International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Multimedia Classification Using DT and SVM-based Multimodal Classifiers: A Metadata-Based Approach

Shwetha[1], Dr Ashokkumar TA[2], Jeevitha S[3]

[1]PG Computer Applications, CMR University, Bengaluru, India
Email: *shwetha.d[at]cmr.edu.in*

[2]Professor, CMR University, Bengaluru, India
Email: *ashokkumar.t[at]cmr.edu.in*

[3]Assistant Professor, Excel Engineering College, Salem, India
Email: *gobi.rosary[at]gmail.com*

Abstract: ***The rapid growth of web-based multimedia repositories has created significant challenges in the efficient organization, indexing, and retrieval of digital content. Traditional content-based classification methods often require high computational resources, making metadata-driven approaches an attractive alternative. This study proposes a metadata-based multimedia classification framework that employs two supervised machine learning algorithms, Decision Tree (DT) and Support Vector Machine (SVM), to categorize multimedia video files. Metadata are extracted using the Media Info tool and subsequently preprocessed to remove redundant and irrelevant attributes, improving data quality and classification reliability. Following feature selection, twenty-two significant metadata attributes are retained for model development. The dataset is partitioned into training and testing subsets using a 60:40 ratio to ensure consistent performance evaluation. Both classifiers are assessed using standard metrics, including classification accuracy and confusion matrix analysis. Experimental results demonstrate that metadata provide sufficient descriptive information for effective multimedia classification while eliminating the need for computationally intensive content analysis. Comparative evaluation reveals distinct differences in the predictive capability and classification efficiency of DT and SVM, enabling a comprehensive assessment of their suitability for metadata-driven multimedia categorization. The proposed framework offers a scalable, computationally efficient, and practical solution for multimedia management, supporting applications such as digital libraries, content management systems, multimedia search engines, and intelligent web-based information retrieval.***

Keywords: Multimedia Classification, Metadata Extraction, Machine Learning, Decision Tree, Support Vector Machine, MediaInfo, Feature Selection, Web Multimedia, Content Management, Information Retrieval

1. Introduction

The rapid growth of digital communication, cloud computing, and high-speed Internet has led to an unprecedented increase in multimedia content across video streaming platforms, social media, digital libraries, and educational portals. Efficient organisation, indexing, and retrieval of these resources have become major challenges, making automated multimedia classification an important research area. Multimedia combines heterogeneous data such as text, images, audio, and video, with multimedia video representing the richest form of information. Besides content, multimedia files also contain metadata describing technical characteristics such as file format, codec, duration, resolution, frame rate, bit rate, and file size. Unlike computationally expensive content-based analysis, metadata-based classification offers a faster and more scalable alternative [1].

Traditional approaches to multimedia classification are often labor-intensive, expensive, and susceptible to inconsistent results, highlighting the need for intelligent automated solutions. This study presents a metadata-based multimedia classification framework that utilises metadata extracted from video files using the MediaInfo tool. After preprocessing to eliminate redundant, missing, and irrelevant attributes, the refined dataset is analysed in the KNIME Analytics Platform. The dataset is divided into 60% training and 40% testing samples for model development. Decision Tree (DT) and Support Vector Machine (SVM) algorithms are employed to classify multimedia content based on metadata characteristics. Their performance is assessed using accuracy, precision, recall, F1-score, confusion matrix, and other standard evaluation measures. The proposed framework demonstrates that metadata alone can achieve reliable and computationally efficient multimedia classification, support applications such as digital libraries, multimedia retrieval, video indexing, recommendation systems, and large-scale web content management while providing a scalable solution for intelligent multimedia organization [2]. Each component of the proposed framework is discussed comprehensively in the following subsections [3].

