# Performance Comparison of Various Decision Tree Algorithms for Classification of Advertisement and Non Advertisement Videos

## B. Rebecca Jeya Vadhanam[1]*, S. Mohan[2] , V.V. Ramalingam[3] and V. Sugumaran[4]

[1]Research and Development Centre, Bharathiar University, Coimbatore – 641 046, Tamil Nadu, India;
beckyatsrm@gmail.com
[2]CCIS, AL Yamamah University, Riyadh, Kingdom of Saudi Arabia;
s.mohan77@gmail.com
[3]Department of Computer Science and Engineering, Faculty of Engineering and Technology, S.R.M University,
Kattankulathur – 603203, Tamil Nadu, India,
ramabi1976@gmail.com
[4]VIT University, Chennai Campus, Vandalur – 600127, Kelambakkam Road, Chennai, India,
v_sugu@yahoo.com

## Abstract

**Background/Objectives:** The main objective of the present study is to do the prerequisite process to develop a viewer-friendly electronic embedded system and business beneficial system to promote their products. This can be achieved by classifying the extracted Advertisement (ADD) videos from the Non-Advertisement (NADD) videos which consists of more visual information. **Methods/ Statistical Analysis**: The proposed frame work facilitates to identify the advertisement and non advertisement videos from the live stream television videos are discussed. The Block Intensity Comparison Code (BICC) technique is applied to extract the essential features from the ADD and NADD video frames. The frames are divided into various block sizes to select the best performing block size of the frame. The 8x8 frame size has been chosen as the promising block size to conduct the experiments. An extensive experimental analysis has been demonstrated with different classifier and a comparative study also reported. **Findings**: Decision tree algorithm (C4.5) has been employed to identify the vibrant features and these features are taken as the input to the various decision tree algorithms, namely J48, J48graft, LM tree, Random tree, BF tree, REP tree and NB tree to classify the video genre. A broad investigation has been made by a random tree algorithm which produced better predictive performance than the other algorithms. The training and the optimization of random tree model with their essential parametric measures are reported. Based on the overall study, random tree with BICC feature was found as the most preferred classification algorithm that achieved the 92.08% than the other algorithms. The classification capability and the performance evaluation of random tree algorithm with block intensity comparison code is reported and discussed for further study. **Application/Improvements:** The performance of the classifier can also be improved with other novel features.

**Keywords:** Advertisement (ADD) Videos, Block Intensity Comparison Code (BICC) Features, Classification, Non-Advertisement (NADD) Videos

---

*\*Author for correspondence*

# 1. Introduction

Advertisement monitoring system is a revolutionary change in internet advertising, television advertising and smart phone advertising. Advertising is a good communication idea, consisting both accurate and influential messages enable to reach the targeted customers. However, current television video system has two issues such as, the growing number of television viewers and advertisement channels. Therefore, it is very essential need to find a viewer-friendly electronic embedded system to provide a enjoyable moment for the television viewers and business beneficial system to promote their products. Advertisement detection enables to locate and skip the ADDs in live stream videos for watching the desired video programs effectively. In this present study, the work is mainly concentrated to develop a system that automatically recognizes and distinguishes the ADD videos from the NADD videos. The proposed system can later be extended with the help of other new features or existing features like color, edge, motion and histogram for classifying other types of videos. The proposed system for automatic TV advertisements detection is an essential approach for all kinds of multimedia analysis applications such as skipping advertisements for viewers and providing assistance for program segmentation. It also helps to monitor broadcast time of target advertisements and relaying the suitable advertisements to target customers. There are too many varieties of advertisements that are available in the TV stream. The varieties include movie, cartoons, sports and few random abstracts. First, it is a challenging task to find a common feature amongst them. Next, differentiating the definition of ADD videos and NADD videos is also another challenging task. In the proposed work, the intensity comparison technique is used for feature extraction and Decision tree classifier algorithms, namely J48, J48graft, LM tree, Random tree, BF tree, REP tree and NB tree have attempted to detect the advertisement clips from the general programs.

