iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52950

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Digitizing Notes using Optical Character Recognition and Automatic Topic Identification and Classification using Natural Language Processing

Soham Kulkarni[1], Rhushabh Madurwar[2], Rushikesh Narlawar[3], Anuj Pandya[4], Namrata Gawande[5]
[1, 2, 3, 4]*Student,* [5]*Asst. Prof. Department of Computer Engineering Pimpri Chinchwad College of Engineering Pune, India*

Abstract: ***In today's world digital documents are a major part of everyone's life as they have a wide scope of usage. However handwritten notes still contain loads of important and valuable information. In our research, we explore the different methods of Optical Character Recognition, or OCR which can be used for digitizing manual notes. Along with it we deep dive into the concept of Topic Detection and Identification and methods to implement it which are useful for extracting the crux of any document or piece of information. With the aim of integrating both processes into a single system, we study various algorithms involving neural networks like ANN, RNN, and CNN, and methods such as Tesseract, KNN, and LSTM that are used for implementing OCR while techniques such as K means clustering, TF-IDF, LDA and LINGO have been employed to perform topic detection and identification. Based on our study and results from various papers, we have decided to use CNN for OCR.***
Keywords: Optical Character Recognition, Natural Language Processing, Convolutional Neural Network, Support Vector Machine, Deep Learning.

I. INTRODUCTION

Handwritten notes and documents are a ubiquitous part of our world and have invaluable practical worth. Even though documents in digital format are being widely used and are being rapidly adopted in major applications and domains especially since the Covid-19 pandemic, still a large amount of information and data remains in the form of manual handwritten documents. Thus, extracting this information from these physical documents and identifying important parts from it is a very crucial job. This process can be practically performed using Machine Learning and various subdomains such as Image Processing, Natural Language Processing, etc which fall under it. Therefore for the conversion and extraction of these manually handwritten documents techniques like Optical Character Recognition (OCR), Topic Detection, and Topic Identification are being widely used presently.

Optical Character Recognition or OCR involves using technology or a model for the conversion of images in typed, handwritten, or printed text format to a digital format from formats such as a scanned document, image of the document, etc. It is mainly used to convert physical documents or hard copies of documents to soft copies where they can later be stored in a database or repository and can be edited using word processors. These converted documents can later be used for applications like feature recognition, pattern recognition, feature extraction, etc. Many approaches and techniques have been explored for the purpose of OCR and Topic Detection. Deep Learning concepts involving neural networks such as ANN[20], RNN[16], CNN[5], and Machine Learning algorithms such as KNN[21], SVM[22], etc have been employed for implementing OCR. Even though OCR is an important part of converting and extracting data from handwritten documents, topic detection and identification also form a very crucial part of the overall process. Natural Language Processing or NLP is the sub domain of Artificial Intelligence which combines computational power and linguistics to create systems that understand, extract and analyze meaning of text and human speech. It analyzes the semantics, syntax and pragmatic features of that text by implementing processes such as topic detection and identification.Along with this various other techniques and frameworks have been proposed for creating systems for specific applications based on either of the mentioned processes. The goal of this paper is to research the wide range of methods being used for the mentioned processes and present the state-of-the-art techniques and results achieved so far by reviewing the related work and drawing a comparison amongst them. The findings and outcomes obtained will further be used in research and implementation of a system which integrates the two procedures and provides the users of the application a single solution for digitizing handwritten notes or documents and classifying them according to their respectives topics.

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

In order to learn in detail about the domains and the previous research done related to various aspects of the field, several studies and research papers were referred to.

Chin-Yew Lin in 1995 [1] introduces “concept counting” for the automatic topic identification of text. Instead of simply counting words which can lead to unsatisfactory results, this paper proposes a system where a concept is identified using generalization taxonomy. A prototype system to test this proposed algorithm was implemented, resulting in a recall of 0.32 and a precision of 0.35. The author did not use linguistic text processing tools and hence the results could be further improved.

