ISSN (Print): 2321 -4058, (Online): 2321 -4392

THE INTERNATIONAL JOURNAL OF ADVANCES IN COMPUTER SCIENCE AND CLOUD COMPUTING (IJACSCC)

VOLUME NO. 11

ISSUE NO. 3

SEPTEMBER - DECEMBER 2023

ENRICHED PUBLICATIONS PVT. LTD

S-9, IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-47026006

ISSN (Print): 2321 -4058, (Online): 2321 -4392

The International Journal of Advances in Computer Science and Cloud Computing (IJACSCC)

Aim & Scope

The International Journal of Advances in Computer Science and Cloud Computing (IJACSCC) is a peer reviewed Journal in the field of Computer Science and Engineering. IJACSCC is an international forum for scientists, researchers and engineers involved in all aspects of Computer Science and Cloud Computing to publish high quality, referred papers. The Journal offers survey and review articles from experts in the field, promoting insight and understanding of the state of art, and latest trends in the field. The content include original research and innovation ideas, applications from all over the world. All published papers are also available freely with online full-text content.

ISSN (Print): 2321 -4058, (Online): 2321 -4392

ISSN (Print): 2321 -4058, (Online): 2321 -4392

International Journal of Advances in Computer Science and Cloud Computing

(Volume No. 11, Issue No. 3, September - December 2023)

Contents

ISSN (Print): 2321 -4058, (Online): 2321 -4392

Workload Multivariate Prediction by Vector Autoregressive and the Stacked LSTM Models

[1] **Soukaina Ouhame,** [2] **Youssef Hadi,** [3] **Fettah Akhiat,** [4] **El Hassan Elkafssaoui**
[1]MISC Laboratory, Ibn Tofail University, Kenitra Morocco
[2]LAGA Laboratory, Ibn Tofail University, Kenitra Morocco
[3,4]IMEMA Laboratory, Ibn Tofail University, Kenitra Morocco
E-mail: [1]soukaina.ouhame@uit.ac.ma, [2]hadi@uit.ac.ma

ABSTRACT

In cloud computing services, Infrastructure as a service (IaaS) is a service model that provides virtual computing resources in the form of hardware, networking, and storage services to the end users as needed in an elastic manner. However, cloud-hosting platforms introduce several minutes delay in the hardware resource allocation. The obvious solution to this issue is to predict the future need of computing resources and allocate them before being requested. This paper represents a hybrid method for predicting multivariate workload based on the Vector Autoregressive (VAR) model and the Stacked Long Short Term Memory (LSTM) model. In the proposed method, two metrics are used: CPU and memory usage, the VAR model is used to filter the linear interdependencies among the multivariate time series, and the stacked LSTM model to capture nonlinear trends in the residuals computed from the VAR model. The proposed hybrid model is compared with other hybrid predictive models: the AR-MLP model, the RNN-GRU model and the ARIMA-LSTM model. Results of experiments show superior efficacy of the proposed method over the other hybrid models.

Keywords- Cloud Computing, Multivariate Workload Prediction, Vector Autoregressive, VAR, Long Short Term Memory, Stacked LSTM.

I. INTRODUCTION

The most important benefits of cloud computing is the ability to provide flexible and fast IT resources on demand. Currently, the majority of cloud providers offer scalable services that automatically provide computer resources (such as CPU, memory, and storage). However, the scaling time mainly introduces a delay of several minutes, the start-up time to initialize a new VM. It is important to determine the exact amount of resources in advance to reduce scaling time.

Consequently, workload forecasting is the key solution to solve this problem.

Many prediction methods have been used in the literature, Jiang et al [1] explains that the behavior of web and data center workloads can be modeled through time series models. Chen et al [2] introduces a new prediction method using an enhanced fuzzy neural network method. Islam et al [3] uses Neural Network (NN) and Linear Regression (LR) algorithms and the sliding window technique in order to develop a new workload prediction strategy. Liu et al [4] presents a novel forecasting approach, which classifies the workloads and selects one prediction model among Support Vector Regression (SVR) model and Linear Regression (LR) model according to the workload features. Calheiros et al [5] uses the Auto Regressive Integrated Moving Average (ARIMA) model to predict cloud workload for Software as a Service (SaaS) providers. Zhang [6] proposes a hybrid approach to time series forecasting using both ARIMA and ANN models.

Recently, deep neural networks have been intensively used in workload prediction. D. Janardhanan and E. Barrett introduces a hybrid model, Long Short Term Memory neural network (LSTM) and ARIMA model [7] for CPU workload prediction.

In [8], the authors indicate that the LSTM model could solve the issues faced by cloud systems, as it is fragile and costly in the event of issues such as dynamic scaling of resources and energy consumption. The authors state that if it would be possible to determine the precise future workload of a server, resources can be adjusted according to demand and thus maintain both quality of service and reduce energy consumption.

In this paper, we propose a multivariate workload prediction model based on the Vector Autoregressive model and Stacked LSTM neural network. The Vector Auto Regressive model analyzes multivariate time series and predicts their future values, then the residues are calculated from the VAR model and used as inputs for the next stacked LSTM model, which simultaneously forecasts future resource values.

The contributions of this paper are listed as follows:

- We provide the fundamental definitions and necessary notions for building the Vector Autoregressive and LSTM models.
- We present the multivariate workload prediction algorithm.
- To conduct experiments and evaluate the proposed method, we use real world workload traces of GWA- T-12 Bitbrains service provider [9]. The paper is organized as follows. Section II presents the proposed method. The experimental results and analysis are shown in Section III. Finally, Section IV concludes the paper.

II. PROPOSED METHOD

In general, time series data is composed of 2 portions the linear and the nonlinear portion [6]. Thus, we can express as follows:

$$x_t = L_t + N_t + \varepsilon_t \quad (1)$$

L_t Represents the linearity of data at time t, while N_t signifies nonlinearity. The ε value is the error term.

The Vector Autoregressive (VAR) model is used for predicting multivariate time series. The model works decently on linear problems. Alternatively, the Stacked LSTM model can capture nonlinear trends in the dataset. Therefore, the two models are consecutively combined to comprise both linear and nonlinear tendencies in the model.

First of all, we introduce some background concepts of the two models. After that, we describe the proposed model for multivariate workload prediction in the cloud.

A. The Vector Autoregressive model

VAR models introduced by Sims [10] is a univariate model extension for predicting multivariate time series. The structure is that each variable is a linear function of past lags of itself and past lags of the other variables.

In this paper, we have to predict two time series (CPU and Memory usage) therefore the vector autoregressive model is as follows:

$$y_1(t) = a_1 + w_{11}y_1(t-1) + w_{12}y_2(t-1) + e_1(t-1) \quad (2)$$

$$y_2(t) = a_2 + w_{21}y_1(t-1) + w_{22}y_2(t-1) + e_2(t-1) \quad (3)$$

Where $y_1(t)$ and $y_2(t)$ are the CPU and Memory usage at moment t, $y_1(t-1)$ and $y_2(t-1)$ are the CPU and Memory usage at moment $t-1$ (here the lag value is 1)

a_1 a_2 Are the constant terms, w_{11}, w_{12}, w_{21} and w_{22} are the coefficients, and e_1 e_2 are the error terms.

Note that the two variables used have to be of the same order of integration ($I(0)$) stationary.

Before we can estimate a bivariate VAR model for the two series, we must specify the order p.

VAR order selection

The most common approach for model order selection involves selecting a model order that minimizes one or more information criteria evaluated over a range of model orders; Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) or Hannan-Quinn Criterion (HQC). In this paper, we resolve to use the AIC metric to estimate parameters.

$$AIC = -2\ln(\hat{L}) + 2k \quad (4)$$

The $\ln(\hat{L})$ notation is the value of the likelihood function, and k is the degree of freedom, that is, the number of parameters used. A model that has a small AIC value is generally considered a better model.

VAR model estimation

The Ordinary Least Squares (OLS) method or the Maximum Likelihood (ML) method can be used to estimate the parameters of VAR model. In this paper, we use The OLS estimator. The residual values are computed and entered to the subsequent LSTM model. As the VAR model has identified the linear trend, the residual is assumed to comprise the non- linear features.

$$x_t - L_t = N_t \quad (5)$$

B. The Long Short Term Memory model

Neural Networks are known to perform well on nonlinear tasks. Because of its versatility due to large dimension of parameters, and the use of nonlinear activation functions in each layer, the model can adapt to nonlinear trends in the data.

To understand the LSTM model, the mechanism of Recurrent Neural Networks (RNN) should first be discussed. The RNN is a type of sequential model that performs effectively on time series data. It takes a sequence of vectors of time series data as input X = [x_1 , x_2 , x_3 ,......,x_n] and outputs a vector value computed by the neural network structures in the model's cell, symbolized as A in Fig. 1, Vector X is a time series data spanning t time periods. The values in vector X is sequentially passed through cell A . At each time step, the cell outputs a value, which is concatenated with the next time step data, and the cell state C . The output value and C serve as input for the next time step. The process is repeated up to the last time step data.

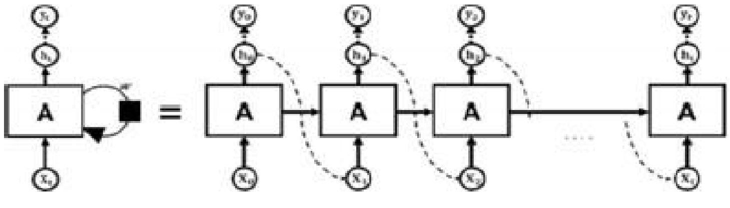


Fig. 1 Structure of Recurrent Neural Network

The A cell in Fig. 1 can be substituted with various types of cells. In this paper, we select the standard LSTM cell with forget gates, introduced by F. Gers et al. (1999) [11]. The LSTM cell paper comprises four interactive neural networks, each representing the forget gate, input gate, input candidate gate, and the output gate. The forget gate outputs a vector whose element values are between zero and one. It serves as a forgetter that is multiplied to the cell state C_{t-1} from the former time step to drop values that are not needed and keep those that are necessary for the prediction.

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \qquad (6)$$

The σ function, also denoted with the same symbol in Fig.2, is the logistic function, often called the sigmoid. It serves as the activation function that enables nonlinear capabilities for the model.

$$\sigma(X) = \frac{1}{1-e^{-x}} \qquad (7)$$

In the next phase, the input gate and the input candidate gate operate together to render the new cell state C_t, which will be passed on to the next time step as the renewed cell state. The input gate uses the sigmoid as the activation function and the input candidate utilizes the hyperbolic tangent, each outputting it and C'_t. The it selects, which feature in C' should be reflected in to the new cell state C_t.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \qquad (8)$$

$$C_t = \tanh(W_c.[h_{t-1}, x_t] + b_c) \qquad (9)$$

The tanh function in Fig.2 is the hyperbolic tangent. Unlike the sigmoid, which renders value between zero and one, the hyperbolic tangent outputs value between -1 and 1.

$$\tanh(X) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \qquad (10)$$

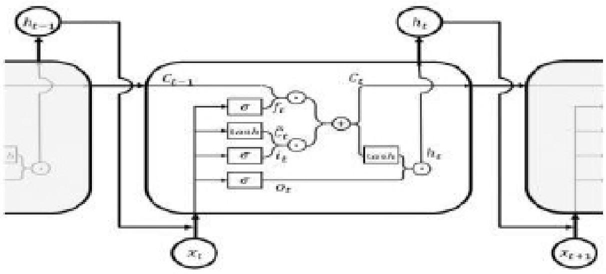


Fig.2 Inner structure of a Long Short-Term Memory cell

Finally, the output gate decides what values are to be selected, combining o_t with the tanh applied state C_t as output h_t. The new cell state is a combination of the forget-gate applied former cell state C_{t-1} and the new tanh applied state C_t.

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_0) \qquad (11)$$

$$C_t = f_t.C_{t-1} + i_t.C'_t \qquad (12)$$

$$h_t = o_t.\tanh(C_t) \qquad (13)$$

The cell state C_t and h_t output will be passed to the next time step, and will go through a same process.

Stacked LSTM architecture can be defined, as an LSTM model comprised of multiple LSTM layers.

An LSTM layer above provides a sequence output rather than a single value output to the LSTM layer below (Fig.3).

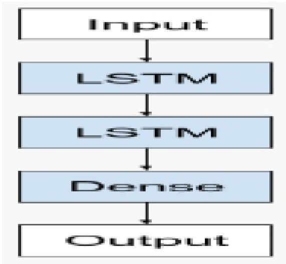


Fig.3 Stacked Long Short-Term Memory Architecture

C. The proposed VAR-LSTM model

Algorithm 1: VAR model fitting algorithm

Input: Y1: CPU usage time series
Y2: Memory usage time series
P_{max}: the max lag order
Output: Residual: residual values of the two series

1. If Y1 and Y2 are not stationary then
2. Y1=Difference (Y1)
3. Y2=Difference (Y2)
4. End If

Grouping the two time series into an unique multivariate time series Y

5. Y= concatenate (Y1, Y2)

Convert into input/output with the percentage of 80%

6. Train, Test = divide (Y, 0.8)
7. Select_order (maxlags = Pmax)
8. P_lag = order lag with least AIC value
9. Mfit = fit VAR (P_lag)
10. for all data in Test do
11. Residual = Test - predict (Train, Mfit)
12. Return Residual.

