Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2025 (43) 3** : 1–10.
ISSN-0037-8712
doi:10.5269/bspm.78536

# Reliable ECG Classification Using RNN and GRU Architectures for Arrhythmia Detection

Bhagyashree Saikia*, Pooja Debbarma, Rajesh Kumar, M. M. Dixit

ABSTRACT: Cardiac disease has become a severe threat to public health as it has become a leading cause of mortality in India. Electrocardiogram (ECG) signal classification plays a pivotal role in early detection of cardiac arrhythmias, potentially reducing morbidity and mortality associated with cardiovascular diseases. This study presents a comparative analysis of Recurrent Neural Network (RNN) and Grated Recurrent Unit (GRU) architectures for ECG signal classification using the MIT-BIH Arrhythmia dataset. After preprocessing and class balancing, both models were trained to classify five heartbeat types. Experimental results show that GRU model achieved a significantly higher test accuracy (98.61%) compared to the RNN (83.39%), demonstrating its potential for real-time cardiac monitoring applications in diagnostic systems and wearable devices.

Key Words: Recurrent neural network, grated recurrent unit, preprocessing, test accuracy.

## Contents



## 1. Introduction

Cardiovascular diseases (CVDs) become a leading cause of death globally. Early detection and management are vital in reducing adverse outcomes. The ECG is a non-invasive tool to monitor heart rhythms and identify arrhythmias. [2] Manual analysis of ECG signals is time-consuming and error-prone, especially in large-scale monitoring scenarios. This has led to increased interest in automated ECG classification using machine learning and deep learning techniques. [1] Recent advances in deep learning, particularly in Recurrent Neural Networks (RNN) and Gated Recurrent Units (GRU), have shown promising results in time series signal classification tasks, including ECG data. [7] [8] RNNs are known for handling sequential

* Corresponding author.
2010 *Mathematics Subject Classification:* 92C55.
Submitted August 21, 2025. Published October 07, 2025

data, but they suffer from issues like vanishing gradients. GRUs, as a variant of RNNs, mitigate these issues while being computationally efficient. [5] [6] This study compares RNN and GRU models on the MIT-BIH Arrhythmia dataset to identify the model better suited for reliable ECG classification.

## 2. Literature Review

Electrocardiogram (ECG) classification plays a crucial role in the early diagnosis of arrhythmias and cardiac disease. Traditional ECG classification approaches have predominantly relied on domain-specific feature engineering, such as ex tracting time-domain, frequency-domain, or wavelet-based features, followed by classification using algorithms like Support Vector Machines (SVM) or Random Forests. Although such methods have achieved moderate accuracy, they require substantial manual intervention and expert knowledge, and often struggle to generalize across noisy recordings and inter-patient variability. In recent years, the advent of deep learning has facilitated the automatic extraction of discriminative features directly from raw or minimally preprocessed ECG signals, leading to notable improvements in classification accuracy, robustness, and adaptability across diverse datasets. [3]

Recurrent Neural Networks (RNNs) have been widely applied to ECG classification tasks owing to their ability to model sequential dependencies in time series data. Schwab et al. developed an RNN ensemble enhanced with an attention mechanism to classify arrhythmias from single-lead ECG recordings, achieving an average F1-score of 0.79 on an unseen test set while also offering interpretable decision-making through attention weight visualization. [7] Traditional RNNs often face challenges such as the vanishing gradient problem and slow convergence during training. Gated Recurrent Units (GRUs), as a streamlined variant of LSTM networks, address these issues by enabling faster training and requiring fewer parameters, while still effectively modelling long-term dependencies. For instance, Yao et al. proposed a hybrid CNN-GRU architecture incorporating wavelet-based denoising and data augmentation, which achieved sensitivity, accuracy, and F1-scores of 99.33%, 99.61%, and 99.42%, respectively, on the MIT-BIH Arrhythmia Database. [10]

