International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Gross Enrolment Ratio Prediction using Artificial Neural Network

Hussain. J, Rosangliana. D, Vanlalruata

Abstract**: *The present paper deals with Gross Enrolment Ratio (GER) prediction in higher education within the state of Mizoram, India. The data used in this study are obtained from the yearly report of All India Survey on Higher Education (AISHE), published by Ministry of Human Resource Development (MHRD), Govt. of India and Statistical Handbook of various years published by Dept. of Economics & Statistics, Govt. of Mizoram. In this study, a soft computing technique known as Artificial Neural Network (ANN) is implemented for prediction of GER in higher education. The data obtained are analyzed and categorized into two classes known as the input and target data. The input data represent the years from 1968 to 2017. The target data represents the enrolment details corresponding to the input year such as, male enrolment, female enrolment, eligible population and GER. After generating these input data and target data, an ANN is used for building a model. The model has been trained and tested using 530260 student enrolment data for the period of 50 years. In order to obtain an accurate GER prediction, the accuracy of four architectures of ANN known as Back Propagation (BP), Radial Basis (RB), Recurrent Neural Network (RNN), and Feed Forward Neural Network (FFNN) are compared. The comparison is carried out by performing a prediction on the known data set. It is found that BP Neural Network with 50 hidden neurons and a learning rate of 0.1 gives the best prediction accuracy. Therefore, to predict the future GER, a BP neural network is implemented in this study. The main focus of this study is to analyze the pattern of enrolment and to predict future GER, as GER is a primary indicator for the status of higher education. The result obtained may help policymakers to take suitable decision to increase GER in higher education within the state of Mizoram so as to contribute to the govt. of India's target of 30% by the year 2020.***

Keywords : GER prediction; Neural Network; Back Propagation; Student Enrolment Prediction; Classification; Soft Computing; Forecasting

I. INTRODUCTION

The state of Mizoram are only a little more than 50 years old, however, within this short span; the state has made rapid achievement. [1]

In literacy rate, Mizoram recorded 88.8% in the year 2001 and with a phenomenal increase in the subsequent decade, the figure for 2011 is as high as 91.33 percent, surpassed only by Kerala (93.91%) and Lakshadweep (92.28%). Despite this rapid increase in literacy rate, student enrolment in higher education is comparatively low. This may be due to poor economic condition of the state. Often families are unable to financially support their children for higher education. Some students are forced to go for jobs to support their families, which has a negative impact on enrolment in higher education. The proportion of budget allocation to higher education has also not changed much since 2007-08. The highest percentage of allocation since 2007-08 was in 2017-18 spending 1.62% of the total budget on higher education. Budget allocation percentage out of total budget for the past 12 years is given in below in Figure 1:

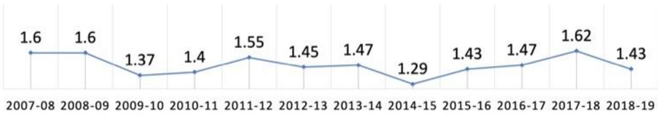

Figure 1: Budget allocation for the past 12 years

Higher Education Institutions within the state has been unevenly distributed among its districts. 69% of the total number of the institutions is established in the capital district of Aizawl. This may be due to rapid urbanization. However, access to higher education in rural areas is becoming more difficult due to a limited number of seats. The district wise distribution of institution as on 30th Sep. 2017 is given in Figure. 2 below:

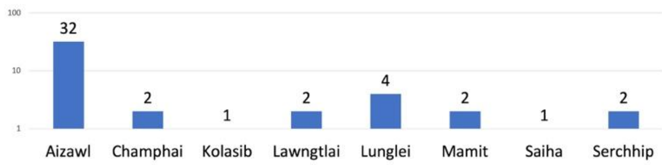

Figure 2: District wise distribution of institutions

All the above factors may be possible causes for declining higher education enrolment in Mizoram.

GER projection is an important requirement. This projected GER provides a vital source of information for decision making, education planning and budget planning [2]. But, predicting GER depends on many external factors.

