International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

# An Efficient Optimization of Multimodal Web Page Genre Classification Based on Objects Using LR-YoloV4 and (BM)2-CWRNN Deep Learning Techniques

**Manjunath Pujar[1], Monica R Mundada[1], Sowmya B J[2], Supreeth S[3] and Shruthi G[3], Rohith S[4]**

[1]Dept. of CSE, M.S. Ramaiah Institute of Technology, Bengaluru, 560054, India
[2]Dept. of Artificial Intelligence and Data Science, M.S. Ramaiah Institute of Technology, Bengaluru, 560054, India
[3]School of Computer Science and Engineering, REVA University, Bengaluru, 560064, India
[4]Department of Electronics & Communication Engineering, Nagarjuna College of Engineering & Technology, Bengaluru, 562110, India.
* Corresponding author's Email: researchrit1985@gmail.com

*Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##*

**Abstract:** Web data mining has emerged as a convenient and crucial platform for extracting valuable data. In order to upload and download data, users prefer to use the World Wide Web. Therefore, an alternative way is offered by the web classification for supporting effective information retrieval on the Web multimedia data. In this study introduces a video analysis that involves selecting representative frames from a video sequence. Manhattan distance, also known as taxicab distance, is one of the distance metrics used in keyframe extraction. The video quality measure involves comparing the content of the video ad to a reference, such as a non-advertisement video or an ideal ad. SSIM quantifies the structural similarity between the reference and the ad in terms of luminance, contrast, and structure. To identify and categorize objects in video or image, often bounding boxes are drawn around the detected objects. The purpose of YOLOv4 is to design an object detector that operates efficiently in producing systems and can be easily trained and used. The Blue Monkey (BM) algorithm is a novel optimization metaheuristic algorithm that is inspired by the efficient performance of blue monkey swarms in nature to enhance video quality. The various machine learning classifiers were chosen for classification, named BM2-CWRNN. The extracted features from the video, the web pages are considerably categorized by the classifier as per their corresponding domain. The publicly accessible Web classification URL datasets are utilized. The results attained the proposed CWRNN are contrasted with the Brownian motion algorithms. The experimental results indicated that the classification accuracy is higher. The accuracy rates are attained via the proposed BM2-CWRNN and the web pages are effectively classified consistent with their classes.

**Keywords:** Web content classification, Structural similarity, Video quality, Key frame extraction, Brownian motion, Complex wavelet recurrent neural network

## 1. INTRODUCTION

The Web has become the largest publicly accessible provider of data in the world due to its rapid growth over the past few decades. Unlabeled, distributed, heterogeneous (mixed media), semistructural, time-varying, and large-dimensional web data are some of the attributes of web data. Web data mining refers to the procedure of extracting significant data and patterns from the arrangement of web hyperlinks, the substance present on web pages, and the usage data. This opportunity arises because of the unique traits associated with such data-mining techniques. Web mining can be classified into three types: Web structure mining, Web usage mining, and Web content mining. Web content mining plays a significant role in the retrieval of information among these categorizations. Web content mining involves the use of data mining methods on web pages to extract valuable information from various forms of content, such as text, image, video, and audio. This field is considered a sub-

*E-mail:author's email*

area of web mining and draws on the existing field of information retrieval [3, 4]. Web Content Mining is widely practiced through two essential techniques: rule-based methods and machine learning-based methods. When predicting unseen content formats, models relying on rules are incapable of generalizing effectively. Machine learning-based models outperform pre-trained formats of content by displaying strong performance. In contrast, these models have resilience to unknown news item styles, as they are capable of creating decision boundaries that effectively differentiate between different classes [5]. However, machine learning is prone to errors. Assume that the training will be done with a machine learning method and data sets that are too tiny to be inclusive. A biased training set will result in biased predictions. As a result, irrelevant data is displayed at the end of the request. To overcome this difficulty, a web content mining system is proposed in this research methodology with the aid of the ZSS clustering technique and RBF-DBN similarity matching.

## 2. LITEATURE SURVEY

S. Markkandeyan and M. Indra Devi [6] proposed a highly effective approach for classifying web pages using machine learning techniques. Web page collections were considered input or training data. The training data were pre-processed and feature extracted to get initial feature set. Unrelated and noisy features were removed using Principal Component Analysis (PCA). The selected reduced features were applied to the evaluator and search methods. The intermediate feature set was used for the application of the Meta-Web page classifier based on attribute selection (ASC). The Web classified output was derived from the final set of features. The experimental evaluation resulted that the presented method achieved 97% accuracy for the course data set.

Farman Ali et al. [7] developed a system for classifying web content that uses fuzzy ontology and Support Vector Machine (SVM) technology. The system utilizes censored Web blacklists to categorize Uniform Resource Locators (URLs) into adult and medical URLs, resulting in improved accuracy and speed. The unclear ontology concept is used to extract website types, including adult content, traditional content, and medical content, and block red image content. Clinical trials have shown that the system is effective in detecting and blocking adult content.

Daniel López-Sánchez et al. [8] presented a multimedia web mining technique applied with transfer learning and deep neural networks. The multimedia content extractor successfully retrieved all images from a specific webpage and eliminated those that lacked valuable information. The extracted images were given to the deep feature extractor which was trained by transfer learning. The web category label was produced as an output by the classifier. The classification process involved the use of feature descriptors in the classifier. Through experimentation, the proposed strategy was shown to be effective, using a variety of classification methods and features from different depths within the deep model.

Daniel López-Sánchez et al. [9], developed a web page categorization system using deep learning and metrics, focusing on visual content. Visual content from the website was extracted by the visual content extractor block. The features of the extracted content were extracted by the feature extractor block trained by transfer learning. Distance-based classification was made more suitable with the generated image-descriptors through metric learning. The final classification module assigned a class label to each descriptor and categorized the entire website. From the experimental evaluation, it was known that the presented approach outperformed all the existing methods in terms of categorization accuracy.

