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.2023.53153

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

Form Based Document Understanding Using Sequential Model

Khushi Rathi[1], Aniket Patil[2], Nikita Mahadik[3], Sakshi Ahire[4], Abhilasha Shinde[5]
Students, Department of Information Technology Engineering
Asst. Professor, Department of Information Technology
Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: ***DocFormer is a teachable transformer-based multi-modular start-to-finish model for a number Visual Report Figuring out projects. Additionally, Doc Former allows the model's capacity to hyperlink textual content and visible tokens through sharing realized spatial throughout modalities. For the learn about of quite a number algorithms and pc vision, it is a critical location of research. Some humans fabricate archives and have interaction in unlawful things to do due to the fact the cutting edge report detection gadget is ineffective. In the proposed system, pretend files can be discovered in two ways. As the recognition of deep mastering science has grown, file AI, which consists of matters like record design analysis, visible records extraction, record visible query answering, file photo classification, and so on, has made large growth in current years. We are principally focusing on the elements that contributed to crimes in preceding years when it comes to the elements that make a contribution to the prevalence of crimes, such as the identification of the culprit and different factors.***

Keywords: CrimeAnalyze, OCR, Feature Extraction, SVM, Accuracy.

I. INTRODUCTION

Form files frequently have extra complicated layouts with structured objects such as tables, columns, and textual content blocks. Crimes are social nuisances and have a direct affect on society. Governments spend a lot of cash to stop crime thru regulation enforcement. Today, many regulation enforcement organizations have massive quantities of crime-related facts that need to be processed in order to be modified into beneficial information. The proposed machine can discover whether or not a record is authentic and efficient. The photograph processing aggregate works very effectively and the effects received are accurate. The machine is skilled on counterfeit archives and how to forestall them the usage of a phrase processor. Machine studying functions examine from enter statistics and use automatic optimization methods to consistently enhance output accuracy.

Sequence Modelling is the potential of a pc application to model, interpret, make predictions about or generate any kind of sequential data, such as audio, textual content etc.

It is inappropriate to use a sequential mannequin when:

1) Any of your layers has a couple of inputs or more than one outputs
2) You want to do layer sharing
3) Your mannequin has more than one inputs or outputs
4) You prefer non-linear topology (e.g. a residual connection, a multi-branch model)

Deep learning, which is a subset of computing device learning, is truly a neural community with three or extra layers. Although they are a ways from matching the abilities of the human brain, these neural networks provide it the potential to "learn" from a massive quantity of statistics and make an effort to imitate its behaviour. Although greater hidden layers can assist enhance accuracy, a neural community with simply one layer can nonetheless make difficult predictions. Deep studying makes it feasible for several artificial brain (AI) functions and offerings to enhance automation with the aid of carrying out bodily and analytical duties besides the want for human intervention. Deep getting to know science is in the back of every day merchandise and offerings like digital assistants, voice-activated TV remotes, and credit score card fraud detection.

There was once actually want of constructing such machine will assist to limit the work at technician who does records entry eg. in banking sector, at university admission system etc.

©IJRASET: All Rights are Reserved

Along with that this device is going to become aware of the crook file as properly as the furnished record is pretend or not there used to be no any preceding machine which used to be imparting distinct offerings in a system. Introducing such machine will decrease time at any kind of statistics entry which are associated to record or form. As properly as by way of detection of crook report of candidate technician can without difficulty debar or take delivery of the structure or record of candidate.

