ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 9, September 2015**

# Performance Analysis of Meta Classifiers Algorithms using Yeast Dataset

R.Nithya, P.Manikandan, Dr.D.Ramyachitra,

M.Phil Research Scholar, Dept. of Computer Science, Bharathiar University, Coimbatore, Tamil Nadu, India

Ph.D Research Scholar, Dept. of Computer Science, Bharathiar University, Coimbatore, Tamil Nadu, India

Assistant Professor, Dept. of Computer Science, Bharathiar University, Coimbatore, Tamil Nadu, India

**ABSTRACT:** The Classification technique calculates the categorical and prediction models to predict continuous valued functions. Usually, classification is the process of organizing data into categories for its most valuable and proficient use. The data classification technique makes important data that is easy to find and recover. In this paper the performance of three Meta classifiers algorithms namely Bagging, Decorate and Dagging are analyzed. The yeast dataset is used for estimating the performance of the algorithms by using the Training Set. Finally, the comparative analysis is performed by using the factors such as the classification accuracy and error rates on all algorithms.

**KEYWORDS:** Meta classification, Bagging, Decorate, Dagging, Training set

## I. INTRODUCTION

The Classification method is an important data mining feature which analysis the dataset with huge applications. It is used to classify each item in a set of data into one of predefined set of groups or classes. The Classification algorithm plays a critical role in text categorization. The plan of the classification system is to make a model in training dataset to calculate the class of potential objects whose class is not identified. Classification is a characteristic data mining technique based on machine learning. Hence, the aim of classification is to properly estimate the assessment of a designated discrete class variable, given a vector of attributes [1].

In this paper, the evaluation is made to find out which analysis option is the best for Meta classifiers algorithm by comparing three major algorithms namely: Bagging, Decorate and Dagging. In the analysis, selection process involves four kinds of parameter like supplied test set, training set, percentage spilt and cross validation. This paper considers for training set which is used to analyze the data set values. The Yeast dataset is used for evaluation of those algorithms.
The paper is organized as follows: Section 2 explains the literature review, Section 3 explains the methodology for the Yeast dataset and Section 4 explains the experimental result. Finally, Section 5 gives the conclusion and future work.

## II. LITERATURE REVIEW

Davidh. Wolpert et al., chooses bagging algorithm for estimating the generalization errors by using its techniques and argues that the algorithm is best in case of resultant estimation which gives often more accuracy while comparing with the conventional cross validation algorithms. More precisely the bagging algorithm is applicable only for small training data sets [2].

XingquanZhu et al., compares the traditional bagging (TB) with the proposed lazy bagging (LB) design for reducing the classification errors by building bootstrap replicate bags according to the characteristics of each test instances while investigating with 35 real world benchmark data sets. The comparison outcome shows the performance of both the approaches while LB performs better then TB for significant reduction of classification errors [3].
S.B. Kotsiantis et al., proposed model trees together with bagging for solving classification difficulty. Bagging is a type of regression technique for classification problems which consider the conditional class probability function and

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

searches for a model-tree approximation method. The model tree generates the greatest approximate probability values which are taken as predicted class during the classification. The performance comparison has been done with other model based decision tree approaches on standard benchmark datasets and proves the proposed approach is best in many cases [4].

Mittal C. Patel et al., proposed DECORATE the most popular ensemble learning method which can be used for strong learner to build diverse committees in a straightforward plan. The Artificial Neural networks (ANN) are very flexible with respect to finding the incomplete missing and noisy data and also makes the data to use for dynamic environment. The ANN is dependent on how best is the configuration of the net in terms of number of weights, layers and neurons. This paper presents DECORATE with ANN as a base classifier used to classify data from UCI repository. This experiment is conducted on the public datasets and the analysis results show that the DECORATE ensemble of ANN improves the performance of classification obviously [5].

