CHAPTER 1

INTRODUCTION

1.1 GENERAL

Software quality is considered to be of great importance in the area of software engineering and development. In order to increase the efficiency and the quality of software modules, software defect prediction is used to identify defect prone modules and this helps in achieving high software reliability. Any software projects which possess numerous defects lacks quality and thus techniques and methodologies for predicting the defects aids to reduce the faultiness during software testing process which results in high quality software product. Prediction models with input as software metrics, can predict number of defects in software modules. Software metrics are attributes which includes process, product or source code metrics of the software system. There are many metrics that has proved their value for system maintenance and modification and hence selecting metrics plays a key role in software defect prediction. In this research work, the dataset that is made public by D’Ambros et al. is employed to evaluate bug prediction techniques (Marco D’Ambros et al. 2016; Marco D’Ambros et al. 2010). Basically, the considered dataset includes seventeen product based object oriented metrics collected at class-level for Java based Systems and another set of product metrics includes static code metrics that are available in Promise Repository (Menzies et al. 2012).
1.2 SOFTWARE DEFECT PREDICTION MODEL

Software fault prediction is always a complex area of research and software practitioners and researchers have carried out numerous ways to predict where the fault is likely to occur in the software module and their varying degrees of success. These prediction studies results in fault prediction models and it allows software personnel to concentrate on the defect free software code, thereby resulting in software quality improvement and employing the better utility of the resources. The international standard for evaluating the software quality is ISO/IEC 9126. Based on this ISO/IEC 9126 standard, the characteristics of software quality are with respect to internal and external metrics. The key characteristics include – efficiency, usability, reliability, maintainability, functionality and portability. Internal metrics focus only on the product itself without considering its behaviour, whereas, external metrics focus on the behaviour of the product. When software quality comes into picture, then Software Defect Prediction (SDP) plays a major role. Software is described to be of high quality when it is defect free. This research work mainly concentrates on the internal metrics of the system which includes the source code of software systems and not their functions or behaviour of the system.

It is to be noted that for the past two decades, several researchers focus on developing fault prone software’s as well identifying methodologies to detect the software affected by various types of defects. The prediction models developed by the researchers perform automatically for software defect prediction before carrying out the manual evaluation process. The developed predicted models should be more effective than that of the non – predicted models. Figure 1.1 shows the fundamental block diagram of the basic software defect prediction model.
1.3 IMPORTANCE OF SOFTWARE FAULT PREDICTION AND NEED FOR THE RESEARCH STUDY

In the growing and developed software engineering field, software quality is of high importance. Also, it should be noted that constructing software modules are highly expensive basically. As a result, to enhance the efficient resource allocation, efficiency and the general throughput of quality assurance and testing, it is highly important to predict software defects and assure that the developed software modules are defect free. The prediction of software defects at an early stage will make the company professionals to deliver a quality product to the end customers, as the cost incurred for the development play a major role. The following are the common problems observed in using existing predictor model for defect prediction:

- Defect predictors are noted to converge in a premature manner
- Occurrence of local and global minima during the prediction process
➢ Reliability not guaranteed
➢ Presence of stagnation during the progress of prediction
➢ High computational cost
➢ Interpretability of the system not ascertained

All the above mentioned are noted to be the open research problems in the existing predictor models. Thus, this research work developed neural network predictor models for performing software defect prediction which are reliable, adaptive and scalable that employs the advantages of proposed association rule mining, population – based evolutionary optimization algorithms and neural network architectures for improving the software defect prediction process. The developed models are noted to avoid the local and global minima problems, which were always an open research problem in the predictive models due the applicability of proposed evolutionary optimization approaches for tuning the weights of the developed neural network models. As well, these approaches guarantee on the reliability and interpretability of the system by decreasing the error rate. Thus this research work aims to address the occurred open research problems of the work noted.

The proposed models results in highest probability of software defect prediction rate by minimizing the mean square error for the considered NASA Promise repository datasets than the existing works due to the following:

➢ Convergence of the algorithms at an early duration with no compromise on the prediction rate
➢ Improved reliability of the system
➢ Locating an optimal solution at a feasible position
➢ Specified constraints being satisfied
Complete avoidance of local and global minima

Improving the accuracy and area under curve characteristics

Maintaining a balance between exploration and exploitation mechanism for tuning the weight values of neural network architecture models.

