IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 September 2023, accepted 7 October 2023, date of publication 19 October 2023, date of current version 26 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3325888

RESEARCH ARTICLE

Comparison of Machine Learning Classification and Clustering Algorithms for TV Commercials Detection

EMAN ABDELFATTAH[1] AND SHREEHAR JOSHI[2,3]

[1]School of Computer Science and Engineering, Sacred Heart University, Fairfield, CT 06825, USA
[2]School of Theoretical and Applied Science, Ramapo College of New Jersey, Mahwah, NJ 07430, USA
[3]The Boring Company, Las Vegas, NV 89169, USA

Corresponding author: Eman Abdelfattah (abdelfattahe@sacredheart.edu)

ABSTRACT One of the essential aspects of broadcast monitoring is to detect and consequently extract commercial blocks in telecast news videos. The research carried out until now have based their work almost entirely on preconceived characteristics that are associated with a channel. With the advertisers constantly looking to work around the existing policies, the reliance on the nature of channels during an advertisement does not suffice. The other approach towards identifying a commercial is by frequentist approach. However, it is often the case that sponsored programs and other programs share similar time in any specified hour, rendering the frequentist approach almost useless in the process. As such, this paper uses machine learning based approach which is more generic and can employ inherent differences that commercials have over their non-commercial counterparts for classifying and clustering commercials in the news videos. The datasets which contain 90 hours of recordings from five different news channels from US, England and India have been used to train and test nine different classifiers – K Neighbors, Support Vector Machine, Decision Tree, Random Forests, Ada Boost, Gradient Boost, Gaussian NB, Linear Discriminant Analysis, and Quadratic Discriminant Analysis – and five different clustering algorithms – K Means, Agglomerative, Birch, Mini-Batch K Means, and Gaussian Mixture. Our results show that the Random Forests outperforms all the other classifiers used with respect to F_1 score and median time to train and test on each of these datasets that consists of features of shots extracted from 18 hours of video. Similarly, Mini Batch K Means was found to perform the best for forming clusters of news and commercials.

INDEX TERMS TV commercial detection, machine learning, classification, clustering.

I. INTRODUCTION

The big corporate houses resort to TV commercials in their bid to sell the products to as many people as possible even to this day. In fact, the largest share of investment in advertisement is still in TV commercials. It may seem absurd that a second of broadcasting of an advertisement during the airing of Super Bowl was priced as high as 5 million dollars in 2020. However, considering the fact that the number of viewers has been northwards of 110 million in recent years, the investment is likely to return huge on the investment as there is a good chance fair few percentages of these viewers will be converted into lifelong customers. Commercials help sustain the TV stations as well, with the amount of revenue that is generated through it. Since hefty money is involved in these commercials, and because it has the potential to reach out to a huge number of people, policies are in place to regulate TV advertisements. These policies have not only avoided the misuse of commercials but have also helped shape the nature of such commercials. As for instance, there can only be certain number of shares of TV advertisements in an hour of broadcasting.

TV commercials show distinct features from other forms of programming. Techniques like loud sounds, high cut rates and high motion scenes are strategically applied to attract the audience and keep them hooked throughout the

The associate editor coordinating the review of this manuscript and approving it for publication was Qilian Liang.

span of the advertisement. Other features may result from the requirement of the legislation; removal of TV logo altogether during the broadcasting of advertisement is one such example. The onset of digital era in television has spurred the developers to build systems that automatically detect TV commercials. Such detection systems may fulfill several other needs as well. They could help remove the transmitted commercials from the recorded TV programs of the viewers. In addition, viewers can test for themselves if the TV advertisement standards are good enough. Likewise, the system could also be useful to the advertisers to monitor if their contract is violated by the broadcasters. A broadcaster is expected to oblige to certain contractual agreements. They need to air certain ads for the agreed amount of time and this system would help even the advertisers to monitor whether the broadcasters would stay true to their commitment.

Knowledge-Based Approaches: These methods exploit the inherent differences between a commercial and other segment of videos such as the existence of delimiting black frames and silence periods, high video cut rates and faster motion replacements of commercials, disappearance of channel logo and higher audio intensity during the commercials, and text detections to identify commercials from non-commercials block. Most of these approaches rely on channel-specific assumptions and only a few on algorithm-based generic approach.

Repetition-Based Approaches: These methods rely on fingerprinting video segments of commercials and matching them as they are broadcasted at distinct times after storing their fingerprint to identify the occurrence of a commercial block from their non-commercial counterparts. These methods are more generic than the knowledge-based approaches but more computationally expensive as it requires computing and storing of information of each of the commercials being displayed in the specific channel.

Our research work falls under the knowledge-based approach in that various machine learning algorithms have been deployed to train and test on data that have been extracted based on existing knowledge of the commercials. The machine learning algorithms have been selected based on their past performances in various research within and outside of this specific domain. The classification algorithms used in this research are Decision Trees, Random Forests, Linear Support Vector Machine (Linear SVC), K Nearest Neighbors (K Neighbors), Gaussian Naïve Bayes, Adaptive Boosting (Ada Boost), Gradient Tree Boosting (Gradient Boosting), Linear Discriminant Analysis (LD Analysis) and Quadratic Discriminant Analysis (QD Analysis); and the clustering algorithms used are K Means, Mean Shift, Density-based spatial clustering of applications with noise (DBSCAN), Gaussian Mixture, and Agglomerative Hierarchical Clustering.

A. RELATED WORKS

The detection of advertisement in broadcast is not a new concept and efforts have been made to find optimal methodology over the course of time. Knowledge based method and frequentist (repetitive) method are the most widely adopted approaches. Frequentist approach assumes that the commercial segments repeat in higher frequencies compared to the news, discussions, and other non-commercials. These approaches work on the principle of fingerprinting and hashing of audio-visual features. Here, the assumption is that the commercials are the video segment that are aired more frequently than other forms of programs [1], [2].

Knowledge based methodologies seem to be the preferred approach deployed to filter the advertisements. Ramires *et. al.* published an article that illustrated the means to identify a segment as commercial by taking exclusively the audio features into account [3], while a video-only approach was used by Qian *et. al.*, in which they used color histograms and Short Time Energy (STE) to separate advertisements [4]. Similarly, Sadlier *et. al*. proceeded with the assumption of the presence of blank frames and depression in audio levels. They used MPEG coded video features to filter the advertisements [5].

Classifying a segment as advertisement has also been done based on whether a logo is present on the channel when the advertisement is being telecasted. In countries with stricter ad-policy, the product of logos is displayed during the commercials, while the logo of the channel is not in display. Several techniques have been developed to detect an advertisement by the presence of logos [6], [7], [8]. In some cases, there are no logos at all. This characteristic has also been used to distinguish ad in broadcasting segment [9].