II. LITERATURE SURVEY

Table 2.1: Literature Survey

Sr No.	Title of The Paper	Method	Metrics	Advantages	Disadvantages	Uses
1.	Chen-Yu Lee, Chun Liang Li, Timothy Dozat, Vincent Perot, Guolong Su, Nan Hua, Joshua Ainslie, Renshen Wang, Yasuhisa Fujii, Tomas Pfister. Form Net: Structural Encoding beyond Sequential in form document Information Extraction-2022	The approach with a smaller model size and fewer training data before	FormNet-A3 the maximum 97.28% F1 rating possible	With a smaller model size and fewer pretraining data, Form Net outperforms current techniques.	We suggest Form Net as a structured-aware sequence model to help forms be serialised more effectively.	The console dated receipt dataset for post OCR parsing
2.	Mrs G Chandra Prabha, E. Jeevitha, B. Shwetha . Fake Education Document detection using Image Processing and Deep learning – 2021	It is suggested to use deep learning to identify ink mismatches in hyperspectral document photos.	With 98.88% accuracy and 98.93% F1 score	Advantages of using social media and analysis of various techniques that could help in detecting fake news	The inaccuracy of these CT- BERT and ROBERT models' predictions is addressed by this method.	The used during college admissions are done using scan copies from other genuine resource
3.	Lei Cui, Yiheng Xu, Tengchao Lv, Furu Wei. DOCUMENT AI: BENCHMARKS, MODELS AND APPLICATIONS - 2021	a page segmentation method of handwritten historical document images based on CNN	The classification accuracy of 95% is obtained in the Arabic dataset	Document AI is a very challenging task and has attracted widespread attention in related research areas.	Traditional rule-based methods often require large labour costs, and these manually summarized rules are not scalable	Document AI uses advance d AI technologies such as computer vision, NLP, as well as deep learning models
4.	Srikar Appalaraju , Bhavan Jasani , Bhargava Urala Kota. Doc Former: End-to-End Transformer for Document Understanding -2021	Various combinations of spatial, text, and picture attributes have been employed in approaches in the literature to comprehend and extract information.	Doc Former achieves state ofthe-art performance of 96.17%	VDU is a difficult problem that attempts to comprehend documents in their many layouts and formats (forms, receipts, etc.).	Additionally, DocFormer distributes spatial embeddings learnt across modalities, making it simple for the model to connect text to visual tokens and vice versa.	Text alone, or even just text plus spatial data, is insufficient for this goal..
5.	M.Naresh Babu ; S.N.Vyshnavi , S.Keerthi , P. Divya Sree , D. S. V. Naga Prasad . Crime type and occurrence prediction using Machine learning algorithm – 2022.	This technique aids in determining the association between two numerical numbers or variables.	The accuracy of this grouping is 90%.	The initialization of optimal value is not required.	In order to categorise different criminal patterns, certain machine learning methods, such as Naive Bayes, are used in this work. When compared to precomposed works, the accuracy attained was rather high.	A non-parametric technique for classification and regression is the k closest neighbour's algorithm (kNN).

6.	B. Sivanagaleela, S. Rajesh . Crime Analysis and Prediction Using Fuzzy C-Means Algorithm - 2019	The use of an analytical method is to classifying the crime data based on the crimes using these data.	The used to identify the accuracy of the crime and performance of the crime speed to control the rate.	provide better assistance to avert and take action against crime prior to its occurrence with the accumulated data.	predictive policing is that it can produce biased results	The data can be collected from the Data World website it maintains the district wise crime data and also consists of various crimes such as Kidnapping, Murder, Theft, Robbery etc.
7.	Vijayshree B. Nipane, Poonam S. Kalinge, Dipali Vidhate, Kunal War, Bhagyashree P. Deshpande. Fraudulent Detection in Credit Card System Using SVM & Decision Tree – 2016	A support vector machine is a type of method utilized in classification and pattern recognition.	The rate of accuracy has reached to 59%.	Support Vector Machine (SVM) & decision tree artificial intelligence concepts are being applied to solve the challenge in the contemplate system for fraud detection.	Credit card fraud is a major source of financial losses in this paper's current scenario; it affects both businesspeople and individual clients.	For classification, regression, and various other learning tasks, (SVMs) are a well-liked machine learning technique.

