*Project Report on*

# NewsFlix

*Submitted in partial fulfillment of the requirements for the award of the degree of*

## Bachelor of Technology

*in*

***Computer Science and Engineering***

**By**
**Abhinav Sobi (U2103008)**
**Alan Joseph (U2103021)**
**Basil Eldho Joseph (U2103057)**
**Daniel Robin (U2103072)**

**Under the guidance of**
**Dr. Mary Priya Sebastian**

**Department of Computer Science and Engineering**
**Rajagiri School of Engineering & Technology (Autonomous)**
**(Parent University: APJ Abdul Kalam Technological University)**
**Rajagiri Valley, Kakkanad, Kochi, 682039**
**April 2025**

# CERTIFICATE

*This is to certify that the project report entitled* ***"NewsFlix"*** *is a bonafide record of the work done by* ***Abhinav Sobi (U2103008), Alan Joseph (U2103021), Basil Eldho Joseph (U2103057), and Daniel Robin (U2103072)****, submitted to the Rajagiri School of Engineering & Technology (RSET) (Autonomous) in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology (B. Tech.) in "Computer Science and Engineering" during the academic year 2024-2025.*

**Project Guide**
Dr. Mary Priya Sebastian
Associate Professor
Dept. of CSE
RSET

**Project Co-ordinator**
Mr. Harikrishnan M
Assistant Professor
Dept. of CSE
RSET

**Head of the Department**
Dr. Preetha K G
Professor
Dept. of CSE
RSET

# ACKNOWLEDGEMENT

We wish to express our sincere gratitude towards **Rev. Dr. Jaison Paul Mulerikkal CMI**, Principal of RSET, and **Dr. Preetha K G**, Head of the Department of Computer Science and Engineering for providing us with the opportunity to undertake our project, ”NewsFlix”.

We are highly indebted to our project coordinators, **Mr. Harikrishnan M**, Assistant Professor, Department of Computer Science and Engineering, and **Ms. Sangeetha Jamal**, Assistant Professor, Department of Computer Science and Engineering, for their valuable support.

It is indeed our pleasure and a moment of satisfaction for us to express our sincere gratitude to my project guide **Dr. Mary Priya Sebastian**, Associate Professor, Department of Computer Science and Engineering, for her patience and all the priceless advice and wisdom she has shared with us.

Last but not the least, we would like to express our sincere gratitude towards all other teachers and friends for their continuous support and constructive ideas.

**Abhinav Sobi**
**Alan Joseph**
**Basil Eldho Joseph**
**Daniel Robin**

# Abstract

The NewsFlix project aims to revolutionize news consumption by automatically transforming newspaper articles into short, engaging videos. Using advanced image processing techniques, the system extracts and summarizes articles from uploaded images of newspaper pages, categorizes them, and generates videos incorporating relevant visuals, audio summaries, and captions. This approach makes news more accessible, particularly for individuals with visual impairments, senior citizens, or younger audiences who prefer multimedia formats over traditional reading. Compared to existing methods, NewsFlix offers a seamless integration of text-to-video conversion, enhancing user engagement and content accessibility. By leveraging automation and multimedia, the project addresses the growing demand for quick, digestible news content while making printed news more relevant in the digital age.

# Contents

# List of Abbreviations

**BERT** – Bidirectional Encoder Representations from Transformers

**T-BERTSum** – Text-BERT Summarization

**NTM** – Neural Topic Model

**CNN** – Convolutional Neural Network

**LSTM** – Long Short-Term Memory

**BoGAN** – Bottom-Up Generative Adversarial Network

**LFCN** – Local Feature Convolutional Network

**BSDS500** – Berkeley Segmentation Data Set 500

**NYUD** – New York University Depth

**TTE** – Text-Text Encoder

**LFC** – Local Feature Convolution

**THUCNews** – Tsinghua University Chinese News Dataset

**MCNews** – Multi-Class News Dataset

**LLM** – Large Language Model

**OCR** – Optical Character Recognition

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Background

Media consumption around the world has changed ironically and significantly, especially following the advent of digital media. Although newspapers have traditionally been a main source of information, they now have the problem of not reaching out to all audiences. In addition, younger generations and people with visual impairments or shorter concentration spans generally find traditional print articles less appealing. Many young people-especially-prefer quicker, visual platforms, in particular with short videos, to long articles. In such a scenario, newspapers generally find it difficult to engage these demographics.

The efforts for NewsFlix seek to modernize the content of news and newspapers through the conversion of news stories into engaging, online, and brief video formats. This project will make the news more appealing and accessible to a wider audience, especially those who would have difficulty reading small print or would prefer a multimedia format.