Manuscript published on January 30, 2020.
* Correspondence Author
Hussain. J*, Department of Mathematics and Computer Science, Mizoram University, Aizawl, India. Email: jamal.mzu@gmail.com
Rosangliana. D, Department of Mathematics and Computer Science, Mizoram University, Aizawl, India. Email: drosangliana@gmail.com
Vanlalruata, Department of Mathematics and Computer Science, Mizoram University, Aizawl, India. Email: ruata.mzu@gmail.com

Through a broad range of studies, the variables below consistently prove to be statistically significant to GER [3]: family income, parent's level of education, tuition, student aid levels, and student's academic aptitude.

It was found that the tuition fee level becomes less significant when family income increases [4]. There were also studies on the impact of unemployment rates on enrollment: higher unemployment rates result in lower college enrollment [5]. In this case, some small change in one of the conditioning factors may bring in a sizeable change in the GER figure of students at a given point in time.

According to the latest report of Govt. of India [6], Indian GER in higher education has registered increase from 24.5% in 2015-16 to 25.2% in 2016-17, Govt. of India aims to attain a GER of 30% by 2020, it is still much behind the target. In particular, the GER for the state Mizoram GER for 2017-18 is only 22.9%.

The GER for the state of Mizoram for the last 8 years is given in Table 1 below:

Table 1: GER of Mizoram during the last 8 years

Years	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18
GER	21.6	19.0	22.2	23.2	23.3	24.1	24.5	22.9

In this study, GER for the state of Mizoram is analyzed and prediction is made using an artificial neural network. Male enrolment, Female enrolment, Eligible population and GER are the input parameters for training the ANN model. Once the model is trained with the proposed four ANN architectures, experiments are performed using the dataset based on the performance of this prediction, a particular architecture which is found to be most accurate from the validation is selected to predict the future GER for the state of Mizoram. Four ANN Architectures namely, BP, RNN, RB and FFNN were trained and tested for identifying the best architecture. Which will be use to predict the future GER for the state.

There has been an increase in interest of bio-inspired computational algorithms which is also commonly termed as computational intelligence (CI), in discovering knowledge by means of analyzing and prediction of data. Artificial Neural Networks (ANNs) have been developed as a model parallel to the biological neural networks of the human brain out of the various CI techniques [7]. The underlying processes and relationships of a student enrolment from the past year dataset often have chaotic properties. Unlike statistical and mathematical techniques which depend on influencing factors, the success of ANN prediction accuracy depends on parameters adjustment [8]. Therefore, to discover knowledge from the past enrolment data and performing student GER prediction in Mizoram for higher education, ANN is implemented. The prediction accuracy is compared among the four different architecture known as BP, RB, RNN, and FFNN. The prediction accuracy comparison is validated in two phases. In the first phase, a combination of Mean Square Error (MSE) and Root Mean Square Error (RMSE) is used. The second phase validation is performed by three validation techniques namely as K-Fold Cross-Validation, Leave-One-Out Cross-Validation (LOOCV) and Random Subsampling. The two-phase validation confirmed BP neural network (BPNN) architectures is the most suitable among the other architecture considered. Therefore, enrollment prediction is performed using the BPNN.submission to the journal, rectification is not possible.

II. RELATED WORKS

Prediction with Time series method has wide applications in different areas namely economics, engineering, management, medicine, science, and technology etc. [9].

Data mining techniques were used by Kabakchieva [10] for Predicting Student Performance. The principal motive is to reveal the high potential of data mining applications for university management.

Ramaswami and Bhaskaran [11] had developed a predictive data mining model for identifying slow learners and investigate the factors influencing students' academic performance, using the CHAID decision tree algorithm.

Neural Networks, Multivariate Adaptive Regression Splines, and Classification trees were used by Yu *et. al.* [12] to explore student retentions

The objective of this paper is to provide an overview on the data mining techniques that have been used to predict students' performance Machine by Shahiri and Husain [13].

The performance of distance learning students from Computer Science had also been predicted by Kotsiantis *et. al.* using classification algorithms namely Perceptron-based Learning, Instance-Based Learning, Decision Tree, Bayesian Net, and Rule-learning. [14]

Neural network and Fuzzy time-series approaches were applied by Song and Chissom [15] in the field of Artificial Intelligence for prediction of students' enrolments.

Guo [16] applied survival ratio techniques to predict the students' enrolment number of a four-year university program.

