Received: October 10, 2022. Revised: November 18, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Web Page Recommendation System Based on Text and Image Pattern Extraction Classification Model

Chaithra[1,2]* **Lingaraju Gowdru Malleshappa[3]** **Jagannatha Sreenivasaiah[4]**

[1] *School of Computer Science & Engineering, Reva University, Bengaluru, India*
[2] *M S Ramaiah Institute of Technology, VTU, Bengaluru, India*
[3] *Department of Information Science and Engineering, East Point College of Engineering & Technology, Bengaluru, India*
[4] *Department of Computer Applications, M S Ramaiah Institute of Technology, MSRIT Nagar, Bengaluru, India*
* Corresponding author's Email: chaithra81@gmail.com

Abstract: Classification and analysis are improved factors for the real-time automation system. We need to analyze the attribute level in the area to predict the type of web page that can be predicted. This paper proposed a novel feature prediction model and the classification algorithm to estimate the attribute level of data and its probabilistic image to analyze the type of content that can be predicted on the page. This is to classify the different types of web pages which is better to perform the web page recommendation. For this process, dynamic lexical magnitude pattern (DLMP) system and correlated boltzmann machines (CBM) based classification model are used to analyze the attribute and image of the page area. The dataset consists of collections of attributes and images at various data samples for a page. In the DLMP-PGDT-based feature analysis method; the extract of the attribute and image in various texture patterns are analyzed and framed as the pattern for the given dataset. Then from that, an improved neural network architecture based on the block probability analysis is used to classify the data pattern to predict the class of web page according to the features of the dataset. This classification model assists were to recommend the content. The result analysis presents the comparison result of the proposed work with the web page classification technique like TrAdaBoost algorithm where proposed technique achieved accuracy of about 98%.

Keywords: Web page classification, Dynamic lexical magnitude pattern (DLMP), Phonetic features, Pixel granulometric differential texture (PGDT), Correlated boltzmann machines (CBM).

1. Introduction

In the web content field, the prediction of text with respect to the image of the page guides the selection of content to predict. The bitwise-based algorithm for recommending the web page is successfully done in the work [1]. On the web, page recommendations can be done with respect to text or Images. The recommendations are done using an image by applying an Enhanced hybrid semantic algorithm [2].

The personalized recommendation system increases the transaction opportunities in e-commerce using the big data analytics [3]. To enhance the prediction model and the feature analysis of texture pattern from the input properties of the image and the text data samples can be predicted by optimization algorithm to select the best attributes. In the web page data classification process, the image and the text data samples are used for estimating the type that can predict that content and to predict its growth level of it. This form of training part increases the time complexity and the space complexity of the classification model. This can be overcome by optimal feature selection with the texture pattern-based data classification approach. A [4] Web page recommendation system provides the services to the users with the required information. The users can

Received: October 10, 2022. Revised: November 18, 2022.

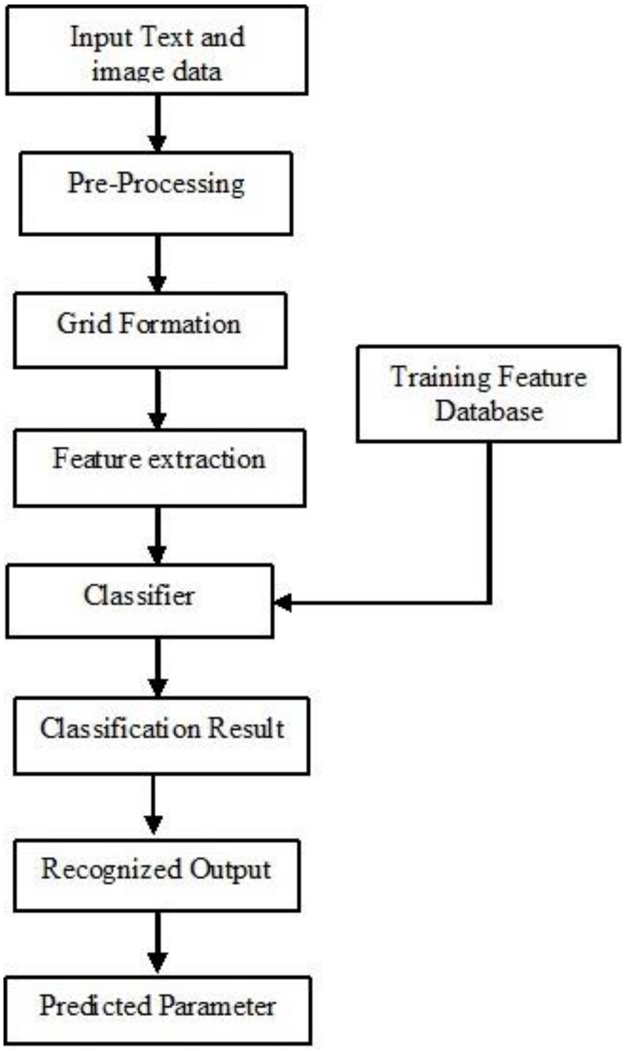


Figure. 1 Basic architecture diagram of data analysis

obtain the information easily by using the web page recommendation system. A hybrid semantic algorithm used for the web image retrieval based on the ontology classification is proposed supporting to the query of the user.