A NewsFlix user may take a picture of the newspaper page. With the use of sophisticated image processing like edge or object detection, each story is identified and pulled from the newspaper page. Each story is analyzed, summarized, and classified into politics, sports, and general news. Short video production involves choosing articles from each summary category. Each video features key images, audio summaries, and captions that render the news more appealing and digestible.

By turning news articles into short videos, NewsFlix not only widens the scope of printed media but also bridges the divide between classical news and the modern digital expectations of the audience. This quite modern-aligned way of bringing news caters to a wider audience by keeping them informed in an interactive manner.

### 1.2 Problem Definition

To develop a system to convert traditional news articles for consumption via video in shorter length, therefore making it much more audience friendly, who are interested in multimedia formats, it has become a challenge for audiences who are not engaged with content coming from traditional newspapers. The system will help the media blossom, in the same breath making it much easier for the blind to enjoy for those around them and limiting attention spans.

### 1.3 Scope and Motivation

The proposed system, NewsFlix, aims to enhance the accessibility and engagement of traditional news by enabling the automatic transformation from newspapers to short video clips. The platform allows users to upload images of newspaper pages; the system employs the following steps for each article: detects individual articles, extracts the text, summarizes it, categorizes the news, and creates videos with associated visuals accompanied by audio narration and captions. This project encompasses the entire process from recognizing the articles to generating the videos and sharing them on social sites, giving the users a fresh take in reading print media-oriented content.

Traditionally catering for their diverse audiences, and building a bridge: it's a two-way traffickers. Many people find printed versions of the news difficult to read, perhaps because the nature of the text is so additional effort for the eye to rehearse. Great short-form video on social media reflects a larger audience who simply want short, engaging information; thus, transforming news stories into short engaging videos through NewsFlix can get around this gap-thereby making news easily consumable for other audience segments. This aims to shift modern news consumption to allow newspapers in broad plan to reach diluted modern audiences with the current digital habits on their fronts.

## 1.4 Objectives

1. **Article Recognition and Extraction:** Develop a system that can automatically recognize and extract individual articles from uploaded images of newspaper pages using edge or object detection techniques.

2. **Text Extraction and Summarization:** Implement processes to extract and summarize text from each identified article, providing concise and easily understandable content.

3. **News Categorization:** Classify each article into relevant categories, such as politics, sports, or general news, to enable users to easily select and filter topics of interest.

4. **Automated Video Generation:** Create short, engaging videos based on the summarized article content, incorporating relevant visuals, audio narration, and captions.

5. **User-Friendly Interface and Social Media Integration:** Provide an intuitive interface where users can upload images, view summaries, select categories, and export videos to social media platforms.

6. **Accessibility and Engagement Enhancement:** Make news content more accessible and appealing, especially for audiences with visual impairments or those who prefer multimedia formats over traditional reading.

## 1.5 Challenges

Some key challenges involved in the project are in accurately recognizing and extracting text from varied and sometimes low-resolution quality newspaper image pages which can be subject to lighting and quality challenges. Meanwhile, creating a system to summarize advanced articles correctly and generate videos in an appealing manner, with fitting visuals and audio, takes a lot of computational power and an appointment on fine-tuning. Ensuring scalability and easy usability of the platform in processing various news genres and types of content is another challenge.

### 1.6 Assumptions

1. The newspaper images provided for processing will be of reasonable quality, with clear and legible text.

2. The system will have access to a sufficiently large database of images and visuals to generate relevant video content for news articles.

3. Users will have a stable internet connection for uploading images and exporting videos to social media platforms.

4. The news provided to the system must be genuine, accurate, and free from any misinformation, as the system operates under the assumption that the input data is reliable and trustworthy.

### 1.7 Societal / Industrial Relevance

The project is aimed at producing lively news videos, instead of text articles, boosting accessibility of information for people with visual impairments, elderly populations, and young ones with short attention spans. This makes for greater inclusion and bridges the widening technology gap.

Industrially, the project can transform how news agencies and content creators engage with audiences by offering a new format for news distribution on social media. It can also be applied in education, marketing, and other industries where transforming text into multimedia content can increase reach and engagement.

### 1.8 Organization of the Report

The report consists of three main chapters. Chapter 1, Introduction, summarizes the background, problem definition, scope, and motivation for the project. It also discusses objectives, challenges, assumptions, and social or industrial relevance. The second chapter, Literature Review, explains the detailed literature review regarding text summarization, edge detection, video generation, and also discusses the gaps in current research. Chapter 3 System Design outlines the project at a high level, then details the structural design into subtasks and covers methodologies and tools used, work allocation to the team members, and the timeline.