In[1] the proposed framework mainly focused few components to identify the advertisement videos. The video analysis was taken based on these components such as, syntactic and semantic video analysis, visual links, tags and product categories and highlighted keywords. In[2] an architecture that recommend custom-built advertisements for IDTV, Internet, and mobile devices. In[3] demonstrated clearly about the inserted ADD video content, semantic video content analysis and the virtual information of inserted ADD videos. In[4] the significance of principal component investigation for dimensionality reduction was outlined and used to minimize the data set. An effort has been made to draw out the temporal relationship of frames in videos through Hidden Markov Model (HMM) and the block intensity comparison code. The outcomes got from the tests demonstrate that the BICC features have performed well when contrast with different components like, edge, movement and histogram features. In the review, a novel text frame classification strategy was exhibited by Probable Text Block Selection (PTBS), Probable Text Pixel Selection (PTPS), Mutual Nearest Neighbor based Symmetry (MNNS). Moreover, the review represented the adequacy of existing text identification systems considered to misclassify non-text frames at both block and edge levels[5]. In[6] proposed an algorithm based on projective geometric invariant for feature tracking and also mentioned a new method for inserting a virtual advertisement into a video sequence. In[7] explained well about the feature selection process using decision tree algorithm. The rest of the paper is organized as follows. Section 2 describes the materials and methods of the study. Section 3 discussed the performance analysis of the classifier. Finally, section 4 concludes the paper.

# 2. Materials and Methods

The present study exposes the problem based on the visual perspective of the video and accordingly it is a very essential context to the recognition of the user. The observer perceives a video through the variations of the intensity values of the pixels. The video genre can be differentiated from the other genre by the individual attributes of their own. Further, the intensity distribution in the video frame will change in the rate of pixel values between the video classes. Regarding this investigation, the proposed work concentrated on a proficient method for classifying the video classes based on the intensity distribution in a frame. The BICC technique is applied on each frame to derive the features. A TV tuner card associated with a PC framework is utilized. The setup document has been introduced to record the recordings. The recordings are recorded in MPEG format of size 1024×1024. The

videos are recorded from different Tamil channels specifically from the TV live stream videos. A wide range of ADD videos are recorded under the ADD class. The news videos, sports videos, cartoon videos, movie videos, music videos, cookery shows, dance shows and adventure shows are recorded under the NADD class. The advertisement videos and non advertisement videos have been processed and segmented into 10 seconds videos. Then, the image frames were extracted at the rate of 25 frames per second from each 10 seconds video as shown in Figure 1.



**Figure 1.** Extraction of frames from video.

Totally, there are 20000 individual frames taken for the experiment and 5000 frames are taken as the test data. In the proposed work, each frame is divided into various block sizes like *2×2, 3×3, 4×4, 5×5, 6×6, 7×7, 8×8, 9×9* and *10×10* of the frame size *320×240*. The average intensity values are calculated for each block of a frame and compared with all the other blocks in the frame. Absolutely, there are 20000 individual frames taken for the test and 5000 frames are taken as the test information. In the proposed work, every frame is partitioned into different block sizes like 2×2, 3×3, 4×4, 5×5, 6×6, 7×7, 8×8, 9×9 and 10×10 of the casing size 320×240. The average intensity values are calculated for each block of a frame and compared with all the other blocks in the frame. The blocking pattern leads an efficient approach to improve the performance evaluation of the video classification. The database has been created for advertisement and non advertisement frames. The overall work has been exemplified in Figure 2.
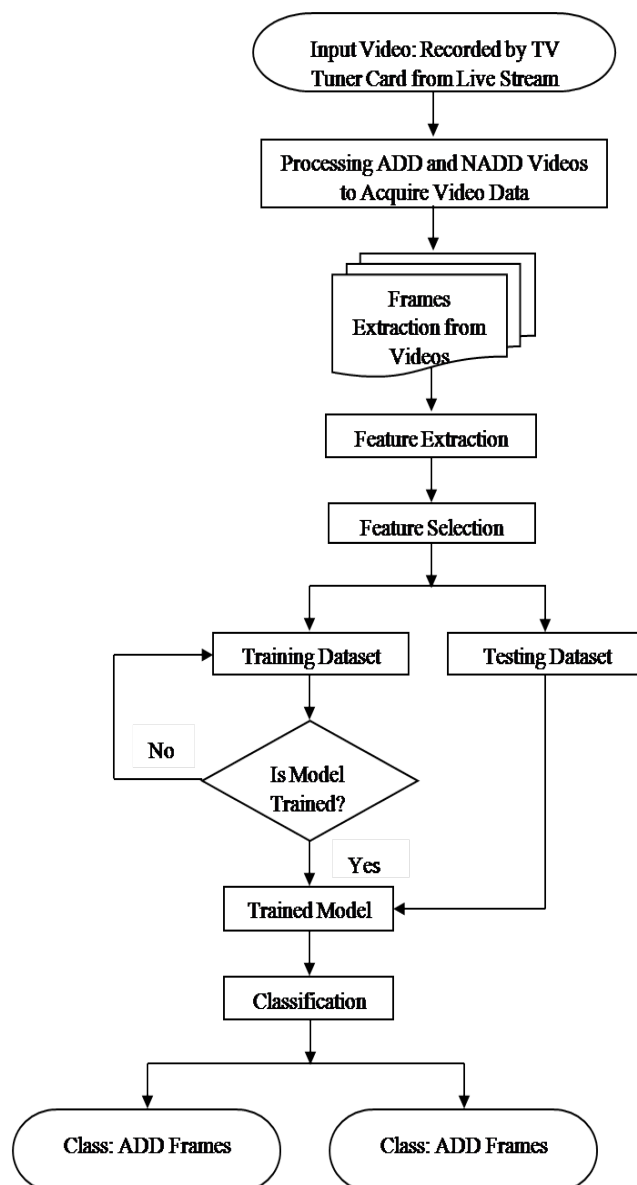