By considering that, the following objectives are identified to be discussed are as follows:

1 To implement a novel dynamic lexical magnitude pattern (DLMP) based texture pattern extraction algorithm for representing the feature of input data samples.
2 To enhance the classification performances by arranging the optimal feature attributes that are relevant to the texture data.
3 To validate the text and image kind of feature attributes for predicting and forecasting the range of web page cultivation based on the probabilistic condition.
4 To implement the texture classification method by using correlated boltzmann machines (CBM) classifier.
5 To validate the performance of the proposed classification model based on the statistical parameters and the comparison result.

The overall description of the paperwork with algorithm statements and the comparative study are can be organized in the following subsections. In that, the related works for the texture classification are surveyed and reviewed in section 2. Section 3 explains the algorithm description and the steps involved in the proposed DLMP-PGDT techniques for web page data prediction. Section 4 validates the performance of the proposed work by estimating the statistical parameters and presented the comparison of classification results with the other traditional classification models. The justification and the future work were concluded in section 5.

2. Literature review

A detailed review of existing methods in web page prediction and data extraction is presented in this section. These are all mainly focused on the process of feature extraction, optimization, and the classification of data samples that are followed in the flow of work.

By considering that, [5] presented by compost method to improve the feature prediction by mitigating the error rate deficit stress for the web page. This sub-divides the data texture and predicts the error rate with respect to it. The web page recommendation has been done using extended techniques of collaborative filtering by using case-based reasoning. The error level of the proposed method found to be with the minimum miss and fallout rates. The F1 measure can still be improved as future work. The web page recommendation is tested using a movie lens and a jester dataset where collaborative filtering concept is used in the recommendation [6]. In future recommendation can be enhanced without compromise of the privacy and its information leakage issues. The recommendation system for financial planning is proposed in this work by using collaborative filtering based on the hybrid approach. Several drawbacks of the existing web page recommendation system like data sparsity and new user and item cold start problems are removed in this work [7]. In [8] proposed web page order is based on the URL. In this work, N-gram-based contents are functioning based on the historical data samples removed from the webpage, SVM and maximum entropy are used for the clustering purpose. The ODP and WEBKb dataset is used as the dataset in this experiment. The Model is working good compare to all the gram model and SVM classifier is used instead of logistic classifier. Still there is a scope for the improvement of F1 measure.

In [9] this work introduced recommendation model of the web page proposed by taking the sequential information's of the usage pattern of the web pages. The fuzzy clustering is also used in this proposed work. This proposed model is validated with the dataset MNSBC which is a real world dataset and obtained results are tested with the necessary parameters. The experimental results show our proposed model offers better performance over the existing works with a high accuracy rate. The privacy, trust and social networks with hybrid type of the intelligent systems can be included in the proposed work.

The proposed model performs the web page classification to help the web page users to access the required data. The data extraction is done to reduce noisy data, this improves the performance of different classification models and compares the classification performance [10] and estimates the different performance levels using the support vector machine, K nearest neighbor, logistic regression, Naïve Bayes, and random forest. Where SVM outperforms all the other classifiers. Similarly to that, [11] proposed an e-commerce product model for the web page recommendation by adding neighbor factor, time function and also added dynamic selection model to select adjacent object set. In the proposed technique RNN and mechanism of attention are combined for the web page recommendation. This estimation helps to improve the web page recommendation accuracy. Still there is scope for improving performance.

In [12] this work URL is categorized using machine learning techniques. Along with the category predictions, different privacy levels are used with three kind of RSA key length on the onion routing. [13] Proposed an HSC known as heuristic sequence clustering algorithm and a hybrid tree-based sequence clustering for web page recommendation. The dataset used for performing the test are CTI, BMSWebView1, BMSWebView2 and MSNBC datasets. Still scope for improving the performance.

From this, [14] proposed recommender systems review as the web is a huge source of information which includes huge learning materials and evolving pedagogy. Various research papers are collected and a review of the paper is written in this work analysing the web page recommendation system using various techniques. In [15] the paperwork E learning recommendation system based on ontology is proposed for the users cold start problem. Collaborative and content-based faltering techniques are used. The proposed system is mainly based on ontology. More work can be done on the learner's behaviour using machine learning techniques.

In [16] the proposed query optimization method based on online teaching and course recommendation. Classification of action verbs are used to increased accuracy. The proposed method also provides an analytical examination between the proposed and traditional methods. As the future work swarm intelligence and genetic algorithms can be used to improve accuracy of the course recommendation and query optimization which is online.