Attention mechanisms have been shown to significantly enhance the performance of GRU-based models in ECG arrhythmia classification. Sun proposed a hybrid architecture that integrates CNNs with attention-based recurrent networks, demonstrating improved accuracy and interpretability by allowing the model to focus on diagnostically important regions of the ECG signal. [9] Similarly, Ebrahimi et al. highlighted in their review that such hybrid models, which leverage convolutional layers for spatial feature extraction and recurrent units such as GRUs for temporal sequence modelling, are increasingly effective in arrhythmia detection tasks. GRUs have also been reported to achieve performance comparable to or better than LSTMs, while maintaining training efficiency due to their reduced parameter complexity. Although simpler recurrent architectures, such as bidirectional Elman RNNs, have occasionally shown competitive results, the consensus indicates that GRUs provide a favourable balance between computational efficiency and predictive performance. [4]

While GRU-based architectures, particularly those enhanced with hybrid designs and attention mechanisms, have shown notable performance improvements, there remains a lack of direct, controlled comparisons with traditional RNNs. In many studies, differences in datasets, preprocessing methods, or training procedures make it difficult to isolate the effect of the architecture itself. To address this limitation, the present work conducts a rigorous head-to-head comparison between RNN and GRU models for ECG classification, ensuring identical datasets, preprocessing pipelines, and training settings. This approach enables fair, replicable evaluation of their respective strengths and weaknesses.

## 3. Methodology

The methodology employed in this study focused on a comparative analysis of Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU) models for automatic ECG signal classification. The process involved several key steps to ensure a reliable and fair evaluation. The process includes data collection, preprocessing, model design, training setup, and evaluation.

### 3.1. Data Acquisition

The dataset used in this study is the MIT-BIH Arrhythmia Database, a well established resource in the field of biomedical signal processing. It comprises annotated ECG recordings that represent a wide variety of normal and abnormal heartbeats. The recordings were originally sampled at a higher frequency, but for this study, all ECG signals were resampled to a uniform sampling frequency of 125 Hz. This down sampling aligns with typical data acquisition settings in portable or wearable cardiac monitoring devices and helps to reduce computational requirements without compromising diagnostic accuracy.

- (0) Normal beats (N): 90,587 instances
- (1) Atrial premature beats (S): 2,776 instances
- (2) Premature ventricular contractions (V): 6,344 instances
- (3) Fusion of ventricular and normal beats (F): 803 instances
- (4) Paced beats (Q): 8,039 instances

These five categories were chosen to reflect a broad spectrum of clinically significant arrhythmias and normal cardiac behavior, providing a realistic and comprehensive benchmark for model training. Before training the deep learning models, several preprocessing steps were performed to ensure that the ECG data were clean, balanced, and suitable for learning. The raw signals were first cleaned to remove noise and irrelevant information. The data was then normalized to ensure that all features were scaled to a common range, which helps to improve the efficiency and stability of the training process. Since the original dataset was imbalanced.

### 3.2. Data Preprocessing

One of the main difficulties challenges of working with real-world medical data is class imbalance. The original class distribution in the MIT-BIH Arrhythmia dataset is heavily skewed toward normal beats.

Table 3.1: Distribution of Heartbeat Classes in the train dataset

| Label | Number of Beats | Description |
|---|---|---|
| 0.0 | 72,470 | Normal beats |
| 4.0 | 6,431 | Fusion of paced and normal |
| 2.0 | 5,788 | Premature ventricular contraction |
| 1.0 | 2,223 | Atrial Premature |
| 3.0 | 641 | Fusion of ventricular and normal |

Table 3.1 shows the number of ECG beats for each type of heartbeat used in this. Most of the beats are normal, while abnormal types like premature and fusion beats appear much less often. This imbalance needs to be considered to ensure the models can accurately detect all heartbeat types.

Table 3.2 shows the number of ECG beats for each class in the test dataset. It is evident that the dataset is skewed, with many normal beats (18,117) compared to fewer samples in other categories. Premature ventricular contractions (PVCs) have 1,448 samples, while atrial premature beats and fusion beats have even fewer examples (556 and 162, respectively).