Magdi Mohamad et al. in 1997 [2] describe a novel approach in which the spacing and size difference of segmented characters is accounted for. The authors state that the spatial relationships between overlapping segmented images as well as inter-character relationships can be used to correctly match strings.

Shefali Arora et al. in 2018[3] highlighted how deep learning is used for the classification of images. In simple terms, the images of handwritten notes are classified and processed. Two methods that are feed-forward neural network and CNN are used for the purpose of feature extraction and training of models respectively. Further, it also stated that when it comes to the classification of handwritten notes the performance shown by the CNN is more than the feed-forward neural network.

Karez Abdulwahhab Hamad et al. in 2005 [4] analyzed optical character recognition and divided it into four parts. The first part of the analysis highlighted the challenges faced for the purpose of recognizing the quality of images and the fonts of the characters etc. The second part of the analysis highlighted the various layers like pre-processing, segmentation, normalization, feature extraction, classification and post-processing for the working of optical character recognition. In the third part, the analysis highlighted the development and various applications and uses of OCR. Finally the fourth part highlighted the history of how optical character recognition came into being used.

Sara Aqab et al. in 2020[5] states how the handwritten recognizing system has been developed. CNN is introduced in OCR which makes it more accurate just like humans visualize & recognize characters and symbols of different patterns and styles.

Hazen T.J. et al. in 2007 [6] differs between topic identification based on voice and topic identification based on text. The topic identification system was supervised learning-based, but it also made clear how important unsupervised methods are because unlabeled data will continue to accumulate over time.

Jamshed Memom et al. in 2020 [7] Describe the history and development of OCR (optical character recognition), which spanned eight decades. The technique was tested using many other languages and manuscript types, including English, Persian, Urdu, Hindi, and many older manuscripts, and the findings were emphasized. Additionally, it explained why RNNs, CNNs, and other machine learning techniques are gaining popularity. These techniques have evolved from earlier iterations of k-nearest neighbor, SVM, random forests, etc. to more complex, advanced neural networks that further boost performance and accuracy.

Burcu Caglar Gencosman et al. in 2014 [8] focused mainly on enhancing search engine results. The content-based and content-ignorant algorithms determine the search engine user behavior and accordingly cause the errors etc.

Srivastav, A. et al. in 2022 [9] surveyed and highlighted the concept of topic modeling and identification on news articles in English and Hindi languages with a focus on using it in applications like newspapers. The data is preprocessed and converted into a bigram or dictionary using the NLTK python library.

The topic identification and classification are done by computing cosine similarity and coherence between news topics in both English and Hindi using three techniques of LDA, HDP, and Doc2Vec.

Sang-Woon Kim et al. 2019 [10] proposed a system to classify research papers based on similar topics to help in finding papers related to each other when researching a particular subject. First data is preprocessed and processed data is converted to a dictionary formed based on high frequency and top-N keywords and a group of top 10,20,30 keywords is created. Then LDA is applied to this keyword dictionary to extract topics and perform topic modeling . Then TF-IDF technique is applied to the data to find the word frequency and to find the important keywords in the papers. Then finally K-means clustering is applied to form clusters of similar words and classify keywords of various papers based on similarity in the subject of the paper. Thus a system for effective research paper classification is proposed for faster and more efficient paper searching.

Guixian Xu et al. in 2019 [11] demonstrated how they handle subject tracking and article and document organization on a broad scale. The topics are first extracted using LDA, then topic modeling is then carried out. The Gibbs Sampling algorithm is then applied to topic modeling's parameter estimation. Following topic extraction, topic tracking is carried out by determining how closely related subjects are by utilizing time as a parameter.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