# Chapter 2

# Literature Survey

## 2.1 T-BERTSum: Topic-Aware Text Summarization Based on BERT

This paper explores the deployment of T-BERTSum-based topic-aware text summarization by investigating the improvement of coherence and relevance of social media-based and other long text sources. It also identifies challenges within the scope of summarization techniques, including the unclear long-distance dependencies and scaffolding upon a precise representation on the topics. For this reason, the authors present an extant/extractive, two-stage framework where the combination of extractive and abstractive summarization takes place supported with topics from Neural Topic Model that will guide the summarization process. An emphasis is in making topical embeddings and revisions made to BERT work in harmony for semantic enhancement and a reduction in repetition in idea summaries. [1]

The T-BERTSum architecture consists of three main functions: topic inference, sentence extraction, and summary generation. In other words, the model is based on a transformer architecture. The input text is read by the encoder to produce contextual representations; the top topics are then used by the transformer to generate the final summaries by the decoder. An efficient combination of cutting-edge tools and frameworks ensures faster computations and batchvprocesses. Being state-of-the-art in summary quality and factual relevance from its trial evaluations on benchmark datasets CNN/Daily Mail and XSum, T-BERTSum shows qualitative assessments that showcase its ability to further improve text summarization for a range of applications and, in turn, open avenues for efficient information retrieval and content management schemes.

## 2.2 An Adaptive and Robust Edge Detection Method Based on Edge Proportion Statistics

In this paper, the authors propose a new method designed for real-time applications of edge detection, which puts emphasis on the enhancement of accuracy and efficiency of image processing. Also, the authors described the shortcomings of the existing edge detection methods- incapability of adaptively learning from high-level representations and their computationally expensive algorithms, thus rendering them unfit for embedded systems and mobile devices.[2]

It employs edge proportion statistics and two-dimensional entropy to automatically adjust the gradient threshold in order to yield a clean region of contiguous edge segments instead of a binary edge map. This method accomplishes better representation of features of imagery and allows facilitating high-level functions such as line fitting, ellipse detection, and image segmentation.

The authors compare their method to state-of-the-art methods in the BSDS500 and NYUD dataset and demonstrated quite impressive performance with lower computational burdens. The study elaborates upon how vital local cues are for edge detection-like, colors, and gradients while admitting that getting such significant information is rather a problem.

The technical implementation of this is projected around three principal steps, namely, adaptive thresholding with the help of 2D entropy and the use of curvature predictive techniques in connecting edge points into segments. The findings confirm that the proposed method improved edge detection performance, thus opening avenues for further research to augment its capabilities in representing complex image content.

Thereby this research contributes nicely to the domain of image processing through an ingenious development of a robust edge detection framework that adjusts precision with computational efficiency, thus apt for real-time scenarios.

## 2.3 Two-Step CNN Framework for Text Line Recognition in Camera-Captured Images

The paper introduces the Two-Step CNN Framework for text line recognition from camera images concerning identity papers. It points out the problems posed by existing methods, which often do not handle the distorting conditions introduced during the process of image acquisition from mobile devices. For such other issues, the proposed dual artificial learning neural network separates character segmentation from character recognition and hence enhances efficiency at a much higher level. [3]

This framework is made specifically for low-resolution images with complex backgrounds and ensures a faster recognition rate in several languages and fonts. This discussion foregrounds the significance of "on-the-device" work, where security issues associated with cloud-based solutions are diluted and properly mitigated, especially if sensitive personal data is contained within identity documents.

The proposed architecture of the system includes a character segmenter and a character recognizer, both geared to operate under the parameter restrictions imposed by mobile and embedded systems. While the segmenter is built as a language-independent one, the recognizer is molded in a way so as to allow customization for specific languages, proven versatile through experiments on Cyrillic, Armenian, and Chinese text lines.

The synthetic data employed in training the model for real-life applications also demonstrate the feasibility of such technical implementations. The comparison between the current methods and the past methods, including Tesseract, has certainly shown that the current techniques are superior compared to the latter with reference to both speed and accuracy. The experimental results show that the framework can significantly reduce processing times while at the same time guarantee a high degree of integrity for the data, making it a strong candidate for efficient text line recognition on mobile platforms.