By considering that, [17] tag-based recommendation is done using the two different types of the algorithm such as ontology and Honey Bee algorithms. The data set used in this work is the MIND dataset. This has given an accuracy of 94.17% better than the already existing algorithm. Efficiency is estimated by using normalized discounted cumulative gain (NDCG). Still there is possibility of increase the performance by accuracy. In [18] the web page recommendation system is proposed in this work hybrid recommender is proposed which is a combination of content-based and collaborating filtering. The self organising maps of the neural network methods have been used along with the movie database. The performance of this novel technique is better than the existing filtering techniques. Still there is possibility of increase the performance by accuracy. In [19], the author proposed hybrid filtering techniques of the collaborating and content which uses the concept of the three phases, where the proposed hybrid technique is better than the conventional techniques. In the future online recommender system which checks acceptance level by the real users to provide recommendations can be implemented. From these, [20] proposed a review of the various techniques of web content mining. There are three different types of web mining, content mining, usage mining, and the final one is structure mining, where in the paper importance is given to web content mining. The techniques to solve the problems using web content mining and types of data are discussed in this paper. In [21] author presented a query knowledge-based query recommendation on the user interaction of types of river flow. According to the coverage region, the position and coordinates information are chosen for the picture flow analysis. The information is then simulated by producing random position changes and constructing a situation where certain data is missing in various modules. [21] Was validated using the procedures currently in use. For this analysis, UQSCM-RFD performance is checked for a few arbitrary queries in the SIGIR dataset. The SIGIR

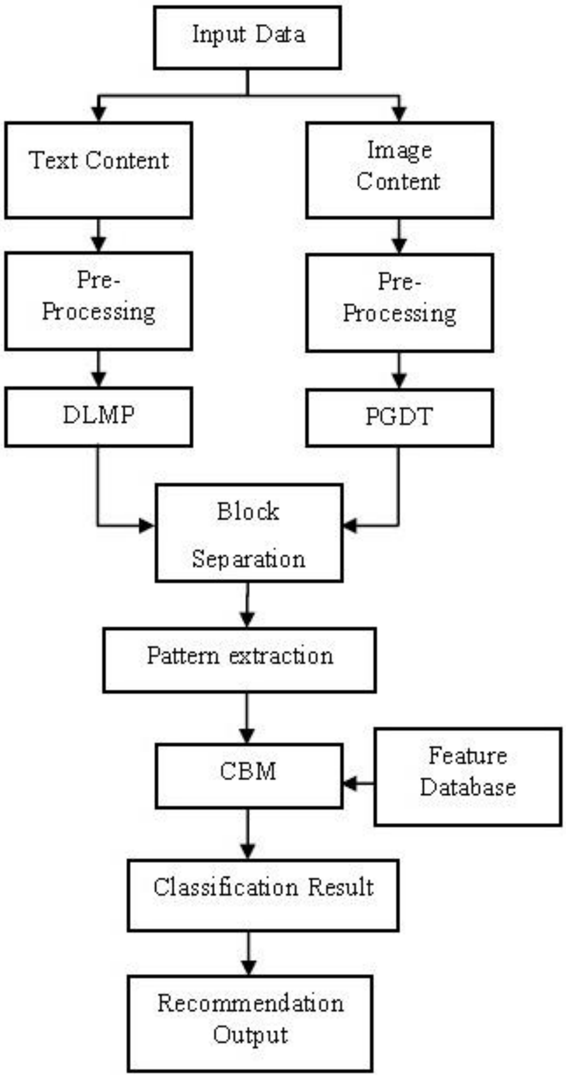


Figure. 2 Block diagram of the proposed model

dataset comprises 502,836 documents, 1696 queries and 2446 terms in queries, and 261,712 terms in documents. Still there is possibility of increase the performance by accuracy.

3. Proposed methodology

The proposed model of image and attribute features based on the probabilistic parameters are explained in this section. Improving predicted recognition accuracy using feature-based data format is the main goal of this effort, to implement novel techniques such as dynamic lexical magnitude pattern (DLMP) system and PGDT techniques are proposed.

The architecture of the proposed model of feature classification in Fig. 2 is processed in two functionalities such as,

- Pre-processing,
- Pattern classification and validation

The provided testing image-based data is first pre-processed to normalize the data with improved clustering and the best feature selection. To extract the data's patterns, the blocks of the pre-processed data are then separated, and the DLMP-PGDT pattern is used. The classifier is then used to determine whether it meets the relevance ratio.

3.1 Data pre-processing

One of the primary and crucial steps in any data processing application is often pre-processing of the data. Because it is crucial to guarantee that the succeeding processes will be more precise. Furthermore, it lessens the influence of any artefacts that can compromise classification accuracy. The input data is extracted from the website. Creating data clusters to reduce the feature set as effectively as possible is an effective pre-processing technique. It removes lost important spots and improves picture flow details to provide distinct texture patterns for subsequent processing. Additionally, normalization is used to remove unnecessary information from the data, integrating it. The following equation may be used to describe the loss in this model as E_{xy}:

$$E_{xy} = \begin{cases} C_{ij}, & if\left(mean\left(T_{ij}\right) > I_{xy}\right) \\ 0, & Otherwise \end{cases} \tag{1}$$

Where, I_{xy} represented the data key points for all x, y.
$x = \{1,2, \dots M\}$; Where, M is the row size of data.
$y = \{1,2, \dots N\}$; Where, N is the column size of data.
The following equation is used to execute the sharpening after obtaining the input data:

$$I_e(x, y) = I_{in}(x, y) + \lambda H(x, y) \tag{2}$$

Where H(x,y) denotes the high pass filter mask for clustering and λ denotes the tuning filter value. The data is then divided into cells using the following equation:

$$T_{ij} = I_e(x - 1{:}\, x + 1, y - 1{:}\, y + 1) \tag{3}$$

The average difference value in T_{ij} is then calculated with regard to the center key point of the mask matrix I_c and the size of the filter mask K. Here, Fig. 3 representation of the mask matrix's index is used where the matrix size is available in both 3 x 3 and 5 x 5 matrix types. With regard to the smallest mask size, filtering performance may be improved. By executing a greater key point reconstruction with

Received: October 10, 2022. Revised: November 18, 2022.

a smaller number of lost key points, this method of filtering lowers the loss ratio value.