Prem Melville et al., presents a new method for generating ensembles using DECORATE algorithm that directly constructs diverse hypotheses using additional artificially-constructed training examples. The technique is easy, common meta-learner that can use any strong learner as a base classifier to build diverse group. The experimental results uses decision-tree induction as a base learner for demonstrating that this approach consistently achieves higher predictive accuracy than both the base classifier and bagging. The DECORATE also obtains higher accuracy than boosting early in the learning curve when training data is limited [6].

Hafida Bouziane et al., investigates a meta-learning approach for classifying proteins into their various cellular locations based on their amino acid sequences. The meta-learner system based on kNearest Neighbors algorithm as base-classifier, has shown good performance in this context as individual classifier and DECORATES as meta-classifier using cross-validation tests for classifying Escherichia Coli bacteria proteins from the amino acid sequence information. This paper also reports a comparison against a Decision Tree induction as base classifier. The experimental results show that the kNN-based meta-learning model is more efficient than the Decision Tree-based model and the individual k-NN classifier [7].

Sanjay Kumar et al., proposed the combination of five supervised machine learning algorithms such as Classification and Regression Tree (CART) Logitboost, Dagging and Bagging Adaboost for classification of credit card information. The result shows the evidence for researchers to detect fraud in credit card. The experimental result also shows the performance analysis of different metalearning algorithms and also compared on the basis of misclassification and correct classification time. The smaller misclassification reveals that bagging algorithm performs better classification of credit card fraud detection technique [8].

Nida Meddouri et al., presents the supervised classification is a spot/task of data mining which consist of building a classifier from a set of instances labeled with their class and then predicting the class of new instances with a classifier. The supervised classification approaches were proposed namely Formal Concept Analysis and Induction of Decision Tree. This paper discusses about supervised classification based on Formal Concept Analysis and present methods based on concept lattice or sub lattice. The proposed new approach builds only a part of the lattice base on the concepts. The concepts are used as classifier in parallel combination using voting rule. Thus proposed method is based on Dagging of Nominal Classifier. Experimental results are given to prove the interest of the proposed method [9].

R. MahaLakshmi., reviews how to apply meta learning techniques to provide a comprehensive evaluation of different classification techniques in meta categorization. The number of cases classified correctly provides us with an estimate of the accuracy of the model. The main aim is to find highly accurate models that are easy to understand and achieve efficiency when dealing with large Datasets [10].

## III. METHODOLOGY

By using the Meta classification technique we find the best algorithm for the yeast dataset based on the training set. The flow diagram for the comparative analysis is shown in Fig 1.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 9, September 2015**

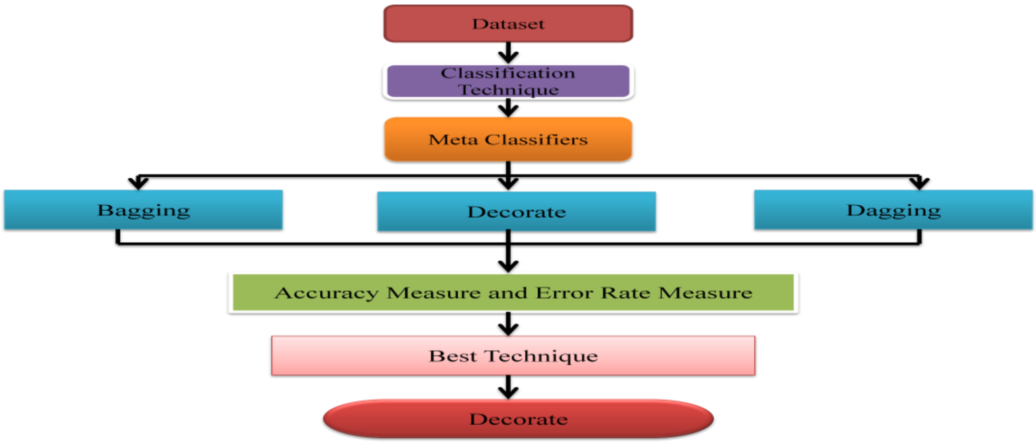


Fig 1: Flow diagram for comparative analysis of Meta classification technique.