Mehtab Afzal et al. [10] conducted an examination of video categorization using class predictive classifiers and class-specific approaches. The first trained based on the visual features of online videos is the content-based category prediction (CNC) classifier. Testing two different video files shows that the plan is suc cessful and improves the performance of
the online video distribution.

## 3. PROPOSED SYSTEM

### *A. Problem Statement*

Existing research methodologies have some drawbacks, which are described as follows.

- Innocent usage of terms in a document's web page link community can reduce precision. Due to the loud and unrelated web pages surrounding the documents.
- Major Issues in Multimedia Web Mining encompasses various techniques such as Multimedia data retrieval, similarity search, dimensional analysis, classification, prediction, and association mining.
- The integration of mining techniques with content-based retrieval and similarity search enables effective multi-media data mining.
- Existing work has the problem of improving the accuracy if many objects are serrated and some objects are dense. This will be easy to misallocate.
- In the existing classification algorithm, the back-propagation technique results in longer training time, providing only the linear solutions and large numbers of samples are required to achieve good performance.

- In existing work, the efficiency and the time efficiency of object detection in images will be low due to the less algorithm complexity.

*B. BM$^2$-CWRNN Approach*

Web page classification is more difficult than text classification because web pages include images, video, text, hyperlinks, and much more information. Webpage classification is in the area of machine learning where learning is based on Webpages. In recent years, the number of web pages on the Internet [WORLD WIDE WEB] is increasing. The task to find Web pages which present information to satisfy requirements by hyperlink is difficult. Therefore, a search engine is required. However, an efficient and accurate method is required for the search. In this work, a method is proposed for webpage classification using Multimedia content (video) for better web page classification. The proposed work contains the following steps,

- Web Content Mining Dataset
- Data Preprocessing
  - Remove repeated URL
  - Find active link using response
- Multimedia Content Extraction
- Video Preprocessing
  - Frame Conversion
  - Key Frame Identification
    - Taxicab Distance -Fidelity
- Irrelevant Video (Ads) Identification
  - Structural Similarity Index Measure (SSIM)
- Object Detection
  - Log Radial-You Only Look Once V4 (LR-YOLOV4)
- Feature Extraction
  - Aesthetic Visual Features
  - Deep Learning Features
- Classification
  - Brownian motion-based Blue Monkey-Complex wavelet repeat neural network ((BM)$^2$–CWRNN)

The proposed system will start from the Web Classification URL data set as input. This dataset can be accessed by the public on the Internet. This dataset comprises the website's URL and its Category. Next, the data preprocessing step will be carried out. This step contains the removal of the repeated URL and the finding of the link. Next, the web content extraction phase will be carried out. In this, the web pages will be loaded using the URL from the dataset. Next, the web scraping will be done using the package in Python. Web scraping refers to the process of extracting substantial volumes of data from websites through automation. Websites contain unstructured data. Web scraping facilitates the gathering of unstructured data and the organizing of it into a structured format. After extraction of web content, it contains all types of data, such as images, videos, text, links, and audio. The proposed work concentrated on Multimedia Content that is a Video file.

Next, in the pre-processing step like Frame Conversion, the videos will be converted into frames and in the Key Frame Identification, more discriminate information from the video is obtained using the Taxicab distance-based Fidelity technique to extract video from links. In this, to improve the recognized region of action in continuous action video, the Taxicab distance will be replaced in the existing Fidelity technique. After that, irrelevant videos such as ads will be eliminated using the Structural Similarity Index Measure (SSIM). Next, the objects will be detected from relevant videos using Log Radial-You Only Look Once V4 (LR-YOLOV4). This step is mainly used to extract the objects from the background to differentiate the genre of a URL link. In this process, the input images are divided into grids. If the grid levels are large, it may result in inaccurate detection of objects in the image. In order to address this issue, the grid levels are reduced by using the function, which results in accurate prediction and reduced loss function. Therefore, the proposed method is termed LRYOLOV4.

Next, the feature extraction step will be performed. In this, the Feature (Colour, Texture, Saliency Map, and Edge BoAP) and (shaped, edges, or motion) are extracted. Finally, these extracted features and their class will be given to the classification algorithm to perform web page classification using Brownian Blue Monkey - Complex Wavelet Recurrent Neural Network ((BM)$^2$–CWRNN). This Brownian motion-based Blue Monkey algorithm will be hybrid with Complex Wavelet Recurrent Neural Network to optimize the weight parameters. Here, to improve the position updation of the Blue Monkey, the Brownian motion technique will be replaced in the existing Blue Monkey Optimization algorithm. Furthermore, in order to enhance the precision of categorization and mitigate the growing intricacy of parameters, and avoid slow convergence, the complex wavelet radial basis kernel will be used. In this phase, the proposed algorithm will be another existing deep learning technique, Deep Belief Network, with result parameters like Accuracy, Precision, and Recall. The figure 1 demonstrates the framework that has been suggested for Web Content Mining Systems.