Several researchers have used machine learning techniques by pre-training classifiers on audio-visual features extracted from commercials. These techniques include simple threshold-based classification (MPEG features) [10]. The use of finite state machines has yielded better result [11], [12]. The experimental results from the work of Wang and Gou in particular yielded recall of 96.47% and precision of 97.27%, which when compared to the existing main approaches of the time was far superior. SVM [13] and more recently an interactive ensemble learning method called Tri-Adaboost have also been used for automatically classifying a segment as advertisement [13].

Mourya *et. al*. approached the problem by converting the video frames into images and used a technique called block-based background subtraction along with frame retrieval, color conversion and frame comparison to mark the beginning and the end of an ad. As the ad would be detected, channels were automatically switched onto the next one [14].

Waseemullah *et. al*. came up with a unique strategy to solving the problem. They used Base64 encodings to separate commercials in broadcast segment. The reason behind the novelty was due to the typical nature of Pakistani commercials. There could be silence even in the middle of the advert segment and the past works could not be properly implemented. Also, due to lack of media legislation, blanks cannot be properly identified to mark the starting and ending points and the channel logo persists even during the airing of commercials. Instead, video frames were converted into texts and the claim is that computational cost is greatly reduced in this framework. They were able to detect ads with an accuracy of 60% [15].

Deep Learning approaches have also been applied to detect advertisements. Luo et al. [16] and Hossari et al. [17] have achieved better results through deep learning algorithms. Toheed *et. al.* employed AlexNet deep learning model and WKNN for advertisements detection and removal from the broadcasts. They attained an average accuracy of 99.37% [18].

Vyas [19] proposed a method that they claim outperforms all the previous works based on audio-visual feature extraction. Audio feature was extracted by Mel-Frequency Cepstral Coefficient (MFCC) bag of words and overlaid text distribution defined video features. These features were categorized with SVM classifiers. Experimented on videos spanning over 54 channels acquired from 3 different channels, the process yielded an astonishing F_1-score of 97%.

Du et al. proposed a new clustering technique [20].The authors used eight synthetic datasets and ten real-world datasets in their experiments. This proposed approach aimed to enhance the performance of clustering when there are some overlapping regions between different clusters.

Pei et al. introduced a new clustering technique [21]. The authors used 18 real-world datasets, with most of them being face images datasets. The proposed algorithm achieved comparable results with existing clustering methods but with lesser time complexity.

Table 1 summarizes the comparison of the approaches used, and the machine learning (ML) techniques employed when applicable, and the types of features utilized between the multiple related research works and this study.

B. MOTIVATIONS AND RESEARCH PROBLEM

News channels across the globe vary in the way commercials are presented and channel specific features can hardly be generalized. Similarly, the frequency with which commercials are shown can greatly vary based on national laws and the agreements made between the channels and the companies whose products are being advertised. To account for these shortcomings, machine learning algorithms, which are both generic and equally efficient, have been deployed in this research to detect commercials.

To the best of the knowledge of the authors, no commercial detection works in the past have tested the efficiency of their system on news channels from more than one country. The algorithms in this research, however, have been trained and tested on data from five different news channels from three different countries. This ensures that the results of this research are more generalizable than any other similar work in the past. In addition to the detection of commercials, an effort for clustering, which is relatively uncommon compared to the task of classification in this domain, has also been made in this research. The clustering algorithms also have been compared to each other across the channels from three different countries in the same way classification models are done.

Lastly, various state-of-the-art audio and video features that have been proposed in various stages of the development of TV Commercials Detection have been used to assess the efficiency of the algorithms used in this research. As such, the authors believe that this research can provide a solid foundation with sufficient breadth and depth for machine-learning based generic TV Commercials Classification and Clustering.

II. FEATURES OVERVIEW

TV commercials have a specific set of characteristics designed with the main aim of attracting their audiences. One of them is high cut rates in their scenes, which are animated using multiple short video segments transitioning abruptly or with fades and dissolves. Another such feature is the presence of text with short descriptions of the product, which might change its position as the commercial progresses. Their audio levels are also usually loud to make them stand out from the non-commercial blocks and usually consist of jingles such as the music of a brand or other background music that suits the scenes. Commercials are usually found in a block with a separator in the beginning and the end of the block and in most of the countries, they do not contain the logo of the news channel they are being displayed at. Not to be forgotten, TV Commercials usually have a typical time duration of 45 - 60 seconds and within the range of the multiple of 5.

Taking these distinct characteristics of commercials into consideration, each of the five datasets used for this research consists of around 30000 instances of seven audio and five visual features extracted from video segments - calculated using RGB Color Histogram Matching - from CNN, CNNIBN, NDTV, TIMESNOW, and BBC news channels [22]. The audio and visual features are as follows:

- Audio Features
- Short Time Energy (STE)
- Zero Crossing Rate (ZCR)
- Spectral Centroid
- Spectral Flux
- Spectral Roll-Off Frequency
- Fundamental Frequency
- Video Features
- Motion Distribution
- Frame Difference Distribution
- Text area Distribution
- Bag of Audio Words
- Edge Change Ratio

III. MACHINE LEARNING ALGORITHMS

The description of nine different classification and five different clustering algorithms have been provided in their subsections below:

TABLE 1. Comparison of this work with other related works.

ARTICLE	APPROACH	ML	CLASSICATION	CLUSTERING	FEATURES
Covell et al. [1]	Repetition-based	X	X	X	Audio and Visual
Duygulu et al. [2]	Repetition-based	X	X	X	Audio and Visual
Sadlier et al. [5]	Knowledge-based	X	X	X	Audio and Visual
Feng et al. [7]	Knowledge-based	X	X	X	Audio and Visual
Banić [9]	Knowledge-based	X	X	X	Audio and Visual
Dimitrova et al. [10]	Knowledge-based	X	X	X	Audio and Visual
Hua et al. [12]	Knowledge-based	✓	✓	X	Audio and Visual
Mourya et al. [14]	Knowledge-based	X	X	X	Audio and Visual
Waseemullah et al. [15]	Knowledge-based	X	X	X	Audio and Visual
Luo et al. [16]	Knowledge-based	✓	✓	X	Audio and Visual
Hossari et al. [17]	Knowledge-based	✓	✓	X	Visual
Toheed et al. [18]	Knowledge-based	✓	✓	X	Audio and Visual
Vyas et al. [19]	Knowledge-based	✓	✓	X	Audio and Visual
This work	**Knowledge-based**	✓	✓	✓	**Audio and Visual**

A. CLASSIFICATION ALGORITHMS

Random Forests [23] is a classification model that consists of multiple numbers of decision trees - with very low to no correlations to one another - and uses the results of these trees to make the final voting decision. It is inherently based on the concept of "wisdom of crowds" in that it uses the majority voted class as the final class for its prediction. This ensures that while the output of any given decision tree might be incorrect or off from the actual value, the cumulative result based off results of all the decision trees will be more accurate or closer to its correct value.