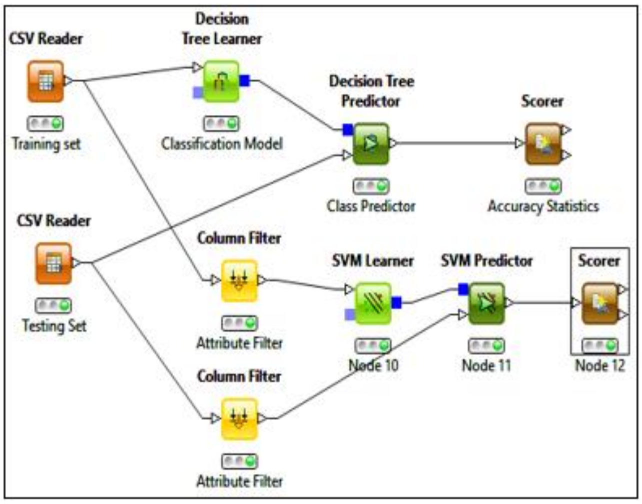


Figure 1: System model of the proposed methodology

The proposed framework begins with metadata extraction and preprocessing from web multimedia video files. Instead of analysing computationally intensive content such as video frames, images, and audio signals, the framework uses metadata that describe the technical and structural characteristics of multimedia files. Metadata are extracted using the MediaInfo tool, which initially generates 27 descriptive attributes related to video, audio, images, text, and file properties [4].

To improve classification efficiency, a feature selection process evaluates the relevance of each attribute. Five attributes—Codec ID/Info, Frame Rate Mode, Color Space, Scan Type, and Compression Mode—are removed because they contain identical values across all records and provide no discriminatory information. The remaining 22 metadata attributes, including duration, bit rate, frame dimensions, aspect ratio, image properties, textual statistics, file size, and class label, are retained for analysis [5].

The refined metadata are exported in CSV format and subjected to preprocessing, including data cleaning, missing value imputation, consistency checking, and removal of irrelevant records. Numerical missing values are replaced using the mean, while categorical values are imputed using the mode. The resulting high-quality metadata repository provides an efficient and scalable input for supervised machine learning, enabling accurate multimedia classification with significantly lower computational cost than conventional content-based approaches [6].

2. Classification Model

The effectiveness of a multimedia classification system depends on the capability of the selected classification model. This study employs two supervised machine learning algorithms, Decision Tree (DT) and Support Vector Machine (SVM), to classify web multimedia videos using extracted metadata. Both models are trained and tested under identical experimental conditions to ensure a fair performance comparison. Their effectiveness is evaluated using standard classification metrics, including accuracy and confusion matrix analysis [7].

The Decision Tree classifier constructs a hierarchical tree by recursively partitioning the dataset using the most informative metadata attributes. The model development involves two stages: attribute selection and classification rule generation. Attribute selection identifies the metadata feature that provides the highest discriminatory power using the Information Gain criterion. The training dataset consists of 247 labelled multimedia records belonging to three categories: Sports (85), News (100), and Entertainment (62). Information Gain is calculated for each metadata attribute to determine the optimal splitting feature, enabling the generation of accurate and interpretable classification rules for multimedia categorization [8].

$$Info_{A(D)} = \sum_{i=1}^{n} \frac{ID_i I}{IDI} * Info\ (D_i) \ldots\ldots (2)$$

$\text{Info}_{IR}\ (D) = -\frac{85}{247}\ (-\ \frac{26}{26}\log_2(\frac{26}{26}) - \frac{0}{26}\log_2\frac{0}{26}) + -\frac{100}{247}\ (-\ \frac{61}{91}\log_2(\frac{61}{91}) - \frac{30}{91}\log_2\frac{30}{91})$

$+ -\frac{62}{247}\ (-\ \frac{77}{95}\log_2(\frac{77}{95}) - \frac{18}{95}\log_2\frac{18}{95})$

= 0.240+ (-0.096) + (-0.175)

= 0.511

After obtaining the expected information, the Information Gain of the selected attribute is determined using Equation (3)[10].

