

# Code Blooded 3.0

**Text Extraction**

**TEAM NAME – METALLIANS**



**Team Members :-**

Hemant Kumar Yadav  
Dhanjay Mahtha  
Shivam Goyal  
Nitin Kumar Mishra  
Ishaan Jain  
Ashutosh Singh

# Problem Statement

Create an Application that can do below

Upload a document using a UI based WebApp

Have a backend engine that will accept/validate and process the document

Extract contents of the document after processing

Display the list of documents in UI along with status

Have a side by side UI to display both document and extracted content of the document in a Form.

# Website Features

**Text from Image**

**Copy Text**

**Support for  
multiple  
Languages**

Image to PDF

# Different Languages Supported

## Steps to Scan Text

- 1.Upload an Image.
- 2.Select the Language.
- 3.Click on Scan.

## Metallians OCR

Scan text from images

Upload Image

English ▼

Scan

# Metallians

Scan text from images

Upload Image

English

हिन्दी निबन्ध, पत्र-लेखन एवं कहानियाँ

## मेरी आत्मकथा

(मैं एक लड़का)

मेरा नाम तरुण वर्मा है। वर्मा मेरा जातिगत नाम है। मैं 10 वर्ष का हूँ और 5वीं कक्षा में पढ़ता हूँ। प्रतिदिन मैं स्कूल बस से विद्यालय जाता हूँ। हमारे गिरे मुझे सुझाव कल्पने खलना पड़ता है। गर्मियों में तो कोई बरिदाई नहीं होती लेकिन सर्दियों में बहुत बार होना है।

मेरे मित्राकी सरकारी सैकरी में है। मे एक अधिकाारी है। सरकार की ओर से उन्हें एक घर मिला हुआ है। वेक पर बहुत सुंदर है। उसमें मैं अपनी माँ, पिताजी, बहन और बहू को साथ रखता हूँ। मेरी माँ बहुत अच्छी है। मे हल सब चर्ट-कटनी को पकड़ो-सिधारी है और हम सबको लिए खीरबन बछाई है। उनका अपनाब हुआ खाक बहुत ज्योतिष होता है। मे सबसे अच्छा राजम बनाता बनाती है।

मेरी एक छोटी बहन है, नाम है मरवा। बहुत खलारी और बापूरी है। एकभी मीठी का उठा। माई छोटी है, बहुत प्यारी बानी। माता से बंधे पाने लगेसे। मैं उसे बहुत प्यार करता हूँ। उसे खिचड़ी, पकौडिया, मिथुनीने ज्योति लखर देता हूँ। अभी बहुत दिनों पहले बीमार पड़ गई तो हमने सो प्रण ही भुख पाने। जब खोलती है तो जैसे खुश आ जाती है। पानकी गली बस नाचती रहती है।

मेरा माई सुलारी बापू है, वह खलारी काका का साथ है। वह मेरे दो बहुत में पढ़ता है। वह खेलकूद तथा पढ़ाई में लक्ष्य प्रयोग करता है, उसे अर्द्धा शिक्षा की इकम भी मिला है।

3

COPY TEXT

हिन्दी निबन्ध, पत्र-लेखन एवं कहानियाँ

## मेरी आत्मकथा

(मैं एक लड़का)

मेरा नाम तरुण वर्मा है। वर्मा मेरा जातिगत नाम है। मैं 10 वर्ष का हूँ और 5वीं कक्षा में पढ़ता हूँ। प्रतिदिन मैं स्कूल बस से विद्यालय जाता हूँ। इसके लिये मझे सबह जल्दी उठना पड़ता है। गर्मियों में तो कोई



SUMMIT  
TECHNOLOGY GROUP

# Different languages

Support for more than 14 Languages.