Bandyopadhyay and Chattopadhyay [17] predicted the population of India with the ANN model. The results showed that the correlation between actual and predicted values is very high (0.94 and 0.98 for males and females respectively).

III. METHODOLOGY

In this paper, we use a multi-layer neural networks with one hidden layer. The input data to the network is composed of 50 years data (1968-2017) and each input year corresponds to 4 parameters namely Male enrolment, Female enrolment, eligible population and GER. Thus, the input data consists of 50 X 4=200 features. A network execution during the training is inspected so as to avoid over training. The methodology starts with the initial data gathering and preprocessing, where relevant information about the GER were identified. The model for the prediction is build using two stage known as training and validation. Once the validation results are found to be acceptable, then the model is used for student enrolment and GER prediction else the ANN is tune again. The overall graphical representation of the block diagram is given in Figure 2 below.

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

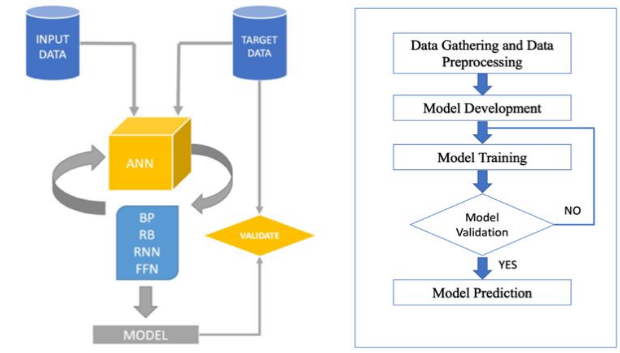

FIG 3: ANN Graphical representation of GER prediction

A. Data Collection and Pre-processing

Data collection and pre-processing are the initial phases of this work. Obtaining a piece of vital information in relation to GER data is by itself a challenging task for a state like Mizoram, where information scatters in horizons. After analysing various sources of data; two sources are selected for training our ANN model. The first source is AISHE and the second is Mizoram Statistical Handbook from 1968-2017 (50 Years enrolment data).

MHRD Govt of India have started collecting higher education statistics from the year 2010 through its online portal. Data collected from the above sources are pre-processed, the following input and target parameters given in Table 2, are consider for prediction of GER for the state of Mizoram.

Table 2: Parameter for training prediction model

INPUT	TARGET			
Year (1968-2017)	Male Enrolment	Female Enrolment	Eligible Population	GER

B. Model Development

In model development, four architectures of ANN is considered, whose prediction performance, are evaluated. From experimental analysis it is found that momentum value of 0.8, learning rate of 0.1 with sigmoid transfer function accomplish the best execution results in considering the time and space complexity. Further, the number of neuron in the hidden layer is varied from 10 to 70 to decide the optimal architecture. In this hit and trial method, a 50 neural in the hidden layer is found to be most ideal for all the analyzed architecture. The parameter details for the prediction model is given in the below Table 3.

Table 3: ANN Prediction Model parameter

No. of Layers	3 (1 input, 1 hidden, 1 output)
No. of Neuron in hidden layer	50
Input Node	4 (Male enrolment, Female enrolment, eligible population and GER)
Momentum value	0.8
Learning rate	0.1
Transfer function	*Sigmoid*
Optimization Algorithm	*Gradient descent*

The matrices used to assess the performance of the analyzed neural network architectures, include MSE, RMSE, Accuracy, Specificity and Sensitivity. Based on this matrix one ANN architecture is selected for preforming GER prediction. In additional it is essential to note that over-training the ANN model can seriously deteriorate prediction accuracy. To counter this, an optimization algorithm known as gradient decent is implemented.

C. Model Training

In this phase of model training, the input data format is given in Table 2. This data are fed to each ANN architecture to evaluate their performance. During training, the inter-unit connections between the input and hidden layer nodes are optimized using gradient descent. The epoch continues to increment until the error in the predictions result is minimized and the network reaches the specified level of accuracy. Once the network is trained and tested it can be given new input information (year) to predict the output (GER). Individually all architectures namely BP, RB, RNN, and FFNN were tuned with the parameter given in Table 3 so that all the four network architectures are in identical setup. This is important to minimize the diversion and to make the comparison possible. The generated model from this training phase is to be evaluated by validation and testing. ANN represents a promising modeling technique, especially for our enrolment data sets which have a non-linear relationship.