S. R. Vispute et al. in 2013[12] proposed a system for retrieving documents and providing personalized documents to end users with the help of their browsing history.
The paper provides a categorization of Marathi documents by using a clustering algorithm called Lingo clustering which is based on VSM. The system got an accuracy of 91.10 % on a dataset of 107 Marathi documents from three different categories.
Shunji Mori et al. in 1992[13] has given a brief about the research and development of OCR throughout history. The paper is divided into two parts: historical development and R&D, which are further divided into structural analysis and template matching. The paper has also provided their view on neural networks, expert systems, and the future scope of OCR.
Joris D'hondt et al. in 2011[14] proposed an innovative technique to divide a textual document into more components by using the coherence function, which is based on lexical chains and provides a coherence graph of documents as output. The proposed methodology in the paper has given the best results in randomized test scenarios and has outplayed other identification techniques.
Pema Gurung et al. in 2017[15] provided the usage of cluster analysis for document collection of various sizes. The paper has given a brief study of the K -means clustering algorithm for topic identification and provided a comparative study of the results of cluster analysis for small and large documents.
Bhagyashree P V et al. in 2019[19] used an advanced deep learning technique DAG-CNN(Directed Acyclic Graph Convolutional neural network) for handwritten character recognition. The given method overcomes some of the disadvantages of CNN, like misclassifying identical cursive words. As DAG is an acyclic-directed graph it has multiple inputs and outputs, and thus each and every layer is connected to the final layer with the help of skip connections. This allows various types of features for contributing to the overall performance.

III. TECHNIQUES

A. Neural Networks

A neural network is a technique in the field of Artificial Intelligence which involves training machines to process data inspired by the biological neuron structure in the human brain. It falls under the specialized domain of Deep Learning and uses interconnected nodes present in a layered structure to process the given input. It generally contains input, output, and multiple hidden layers. Some of the types of neural networks used primarily for the purpose of OCR include Artificial Neural Networks (ANN)[20], Convolutional Neural Networks (CNN)[18], Recurrent Neural Networks (RNN)[16], and similar subtypes. Fig.1 [29] below shows the structure of a neural network with majorly an input layer, output layer and multiple hidden layers.

B. K-Nearest Neighbor (KNN)

KNN or K-Nearest Neighbour is a powerful supervised machine learning algorithm used for classification and regression problems. It involves the classification of data points by forming clusters with 'k' surrounding neighbors and then doing a majority voting process in the respective cluster to determine how the data point can be classified. It has a variety of applications and is also used for implementing OCR.[21]. Fig 2 [28] below shows the clustering done by KNN algorithm.

C. Support Vector Machine (SVM)

Support Vector Machine or SVM is a supervised machine learning algorithm that is used for regression, classification, and outlier detection. The hyperplane is the optimum decision boundary that divides the data into different categories and classes. Each data point is plotted into an n-dimensional plane and classified into a particular topic. The goal is to find a hyperplane that has a maximum margin so that current, as well as future data points, can be classified efficiently. SVM also is popularly used for converting handwritten notes to digital notes using OCR.[22]. Fig. 3 [31]below shows classification done by SVM using hyperplane and support vectors.

D. K-means Clustering

K means is a popular unsupervised machine learning technique used for finding hidden similarities and inferences in unlabelled data. It involves creating clusters of similar data around computed centroids based on the euclidean distance of all points from the respective centroids when the data is plotted.
It executes multiple iterations with the data points to form the most efficient clusters involving maximum data points with similar characteristics and nature. It is extensively used for the purpose of topic detection in various applications.[15]. Fig. 4 below shows clustering done by K-means clustering algorithm.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

E. TF-IDF

Term Frequency measures the frequency of occurrence or count of a particular word in a document. The length and generality of the word affect the result so the term frequency has to be normalized. Each document is vectorized on vocabulary to create a generalized vector for any possible word in the corpus.

$$TF = \frac{count\ of\ term\ in\ document\ (t)}{total\ number\ of\ terms\ in\ document\ (d)}$$

Inverse Document Frequency measures the importance of a word in the document which TF does not consider. It provides weightage to each word based on its frequency and while TF is document-specific, IDF is constant for a corpus.

$$IDF = \frac{Total\ number\ of\ documents\ in\ corpus\ D}{number\ of\ documents\ containing\ term\ t}$$

Both these methods are very popular and extensively used for the purpose of topic detection under the domain of Natural Language Processing (NLP).