# Optical Character Recognition

## Metallians OCR

Scan text from images

Upload Image

English



Scan

### Hemant Kumar Yadav

LinkedIn : <https://www.linkedin.com/in/hytesting>  
Github : <https://github.com/hytesting>

Email : [hytesting@gmail.com](mailto:hytesting@gmail.com)  
Mobile : +917933034299

#### PROJECTS

**Hierarchical Attention Networks for Document Classification** | *Python, Keras, Kaggle IMBD Dataset*

- Designed a CNN with 128 filters achieving an accuracy of **89%**.
- Improved the accuracy to **90.4%** using BNN and attention-based LSTM Encoder.
- Reduced the training time by **82%** using Hierarchical attention network.

**Exploratory Data Analysis on AMEX-Default Prediction Dataset** | *Numpy, Pandas, Matplotlib, Python*

- Collaborated as a part of team to perform EDA on AMEX dataset of size **30+ GB**.
- Performed data preprocessing techniques to handle the missing values.
- Successfully plotted **15+** key insights about the data patterns.

**Visualizing MNIST Dataset with PCA and t-SNE** | *Python, Scikit-Learn, Matplotlib*

- Implemented PCA to project dimensions from **784 to 2** which led to different clusters.
- Improved the cluster details using t-SNE with PCA reduced data of 50 dimensions.

#### ACHIEVEMENTS

- Selected for Amazon ML Summer School 2022.
- Rated **4 star** on CodeChef | Max. Rating **1904** | [Profile Link](#)
- Rated **Specialist** on Codeforces | Max. Rating **1449** | [Profile Link](#)
- Global rank **249/16k** in CodeChef Starters 47.
- Global rank **289/16k** in Codechef SharkDown 2021 - Elimination Parallel
- Global rank **290/16k** in CodeChef Starters 49.
- Global rank **1257/23k** in Codeforces Round 794 (Div. 2).
- Solved 700+ problems of Data Structures and Algorithms.

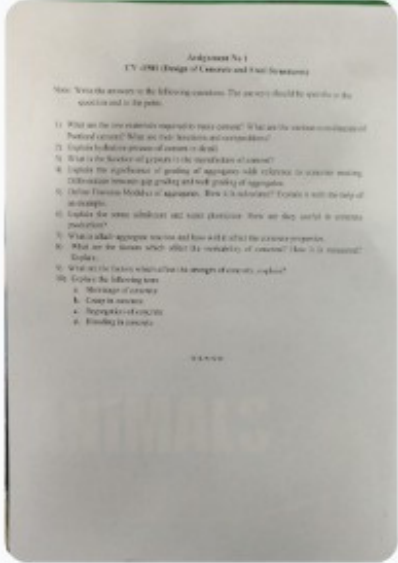
Reading a normal single sided Resume and then segregating it into paragraphs.

# Scanning from JPG

## Metallians OCR

Scan text from images

Upload Image English Scan



**COPY TEXT**

Assignment No 1  
CV -1501 (Design of Concrete and Steel Structures)

Note: Write the answers to the following questions. The answers should be specific to the question and to the point.

1) What are the raw materials required to make cement? What are the various constituents of Portland cement? What are their functions and compositions?

# Generated as PDF

File | C:/Users/HEMANT/Desktop/Metallians/Test1.pdf

1 / 1 | 100% + | [ ] [ ]

Assignment No 1  
CV -1501 (Design of Concrete and Steel Structures)

Note: Write the answers to the following questions. The answers should be specific to the question and to the point.

1) What are the raw materials required to make cement? What are the various constituents of Portland cement? What are their functions and compositions?

2) Explain the hydration process of cement in detail.

3) What is the function of gypsum in the manufacture of cement?

4) Explain the significance of grading of aggregates with reference to concrete making. Differentiate between gap grading and well grading of aggregates.

5) Define Fineness Modulus of aggregates, How it is calculated? Explain it with the help of an example.

6) Explain the terms admixture and super plasticizer. How are they useful in concrete production?

7) What is alkali aggregate reaction and how will it affect the concrete properties?

# Webinar

## Ground Rules

### FUTURE IMPLEMENTATIONS

- Generate editable text PDF.
- Currently we are returning JSON File for thr entire OCR and planning to implement Elastic search to make our repretation searchable.



# Thank you

TEAM – METALLIANS