Decision Tree [24] is an algorithm that uses a set of choices, actions, and their consequences, much like a large network of conditional control statements, to make a prediction for the given task. It consists of a root node (the starting point), which splits up into two decision nodes, which often splits up continuously - just like the root node making distinct branches, until it reaches the leaf node which decides the outcome of the classifier. This model usually requires less data cleaning than other modeling techniques but fails on large datasets due to the limitation on producing only binary outcomes on decision nodes.

Support Vector Machine [25] or simply SVM or SVC, is a classification and regression algorithm that works by finding one or a set of hyperplanes that divide the data from different classes into distinct compartments. The hyperplanes are chosen in such a way that they maximize the distance between them and the data points belonging to different classes all the while ensuring that the points belong to their correct compartments. In cases where the points cannot easily be divided with a 2D hyperplane (a line), the points are mapped to higher dimensions in the process which is referred to as kernelling.

K Nearest Neighbors [26], [27], [28] is a supervised machine learning algorithm that works based on the assumption that the data points that belong to the same class fall near each other. Its efficiency varies greatly based on the use of a crucial variable - the number of nearest points (K) - that is used for comparison or calculation of the value of the new data point. If the task is classification, majority voting is used for the label of the K nearest points and if the task is regression, average values of the K nearest points are calculated to find the output class or value.

Naive Bayes is a classification algorithm that works based on the assumption that the presence of a feature in a class is independent of the presence of all the other features in the same class. Equation (1) uses the Naive Bayesian to estimate the probability of a data point falling into each class and the class that has the highest probability is chosen as the outcome. *Gaussian Naive Bayes* is a variant of the classifier used when the values of its features are normally distributed.

$$P(c|x) = \frac{P(x \mid c) .P(c)}{P(x)} \tag{1}$$

where,

P(c|x)	is the posterior probability.
P(x|c)	is the likelihood.
P(c)	is the class probability.
P(x)	is the predictor prior probability.

Boosting is a class of algorithm in which individual models are added one after another, correcting the discrepancies in the final outcomes of their predecessor models, until the final model reaches the output that is closest to the desired output. By evaluating the output of every iteration, the algorithm tries to reduce the bias error by refining models that can understand the patterns in the data. *Adaptive Boosting* [28] (or simply Ada-Boost) and *Gradient Tree Boosting* [29] (or Gradient Boost) are the two different variants of Boosting algorithm, and they differ in that Ada-Boost refines its models by minimizing the exponential loss function while Gradient Boost does the same by minimizing any differential loss function.

Lastly, Discriminant Analysis is a statistical measure that uses numerical values from the one or more features of a class to group them into different categories that do not overlap with each other [30]. *Linear Discriminant Analysis* (LDA) and *Quadratic Discriminant Analysis* (QDA) differ in that while approaching the Bayes Classifier and forming a decision boundary the former assumes that all the classes have the same covariance matrix while the latter assumes that the classes in the dataset have distinct covariance matrix [31], [32]. As such, LDA is less flexible because of the limited number of parameters to calculate, and QDA is computationally expensive because of the need to calculate separate covariance matrices and this complexity increases as the number of classes and features in the dataset increases.

B. CLUSTERING ALGORITHMS

K Means is one of the simplest clustering algorithms that employs a centroid-based technique to find the K number of pre-specified clusters. It calculates the centroid of different randomly assigned K number of groups based on the initial data points and iteratively updates it as new points are added until the data points converge and the position of the center stops moving. The major advantage of K Means algorithm is that it is relatively fast and hence scalable for large datasets, but it fails when outliers are present in the dataset [33].

Mean Shift is also a centroid-based algorithm that builds upon the principle of kernel density estimation (KDE). It is non-parametric in that, unlike K Means, it doesn't require the number of clusters to be prespecified before the clustering task is initiated. Data points are iteratively shifted to the nearest peak on the KDE surface as centroid is calculated based on the mode of the clusters. As such, it can fit well on clusters with non-convex shapes. However, it has a major disadvantage that it cannot distinguish between meaningful and non-meaningful modes and hence might result in forming more clusters than necessary [34].

The Density-based spatial clustering of applications with noise, or simply DBSCAN, algorithm is like the Mean Shift algorithm in that it also works by segregating areas with high and low densities and does not require the number of clusters to be prespecified. The major advantage that DBSCAN has over Mean Shift is that it can ignore outliers unlike Mean Shift, which puts them in a cluster even if they are completely different from other data points. DBSCAN, however, suffers greatly when clusters are of varying density as this leads to the parameters - distance threshold and number of neighborhood points - not being able to represent well for all the clusters [35].

Gaussian Mixture Models are clustering algorithms that in addition to taking centroid in calculation like K Means algorithm, also uses the covariance structure of the latent Gaussians. It uses the Expectation Maximization Algorithm, which begins with randomly assigned Gaussian parameters which are repeated until the expectation and maximization steps converge. It has an added advantage over K Means that it can identify clusters correctly even if they are not circular, but it is computationally expensive as it needs more parameters to train [36].

Lastly, Agglomerative Hierarchical Clustering is an algorithm in which each data point starts with its own cluster, and they merge as we move higher up in the cluster on every succession. To merge the clusters, average linkage is compared and the ones with the smallest value for it are merged first on the bottom. This ensures that data points which are highly like each other form clusters well before dissimilar points are compared. Like Mean Shift and DBSCAN algorithms, the number of clusters do not need to be pre-specified [37].

IV. PERFORMANCE METRICS

Accuracy is one of the simplest measures of assessing the efficiency of a machine learning algorithm. However, in our case, since the number of instances of commercials and non-commercials are not similar, they can yield misleading results as even if a model blindly predicted all the instances as the occurrence of the mode class, it would still appear to be performing good. Hence, accuracy is not chosen as the performance metric in this research.