### 3.3. Dataset Splitting Strategy

To train the model effectively and evaluate it fairly, the dataset was split into parts. Eighty percent of the data was used for preprocessing and training, while 20% was reserved as a separate test set for final evaluation. The dataset con tained five heartbeat types: Normal (90,587), Atrial Premature (2,776), Premature Nodal Contractions (6,344), Fusion of Ventricular and Normal (803), and Paced beats (8,039). This split ensured balanced and unbiased model development. From the pre-processed data, further splitting was done as follows:

Table 3.2: Number of images per class in test dataset

| Label | Number of Beats | Description |
|---|---|---|
| 0.0 | 18117 | Normal beats |
| 4.0 | 1608 | Fusion of paced and normal |
| 2.0 | 1448 | Premature ventricular contraction |
| 1.0 | 556 | Atrial Premature |
| 3.0 | 162 | Fusion of ventricular and normal |

- Training set: 80% of the pre-processed data
- Validation set: 10%
- Test set: 10%

For the new dataset, a similar splitting strategy was adopted. The training set contained 87,553 samples (80%), while both the validation and test sets comprised 10,945 (10%) and 10,946 (10%) samples, respectively. This systematic partitioning helped maintain a balanced representation of classes across all subsets, which is crucial for training robust models and ensuring fair, unbiased evaluation.

## 4. Results and Discussion

### 4.1. Performance Analysis

*4.1.1. Performance Analysis of the RNN Model.* The MIT-BIH Arrhythmia dataset was used to train and assess the Recurrent Neural Network (RNN), with an 80/10/10 split between training, validation, and testing. The model was trained over 10 epochs with the goal of classifying ECG signals into five groups: four types of irregular heartbeats and one normal heartbeat. The performance of the RNN model is reviewed in this part, with an emphasis on important measures including accuracy, loss, and performance across classes. Training Progress:

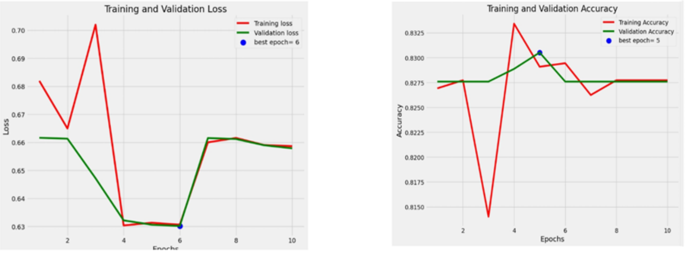
- Epoch 1: Initial training accuracy was 43.82% with a loss of 1.4715, while validation accuracy reached 82.24%.
- Epoch2-6: Accuracy improved steadily, stabilizing at around 82.76%. The training loss decreased, showing the model was learning effectively.
- Epoch 7-10: The accuracy remained constant at 82.72%, and the learning rate was reduced progressively from 0.0010 to 0.00004.

Test Performance: Test Accuracy: 83.39% and Test Loss: 0.625

*4.1.2. Performance Analysis of the GRU Model.* The GRU-based model was trained using the MIT-BIH Arrhythmia dataset with an 80/10/10 split for training, validation, and testing. The model was trained for 20 epochs with the use of early stopping, learning rate reduction, and model checkpoints as callbacks. Training Progress:

- Epoch 1: The model started with a 43.82% accuracy and a loss of 1.4715. Training improved significantly over subsequent epochs.
- Epoch 2-10: The accuracy gradually increased, reaching 99% by the end of epoch 10 with a loss of 0.0555 by epoch 20.

Test Set Performance:



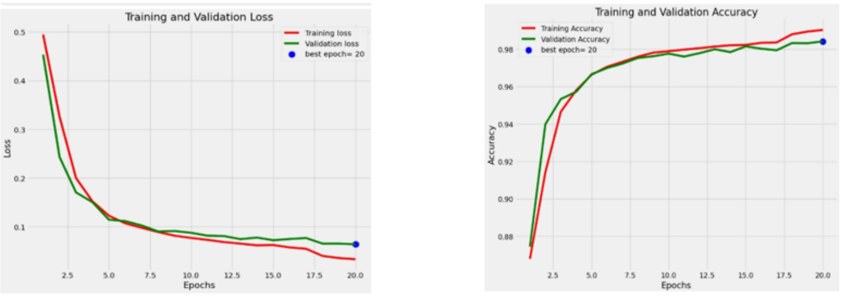

(a) Training and validation losses

(b) Training and validation accuracy

Figure 1: Training validation losses and accuracy for RNN

- Test Accuracy: 98.61%
- Test Loss: 0.0555

The GRU model performed exceptionally well on the test set, with a 98.61% accuracy.




(a) Training and validation losses

(b)Training and validation accuracy

Figure 2: Training validation losses and accuracy for GRU

### 4.2. Evaluation Metrics

*4.2.1. Evaluation Metrics for RNN.* The model achieved an accuracy of 0.83, indicating that 83% of the ECG beats were correctly classified.