*A. Dataset*

The Yeasts are atomic single-celled organisms belonging to the fungi -the taxonomic set that also includes mold and mushrooms [11]. Yeasts are very small, typically 5 to 10 microns (1 micron = 10-4centimeters) which is around 5 times the size of most germs [12]. The yeast datasets has been collected from the keel repository. This dataset contains 201 instance and 9 attributes. The data mining tool weka is used for analyzing the performance of the Meta classification algorithm.

*B. Classification*

The classification techniques collect the data into the classes on the source of their variation. Some of the classification methods or classifiers are the Neural Network Classifier, Naïve Bayes Classifier and so on. Every one of the technique create use of the learning algorithm that generates the model that best fits the relationship between the predictors and the prediction [13]. In this paper the Meta Classifiers algorithms are calculated to predict which of the algorithm is most suitable for the Yeast dataset. In the Meta classification technique three algorithms are compared that is Bagging, Dagging and Decorate to find out which one fits effectively for the Yeast dataset.

*C. Meta classifiers*

Meta is one of the classification methods. In this paper three Meta classification algorithms are used for finding the best algorithm for the Yeast dataset and they are as follows.

1. Bagging
2. Decorate
3. Dagging

*d. Bagging*

Bagging is well known re-sampling ensemble methods that generate and combine a diversity of classifiers using the same learning algorithm for the base-classifiers [14]. Bagging, a sobriquet for bootstrap aggregating is an ensemble method for improving unstable estimation or classification technique. Breiman motivated bagging as a variance reduction technique for a given base method, such as decision trees or methods that do variable selection and fitting in a linear type [15].

*A. Decorate*

DECORATE is a meta-learner for building diverse ensembles of classifiers by using specially constructed artificial training instance. Comprehensive experiments have demonstrated that this technique is consistently more accurate than the Random Forests, Base classifier and Bagging [16].

*B. Dagging*

This meta classifier creates a number of displace, stratified folds out of the data and feeds each chunk of data to a copy of the supplied base Techniques. Prediction is made via majority vote, while all the generated base classifiers are put

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 9, September 2015**

into the Vote meta classifier. It is useful for base classifiers that are quadratic or worse in time performance, about number of instances in the training set [17].

In this paper the experimental measures is evaluated by using the performance factors such as the classification accuracy and error rates. And also we find out the comparative analysis for the Yeast dataset to predict the best algorithm. The accuracy measure and the performance factors by class for the Meta classifiers is depicted in Table 1.

From the Table 1, it is inferred that Decorate Algorithm on Training Set, the TP Rate, FP Rate, Precision, Recall, F-Measure and ROC curve are higher than other two algorithms such as the Bagging and Dagging.

| Algorithms | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Curve |
|---|---|---|---|---|---|---|
| Bagging | 0.736 | 0.079 | 0.734 | 0.736 | 0.734 | 0.949 |
| Decorate | 0.886 | 0.041 | 0.888 | 0.886 | 0.885 | 0.978 |
| Dagging | 0.358 | 0.257 | 0.291 | 0.358 | 0.292 | 0.77 |

Table 1: Comparison of performance factors for Meta classifiers algorithms

| Algorithms | Correctly Classified | Incorrectly Classified |
|---|---|---|
| Bagging | 73.6318 % | 26.3682 % |
| Decorate | 88.5572 % | 11.4428 % |
| Dagging | 35.8209 % | 64.1791 % |

Table 2: Comparison of accuracy measure for Meta classifiers algorithms

| Algorithms | MAE | RMSE | RAE | RRSE |
|---|---|---|---|---|
| Bagging | 0.1009 | 0.2049 | 58.2024 | 69.7925 |
| Decorate | 0.0606 | 0.1498 | 34.9401 | 51.0366 |
| Dagging | 0.1794 | 0.2943 | 103.4594 | 100.2669 |

Table 3: Comparison of error rate measures for Meta classifiers algorithms

From the Table 2, it is inferred that the Decorate algorithm has higher classification accuracy than the other classification algorithms such as the Bagging and Dagging. From the Table 3, it is inferred that the Decorate classification algorithm has the lowest error rates than the other classification algorithms such as the Bagging and Dagging.