The input of the algorithm are the CPU and Memory usage time series, which are formed using the historical data of the workload. The stationarity of each time series is checked by using the augmented Dickey–Fuller (ADF) test; if they are not stationary, we differentiate them.

In the following step, we select the lag order with the t t least value for model fitting, then we compute the residual values.

Algorithm 2: LSTM model training algorithm

Input: Residual: Residual values of the VAR model
N_step: the lag step between each input and output
Output: trainPred, testPred: the predicted train and test data of the multivariate time series.
{Phase1: Data preprocessing}

1. Normalize Residual data

```
        Convert into input/output with the percentage of
80%
2.    train_LSTM, test_LSTM = divide (Residual, 0.8)
3.    X_train, y_train = split (train_LSTM, N_step)
4.    X_test, y_test = split(test_LSTM, N_step)
```

III. RESHAPE TRAIN AND TEST INPUTS DATA

```
{Phase2: Determine model parameters}
6.    Define model
7.          add LSTM (units = 100, activation = 'relu',
input_shape = (N_steps, n_features))
8.        add LSTM (units = 100, activation = 'relu')
9.        add Dense (n_features = 2)

{Phase3: Model fitting & estimation}
11.     Repeat
12.         Forward_propagate model with X_train
            Backward propagate model with y_train
13.         Update model parameters
14.      MSE, MAE=evaluate_model(X_train,y_train)
15.        If MSE converged:
16.            End Repeat
17.    MSEt, MAEt =evaluate_model (X_test, y_test)

{Phase4: LSTM model prediction}
18.        trainPred = predict (X_train)
19.            testPred = predict (X_test)
20.    Return trainPred, testPred
```

The LSTM prediction algorithm works in four main phases : Data preprocessing, fixing model parameters, Model fitting and estimation, and model prediction.

As cited before, the residual values calculated by the algorithm1 are entered to the LSTM model. Since LSTMs are sensitive to the scale of input data, the residual data must be normalized, and then the data is divided into train and test data. The LSTM input layer must be three dimension, then the input data are reshaped into a three-dimensional array with three time steps for each sample, and two features at each time step.

We used a stacked LSTM network with 100 LSTM blocks (or neurons) where the number of timesteps and parallel sets (features) is specified for the input layer via the input_shape argument.

The ReLu (Rectified Linear unit) activation function is used for LSTM blocks.

$$f(x) = x^{+} = \max(0, x) \qquad (14)$$

Where x is the input to a neuron.

ReLu is less computationally expensive than tanh and sigmoid because it involves simpler mathematical operations. Moreover, the problem of vanishing gradient can be greatly reduced using the ReLU family of activation functions.

The cross-validation tests the model's performance to predict new data that was not used in estimating it, in order to flag problems like overfitting or selection bias [12]. We have used a validation split of 15%.

ADAM optimization algorithm is used for stochastic gradient descent for model training.

The network is trained for 200 epochs with batch size of 1.

IV. EXPERIMENTAL EVALUATION

A. Experiment dataset

The dataset contains the performance metrics of 1,750 virtual machines in a Bitbrains distributed data center, a service provider specializing in managed hosting and business computing.

Each file contains the performance metrics of a virtual machine. These files are organized according to traces: fastStorage and Rnd.

In this paper, we choose the first trace; fastStorage: the trace consists of 1250 virtual machines connected to storage area network (SAN) devices. Each file consists of a set of lines; each line represents an observation of the performance metrics of a virtual machine every 300 milliseconds since 1970-01-01. We have selected two columns "CPU usage" and "Memory usage" and 2800 observations for our workload prediction model.

Fig.4 presents the CPU usage and Fig.5 the Memory usage.

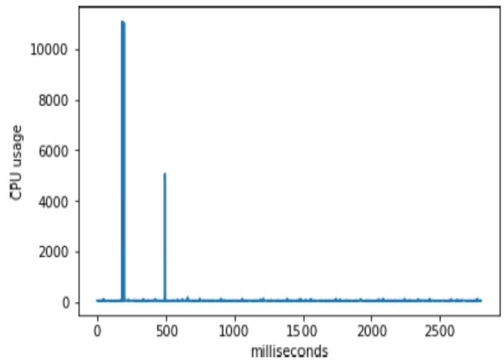


Fig.4 CPU usage by milliseconds

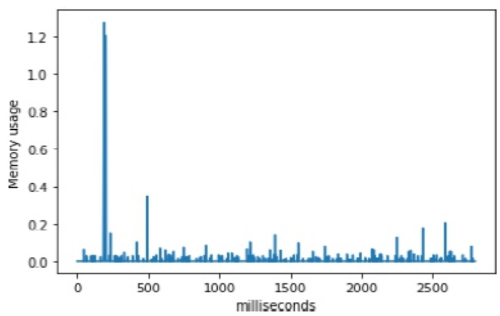


Fig.5 Memory usage by milliseconds

B. Analysis of the results

The resulting time series Y1 and Y2 are stationary, then, we do not need to perform a difference operation.

In the first step, 80% of the multivariate time series Y is used for training and 20% for testing as indicated in the algorithm1.

By choosing Pmax = 4 as the max lag order we have the following table (Table 1).

TABLE I VAR MODEL ORDER SELECTION

	AIC	BIC
0	38.60	36.61
1	36.31	36.32
2	36.05	36.08
3	36.00*	36.04*
4	36.00	36.05

Where BIC stands for Bayesian Information Criterion, which also estimates the quality of a model.

As the third lag order has the least value of AIC then it will be used for fitting our VAR model.

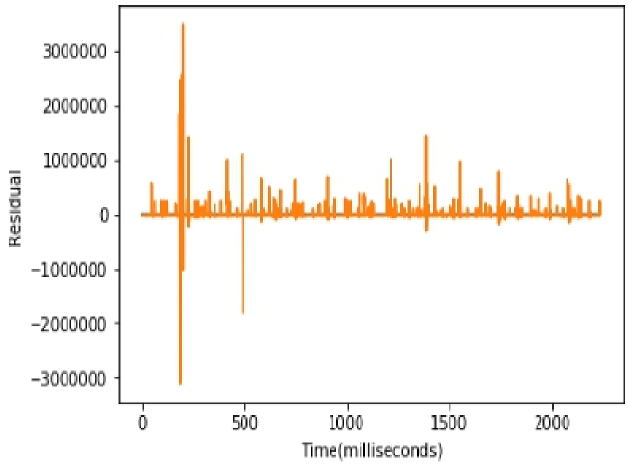


Fig.6 Residuals of VAR model

The data points are generally around 0, as the input is a residual dataset (Fig. 6).

In the second step, 80% of residual data is used for LSTM training and 20% for testing.

To evaluate our model, the Mean Squared Error (MSE), Root Mean Squared Error (RMSE) values, and the Mean Absolute Error (MAE) values of the prediction were calculated.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y'_i)^2 \quad (15)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y'_i\right| \quad (16)$$

Where y_i the output value and y'_i the predicted value The MSE learning curve of both of the train and test data are close to each other, the same for the MAE learning curve (Fig. 7, 8).The MSE values of train and test data have small variations, which means the model has been generalized adequately (Table II) 2).

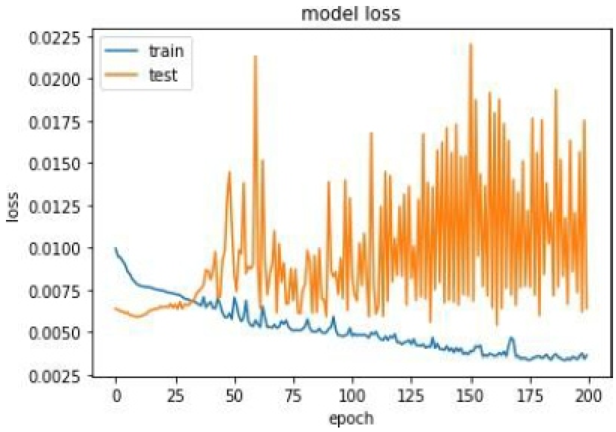


Fig.7 MSE learning Curve

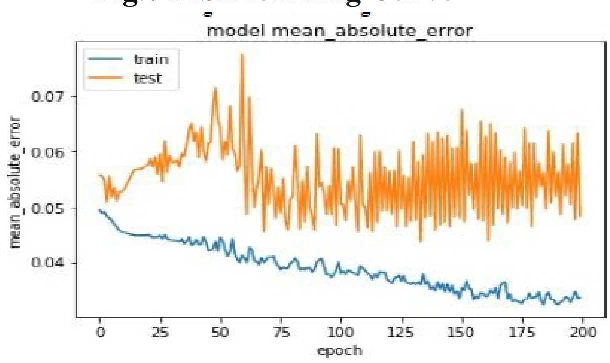


Fig.8 MAE learning Curve

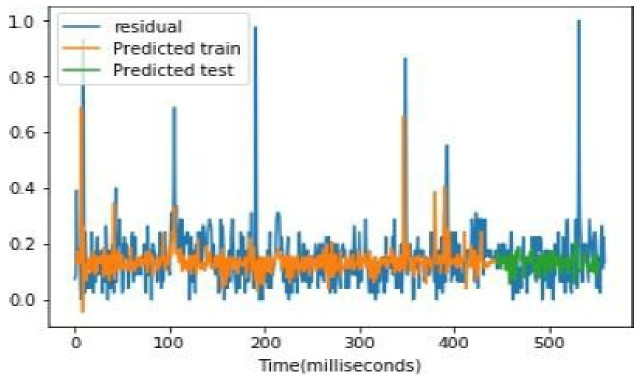


Fig.9 Predicted train & test data of CPU usage

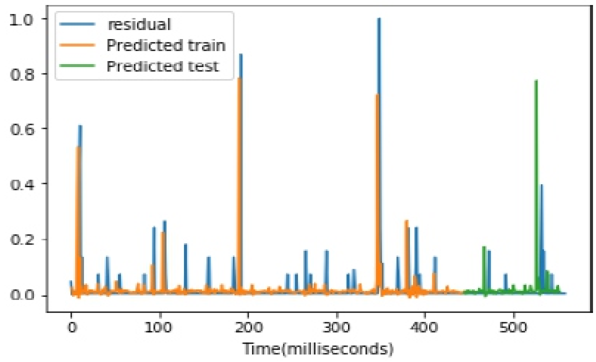


Fig.10 Predicted train & test data of Memory usage

The proposed model indicates lower values of MSE and MAE compared to the ARIMA-LSTM model, the RNN-GRU and the AR-MLP models.

The RMSE value is much lesser in ARIMA-LSTM model. However, the proposed model still also shows inferior values of RMSE compared to the remaining models (Table II).

TABLE II 1VAR-LSTM MODEL PERFORMANCE RESULTS AND ITS COMPARISON

	The propose d model	ARIMA-LSTM (CPU prediction)	RNN-GRU	AR-MLP
MSE-train	0.00420	0.00832	0.00634	0.00464
MSE-test	0.00760	0.00840	0.01130	0.01130
MAE-train	0.03663	0.04404	0.05031	0.04155
MAE-test	0.04336	0.04296	0.06201	0.05413
RMSE-train	0.06482	0.06575	0.07964	0.06812
RMSE-test	0.08721	0.06633	0.10632	0.10633

V. CONCLUSION

An important feature of cloud computing is the ability to determine allocated resources based on actual usage. However, this operation requires a start- up time for resource allocation. In order to reduce this time, it is essential to plan in advance the amount of resources needed for the future.

In this paper, we adopted the VAR-LSTM hybrid model for multivariate workload prediction in an attempt to first filter out linearity in the VAR modeling step, then predict nonlinear tendencies with the LSTM recurrent neural network.

The proposed approach was tested using actual data from Bitbrains data. The results are positive and show that the proposed method is more effective than the other hybrid model .

REFERENCE

[1] Jiang, Y. et al. 2013. Cloud analytics for ca1pacity planning and instant VM provisioning. IEEE Transactions on Network and Service Management. 10, 3 (2013), 312–325. DOI:https://doi.org/10.1109/TNSM.2013.051913.120278

[2] Chen, Z. et al. 2015. Self-adaptive prediction of cloud resource demands using ensemble model and subtractive- fuzzy clustering based fuzzy neural network. Computational Intelligence and Neuroscience. 2015, (2015). DOI:https://doi.org/10.1155/2015/919805.

[3] Islam, S. et al. 2012. Empirical prediction models for adaptive resource provisioning in the cloud. Future Generation Computer Systems. 28, 1 (2012), 155–162. DOI:https://doi.org/10.1016/j.future.2011.05.027.

[4] Liu, C. et al. 2017. An adaptive prediction approach based on workload pattern discrimination in the cloud. Journal of Network and Computer Applications. 80, (2017), 35–44. DOI:https://doi.org/10.1016/j.jnca.2016.12.017.

[5] Calheiros, R.N. et al. 2015. Workload prediction using ARIMA model and its impact on cloud applications' QoS. IEEE Transactions on Cloud Computing. 3, 4 (2015), 449– 458. DOI:https://doi.org/10.1109/TCC.2014.2350475.