Minimizing the mean square error of the network models during prediction process

Minimizing the time complexity of the predictor model

In precise, this research work contributes to prediction application by developing effective, reliable and scalable proposed defect predictor models that are based on population based stochastic evolutionary algorithms and biological modeling of the human brain. The proposed defect predictor models are used to facilitate the software fault prediction process with complete guarantee on the specified set limits in an effective way for practical implications. Besides the contribution to software defect prediction research, this thesis also contributes to neural network modeling and evolutionary algorithm research. The effectiveness based on the defect predictor models are compared with few of the other existing and proposed algorithms based on the computed simulation results for its proper validation and prove its reliability.

### 1.4 CONTRIBUTIONS OF THE RESEARCH WORK

The major contributions of the research are:

i. A novel approach based on relational association rules with the combination of Naive Bayes approach for performing predictive defect classification. The Naive Bayes approach which is to be employed is simple but yet is powerful. The primary Naive Bayes assumption is that for a given class, the features are conditionally independent. Even
if the features are not independent, each feature is considered as independent in terms of how it contributes to the classification of the set.

ii. Proposed emotional ELMAN neural network (EENN) classifier and hybrid Gravitational Search Algorithm (GSA) – Charged System Search Algorithm (CSSA) to perform effective software defect prediction. The weights of the proposed EENN classifier is tuned employing the proposed hybrid GSA – CSSA approach and the accuracy is computed.


iv. A defect prediction approach based on Radial Basis Function Neural Network (RBFNN) and the novel Adaptive Dimensional Biogeography Based Optimization (ADBBO) model. The weights of the radial basis function neural network are tuned employing the novel ADBBO to compute optimal weights for the training process.

Each one of the developed defect predictors contributes a distinct methodology for performing effective and efficient prediction operation in a cooperative manner rather than a competitive manner. The result of the defect prediction rate and their learning and training performance employing these proposed predictive models are highly accurate in an intelligent manner. MATLAB mathematical software is used to carry out simulation of the proposed evolutionary algorithms and neural network architecture models. Numerical simulations are carried out and detailed results are presented in the respective chapters to analyze the performance and validity of all the proposed approaches. The simulated results computed using the proposed neural network predictor models are compared with the earlier developed models in the literature to prove their validity and that of accuracy rate.
1.5 ORGANIZATION OF THE THESIS

The thesis is organized into seven chapters including this discussed chapter 1. An outline of the forthcoming chapters is as given below:

Chapter 2 presents the literature review on the existing models for software fault prediction, various neural network architecture models, details on methodologies available for optimizing the neural network models and background on rule mining approaches.

Chapter 3 details a proposed predictor model employing relational association rules on the metrics data based on user defined confidence and support during training stage and integrates it with traditional Naive Bayes at testing stage to predict whether a software module is defective or non-defective.

Chapter 4 presents a novel approach based on proposed emotional ELMAN neural network (EENN) predictor and hybrid Gravitational Search Algorithm (GSA) – Charged System Search Algorithm (CSSA) to perform effective software defect prediction. Five public datasets are employed in this research contribution to validate the proposed hybrid GSA – CSSA based emotional ELMAN NN classifier model. The simulated results prove the effectiveness of the proposed software predictor model over the other methods available in the previous literature for the considered same datasets.

Chapter 5 software defect predictor model is developed by employing the versatile neural network algorithm – Back Propagation Neural Network (BPN) algorithm that employs gradient descent learning rule and another neural model – Kohonen Self Organizing Feature Maps (K-SOM) that employs clustering approach to perform the defect diagnosing process. The proposed approach is applied on NASA Promise repository datasets to validate its methodology.

Chapter 6 presents a prediction approach based on Radial Basis Function Neural Network (RBFNN) and the novel Adaptive Dimensional Biogeography
Based Optimization (ADBBO) model. The weights of the radial basis function neural network are tuned employing the novel ADBBO to compute optimal weights for the training process. The proposed approach is applied on NASA Promise repository datasets to validate its methodology. Simulation results prove that this approach outperforms all the other existing proposed predictor models available in the literature.

Chapter 7 elucidates the conclusion of the contributions made in the thesis work and as well the suggestions for future scope.

Appendix, list of references and a list of papers published based on this research work are provided at the end of this research thesis report.

1.6 SUMMARY

This chapter presented an introduction to the software defect prediction model and as well the need of fault prediction in developed software modules during the software development process. The chapter provides an in-depth validation and justification for the motivation towards this research work and the need of defect predictor models based on machine learning algorithms. This chapter further presented the major contributions towards the research problem undertaken in the area of software defect prediction and as well sketches the details on the outline of this thesis work.