Finally, $I_f(x, y)$, the pre-processed cluster output of data, is used to further extract patterns from the data. The following is an illustration of how the proposed CBM technology operates:

i-1, j-1	i, j-1	i+1, j-1
i-1, j	i, j	i+1, j
i-1, j+1	i, j+1	i+1, j+1

(a)

i-2, j-2	i-1, j-2	i, j-2	i+1, j-2	i+2, j-2
i-2, j-1	i-1, j-1	i, j-1	i+1, j-1	i+2, j-1
i-2, j	i-1, j	i, j	i+1, j	i+2, j
i-2, j+1	i-1, j+1	i, j+1	i+1, j+1	i+2, j+1
i-2, j+2	i-1, j+2	i, j+2	i+1, j+2	i+2, j+2

(b)

Figure. 3 Indexing of mask-matrix for data pre-processing: (a) 3 × 3 matrix and (b) 5 × 5 matrix

Algorithm 1: DLMP-PGDT algorithm

Input: Input Data T_D
Output: Features of attributes$F_D(s)$.
For i = 1 to M // run the loop for 'M' no. of iteration.
Initialize attributes 'y'and the weight value 'ω_i'
Calculate Potential of the attributes P_t^n
Estimate the likelihood of the attributes by $L_{1:i}^m = L_{1:i-1}^m \times L_i^m$
Update weight value, $\omega_i(n+1)$
Update Attributes, $y(n+1)$
Find maximum likelihood, $m_i^* = max(L_{1:i}^m)$
Find maximum relevance value, $\omega_i^*(n)$
If $(m_i^* > m_{i-1}^*)$, then
Update weight value of attributes and get best relevance value to form feature set.
If $(L_{1:i}^m)$ > 0, then
$s_v = \{s_{v-1}, i\}$
End if
Else
Continue for loop 'i'.
End If
$F_D(s) = T_D(s_v)$
End 'i' Loop

3.2 Pattern extraction

The DLMP system approach is used to process the filtered data after pre-processing. It is also a crucial step in the process of gathering the data needed to define each class of data by extracting the most pertinent information. Here, pattern extraction is mostly carried out to improve the classification process' overall accuracy and efficacy. With regard to the boundary of the data pattern, the zero padding is initialized with the size of 2 rows and columns in this approach, which uses the filtered data as the input for pattern extraction. The 5 x 5 mask may then be used to represent the window size to extract patterns from the input. These data cells are can be analyzed from five different angles of the data attributes that are can be indexed as$\{+90^0, +45^0, 0^0, -45^0, -90^0\}$.

$$I_{\alpha_L}(x, y) = \sum_{x=-N_1}^{N1} \sum_{y=-N_2}^{N_2} |I_W(x, y)| \times f_1(\alpha_L, \alpha_U, r) \quad (4)$$

Where, $f_1(\alpha_L, \alpha_U, r) = \begin{cases} 1 & if\ \alpha_L \leq \alpha_U < r \\ 0 & else \end{cases}$

$\alpha_L = \{+90^0, +45^0, 0^0, -45^0, -90^0\}, \alpha_U = \alpha_L - 45^0$ // 'r' is the size of mask-matrix. According to these index points, the neighboring features of the data samples are can be calculated from the Eq. (5). This can be represented as in the term of α_k.

$$\alpha_k = \{I_M((i-1)to\ (i+1), (j-1)to\ (j+1))\} \quad (5)$$

Following that, the set of key points for the closest neighborhoods are projected from the collections of boundary key points for the estimated average value of the neighborhood characteristics.μ_k, which is

$$\mu_k = \frac{1}{L} \sum_{a=1}^{L} \frac{|\alpha_k(a) - I_M(i,j)|}{I_M(i,j)} \quad (6)$$

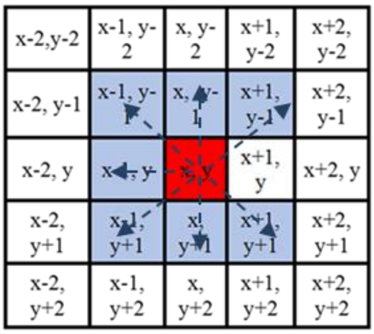
Similarly the average difference 'μ_c' between the centre value of the mask and the boundaries of it can be calculated as in the Eq. (7).

$$\mu_c = \frac{1}{L} \sum_{a=1}^{L} \frac{|\alpha_k(a) - I_c|}{I_c} \quad (7)$$

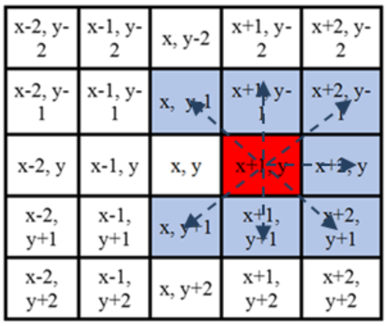
The extracted binary stream of the separated mask is produced by computing the mean values of μ_k and μ_c based on their sign differences for each iteration, as illustrated below:

$$S = \begin{cases} 1, & if(\mu_k > \mu_c) \\ 0, & Otherwise \end{cases} \quad (8)$$

Received: October 10, 2022. Revised: November 18, 2022.