[6] Peter Zhang 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 5, (2003), 159–175. DOI:https://doi.org/10.1016/S0925- 2312(01)00702-0.

[7] *Janardhanan, D. and Barrett, E. 2018. CPU workload forecasting of machines in data centers using LSTM recurrent neural networks and ARIMA models. 2017 12th International Conference for Internet Technology and Secured Transactions, ICITST 2017. (2018), 55–60. DOI:https://doi.org/10.23919/ICITST.2017.8356346.*

[8] *Kumar, J. et al. 2018. Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model for Cloud Datacenters. Procedia Computer Science. 125, (2018), 676–682. DOI:https://doi.org/10.1016/j.procs.2017.12.087.*

[9] *The Grid Workloads GWA-T-12 Bitbrains: http://gwa.ewi.tudelft.nl/datasets.*

[10] *Sims, C.A. 1980. Macroeconomics and Reality. Econometrica. 48, 1 (1980), 1. DOI:https://doi.org/10.2307/1912017.*

[11] *Cummins, F. et al. 2000. Learning to Forget: Continual Prediction with LSTM. Neural Computation. June 2016 (2000). DOI:https://doi.org/10.1197/jamia.M2577.*

[12] *Cawley, G.C. and Talbot, N.L. 2010. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. Journal of Machine Learning Research. 11, (2010), 2079–2107. DOI:https://doi.org/10.1016/j.biopha.2003.08.027*

ISSN (Print): 2321 -4058, (Online): 2321 -4392

Enhanced Encryption Approach for Data Security and user Validation using AES and Blowfish Algorithm in Cloud

[1] Rajni Kumari, [2] Anupama Gupta, [3] Kaur Baldip

[1]Computer Science and Engineering Department CT Group of Institutions, Shahpur Campus, Punjab, India

[2]Computer Science and Engineering GZSCCET, Punjab,India

3Computer Science and Engineering Department LLRIET, MOGA,Punjab,India

E-mail: [1]rjni786@gmail.com, [3]baldip20@gmail.com

ABSTRACT

Cloud computing is a process that enables its users to access numerous services which are Saas, Paas and Iaas. Many of the individuals who use the framework to offer their services to different users usually try to manipulate and manage their data. One of the challenges in cloud computing is that the data that has been accessed with the aid of different resources need the access of the user on the access control model purposely for the integrated management. The issue of security has been of major concern in cloud computing and due to this, many researchers have decided to pursue the topic in detail. Numerous security issues come up as a result of outlining the importance of cloud computing.

Keywords - Cloud Computing, Security, OTP, Saas, Paas, Iaas.

I. INTRODUCTION

Cloud computing is primarily used to offer many services and also act as a platform for computer applications. The services that are provided by cloud computing has been used in several fields such as personal and business purposes. In many times the individuals who are using the cloud computing do not have the knowledge of where the software is located as well as its properties. In our case, the data that is used has been stored in a form that can be processed through the internet. Cloud computing examples include; web mails, social network sites, online business application and many cloud storage applications. The amount of information from each work area or from various gadgets is normally put aside in distributed computing because the information cannot be found on the web. The use of cloud computing does not put into consideration the handling that has been made for the end user level that all is in line with the happenings that the utilization information of the applications for mainly customer loyalty and developing of associations.

1.1 Deployment of cloud services

Many of the cloud services that can be used arise from different cloud practices. The services are beneficial from the private cloud, community cloud, public cloud, and hybrid cloud. The communication, email services and the online media storage services which are some of the cloud providers operate and provide services through the public cloud. Different services which can be found on enterprises can be utilized by the public cloud. Some of the cloud deployment models include:

1.1.1 Public Cloud: This type of cloud is mainly used by different services by use of various resources which can be found in different locations. The services have been offered by the third-party providers that are rated by checking their quality and the number of individuals using them.

1.1.2 Community Cloud: This is a type of cloud that is used by different organizations that need a framework which is the same. They normally use it in putting in order their products. The model requires the companies to develop a strategy that allows them to share their resources. Even though community cloud is expensive than public cloud, it provides high security, better services and privacy for the individuals who use it.

1.1.3 Hybrid Cloud: The cloud involves the use of services that have been provided by two different clouds but for the purpose of serving a single company. Hybrid cloud involves the combination of two cloud to perform the required services. There is the bringing together of both the public and the private could form a single cloud that can be used to render better services. The cloud is primarily used so as the virtual cloud services can be deployed on the physical hardware units.

1.1.4 Private Cloud: This type of cloud is mainly used by large firms that need high security and data storage for the different applications of the firm. The cloud is mainly custom made to fit the requirements of the company that is using it. The software and the hardware of the cloud can be upgraded or degraded to fit the demands of the organization.

1.2 Challenges in Cloud Computing: The performance of cloud computing can be affected by numerous challenges. Some of the challenges are:

a. availability
b. data security and integrity
c. complexity
d. authentication and Validation

1.3 Major Risks of cloud computing security: Cloud computing is a large network that is used in carrying out various cloud services. There are different users who use the cloud services. Cloud computing has different security problems such as application security, data availability and integrity authentication and validation and data encryption. Since cloud computing can be accessed remotely by use of the internet, security issues arise that put in danger the cloud resources up to the extent that the integrity and authentication are questioned. The issue of security in cloud computing is something that is new and due to this, a lot of research can be done so as to provide the best security for the framework. The following are cloud computing issues in security;

i. Data Integrity
ii. Proper usage metering
iii. Abusing the cloud's resources
iv. Security attacks
v. Cloud Services Availability
vi. Legal and Regulatory Risks
vii. Data Leakage and data security

1.4 Important Security Issues in the Cloud:

The development in technology has made it easier for many service providers to attract more customers by use of various resources making them to easily get their information from anywhere by use of the internet. Cloud computing is the best place where users can find the best services. Due to the high-security risks that are associated with cloud service, many users fear using the platform. Some of the security issues that are linked with cloud data include;

Integrity: one of the major factors that play an important role in data storage in the cloud service is integrity. The main purpose of the integrity in cloud service is to ensure that data which is sent from the cloud storage is not corrupted in any manner. This means that no alterations have been made on the data by an individual who has not been authorized. Since there can be a system failure, the data that has been worked on can be stored in a portable gadget which makes its retrieval easy.

Availability: the main role of availability is to make sure that any resources that have been used in cloud computing can be utilized for the services and ensures that there are not attacks. This means that when an individual who uses the system tries to get any information from the cloud server, it must be available for it to be accessed.

Confidentiality: the main role of confidentiality is to make sure that data is not accessed by unauthorized individuals. In situations where unauthorized people can have access to information then confidentiality would have been lost.

This can manifest itself either electronically or physically. Social engineering is the main way through which physical confidentiality can occur. In circumstances where the customers and any other users do not encrypt their communications, electronic confidentiality occurs.

II. REVIEW LITERATURE

2.1. Jongo Mun et al,[1] "An Improvement of User Authentication Framework for Cloud Computing" one of the most recent innovation give the various administrations to the customers by use of the internet is cloud computing. The users have the ability to store and even access their data from any location in the world through the internet. This study tries to show an approach that can be used to give security to cloud services by use of customer confirmation model. User authentication has been given by taking into consideration the cryptography model that involves different mathematical formulas with an aim of ensuring that the information of the person using the software cannot be accessed easily. According to the author of the paper, user authentication gives a high security.

2.2. Ahmed Dheyaa Basha et al [2] "Mobile Applications as Cloud Computing: Implementation and Challenge" since cloud computing and other versatile application have been linked with each other, various administrations that are related to cloud condition can be used by use of the portable application. There are many portable applications that are used mainly for cloud security. The paper outlines the administrations in detail. Also, various types of challenges have been dealt with in this paper.

2.3 Uma Somani et al. [3] The author shows how the individuals who use cloud computing have the validity of using different cloud computing services which can be found in the cloud environment. The main aim of this is to ensure that cloud services are utilized in the best manner. Some of the programming issues and those of different equipment have been solved by cloud computing. Also, the problem of pilling of stock in various organizations has been solved by cloud computing. This study outlines how the issue of information stockpiling has been handled by the use of the encryption model. This is done with the main aim of ensuring that the information that has been put away in the distributed location is done in a scrambled manner. Putting away of information through scrambling utilizes the computer signature calculations and mostly the Rivest, Adi Shamir (RSA) calculation. According to the calculations, utilization of hashing has been done with the aim of 0-changing the information by taking into consideration 128- piece key. Other 16-sub keys have been created from the 128-piece key which is used to change plain information to figure information. This creates an arrangement that offers the best security for the information that has been moved from the cloud storage.

2.4 Ru Wei Huang et al. [4] “Design of Privacy- Preserving Cloud Storage Framework" proposed Privacy security is a key issue for distributed storage. Encryption is an entrenched innovation for ensuring delicate information. Be that as it may, it gives some new issues: how information proprietor and capacity specialist organization to work on encoded information? How to decrease information proprietor's outstanding task at hand of information administration and bolster information sharing in the meantime?

In the event that they can't be anything but difficult to settled, encryption will convey a great deal of inconveniences to individuals. In this paper, we structure a security safeguarding distributed storage system to tackle those issues. We plan a communication convention among members, receive key inference calculation to create and oversee keys, utilize the mix of symmetric and lopsided encryption to shroud the touchy information of clients, and apply Bloom channel to acknowledge figure content recovery.

A framework dependent on the structure is figured it out. The paper breaks down the achievability of the structure from the execution of Bloom channel, the running overhead of the framework and the protection security of the system.

2.5 Farzad Sabahi. Et al [6] “Research on Privacy- Preserving Cloud Storage Framework Supporting Cipher Text Retrieval” The author of this study tries to give more information on the problem of stockpiling mainly in the distributed computing stockpiling. Security plays an important role in cloud computing because of the capacity of information from all across the world. The author of this study shows various problems that have been experienced by clients who have decided to transform their businesses' information to cloud foundation. Some of the security problems that have been handled in this paper include; assaults, trustworthiness and accessibility. The writer of the paper outline various ways that can be used to deal with the identified security issues. According to the plan of handling the issues, information security has the ability to accommodate media that has been stored in cloud servers. .

III. RESEARCH METHODOLGY

In this section, cloud computing save channel has been utilized specifically for the user and data verification. There are numerous stages that have been developed to help in user registration in cloud computing. The hybrid credential is used in processing the personal information of users during the registration procedure. The credential has been stored in the system so that it can be processed later by use of the hybrid encryption model. Each user is given a confidential key to ensuring that the information is not accessed by unauthorized people. The initialization of user login is completed after the registration process. Initialization of the user's login is done through the verification of the one-time key validation procedure. The login of the user's credentials has a secret key which is given to the user. The use of the key verifies the user and the one-time key is sent for verification. When the login process is over, the information that has been stored by the user on the distributed storage has been moved and the information has been encoded which is one way of encryption. The main aim of doing the entire process is to ensure that there is a high level of information privacy.

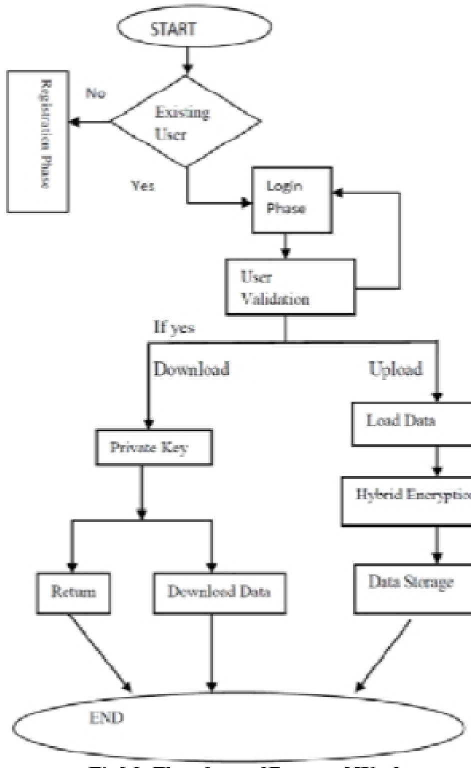


Fig3.1. Flowchart of Proposed Work

IV. RESULTS AND DISCUSSION

Sr, No	Files (in bytes)	Fully Homomorphism	
	Time	(In ms)	DES
1	2681	132	118
2	3431	165	142
3	4386	285	256
4	5546	365	343
5	22588	606	555

Table 1: Computation time for encryption of different size files.

The table below shows the value for computation duration pf various size files. Fully homomorphism and DES encryption models have been evaluated for the computation time for the intended job.