III. METHODOLOGY

A. Existing Methodology

The problem formulation presented in the given text involves sequential tagging of tokenized words in a form document, where each token is assigned a key entity class using the "BIOES"[1] scheme. The proposed approach is to use a long-sequence transformer extension called ETC as the backbone for the sequence model, which can handle potentially long documents. To address the difficulty of recovering from serialization errors that occur when an entity sequence crosses multiple spans of a form document, the authors propose two novel components: Rich Attention and Super-Tokens[1]. Rich Attention captures both the semantic relationship and spatial distance between every pair of tokens, while Super-Tokens model local relationships between pairs of tokens that might not be visible to each other or correctly inferred in an ETC model after suboptimal serialization.

B. System Architecture

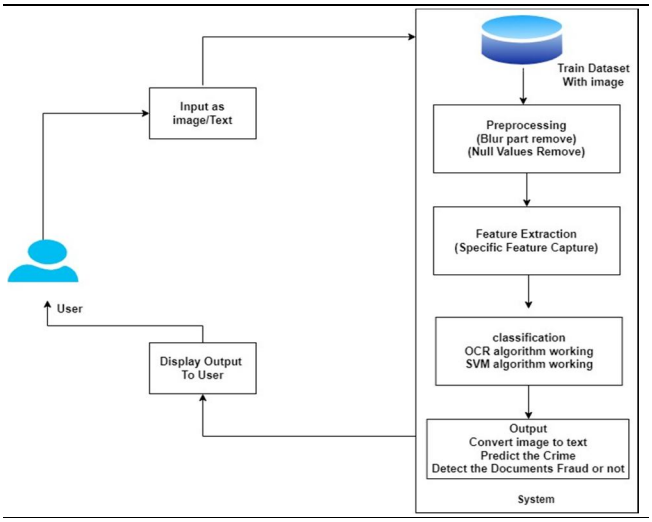


Figure 3.1 : System Architecture

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

C. Image Extraction

The first module basically comprises of Image Extraction using OCR[1]. OCR, or Optical Character.

Recognition, is a technology used to convert images containing printed or handwritten text into machine readable text. After an image file has been processed through an online OCR service, the extracted text can be edited using word processing software such as Microsoft Word. Additionally, OCR technology[1] can also use machine learning algorithms to improve its accuracy over time. Machine learning algorithms use pattern recognition techniques to identify and analyze patterns within data, and they can learn to recognize new patterns and improve their accuracy as they encounter more data. OCR software can also be trained to recognize specific fonts or handwriting styles, making it more accurate and efficient at recognizing specific types of text. However, it is important to note that OCR technology may still make mistakes, particularly with poor quality scans or illegible handwriting. Therefore, it is always important to review and verify the accuracy of OCR-generated data before using it for critical business applications.

1) OCR Process

The process of how OCR works may vary slightly among OCR software applications, but it follows a few general steps.

First, the scanner reads a physical paper document and creates a scanned image, typically in black and white. The OCR engine then applies pre-processing techniques like de-skewing, binarization, zoning, and normalization to correct errors and improve the accuracy of the scanned image. Next, AI tools like pattern matching and feature extraction are used to identify original characters from the scanned image or document. Finally, the OCR software converts the extracted data into electronic documents, and advanced OCR systems may compare the data against a glossary or library of characters to ensure maximum accuracy. It is important to note that OCR technology may still make errors, especially with poor quality scans or illegible handwriting, so it is crucial to verify the accuracy of OCR generated data before using it for important business applications. In this module after extraction of words on the pan card some information is important like Pan Number, Date of birth, Name and Father's name . That information is mapped and stored in a CSV file so we can have unstructured data in a structured format.