(a)



(b)

Figure. 4 Phasor diagram of proposed pattern: (a) Estimate α_c and (b) Estimate α_k

Following that, the binary streams are used to determine the appropriate decimal value of B, which is shown as follows:

$$B = \sum_{k=0}^{n}\left(2^{k-1} \times S\right) \tag{9}$$

Therefore, it is projected that the maximum key point progression for each key point is γ_k, as shown below:

$$\gamma_k(x,y) = \max_{\alpha_L}\left(I_{\alpha_L}(x,y)\right) \tag{10}$$

Following that, the binary code mapping is carried out to extract the pattern vectors from the input, as seen in the following example:

$$I_p(x-2,y-2) = B \oplus \sum_{i=0}^{P} 2^i \times f_2\left(I_{Ref}(s,t), I_{\gamma_1}(s,t), I_{\gamma_2}(s,t)\right) \tag{11}$$

Where, $I_{Ref}(s,t) = I_W(x+t, y+t), \;\; \forall t = (-1:1)$

$$f_2(p,q,r) = \begin{cases} 1, & if\,(p < r \;\&\; q > O \& \; q > r) \\ 0, & else \end{cases} \tag{12}$$

Algorithm II provides a full explanation of the proposed DLMP-PGDT based pattern extraction process. The identification of neighborhood critical points with regard to various projection angle differences is shown in Fig. 4's phasor depiction of the DLMP-PGDT approach.

In this, the highlighted red marked region of the matric was used as the reference point to calculate the value of α. To do this, the neighborhood key points for the coordinate location of (x, y) may be represented for computing the value of "$'\alpha_k'$," and the estimated matrix center value can be represented as "α_c." The assessment of the current matrix's size is then shown by the projection angles of the arrow. The border key point and the center key point were not taken into account in this instance for ‘α_c’. The geometrical aspects are also taken from the data once the patterns are filtered data I_f. In this method, the data matrix's orientation θ is determined and may be rotated about to the updated angle to provide an exact orientation. The orientation is determined as follows:

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2\mu_{11}}{\mu_{20}-\mu_{02}}\right) \tag{13}$$

Where, μ_{ij} signifies the central instants of the data, which is projected as shown in underneath:

$$\mu_{ij} = \sum_i \sum_j \left(\left(I_f(x-x_\mu)\right)^i \left(I_f(y-y_\mu)\right)^j\right) \tag{14}$$

Where, the coordinate points ‘x’ and ‘y’ for each data points in the matrix are can be represented as x_μ and y_μ are estimated as follows:

$$x_\mu = \frac{I_f(1,0)}{I_f(0,0)} \tag{15}$$

$$y_\mu = \frac{I_f(0,1)}{I_f(0,0)} \tag{16}$$

$I_f(0,1)$ and $I_f(1,0)$ are the neighbor key points for the angles of 0° and 90°, respectively, where $I_f(0,0)$ is the data mask's central key point. In order to determine the data matrix's correct orientation depending on θ, the data matrix may then be rotated in relation to the updated, as illustrated below:

$$\theta' = \begin{cases} 90-\theta, & +90 \geq \theta \geq 0 \\ -90-\theta, & -90 \geq \theta < 0 \end{cases} \tag{17}$$

So, using the prior position and movement speed, the higher peaks and lower peaks in the picture data are generated. Using the x and y coordinates of the important locations in the data, the vertices of the higher peaks and lower peaks are then computed. These are used to extract the output geometrical key points F_V from the filtered data.

 DOI: 10.22266/ijies2023.0228.29

Received: October 10, 2022. Revised: November 18, 2022.

Algorithm 2 is used to precise categorize the recognized data after extracting the patterns from the history data of web page movement. Based on the data's supervised histogram feature vectors, it conducts several classification algorithms. By evaluating the weight of each neuron on the webpage, the texture classification is mostly carried out here to enhance the recognition process. This is to predict the network connectivity between the layers of NN and the feature attributes. The relevancy of the image and the text attributes are can be estimated as the binary value of '1' and '0'. By utilizing the patterns that were recovered during the earlier stage, this form of categorization can effectively increase the accuracy rate. The quantity of input data is trained in this case, and patterns are identified with a higher rate of accuracy.