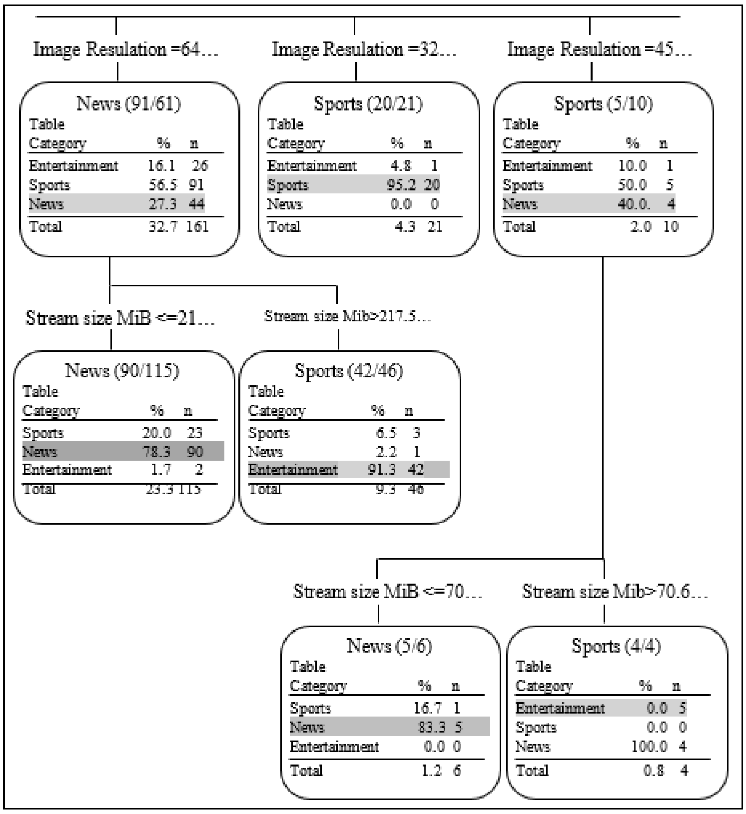
Gain (A)= Info (D)- Info_A(D)……(3)

Gain (IR) = Info(D) - InfoIR(D)
Gain (IR) = 0.666 - 0.511 Gain (IR) = 0.155 bits

Table 1: Information Gain of Multimedia Metadata Attributes

S. No	Multimedia Metadata Attribute	Information Gain
1	Video Duration	0.116
2	Video Bit rate kbps	0.126
3	Maximum bit rate kbps	0.13
4	Width Pixels	0.111
5	Height Pixels	0.115
6	Display aspect ratio	0.125
7	Bits/(Pixel*Frame)	0.138
8	Stream size MiB	0.142
9	Audio Duration	0.116
10	Audio Bit rate kbps	0.125
11	Maximum bit rate kbps	0.128
12	Stream size MiB	0.14
13	Image Resolution	0.155
14	Image Height	0.115
15	Image Width	0.111
16	Text Page	0.109
17	Word count	0.116
18	Character count	0.118
19	Line count	0.099
20	Paragraph count	0.11
21	Size in kbps	0.121

After calculating the Information Gain for all selected metadata attributes, the dataset is partitioned into subsets, and the attribute with the highest Information Gain is selected as the optimal splitting node. This process is repeated recursively until the complete Decision Tree is generated. In the proposed model, Image Resolution produces the highest Information Gain and is therefore chosen as the root node [13].



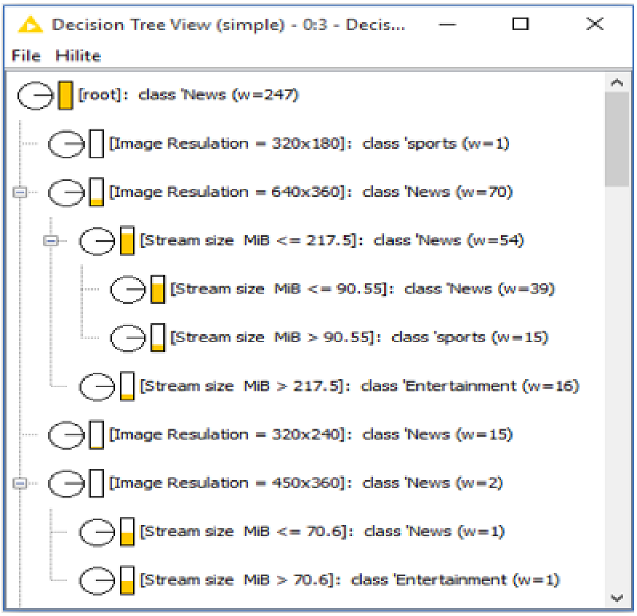


Figure 2(a) and (b): Tree structure result of the DT classification model and Classification rules

