

Test and Image Plagarism Detection

Sk. Shehanaj, R.Gayathri, V. Jahnavi, P. Pavan

*(Under the guidance of Dr. A. Balaji, Professor, Head of the Department,
Department of Computer Science and Engineering, Tirumala Engineering College)*

ABSTRACT

Plagiarism is a serious concern in academia and creative industries, where originality of content is critical. Text and image plagiarism detection are two techniques used to identify instances of copying or unattributed use of existing content. Text plagiarism detection involves comparing the text of a document with a database of existing texts to identify similarities and possible instances of copying. This technique is commonly used in academic settings, journalism, and other fields where originality of content is essential. Image plagiarism detection, on the other hand, involves identifying whether an image has been copied or altered from an existing source.

Keywords— *Cancer, recognition, Logistic Regression, KNN, SVM, Random Forest, Decision Tree*

I. INTRODUCTION

Plagiarism detection is the process of identifying whether a piece of content has been copied from an existing source without proper attribution or permission. Text and image plagiarism detection are two of the most commonly used techniques for identifying plagiarism in written and visual content, respectively. Text plagiarism detection involves comparing the text of a document to a database of existing texts to identify any instances of copying.

There are several software tools available that can perform this task automatically, using algorithms to identify similarities between the text being analyzed and existing sources. These tools are widely used in academia, journalism, and other fields where originality of content is critical. Image plagiarism detection, on the other hand, involves identifying

whether an image has been copied or altered from an existing source. This can be done using a variety of techniques, including reverse image search, which involves uploading an image to a search engine to identify any instances of its use elsewhere on the internet. Other methods include watermarking, where a unique digital signature is embedded in an image to identify its origin, and image analysis algorithms that can identify similarities between images and identify whether they are original or copied. Both text and image plagiarism detection are important tools for maintaining the integrity of content and ensuring that creators are properly credited for their work. They are particularly useful in academic settings, where plagiarism can result in severe consequences, including failing grades, suspension, or expulsion. They are also useful in the publishing and media industries, where originality of content is essential for maintaining credibility and avoiding legal issues. standards of academic and creative integrity.

II. Literature survey

This paper gives a brief idea about classification, the classification is done based on language in the documents. Languages are classified as Mono-lingual and cross-lingual or multi-lingual. Mono-lingual plagiarism detection identifies and extracts texts from the document and detects language of same kind i.e English-English plagiarism. Cross-lingual or multi-lingual plagiarism detection also deals with identification and extraction of text from document and detects language of different kinds i.e English-Arabic plagiarism. Based Plagiarism Detection for Flowchart Figures in Texts does pre-processing by determining.

III. PROPOSED SYSTEM:

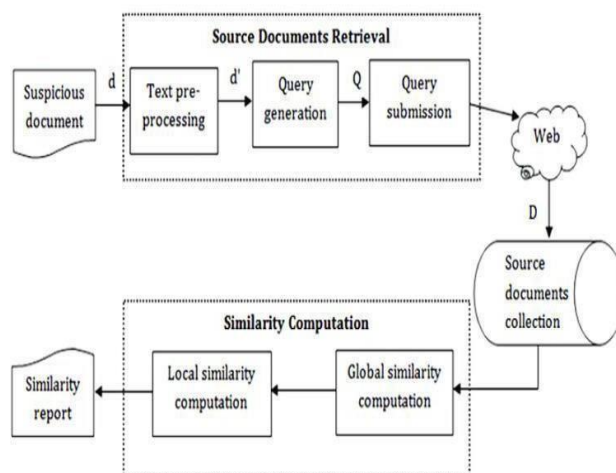


Fig: Proposed System

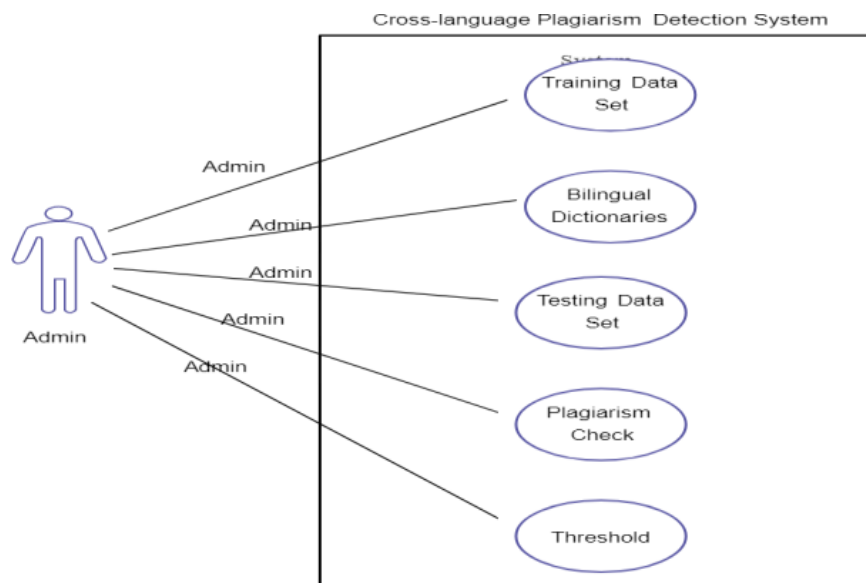
The proposed system is as shown in above figure. At first, the database is collected from cancer organizations photos and images in database, the data analysis approach selects the images with the most comparable correlations to the query image.