F. Latent Dirichlet Allocation (LDA)

LDA is a topic modeling technique that is used to identify abstract topics from a collection of documents.It involves the unsupervised classification of the documents to form groups of similar topics in the data naturally. It associates words of different documents with the assumption that topics similar to each other will use words that are similar and associated with one another. This helps in mapping a probability distribution for topics which facilitates topic detection and identification process.[10] . Fig. 5 [30] below shows an overview of the functioning of the LDA algorithm.

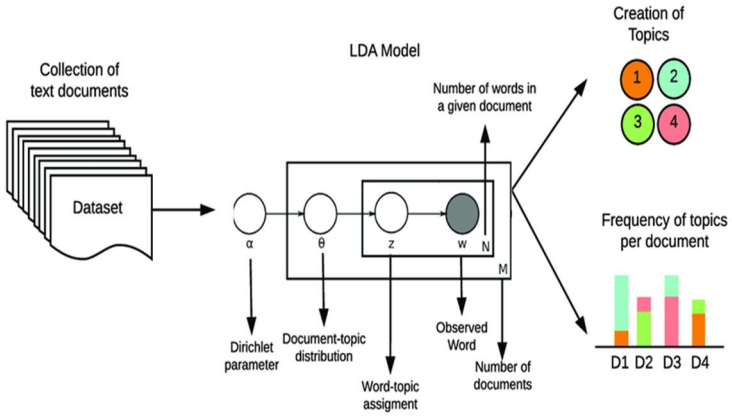


Fig.1. Latent Dirichlet Allocation

IV. METHODOLOGY

OCR is generally done in two major steps -

1) *Text detection:* Detecting the position of the text i.e. the words and letters on a page and drawing bounding boxes around it. It could be a very densely populated document or a sparsely worded document. After detection, the next step is to identify the word.
2) *Recognition:* There are three approaches that can be taken here:

a) Computer Vision Techniques

This involves using a lot of image transformations such as contour detection, gaussian blur, filters, and then finally image classification on that. It is an Open-CV heavy task and requires a lot of fine-tuning. The main problem here is that it is very difficult to generalize. Every type of document is different with respect to the lighting, the tilt of the picture, the clarity, noise, size, etc. Hence what works for a specific type of document won't work for others. Hence these methods are generally not preferred.

b) Standard Deep Learning

These are the popular approaches such as SSD, YOLO, and Mask RCNN which are frequently applied. In this, we apply a general deep learning model to all kinds of documents and get a robust solution. This approach is simple and easy to use.

c) Specialized Deep Learning

These are specially curated methods that are highly accurate and now dominant in text identification such as EAST, CRNN, and STN.

We will be using these specialized Deep Learning methods to curate our system to recognize handwritten notes.

Our system, as shown in Fig.6, consists of two modules, OCR and NLP.