Precision is another measure of efficiency, which can be calculated with the formula given below:

$$\mathrm{Precision} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FP}) \tag{2}$$

where,

TP	= True Positives
FP	= False Positives

Similarly, recall value can be found with the formula given below:

$$\mathrm{Recall} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}) \tag{3}$$

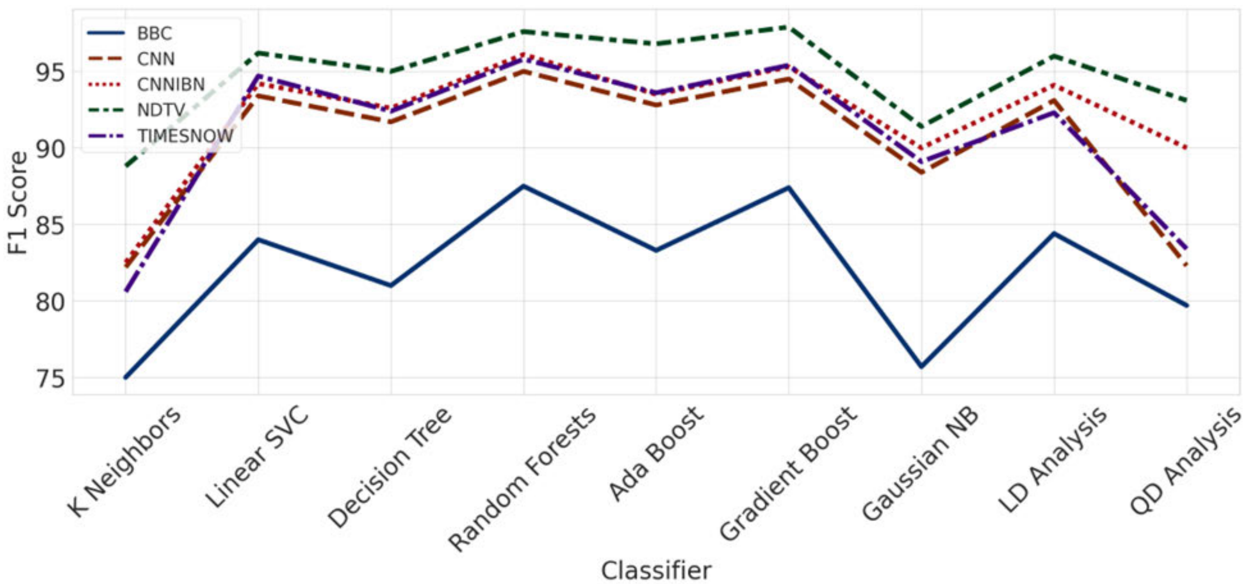

FIGURE 1. F_1 scores of the classifiers in five different channels.

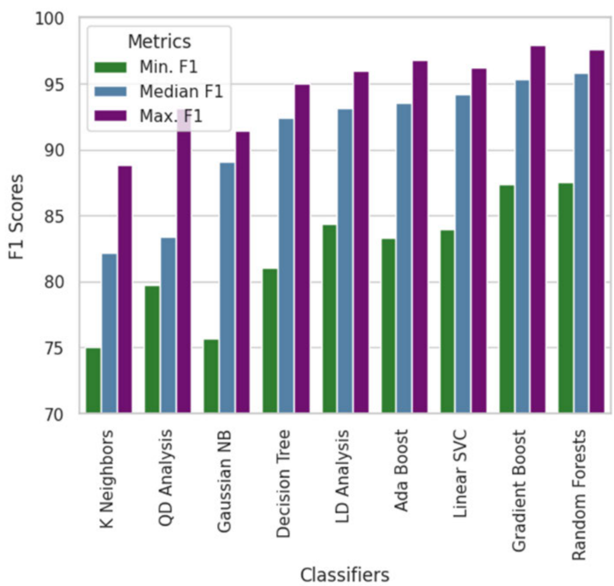

FIGURE 2. Minimum, median, and maximum F_1 scores of the classifiers.

where,

TP = True Positives
FN = False Negatives

Choosing precision over recall yields results that penalize making incorrect predictions while ignoring the correct predictions that are missed. In contrast, when recall is chosen, results are swayed by the number of predictions that are missed alone while ignoring the wrong predictions made. For a comprehensive analysis of the machine learning algorithms in identifying commercial and noncommercial data points, both metrics need to be taken into account. As such, F_1 score, a metric that takes precision and recall both into consideration, has been used as the performance metric for the first phase of this research. F_1 score can be calculated with the formula given below:

$$F_1\text{score} = 2 * \text{Precision} * \text{Recall}/(\text{Precision} + \text{Recall}) \quad (4)$$

As for the second phase of this research, Silhouette Score (S) is chosen to measure the efficiency of the clustering algorithms due to their ability to measure the distinctness of the clusters formed from the algorithms used. It can be calculated using the formula given below:

$$S = (x_2 - x_1)/\max(x_1, x_2) \quad (5)$$

where,

x_1 = mean distance between points of a cluster
x_2 = mean distance between clusters

V. EXPERIMENTAL RESULTS AND DISCUSSION

The research consisted of two distinct phases: the first phase comprised of classification with nine different classifiers and the second phase consisted of clustering with five different clustering algorithms on data from five different channels distinctly.

A. CLASSIFICATION (PHASE 1)

The data from each of the channels were split into train and test sets with ratio of 80:20. The classifiers were then trained using the data from the train set and evaluated on the test set.

The F_1 scores of all the classifiers in all channels data are displayed in Fig. 1. While the F_1 score for any classifier varied between the datasets, these classifiers performed in a very similar manner with respect to others in each of these datasets. For instance, all the classifiers performed better in the data from NDTV than they did in any other datasets. Even