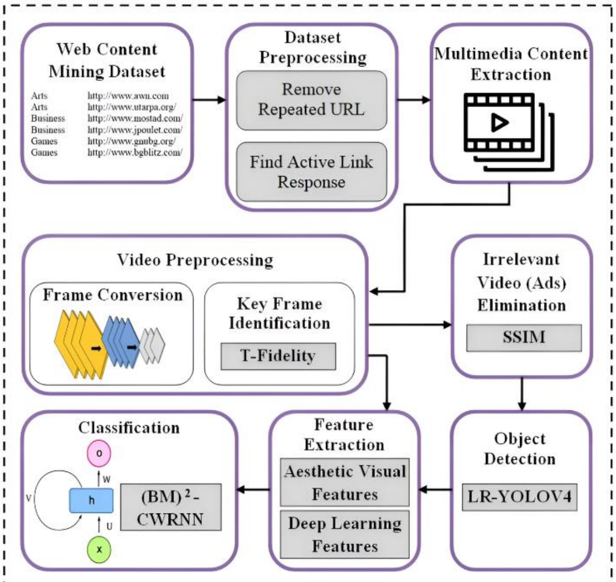


Figure. 1 Proposed Improved Web Content Mining Systems

## 4. EXPERIMENTAL SETUP

### *A. Dataset Details*

The data for this research framework are obtained from the publicly accessible Kaggle dataset on Web Classification URLs. The dataset contains the URL and category of the website. In this situation, 80% of the data is allocated for the training stage while the remaining 20% is allocated for the testing stage. The work suggested is carried out using PYTHON as the operating platform using an Intel core i5/i7 processor operating at a CPU frequency of 3.20 GHz and offering a 4GB RAM capability.

### *B. Preprocessing*

Preprocessing is a crucial stage in web content mining, since web data are typically confusing and noisy. The unstructured and undesired data form is ignored during the processing of web data by transforming the raw data into a useful and efficient shape. Thus, data pre-treatment greatly improves the quality of the data, leading to accurate data mining. In the proposed work, the preprocessing phase comparison to Convolutional Neural Networks consists of two steps: identifying the active link based on the response received and eliminating any duplicated URLs.

The website's links are referred to as active links. The user is redirected to another link when they click on the link. As a result, these active connections are accepted, while the inactive ones are not. The term "removal of repeated URLs" refers to the process of eliminating duplicate or identical links from web content. The pre-processing stage is finished before the web content extraction is carried out. In order to extract web data, the URL links from the dataset are first used to load the web sites. Ads, emoticons, data replicas, and other pertinent and unnecessary information are frequently embedded in online sites. The efficient extraction of web content is accomplished through the use of web scraping. Web scraping is the process of collecting an enormous amount of data from web sites. Most of the web data are in an unstructured format. Web scraping is used to collect and store this unstructured data in an organized fashion. The Web Content Extraction contains a vast amount of information. The extracted data are present in the form of texts, images, videos, along with audio. The AF is concentrated on the proposed work. Thus, from the multimedia content, the AF is fetched out.

#### *a) Key frames are extracted using the clustering technique*

Keyframe extraction is a video analysis process that involves selecting representative frames from a video sequence to summarize its content as depicted in figure 2. Taxicab distance, also known as Manhattan distance, is a distance measurement used in keyframe extraction. The distance between two points is determined by adding the absolute differences of their coordinates. To perform keyframe extraction using Taxicab distance-based fidelity techniques, you typically follow these steps:

- Frame Selection: Divide the video into individual frames or shots.
- Feature Extraction: Extract relevant features from each frame. These features can include color histograms, motion vectors, or any other descriptors that help to represent the content of the frame.

The taxi distance between two frames (F1 and F2) can be calculated using the following formula.

$$Taxicab\ distance = \Sigma|F_1(pix) - F_2(pix)| \quad (1)$$

For all pixels pix in the frames. Here, $|F_1(pix) - F_2(pix)|$ represents the pixel values at the corresponding positions in the two frames, and the sum is taken over all pixels in the frames. The steps for keyframe extraction using Taxicab distance-based fidelity techniques are as follows.

- Compute the taxi distance between each pair of frames in the video.
- Generate a dissimilarity matrix in which each element (i, j) depicts the taxi distance between frame i and frame j.
- The selected keyframes are those that best represent the changes in the content in the video based on the Taxicab distance metric. You can adjust the

threshold to control the granularity of keyframe selection. Keep in mind that taxi distance is just one of many metrics that can be used to extract keyframes, and the selection of the metric is determined by the particular needs of your application [11].

$$Taxicab\ distance(d_i) = \Sigma|F_i - F_{i-1}| \quad (2)$$

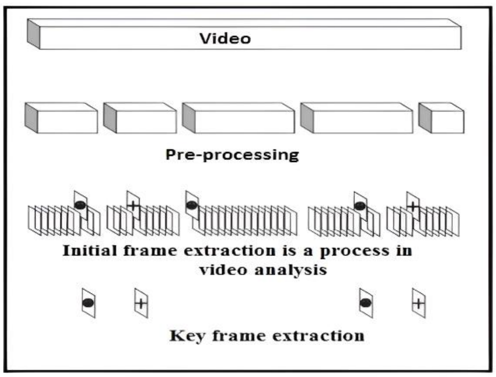


Figure. 2 Steps involved in Key frame extraction

This paper provides a detailed assessment of the video frame features and artificial intelligence approaches used for static video summarization around the world. The process of selecting representative frames from a video sequence usually involves three stages: detecting shot boundaries and the main phase of selecting key frames. Assessing the excellence of frame extraction poses a difficulty because of the subjective aspect of evaluators' opinions. In addition to conventional precision/recall values, less frequently utilized metrics are discussed.

This paper presents a comprehensive analysis of video frame characteristics and global implementation of artificial intelligence techniques for static video summarization. The process of extracting significant frames from a video sequence typically involves three stages: shot boundary detection, followed by the main phase of key frame selection. In addition to the conventional precision/recall metrics, less frequently employed estimators are also discussed.