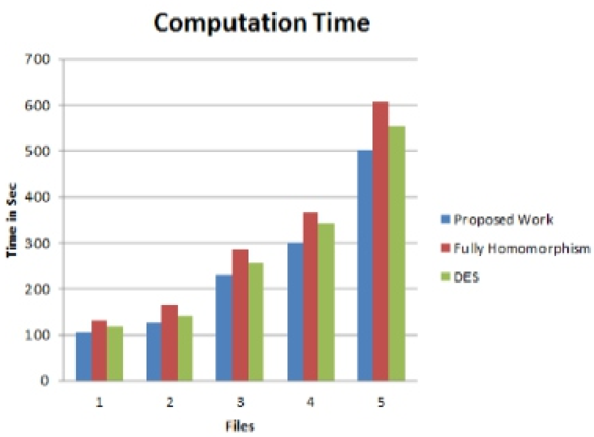


Fig.4.2.Compression graph for computation time for encryption

The above figure is a comparison graph for the computation time of the intended approach and fully homomorphism and DES model.

Sr, No.	Files (in bytes)	Proposed Work	Huffman Coding	RLE
		Compressed files size (In bytes)		
1	2681	1910	2346	2576
2	3431	2430	2995	3068
3	4386	3230	3964	4018
4	5546	4168	4864	5264
5	22588	14044	17048	19428

Table2: Compressed File Size after compression of data of various sizes

The table above is a representation of the value of the compressed files for the encryption of various size of files. The compressed values were evaluated mainly for work and fully homomorphism and DES encryption model.

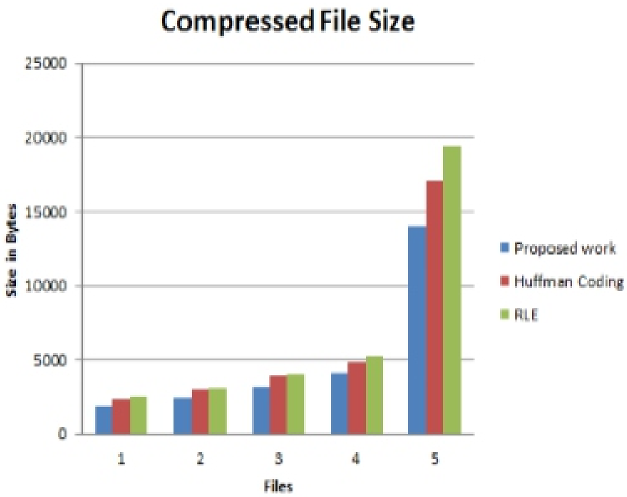


Fig 4.3 Comparison graph for compression file size after compression.

The figure above is a representation of the comparison graph for compressed file size by using proposed approach and Huffman coding and RLE compression model.

V. CONCLUSIONS

1. Data storage has been done on the cloud for the given work. Each user has been given a different name and different email ID for security purposes. The request by the user to access the system is transmitted to the cloud whereby the secret key is sent by the cloud to the individual using the system and if it is verified, then the user has the ability to access the account. After login, the user can use the hybrid encryption to securely store data in the cloud.
2. The hybrid encryption deployed is that of the arithmetic and logarithmic formula which is necessary for encryption/ description process of any data which is removed from the confidential information to the cipher text.
3. Security of data is ensured by these operations since they lack key operations which are moved to the individual using the system.
4. The detailed work makes it possible for the storage of data in the cloud. The work also has the ability to minimize the storage capacity of data. Performance evaluation has been made possible by purposed work, encryption size and the time of computation.

5. The proposed project uses fully homomorphism and DES approach to provide 15% efficiency compared to the work that has been done before.

FUTURE WORK

This can be made possible by implementing encryption and future reference for the best security of data in the cloud. This project which is to be done by efficiently utilizing the cloud environment.

REFERENCES

[1] Jongho Mun, Jiye Kim, Dongho Won "An Improvement of User Authentication Framework for Cloud Computing", Journal of Computers, December 15, 2015, pp 446-454.

[2] Ahmed Dheyaa Basha, Irfan Naufal Umar, and Merza Abbas,Member, IACSIT "Mobile Applications as CloudComputing: Implementation and Challenge", pp. 7865-7564, IEEE, 2013.

[3] Uma Somani, "Implementing Digital Signature with RSA Encryption Algorithm to Enhance the Data Security of Cloud in Cloud Computing," 2010 1st International Conference on Parallel, Distributed and Grid Computing (PDGC - 2010).

[4] RuWei Huang,Si Yu, Wei Zhuang and XiaoLin Gui,"Research on Privacy-Preserving Cloud Storage Framework Supporting Cipher text Retrieval" 2011 Ninth International Conference on Grid and Cloud Computing.

[5] Farzad Sabahi, "Cloud Computing Security Threats andResponses," IEEE Trans. on Cloud Computing., vol. 11, no. 6, pp. 670 { 684, 2002.}

[6] Cong Wang, Qian Wang, Kui Ren and Wenjing Lou "Ensuring Data Storage Security in Cloud Computing." IEEE 2009.

[7] Mehmet Yildiz, Jemal Abawajy, Tuncay Ercan and Andrew Bernoth "A Layered Security Approach for Cloud Computing Infrastructure" 2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks. © 2009 IEEE.

[8] Jianfeng Yang, Zhibin Chen "Cloud Computing Research and Security Issues" Vol. 978-1-4244-5392-4/10/$26.00 ©2010 IEEE

[9] Sang-Ho Na, Jun-Young Park, Eui-Nam Huh "Personal Cloud Computing Security Framework" 2010 IEEE Asia- Pacific Services Computing Conference.

[10] Mohammed Achemlal, Sa ıd Gharoutand Chrystel Gaber "Trusted Platform Module as an Enabler for Security in Cloud Computing", Vol 978-1-4577-0737-7/11/$26.00 ©2011 IEEE.

[11] Yong Pan "Research on dependability of cloud computing systems" IEEE Conf. on Reliability, Maintainability and Safety (ICRMS), 2014 ,pp. 435 – 439.

[12] Idris, A.S. "The readiness of Cloud Computing: A case study in Politeknik Sultan Salahuddin Abdul Aziz Shah, Shah Alam" IEEE Conf. on Computational Science and Technology (ICCST), 2014, pp. 1 – 5.

[13] Shaikh, F.B. "Security threats in cloud computing" IEEE Conf. on Internet Technology and Secured Transactions (ICITST), 2011, pp. 214 – 219.

[14] Almorsy, M. "Collaboration-Based Cloud Computing Security Management Framework" IEEE Conf. on Cloud Computing (CLOUD), 2011, pp. 364 – 371.

ISSN (Print): 2321 -4058, (Online): 2321 -4392

Linac Quality Assurance Software

[1] Aleksei E. Zhdanov, [2] Leonid G. Dorosinsky
[1,2]Ural Federal University named after the first President of Russia B.N.Yeltsin, Russia
E-mail: jjj1994@yandex.ru

ABSTRACT

For radiotherapy, quality assurance is necessary to provide precise and accurate radiation treatment. Besides treatment plans and patient positioning, machine status is another vital issue that affects radiotherapy.

In this article, we perform Linear Accelerator (Linac) quality assurance (QA) software based on MATLAB. All dates and images which were used in this article were taken by Linac of radiation therapy department of University Clinic of Erlangen. We have tested the accuracy of beam isocenter, conformity of light field and irradiation field, multileaf collimator (MLC) leakage and dynamic MLC positions relative to rotation, and in this article we will include the results of these tests and analyze machine status by parameters mentioned above. Finally, we will discuss about the validity of existing machine quality assurance program.

Keywords - linear accelerator, quality assurance, image processing, multileaf collimator, light field, irradiation field

I. INTRODUCTION

Quality assurance is designed to ensure that radiation is delivered correctly. Machine characteristics are measured on a daily, monthly and annual basis, and compared to tolerance levels (unacceptable in the long term) and action levels (unacceptable at any time). Different devices are used to measure these QA checks [1]. Daily checks can use less precise but less cumbersome instruments, while monthly and annual checks use more precise instruments.

The objective of patient safety is to ensure that exposure of normal tissue during radiotherapy be kept as low as reasonably achievable consistent with delivering the required dose to the planning target volume. This forms part of the objective of the treatment itself. The measures to ensure quality of a radiotherapy treatment inherently provide for patient safety and for the avoidance of accidental exposure. Therefore patient safety is automatically integrated with the quality assurance of the radiotherapy treatments.

An assessment of clinical requirements in radiotherapy shows that a high accuracy is necessary to produce the desired result of tumor control rates that are as high as possible, consistent with maintaining complication rates within acceptable levels.

The objective of this article was to develop the clinical daily quality assurance software for Linac by analyzing 7th measurements: isocenter accuracy measurement, FOV accuracy measurement, MLC leakage measurement, dynamic MLC Banks Position Measurement, leaf position measurement, one picket test, MLC leakage ratio.

II. METHOD AND MATERIAL

A. Film

Oncology EDR2 Film is a high-contrast film that is coated on both sides with a very fine grain emulsion having low image noise. It is excellent for relative dosimetry (field uniformity, equipment characterization: field shapes, port openings, multi- leaf collimators). Before irradiation, to remove any vibrations we strongly set EDR2 film. Thereafter, we puncture reference point with needle on cover, so as to develop a small dot for reference for the analysis. The reference points are located on light field vertex or longitudinal direction.

B. Electronic Portal Imaging Device

The use of electronic portal imaging devices (EPIDs) is a method for the dosimetric verification of radiotherapy plans, both pretreatment and in vivo. The aim of this study is to test a 2D EPID-based dosimetry algorithm for dose verification of some plans inside a homogenous and anthropomorphic phantom and in vivo as well. Quality assurance based on EPID offers high spatial resolution, fast image acquisition, and digital output [2].

C. Isocenter Accuracy Measurement

A purpose of accuracy measurement are to evaluate the isocenter accuracy of radiation beam relative to the Linac mechanical axis using a simplified star shots test, and to determine the diameters of isocenter. The films have to located between two plastic blocks for fix and place it corresponding to the expected beam isocenter by laser. Isocenter measurement for gantry-rotation and ring-rotation. During gantry-rotation we set gantry angle equal to 0, 70, 140, 210, 280, 350 degree and irradiate field size equal to 10x50 mm2. During ring-rotation we make the film positioning direction towards the beam portal and repeat the previous experiment. Thereafter, film is then scanned with scanner for analysis [3].

D. FOV Accuracy Measurement

We allocate place film towards beam portal, whereupon we puncture four dots corresponding to corners light-field on film. To irradiate installing film we set dose equal to 300 MU. Thereafter, to achieve adata for film analyze we put film into a scanner. Thus, we repeat test with different MLC window size to the same and irradiate with dose of 300 MU.

E. Dynamic MLC Banks Position Measurement

In this test, we verify the position of MLC banks relative to gantry- and ring-rotation by film measurements [4].

- MLC-bank position with gantry-rotation: We set MLC window to half-open, and irradiate each time 300MU from gantry angle of 0 and 180 degree. If MLC banks stay stationary with gantry rotation, no overlap should occur in the film. Film will be later analyzed.
- MLC-bank position with ring-rotation: Like gantry-rotation test, we set MLC window to half-open, and irradiate 300MU each time with ring rotation of 90 degree. If ring rotation is precise and MLC banks stay stationary, no overlap area should occur as well.

F. Leaf Position Measurement

The inspection of the static leaf positions of MLC devices is essential for safe radiotherapy deliveries in both static and dynamic modes. MLCs ensure the accurate dose delivery. Picket fence tests are to measure the MLC errors weekly [2]. A “picket” is the line formed by several MLC pairs all at the same position. There is usually some ideal gap between the MLCs.

In this experiment, we verified the delivered dose with the picket fence pattern. This test is to verify the position of the MLC leafs for different gantry- rotation with static leafs and under leaf motion. The accuracy of MLC positions during radiotherapy is important as even small positional deviations can translate into considerable dose delivery errors. This becomes crucial when radiosensitive organs are located near the treated volume [5].

G. Shape Extraction Algorithm

We used MATLAB to extract specific shapes using extraction algorithm. This algorithm is working as the following: the first step is to convert the image to binary image based on specific threshold, the second step is to extract all the white objects and the black objects with their area value, their center coordinate of each object, and the dimensions of them, and the third step is to surround each object with a square o identify the objects.

Quality assurance is designed to ensure that radiation is delivered correctly. Machine characteristics are measured on a daily, monthly and annual basis, and compared to tolerance levels (unacceptable in the long term) and action levels (unacceptable at any time). Different devices are used to measure these QA checks [1]. Daily checks can use less precise but less cumbersome instruments, while monthly and annual checks use more precise instruments.

III. RESULT AND ANALYSIS

A. Isocenter Accuracy Measurement

To define isocenter sphere we use computer-based image analysis in MATLAB [2]. Isocenter accuracy measurement consists of three steps. Firstly, we enhancing image contrast Fig. 1a. Fig. 1a shows originally image getting by scanning the film. Secondly, we identify and fit central axis of each unit star shot.

The result of this action is shown in Fig.1b. Thirdly, we carry a calculation is based on a line configuration of central axis that we use to define smallest hitting circle. The result of this algorithm is shown on Fig.1c. As a result, the program displays image Fig.1c and a diameter value of the isocenter sphere. The program demonstrates a following result of isocenter diameters are1.56 mm (gantry rotation) and 0.54 mm (ring rotation).

B. FOV Accuracy Measurement

After scanning the film we extract light-field and irradiation field from original image (Fig.2a) by calibration image contrast. The result of this extraction is shown on Fig.2b and Fig.2c. Thereafter, the program analyzes field corners and borders of the resulting images and computes the border length of both fields. The resulting data is places in TABLE I.