Algorithm 2: Classification algorithm
Input: Features matrix of the training set $F_D(s)$ **Output:** Result from classifier $V(k)$ The initialization of the feature attributes as $F_D(s)$, $F_D(s) = \{T_{D1}(s), T_{D2}(s), \ldots, T_{Dm}(s)\}$ // Initialize the feature properties. The layers in the neural network are can be defined as the combination of data sequence that are can be represented as. $X_D(s) = \begin{bmatrix} F_{D1}(s) \\ F_{D2}(s) \\ \cdots \\ F_{Dm}(s) \end{bmatrix}$ // Matrix arrangement for input layer in the Block separation. The feature blocks are arranged from the matrix values are can be estimated as $F\left(X_D(s).X_D^*(s)\right)$. Arrange the kernel function of the classification model, K_m From the kernel function, the relevancy between the features are calculated as t_n. $t_n = F^T \omega_n$ // Texture relevancy. 'ω_n' weight value of attributes. $u_n = F^T \omega_n$ // Texture relevancy. Calculate the maximum value of matching between the training set and the testing feature attributes as $\widehat{T_s}$ Where, the relevance factor $X_b^{\bar{d}} \in R^{(T-T_p)M}$ Where, 'P' and 'Q^T' – Predicted component. The predicted label can be representing by $V(k) = \frac{d_{ij}}{R_j - R_i}$ Where, d_{ij} – Distance matrix for 'i' and 'j' of the relevance matrix 'R'.

4. Results analysis

The results of the proposed method of web page classification based on the image and attribute parameters are validated and compared with the existing methods. The performance of proposed classification model was calculated as the statistical parameters of texture based classification model that are presented as the table result and the graph plot of those parameters. The Python tool's version 3.8 implementation and testing included this.

Performance measures

The parameters that are used for the performance metric-based analytical process can be calculated by the statistical probability between the truly classified samples of data and the no. of misclassification results. These are all calculated from the arrangement of the confusion matrix that is evaluated by the comparison of classified result with the ground-truth of the dataset.

$$Sensitivity, TPR = \frac{True\ Positive\ (TP)}{Total\ No.of\ Positive\ samples} \quad (18)$$

$$Specificity, TNR = \frac{True\ Negative\ (TN)}{Total\ No.of\ Negative\ samples} \quad (19)$$

$$Jaccard, J = \frac{TP}{TP+FP+FN} \quad (20)$$

$$Dice\ Overalap, D = \frac{2J}{J+1} \quad (21)$$

$$Precision, P = (1 - FDR) = \frac{TP}{TP+FP} \quad (22)$$

$$Recall, R = (1 - FNR) \quad (23)$$

$$F1\ Score, F\_S = \frac{2TP}{2TP+FP+FN} \quad (24)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FN)(TP+FP)(TN+FN)(TN+FP)}} \quad (25)$$

$$Accuracy, Acc = \frac{Total\ correct\ labels}{Total\ No.of\ Samples} \quad (26)$$

$$Error\ (\%) = (100 - Accuracy\%) \quad (27)$$

$$Cohen's\ Kappa = 1 - \frac{(1-P_o)}{(1-P_e)} \quad (28)$$

Where, P_e – Hypothetical probability and the P_o – Probability of Relative observation, Fig. 5 and Table 1 show the comparison chart and the table result for the parameters of F1_score, Mathew's correlation coefficient (MCC), sensitivity, specificity, and precision of the proposed techniques. This represents that the DLMP-PGDT pattern extraction model in the CBM classification model of the proposed feature classification method achieved better classification performance than the existing method of TrAdaBoost [22].

The kappa coefficient and accuracy comparison between the current and suggested classifier models was shown in Fig. 6. This analysis demonstrates that, when compared to the current approach, the accuracy of the suggested method was boosted to 0.983 and the kappa coefficient was enhanced to 0.981. The comparative results for the parameters of the FPR and error rate are shown in Fig. 7. Table 2 plots the AUC value to represent the texture classification result for proposed web page prediction model.

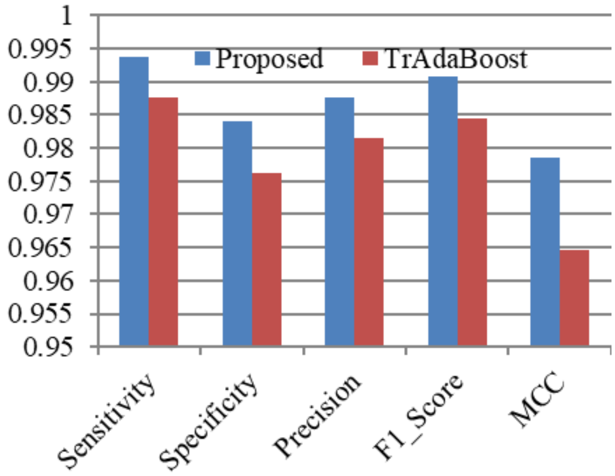


Figure. 5 Performance analysis

Table 1. Comparative analysis of performance among proposed and existing techniques

Parameters	***Proposed***	***TrAdaBoost***
Accuracy	0.98	0.97
Error rate	0.02	0.03
FPR	0.01	0.02
F1_Score	0.99	0.98
Kappa Coefficient	0.98	0.97
MCC	0.98	0.96
Precision	0.99	0.98
Specificity	0.98	0.97
Sensitivity	0.99	0.99