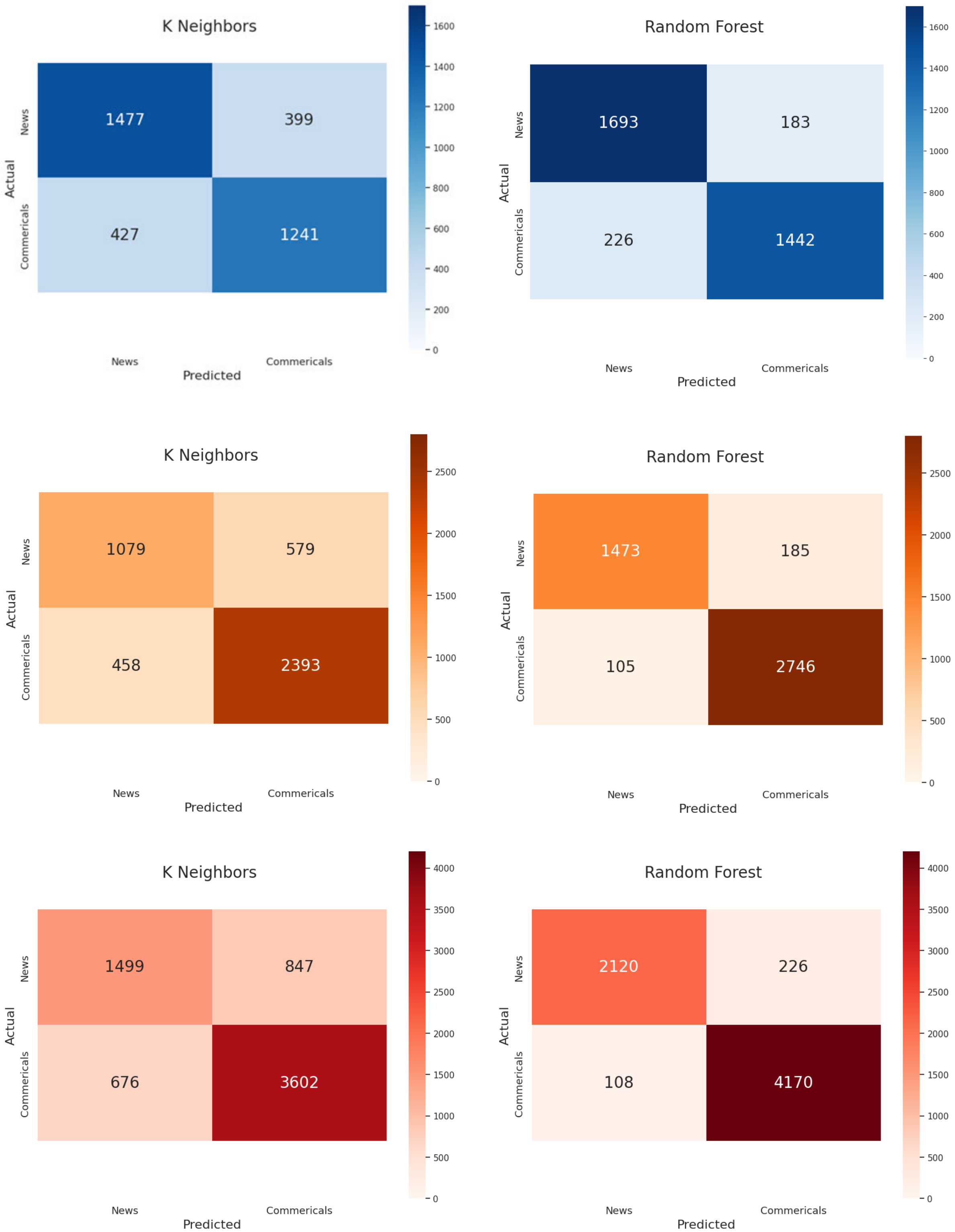

FIGURE 3. **Confusion matrices of the best and worst performing classifier in different channels in the following order: BBC (blue), CNN (orange), CNNIBN (red), NDTV (green), and TIMESNOW (purple).**

more interesting is the fact that the F_1 scores of the classifiers were considerably higher for Indian (NDTV, TIMESNOW, CNNIBN) and American (CNN) news channels than the same for the British (BBC) news channel. This can be attributed to the fact that the Indian and American news channels are very similar to one another in that they have around eight text blocks in total on top and bottom of the screen compared to only three text blocks - mostly in the bottom - in the British

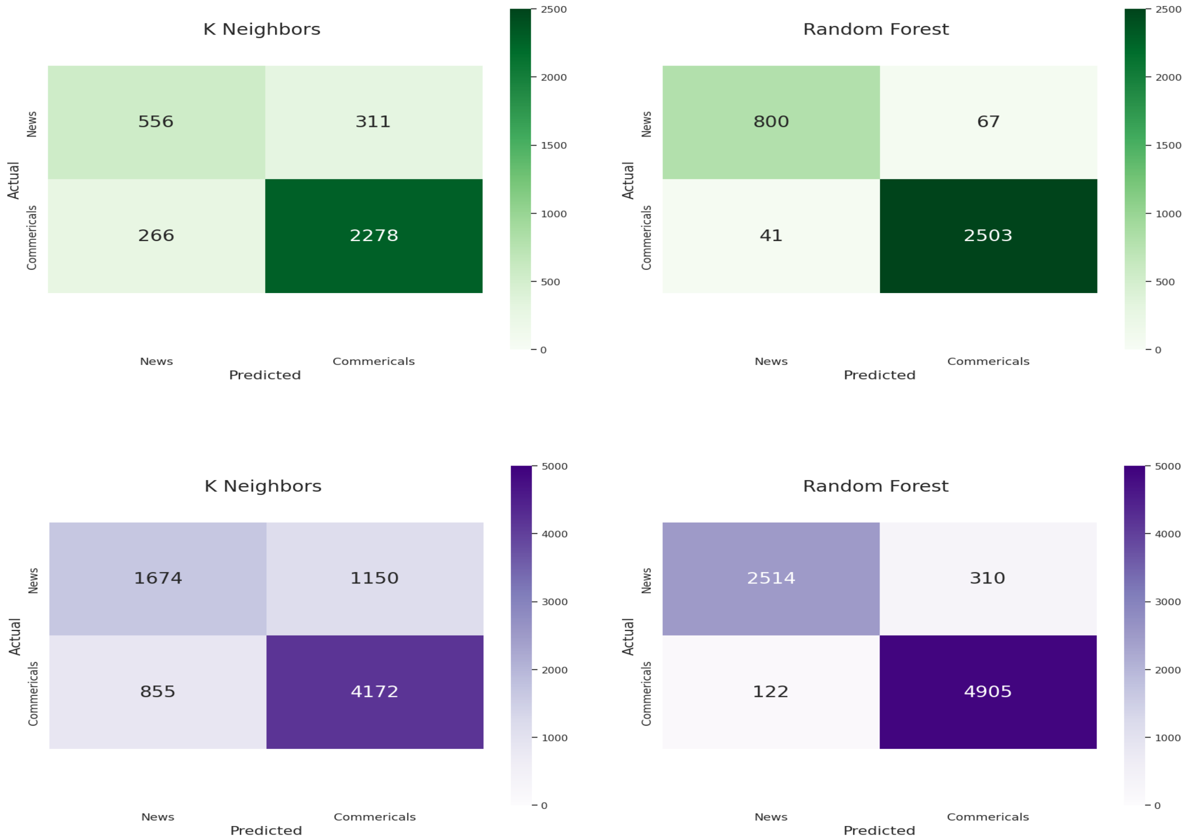

FIGURE 3. ***(Continued.)*** **Confusion matrices of the best and worst performing classifier in different channels in the following order: BBC (blue), CNN (orange), CNNIBN (red), NDTV (green), and TIMESNOW (purple).**

channel while broadcasting the news. Such a difference not only led the algorithms to better identify the context of the video but also helped in discerning inherent differences between news and commercials.

To assess the efficiency of each of classifiers, median F_1 scores were calculated. Median value was chosen rather than mean to ensure that the outlier from BBC did not affect the scores and their comparison. It was found that Random Forests had the highest median F_1 score of 95.9 followed by Gradient Boost, Linear SVC, LD Analysis, Ada Boost, Decision Tree, Gaussian NB, QD Analysis, and K Neighbors with median F_1 scores of 95.3, 94.2, 94.1, 93.5, 92.3, 89.1, 83.4, and 82.2 respectively.