- Class 0 (Normal): The RNN handled normal heartbeats very well, with perfect recall and a strong F1-score of 91%.
- Class 1 (Atrial Premature): The model didn't detect any beats from this class, with both recall and F1-score at 0%.
- Class 2 (Premature Ventricular Contraction): Performance was the same here—no correct predictions, resulting in 0% scores.
- Class 3 (Fusion): The model failed to recognize any fusion beats, again scoring 0% for both metrics.
- Class 4 (Paced): Paced beats were also completely missed, showing no detection capability for this class.

Table 4.1 RNN Model Classification Report:

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0.0 (Normal) | 0.83 | 1.00 | 0.91 | 9059 |
| 1.0 (Atrial Premature) | 0.00 | 0.00 | 0.00 | 278 |
| 2.0 (Premature Ventricular Contraction) | 0.00 | 0.00 | 0.00 | 724 |
| 3.0 (Fusion) | 0.00 | 0.00 | 0.00 | 81 |
| 4.0 (Paced) | 0.00 | 0.00 | 0.00 | 804 |
| Accuracy | ----- | ------ | 0.83 | 10,946 |
| Macro avg | 0.17 | 0.20 | 0.18 | 10,946 |
| Weighted avg | 0.68 | 0.83 | 0.75 | 10,946 |

*4.2.2. Evaluation Metrics for GRU.*

- Class 0: Perfect performance with 100% recall and 99% F1-score.
- Class 1: A slight dip in performance, achieving 76% recall, but the F1 score remained strong at 83%.
- Class 2: Achieved high performance with 96% recall and 96% F1-score.
- Class 3: Despite a 77% recall, the F1-score remained good at 79%.
- Class 4: Excellent results with 99% recall and 100% F1-score.

Table 4.2: GRU Model Classification Report:

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0.0 (Normal) | 0.99 | 1.00 | 0.99 | 9059 |
| 1.0 (Atrial Premature) | 0.92 | 0.76 | 0.83 | 278 |
| 2.0 (Premature Ventricular Contraction) | 0.96 | 0.96 | 0.96 | 724 |
| 3.0 (Fusion) | 0.83 | 0.77 | 0.79 | 81 |
| 4.0 (Paced) | 1.00 | 0.99 | 1.00 | 804 |
| Accuracy | ----- | ------ | 0.99 | 10,946 |
| Macro avg | 0.94 | 0.89 | 0.92 | 10,946 |
| Weighted avg | 0.99 | 0.99 | 0.99 | 10,946 |

## 4.3. Confusion Matrix Analysis

*4.3.1. Confusion Matrix for RNN.* In the below fig.3 shows the confusion matrix for the RNN model used to classify different types of ECG heartbeats. The model performs very well in identifying normal beats, with a high recall of 98.98%. It also shows excellent accuracy for detecting atrial premature beats and fusion beats, with almost perfect scores. However, it struggles with identifying premature ventricular contractions only 22.35% of these are correctly classified. Most of them are wrongly labelled as normal, which highlights the model's weakness in detecting certain types of abnormal heartbeats.

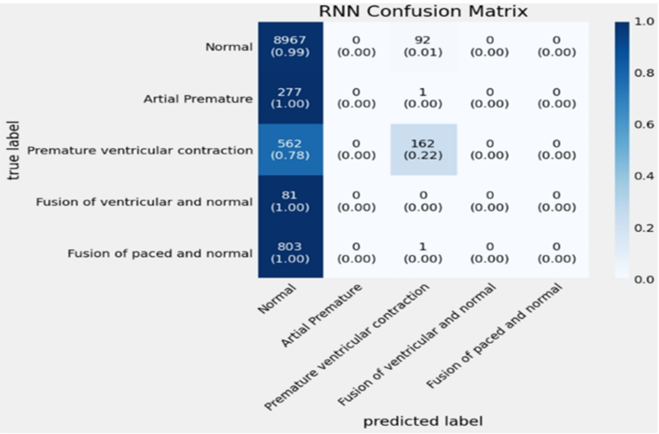



Figure 3: Confusion Matrix for RNN "testing set"