*b) Identification of inappropriate videos (ads) using the Structural Similarity Index Measure*

Identifying irrelevant video ads using the Structural Similarity Index Measure (SSIM) measure involves comparing the content of the video ad to a reference, such as a non-advertisement video or an ideal ad. SSIM quantifies the structural similarity between the reference and the ad in terms of luminance, contrast, and structure. Identifying irrelevant video ads using the Structural Similarity Index Measure (SSIM) can be a complex task and typically involves comparing the structural features of the ad to a reference image or video frame. It is often used in image processing and computer vision to assess the quality of image compression, denoising, and other image enhancement techniques. SSIM takes into account 1. A more thorough evaluation of image quality can be achieved by analyzing luminance, contrast, and structure, rather than relying on basic methods such as Mean Squared Error (MSE). The output of this process is a numerical value ranging from -1 to 1, where a score of 1 indicates that the images are completely similar. Higher SSIM values typically indicate better image quality. The formula for the structural similarity index measure is as follows:

$$SSI = (2 * \mu x * \mu y + C1) * (2 * \sigma xy + C2)\ /(\mu x^2 + \mu y^2 + C1) * (\sigma x^2 + \sigma y^2 + C2) \quad (3)$$

Where:

- The means of the reference image and the ad image are µx and µy, respectively.
- The variances of the reference image and the ad image are denoted as σx and σy, respectively.
- σxy represents the statistical measure of the association between the reference image and the ad image. The division is stabilized using constants C1 and C2.

*C. Object Detection using Log Radial-You Only Look Once V4 (LR-YOLOV4)*

In computer vision, object detection involves recognizing and positioning objects in an image or video. It is commonly used in applications like self-driving cars, surveillance, and image recognition. Object detection algorithms typically use deep learning techniques such as Convolutional Neural Networks (CNNs), which are commonly used to identify and categorize objects within an image, typically resulting in the creation of bounding boxes around the detected objects. A few widely known frameworks for detecting objects include YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), and Faster R-CNN (Region-Based Convolutional Neural Network). The YOLOv4 aims to enhance the operational efficiency of object detection systems and optimize them for parallel computing. The main goal of YOLOv4 is to ensure easy training and usage of the intended object detector.

Figure 3 demonstrates the process of YOLO in transforming the given image into an image where a bounding box is assigned. Initially, the input photo is resized, followed by the execution of a convolutional neural network in the second step. Then, non-max suppression is performed, resulting in the generation of an image that is marked with a bounding box.

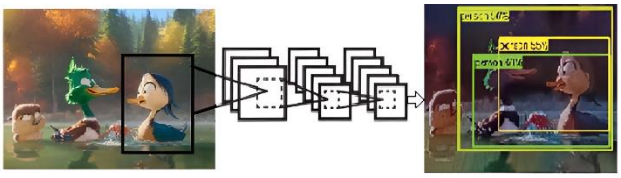

Figure 3: Process of YOLO in transforming the given image

The Confusion Matrix is a useful tool for evaluating the performance of machine learning classification problems, as it provides a measurement of the output results in the form of multiple classes. This matrix is a reliable technique to predict evaluation, as it shows and compares the actual and predicted values in the model. The F1 score, accuracy, precision and recall [12] are prediction models used to obtain the outcomes of the evaluation matrix. Figure 4 depicts the representation of this confusion matrix.

| | | Actual Values | |
|---|---|---|---|
| | | 1 (*Positive*) | 0 (*Negative*) |
| Predicted Values | 1 (*Positive*) | TP (*True Positive*) | FP (*False Negative*) |
| | 0 (*Negative*) | FN (*False Negative*) | TN (*True Negative*) |

Figure 4: Component of Confusion Matrix Display

The amount of data that are positive and precisely forecasted as positive is referred to as true positive. False Positive refers to data that are negative but are incorrectly forecasted as positive. The amount of data that is positive but incorrectly forecasted as negative is referred to as False Negative. The amount of data that are both negative and accurately expected to be negative.

**Accuracy:** Accuracy refers to the proportion of relevant objects correctly identified among all selected objects. Precision, on the other hand, can be described as the comparison between correctly distinguished objects and the total number of existing items, while the error rate represents objects that were mistakenly recognized out of the total existing objects [20].

$$Accuracy = \frac{TP+TN}{TP+TN+FF+FN} \times 100\% \quad (4)$$

**Precision:** The precision [21] of the model or system is determined by evaluating its ability to provide the precise information desired for the prediction outcomes.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

**Recall:** Recall is defined as the ratio of accurately recognized items or true positives in relation to all positive information. A comprehensive analysis indicates that the system or model developed can categorize various kinds of object [22] in an effective and precise manner.

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

**F1 Score:** Precision, recall and the F1 score [23] provide a visual representation of the relative impact.

$$F1\ Score = 2 \times \frac{Recall \times Presis}{Recall+Presis} \times 100\% \quad (7)$$

*D. Video feature extraction using the combination of Aesthetic and Deep Learning features*

It has been investigated that the evaluation of video aesthetics heavily relies on the color combination properties and motion of the video, which can be analyzed using deep learning techniques. However, extracting these features is a subjective task as it varies based on individual psychology and emotions while viewing, comprehending, and evaluating videos. The use of deep learning can significantly enhance the accuracy of feature extraction, thus improving the assessment of video aesthetics [13].

Here is a step-by-step process to extract these features:

- ➢ Aesthetic Features: Use image processing techniques or models trained for aesthetic assessment to extract features like color harmony or balance.
- ➢ Visual Features: Compute color histograms, texture features using methods such as Gabor filters, and shape descriptors (e.g., using contour analysis).
- ➢ Deep Learning Features: Load a pre-trained CNN model and remove the classification head. Images are passed through the network to extract features from intermediate layers or the final layer.

These features can be used as high-level representations of images.

- ➢ Combine Features: You can concatenate or combine the extracted features from the previous steps into a feature vector.
- ➢ Application: Use these combined features for different computer vision purposes such as classifying images, detecting objects, or retrieving images based on their content.