D. Model Validation

Validation strategies in machine learning are utilized to get the error rate of the ANN model. This validation enables us to compare the performance of our ANN Architecture.

In this paper, we implement a two-phase validation technique. The first phase is composed of MSE, RMSE, Accuracy% and precision% validation. The second phase is composed of three validation techniques. These techniques are based on a bias, precision, and accuracy in statistical estimation. The MSE and RMSE are given in Equation 1 and 2 respectively.

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

$$\text{RMSE} = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (2)$$

Where y_i is the real data and $\hat{y}_i$ is the predicted information. Both MSE and RMSE express average model prediction. RMSE does not necessarily increase with the variance of the errors rather it increases with the variance of the frequency distribution of error magnitudes.

The second phase validation are implemented as follows:

I. K-Fold Cross-Validation: In this system, k-3 folds are utilized for training and the remaining one is utilized for validation. The schema is depicted in below Figure 4.

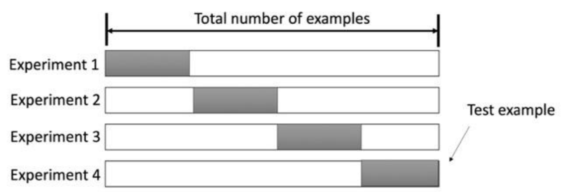

FIGURE 4: K-Fold Cross-Validation

Retrieval Number: E6873018520/2020©BEIESP
DOI:10.35940/ijrte.E6873.018520
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The preferred standpoint is that 75% of enrolment data are utilized for training and the remaining 25% for validation. The error rate of the model is the average of the error rate of every cycle. Therefore, the number of cycles is directly proportional to the number of error. The error rate is improved by utilizing the stratification strategy.

II) Leave-One-Out Cross-Validation (LOOCV): In this process, the majority of the information is utilized for training and one record is utilized for validation. In this validation the dataset is divided into M parts, each of this M part is rehashed for N times if there are N records. The preferred standpoint is that the 9/10 information is utilized for training and 1/10 is utilized for validation. The error rate of the model is the average of the error rate of all iteration. The following Figure 5 represents the LOOCV validation technique.

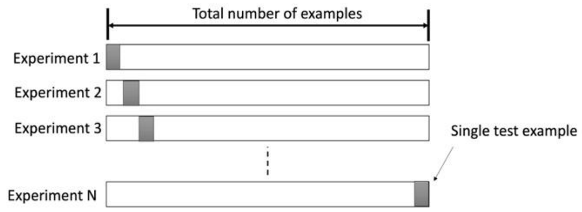

FIGURE 5: Leave-One-Out Cross-Validation (LOOCV)

III) Random Subsampling: In this technique, different arrangements of information are randomly chosen from the dataset and consolidated to form a test dataset. Random-subsampling does not ensure all instances are used for training and validation. However, more number of loop increase the possibilities of the instances being utilized. In our enrolment data a random of 8 instances are selected for training. Among these one instance is used for validation. The average of the error rate of all iteration is the error rate of the model. The following Figure 6 represent a random subsampling validation technique, applied to our ANN model.

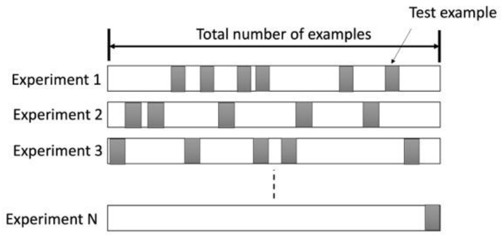

FIGURE 6: Random Subsampling

E. Model Forecasting

The objective of this model forecasting is to estimate the value of GER for the coming years 2018, 2019 and 2020. Often a trend exists that a series increases, decreases or remains at a constant level with respect to time. Therefore, the time is taken as an input feature for prediction. Model forecasting implemented with ANN is found to be the most suitable as ANN is optimized for pattern recognition.

IV. RESULT AND DISCUSSION

Tables 4 below shows the prediction results of the four machine learning architecture using the validation phase 1 described as above. Apart from MSE and RMSE an added parameter namely accuracy (%) and precision (%) is taken into consideration. Each of the ANN architecture is compared against Accuracy (%) Precision (%) MSE (%) RMSE. The result listed in Table 4 clearly identify that BP neural network is the most effective architecture among the all.