2) OCR Algorithm[1]

```
# Input: An image containing text
# Output: The recognized text from the image
 # Step 1: Image Preprocessing
Image = load_image("image.jpg")
grayscale_image = convert_to_grayscale(image)
thresholded_image=apply_threshold(grayscale_image)

# Step 2: Text Detection
regions = detect_text_regions(thresholded_image)
text_boxes = extract_text_boxes(regions)

# Step 3: Character Segmentation
characters = []
for each text_box in text_boxes:
character_images= segment_characters(text_box)
characters.append(character_images)
# Step 4: Character Recognition
recognized_text = ""
for each character_image in characters:
recognized_char=recognize_character(character_image)
recognized_text += recognized_char
```

©IJRASET: All Rights are Reserved

```
# Step 5: Output
print("Recognized Text: " + recognized_text)
```

D. Authentication

It is the process of verifying and confirming the authenticity and validity of a document, ensuring that it is genuine, credible, and reliable.

1) Easy OCR

Easy OCR is popular open-source OCR library that supports multiple languages, including English, Marathi ,Hindi etc.

a) Algorithm used for extraction in EasyOCR

EasyOCR is a deep learning-based OCR library that uses Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for extracting text from images and performing OCR. The library is built on top of the PyTorch deep learning framework and uses a combination of pre-trained models and fine-tuning techniques to achieve stateof-the-art performance on various OCR tasks.

The following are the main algorithms used by EasyOCR for extracting words and information about them:

- Object Detection: EasyOCR uses a pre-trained object detection model to locate the text regions in the input image. The object detection model is based on the Single Shot Detector (SSD) architecture and is trained on a large dataset of images containing text regions.
- Text Recognition: Once the text regions are detected, EasyOCR uses a deep learning-based text recognition model to extract the text from each region. The text recognition model is based on a combination of CNNs and RNNs and is trained on a large dataset of images containing text.
- Language Identification: EasyOCR can automatically identify the language of the extracted text using a language identification model. The language identification model is based on a
- character-level CNN and is trained on a large dataset of text in various languages.
- Confidence Estimation: EasyOCR computes a confidence score for each extracted word, which represents the probability that the OCR engine correctly recognized the word. The confidence score is based on the output probabilities of the text recognition model and is computed using a beam search algorithm.

Overall, EasyOCR combines several deep learningbased algorithms to perform OCR and extract text from images with high accuracy and efficiency.

b) Why parallel processing is require in EasyOCR

EasyOCR is a deep learning-based OCR library that uses Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for text recognition. Training and running these deep learning models can be computationally intensive and require a significant amount of processing power.

GPU (Graphics Processing Unit) acceleration can significantly speed up the training and inference process for deep learning models. GPUs are designed to handle large amounts of parallel computations, making them well-suited for training and running neural networks.

In contrast, CPUs (Central Processing Units) are better suited for sequential computations.

By using GPUs, EasyOCR can perform text recognition tasks much faster than if it were running on a CPU. This is particularly important for realtime applications where speed is critical, such as in mobile devices, cameras, or other OCR systems that require fast processing.

In addition, GPUs can handle larger batch sizes, which can further speed up the training and inference process. This is because GPUs can process multiple images or data points in parallel, allowing the neural network to be trained or evaluated more efficiently.

Overall, using GPUs in EasyOCR can improve its performance and make it more suitable for real-time and high-volume OCR applications.

c) Implementation

After extraction of all the words we authenticate the particular words are present on the document or not if the required words are missing or any wrong information is edited on the document then we are concluding the document is fake using EasyOCR.

E. Criminal Analysis

The third module will find the criminal record of a person and if criminal record found then it will show list of crimes done by the person. For that purpose, we had created a dummy database which consist of Name of the person , Pan No , Date of Birth , Criminal ID , type of crime and location .

This data is stored in a SQLite database, which was then queried and from that query we are analysing that the person is criminal or not.

IV. RESULT

Below image shows the home page where there are three options: Img to text extraction , Criminal analysis , fake document detection.



Figure 4.1 : Home Page

By clicking on first option i.e. img to text extraction will display the page as shown in below image.