Table 4.3: The table reports the overall values of accuracy TPR, TNR, FPR, FDR, F1-Score

| Class | TPR (Recall) | TNR | FPR | FDR | F1 Score |
|---|---|---|---|---|---|
| Normal | 98.98% | 57.27% | 42.73% | 8.57% | 94.04% |
| Atrial Premature | 99.64% | 100.00% | 0.00% | 0.00% | 99.82% |
| Premature Ventricular Contraction | 22.35% | 99.97% | 0.3% | 1.81% | 35.63% |
| Fusion of Ventricular Contraction | 100.00% | 100.00% | 0.00% | 0.00% | 100.00% |
| Fusion of Paced and Normal | 99.88% | 99.98% | 0.02% | 0.25% | 99.00% |

*4.3.2. Confusion Matrix for GRU.* In the above fig. 4 shows the GRU model's confusion matrix for ECG classification, where it performs excellently on normal and premature ventricular contractions. The model correctly classifies most heartbeat types, especially improving detection of critical arrhythmias. Atrial premature and fusion beats show moderate accuracy, with some misclassifications. Overall, GRU outperforms the RNN in balanced and accurate heartbeat classification.

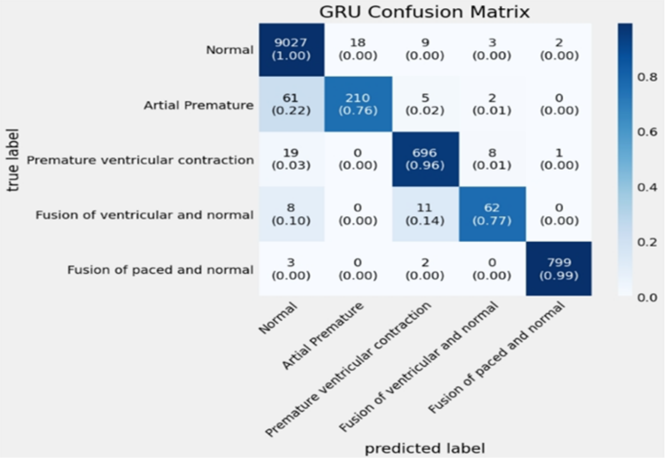



Figure 4: Confusion Matrix for GRU "testing set"

Table 4.4: The table reports the overall values of accuracy TPR, TNR, FPR, FDR, F1-Score

| Class | TPR (Recall) | TNR | FPR | FDR | F1 Score |
|---|---|---|---|---|---|
| Normal | 99.65% | 95.18% | 4.82% | 0.99% | 99.32% |
| Atrial Premature | 75.54% | 99.79% | 0.21% | 7.93% | 83.46% |
| Premature Ventricular Contraction | 96.13% | 99.82% | 0.18% | 2.24% | 97.26% |
| Fusion of Ventricular and Normal | 76.54% | 99.74% | 0.26% | 27.91% | 84.67% |
| Fusion of Pacec and normal | 99.38% | 99.89% | 0.11% | 1.35% | 99.36% |

In this table 4.4 shows how well the GRU model identifies different types of heartbeats. It performs almost perfectly on normal and paced beats, with F1 scores over 99%. For atrial premature and fusion beats, the accuracy is slightly lower, meaning a few were misclassified.

### 4.4. Comparative Analysis of RNN vs GRU

Both models were evaluated under identical experimental conditions, such as the same training, validation, and testing splits (80/10/10), preprocessing methods, and performance metrics, in order to determine how well the Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU) models classified ECG signals. Using the MIT-BIH Arrhythmia dataset, the goal was to identify the model with the best diagnostic abilities for classifying heartbeats.

Table 4.5: Comparative Analysis of RNN and GRU Model Performance

| Metric | RNN | GRU |
|---|---|---|
| Test Accuracy | 83.39% | 98.61% |
| Test Loss | 0.625 | 0.0555 |
| F1-Score (Macro Avg) | 0.18 | 0.92 |
| Misclassification | High (Minor classes | Low |
| Training Epochs | 10 | 20 + Early Stopping |
| Training Time | Lower | Higher |

The GRU model clearly outperformed the RNN in terms of accuracy, loss, and F1-score. Its superior performance can be attributed to its ability to capture temporal dependencies more effectively and manage long-range contextual information in ECG signal sequences. The RNN model, while simpler, suffered from vanishing gradient issues and struggled to correctly classify less frequent arrhythmia types.