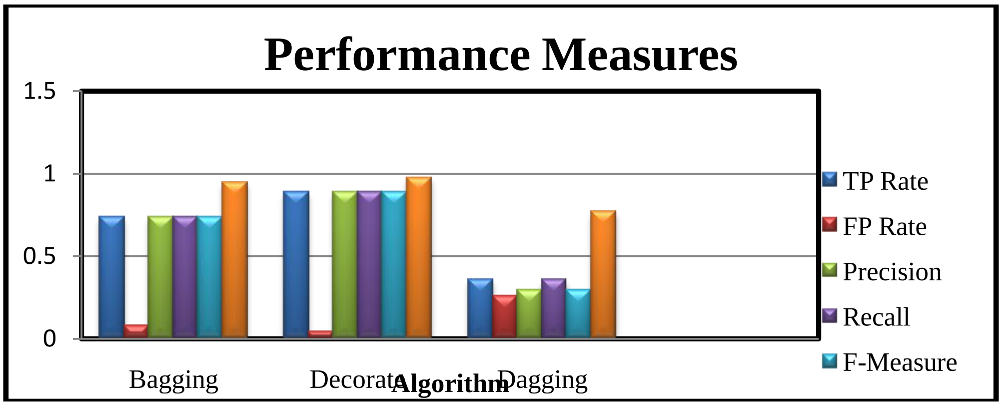


Fig 2: Comparison of performance factors for Meta classifiers algorithm

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 9, September 2015**

From the figure 2 it is inferred that the Decorate algorithm performs better than the other algorithms based upon the Performance factors such as the TP Rate, FP Rate, Precision, Recall, F-Measure and ROC Curve.

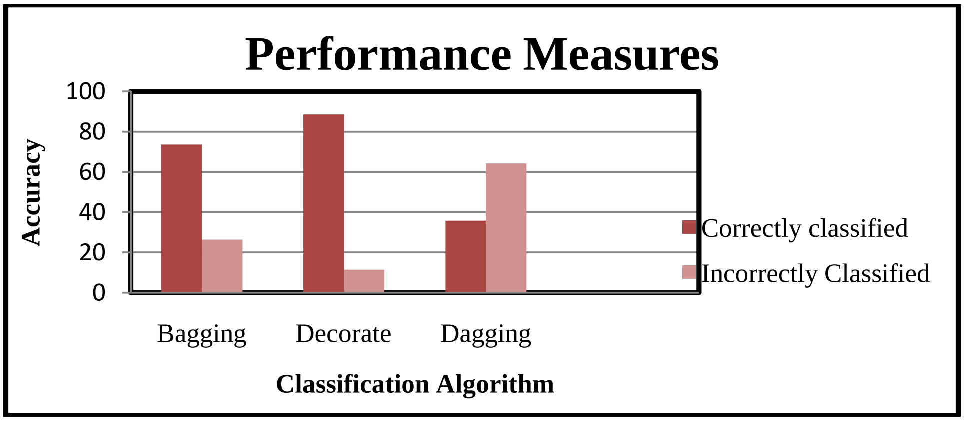


Fig 3: Comparison of accuracy measure for Bayes classifiers algorithms

From the figure 3 it is inferred that the Decorate algorithm performs better than the other algorithms based upon the Performance factors such as the correctly classified and incorrectly classified.