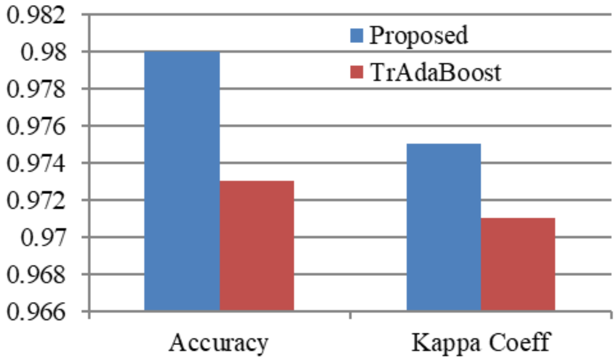


Figure. 6 Comparison results of kappa coefficient and accuracy for the existing and proposed model

Figure. 7 Comparison results for the parameters of the FPR and error rate

Table 2. AUC analysis

Methods	***AUC***
Proposed	0.965011
Bayes	0.739617
TrAdaBoost	0.922684

Table 3. Experimental analysis for accuracy

Methods	***Accuracy***
Proposed	0.98
Bayes	0.842
TrAdaBoost	0.97

4.1 Accuracy

The comparison of Accuracy from the classification output is shown in Table 3. For the parameters of data pattern-based feature characteristics, geometrical aspects of the input

Received: October 10, 2022. Revised: November 18, 2022.

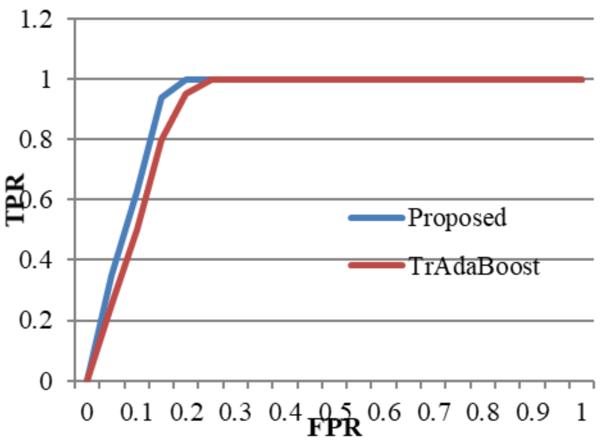


Figure. 8 ROC analyses

Table 3. Comparative analysis of performance among proposed and existing techniques

Methods	***Accuracy***
Proposed	98%
Novel Web page Recommender System [23]	61%

picture, and text data, the suggested DLMP-PGDT technique outperformed the other classification models in this regard. The efficiency of the proposed work can be expressed in the statistical parameters in the range of 0 to 1. This can also have expressed in terms of percentage by multiplying it by 100 for the ratio value of the probabilistic data analysis. This represents the amount correctly classified in the DLMP-PGDT method.

4.2 FPR and TPR

The FPR (false positive rate) and the TPR (true positive rate) of the referred to as the probability of samples that are truly detected as the positive class and the probability of samples that are misclassified as positive class respectively. The comparison chart with respect to the FPR vs TPR was displayed in Fig. 8 as the receiver operating curve (ROC) to represent the efficiency of the proposed classification model.

The overall experimental analysis results depicted that the proposed DLMP-PGDT-geometric feature extraction based classification technique provides an improved results compared to the other techniques.

In Table 3 the proposed system is compared with the existing novel web page recommender for the anonymous users using clustering of the web pages where in the filtered data 13745 sessions along with the 683 page views for performing the recommendation.

5. Conclusion

The proposed work presented a novel pattern extraction-based classification method for image feature data forecasting and recognition. For this purpose, various data processing techniques are employed in this work at the stages of pre-processing, block separation, pattern extraction, and classification. Initially, the web page data cluster was implemented to reduce the loss of key points and to integrate with the input data. During this process, the mask matrix is constructed in the form of 3×3 and 5×5, which ensures the reduced loss / error ratio. After pre-processing the data, block separation is performed to increase the overall efficiency of recognition. Here, the DLMP-PGDT technique is utilized to extract the most useful patterns by computing the intensity of the centre key point with its neighboring key points. This type of feature extract increases the overall accuracy of the data recognition system. At last, the classifier is deployed to classify whether the data is relevant to the prediction or not based on the extracted feature vectors. In this paper, traditional classification model was compare with the proposed mechanism, where the results depicted that the combination of PGDT-CBM technique outperforms the other techniques.

The proposed system classifier has given the result with accuracy of about 98%. The model is better in performance than that of the various existing models as shown in the comparison results. Also, it efficiently improves the performance of classification with improved recognition rate, accuracy, and reduced error value.

The future work of this proposed model was can be improved by proposing the optimal classification model with the parameters of image and the other data pattern based texture feature based techniques.

Conflicts of interest

The authors declare that there is no conflict of interest

Author contribution

Chaithra, Lingaraju Gowdru Malleshappa and Jagannatha Sreenivasaiah contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

Received: October 10, 2022. Revised: November 18, 2022.

References

[1] C. K. Leung, F. Jiang and J. Souza, "Web page recommendation from sparse big web data", In: *Proc. of International Conference on Web Intelligence (WI)*, Santiago, Chile pp. 592-597, 2018.