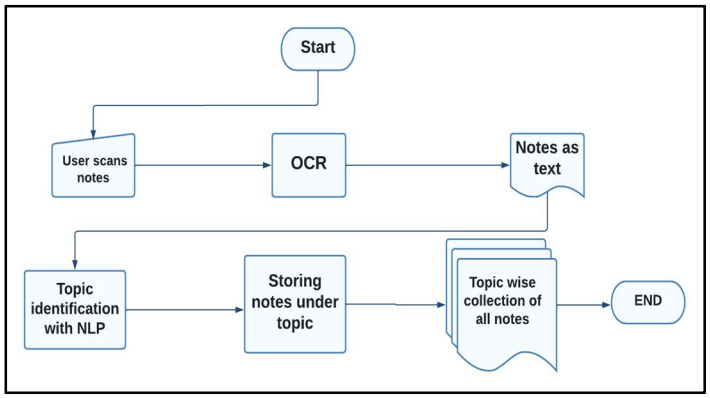


Fig 2. Block Diagram of the proposed system

- *Image Acquisition:* The system initially involves scanning handwritten or manual documents or notes using a mobile phone's camera which is assumed to be of decent quality.
- *Text Detection:* A significant step in the pipeline which is used to determine if text is present in the given image or not. And if present, its coordinates are remembered. This is done using text localization and verification. Usually, bounding boxes are added to the regions where the text is identified.
- *Transformation:* This is an optional step used to extract and clean the detected text so as to provide the deep learning model with quality inputs. It handles all kinds of distortion in texts, such as removing tilt, and aligning it horizontally, etc.
- Text Recognition: This is where a neural network is applied to actually recognize the text, and convert it into digital form.
- *Final Text:* The final text obtained usually isn't 100% accurate and perfect, and hence NLP can be used to fix mistakes and misidentifications. For example, the word "dictionary" may be recognized as "dictlonary" or "dlctlonany" where the letter "i" is confused with the letter "l" due to vast differences in handwriting. Such words can be rectified using NLP models.

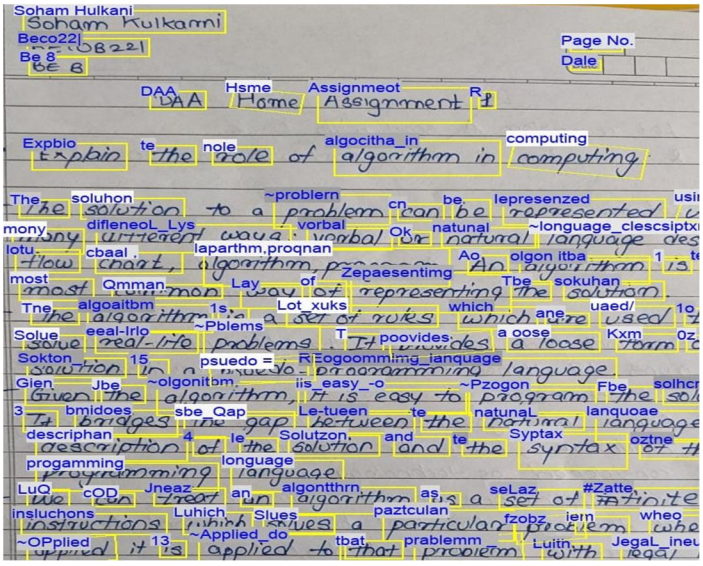


Fig 3. Example of OCR

After the conversion, topic detection and identification are done by employing Natural Language Processing or NLP techniques such as Linear Discriminant Analysis (LDA), Term Frequency - Inverse Term Frequency (TF-IDF), or similar methods.
Topic identification can be done using two techniques- topic modeling and topic classification. Topic modeling uses unsupervised machine learning to group together documents with similar words, keeping in mind their relations.
Topic classification on the other hand uses supervised machine learning to identify what topic a document belongs to on the basis of the previous training documents provided to it. It is classified into three types: Rule-based system, Machine Learning system, and Hybrid system.
Hence using topic classification, the notes are classified under their respective topics. The user through the app or platform will be able to access and store topic-wise classified collections of notes for various subjects and respective relevant topics.

V. RESULTS

From the survey conducted the performance of different algorithms is compared in the given table:

Table 1. Analysis of Results

Sr. No.	Topic	References	Algorithms/Technique	Accuracy
1.	OCR	R.Parthiban, et al., 2020[16]	RNN	90%
2.	OCR	Chirag Patel, et al., 2012[17]	Tesseract Transym	70% 47%
3.	OCR	Usha Tiwari, et al., 2019[20]	ANN	90.5%
4.	OCR	Gaurav Y. Tawde, et al., 2013[21]	KNN	93.8%
5.	OCR	Sara Aqab, et al., 2020 [5]	CNN	83.4%
6.	OCR	Nasien, Dewi, et al., 2010 [22]	SVM	88%
7.	OCR	Arora, Shefali, et al., 2018[3]	FFNN CNN	90% 95.63%
8.	OCR	Srivastava, S. et al. 2019[18]	CNN	95.71%
9.	NLP	Srivastav, A., et al. ,2022[9]	LDA Doc2Vec	93.21% 67.55%
10.	NLP	Burcu Caglor Gencosman., et al., 2014[8]	Character n-gram method	98.59% 92.96%
11.	Topic Identification	Joris D'hondt, et al., 2011[14]	Square Segment Identification (SSI) algorithm	74.5% 65.5%
12.	Topic Detection	Sonal Jain, et al., 2010[27]	Ontological Approach	82.3%