## 5. Conclusion

In this manuscript, we developed and assessed deep learning models RNN and GRU for the classification of ECG signals using the MIT-BIH Arrhythmia dataset. In terms of accuracy and capacity to identify both normal and abnormal heartbeat classifications, the GRU model fared better than the conventional RNN. The GRU showed good performance across all five heartbeat categories, achieving approximately 98.61% test accuracy, whereas the RNN had trouble with minority heartbeat types. This suggests that for ECG signal classification tasks, GRU is a more dependable and effective model. Our results demonstrate the potential of sophisticated recurrent neural networks in creating intelligent medical systems that target the early identification of cardiovascular anomalies. Subsequent investigations may examine evaluating the model on a wider range of datasets or including attention mechanisms to enhance performance.

## Acknowledgments

We thank the organizers of the "International Conference on Mathematical Sciences and Computing– Innovations and Applications (ICMSC-2025)," jointly organized by the Department of Mathematics, North Eastern Regional Institute of Science and Technology (NERIST) and the Department of Mathematics, National Institute of Technology (NIT) Uttarakhand, held over 26–28 June 2025.

## References

1. Abubaker, Mohammed B., and Bilal Babayiğit, *Detection of cardiovascular diseases in ECG images using machine learning and deep learning methods*, IEEE transactions on artificial intelligence 4, no. 2 (2022): 373-382.

2. Acharya, U. Rajendra, Shankar M. Krishnan, Jos AE Spaan, and Jasjit S. Suri, eds., *Advances in cardiac signal processing*, Vol.8.Berlin: springer, (2007).

3. Ansari, Y., Mourad, O., Qaraqe, K., and Serpedin, E., *Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023*, Frontiers in Physiology, 14 (2023), 1246746.

4. Ebrahimi, Zahra, Mohammad Loni, Masoud Daneshtalab, and Arash Gharehbaghi, *A review on deep learning methods for ECG arrhythmia classification*, Expert Systems with Applications: X 7 (2020): 100033.

5. Liu, Xinwen, Huan Wang, Zongjin Li, and Lang Qin., *Deep learning in ECG diagnosis: A review*, Knowledge-Based Systems 227 (2021): 107187.

6. Lynn, Htet Myet, Sung Bum Pan, and Pankoo Kim., *A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks*, IEEE Access 7 (2019): 145395-145405.

7. Schwab, Patrick, Gaetano C. Scebba, Jia Zhang, Marco Delai, and Walter Karlen, *Beat by beat: Classifying cardiac arrhythmias with recurrent neural networks*, In 2017 computing in cardiology (CinC), pp. 1-4. IEEE, 2017.

8. Singh, Shraddha, Saroj Kumar Pandey, Urja Pawar, and Rekh Ram Janghel, *Classification of ECG arrhythmia using recurrent neural networks*, Procedia computer science 132 (2018): 1290-1297.

9. Sun, Jie, *Automatic cardiac arrhythmias classification using CNN and attention-based RNN network*, Healthcare Technology Letters 10, no. 3 (2023): 53-61.

10. Yao, Guoliang, Xiaobo Mao, Nan Li, Huaxing Xu, Xiangyang Xu, Yi Jiao, and Jinhong Ni, *Interpretation of electrocardiogram heartbeat by CNN and GRU*, Computational and Mathematical Methods in Medicine 2021, no. 1 (2021): 6534942.

*Bhagyashree Saikia,*
*Department of Mathematics,*
*North Eastern Regional Institute of Science and Technology,*
*India.*
*E-mail address:* bhagyashreesaikia16@gmail.com

*and*

*Pooja Debbarma,*
*Department of Electronics and Communication Engineering,*
*North Eastern Regional Institute of Science and Technology,*
*India.*
*E-mail address:* dbpooja25@gmail.com

*and*

*Rajesh Kumar,*
*Department of Electronics and Communication Engineering,*
*North Eastern Regional Institute of Science and Technology,*
*India.*

*E-mail address:* rk@nerist.ac.in

*and*

*M. M. Dixit,*
*Department of Mathematics,*
*North Eastern Regional Institute of Science and Technology,*
*India.*
*E-mail address:* mmd@nerist.ac.in