Correlation levels at this step are used to report on the tested picture plagiarism, and the expert is responsible for the ultimate interpretation of the results. Training and testing are the two main components of the system as it is currently envisioned.

They are seen as using the Histogram in the learning phase and the modelling done by this network in the testing phase for the recognition stage in the train phase. Based on correlation rates between query photos and images in database.

the data analysis approach selects the images with the most comparable correlations to the query image. Correlation levels at this step are used to report on the tested picture plagiarism, and the expert is responsible for the ultimate interpretation of the results.

The architecture of proposed system is shown in Fig. Based on the research article by Senosy Arrish et al [11] this system focuses entirely on flowchart images. This system works by comparing the shape of the query image with those in the database as shown in Figure 1.



This method begins by pre-processing the flowchart to get maximum result, reducing error and better accuracy in plagiarism detection. The pre-processing has 3 stages which is thinning the images, removing connected lines and removing texts. The database consists of 2 sets. The 1st set is for storing figures and the 2nd set is for storing preprocess figure. Each of these figures has an ID which makes it easier to recall in the comparison stages.

The main objective of the project is to develop a facility that can easily and accurately predict cancer. The machine learning makes it easier with the algorithms. The proposed system is developed using the symptoms that a cancer victim should have. These can be easily understandable to the user. The existing system is developed using the cell radius, colour etc. which are the values of the tumor. It gives accurate results but the victim is not able to identify the cell parameters. Hence the proposed system will be advantageous and friendly to predict the presence of cancer.

IV. TRAINING AND TESTING THE DATASET

The Testing in smart governance through big data involves ensuring that the data-driven solutions implemented are accurate, reliable, and secure. Here are some of the key considerations for testing in smart governance through big data: Data Quality Testing: Ensure the data is accurate, complete, and reliable. This includes testing the data sources, data transformation processes, and data storage mechanisms. Performance Testing: Test the performance of the data processing and analysis systems to ensure they can handle large volumes of data and produce results in real-time.

V. RESULT

Among all those algorithms, the best performance is selected. For mainly three types of images.

Fig: Copied image



Fig: Original image

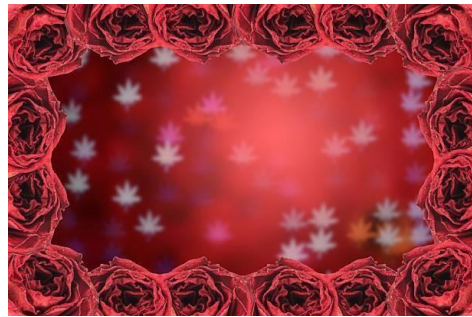


Fig: Original image



Fig: Original image

CONCLUSION

This Plagiarism is a serious issue in academia and other professional settings, as it involves presenting someone else's work as one's own. Text and image plagiarism are two common forms of plagiarism that can be detected using various methods.

Text plagiarism detection involves comparing a given text against a database of known sources to identify any matching content. This can be done manually or using plagiarism detection software that relies on algorithms to compare the text against a database of sources.

REFERENCES

- [1] G. Sruthi, Ch. Likitha Ram, M. Koushik Sai, BhanuPratap Singh, Nikhil Majhotra, Neha Sharma; "Cancer prediction using machine learning"; 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)
- [2] Wunsch, Georgios, RuiXu, "Multiclass Cancer Classification Using Semisupervised Ellipsoid ARTMAP and Particle Swarm Optimization with Gene Expression Data", Jan- March, 2007, IEEE/ACM Transactions on Computational Biology and Bioinformatics (Volume:4, Issue: 1)
- [3] K Kourou, Konstantinos P.Exarchos, Michalis V.Karamouzis, "Machine learning applications in cancer prognosis and prediction", 2015.
- [4] C. Suguna, S. P. Balamurugan; " Diagnosing Cervical cancer using machine learning methods"; 2022 International Congress On Human – computer interaction", Optimization and Robotic Applications (HORA)
- [5] Aman Sharma & Rinkle Rani, "A Systematic Review of Applications of Machine Learning in Cancer Prediction and Diagnosis", Survey Article, Published: 25th January, 2021
- [6] Nimmagadda Satyanarayana Murthy & Chaitanya Bethala, "Review paper on research direction towards cancer prediction and prognosis using machine learning and deep learning model", Original Research, Published: 29 March 2021
- [7] Dakhaz Mustafa Abdullah, Adnan Mohsin , Amira Bibo Sallow, "Lung cancer Prediction and Classification based on Correlation Selection method Using Machine Learning Techniques".
- [8] YawenXiao, JunWu, ZongliLin, XiaodongZhao , "A deep learning-based multi-model ensemble method for cancer prediction", 3 October 2017
- [9] Aman Sharma & Rinkle Rani, "A Systematic Review of Applications of Machine Learning in Cancer Prediction and Diagnosis", Survey article, Published: 25 January 2021.
- [10] Akshitha Shetty; Vrushika Shah, "Survey of Cervical Cancer Prediction Using Machine Learning: A Comparative Approach", 18 October 2018, 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)