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

VI. CONCLUSION AND FUTURE SCOPE

This paper provides a thorough survey of a wide range of techniques and algorithms which have been employed for the purpose of OCR and Topic detection respectively. Although OCR is not a relatively new field of research, it has a lot of scope and room to be improved upon. We observe that using deep learning methods involving neural networks and their variants provides greater accuracy than some other ML techniques. We also dive deep into the concept of Topic Detection and Identification and methods to implement it which are useful for extracting the crux of any document or piece of information. Various NLP techniques and ML models are explored which perform topic detection and identification for specific applications.

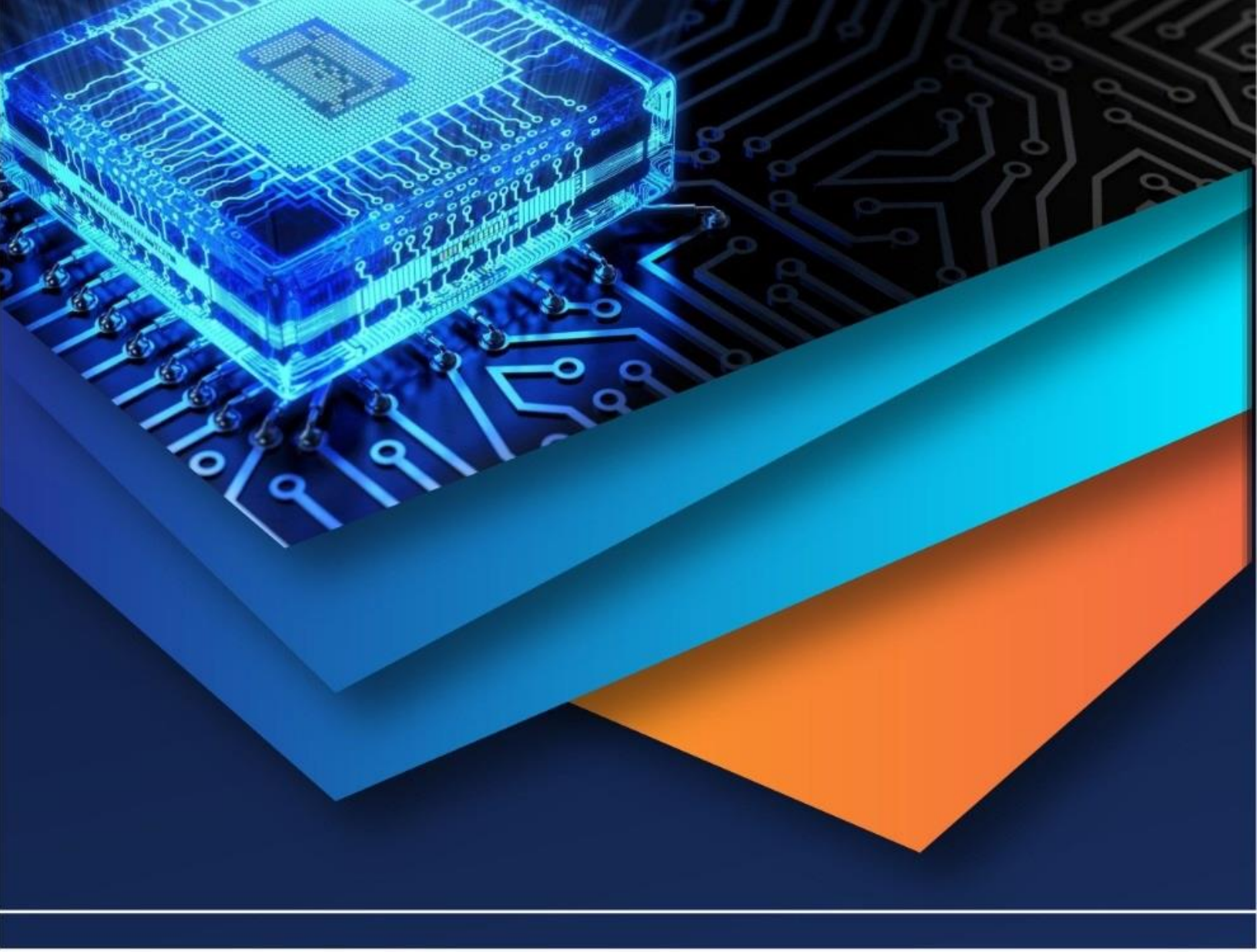
The objective is to propose a system that integrates both processes into a single system. The aim is to use a deep learning neural network-based approach for first digitizing documents using OCR and then apply NLP techniques to implement topic detection and identification on the digitized data. The system can then be made available for general usage through an app or a platform that provides these functionalities. Such a system can prove to be extremely useful in places such as educational institutes for sharing important study material and documents. A future application of this system may be used for automatic answer paper checking as well but such an application will require good accuracy and minimal error to be applicable for practical use. Such use cases can thus help in expanding and developing the educational sector with extended applications in other sectors as well.