[2] G. Deepak and J. S. Priyadarshini, "Personalized and enhanced hybridized semantic algorithm for web image retrieval incorporating ontology classification, strategic query expansion, and content-based analysis", *Computers & Electrical Engineering*, Vol.72, pp. 14-25, 2018.

[3] H. Chen, "Personalized recommendation system of e-commerce based on big data analysis", *Journal of Interdisciplinary Mathematics*, Vol. 21, No. 5, pp.1243-1247, 2018.

[4] G. Deepak and J. S. Priyadarshini, "A hybrid semantic algorithm for web image retrieval incorporating ontology classification and user-driven query expansion", In: *Proc. of Advances in Big Data and Cloud Computing*, Springer, Singapore, pp. 41-49, 2018.

[5] J. Bhavithra and A. Saradha, "Personalized web page recommendation using case-based clustering and weighted association rule mining", *Cluster Computing*, Vol. 22, No. 3, pp. 6991-7002, 2019.

[6] S. K. Raghuwanshi and R. K. Pateriya, "Collaborative filtering techniques in recommendation systems", *Data, Engineering and Applications*, Springer, Singapore, pp. 11-21, 2019.

[7] N. Pereira and S. L. Varma, "Financial planning recommendation system using content-based collaborative and demographic filtering", *Smart Innovations in Communication and Computational Sciences*, Springer, Singapore, pp. 141-151,2019.

[8] V. S. Rathore and N. Singh, "Plan of Proficient URL Based Web Page Classification Utilizing NLP", *International Journal of Research in Engineering, Science and Management*, Vol. 2, No. 10, 2019.

[9] R. Katarya and O. P. Verma, "An effective web page recommender system with fuzzy c-mean clustering", *Multimedia Tools and Applications*, Vol. 76, No. 20, pp. 21481-21496, 2017.

[10] S. Lassri, E. H. Benlahmar and A. Tragha, "Web Page Classification Based on an Accurate Technique for Key Data Extraction", In: *Proc. of International Conference on Advanced Intelligent Systems for Sustainable Development*, Springer, Cham, pp. 1124-1131, 2020.

[11] K. Wang, T. Zhang, T. Xue, Y. Lu and S. G. Na, "E-commerce personalized recommendation analysis by deeply-learned clustering", *Journal of Visual Communication and Image Representation*, Vol. 71, p. 102735, 2020.

[12] I. H. Liu, Y. L. Chang, J. S. Li and C. G. Liu, "Differential Privacy Protection with Group Onion Routing based on AI-based URL Classification", In: *Proc. of the 2020 ACM International Conference on Intelligent Computing and Its Emerging Applications*, pp. 1-6, 2020.

[13] H. Singh and P. Kaur, "An effective clustering-based web page recommendation framework for e-commerce websites", *SN Computer Science*, Vol. 2, No. 4, pp. 1-20, 2021.

[14] J. Joy and R. V. G. Pillai, "Review and classification of content recommenders in E-learning environment", *Journal of King Saud University-Computer and Information Sciences*, 2021.

[15] J. Jeevamol and V. G. Renumol, "An ontology-based hybrid e-learning content recommender system for alleviating the cold-start problem", *Education and Information Technologies*, Vol. 26, No. 4, pp. 4993-5022, 2021.

[16] M. S. Rafiq, X. Jianshe, M. Arif and P. Barra, "Intelligent query optimization and course recommendation during online lectures in E-learning system", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 12, No. 11, pp.10375-10394, 2021.

[17] N. Roopak and G. Deepak, "KnowGen: a knowledge generation approach for tag recommendation using ontology and honey bee algorithm", In: *Proc. European, Asian, Middle Eastern, North African Conference on Management & Information Systems*, pp. 345-357. Springer, Cham, 2021.

[18] Y. Afoudi, M. Lazaar and M. A. Achhab, "Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network", *Simulation Modelling Practice and Theory*, Vol. 113, p. 102375, 2021.

[19] S. Sharma, V. Rana and M. Malhotra, "Automatic recommendation system based on hybrid filtering algorithm", *Education and Information Technologies*, Vol. 27, No. 2, pp. 1523-1538, 2022.

[20] P. H. Shah, and B. Pandit, "A Review: Web Content Mining Techniques", *Data Engineering for Smart Systems*, pp.159-172, 2022.

[21] G. Deepak and A. Santhanavijayan, "UQSCM-RFD: a query–knowledge interfacing approach

Received: October 10, 2022. Revised: November 18, 2022.

for diversified query recommendation in semantic search based on river flow dynamics and dynamic user interaction", *Neural Computing and Applications*, Vol. 34, No. 1, pp. 651-675, 2022.

[22] L. Yan, R. Zhu, Y. Liu and N. Mo, "TrAdaBoost based on improved particle swarm optimization for cross-domain scene classification with limited samples", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 11, No. 9, pp. 3235-3251, 2018.

[23] R. Wagh and J. Patil, "A Novel Web Page Recommender System for Anonymous Users Based on Clustering of Web Pages", *Asian Journal for Convergence in Technology (AJCT)*, Vol. 5, No. 1, pp. 1-4, 2019.