2014. The results of the test data are displayed in the form of the two-dimensional confusion matrix. The confusion matrix having a column and a row for each class and each element of the matrix is shown the number of test example for which predicted class is the column, and actual class is the row.

Total 264 instances and ten features are used in this study. These features or attribute are used as input to the machine learning technique for classification of rolling element bearing defects. The features included ten statistical time domain features such as kurtosis, skewness, RMS, standard deviation, variance, peak to peak, maximum value, minimum value, crest factor and form factor [17].

The testing results of ELM, rotation forest, RF and RS using 10-fold cross validation and test set are shown in Table 2-5 respectively. Total 264 number of cases are obtained out of these 66 instances are considered as each healthy bearing (HB). Bearing with inner race defect (IRD), bearing with outer race defect (ORD), and ball defect (BD).

Table 2 Confusion matrix for ELM

Using Test Set (90 instances)					10 Fold cross validation (264 instances)				
HBD	IRD	ORD	BD	Classified as	HB	IRD	ORD	BD	Classified as
18	0	0	0	HB	66	0	0	0	HB
0	20	0	0	IRD	0	65	1	0	IRD
2	0	24	0	ORD	0	2	64	0	ORD
0	5	0	21	BD	0	0	9	57	BD

Table 3 Confusion matrix for rotation forest

Using Test Set (90 instances)					10 Fold cross validation (264 instances)				
HB	IRD	ORD	BD	Classified as	HB	IRD	ORD	BD	Classified as
18	0	0	0	HB	56	0	1	9	HB
0	20	0	0	IRD	0	66	0	0	IRD
0	0	26	0	ORD	0	1	65	0	ORD
8	0	0	18	BD	10	0	0	56	BD

Table 4 Confusion matrix for RF

Using Test Set (90 instances)					10 Fold cross validation (264 instances)				
HB	IRD	ORD	BD	Classified as	HB	IRD	ORD	BD	Classified as
17	0	0	1	HB	56	0	0	10	HB
0	20	0	0	IRD	0	66	0	0	IRD
0	0	26	0	ORD	0	0	66	0	ORD
8	0	0	18	BD	11	0	0	55	BD

Table 5 Confusion matrix for RS

Using Test Set (90 instances)					10 Fold cross validation (264 instances)				
HB	IRD	ORD	BD	Classified as	HB	IRD	ORD	BD	Classified as
17	0	0	1	HB	54	2	0	10	HB
0	20	0	0	IRD	0	66	0	0	IRD
0	0	26	0	ORD	0	0	66	0	ORD
8	0	0	18	BD	19	0	0	47	BD

Table 6 shows the comparison of training efficiency or correctly classify instances (CC); incorrectly classify instances (IC), root mean square error (RMSE), and time of training in second. The ELM with test set gives 94.12% classification efficiency and time for training also less than the other ensemble techniques used in this study. Second best classification accuracy and training time have been achieved by rotation forest, which is 91.11% and 0.17*s*.

Table 6 Comparison of performance measure of ELM, rotation forest, RF and RS

Parameters	Test set (%)		Ten-fold (%)		RMSE		Time (s)
	CC	IC	CC	IC	Test set	10-fold	
ELM	94.12	5.88	93.22	9.78	0.1475	0.1375	0.015
Rotation Forest	91.11	8.89	92.04	7.95	0.1675	0.1669	0.17
RF	90	10	92.04	7.95	0.1778	0.1743	0.12
RS	90	10	88.25	11.74	0.1943	0.1946	0.03

5. Conclusions

In this study, extreme learning machine and ensemble techniques such as rotation forest, random forest and random subspace have proposed for the classification of rolling element bearing defects. An experiment have been conducted to acquire the vibration response of the healthy bearing, bearing with inner race defect, bearing with outer race defect and ball defect. Ten statistical features are extracted from the vibration signal which is used as input to machine learning techniques. Following conclusions are made base on the study:

(a) Extreme learning machine provide good learning efficiency in comparison to Rotational forest, random forest, and random subspace.
(b) Training time of ELM is smaller than the other used learning techniques.
(c) The used techniques also provide good results when the number of instances is small.

REFERENCES

1. Yang, Y., Yu, D. and Cheng, J. A Fault Diagnosis Approach for Roller Bearing Based on IMF Envelope Spectrum and SVM, *Measurement*, **40** (9-10), 943–950, (2007).
2. Lei, Y., He, Z. and Zi, Y. A New Approach to Intelligent Fault Diagnosis of Rotating Machinery, *Expert Systems with Applications*, **35** (4), 1593–1600, (2008).

3. Prabhakar, S., Mohanty, A.R. and Sekhar, A.S. Application of Discrete Wavelet Transform for Detection of Ball Bearing Race Faults, *Tribology International*, **35** (12)**,** 793–800, (2002).
4. Paya, B.A, Esat, I.I. and Badi, M.N.M. Artificial Neural Network Based Fault Diagnostics of Rotating Machinery Using Wavelet Transforms As a Preprocessor, *Mechanical System Signal Processing*, **11** (5), 751–765, (1997).
5. Purushotham, V., Narayanan, S.and Prasad, S. A. N. Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition, *NDT & E International,* **38** (8)**,** 654–664, (2005).
6. Vyas, N. and Satishkumar, D. Artificial neural network design for fault identification in a rotor-bearing system, *Mechanism and Machine Theory*, **36** (2), 157–175, (2001).
7. Kankar, P.K., Sharma, S.C. and Harsha, S.P. Fault diagnosis of ball bearings using machine learning methods, *Expert system with applications*, **38** (3), 1876–1886, (2011).
8. Kankar, P.K., Sharma, S.C. and Harsha, S.P. Rolling element bearing fault diagnosis using wavelet transform, *Neurocomputing*, **74**(10), 1638–1645, (2011).
9. Kavathekar, S., Upadhyay, N. and Kankar P.K. Fault Classification of Ball Bearing by Rotation Forest Technique, *Procedia Technology,* **23,** 187–192, (2016).
10. Sharma, A., Amarnath, M. and Kankar P.K. Novel ensemble techniques for classification of rolling element bearing faults, *Journal of Brazilian Society of Mechanical Science and Engineering*, (2016). doi:10.1007/s40430-016-0540-8.
11. Huang, G., Zhu, Q. and Siew, C. Extreme Learning Machine: A New Learning Scheme of Feed forward Neural Networks, *IEEE International Joint Conference on Neural Networks,* **2,** 985–990, (2004).
12. Huang, G. B., Zhu, Q. and Siew, C. Extreme learning machine: Theory and applications, *Neurocomputing*. **70** (1-3),489–501, (2006).
13. Huang, G.-B., Zhou, H., Ding, X. and Zhang, R. Extreme learning machine for regression and multiclass classification, *IEEE Transaction on System, Man and Cybernetics. Part B,* **42** (2), 513–529, (2012).
14. Rodríguez, J.J. and Kuncheva, L. I., C.J. Alonso, Rotation forest: A New classifier ensemble method, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28** (10), 1619–1630, (2006).
15. Breiman, L. Random Forests, *Machine Learning*, **45** (5), 1–35, (1999).
16. Ho, T. K. The random subspace method for constructing decision forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20** (8), 832–844, (1998).
17. Vakharia, V., Gupta, V. K. and Kankar, P.K. A multiscale permutation entropy based approach to select wavelet for fault diagnosis of ball bearings, *Journal of Vibration and Control*, **21** (16), 3123-3131, (2014).