The lowest, median, and highest values of F_1 scores of all the classifiers have been displayed in Fig. 2. It is evident from Fig. 2 that the Random Forests and Gradient Boost both have the maximum F_1 score of value greater than 97 and minimum F_1 score of value greater than 85. Similarly, Linear SVC, LD Analysis, Ada Boost, and Decision Tree have maximum F_1 score in between 95 and 97 and minimum F_1 score in between 80 and 85; and Gaussian NB, QD Analysis, and K Neighbors have their maximum F_1 score of value less than 95 and minimum F_1 score of value less than 80. As such, we can fairly infer that Random Forests and Gradient Boost are the top performing classifiers for the task of TV Commercial Detection.

The confusion matrices, representing the correctly and incorrectly identified classes of the best and the worst performing classifiers for each of the news channels have been shown in Fig. 3.

Looking at the F_1 scores alone to assess the efficiency of the classifier can be misleading, however. For the real-time detection of a commercial block, a classifier must be equally fast in training from the stream of incoming data. Taking this into consideration, the time taken by the classifiers to train and test on five different datasets were recorded and the minimum, median, and maximum times have been shown in Fig. 4.

As apparent in Fig. 4, Gradient Boost took the longest to train and test with a median of 131 seconds followed by Random Forests, Ada Boost, Linear SVC, Decision Tree, K Neighbors, LD Analysis, QD Analysis, Gaussian NB with the median train and test time of 31, 27, 21, 8, 4, 1, 0.7, and 0.3 seconds respectively. The time that Gradient Boost took can prove extremely expensive in the real time detection of the commercial blocks where video streams are rapidly changing. As such, Random Forests becomes the only algorithm that stands the test of time and efficiency both and hence is chosen as the most favorable machine learning model for classification of commercial. Precision, Recall, F_1 Score and Time for this phase of the research has been tabulated in Table 1.

TABLE 2. Evaluation metrics of classification algorithms.

CHANNEL	CLASSIFIER/METRIC	PRECISION	RECALL	F_1 SCORE	TIME (IN SECONDS)
BBC	K Neighbors	75.7	74.4	75.0	1
	Linear SVC	89.0	79.4	84	7
	Decision Tree	80.7	81.5	81.0	2
	Random Forests	88.7	86.5	87.5	6
	Ada Boost	90.3	77.3	83.3	7
	Gradient Boost	87.7	87.1	87.4	31
	Gaussian NB	73.7	77.8	75.7	0
	LD Analysis	89.3	80.0	84.4	0
	QD Analysis	82.1	77.4	79.7	0
CNN	K Neighbors	80.5	83.4	82.2	2
	Linear SVC	92.4	94.4	93.4	9
	Decision Tree	91.6	91.8	91.7	3
	Random Forests	93.7	96.3	95.0	11
	Ada Boost	92.5	93.1	92.8	10
	Gradient Boost	93.9	95.2	94.5	52
	Gaussian NB	88.3	88.5	88.4	0
	LD Analysis	91.7	94.6	93.1	0
	QD Analysis	71.2	97.6	82.3	0
CNNIBN	K Neighbors	81.0	84.2	82,5	5
	Linear SVC	93.5	94.8	94.2	13
	Decision Tree	93.1	92.1	92.6	5
	Random Forests	94.9	97.1	96.1	16
	Ada Boost	93.1	93.9	93.5	15
	Gradient Boost	94.4	96.2	95.3	75
	Gaussian NB	89.3	90.8	90.0	0
	LD Analysis	92.8	95.4	94.1	0
	QD Analysis	86.4	93.9	90.0	0
NDTV	K Neighbors	88.0	89.5	88.8	1
	Linear SVC	95.9	96.5	96.2	5
	Decision Tree	94.9	95.0	95.0	2
	Random Forests	96.6	98.6	97.6	7
	Ada Boost	96.1	97.6	96.8	8
	Gradient Boost	97.4	98.4	97.9	38
	Gaussian NB	94.8	88.2	91.4	0
	LD Analysis	95.4	96.6	96.0	0
	QD Analysis	90.8	95.6	93.1	0
TIMESNOW	K Neighbors	78.4	83.0	80.6	6
	Linear SVC	93.5	96.0	94.7	16
	Decision Tree	92.2	92.7	92.4	6
	Random Forests	94.1	97.6	95.8	21
	Ada Boost	92.5	94.8	93.6	18
	Gradient Boost	94.0	96.7	95.4	94
	Gaussian NB	89.1	89.2	89.1	0
	LD Analysis	92.3	96.3	92.3	0
	QD Analysis	72.9	97.5	83.4	0

B. CLUSTERING (PHASE 2)

In the beginning of this phase of the research, the data from five different news channels were first standardized by removing the mean and scaling them to their unit variance. Then, the principal components of these standardized data were obtained as shown in Fig. 5 and the first n components which explained 70% of the variability of the data were selected to train the clustering algorithms, where n turned out to be around 50 for all these data from different news channels. This task of principal component analysis and pruning not only compressed the data by reducing the number of dimensions but also ensured that noises were removed, and clustering algorithms took much less time to train.

The pruned data was then fed to the clustering algorithms. All the algorithms were trained for 30 times on each of the datasets and their average was recorded. For the algorithms that required number of clusters (n) to be prespecified, they were trained for the values of n ranging from 2 to 20 based on the findings from [1] for each of the number of iterations – 30 – to find the optimum value of n. The median value of the Silhouette scores across all the five datasets are shown in Fig. 6.

It is apparent from Fig. 6 that K Means obtained the highest median Silhouette Score of 0.71 followed by Mean Shift, DBSCAN, Gaussian Mixture, and Agglomerative Hierarchical Clustering with Silhouette Score of 0.56, 0.55,

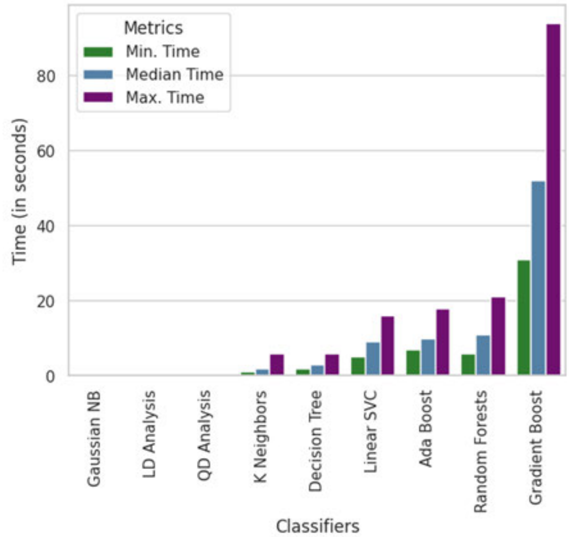

FIGURE 4. Minimum, median and maximum times taken by classifiers to train and test.