## 2.4 A Long-Text Classification Method of Chinese News Based on BERT and CNN

The authors of the paper introduce the Local Feature Convolutional Network (LFCN) model for better classification of Chinese long-text news articles. In doing so, they had brought notice to the challenges that most previous methods have. Some of these include the inability of BERT to accommodate long documents and the inability to modify the lengths of sentences targeted by long documents. In order to solve the above problems, they proposed an algorithm known as Dynamic LEAD-n (DLn), which extracts the most relevant sentences out of a long text and reduces it into the maximum input length of 256 tokens.[4]

The LFCN model merges further into the BERT model to obtain representations of the feature vectors at the sentence level that capture both global and local features of the text. There is a textual-text encoder layer providing input as text vectors, and an LFC(layer) as a feature extractor to pull out salient features, such as key phrases. The final model merges these feature vectors into a single representation for classification.

The experiments were carried on the two datasets, namely, THUCNews and MCNews, to reveal the effectiveness of the proposed model. Evaluation metrics comprised accuracy, weighted precision, recall, and F1 score, suggesting that the model performed better than the common classification methods. The work emphasizes the need to tackle the challenges posed by lengthy Chinese texts and presents a very effective framework to improve text classification tasks in NLP.

The architecture is designed to capture both local and global features while coming to terms with the text context. BERT with its ambience was fused with convolutional networks to seek efficiency and accuracy in Chinese news text classification, which brightened the future of work in the domain.

## 2.5 Scripted Video Generation With a Bottom-Up Generative Adversarial Network

This paper proposes a Bottom-Up Generative Adversarial Network capable of generating videos from natural language descriptions with the goal of overcoming the semantic-alignment and temporal-coherency challenges that go hand-in-hand in video synthesis. They argue for a valid claim that constraints on existing methods are now enforced, especially to ensure consistency between the generated and input texts when dealing with more complex cases that concern more than one entity.

To solve the above problems, a sophisticated mechanism comprising a video generator, a region-level semantic alignment module, and two discriminators working to ensure frame-level and video-level coherence is proposed through the BoGAN architecture. This multi-faceted approach smoothly allows the BoGAM to capture local and global semantic relationships and generate frames that are not only visually coherent but also semantically coherent with the given script.

This framework is explicitly developed to handle the complexities of video generation, using a bi-directional Long Short-Term Memory (LSTM) network for text encoding. The purpose of this is to enhance the model's ability to understand and encode the semantics of the input. The region-level module focuses on the contributing sub-regions of the video, using wider search regions based on the most prominent words. The frame-level discriminators make sure that the generated video appears realistic, and that its frames link well with each other and maintain coherence.

Discussions on the technical details of implementation indicate that a "two-pass training" process has been used, which greatly improves performance on both synthetic and real-world datasets, indicating the effectiveness of the approach as compared to simultaneous training methods. Results indicate BoGAN state-of-the-art performance in video generation, underscoring the technical maturity of this approach as a serious solution for NHS video synthesis that faithfully aligns with natural language descriptions.

## 2.6 Summary and Gaps Identified

The second chapter of our report showcases advancements in different technical domains pertaining to our project. They are hot research topics of today's world. The first mentioned is a topic-aware text summarization model called T-BERTSum. It is based on the BERT encoding and neural topic models, which aim to enhance topical relevance and coherence, giving state-of-the-art results in datasets like CNN and Daily Mail.

Next we documented an adaptive edge detection method based on 2D entropy and curvature prediction for real-time image processing. It is superior in accuracy & computational efficiency compared with BSDS500 and NYUD datasets.

The chapter also details an OCR framework called Two-Step CNN Framework for text line recognition in low-quality, camera-captured images, particularly for identity documents. By combining character segmentation & recognition, the system outperforms Tesseract in speed and accuracy. It also ensures secure on device processing.

We also mention a long-text classification model for news articles in Chinese powered by the Local Feature Convolutional Network (LFCN). It helps BERT's shortcomings concerning sequence lengths. The method is tested for classification of the THUCNews and MCNews datasets.

Finally, the chapter details about BoGAN, a Bottom-Up Generative Adversarial Network. It performs video generation from natural language descriptions. Its integration of LSTMs & dual discriminators ensures semantic alignment & temporal coherence, achieving great performance on multiple datasets.

**Gaps Identified**

1. **Limited Accuracy in Article Recognition and Extraction**

   While the *NewsFlix* system uses image processing techniques to recognize and extract articles from newspaper images, challenges remain in accurately processing low-quality images or poorly scanned newspaper pages. Variability in print quality, lighting conditions, and text formatting can affect the reliability of the article extraction, limiting the system's accuracy.

2. **Lack of Contextual Understanding in Text Summarization**

   Although the system utilizes summarization algorithms to condense news articles, these models may still face difficulties in capturing the full context of complex or nuanced articles. This can result in summaries that miss important details or fail to adequately convey the core message of the article, affecting user understanding.