**Figure 2.** System architecture.

## 2.1 Feature Extraction

The similarity and difference of the ADD and NADD frames are identifies through the visual features. The visual features include the audio, text, images, motion and colour, etc. Here, Block Intensity Comparison Code (BICC) is applied on various block size of frames. The promising block size 8×8 has been selected to draw the useful features to carry out the experiment further. There are 64 highlighted features got from the 8×8 piece of every ADD and NADD outlines. The drawn 64 features were considered for the

present study. These features are the best confirmation for both static and dynamic properties. Grouping or distinguishing legitimate video by utilizing BICC features that gives important and discriminative information is favourable for high request exactness. The pseudo code of the component extraction process is given beneath.

Step 1: Each image or frame is divided into $K \times K$ blocks, where $K = 2, 3, 4, 5, 6, 7, 8, 9, 10$. Each of size $M/K \times N/K$, where $M, N$ is the size of the image.

Step 2: Select $8\times8$ block of image of size $320 \times 240$ used for the experimental study and test. The average intensity value is calculated for each block of a frame and compared with every other block in the frame. Average Intensity Value, Where $q=16$ here;

Step 3: Feature vector has been designed as follows:

$$Y [((i-1)*M) + 1: ((i*M),(j-1)*N)+1):(j*N)], \qquad (1)$$

Where, $M \times N$ size of the image. i, j is the average intensities of $i^{th}$ and $j^{th}$ block respectively.

Referring to Figure 3 the blocking example of the frame is utilized to identify whether the change is available or truant in every square of the edge. Blocking design likewise enhances the effectiveness of the elements to accomplish the best order exactness. The human visual view of the entity relies on upon the intensity variations. In view of this unique circumstance, the intensity changes between blocks of a frame in a video are represented by using block intensity comparison code. BICC has been utilized to produce the vector for $8\times8$ blocks of frames comprising of 1's and 0's which are utilized as feature vector.
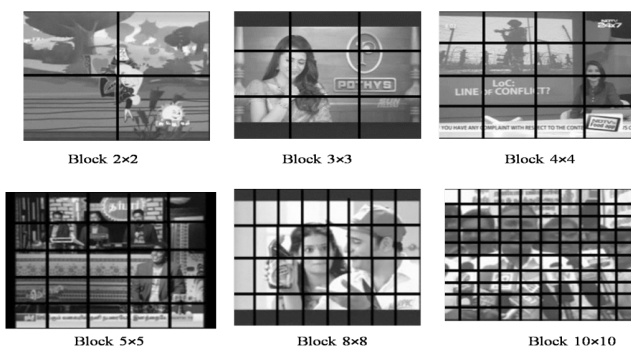