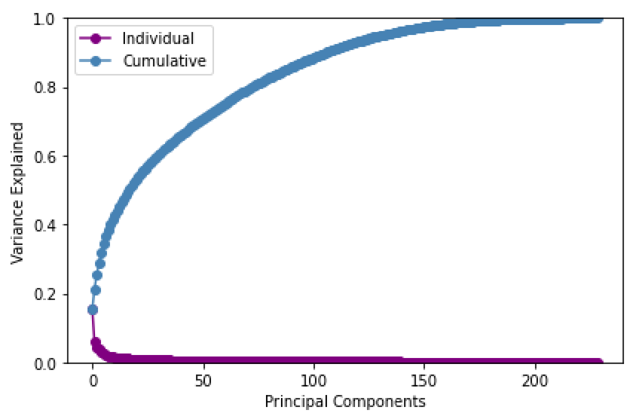

FIGURE 5. Variance explained by the principal components.

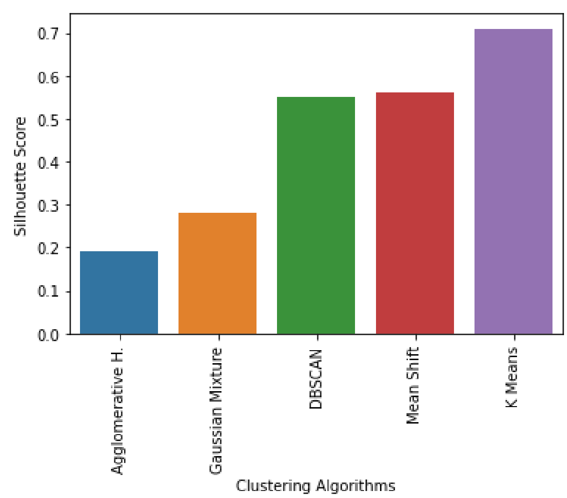

FIGURE 6. Median silhouette scores for the clustering algorithms.

0.28, and 0.19 respectively. It appears that K Means have greatly availed from the tasks of outlier removal and dimensionality reduction which prevented mean values of the clusters from being too close (something that K Means fails on) during the earlier stage of this phase of the research.

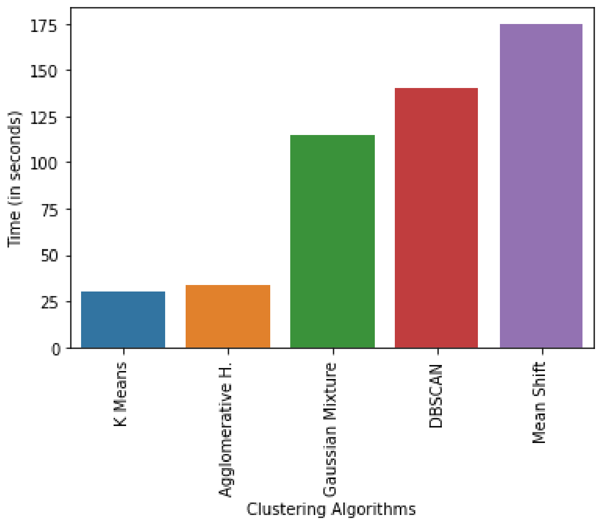

FIGURE 7. Median time taken by the clustering algorithms.

Although clustering itself has very little use in the real-time detection of commercials, no assessment of any algorithms can be complete without analyzing the time they take to perform any task. As such, the time taken by the clustering algorithms to train and form the clusters were recorded and are displayed in Fig. 7.

It is clear from Fig. 7 that K Means took the shortest amount of median time of 30 seconds followed by Agglomerative Hierarchical Clustering, Gaussian Mixture, DBSCAN and Mean Shift with median time of 34, 115, 140, 175 seconds respectively. K Means, being the simplest algorithms of all in terms of the number of computations it does, leaves all the other algorithms behind in terms of time as well.

VI. CONCLUSION

In this research, nine different machine learning classification algorithms and five different clustering algorithms were applied to the data from five different news channels from three different countries. Random Forests was found to be the most efficient classifier with median F_1 score of 95.9% and taking only 31 seconds to come with its results. Similarly, K Means Clustering proved to be the most efficient clustering algorithms, taking only 30 seconds to form clusters with median Silhouette score of 0.71.

The work, at its present stage, is far from complete. However, the authors believe that this research will provide a strong foundation for the machine-learning based generic approach for TV Commercials Detection and Clustering. In our future work, channel data from a greater number of countries will be used to train and test the efficiency of the models. In addition to news channels, channels with other different genres both from traditional linear television broadcast and online viewing platforms like YouTube and

Netflix will also be used to better generalize the performance of the models already used in this research along with deep learning models.

REFERENCES

[1] M. Covell, S. Baluja, and M. Fink, "Advertisement detection and replacement using acoustic and visual repetition," in *Proc. IEEE Workshop Multimedia Signal Process.*, Oct. 2006, pp. 461–466.

[2] P. Duygulu, M.-Y. Chen, and A. G. Hauptmann, "Comparison and combination of two novel commercial detection methods," in *Proc. Int. Conf. Multimedia Expo*, vol. 2, 2004, pp. 1267–1270.

[3] A. Ramirez, D. Cocharro, and M. Davies, "An audio-only method for advertisement detection in broadcast television content," Tech. Rep., 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1811.02411

[4] Q. Xia and G. Tang, "Research on TV advertisement detection base on video shot," in *Proc. 3rd Int. Conf. Syst. Sci., Eng. Design Manuf. Informatization*, Chengdu, China, Oct. 2012, pp. 245–248.

[5] D. A. Sadlier, S. Marlow, N. O'Connor, and N. Murphy, "Automatic TV advertisement detection from MPEG bitstream," *Pattern Recognit.*, vol. 35, no. 12, pp. 2719–2726, Dec. 2002.

[6] B. Zhang, B. Feng, P. Ding, and B. Xu, "TV commercial detection using constrained Viterbi algorithm based on time distribution," in *Proc. 9th Int. Conf. Fuzzy Syst. Knowl. Discovery*, Chongqing, China, 2012, pp. 2010–2014.

[7] Z. Feng and J. Neumann, "Real time commercial detection in videos," Tech. Rep., 2013. [Online]. Available: https://fengzheyun.github.io/downloads/projects/before2015/Comcast-2013.pdf

[8] L. Meng, Y. Cai, M. Wang, and Y. Li, "TV commercial detection based on shot change and text extraction," in *Proc. 2nd Int. Congr. Image Signal Process.*, Tianjin, China, Oct. 2009, pp. 1–5.

[9] N. Banić, "Detection of commercials in video content based on logo presence without its prior knowledge," in *Proc. 35th Int. Conv. MIPRO*, Opatija, Croatia, May 2012, pp. 1713–1718.