Remember that the choice of features and methods may vary depending on the specific task you are working on, and you may need to fine-tune or train models for certain applications. Additionally, dimensionality reduction techniques such as PCA or t-SNE can be applied to reduce feature space, if needed.

### *E. Brownian motion and process of diffusion*

At time t, the particle's position x and its position y, starting at position x, must be determined to determine the particle's transition density p(x, y, t). The relationship between Brownian motion and the diffusion process was established by the diffusion equation and Einstein [14].

The function p (0, y, 0) is equal to d (y) in our scenario. Additionally, consider the scenario where the particle transitions from y to y+h during a time interval q. This event is characterized by the transition density r(q, h). Consequently, we have the following relationship:

$$p(0,y,t+\theta) = \int_{-\infty}^{\infty} p(0,y-h,t)\rho(\theta,h)dh = \int_{-\infty}^{\infty} \left(p - p_y h + \frac{1}{2}p_{yy}h^2 + \cdots\right)\rho dh \quad (8)$$

Given that:

- r is a transition density $\int_{-\infty}^{\infty} \mathrm{rdh} = 1$
- r(h,q) = r (q, h) by symmetry, so $\int_{-\infty}^{\infty} \mathrm{hrdh} = 0$ and
- the variance of r, $\int_{-\infty}^{\infty} h^2 \mathrm{rdh} = \mathrm{q}$

These assumptions lead to the following:

$$\frac{p(0,y,t+\theta)-p(0,y,t)}{\theta} = \frac{p_{yy}(0,y,t)}{2} + \text{ higher-order even moments} \quad (9)$$

The following analytical solution in diffusion equation is

$$p(0,y,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} \quad (10)$$

At time t, when the Brownian motion starts at x = 0, it conforms to a normal distribution N (0, t). More specifically, this motion is known as a natural Brownian motion. The stochastic process *X(t)* possesses the following characteristics:

- ➢ $X(0) = x$ almost surely
- ➢ $X(t)-X(s) \sim N(0, t-s), \forall\ 0 \leqslant s \leqslant t,$
- ➢ $\forall 0 \leqslant t1 \leqslant t2 \ldots \leqslant tn$ the random variables $X(t1), X(t2)-X(t1) \ldots$ are independent.

In this study, the Brownian motion method was utilized to calculate the transition densities of the positions using solutions from equivalent diffusion equations. This mathematical theory is extensive. Kolmogorov's equations govern the position y of a particle undergoing a Brownian motion, which occurs within the interval [1, 1] and starts at position x at t = 0. This natural phenomenon comes to a halt when the particle reaches either boundary of the interval.

$$\begin{cases} \partial_t p(t,x,y) = \frac{1}{2}\Delta_x p(t,x,y) \text{ over } \mathbb{R}_+^* \times (-1,1), \\ p(t,x,y) \longrightarrow \delta_y(x), t \to 0, \\ \partial_t p(t,x,y) = \frac{1}{2}\Delta_y p(t,x,y) \text{ over } \mathbb{R}_+^* \times (-1,1), \\ p(t,x,y) \longrightarrow \delta_x(y), t \to 0, \\ p(t,x,y) = 0 \text{ if } |x| = 1 \text{ or } |y| = 1 \end{cases} \quad (11)$$

A stochastic process can be used to establish two random variables: the exit time (t), which is the time when *Xt* reaches the limit, and the exit position (*Xt*).

$$t = \inf\{t, X_t\ 6 \notin [-1,1]\} \quad (12)$$

$$X_t \in \{-1, 1\}$$

When Dirichlet conditions are applied to the segment boundaries, the transition density p can be deduced analytically for the particle's position.

$$p(x,y,t) = \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} \left( \exp\left(-\frac{(x-y-4n)^2}{2t}\right) - \exp\left(-\frac{(x-y-2-4n)^2}{2t}\right) \right) \quad (13)$$

The Brownian motion involving reflection at -1, which operates within the range of -1 to 1, presents a distinct scenario where the particle does not terminate at 1. Instead, it rebounds within the segment and finally stops at -1. In the previous system of PDEs, this is essentially the same as having a Neumann condition at 1 and reflection is defined mathematically using the local-time notion. The reflected and drifted cases can provide the analytical equations for *p(x, y, t)*, as indicated in [14]. In higher dimensions, it is important to mention that the link between PDEs and Brownian motion remains valid. In this context, the interval [-1, 1] is replaced by a domain $D \subset Rd$.

$$\begin{cases} \frac{1}{2}\Delta u + g(x) = 0, x \in D \\ u(x) = \psi(x), x \in \partial D \end{cases} \quad (14)$$

As a result, it is possible to write the expression of *u(x)* using the Feymann-Kac relationship based on the Brownian motion.

$$u(x) = \mathbb{E}_x\left[\psi(X_\tau) + \int_0^\tau g(X_\tau)ds\right] \quad (15)$$

*Xt* is a Brownian motion that terminates when it reaches the boundary *D*. The exit time, denoted as *∂D*, *t=inf {t, Xt 6∉D}*, signifies when this event occurs, while the exit position is represented by *Xt*. If the boundary *D* contains Neumann conditions, it is necessary to reflect *Xt* across this particular section of the boundary. With the inclusion of an extra transportation factor $\mu\Delta u$ in the Poisson equation, *Xt* transforms into a drifted Brownian motion.

## 5. RESULTS AND DISCUSSIONS

In this research, the Python programming language was used to pre-process, build, and build machine learning models, as well as extract webpage data from URL information. Python utilities such as Urllib and

Selenium were utilized to extract content from URLs and create web pages. To speed up and improve the efficiency of mathematical processes, the texts on the webpages and in the Numpy library were preprocessed.