3.2.2 Support Vector Machine Classification Model (150 Words)

Classification rules are generated by tracing every path from the root node to the leaf nodes of the Decision Tree. Each path represents a sequence of decisions based on selected metadata attributes, while each leaf node indicates the predicted multimedia category. Since Image Resolution provides the highest Information Gain, it is selected as the root node, and the remaining nodes are formed through recursive partitioning. The resulting rules provide a transparent and interpretable mechanism for classifying multimedia records. The Support Vector Machine (SVM) is another supervised learning algorithm employed in this study for metadata-based multimedia classification. SVM identifies an optimal separating hyperplane that maximises the margin between different multimedia classes, thereby improving prediction accuracy and reducing misclassification. This characteristic enables the model to generalize effectively to unseen data, particularly in high-dimensional feature spaces. The classifier is trained using the extracted metadata and implemented in the KNIME Analytics Platform. Its performance is subsequently compared with the Decision Tree model using standard evaluation metrics to determine the most effective classifier for multimedia categorization [14].

$$\Phi(w) = \frac{1}{2} \parallel w \parallel^2$$

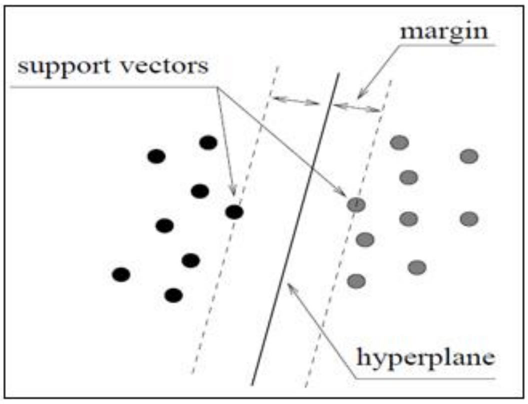


Figure 3: A Linear Support Vector Machine

The SVM classification process consists of two main stages, namely training and testing. During the training phase, multimedia metadata instances are provided to the classifier. Each training instance contains a set of metadata attributes together with its corresponding class label. Based on these labelled samples, the SVM algorithm learns the relationship between metadata features and multimedia categories to construct an effective classification model [15].

After the training process is completed, the learned model is applied to the testing dataset to predict the class labels of previously unseen multimedia instances. The input feature vectors are mapped into a higher-dimensional feature space using an appropriate transformation function, allowing the classifier to identify an optimal separating hyperplane with the maximum margin. This process enables the SVM model to classify multimedia content accurately based on the extracted metadata features [16].

3.3 Classification Analysis

After constructing the Decision Tree (DT) and Support Vector Machine (SVM) classification models, their performance is evaluated using commonly accepted classification metrics. Both classifiers are tested under the same experimental conditions so that their prediction capabilities can be compared fairly. The evaluation is carried out using performance measures such as True Positive (TP), False Positive (FP), Precision, Recall, and F-Measure, which provide quantitative information about the effectiveness of each classification model [17].

Precision indicates the proportion of correctly classified positive instances among all instances predicted as positive, whereas Recall measures the ability of the classifier to identify all relevant instances belonging to a particular class. The F-Measure combines both Precision and Recall into a single metric, providing a balanced assessment of the overall classification performance [18].

In addition to these evaluation measures, a confusion matrix is used to analyse the classification results produced by both the DT and SVM models. The confusion matrix summarises the numbers of correctly and incorrectly classified multimedia instances for each category, making it easier to examine the strengths and weaknesses of the classifiers. Based on these evaluation measures, the performance of the DT and SVM models is compared to determine their classification accuracy, prediction efficiency, and overall suitability for metadata-based multimedia classification [19].

3. Experimental Results and Discussion

4.1 Classification using the Decision Tree Model

To evaluate the performance of the proposed Decision Tree (DT) classification model, experiments were conducted using 247 web multimedia-video metadata instances extracted during the metadata extraction and preprocessing stages. The dataset was divided into training and testing sets according to the methodology described in the previous section. The Decision Tree classifier was trained using the selected metadata attributes, and its performance was evaluated using standard measures, including the number of correctly classified instances, incorrectly classified instances, True Positive (TP), False Positive (FP), Precision, Recall, and F-Measure. These evaluation metrics provide a comprehensive assessment of the classifier's prediction capability for each multimedia category. Precision indicates the proportion of correctly predicted instances, whereas Recall measures the classifier's ability to identify all relevant instances. The F-Measure combines Precision and Recall to provide an overall evaluation of classification performance.