Table 4: ANN Architecture Validation Phase 1.

Method	Accuracy (%)	Precision (%)	MSE (%)	RMSE (%)
BP	97.3	98.6	99.6	98.7
RB	94.4	94.6	93.6	97.2
RNN	95.2	97.6	94.6	93.8
FFNN	93.8	95.6	91.6	97.3

Tables 5, 6 and 7 shows the performance metrics of average sensitivity and specificity for the cross-validation using K-Fold Cross-Validation (4 Fold), Leave-One-Out Cross-Validation (LOOCV) and Random Subsampling for the four architecture respectively. Further the validation clearly shows that BP (BP) is better than the others.

Table 5: ANN architecture validation using 4-Fold Cross-Validation

Indicator	**BP**	RB	RNN	FFN
Sensitivity	0.93	0.84	0.65	0.73
Specificity	0.92	0.95	0.96	0.45

Table 6: ANN architecture validation using LOOCV

Indicator	**BP**	RB	RNN	FFN
Sensitivity	0.99	0.78	0.72	0.78
Specificity	0.87	0.89	0.83	0.5

Table 7: ANN architecture validation using Random Subsampling

Indicator	**BP**	RB	RNN	FFN
Sensitivity	0.97	0.94	0.69	0.76
Specificity	0.97	0.85	0.83	0.48

By analyzing, the results obtained from table 5, 6 and 7, we conclude that BP Neural Network is better as compared with the other compared Neural Network Architecture for predicting the enrolment and GER for the year 2018, 2019 and 2020. The deviation from the actual enrolment for each four type of ANN architecture is analyzed and evaluated. In this pattern, the BP architecture is found to predict value closed to the actual enrolment data. All the other model predict a value greater than the actual known value. The deviation of their prediction accuracy deviate in the trend, starting with BP, RNN, RB and FFN. The student enrolment predictions of four compared ANN architectures are given in Figure 7 below.

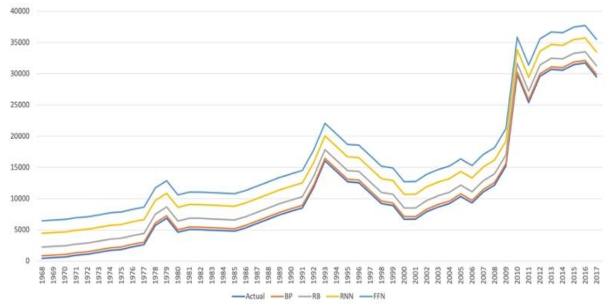

FIGURE 7: Prediction against actual dataset

The student enrolment prediction for 2018, 2019 and 2020 for male, female and total enrolment using BP Architecture is given in Figure 8 below:

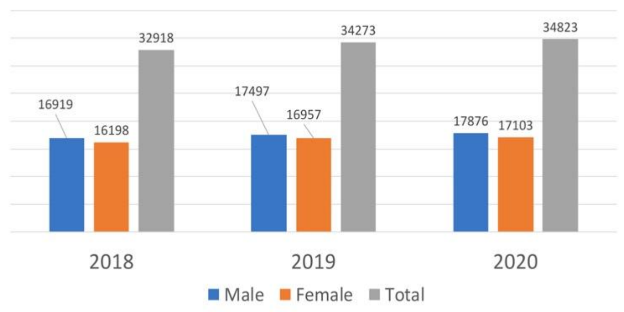

FIGURE 8: Enrolment prediction for the year 2018, 2019 and 2020

The GER prediction for the year 2018, 2019 and 2020 using BP algorithm is given in Figure 9 below:

FIGURE 9: GER Prediction for the year 2018, 2019 and 2020

Further, it was found that both the ability to correctly predict student enrolment and GER the sensitivity increase over time, as the training dataset becomes more balanced. From our experiments, it is found that to achieve 30% GER by the year 2020 in the state of Mizoram we should have 2868 numbers of additional student enrolments in higher education sector. Student enrolment during 2018 is not known yet as annual report of AISHE is still under process.