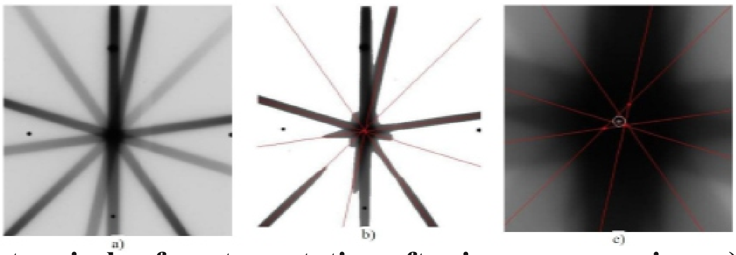


Fig.1 Film Isocenter circle of gantry rotation after image processing; a) contrast enhancing, b) fitting central axis, c) the smallest hittingcircle (zoom-in)

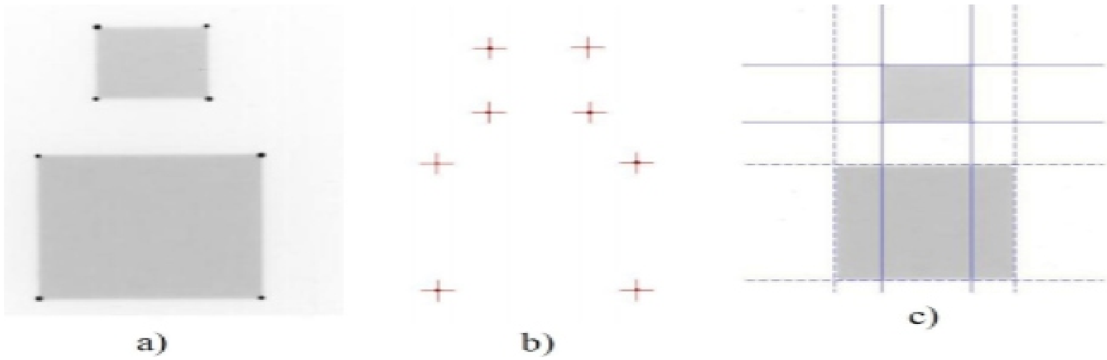


Fig.2 Correlation of the light-field and irrradiation field; a) original film, b) extracted corner of lifht-field, c) extracted corner of lifht-field

Table 1. FOV detection result

Vertex coordinate (cm)	Border length (cm)	
Light field (50*50 mm)	Light field (50*50 mm)	Irradiation field (50*50mm)
(7.92,5.48)	4.87	5.02
(12.79,5.41)	4.96	
(7.85,10.45)	5.04	5.05
(12.90,10.45)	5.03	
Light field (100*100 mm)	Light field (100*100 mm)	Irradiation field (100*100mm)
(5.28,14.41)	9.95	10.14
(15.23,14.33)	9.83	
(5.33,24.24)	9.91	10.13
(15.24,24.19)	9.86	

C. MLC Leakage Measurement

In this article, we performed four tasks to evaluate the MLC leakage. The first task is to perform fully opened MLC. The idea from this step is to see the whole field that the MLC provide. We applied the extraction algorithm in MATLAB to determine the field size and the field dimensions is 9.8298x9.81837 cm2. The second task and third task are to irradiate with the same field size 20x150 from different angles and the reason behind this is because when we irradiate from gantry angle zero degree (the Linac is above the table) the delivered dose is maybe different from irradiating it from 90 degree because of the heavy weight of the Linac. Irradiation of field size 20x150 mm2 with 90 degree of the gantry rotation using the extraction algorithm is 11.7729x96.5835 mm2. Irradiation of field size 20x150 mm2 with 90 degree of the gantry rotation using the extraction algorithm is 11.5443x96.3549 mm2. The standard deviation in x-direction is 0.359 and in y-direction is 0.359. The value of the STD is small and this indicates that the check of irradiating from two different angles is the same and the device work properly.

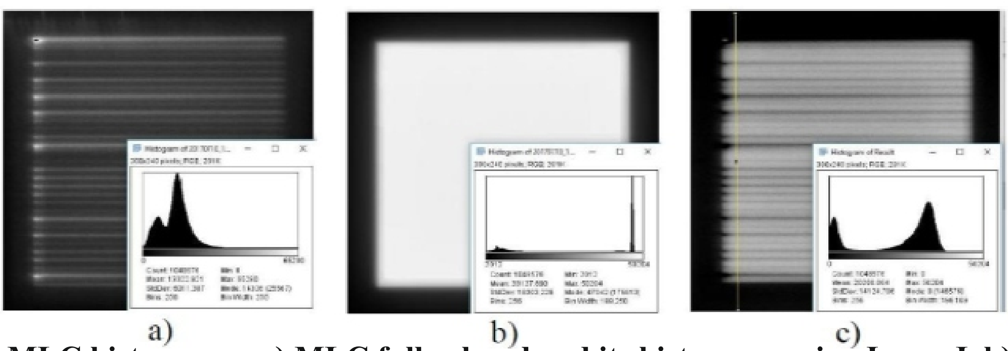


Fig.3. MLC histograms: a) MLC fully closed and its histogram using ImageJ, b) MLC fully opened and its histogram using ImageJ, c) The subtracted image of the fully opened and fully closed MLCs

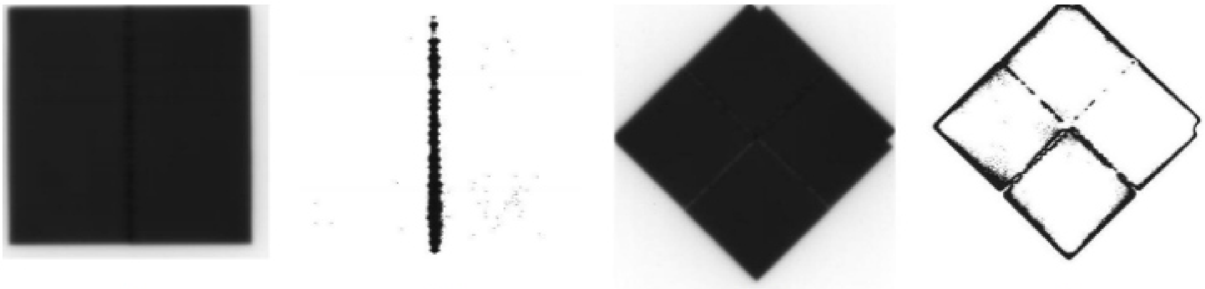
We performed another analysis to prove the MLC leakage using three steps concept as follows: the first step is to perform histogram of the opened MLC and we noticed that all the value around zero and one, the second step is to subtract the closed MLC image (Fig.3a) from the opened MLC image (Fig.3b) and the result is shown in Fig.3c. Fig.3c shows after compare it with the opened MLC in Fig.3b that we lost a percentage of the radiation dose and this proves the MLC leakage.

D. Dynamic MLC Banks Position Measurement

In this test we use both film and EPID for overlapping detection, Fig.4a and Fig.4b shows film scanning result of MLC banks position with gantry rotation and Fig.4c and Fig.4d shows result with ring rotation. From the two films we can see clearly the overlapping and a gap actually occurs.

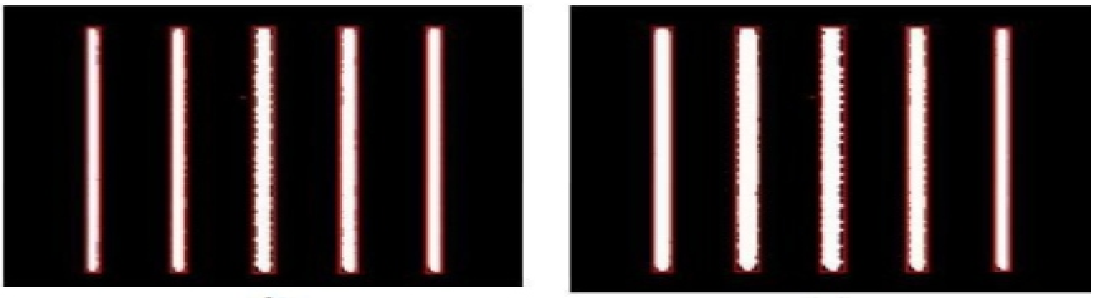
E. Leaf Position Measurement

A common test conducted to measure the positional accuracy of the MLC is the Picket Fence test. This test provides an assessment of the position of each MLC leaf individually and in relation to the alignments of the other leaves. It also shows the actual irradiated gap width between each two leafs.



a) b) c) d)

Fig.4. MLC bank positions: a) With gantry rotation (Initial film), b) With gantry rotation (Overlapping area), c) With ring rotation (Initial film), d) With ring rotation (Overlapping area)



a) b)

Fig.5. The segmented picket fence: a) The segmented picket fence of static MLC using MATLAB, b) The segmented picket fence of VMAT using MATLAB

In this experiment, we applied a Picket Fence pattern in two cases. The first case is static MLC and we measured the leafs positions and their dimensions. We used the previous shape extraction algorithm which can give us the width and the height and the center position of each leaf. Fig.5a shows the result of the static MLC segmentation. Fig.5b shows the result of the VMAT MLC segmentation. The results show that the Picket positions in the both cases (static or rotating MLCs) are almost the same based on the standard deviation values. The width and the height of each Picket is almost the same based on the standard deviation values. But this algorithm is based on using specific threshold to segment the picket and this threshold may exclude the effect of the small variations of the radiation values around the picket, so for detecting these effects we used ImageJ to see this effect. Fig.6a shows the plot profile through the picket fence in the case of static MLC and as we can see, we have high values in the picket's positions and in between must be zero if the MLCs have no leakage and this is not the case. We notice from the plot profile that in between the pickets the values are above zero and this indicates a MLCs leakage but it is small but the position of the pickets are correct. Fig.6b shows the picket fence in the case of VMAT. The result shows using the plot profile graph that the positions of the picket are correct as we compare it with the static case. But because of the rotation of the Linac we can notice the values in between the picket are not stable like the static case. We noticed the shadow in between the pickets and this indicates that there is radiation in the gap between the picket and it is more than the static case because of the speed of the rotation.

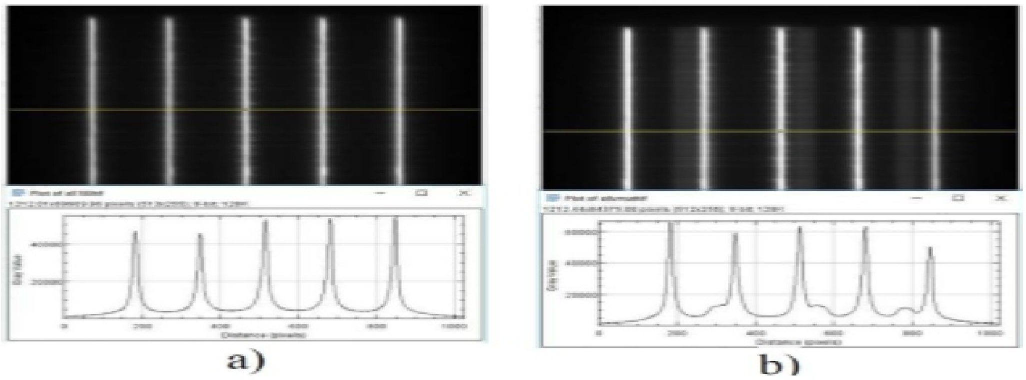


Fig.6. Picket Fence test using: a) Static MLC, b) VMA

F. One Picket Test

In this test, we check one picket of the fence from the static MLCs case. We observe the area on the left side of the picket and we can see the plot profile is not homogeneous but does not show big change at the region along the profile. This means that there is no mechanical problem and the MLCs work properly but there is a leakage. We repeated the same process on the right side of the picket and the result is no mechanical problem and there are no holes.

G. MLC Leakage Ratio

In this test, MLC leakage was measured by using the EPID image. We calculated the leakage ratio by taking the maximum value within the field area when the MLCs are opened or closed. Measured MLC leakage was 0.769% for a 10x10 cm2 field size.

Quality assurance is designed to ensure that radiation is delivered correctly. Machine characteristics are measured on a daily, monthly and annual basis, and compared to tolerance levels (unacceptable in the long term) and action levels (unacceptable at any time). Different devices are used to measure these QA checks [1]. Daily checks can use less precise but less cumbersome instruments, while monthly and annual checks use more precise instruments.

CONCLUSION

From scientific literature we know that MLC leakage consists of transmission through the leaves and leakage between leaves and scattering happens when leakage between leaves occurs, which in film will leave a smooth border. When we analyze isocenter sphere or overlapping area, we set threshold to extract overlapping area simply and ignore the scattering. We think of more precise calculation, which is PSF should be used for exacting actual area being exposed. We calculated the leakage ratio and the result shows that the leakage level is low as it is 0.769%. We noticed small leakage from the MLC using ImageJ tool, which is acceptable but we can not say that closed MLC can prevent any extra radiation. Another and important test is picket fence test which investigate any mechanical problem in the MLCs. We noticed based on the analysis using ImageJ tool that there is no mechanical problems in one picket of the fence and consequently all the other pickets will give the same results. We calculated the picket fence width and height using extraction algorithm using MATLAB and the results show almost the same width and height for all pickets. Finally, the picket's position test in two cases and the result shows same position of the pickets if the MLC is static or moving like in the VMAT case. Quality assurance is an important check to prevent any unnecessary radiation to the patient.