The classification results obtained from the Decision Tree model are presented in Table 2, which summarises the performance of the Sports, News, and Entertainment categories together with their corresponding TP, FP, Precision, Recall, and F-Measure values. The Decision Tree classifier correctly classified 234 of the 247 multimedia instances, while 13 instances were incorrectly classified. The Sports category achieved the highest classification performance with a high Precision value and a low False Positive rate. The News category also performed well, with 99 out of 100 instances classified correctly. The Entertainment category produced satisfactory results, demonstrating the effectiveness of the selected metadata attributes. Overall, the Decision Tree model achieved an accuracy of 94.7%, confirming its effectiveness for metadata-based multimedia classification.

The obtained results demonstrate that metadata extracted from multimedia files provide sufficient descriptive information for effective content classification. The high classification accuracy achieved by the Decision Tree model indicates that the selected metadata attributes successfully capture the distinguishing characteristics of the multimedia categories. Furthermore, the evaluation measures show consistent performance across all three classes, suggesting that the proposed metadata-based classification approach can accurately classify multimedia content while maintaining reliable prediction performance.

Classification using the SVM Model

The Support Vector Machine (SVM) classification model was evaluated using the same dataset consisting of 247 web multimedia-video metadata instances. Experimental results show that the classifier correctly classified 220 instances, while 27 instances were incorrectly classified. The performance of the model was analysed using True Positive (TP), False Positive (FP), Precision, Recall, and F-Measure, and the detailed results are presented in Table 3. Among the three multimedia categories, the Entertainment class achieved the highest Precision and classification accuracy, whereas the Sports category recorded a comparatively low False Positive rate. In the News category, 91 out of 100 instances were classified correctly, while 9 instances were misclassified. Overall, the SVM classifier achieved an accuracy of 89%, demonstrating that it provides satisfactory performance for metadata-based multimedia classification and serves as an effective approach for classifying web multimedia content.

Table 2: Classification result of Decision Tree classification model

S. No	Class Labels	Total Instances	Correctly Classified	Incorrectly Classified	TP	FP	Precision	Recall	F-Measure
1	Sports	85	76	9	76	2	0.974	0.894	0.933
2	News	100	99	1	99	8	0.925	0.99	0.957
3	Entertainment	62	59	3	59	3	0.952	0.952	0.952
	Total	247	234	13	234	13	0.950	0.945	0.947

Classification using the SVM Model

The Support Vector Machine (SVM) classification model was evaluated using the same dataset consisting of 247 web multimedia-video metadata instances. Experimental results show that the classifier correctly classified 220 instances, while 27 instances were incorrectly classified. The performance of the model was analysed using True Positive (TP), False Positive (FP), Precision, Recall, and F-Measure, and the detailed results are presented in Table 3. Among the three multimedia categories, the Entertainment class achieved the highest Precision and classification accuracy, whereas the Sports category recorded a comparatively low False Positive rate. In the News category, 91 out of 100 instances were classified correctly, while 9 instances were misclassified. Overall, the SVM classifier achieved an accuracy of 89%, demonstrating that it provides satisfactory performance for metadata-based multimedia classification and serves as an effective approach for classifying web multimedia content.

Table 3: Classification result of SVM classification model

S. No	Class Labels	Total Instances	Correctly Classified	Incorrectly Classified	TP	FP	Precision	Recall	F-Measure
1	Sports	85	69	16	69	4	0.945	0.812	0.873
2	News	100	91	9	91	16	0.85	0.91	0.879
3	Entertainment	62	60	2	60	7	0.968	0.968	0.93
	Total	247	220	27	220	27	0.921	0.896	0.894

Table 3 presents the comparative analysis of the Decision Tree (DT) and Support Vector Machine (SVM) classification models for web multimedia-video data. The comparison is based on standard performance measures, including classification accuracy, Precision, Recall, F-Measure, and False Positive rate. The results indicate that the Decision Tree classifier achieved better overall performance than the SVM classifier.