**Figure 3.** Sample frames divided into various block Size.

## 2.2 Feature Selection

Information mining techniques are utilized to burrow abundant information to get valuable data from the dataset. The dataset that will be mined may have a bigger size; subsequently, the calculation time will be more to mine the general information. Broadly, the time variable is an exponential function of the measurement of information. In this unique circumstance, the dimensionality diminishment method is utilized to accelerate the decision-making process. The process of feature selection has been executed utilizing the Information Gain and entropy measure[8]. The best performing components are chosen to enhance the exactness of video classification. Initially, the decision tree has been generated by a training data set. The feature which stays on the top of the tree is called root and that feature is the most important feature for classification based on entropy reduction. Then, next nodes down the root were considered. As the number of features increases, the classification accuracy increases up to a certain level, then it starts falling down. Here, all the features that appear in the decision tree have been chosen. With the pruned J48 algorithm, there were 19 well performing features (*h1, h5, h8, h10, h15, h20, h27, h31, h39, h42, h47, h50, h52, h54, h55, h56, h57, h58* and *h61*) selected out of 64 features derived from BICC features of $8\times8$ block size of the frame. The rest of the components are intentionally overlooked for the further study.

## 2.3 Classification Algorithm

The objective of this present study aimed to investigate the performance of different classification algorithm using the WEKA tool for ADD and NADD video classification. Classification algorithms could be compared on the basis of predictive accuracy, robustness, scalability and interpretability criteria. The performance of different techniques must be evaluated to attain the ultimate model for video classification. The best model can be built with the aid of machine learning algorithm to obtain the end results. The performance applicability and their evaluation of decision tree J48, J48 graft, LMT, Random tree, BF tree, Rep tree and NB tree is discussed in the following subsections.

### 2.3.1 Decision Tree J48

In this experiment, the performance of decision tree J48 has been evaluated and compared with other algorithms.

Classification algorithms always find a rule or set of rules to represent the data and classified into classes. The decision tree is a popular and simple structure that uses "divide and conquer" technique to break down a complex decision making process into a collection of simple decisions. The decision tree mechanism is transparent and thereby providing an interpretable solution. Given a database D = {t1, t2, ….., tn} where ti = {ti1, ti2, ….., tih} and the database schema contains the attributes {A1, A2, A3,….., Ah}. It is also given a set of classes C = {1,….., m}. A decision tree computational model associated with D that has the following properties[9].

- Each internal node is labelled with an attribute, *Ai*.
- Each arc is labelled with a predicate that can be applied to the attribute associated with the parent.
- Each leaf node is labelled with a class, *Cj*.

Given a set of classes C = {1, ……., m} with equal probability of occurrence the entropy is - p1 log2 p1 – p2 log2 p2 ………- pm log2 pm where pi is the probability of occurrence of i. Attribute with the lowest entropy is selected as split criteria for the tree. Tree pruning is done is a bottom-up fashion. It is used to improve the prediction and classification accuracy of the algorithm by minimizing over-fitting.

### 2.3.2 J48 Graft

J48graft is an algorithm having purposed to increase the probability of classifying rightly the instances. This algorithm creates only single tree and it reduces prediction error. J48 graft algorithm for generating grafted decision tree from a J48 tree algorithm. The purpose of this grafting algorithm is to increase the probability of correctly classifying instances that fall outside the areas covered by the training data. The grafting technique is an inductive process which adds nodes to infer decision trees with the purpose of reducing prediction errors. The J48 grafting algorithm provides the best general prediction accuracy over a representative selection of the learning process[10].

### 2.3.3 Logistic Model Tree

A Logistic Model Tree (LMT) basically consists of a standard decision tree structure with logistic regression functions at the leaves. The LMT consists of a tree structure that is made up of a set of inner or non-terminal nodes and a set of leaves or terminal nodes. The Logistic Model Tree algorithm makes a tree with binary and multiclass target variables, numeric and missing values. LMT is a combination of induction trees and logistic regression. LMT uses cost-complexity pruning. This algorithm is significantly slower than the other algorithms[11].

### 2.3.4 Random Tree

Random Tree (RT) is an efficient algorithm for constructing a tree with K random features at each node. Random tree is a tree which drawn at random from a set of possible trees. Random trees can be generated efficiently and the combination of large sets of random trees generally leads to accurate models. Random tree models have been extensively developed in the field of Machine Learning to build a suitable and accurate model for video classification[12].