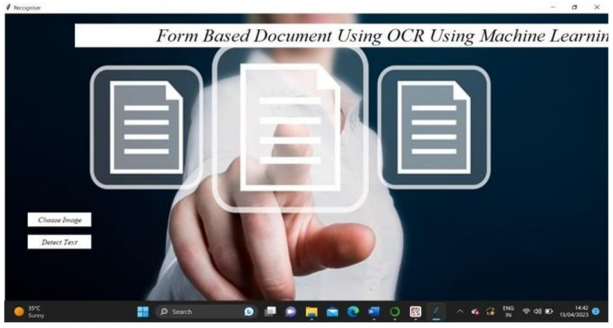


Figure 4.2 : Text Extraction Page

By clicking on choose image button , we can choose image from directory and that image is visible on the screen as shown in below image.



Figure 4.3 : Input to Text Extraction

To extract the relevant information from that selected image click on detect text option .



Figure 4.4 : Extracted Text

By clicking on Fake document detection button on home page, Dialogue box will appear displaying whether the document is fake or real.



Figure 4.5 : Authentication

And lastly by clicking on Criminal analysis button will open a page which display name of the person and whether the person has criminal record or not.



Figure 4.6 : Criminal Detection

If person has some criminal record then there is a option at the bottom of the page to show the list of crime.



Figure 4.7 : Criminal record Result

V. CONCLUSION

In this paper, we existing Doc-Former, a multimodal, always trainable, transformer- based totally model for a range of duties of visible report comprehension. A systematic strategy to figuring out crime is crime evaluation and prediction. The proposed gadget can recognize whether or not a report is true and efficient. Whether the device is skilled on cast documents.

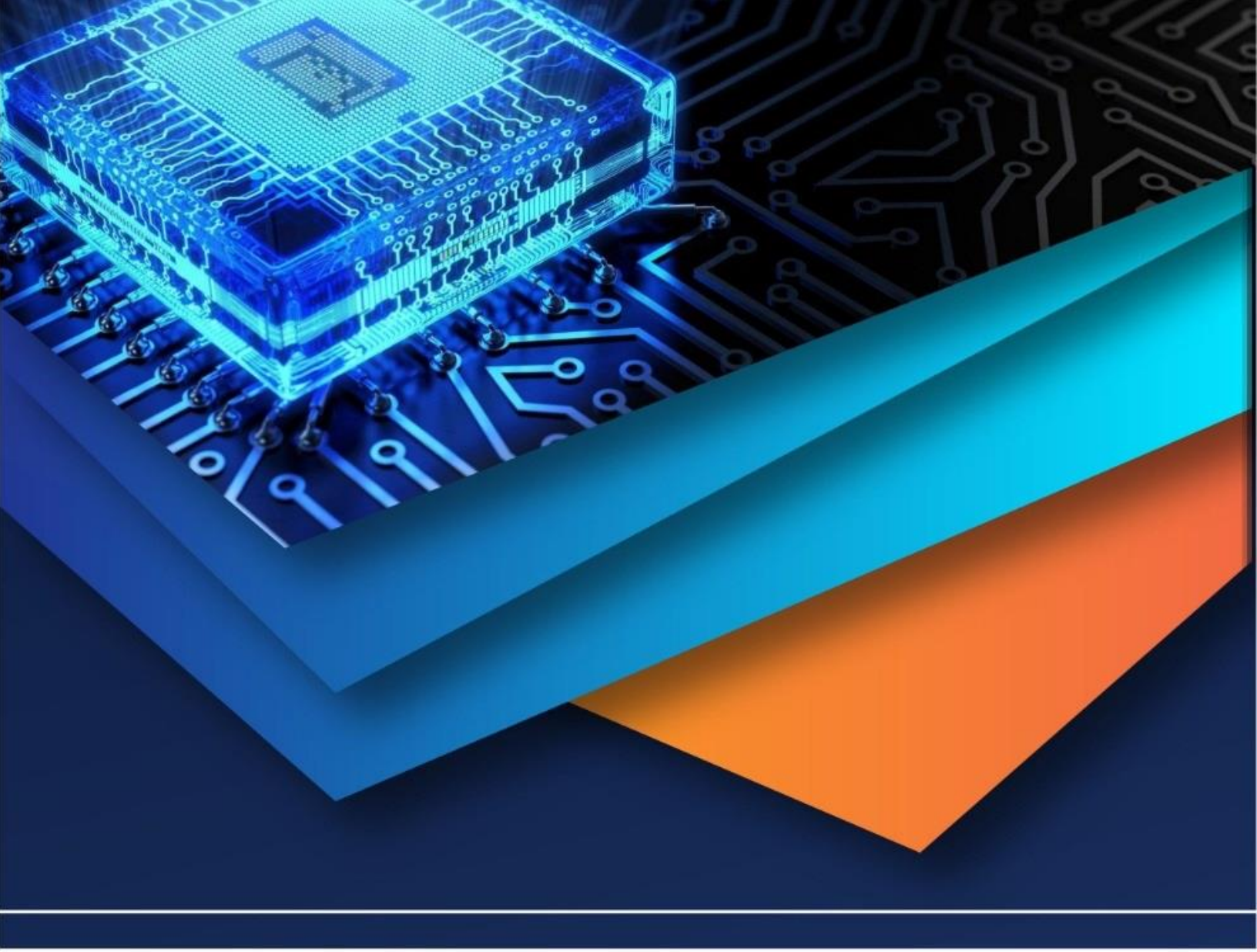
This task has big scope. It is viable to construct such gadget for specific enterprise in accordance to their required fields. Also, it is viable to fill statistics which we get from file at once into the field. We can additionally furnish some points like accepting a couple of language will gives flexibility to customers as properly as technician to extract records from record which will be in quite a number languages and constructing such gadget for special company will decrease time of statistics entry.