Table 4: Confusion matrix of DT and SVM classification result

Confusion Matrix							
Decision Tree Classification Model				Decision Tree Classification Model			
===**Confusion Matrix**===				===**Confusion Matrix**===			
a	**b**	**c**	**←Classified as**	**a**	**b**	**c**	**←Classified as**
76	6	3	| a= Sports	76	6	3	| a= Sports
1	99	0	| b= News	1	99	0	| b= News
1	2	59	| c= Entrainment	1	2	59	| c= Entrainment

The experimental results indicate that the Decision Tree (DT) classification model performs better than the Support Vector Machine (SVM) classification model. Since the web multimedia-video metadata dataset primarily consists of continuous-valued independent attributes, the Decision Tree model achieved higher classification accuracy, whereas the SVM model produced comparatively lower accuracy.

4. Conclusion and Future Work

In this study, web multimedia-videos were classified into different categories using metadata extracted from multimedia content. The extracted metadata were stored in a database and used to train and evaluate the Decision Tree (DT) and Support Vector Machine (SVM) classification models. The experimental results showed that the Decision Tree classifier outperformed the SVM classifier in terms of overall classification accuracy and efficiency. The DT model achieved an accuracy of 94.7%, whereas the SVM model achieved an accuracy of 89%. The comparatively lower performance of the SVM model may be attributed to the characteristics of the metadata features used for classification. Overall, the results demonstrate that the Decision Tree algorithm is a more effective approach for metadata-based web multimedia-video classification. As future work, the classification accuracy of the SVM model can be improved by exploring enhanced feature selection methods, parameter optimization techniques, or additional metadata attributes for web multimedia-video classification.

References

[1] P. M. Kamde and S. P. Algur, "A Survey on Web Multimedia Mining," International Journal of Multimedia & Its Applications, vol. 3, no. 3, pp. 72–84, 2011.

[2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349–1380, 2000.

[3] C. G. M. Snoek and M. Worring, "Multimodal Video Indexing: A Review of the State-of-the-Art," Multimedia Tools and Applications, vol. 25, no. 1, pp. 5–35, 2005.

[4] MediaArea.net, "MediaInfo: Technical and Tag Data for Video and Audio Files," MediaInfo documentation, MediaArea.net. [Online]. Available: https://mediaarea.net/en/MediaInfo

[5] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2011.
[6] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Burlington, MA, USA: Morgan Kaufmann, 2016.
[7] T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.
[8] J. R. Quinlan, "Induction of Decision Trees," Machine Learning, vol. 1, pp. 81–106, 1986.
[9] C. E. Shannon, "A Mathematical Theory of Communication," Bell System Technical Journal, vol. 27, no. 3, pp. 379–423, 1948.
[10] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," Journal of Machine Learning Research, vol. 3, pp. 1157–1182, 2003.
[11] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees. Belmont, CA, USA: Wadsworth International Group, 1984.
[12] J. R. Quinlan, C4.5: Programs for Machine Learning. San Mateo, CA, USA: Morgan Kaufmann, 1993.
[13] S. R. Safavian and D. A. Landgrebe, "A Survey of Decision Tree Classifier Methodology," IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 3, pp. 660–674, 1991.
[14] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, pp. 273–297, 1995.
[15] J. C. Platt, "Fast Training of Support Vector Machines Using Sequential Minimal Optimization," in Advances in Kernel Methods—Support Vector Learning, B. Schölkopf, C. J. C. Burges, and A. J. Smola, Eds. Cambridge, MA, USA: MIT Press, 1998, pp. 185–208.
[16] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, vol. 2, no. 2, pp. 121–167, 1998.
[17] M. Sokolova and G. Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," Information Processing & Management, vol. 45, no. 4, pp. 427–437, 2009.
[18] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," Journal of Machine Learning Technologies, vol. 2, no. 1, pp. 37–63, 2011.
[19] T. Fawcett, "An Introduction to ROC Analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.
[20] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd ed. Hoboken, NJ, USA: Wiley-Interscience, 2006.