### 2.3.5 Best First Tree

In Best-First (BF) decision tree algorithm, the tree expands by selecting the node which maximizes the impurity reduction among all the existing nodes to split. In this algorithm, the impurity could be measured by the Gini index and information gain[13]. BF tree are constructed in a divide-conquer method similar to the standard depth-first decision trees. The basic step for constructing the best-first tree is given below.

- Select an attribute to place at the root node and make some branches for this attribute based on some criteria.
- Split training instances into subsets, one for each branch extending from the root node.
- Constructing process continues until all nodes are pure or a specific number of expansions are reached.

### 2.3.6 Reduced Error Pruning Tree

Reduced Error Pruning (REP) Tree is the simplest and most understandable technique in decision tree pruning. It is a fast decision tree learner, which builds a decision or a regression tree using information gain as the splitting criterion and prunes it using reduced error pruning. Using REP algorithm, the tree traversal has performed from bottom to top and then checks for each internal node and replace it with most frequently class with the most concern about the tree accuracy, which must not reduce. This procedure will continue until any further pruning will decrease the accuracy[14].

### 2.3.7 Naïve Bayes Tree

A Naive Bayes (NB) classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong independence assumptions. Naïve Bayes classifiers can handle an arbitrary number of independent variables, whether continuous or categorical. The algorithm makes predictions using Bayes theorem which incorporates evidence or prior knowledge in its prediction[15,16]. Given a set of variables $X = \{x_1, x_2, \ldots, x_d\}$, the posterior probability can be constructed for the event Cj among a set of possible outcomes $C = \{c_1, c_2, \ldots, c_d\}$. Simply put, X is the predictors and C is the set of categorical levels present in the dependent variable. Using Bayes rule:

$$P(C_j / x_1, x_2, \ldots x_d) \, \alpha \, p(x_1, x_2, \ldots x_{d \setminus} C_j) \, p \, (C_j) \qquad (2)$$

Where, $p \, (Cj \mid x_1, x_2, \ldots, x_d)$ is the posterior probability of class membership.

# 3. Results and Discussion

To investigate the performance of the selected classification algorithm from decision tree family the same experimental procedures have been followed as suggested by WEKA tool. Decision tree classifier algorithms are potentially powerful predictors and characterize the structure of the dataset. In the present study, a large data set contains 20,000 instances of the ADD and NADD class. The classification has been done using 10-fold cross-validation with the default percentage split is 66% for all the classifiers is summarized in the following subsections.

## 3.1 Performance Evaluation of Decision Tree J48

Decision tree J48 algorithm is employed to build the model for class ADD and NADD. Here, the applied algorithm evaluates the BICC feature and prepares a decision table which demonstrates and distinguishes total number of instances into two different classes of ADD and NADD classes. With BICC feature the classification accuracy was generated using J48 decision tree algorithm. The minimum number of objects (variable parameter) required to form a class (M) was varied from 1 to 10000 (total no.of.instances per class) and the corresponding classification accuracy were noted down for further study. The value of 'M' which gives the maximum classification accuracy was fixed and confidence factor was varied from 0 to 1 in steps of '0.1'. The best classification accuracy of 83.69% of the block size 8x8 of frames was achieved with M value of 100 and confidence factor of '0.25'.

## 3.2 Performance Evaluation of J48 Graft

In this experimental study, Class for building model by J48 graft algorithm shows the same level of accuracy which obtained in decision tree J48 classifier algorithm. There was no major difference in their experimental results. But, J48 classifier takes less time to build the model than J48 graft classifier. J48 graft algorithm takes 1.92 seconds to build the model and J48 classifier only takes 1.50 seconds to build the model. Here also, the minimum number of objects (variable parameter) required to form a class (M) was varied from 1 to 10000 (total no.of.instances per class) and the corresponding classification accuracy were noted down for further study. The value of 'M' which gives the maximum classification accuracy was fixed and confidence factor was varied from 0 to 1 in steps of '0.1'. The best classification accuracy of 83.69% of the block size 8x8 of frames was achieved with M value of 100 and confidence factor of '0.25'.