In isocenter accuracy assurance test, diameter of the gantry-rotation isocenter sphere is 13% excessive to expected value and in the ring-rotation isocenter sphere is 27% excessive, which implies slight offset of beam isocenter from gantry- rotation isocenter and ring-rotation isocenter. In field correlation test, difference of edge length of actual irradiation field and light-field is within 2%, so we can conclude that irradiation field conform to light-field quite well. In MLC leakage test, we can see from EPID output that we have MLCs leakage but with small value. The picket fence test proves the correct positions of the MLCs in both static and rotating cases but the rotation of the gantry in VMAT setup affects the result with small dose in between the pickets during the rotation and this leakage is related to the speed of the gantry. In dynamic MLC position test, we can see from the film that the bank overlap occurs for gantry rotation, and void gap for ring rotation. In both tests, we set MLC window half open and the overlap area indicates either deviation of MLC banks position during rotation, or "half-open" does not mean exact half-open. Thus, in this article, we have performed Linac QA software based on MATLAB.

REFERENCES

[1] D. S. Chang, F. D. Lasley, I. J. Das, M. S. Mendonca, J. R. Dynlacht, "Linac Quality Assurance," Basic Radiotherapy Physics and Biology, pp. 151-154, 2014.

[2] B. Sun, S. Yaddanapudi, S. M Goddu, S. Mutic, "A self- sufficient method for calibration of varian electronic portal imaging device," 8th International Conference on 3D Radiation Dosimetry, p. 3, 2015.

[3] T. Depuydt, R. Penne, D. Verellen, J. Hrbacek, S. Lang, K. Leysen, I. Vandevondel, K. Poels, T. Reynders, T. Gevaert, "Computer-aided analysis of star shot films for high-accuracy radiation therapy treatment units," Physics in medicine and biology, p. 57, 2012.

[4] K. S. Jothybasu, a. S. V. Bahl, Amit, G. K. Rath, D. N. Sharma, "Static versus dynamic intensity-modulated radiotherapy: Profile of integral dose in carcinoma of the nasopharynx," J Med Phys, pp. 66–72, 2009.

[5] K. Kainz, "A handbook for teachers and students," Radiation oncology physics, pp. 413-414, 2006.

ISSN (Print): 2321 -4058, (Online): 2321 -4392

Automatic Control System Software for the Gas Pumping Unit with an Anti- Surge Algorithm Based on Trace Mode 6

[1] Aleksei E. Zhdanov, [2] Leonid G. Dorosinsky
[1,2]Ural Federal University named after the first President of Russia B.N.Yeltsin, Russia
E-mail: jjj1994@yandex.ru

ABSTRACT

Continuous gas transportation depends on the stable operation of GPU. However, this stability can be disrupted by a surge. Fighting this phenomenon is a complex technological task that requires a detailed study of the gas pumping unit's operation.

Interaction of the operator with the ACS GPU is carried out through the adopted "man-machine interface", which regulates the order of the operations of preparation, start-up and control of GPU.

Thus, implemented software based on TRACE MODE 6 with an anti-surge algorithm, previously the algorithm was implemented by a third-party device. The described ACS software implements the management of the technological process at all stages and already used on real gas pumping station (GPS).

Keywords - automatic control system, gas pumping unit, centrifugal compressor, surge, anti-surge

I. INTRODUCTION

Gas Pumping Unit (GPU) launch is the most important operation of the compressor station (CS). This is due to the fact that when the GPU is started simultaneously, a very large number of systems are running in the system itself, both the unit itself and the auxiliary systems of the compressor station, from the preparation and correct setting of which it depends how reliable this start-up is. During the transferring process the rotors change state from the working to the installation state. GPU compressor is close to working in the surging zone. Through the supercharger, a large gas flow is carried out at a low compression ratio, which leads to high speeds, especially the recirculation pipelines, which causes them to vibrate [1]. Continuous gas transportation depends on the stable operation of GPU. However, this stability can be disrupted by a surge. Fighting this phenomenon is a complex technological task that requires a detailed study of the gas pumping unit's operation. The article shows the developed ACS GPU software and implemented anti-surge algorithm.

II. METHOD AND MATERIAL

A. Architecture TRADE MODE

TRACE MODE 6.0 is based on a distributed common object model - DCOM, which is the basis of Windows XP. The system modules are divided into Servers and Clients. DCOM. Communication between them is carried out using a distributed common model of DCOM objects. Therefore, individual modules of the system are easily interconnected, and the GPU on the basis of TRACE MODE is easy to maintain, develop and integrate into corporate information systems [2].

The basis of Trace Mode 6.0 is a real-time server that is part of the executive modules (Real-Time Monitors (RTM), NetLink RTM, documentation servers and global recorders). For communication with the Device for Communication with the Object (DFC), interfaces are provided: OPC - client; DDE / NetDDE. Interaction with the Database is carried out through standard SQL / ODBC interfaces and OPC- server.

B. Distribution of Software Components

The software of the local level ACS GPU consists of the software of the controller PLC-03-02 and contains:

- Operating system Windows CE;
- Micro RTM executive module;
- Application software.

On the controllers, the programmable logic controller (PLC) operates under the control of the MICROM executive monitor in the Windows CE environment. The project of applied PLC software is developed in the TRACE MODE tool system.

The software of the controllers implements the functions of collecting information from process sensors, the functions of local automatic control of technological equipment and providing communication via a serial interface and over a local network, and the exchange of data with communication device boards.

C. ACS Control functions

PLC-03-02 controller level:

- Automatic control of starting and stopping the GPU in accordance with the set algorithm by the command received from the operator's automated workstation (AWS) and by the operator's command with substation control house (SCH);

- Remote control of technological equipment and crane strapping GPU according to the given algorithms [3].

Operator's AWS and SCH level:

-AWS level includes the formation of control commands for transmission to the ACS;

- ACS level includes generation of control commands for transmission to the controller of a specific GPU.

D. ACS Information Functions

ACS GPU software provides continuous automatic monitoring of technological parameters of GPU.

ACS GPU software provides the implementation of the following information functions:

- Continuous automatic monitoring of GPU process parameters;
- Display of information on the screen of the operational control panel of the control unit;
- Transfer of information to the level of the GPU. ACS GPU software provides the following inform-ation functions:
- Collection and display of information received from all local ACS;
- Automatic presentation of information on the current values of the monitored parameters with indication of the output beyond the permissible limits and the deviation characteristics (pre-warning or emergency) on mnemonic circuits and in special video frames;

- Automatic display of tripping of interlocks and protections and indication of the state (open, closed) of the crane strapping on the mimic diagram of the unit;
- Presentation of technological schemes to the operator with indication of the status of executive mechanisms, switching devices and stop valves;
- Formation of an array of retrospective information with a specified periodicity with output to the press at the request of the daily report operator;
- Calculation and control of values of regime- technological parameters;
- Control of serviceability of technical means of ACS GPU.

ACS GPU software provides the following information functions:
- Collection and display of information received from the controller of this unit;
- Automatic presentation of information on the current values of monitored control cabinet (CC) parameters with indication of the output beyond the permissible limits and the deviation characteristics (warning or emergency) on mnemonic circuits and in special video frames;
- Automatic display of the triggering of locks and protections, warning and alarm indication [4].

E. Data Transfer Equipment

Two types of computer technology are used in the ACS GPU:
- Personal electronic computers (PCs) are the basis of the higher-level (workstations and servers);
- Programmable controllers in the structure of control cabinets of objects (lower-level).

Higher-level hardware (IBM-compatible personal computers in industrial design) (workstations, servers, network facilities) are powered from the 220V network through uninterruptible power supplies (Smart UPS), which ensure the autonomous operation of the equipment for at least 30 minutes.

Lower-level hardware is a multi-functional complex of software and hardware that has great flexibility in configuration and allows the user to select the necessary hardware for the solution of various control and monitoring tasks by the design layout method, and also to rebuild or expand the complex in case of changing the parameters of the object management.

The general structure providing the interconnection between the main components of ACS GPU (AWS, SCH, Servers) is a two-level local area network system (Fig. 1).

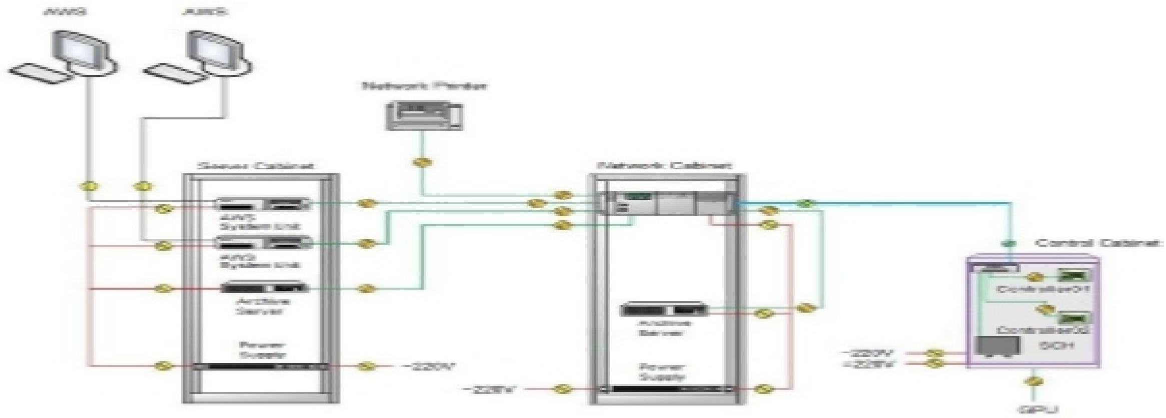


Fig. 1. Interconnection structure between the main components of ACS GPU

III. RESULT AND ANALYSIS

A. AWS and SCH

The frames are the main part of displaying information about the state of the technological process, the functioning of equipment, the issuance of emergency and warning messages to operational personnel, and the management of technological objects. Frames are made taking into account the convenience of perception of output information by the service personnel. The state of the process and the operation of the equipment are visually represented in the form of static and dynamic graphic objects. Static graphic objects (slides) are background images (mnemonic diagrams, tables, etc.), which show fragments of the control and monitoring object. Dynamic elements consist of displaying analog and discrete variables, as well as a simplified image of cranes, pumps, fans, etc. The software provides an approach in which the control system is viewed as a hierarchy of objects - a workshop (a group of technological objects), a technological object, a node, and signals. Objects are described by data structures and all actions in the system are manipulation of objects.

Input analog signals and calculated analog variables are displayed by dynamic graphic objects - parameter display panel (digital display), trend.

Input discrete signals and calculated discrete variables are displayed by dynamic graphic objects - colour indicators, blinkers.

The working area of the dispatcher screen consists of static and dynamic parts. Static elements of the screen never change while the program is running. Dynamic elements contain information about the progress of the technological process and vary depending on the state of the system. Figure 2 shows AWS panel.

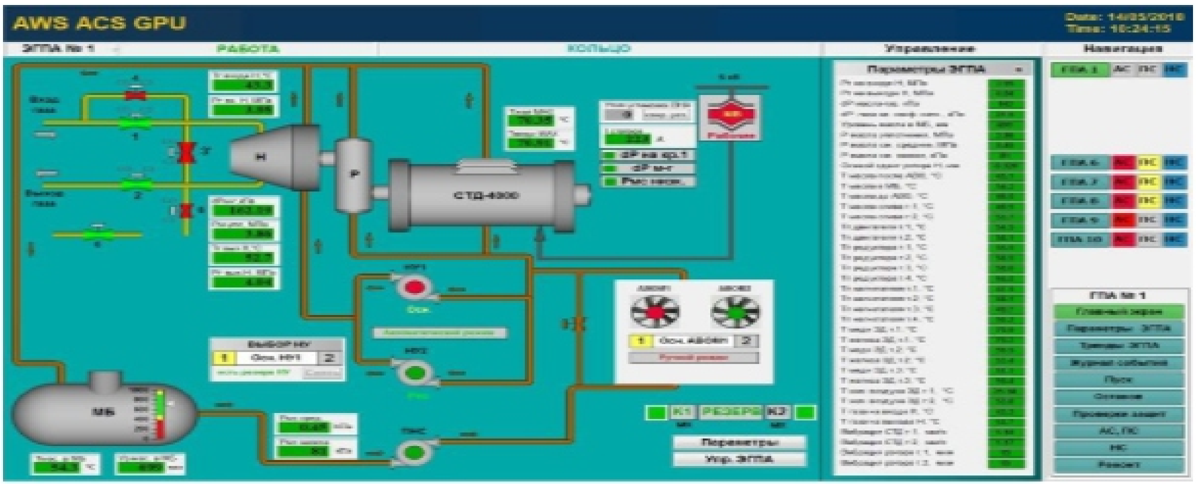


Fig. 2. AWS panel

After startup, the program displays the main screen. From the main screen we can get to the screens of a lower level. There is a static panel on the right side of the screen, at the top of which there are three indicators: an alarm, a warning signal, a malfunction signal, which is highlighted when there are emergency or warning signals, respectively. By clicking on one of the indicators, we can go to the corresponding screen, where you can find out the nature of the signal and acknowledge (update the status) of it. Also in the right part of the screen there is a row of buttons that allow switching to auxiliary screens in accordance with the inscriptions on these buttons. In the work area of the frame of a complex object, its graphic image is placed, including a static image and a set of dynamically changing simple graphic objects. SCH panel constructed according to a similar logic and is shown in Figure 3.