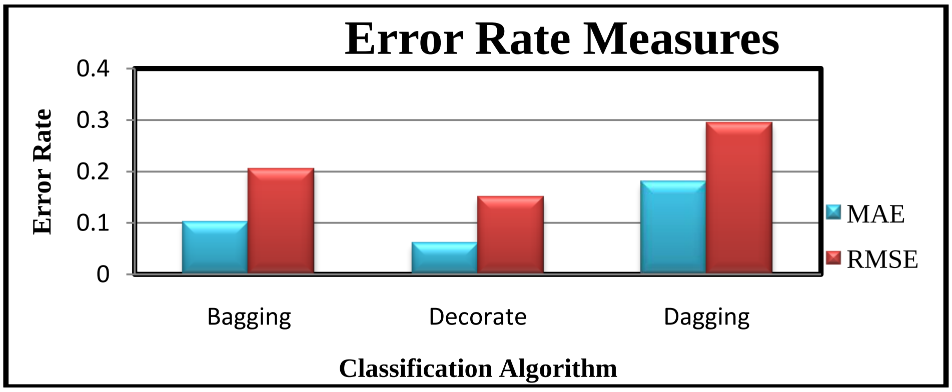


Fig 4: Comparison of error rate measures for Meta classifiers algorithms

From the figure 4 it is inferred that the Decorate algorithm performs better than the other algorithms based upon the Error Rate Measures such as the MAE and RMSE.

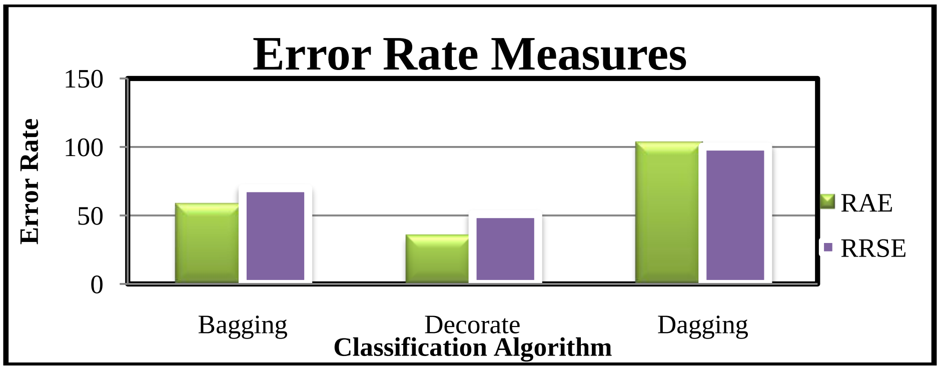


Fig 5: Comparison of error rate measures for Meta classifiers algorithms

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 9, September 2015**

From the figure 5 it is inferred that the Decorate algorithm performs better than the other algorithms based upon the Error Rate Measures such as the RAE and RRSE.

### *A. Statistical Analysis*

For Correctly Classified instances, it is inferred that Decorate algorithm performs 16% better than Bagging and 59% better than Dagging. Similarly for incorrectly classified instances it is inferred that Decorate algorithm performs 56% better than Bagging and 82% better than Dagging.

For TP rate, it is inferred that Decorate algorithm performs 16% better than Bagging and 59% better than Dagging for TP rate. For FP Rate it is inferred that Decorate algorithm perform 48% better than Bagging and 84% better than Dagging. For Precision it is inferred that Decorate algorithm is 17% better than Bagging and 84% better than Dagging. For Recall it is inferred that Decorate algorithm is 16% better than Bagging and 59% better than Dagging. For F-measure it is inferred that Decorate algorithm is 17% better than Bagging and 67% better than Dagging. For Roc curve it is inferred that Decorate algorithm is 2% better than Bagging and 21% better than Dagging.

For MAE, it is inferred that Decorate algorithm performs 39% better than Bagging and 66% better than Dagging. For RMSE it is inferred that Decorate algorithm performs 26% better than Bagging and 49% better than Dagging. For RAE it is inferred that Decorate algorithm performs 39% better than Bagging and 66% better than Dagging. For RRSE it is inferred that Decorate algorithm performs 26% better than Bagging and 49% better than Dagging.

## V. CONCLUSION AND FUTURE WORK

This paper evaluates the performance of three Meta classifiers algorithms namely Bagging, Decorate, Dagging. The Yeast datasets is used to analyze the performance by using training set based on the class attribute. The algorithms are analyzed based on the performance factors such as classification accuracy and error rates. From the experimental results, it is observed that the Decorate algorithm performs better than other algorithms. In the future, the Meta Classification algorithms can be experimented on other datasets to obtain more efficient results. Also the Meta classification algorithms can be calculated by using parameters such as the cross validation, percentage split, and supplied test set.