## 3.3 Performance Evaluation of Logistical Model Tree (LMT)

Class for building a LMT decision tree classifier contains a standard decision tree structure with logistic regression function at the leaves. LMT is significantly outperforming when compared with other models. Minimum number of instances (variable parameter) was varied from the default value 15 to 10000 in steps of 5 and the corresponding classification accuracy was noted down for further experimental study. The best classification accuracy of 91.345% of the block size 8x8 of frames was achieved in 10-fold cross validation with the default percentage split 66%. Looking at this ranking for classification accuracy, one could find LMT and Random tree very close to the top and achieved the best classification accuracy. Table 1 shows the computational efficiency of the selected classifier model. LMT is very slow compared to the other algorithms. This is due to the slow estimation process for the parameters of the logistic model performed by the Logit Boost algorithm. If the computational efficiency is increased, LMT would be the best model for video classification.

## 3.4 Performance Evaluation of BF Tree

In this experimental study, the same training dataset was used to evaluate the performance of BF tree. The Best-first decision tree performs the best split in the tree based on boosting algorithms. Minimum number of instances (variable parameter) was varied from the default value 2 to 10000 in steps of 1 and the corresponding classification accuracy was noted down for further experimental study The best classification accuracy of 83.515% for the block size 8x8 of frames was achieved with M value of 100 in 10-fold cross validation with the default percentage split 66%.

## 3.5 Performance Evaluation of Rep Tree

Observing in the result of accuracy and compilation efficiency, one can also easily find that the REP tree model has taken only minimum compilation time of 0.52 seconds to build the model. A minimum total weight of instances (Variable parameter) was varied from 2 to 10000 in step of 1. The REP tree model has achieved the classification accuracy of 80.64% of the block size 8x8 of frames with M value of 100 in 10-fold cross validation with the default percentage split 66%.

## 3.6 Performance Evaluation of NB Tree

From the experimental study, it is observed that the highest relative error is found in the Naïve Bayes" classifier. NB tree model achieved the lower classification accuracy of 79.61% where the rest of the algorithm achieved ranging 80-92% classification accuracy. An algorithm which produced a lower error rate and maximum level of accuracy will be preferred as it has the more powerful classification capability.

**Table 1.** Compilation time to build models.

| Decision Tree Algorithm | Time taken to build model (s) |
| --- | --- |
| J48 | 1.50 |
| J48 Graft | 1.92 |
| LMT | 157.39 |
| Random Tree | 0.56 |
| BF Tree | 13.31 |
| Rep Tree | 0.52 |
| NB Tree | 28.81 |

The above mentioned various decision tree classifier algorithms evaluate the BICC features and construct the decision table. One could find from the given results, the Random Tree algorithm provides best classification accuracy rates with the best performing features. Though the Random Tree and LMT achieved the same level of classification accuracy, the time complexity is higher in LMT algorithm than Random tree. The time taken for building a model through LMT classifier is 157.39 seconds and for Random tree has taken only 0.56 seconds. Hence, Random Tree classifier is chosen as the best model for video classification of ADD and NADD videos. Further, the study is focused on the Random Tree algorithm to build a better model for video classification. An extensive investigation has made by Random Tree algorithm under the performance measures.

## 3.7 Performance Evaluation of Random Tree

An extensive investigation shows Random Tree model produced a high predictive performance, which is competitively compared with other algorithms. The training and optimization of Random Model Trees scales better than J48, J48 graft, LMT, BF tree, REP tree and NB tree. The performance of the random tree model is evaluated in tuning of parameters. The variations of the classification accuracy with respect to default K-value = 15 is shown in Figure 4.
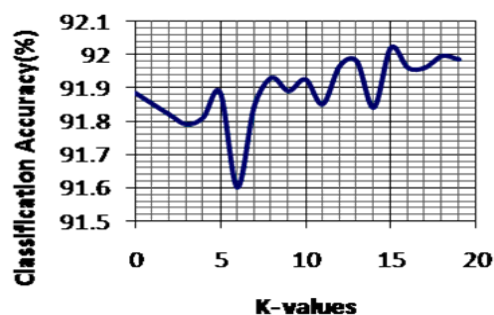


**Figure 4.** Classification performance of random tree model.

The K-values varied from 1 to 19 (total no. of. best performing features) and the corresponding classification accuracies were noted down. The value of K =15, which gives the maximum classification accuracy was fixed with the consideration of the apparent structure of the decision tree and the value of maximum depth was varied from 0 to 19 in steps of 1.