*A. Image and Video Aesthetics display*

Aesthetics is the study that concerns the assessment of beauty and its connection to the mind and emotions [15]. A system to evaluate the aesthetics quality of videotapes utilizes a combination of visual elements based on photographic and cinematographic principles, along with a literacy system that considers the representation question of the videotape. More specifically, the data are obtained by analyzing the arrangement of videotapes from various positions, including both low and high viewpoints. This analysis helps to develop a methodology that combines printed and moving elements to create a bracket for the videotape. For each shot on every videotape, essential frames are assigned specific names.

The analysis of video animation or video animation utilizing deep learning indicates that the analysis is heavily dependent on the movement and blending of colors in the video. Extracting the two features of the video is a very significant responsibility as people exhibit various psychological aptitudes and emotions during their viewing experience, understand and analyze the video [16]. Deep learning offers great opportunities for improved video analysis by increasing the precision of feature extraction. The video display is processed by the hidden layer of the neural network during this procedure [17]. There are several deep neural network models, such as the Recursive Neural Network (RNN), Deep Convolutional Neural Network (DCNN), and Recurrent Neural Network (RNN) are among the neural networks utilized in this context. The process is described below:

1) In order to extract the video into different shots, frames, and keyframes (as shown in Figure 2), preprocessing techniques are crucial. Using deep convolution neural networks, deep learning can automatically extract keyframes [14].
2) Detecting salient features is a crucial aspect of video aesthetics, as it allows for the synthesis of information. Detection of object motion and color contrast within the video.
3) Feature extraction methods are used to enhance the quality of images and videos by removing unnecessary data. Defining different levels of features. Feature extraction is classified according to three levels: low level, middle level, and high level.

*B. Performance analysis of the proposed classification technique*

The proposed BM-CRWNN the Brownian motion and continuous wavelet analysis into recurrent neuron networks, this combination is proposed to capture and temporal dependencies in data. Validation includes various performance metrics such as specificity, accuracy, precision, sensitivity, and the F-measure. False Positive (FP), True Positive (TP), simultaneously with methodologies, such as deep convolutional neural networks (DCNN), a combination of convolutional neural network (CNN), support vector machine (SVM) and Long Short-Term Memory (LSTM) is utilized. To assess the efficiency of the model, the evaluation is carried out by comparing it with existing methodologies.

Table 1: Performance of the suggested BM$^2$-CRWNN in terms of accuracy, sensitivity, and specificity

| **Techniques** | **Performance metrics (%)** | | |
|---|---|---|---|
| | **Accuracy** | **Sensitivity** | **Specificity** |
| **Proposed BM$^2$-CWRNN** | 95.05 | 95.14 | 96.16 |
| LSTM | 94.15 | 95.14 | 95.84 |
| SVM | 91.64 | 92.12 | 93.52 |
| CNN | 89.12 | 90.11 | 91.76 |
| DCNN | 86.15 | 87.25 | 89.12 |

The results of the performance evaluation of the BM-CRWNN are displayed in Table 1 in comparison to other widely used methodologies such as DCNN, CNN, SVM, and LSTM. The table highlights the performance of each method in terms of precision, sensitivity, and specificity. According to the tabulation, BM-CRWNN achieves 95.05% accuracy, 95.14% sensitivity, and 96.16% specificity. However, the average accuracy, average sensitivity, and average specificity achieved by current methods (DCNN, CNN, SVM, and LSTM) are 92.50%, 93.16%, and 94.16%, respectively. These findings indicate that the BM-CRWNN outperforms the other methods in terms of performance metrics. Therefore, the suggested approach effectively detects the different categories of websites by examining video characteristics and also improves the extraction of web content [18].

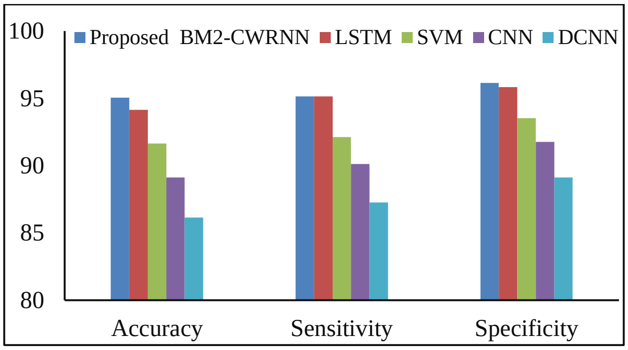


Figure 5: Based on accuracy, sensitivity, and specificity, a graphical representation of the proposed BM-CRWNN is presented.

Based on data sets in Table 2, the classification models developed by DCNN, CNN, SVM, LSTM, and BM2-CRWNN exhibit superior classification performance compared to the existing models. The dataset contains URL information related to inactive web pages, however, by utilizing embedded words and deep learning models, more improved classification models have been achieved. The URL-based approach is used by the BM2-CRWNN model, while the models introduced in this research employ a content-based approach. The proposed model extracts features from URLs. However, the primary objective of this study is to achieve more accurate classification performance.

Based on the graphical analysis presented in Figure 5, it is evident that the proposed methodology continues existing work in terms of metrics. The BM$^2$-CRWNN accurately categorizes websites according to their domains or classes, making it easier to extract essential content from large data sets.