## REFERENCES

1. Mahendra Tiwari, et al., "Comparative Investigation Of Decision Tree Algorithms On Iris Data", International Journal of Advances in Computer Science and Technology, Volume 2, Page No: 30-35, March 2013, ISSN 2320 – 2602.
2. XingquanZhu, et al., "Alazybaggingapproachtoclassification", Pattern Recognition 41 (2008) 2980—2992.
3. Davidh.wolpert, et al., "An Efficient Method To Estimate Bagging's Generalization Error", Kluver academic Publisher, (1997)1-16.
4. S.B. Kotsiantis., et al., "Bagging Model Trees for Classification Problems", PCI 2005, LNCS 3746, pp. 328 – 337.
5. Mittal C. Patel., et al., "Decorate Ensemble of Artificial Neural Networks with High Diversity for Classification", International Journal of Computer Science and Mobile Computing, Volume 2, Page No:134-138, May 2013, ISSN 2320–088X.
6. Prem Melville., et al., "Creating Diverse Ensemble Classifiers", Pattern Recognition nov(2013)1-20.
7. Hafida Bouziane., et al., "Meta-Learning for Escherichia Coli Bacteria Patterns Classification", Proceedings ICWIT 2012, Page No: 1-150.
8. Sanjay Kumar Sen., et al., "Meta Learning Algorithms for Credit Card Fraud Detection", International Journal of Engineering Research and Development, volume 6, Page No: 16-20, March 2013, e-ISSN: 2278-067X, p-ISSN: 2278-800X.
9. Nida Meddouri., et al., "Parallel learning and classification for rules based on formal concepts", Procedia Computer Science 35 ( 2014 ) 358 – 367.
10. R. MahaLakshmi., "A Comparative analysis on persuasive meta classification strategy for web spam detection", IRACST - International Journal of Computer Science and Information Technology & Security, volume 2, Page No: 778-782, August 2012, ISSN: 2249-9555.
11. http://www.livescience.com/42140-what-is-yeast-candida.html
12. https://www.wyeastlab.com/he-yeast-fundamentals.cfm
13. Shabia Shabir Khan., et al., "Evaluation of Knowledge Extraction Using Various Classification Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering , Volume 3,pp. 251-256, June 2013, ISSN: 2277 128X. 12.
14. S. B. Kotsianti., et al., "Combining Bagging, Boosting and Dagging for Classification Problems", Knowledge-Based Intelligent Information and Engineering Systems, Volume 4693, 2007, pp 493-500.
15. http://sfb649.wiwi.hu-berlin.de/fedc_homepage/xplore/ebooks/html/csa/node229.html
16. http://wiki.pentaho.com/display/DATAMINING/Decorate
17. R. MahaLakshmi., "A Comparative analysis on persuasive meta classification strategy for web spam detection", IRACST - International Journal of Computer Science and Information Technology & Security, volume 2, Page No: 778-782, August 2012, ISSN: 2249-9555.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 3, Issue 9, September 2015**

**BIOGRAPHY**

**Ms. R. Nithya** is currently pursuing the M.phil degree in the Department of Computer Science, Bharathiar University, Coimbatore, India. She received the Master of Computer Science (M.Sc) in 2014 from the Bharathiar University. He has published papers in Conferences. Her research interests include data mining, and bio informatics.

**Mr. P. Manikandan** is currently pursuing the Ph.D. degree in the Department of Computer Science, Bharathiar University, Coimbatore, India. He received the Master of Computer Science (M.Sc) in 2013 and Master of Philosophy (M.Phil) in Computer Science degree from the Bharathiar University, in 2014. He has authored a book and also published papers in Conferences and Journals. His research interests include data mining, and bio informatics.

**Dr. D. Ramyachitra** obtained MCA from Madras University in 2000 and M.Phil in Computer Science in 2004 and Ph.D in Computer Science in 2013 from Bharathiar University. She has authored a book and also published several papers Conferences and journals. Her area of interest includes grid computing, image processing and bioinformatics.