VI. ACKNOWLEDGEMENT

We are very thankful to all the teachers who have provided us valuable guidance towards thecompletion of this project work on “Form based document understanding using sequential model”.We express our sincere gratitude towards cooperative department who has provided us with valuable assistance and requirements for the project work .We are very grateful and want to express our thanks to Prof. Abhilasha Shinde for guiding us inright manner, correcting our doubts by giving her time whenever we required, and providing her knowledge and experience in making this project work. We are also thankful to the HOD of our Information Technology department Dr. M. L. Bangare for his moral support and motivation which has encouraged us in making this project work. We are also thankful to our Principal Prof. Dr. A.V.Deshpande, who provided his constant support and motivation that made a significant contribution to the success of this project.

REFERENCES

[1] Chen-Yu Lee, Chun- Liang Li, Timothy Dozat, Vincent Perot, Guolong Su, Nan Hua, Joshua Ainslie, Renshen Wang, Yasuhisa Fujii, Tomas Pfister. FormNet: Structural Encoding beyond Sequential in form document Information Extraction-2022
[2] Mrs G Chandra Prabha, E. Jeevitha, B. Shwetha . Image processing and deep learning-based document detection for fake education – 2021
[3] Lei Cui, Yiheng Xu, Tengchao Lv, Furu Wei. DOCUMENT AI: BENCHMARKS, MODELS AND APPLICATIONS - 2021
[4] Srikar Appalaraju , Bhavan Jasani , Bhargava Urala Kota. DocFormer: End-to-End Transformer for Document Understanding -2021
[5] M.Naresh Babu ; S.N.Vyshnavi , S.Keerthi , P. Divya Sree , D. S. V. Naga Prasad .Crime type and occurence prediction using Machine learning algorithm – 2022.
[6] B. Sivanagaleela, S. Rajesh . Fuzzy C-Means Algorithm for Crime Prediction and Analysis in 2019
[7] Vijayshree B. Nipane, Poonam S. Kalinge, Dipali Vidhate, Kunal War, Bhagyashree P. Despande , Fraudulent Detection in Credit Card System Using SVM and Decision Tree - 2016.



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)