Table 2: Evaluation of the effectiveness of the BM$^2$-CWRNN model through precision, recall, and F-measure measurements

| Techniques | Performance metrics (%) | | |
|---|---|---|---|
| | **Precision** | **Recall** | **F-measure** |
| **Proposed BM$^2$-CWRNN** | 95.75 | 95.64 | 95.25 |
| LSTM | 94.95 | 95.14 | 94.99 |
| SVM | 91.21 | 92.25 | 93.71 |
| CNN | 88.10 | 90.11 | 92.18 |
| DCNN | 86.78 | 87.25 | 89.66 |

Table 2 illustrates the results of the precision, recall and F measurements of BM2-CWRNN. The model's precision, recall, and F-measure are the factors that determine its value, which are reported to be 95.75%, 95.64%, and 95.25%, respectively. In contrast, the existing work has precision rates ranging from 86.78% to 94.95%, recall rates ranging from 87.25% to 95.14%, and F measure rates ranging from 89.66% to 94.99%. These rates are lower than those of the proposed methodology. Therefore, BM2-CWRNN simplifies the complexities and improves the robust data mining from the Web. The metrics rates of the proposed methodology depicted in Figure 6 are higher compared to the rates of existing work.

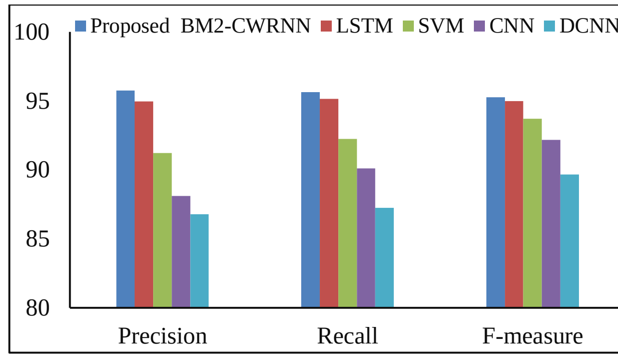


Figure 6: A graphical representation of the proposed BM$^2$-CRWNN based on precision, recall, and F measurement

Table 3: An evaluation of the performance of BM$^2$-CWRNN is conducted using FPR, FNR, and MCC metrics

| Techniques | Performance metrics (%) | | |
|---|---|---|---|
| | **FPR** | **FNR** | **MCC** |
| **Proposed BM$^2$-CWRNN** | 0.14 | 0.27 | 0.74 |
| LSTM | 0.14 | 0.66 | 0.74 |
| SVM | 0.43 | 0.91 | 0.85 |
| CNN | 0.40 | 0.68 | 0.82 |
| DCNN | 0.38 | 0.51 | 0.80 |

Table 3 compares the performance of the proposed method to that of existing methods such as DCNN, CNN, SVM, and LSTM, in terms of metrics such as accuracy, sensitivity, specificity, precision, and others. The findings indicate that the proposed BM2-CWRNN method is notably superior in categorizing online content based on captured video frames. The BM2-CWRNN model proposed in this case demonstrates higher accuracy and reduces the error to a minimum. The suggested BM2-CWRNN model achieves an MCC of 0.74, whereas the DCNN, CNN, SVM, and LSTM models achieve MCC values of 0.80, 0.82, 0.85, and 0.74 respectively. Additionally, the proposed model exhibits an FPR of 87.7% and an FNR of 0.043. According to our research, the suggested approach outperforms conventional models in terms of accuracy, f-measure, false positive rate (FPR), and various other metrics. The performance of the proposed BM2-CWRNN model clearly demonstrates its superiority over existing methods.

Figure 7 shows the proposed model, as well as well-established models such as FNR, FPR, and MCC, along with their respective metrics. This serves to highlight the high reliability of the suggested BM2-CWRNN model in classifying web content from extracted video images. The proposed model effectively reduces errors and achieves fast convergence, thus demonstrating the efficiency of the optimization strategy in minimizing errors according to the defined objective.

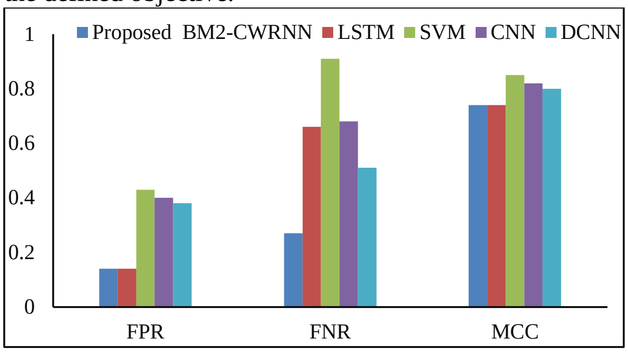


Figure 7: A comparison of the proposed BM2-CWRNN with FPR, FNR, and MCC

Table 4 presents the f1 scores obtained from multiclass classification models, which were trained using the video content found on the Web pages. All models utilized the extracted images as input data, and hyperparameter optimization was performed. The test data provided the f1 scores. Superior classification performance is observed in the BM$^2$-CWRNN model compared to existing models, based on F1 scores.

Table 4: Multiclass classification model results

| Models | F1 Score |
|---|---|
| Proposed BM2-CWRNN | 95.25 |
| LSTM | 94.99 |
| SVM | 93.71 |
| CNN | 92.18 |
| DCNN | 89.66 |

Based on an evaluation of their models' training time and a comparison of their respective f1 scores and training times, it can be concluded that the BM$^2$-CWRNN model surpasses the existing models.

## 6. Conclusions

The Blue Monkey algorithm is a novel approach for extracting key frames from web content in videos. This method, based on web content, uses various techniques such as DCNN, CNN, SVM, and LSTM classifications. The evaluation process includes tasks like data preprocessing, frequency spectrum evaluation, multimedia content extraction, feature extraction, and classification. The proposed approach achieved a precision rate of 95.05%, a sensitivity rate of 95.14%, and a specificity rate of 96.16%. The BM$^2$-CWRNN achieved higher classification accuracy, ensuring web pages are classified consistently. Further experiments are needed to assess the algorithm's application level and explore more video enhancement algorithms.

## Conflicts of Interest

No author has disclosed any conflicts of interest.