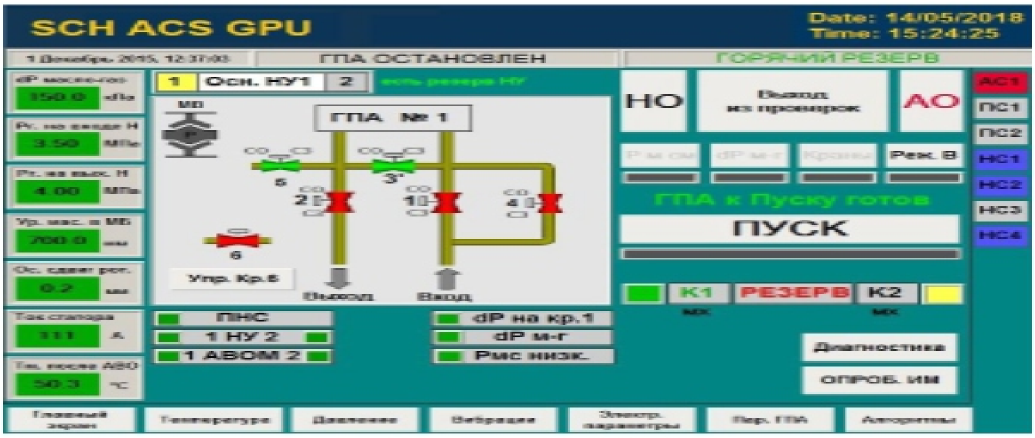


Fig. 3. SCH panel

B. Anti-Surge Algorithm

The work of the anti-surge algorithm (Figure 4) corresponds to the operation of anti-surge protection of the SIGNAL-1M system.

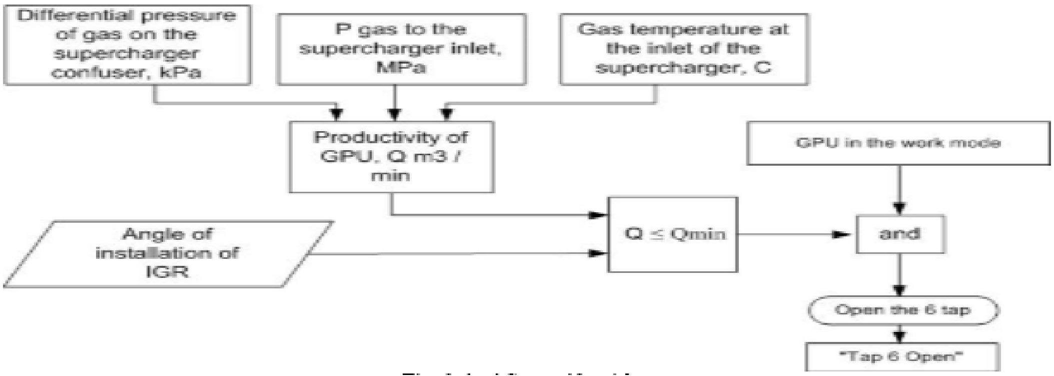


Fig. 6. Anti-Surge Algorithm

The volumetric productivity is calculated by the formula (1) [5].

$$Q = A\sqrt{\frac{\Delta P \cdot T \cdot P_0 \cdot 735{,}56 \cdot K_{ng}}{\gamma_0 \cdot P \cdot T_0}} \qquad (1)$$

Where:

Q - volumetric capacity of the centrifugal supercharger, m3 / min;

A - the consumption factor of the inlet branch pipe of the centrifugal supercharger (2);

$$A = 68{,}4\frac{kg^{1/2} \cdot m^{1/2}}{min \cdot (mm\,Hg)^{1/2}} \qquad (2)$$

ΔP - pressure drop of the gas on the compressor confuser (0.... 0.16 MPa);

T - gas temperature at the inlet of the supercharger (273.... 323 K);

P - gas pressure at the inlet of the supercharger (2 ... 5 MPa);

K_{ng} - coefficient of compressibility of natural gas, K_{ng} = 0,94;

Specific weight of gas 0.73 kg / m3 at normal pressure P_0 = 0.1033 MPa and temperature T_0 = 293 K.

The value of the emergency performance set point is determined by the angle of the input guide rail (IGR) according to Table 1. IGR angle value is entered from the SCH.

In the event of a reduction in the calculated performance value below the emergency set point, a warning alarm and a command to open the crane 6 are formed. The operator closes the crane 6 manually after eliminating the surge of the centrifugal supercharger.

CONCLUSION

As previously noted, use of GPU sensors has rapidly expanded as an effective tool for improving GPU run- life performance. Sensors used in conjunction with temperature and pressure monitoring and surveillance tools allow key GPU operating parameters to be acquired, stored, and evaluated continuously [6].

Table 1. Angle of the input guide rail

IGR Angle, degree	Alarm value, m3/min
+35	97
+25	102
+15	104
+5	106
0	108
-5	112
-15	118
-25	124
-35	130

Also, on a practical basis, use of monitoring and surveillance systems enables the engineer effectively to manage more producing GPU than in the past. Given the decline in availability of skilled personnel in the industry, this benefit alone can be an important reason for using these systems.

The control of GPU is organized by the operator (from the SCH or AWC) of ACS included in the ACS GPU.

Interaction of the operator with the ACS GPU is carried out through the adopted "man-machine interface", which regulates the order of the operations of preparation, start-up and control of GPU.

Thus, implemented software based on TRACE MODE 6 with an anti-surge algorithm, previously the algorithm was implemented by a third-party device. The described ACS software implements the management of the technological process at all stages and already used on real gas pumping station (GPS).

REFERENCES

[1] S. Becker, "Analysis of surge appearances in the automatic control system of gas pumping unit on the basis of Hopf bifurcation," Automation, telemechanization and communication in oil industry, pp. 46-52, 1973.

[2] J. Cheng, M. Grossman, T. McKercher, "Professional CUDA C Programming," Paperback, p. 404, 2014.

[3] G. Wang, A. Zomaya, G. Martinez, K. Li, "Algorithms and Architectures for Parallel Processing," ICA3PP, p. 15, 2015.

[4] D. A. Yuen, L. Wang, X. Chi, L. Johnsson, W. Ge, Y. Shi, "GPU Solutions to Multi-scale Problems in Science and Engineering," Lecture Notes in Earth System Sciences, p. 10, 2013.

[5] L. C. Uren, "Petroleum Production Engineering: Oil field development," The Journal of Geology 65, p. 540, 1957.

[6] J. Vandevier, "ESP—Conclusion: Multiple factors affect electrical submersible pump run life," Washington PulseMagazinePast Issues, p. 1, 2010.

ISSN (Print): 2321 -4058, (Online): 2321 -4392

Issues and Challenges in Software Quality Prediction using Soft Computing Techniques

[1] **Mukesh Kumar Shrivastav,** [2] **R. K. Dwivedi,** [3] **Madhumita Singha**
[1]Research Scholar, Vinoba Bhave University, Hazaribag
[2]Associate Professor, Vinoba Bhave University, Hazaribag
[3]Associate Professor, XISS, Ranchi
E-mail: [1]mukesh.k.shrivastav@gmail.com, [2]rkdwivedi_vbu@yahoo.com,
[3]madhumitasingha@xiss.ac.in

ABSTRACT

Software quality Prediction is area of research where we extract the inherent code attributes to study its impact on software quality attributes using range of techniques to know effectiveness of a particular method. Software quality analysis and prediction focuses on detecting high-risk program modules, allowing practitioners to allocate project resources strategically [1]. It is quite beneficial for testers, developers and project managers to know the modules which are highly risky and fault prone for strategic decisions. However, due to versatility of software, various programming approaches, languages and different platforms it is always challenging to devise standard heuristics and method which is applicable to each type of software. Starting from definition of Software Quality, selection of code attributes for the purpose, selection of method, data sets to test and standardize the outcome is quite challenging. Adoption of agile methodology, cloud oriented developments, mobile developments brings more challenges of quality professionals. The purpose of this paper is to outline the issues, challenges, state of the art and find the gap in research work done in the past.

Keywords - Software Quality, Software Quality Model, Software Product Quality, Software Quality Prediction, Soft Computing Techniques Introduction.

I. INTRODUCTION

A quality software product is cornerstone of all software engineering practices. The quality of a system or product is the degree to which the system satisfies the stated and implied needs of its various stakeholders, and thus provides business value. There has been common practiced approach and believe to improve the quality software product by following standard and tested process models which inherently improves the software quality. However, due to complex requirements, low budget, immature team, short schedule and unorganized software development processes faults are part and parcel of software product. Excessive or irrational schedules are probably the single most destructive influence in all of software [2]. Detection and Prediction of software defects early is major challenge for tester, developers and project managers.

Software quality is defined as The totality of functionality and features of a software product that bear on its ability to satisfy stated or implied needs [ISO/IEC 25000][3]. Software Quality is defined using quality models. Quality Model is a schema to better explain of one's view of quality [4].

Software metrics is measurable attributes of software product which has direct impact on software quality. Software Product metrics can be categorized in to internal, external and quality in use. Internal attributes can be extracted during development phase where as external and in use attributes can be extracted during testing and deployment phase respectively. Software quality Prediction involves selection of software metrics suitable for particular software, its quantification, in depth analysis and testing to find its impact on software quality characteristics in order to achieve quality product. Vijay Kumar et al. studied that there is non-linear relationship between software metrics and defects, which results to a complex mapping [5].

There have been continuous researches to find method and the defects as early as possible in development stages of software development life cycle using various methods and techniques such as Case Based Reasoning, Regression Trees, Bayesian belief networks (BBN) and Support Vector Machines (SVM). Soft computing methods are recent approaches which are being tested and researched for finding this relationship. Code attributes extracted from the software are often imprecise and uncertain so soft computing approaches are quite promising methodology for the purpose.

Soft computing techniques which are being researched and used in software quality prediction techniques are Fuzzy Logic, Neural Network, Genetic Algorithms and some Hybrid techniques which are combination of two or more.

Practically software quality is estimated using either of reliability or maintainability. Reliability can be predicted by finding the presence or absence of defects. Maintainability can be predicted by estimating effort to change any component such as class or method in the software.

In this paper, we have reviewed the research papers, journals and articles published in different journals to know the current state of the art for research and finally tried to find different aspects of Software Quality Prediction and current trends.

II. SOFTWARE QUALITY MODELS

Software Quality Models are set of attributes and subsequent sub-attributes suitable for a software product. Some of popular quality models proposed by earlier scholars are McCall's Quality Model, Boehm's Quality Model, Dromey's Quality Model, ISO/IEC 9126-3 Quality Model and newly added ISO/IEC 25000.

According to T. Davuluru et. al,There are two types of categories of models they are product models and process models. [6] McCall, Boehm, Dromey, FURPS, ISO 9126 come under product quality models and CMMI, Six Sigma comes under process quality models. However, we can add another category as Project Quality models.

Miguel et. al. classifies the evolution of software quality as Basic Model which are adoptable to any software product where as tailor made models are domain specific and derived from Basic model[7].

Selection of a quality model for a software project is a common and real challenge but essential to settle the objective and requirements.

III. SOFTWARE PRODUCT METRICS

The quality of underlying software measurement data is critical. An intelligent selection of software metrics before training a defect model is likely to improve the end result by removing redundant and less important feature [8]. As said by H. Olague et. al Empirical validation of software metrics suites to predict fault proneness in object-oriented (OO) components is essential to ensure their practical use in industrial settings.[9]

Measurable Code attributes are called metrics and are derived from the inherent properties of code. Feature selection is the process of choosing some subset of the features and building a classifier based solely on those. It can remove as many unnecessary features as possible, leaving only those features which are useful in building a classifier. Metrics can be method level or class level.

Lines of code(LOC) is popularly used and controversial metrics for the studies because of threats to its validity. Lines of code depend on selection of programming language. It is also affected by programmer expertise. Comments inside a programs increases or decreases size of code.

McCabe(1976) and Halstead(1977) metrics are method level which was proposed in 1970s but are still widely used[10,11]. Cyclomatic Complexity was given by McCabe gives relative complexity of various designs. It can be used early phases of the development and in practice it is a count of the number of test conditions in a program. Halstead proposed to use tokens which are extracted directly by counting number of operators, operands and constructs in the code and principally attempts to estimate the programming effort.