New algorithm with greater precision and specificity can be achieved by introducing new variables in the training dataset.

V. CONCLUSION

In the light of the above findings, we conclude that low enrolment in higher education within the state of Mizoram may be caused by certain factors which could be summarized as: (a). Due to poor economic conditions, some students are forced to go for jobs to support their families, which has negative impact on enrolment in higher education. (b). Uneven distribution of Institution among the districts leads to low enrolment, (c). Low budget allocation in higher education sector is an important factor of low enrolment. (d) Increase in unemployment leads to low enrolment in higher education. All the above factors accounting for the decline in higher education enrolment needs to be addressed immediately in order to achieve higher GER.

To counter the challenge, increasing budget allocation in higher education sector of the state is one of the most important things which will directly affect the growth of infrastructure as well as filling up of vacant teaching post in the institutions under the administration of state government. This will lead to existence of different type of institutions depending upon demand and more intakes of students in the institutions thereby increasing the enrolment within the state. Sensitization of people with regard to the importance of higher education through media may also be an important step. This may reduce family pressure to force their children for jobs at an early stage which is very common in families of economic weaker section.

REFERENCES

1. State plan at a glance of mizoram: rusa. (2018, January 16). Retrieved from rusa.nic.in/download/152/shep/3101/mizoram_spglance-final.docx
2. Strauss, L. M., & Borenstein, D. (2015). A system dynamics model for long-term planning of the undergraduate education in Brazil. Higher Education, 69(3), 375-397.
3. Patanakul, P., Kwak, Y. H., Zwikael, O., & Liu, M. (2016). What impacts the performance of large-scale government projects?. International Journal of Project Management, 34(3), 452-466.
4. Wilkins, S., Shams, F., & Huisman, J. (2013). The decision-making and changing behavioural dynamics of potential higher education students: the impacts of increasing tuition fees in England. Educational Studies, 39(2), 125-141.
5. Brand, J. E. (2015). The far-reaching impact of job loss and unemployment. Annual review of sociology, 41, 359-375.
6. MHRD. “AISHE Final Report 2017-18”, (2018).
7. Panayotou, T. (2016). Economic growth and the environment. The environment in anthropology, 140-148.
8. Guo, S. (2002, May). Three enrollment forecasting models: issues in enrollment projection for community colleges. In 40th Research and Planning Conference, Pacific Grove, CA.
9. Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited,.
10. Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. Cybernetics and information technologies, 13(1), 61-72.
11. Ramaswami, M., & Bhaskaran, R. (2010). A CHAID based performance prediction model in educational data mining. arXiv preprint arXiv:1002.1144.
12. Yu, C. H., DiGangi, S., Jannasch-Pennell, A., & Kaprolet, C. (2010). A data mining approach for identifying predictors of student retention from sophomore to junior year. Journal of Data Science, 8(2), 307-325.
13. Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. Procedia Computer Science, 72, 414-422.
14. Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2004). Prediction of Student’s Performance in Distance Learning Using Machine Learning Techniques. Applied Artificial Intelligence, 18(5), 411-426.
15. Song, Q., & Chissom, B. S. (1993). Forecasting enrollments with fuzzy time series—part I. Fuzzy sets and systems, 54(1), 1-9.
16. Guo, S. (2002, May). Three enrollment forecasting models: issues in enrollment projection for community colleges. In 40th Research and Planning Conference, Pacific Grove, CA.

17. Bandyopadhyay, G., & Chattopadhyay, S. (2006). An Artificial Neural Net approach to forecast the population of India. arXiv preprint nlin/0607058.

AUTHORS PROFILE

Prof. Jamal Hussain is the Head, in Department of Mathematics and Computer Science, Mizoram University. His specialization is Mathematical Modelling and Computer Simulation of Agricultural, Biological, Ecological and Environmental Systems. Artificial Neural Network and Intrusion Detection Systems.

David Rosangliana is a research scholar in the Department of Mathematics and Computer Science, Mizoram University. He is currently working in Govt. Zirtiri Residential Science College, Aizawl and have a teaching experience of more than 11 years in Under Graduate level.

Vanlalruata is a research scholar in the Department of Mathematics and Computer Science, Mizoram University. His area of interest is Computer Vision and Machine Learning.