REFERENCES

[1] Chin-Yew Lin. 1995. Knowledge-based Automatic Topic Identification. In the 33rd Annual Meeting of the Association for Computational Linguistics, pages 308–310, Cambridge, Massachusetts, USA. Association for Computational Linguistics.
[2] Gader, Paul & Mohamed, Magdi & Chiang, Jung Hsien. (1997). Handwritten word recognition with character and inter-character neural networks. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on. 27. 158 - 164. 10.1109/3477.552199.
[3] Arora, Shefali & Bhatia, M.. (2018). Handwriting recognition using Deep Learning in Keras. 142-145. 10.1109/ICACCCN.2018.8748540.
[4] Hamad, Karez & Kaya, Mehmet. (2016). A Detailed Analysis of Optical Character Recognition Technology. International Journal of Applied Mathematics, Electronics and Computers. 4. 244-244.
[5] Sara Aqab and Muhammad Usman Tariq, “Handwriting Recognition using Artificial Intelligence Neural Network and Image Processing” International Journal of Advanced Computer Science and Applications(IJACSA), 11(7), 2020.
[6] Hazen, T.J. (2011). Topic Identification. In Spoken Language Understanding (eds G. Tur and R. De Mori). https://doi.org/10.1002/9781119992691.ch12
[7] J. Memon, M. Sami, R. A. Khan and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," in IEEE Access, vol. 8, pp. 142642-142668, 2020, doi: 10.1109/ACCESS.2020.3012542.
[8] Burcu Caglor Gencosman, Huseyin C. Ozmutlu, Seda Özmutlu. Character n-gram application for automatic new topic identification. ELSEVIER 26 June 2014.
[9] Srivastav, A., Singh, S. Proposed Model for Context Topic Identification of English and Hindi News Article Through LDA Approach with NLP Technique. J. Inst. Eng. India Ser. B 103, 591–597 (2022).
[10] Kim, SW., Gil, JM. Research paper classification systems based on TF-IDF and LDA schemes. Hum. Cent. Comput. Inf. Sci. 9, 30 (2019). https://doi.org/10.1186/s13673-019-0192-7
[11] G. Xu, Y. Meng, Z. Chen, X. Qiu, C. Wang and H. Yao, "Research on Topic Detection and Tracking for Online News Texts," in IEEE Access, vol. 7, pp. 58407-58418, 2019, doi: 10.1109/ACCESS.2019.2914097.
[12] S. R. Vispute and M. A. Potey, "Automatic text categorization of marathi documents using clustering technique," 2013 15th International Conference on Advanced Computing Technologies (ICACT), 2013, pp. 1-5, doi: 10.1109/ICACT.2013.6710543.
[13] S. Mori, C. Y. Suen and K. Yamamoto, "Historical review of OCR research and development," in Proceedings of the IEEE, vol. 80, no. 7, pp. 1029-1058, July 1992, doi: 10.1109/5.156468.
[14] Joris D’hondt, Paul-Armand Verhaegen, Joris Vertommen, Dirk Cattrysse, Joost R. Duflou, Topic identification based on document coherence and spectral analysis, Information Sciences, Volume 181, Issue 18, 2011,https://doi.org/10.1016/j.ins.2011.04.044.
[15] Gurung, Pema,Wagh, Rupali 2017/03/25 A study on Topic Identification using K means clustering algorithm: Big vs. Small Documents Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 2