[10] N. Dimitrova, S. Jeannin, J. Nesvadba, T. McGee, L. Agnihotri, and G. Mekenkamp, "Real time commercial detection using MPEG features," in *Proc. Int. Conf. Inf. Process. Manage. Uncertainty Knowl.-Based Syst.*, 2002, pp. 481–486.

[11] X. Wang and Z. Guo, "A novel real-time commercial detection scheme," in *Proc. 3rd Int. Conf. Innov. Comput. Inf. Control*, Dalian, China, 2008, p. 536.

[12] X. S. Hua, L. Lu, and H.-J. Zhang, "Robust learning-based TV commercial detection," in *Proc. IEEE Int. Conf. Multimedia Expo*, Amsterdam, The Netherlands, Jul. 2005, p. 4.

[13] N. Liu, Y. Zhao, Z. Zhu, and H. Lu, "Exploiting visual-audio-textual characteristics for automatic TV commercial block detection and segmentation," *IEEE Trans. Multimedia*, vol. 13, no. 5, pp. 961–973, Oct. 2011.

[14] P. Mourya, S. Jadhav, and R. Pawar, "Real time commercials detection in television (TV) program," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 7, no. 5, pp. 198–201, 2018.

[15] W. Nazir, M. F. Hyder, M. A. Siddiqui, and M. Mukarram, "TV ad detection using the Base64 encoding technique," *Eng., Technol. Appl. Sci. Res.*, vol. 11, no. 5, pp. 7605–7609, Oct. 2021.

[16] C. Luo, Y. Peng, T. Zhu, and L. Li, "An optimization framework of video advertising: Using deep learning algorithm based on global image information," *Cluster Comput.*, vol. 22, no. S4, pp. 8939–8951, Jul. 2019.

[17] M. Hossari, S. Dev, M. Nicholson, and K. McCabe, "ADNet: A deep network for detecting adverts," in *Proc. CEUR Workshop*, 2018, pp. 45–53.

[18] A. Toheed, A. Javed, A. Irtaza, H. Dawood, H. Dawood, and A. S. Alfakeeh, "An automated framework for advertisement detection and removal from sports videos using audio-visual cues," *Frontiers Comput. Sci.*, vol. 15, no. 2, pp. 1–4, Apr. 2021.

[19] A. Vyas, R. Kannao, V. Bhargava, and P. Guha, "Commercial block detection in broadcast news videos," in *Proc. Indian Conf. Comput. Vis. Graph. Image Process.*, Dec. 2014, pp. 1–7.

[20] M. Du, J. Zhao, J. Sun, and Y. Dong, "M3W: Multistep three-way clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 29, 2022, doi: 10.1109/TNNLS.2022.3208418.

[21] S. Pei, H. Chen, F. Nie, R. Wang, and X. Li, "Centerless clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 167–181, Jan. 2023.

[22] P. Guha, R. Kannao, and R. Soni, "TV news channel commercial detection dataset, UCI machine learning repository," UCI Mach. Learn. Repository, Tech. Rep., doi: 10.24432/C5990S.

[23] L. Breiman, "Random forest," *Mach. Learn.*, vol. 45, pp. 5–32, 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[24] M. A. Songuist, "Problems in the analysis of survey data, and a proposal," *J. Amer. Stat. Assoc.*, vol. 58, no. 302, pp. 415–434, Jun. 1963.

[25] V. Vapnik and C. Cortes, "Support vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995. [Online]. Available: https://doi.org/10.1007/BF00994018

[26] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is nearest neighbour meaningful?" in *Proc. 7th Int. Conf. Database Theory (ICDT)*, vol. 7, Jerusalem, Israel. Berlin, Germany: Springer, Jan. 1999, pp. 217–235.

[27] N. S. Altman, "An introduction to kernel and nearest neighbor nonparametric regression," *Amer. Statistician*, vol. 46, no. 3, pp. 175–185, 1992. [Online]. Available: https://doi.org/10.2307/2685209

[28] Y. Freund and R. E. Schapire, "Game theory, on-line prediction and boosting," in *Proc. 9th Annu. Conf. Comput. Learn. Theory*, 1996, pp. 325–332.

[29] J. H. Friedmann, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001. [Online]. Available: https://doi.org/10.1214/aos/1013203451

[30] G. B. G. Pereira, L. P. Fernandes, J. M. R. D. S. Neto, H. D. D. M. Braz, and L. da Silva Sauer, "A comparative study of linear discriminant analysis and an artificial neural network performances in breast cancer diagnosis," in *Proc. IEEE ANDESCON*, Oct. 2020, pp. 1–6.

[31] S. Ji and J. Ye, "Generalized linear discriminant analysis: A unified framework and efficient model selection," *IEEE Trans. Neural Netw.*, vol. 19, no. 10, pp. 1768–1782, Oct. 2008.

[32] T. Omar, Z. Wassim, and B. M. Mohamed, "Brain-computer interface: Frequency domain approach using the linear and the quadratic discriminant analysis," in *Proc. 1st Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, Mar. 2014, pp. 346–349.

[33] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1, no. 14, 1967, pp. 281–297.

[34] K. Fukunaga and L. D. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Inf. Theory*, vol. 21, pp. 32–40, 1975.

[35] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining (KDD)*, Aug. 1996, pp. 226–231.

[36] R. O. Duda and P. E. Hart, "Pattern classification and scene analysis," in *A Wiley-Interscience Publication*, 1974.

[37] L. Zhang, Z. Zhu, and Y. Zhao, "Robust commercial detection system," in *Proc. IEEE Multimedia Expo Int. Conf.*, Beijing, China, Jul. 2007, pp. 587–590.

EMAN ABDELFATTAH received the M.S. degree in computer science and the Ph.D. degree in computer science and engineering from the University of Bridgeport, in 2002 and 2011, respectively. She is currently an Assistant Professor with Sacred Heart University, USA, and also an Adjunct Assistant Professor with American Intercontinental University Online. She was an Assistant Professor in computer science with the Ramapo College of New Jersey, USA. She was a professional Assistant Professor in computer science with the School of Engineering and Computing Sciences, Texas A&M University-Corpus Christi, USA. She was also an Adjunct Professor with the Department of Computer Science and Engineering and the Department of Mathematics, University of Bridgeport, USA. Her research results were published in prestigious international conferences. Her research interests include machine learning, network security, and networking.

SHREEHAR JOSHI received the Bachelor of Science degree in computer science and data science from the Ramapo College of New Jersey. Currently, he is a Software Engineer with The Boring Company. He has been actively contributing to the field of machine and deep learning by publishing several research works in renowned national and international conferences. He has been involved in several research in the application of machine learning in fields, such as human activity recognition, the IoT BotNet detection, diagnosis of medical conditions, and textual analysis.