The variation of classification accuracy with respect to maximum depth of the tree and folds is shown in Figure 5 and 6. The best classification accuracy of 92.085 % of the block size 8x8 of frames was achieved with maximum depth value of 19. Then the proposed work has been extended with the tuning parameters of folds. The number of folds varied in the decreased order from 11 to 0 with a fixed value of k =15 and depth = 19. Thus, the random tree classifier is able to achieve the maximum classification accuracy of 92.085 % with the minimum default folds.
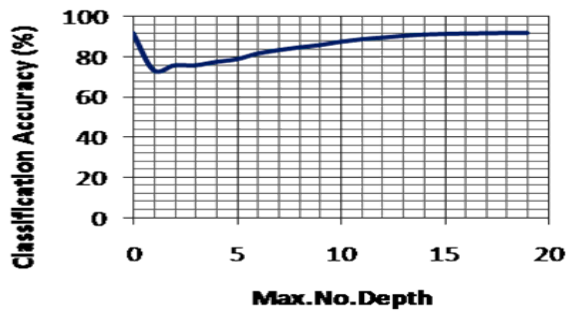


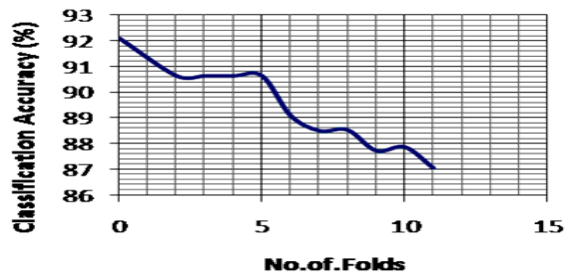**Figure 5.** Performance of random tree depth vs classification accuracy.



**Figure 6.** Performance of random tree folds vs classification accuracy.

With reference of Table 2, one can consider the calculation of true positive values, true negative values, and false positive and false negative values. All measures of sensitivity, specificity and accuracy can be calculated based on these values. Finding the TP rates and FP rates is very essential part to conclude the best model. The TP rate should be close to '1' and FP should be close to '0' for better classification accuracy. From the given Table 2, one can realize the closeness of TP rate to '1' and FP rate to '0'. Thus, both values confirm that the built model is the best one of the other models.

**Table 2.** Detailed accuracy by class.

| Detailed Accuracy by Class | | | | | | |
|---|---|---|---|---|---|---|
| TP Rate | FP Rate | Precision | Recall | F-Measures | ROC Area | Class |
| 0.952 | 0.11 | 0.897 | 0.952 | 0.923 | 0.96 | ADD |
| 0.89 | 0.049 | 0.948 | 0.89 | 0.918 | 0.96 | NADD |
| 0.921 | 0.079 | 0.922 | 0.921 | 0.921 | 0.96 | Weighted Avg. |

## 3.8 Interpretation of the Confusion Matrix

Confusion matrix is a resultant table layout to demonstrate the actual and predicted classifications of the classifier is shown in Table 3. From the observations, 485 data points of ADD class were misclassified as NADD class. And 1098 data points of NADD class were misclassified as ADD class. The training data for two classes of advertisement video data and the non advertisement video data shows the user accuracy of 95.15% and 89.02% respectively.

**Table 3.** Confusion matrix.

| Class | ADD | NADD |
|---|---|---|
| ADD | 9515 | 485 |
| NADD | 1098 | 8902 |

## 3.9 Comparative Study

Table 4 shows the overall classification accuracy of the various machines learning algorithm such as decision treeJ48, J48graft, LM tree, Random tree, BF tree, REP tree and NB tree. One can observe that the Random tree classifier gives the maximum classification accuracy for the proposed problem. Figure 7 and 8 shows the example video frames of ADD and NADD videos.