According to Ankita, McCabe's cyclomatic complexity metric only considers complexity from control graph or conditional statements and not considers complexity is there interaction or coupling between two object classes.[12]

Chidamber and Kemerer (1994) proposed Object Oriented Programming suit popularly known as CK metrics suite[13]. It comprises of Weighted number of methods(WMC), Number of member functions, number of children (NOC), Depth of inheritance tree (DIT), Response for a class(RFC), Lack of Cohesion in methods(LCOM) and coupling between objects (CBO). Using a base of code, they tried to correlate these metrics to the defect density and maintenance effort using covariant analysis.

Abreu et al[14] gave MOOD metrics suite for measurement of Object Oriented Metrics suite. It comprises of Method Inheritance Factor(MIF), Attribute Inheritance Factor(AIF), Method Hiding Factor(MHF), Attribute Hiding Factor(AHF), Polymorphism Factor(POF).

Yadav et al. observed that 49% of the metrics used are OO (Chidamber and Kemerer's (CK) metrics are among the highest), 27% are traditional source code metrics or 24% are process metrics.[15]

According to Lamrani, M., Amrani et al., due to the degree of formalism used to express them, they[CK metrics suite] are still frequently cited in several contributions and improved by other authors for being useful as quality indicators over many characteristics related to complexity, inheritance, coupling, cohesion and messaging.[16]

Identifying the problems of package level analysis in Object Oriented systems Robert C. Martin developed software metrics for a set of classes i.e. packages. The objective of the package level metric is to identify poorly designed packages.[17]

R. C Martin metrics suite consists of Afferent couplings (Ca), Efferent couplings (Ce), Abstractness (A), Instability (I), Distance from the main sequence (D), Package dependency cycles.

Rest of the paper is organized as sections for each of Software quality models, software product metrics, Data sets, Soft computing techniques followed by conclusion.

IV. DATA SETS

Software quality Prediction research requires quality datasets for experiments. However acquiring quality datasets is not feasible due to lack system to capture that data or unwillingness of private parties to share their data due to business policy obligations. In such a scenario, open source datasets are quite suitable and feasible solution. As mentioned, “The NASA Metrics Data Program data sets have been heavily used in software defect prediction experiments and recommended that researchers 1) indicate the provenance of the datasets they use, 2) report any pre- processing in sufficient detail to enable meaningful replication, and 3) invest effort in understanding the data prior to applying machine learners.” [18].

V. SOFT COMPUTING TECHNIQUES

Soft Computing uses special techniques like neural networks, fuzzy logic, support vector machines, and evolutionary computation to minimize the execution periods of functional tasks.[19]

Soft computing is not a closed and clearly defined discipline at present but very promising set of techniques for problems that could not be stated precisely. These methods can exploit tolerance for precision, uncertainty and partial truth to achieve tractability, robustness and low cost solution.

They differ from Hard Computing that it requires precision, certainty and rigor. Soft computing techniques address the problems in their domain in complementary way rather competitive way. The common element of these techniques is generalization through non-linear approximation and interpolation in high dimensional spaces.

Analysis of the selected primary studies revealed that current software defect prediction research focuses on five topics and trends: estimation, association, classification, clustering and dataset analysis. The total distribution of defect prediction methods is as follows. 77.46% of the research studies are related to classification methods, 14.08% of the studies focused on estimation methods, and 1.41% of the studies concerned on clustering and association methods. In addition, 64.79% of the research studies used public datasets and 35.21% of the research studies used private datasets.[20]

A. Artificial Neural Network (ANN)

Artificial Neural Network is inspired by human brain in reasoning and decision making. In 1943, Warren S. McCulloch, a neuroscientist, and Walter Pitts, a logician developed the first conceptual model of an artificial neural network. In their paper, "A logical calculus of the ideas imminent in nervous activity,” they describe the concept of a neuron, a single cell living in a network of cells that receives inputs, processes those inputs, and generates an output.

Khoshgaftaar et al. introduced the use of the neural networks as a tool for predicting software quality[21].

In their study, they presented a large telecommunications system, classifying modules as fault prone or not fault prone. They compared the ANN model with a non-parametric discriminated model, and found that ANN model had better predictive accuracy.

Singh et al.[22] used the ANN for estimation of testing effort using Object Oriented Metrics. They have taken the testing effort as dependent variable where as Object Oriented metrics as independent variables and found that ANN are suitable in finding the non linear relationship between the variables.

Artificial neural networks were originally designed to model the pattern recognition abilities of the brain. They have been used extensively for many practical predictive and data classification tasks. There are many ways neural networks can be trained, and fuzzy logic and genetic algorithm can be used for the purpose. Training and learning help neural networks perform better in unexpected situations. At that time fuzzy values would be more applicable than crisp values.

B. Fuzzy Logic (FL)

Fuzzy Set Theory was first proposed by L. Zadeh in 1965 for knowledge representation where knowledge is imprecise and uncertain. It includes the use of approximate reasoning, by which we model how humans obtain information from imprecise information. A major goal is to obtain a description of human knowledge and experience in a way that can be implemented on machines.

Jaya Pal et. al[23] compared multiple regression and fuzzy logic for software quality prediction on some academic software projects and found positive outcome. It was suggested to collect more data in support of the outcome.

C. Genetic Algorithms (GA)

Genetic Algorithms (GA) were introduced in the late 1960's by John Holland [23]. They are based on the Darwinian theory of evolution whereby species compete to survive and the fittest get a higher chance to remain until the end and produce next generation. GA is a class of robust problem-solving techniques based on a population of solutions, which evolve through successive generations by means of the application of three genetic operators: selection, crossover, and mutation [24].

As mentioned, Researches reporting the use of genetic algorithm in fault proneness prediction application are few and recent [25].

D. Hybrid Approaches

Hybrid techniques are used in complementary way to obtain better output such as Neuro-Fuzzy, the combination of Neural Network and Fuzzy Logic.

"Adaptive neuro fuzzy inference system (ANFIS) model based on these two basic elements of soft computing, neural network and fuzzy logic gives a more accurate measure of reliability than the FIS model, as it reduces error from 11.74%, in case of FIS model, to 6.66% in ANFIS".[27]

In [28,29] Singh et al. used feed forward neural network for software reliability prediction and applied back propagation algorithm to predict software reliability growth trend.

CONCLUSION

We can conclude with following observation that Software Quality Definition is objective based, agenda oriented and depends on the purpose of its constitution. A defined and fixed quality model may not be suitable across domains and applications so it is preferable to have flexibility in selection of quality attributes depending upon the domain and application prepared on consensus with prospective user and stakeholders.

Reliability and Maintainability are most desirable quality attribute in software products. Reliability can be predicted by finding the presence or absence of defects. Maintainability can be predicted by estimating effort to change any component such as class or method in the software. CK Metrics suite is preferable object oriented metrics suite used for the prediction of software quality. McCabe Complexity metrics is suitable for procedural programs and not inadequate to cater Object oriented Paradigm requirements. Standardization of findings and replicating it for other application is major issue for research scholars in the field. In most cases, public data sets provided by Promise database and NASA's MDP has been used for standardization and validation of results. An effective software quality prediction requires intelligent selection of quality standards, quality attributes, code attributes and suitable prediction technique depending upon domain requirements.

Current software defect prediction research focuses on five topics and trends: estimation, association, classification, clustering and dataset analysis. All references mentioned in the paper suggest to carry out more and more researches on Soft computing techniques in order to make it more subtle and replicable.

REFERENCES

[1] Gao, K., Khoshgoftaar, T. M. and Seliya, N. (2011). Predicting high-risk program modules by selecting the right software measurements. Software Quality Journal, 20(1), 3- 42.

[2] Capers Jones, "Applied Software Measurement: Global Analysis of Productivity and Quality", p.403, McGraw Hill Professional(2008).

[3] https://www.iso.org/standard/35733.html

[4] Gupta, D., Goyal, V. K., & Mittal, H. (2011). Comparative study of soft computing techniques for software quality model. International Journal of Software Engineering Research & Practices, 1(1), 33–37.

[5] Vijay Kumar, Arun Sharma, Rajesh Kumar, Applying Soft Computing Approaches To Predict Defect Density In Software Product Releases: An Empirical Study, Computing And Informatics, Vol 32, No 1 (2013).

[6] T. Davuluru, J. Medida and V. S. K. Reddy, "A study of software quality models," 2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014), Unnao, 2014, pp. 1-8. doi: 10.1109/ICAETR.2014.7012958

[7] Miguel, Jose & Mauricio, David & Rodriguez, Glen. (2014). A Review of Software Quality Models for the Evaluation of Software Products. International journal of Software Engineering & Applications. 5. 31-54. 10.5121/ijsea.2014.5603.

[8] Gao, K., Khoshgoftaar, T.M., Wang, H., Seliya, N.: Choosing software metrics for defect prediction: An investigation on feature selection techniques. Software: Practice and Experience. Special Issue: Practical Aspects of Search-Based Software Engineering 41(5), 579–606 (2011), doi:10.1002/spe. 1043

[9] H. Olague, L. Etzkorn, S. Gholston, S. Quattlebaum, "Empirical validation of three software metrics suites to predict fault-proneness of object-oriented classes developed using highly iterative or agile software development processes", IEEE Trans. Softw. Eng., vol. 33, no. 8, pp. 402- 419, 2007.

[10] Thomas J. McCabe, A complexity measure, Proceedings of the 2nd international conference on Software engineering, p.407, October 13-15, 1976, San Francisco, California, USA.

[11] Maurice H. Halstead, Elements of Software Science (Operating and programming systems series), Elsevier Science Inc., New York, NY, 1977.

[12] Garg, Ankita. (2014). An approach for improving the concept of Cyclomatic Complexity for Object-Oriented Programming.

[13] Chidamber, S. and Kemerer. C. (1994). A metric suite for object oriented design, IEEE Transactions on Software Engineering, 20(6)

[14] F.B. Abreu and R. Carapuca, Candidate Metrics for Object- Oriented Software within a Taxonomy Framework,‖ System and Software, vol. 26, no. l, pp. 87-96, Jan. 1994.

[15] Yadav, Harikesh & YADAV, DILIP. (2015). A fuzzy logic based approach for phase-wise software defects prediction using software metrics. Information and Software Technology. 63. 10.1016/j.infsof.2015.03.001.

[16] Lamrani, M., Amrani, Y.E., & Ettouhami, A. (2013). A Formal Definition of Metrics for Object Oriented Design : Mood Metrics.

[17] Hyrynsalmi, Sami & Leppänen, Ville. (2009). A Validation of Martin's Metric 10.13140/2.1.3645.3448

[18] D. Gray, D. Bowes, N. Davey, Y. Sun, B. Christianson, "The misuse of the nasa metrics data program data sets for automated software defect prediction", Evaluation & Assessment in Software Engineering (EASE 2011) 15th Annual Conference on. IET, pp. 96-103, 2011.

[19] Catal, Cagatay & Bayrak, Coskun & Bou Nassif, Ali & Polat, Kemal & Akbulut, Akhan. (2016). Special Issue: Soft computing in software engineering. Applied Soft Computing. 49. 953-955. 10.1016/j.asoc.2016.09.046.

[20] R. S. Wahono. A systematic literature review of software defect prediction: Research trends, datasets, methods and frameworks. Journal of Software Engineering, 1(1):1--16, 2015.

[21] T.M.Khoshgaftaar, E.D.Allen, J.P Hudepohl, S.J Aud,., "Application of neural networks to software quality modeling of a very large telecommunications system," IEEE Transactions on Neural Networks, Vol. 8, No. 4, pp. 902-- 909, 1997

[22] Singh, Yogesh & Kaur, Arvinder & Malhotra, Ruchika. (2008). Predicting Software Fault Proneness Model Using Neural Network. 204-214. 10.1007/978-3-540-69566-0_18.

[23] Jaya Pal and Vandana Bhattacharjee, 2015. "SOFTWARE QUALITY PREDICTION USING FUZZY RULE BASED SYSTEM" International Journal of Current Research, 7, (12), 24181-24185

[24] J. H. Holland. Adaptation in Natural Artificial Systems. University of Michigan Press, 1975

[25] Arning, Andreas, Rakesh Agrawal, and Prabhakar Raghavan. A Linear Method for Deviation Detection in Large Databases, 2nd International Conference on Knowledge Discovery and Data Mining Proceedings, 1996, pp. 164-169.

[26] Rosli, M.M., Teo, N.H., Yusop, N.S., Shahriman, N., & Mohamad Fault Prediction Model for Web Application Using Genetic Algorithm.

[27] Kaswan KS, Choudhary S, Sharma K (2015) Software Reliability Modeling using Soft Computing Techniques: Critical Review. J Inform Tech Softw Eng 5:144. doi:10.4172/2165-7866.1000144

[28] Singh Y, Kumar P (2010) Application of feed-forward networks for software reliability prediction. ACM SIGSOFT Software Engineering Notes35: 1-6.

[29] Singh Y, Kumar P (2010) Redirection of Software Reliability Using feed Forward Neural Networks. International conference on Computational Intelligence and software Engineering.

[30] ISO/IEC (2011). Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models. International Standard ISO/IEC 25010:2011

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).
2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);
3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);
4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

<u>**Address of the Editorial Office:**</u>

Enriched Publications Pvt. Ltd.
S-9,IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-45525005