## Author Contributions

**Mr. Manjunath Pujar** has identified Initial problem identification, algorithm write-up, analysis, drafting of the manuscript, and simulation. **Dr. Monica R Mundada and Dr. Sowmya B J** were responsible for the Literature survey and helped in the initial review process. **Dr. Rohith S** was responsible for the Complexity analysis of the research, evaluation of the research work. **Dr. Supreeth S** and **Dr. Shruthi G,** responsible for the final formatting and applied for the journal. All authors worked together to implement and evaluate the integrated system, and approved the final version of the paper.

## Acknowledgment

Authors acknowledge the support from MSRIT and REVA University, Bengaluru- 560064 for the facilities provided to carry out the research.

## References

[1] M. O. Samuel, A. I. Tolulope, and O. O. Oyejoke, “A Systematic Review of Current Trends in Web Content Mining,” Journal of Physics: Conference Series, vol. 1299, no. 1. IOP Publishing, p. 012040, Aug. 01, 2019. doi: 10.1088/1742-6596/1299/1/012040.

[2] Neha Doomra and RashmiVerma, “Web mining techniques”, IJSRST, vol. 6, no. 3, 2019.

[3] Schenker, A. Kandel, H. Bunke, and M. Last, “Graph-Theoretic Techniques for Web Content Mining,” Series in Machine Perception and Artificial Intelligence. WORLD SCIENTIFIC, May 2005. doi: 10.1142/5832.

[4] P. M. Bharti and T. J. Raval, “Improving Web Page Access Prediction using Web Usage Mining and Web Content Mining,” 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, Jun. 2019. doi: 10.1109/iceca.2019.8821950.

[5] Y. Kim and S. Lee, “SVM-based web content mining with leaf classification unit from DOM-tree,” 2017 9th International Conference on Knowledge and Smart Technology (KST). IEEE, Feb. 2017. doi: 10.1109/kst.2017.7886134.

[6] S. Markkandeyan and M. Indra Devi, “Efficient Machine Learning Technique for Web Page Classification,” Arabian Journal for Science and Engineering, vol. 40, no. 12. Springer Science and Business Media LLC, pp. 3555–3566, Sep. 08, 2015. doi: 10.1007/s13369-015-1844-1.

[7] F. Ali et al., “A Fuzzy Ontology and SVM–Based Web Content Classification System,” IEEE Access, vol. 5. Institute of Electrical and Electronics Engineers (IEEE), pp. 25781–25797, 2017. doi: 10.1109/access.2017.2768564.
[8] D. López-Sánchez, A. G. Arrieta, and J. M. Corchado, “Deep neural networks and transfer learning applied to multimedia web mining,” Distributed Computing and Artificial Intelligence, 14th International Conference. Springer International Publishing, pp. 124–131, Jun. 21, 2017. doi: 10.1007/978-3-319-62410-5_15.
[9] D. López-Sánchez, A. G. Arrieta, and J. M. Corchado, “Visual content-based web page categorization with deep transfer learning and metric learning,” Neurocomputing, vol. 338. Elsevier BV, pp. 418–431, Apr. 2019. doi: 10.1016/j.neucom.2018.08.086.
[10] M. Afzal, X. Wu, H. Chen, Y.-G. Jiang, and Q. Peng, “Web video categorization using category-predictive classifiers and category-specific concept classifiers,” Neurocomputing, vol. 214. Elsevier BV, pp. 175–190, Nov. 2016. doi: 10.1016/j.neucom.2016.06.004.
[11] Bodyanskiy, Y., Kinoshenko, D., Mashtalir, S., & Mikhnova, “On-line video segmentation using methods of fault detection in multidimensional time sequences”, International Journal of Electronic Commerce Studies, 3(1), pp. 1-20.
[12] S. Mashtalir and O. Mikhnova, “Key Frame Extraction from Video,” International Journal of Computer Vision and Image Processing, vol. 4, no. 2. IGI Global, pp. 68–79, Apr. 01, 2014. doi: 10.4018/ijcvip.2014040105.
[13] Phatak, M. V. “Survey: Deep Learning for Video Aesthetics.” International Journal for Research in Applied Science and Engineering Technology 7.0 (2019):138.0–149.0. doi: 10.22214/ijraset.2019.2021.
[14] R. Gómez González and V. Garzó, “Mobility and diffusion of intruders in granular suspensions: Einstein relation,” Journal of Statistical Mechanics: Theory and Experiment, vol. 2024, no. 2. IOP Publishing, p. 023211, Feb. 26, 2024. doi: 10.1088/1742-5468/ad267b.
[15] D. Joshi et al., “Aesthetics and Emotions in Images,” IEEE Signal Processing Magazine, vol. 28, no. 5. Institute of Electrical and Electronics Engineers (IEEE), pp. 94–115, Sep. 2011. doi: 10.1109/msp.2011.941851.
[16] J.-H. Choi and J.-S. Lee, “Automated Video Editing for Aesthetic Quality Improvement,” Proceedings of the 23rd ACM international conference on Multimedia. ACM, Oct. 13, 2015. doi: 10.1145/2733373.2806386.
[17] M. Wang, Y. Ming, Q. Liu, and J. Yin, “Image-Based Video Retrieval Using Deep Feature,” 2017 IEEE International Conference on Smart Computing (SMARTCOMP). IEEE, May 2017. doi: 10.1109/smartcomp.2017.7947017.
[18] M. Pujar, M. R. Mundada, B. J. Sowmya, S. Supreeth, and G. Shruthi, “An Efficient Framework for Web Content Mining Systems Using Improved CD-PAM Clustering and the A-CNN Technique,” SN Computer Science, vol. 4, no. 5. Springer Science and Business Media LLC, Sep. 09, 2023. doi: 10.1007/s42979-023-02137-w.