[16] R. Parthiban, R. Ezhilarasi and D. Saravanan, "Optical Character Recognition for English Handwritten Text Using Recurrent Neural Network," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1-5, doi: 10.1109/ICSCAN49426.2020.9262379.
[17] Patel, Chirag & Patel, Atul & Patel, Dharmendra. (2012). Optical Character Recognition by Open source OCR Tool Tesseract: A Case Study. International Journal of Computer Applications. 55. 50-56. 10.5120/8794-2784.
[18] Mittal, Usha ,Srivastava, Sonal,Chawla, Priyanka 2019,Object Detection and Classification from Thermal Images Using Region based Convolutional Neural Network .Journal of Computer Science Doi - 10.3844/jcssp.2019.961.971
[19] Bhagyashree P V, Ajay James, Chandran Sarvanan, A Proposed Framework for Recognition of Handwritten Cursive English Characters using DAG-CNN. IEEE Doi: 10.1109/ICIICT1.2019.8741412
[20] Tiwari, Usha & Jain, Monika & Mehfuz, Shabana. (2019). Handwritten Character Recognition—An Analysis. 10.1007/978-981-13-0665-5_18.
[21] Tawde, Gaurav Y., Mrs. Jayashree M. Kundargi and Jayashree M. Kundargi. “An Overview of Feature Extraction Techniques in OCR for Indian Scripts Focused on Offline Handwriting.” (2013).
[22] Nasien, Dewi & Haron, Habibollah & Yuhaniz, Siti. (2010). Support Vector Machine (SVM) for English Handwritten Character Recognition. Computer Engineering and Applications, International Conference on. 1. 249-252. 10.1109/ICCEA.2010.56.
[23] Magdi Mohamed and Paul Gader. Handwritten Word Recognition Using Segmentation-Free Hidden Markov Modelling and Segmentation-Based Dynamic Programming Techniques (1996) in IEEE. Doi: 10.1109/34.494644

[24] Aisha Sharaf, Bhagya Viswanath, Kavya Chandran, Nishana Salim, Anju S Oommen. Handwritten Text Recognition and Digitization System. IJIRSET 2019
[25] Manoj Sonkusare and Narendra Sahu. A Survey On Handwritten Character Recognition(HCR) Techniques For English Alphabets. Advances in Vision Computing: An International Journal (AVC) Vol.3, No.1, March 2016. DOI:10.5121/avc.2016.3101
[26] K.Karthick, K.B.Ravindrakumar, R.Francis, S.Ilankannan. Steps Involved in Text Recognition and Recent Research in OCR; A Study. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1, May 2019
[27] S. Jain and J. Pareek, "Automatic Topic(s) Identification from Learning Material: An Ontological Approach," 2010 Second International Conference on Computer Engineering and Applications, 2010, pp. 358-362, doi: 10.1109/ICCEA.2010.221.
[28] Xing, W., & Du, D. (2019). Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention. Journal of Educational Computing Research.
[29] Baek J, Choi Y. Deep Neural Network for Predicting Ore Production by Truck-Haulage Systems in Open-Pit Mines. Applied Sciences. 2020
[30] Buenaño-Fernández, Diego & Gonzalez, Mario & Gil, David & Luján-Mora, Sergio. (2020). Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2974983.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)