**Table 4.** Classification accuracy of various classifiers - a comparative study.

| Name of the Classifier | Variable Parameter | Classification Accuracy(%) |
|---|---|---|
| Decision tree- J48 | Minimum no. of objects | 83.69 |
| J48 graft | Minimum no. of objects | 83.69 |
| LMT(Logistical Model Tree) | Minimum no. of instances | 91.34 |
| **Random Tree** | **K values** | **92.08** |
| BF Tree (Best First) | Minimum number of objects | 83.51 |
| Rep Tree(Reduced Error Pruning) | Minimum total weight of instances | 80.64 |
| NB Tree(Naïve Bayes) | Naïve Bayes Algorithm | 79.61 |



**Figure 7.** Samples ADD video clips taken from advertisements.



**Figure 8.** Samples NADD video clips taken from general programs.

# 4. Conclusion

In the present study, the video classification of ADD and NADD is taken up with the help of BICC features and various decision trees. The results show that the random tree model is the suitable model which is produced the best classification accuracy of 92.08% than the other algorithms. Random Tree approach with BICC features could classify the ADD video frames from NADD video frames with minimum rate of misclassifications error in a huge volume of video data set. The computational simplicity and the best classification accuracy of this approach enable the novel applications such as automatic channel changes, skipping the advertisement videos and live interactive video shows. The proposed work would be extended to build the best classification model to increase the versatility of the video classification system with other feature likes edge, motion and histogram and with new classifiers. Therefore the findings of the study will surely help the busy current generation to skip the nuisance of advertisements to enjoy watching their favourable shows of various television channels.

# 5. References

1. Jinqiao Wang, Min Xu, Hanqing Lu, Ian Burnett. ActiveAd: A Novel Framework of Linking Ad Videos to Online Products, Neurocomputing, 2016; 185:82–92.
2. Pessemier, Deryckere, Vanhecke, Martens. Proposed Architecture and Algorithm for Personalized Advertising on IDTV and Mobile Devices, IEEE Trans. Consum. Electron. 2011; 54(2):709–13.
3. McCoy Scott, Everard Andrea, Polak Peter, Dennis F Galletta. The Effects of Online Advertising, Communications of ACM - Emergency Response Information Systems: Emerging Trends and Technologies. 2007; 50(3):84–88.
4. Kalaiselvi Geetha M, Palanivel S, Ramalingam V. A Novels Block Intensity Comparison Code for Video Classification and Retrieval, Expert Systems with Applications. 2009; 36(3):6415–20.
5. Palaiahnakote Shivakumara, Anjan Dutta, Trung Quy Phan, Umapada Pal. A Novel Mutual Nearest Neighbor based Symmetry for Text Frame Classification in Video, Pattern Recognition. 2011; 44(8):1671–83.
6. Wei Chen, Jeng-Sheng Yeh, Tsorng-Lin Chia. Automatic Insertion of Advertisements into a Video Sequences, Proceedings of the 9th International Conferences on Machine Learning Cybernetics, Qingdao, 2010.

7. Sugumaran V, Muralidharan V, Ramachandran KI. Feature selection using Decision Tree and Classification Through Proximal Support Vector Machine for Fault Diagnostics of Roller Bearing, Mechanical Systems and Signal Processing, 2007; 21(2):930–42.

8. Aha, DW, Kibler, D. Albert MK. Instance-based Learning Algorithms, Machine Learning. 1991; 6(1): 37–66.

9. Quinlan JR. Introduction of Decision Trees, Machine Learning. 1986; 1(1):81–106.

10. Trilok CS, Manoj J. WEKA Approach for Comparative Study of Classification Algorithm, International Journal of Advanced Research in Computer and Communication Engineering. 2013; 2(4):1925–31,

11. Landwehr N, Hall M, Frank E. Logistic Model Trees, Machine Learning. 2005; 59(1-2):161–205.

12. Wang Y, Witten IH. Induction of Model Trees for Predicting Continuous Classes, Proc European Conference on Machine Learning Poster Papers; 1997, p. 128–37.

13. Haijian Shi. Best-First Decision Tree Learning, Master's thesis, University of Waikato, Hamilton, NZ, COMP594, 2007.

14. Quinlan JR. Simplifying Decision Trees, Internal Journal of Human Computer Studies - Special Issue: 1969-1999, the 30th Anniversary. 1999; 51(2):497–510.

15. Fauset L. Fundamentals of Neural Network Architecture, Algorithms and Applications, Prentice Hall, NJ, 1994.

16. Rebecca Jeya Vadhanam B, Mohan S, Sugumaran V. Application of Artificial Immune Recognition System for Identification of Advertisement Video Frames using BICC Features, Indian Journal of Science and Technology. 2016 Apr; 9(14). DOI: 10.17485/ijst/2016/v9i14/86635.