3. **Insufficient Customization for Diverse News Categories**

   The classification of articles into categories like politics, sports, or general news may not always be accurate or comprehensive, particularly for articles that span multiple topics. Existing models may struggle to categorize content correctly, which limits the system's ability to tailor content to user preferences effectively.

4. **Limited Integration of Multimedia Content**

   While the system generates videos by incorporating text-to-speech and visual elements, the integration of multimedia content can be improved. The system may not always generate perfectly synchronized video and audio or choose the most relevant images, reducing the quality and engagement of the resulting videos.

5. **Challenges in Scalability and Real-Time Processing**

   As the volume of news content grows, the system may face difficulties in processing large numbers of articles or generating videos in real-time. Current methods may not be optimized for handling large datasets or producing videos quickly enough to meet the demands of a rapidly evolving news environment.

6. **Inadequate User Personalization**

   The system's current design lacks advanced user personalization features. While users can select categories, it may not offer individualized recommendations based on past behavior or preferences, limiting its ability to provide a truly customized news experience.

# Chapter 3

# System Design

This chapter includes the overall framework and detailed structural design of the project. Work division and timeline of the project are also detailed.

## 3.1 System Architecture

The overall system architecture of the project is shown in Figure 3.1. The framework consists of Article Segmenter, OCR Module, Summarizer, Classifier and Video Generator. A detailed explanation of these five components can be found in Section 3.2.

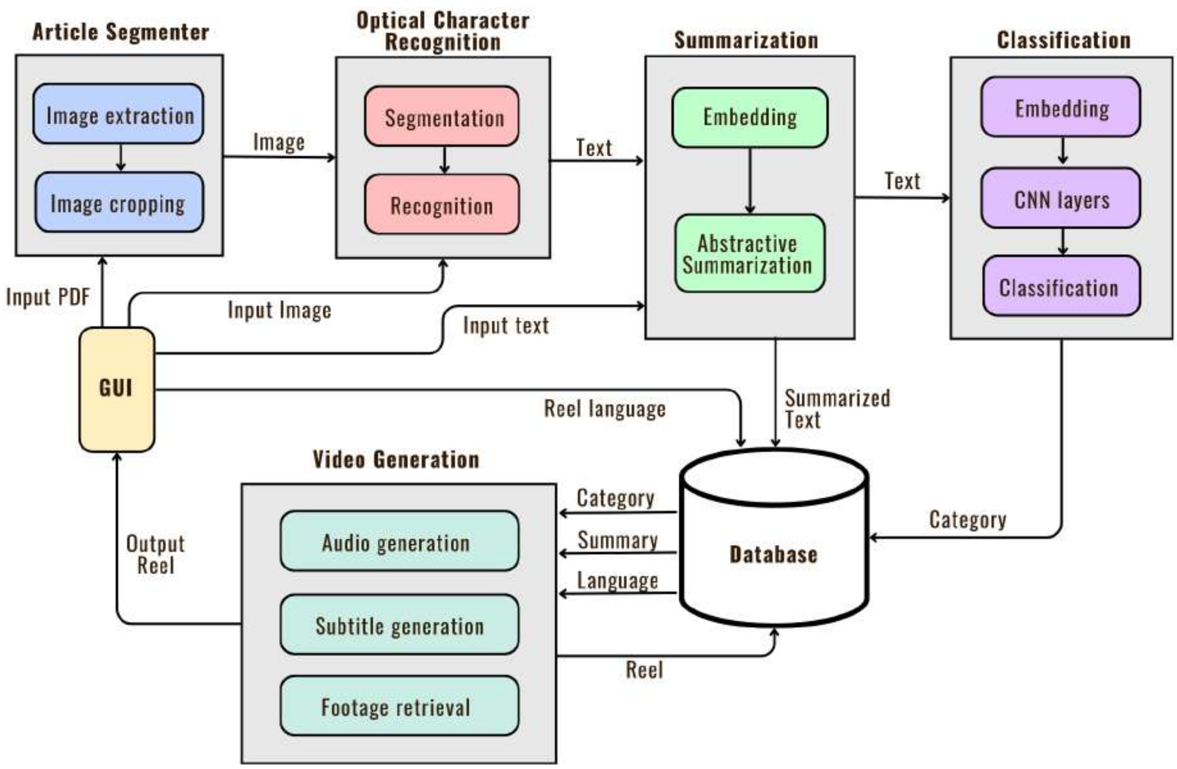



Figure 3.1: Architecture Diagram.

Other than these main modules, there are two other components which are integral for this project. For GUI industry standard React.js library is used. Web based client interface ensures inclusion of diverse devices like desktops, laptops and mobile devices. Database is handled using PostgreSQL, the open-source RDBMS. It is used for storing user authentication data and data related to generated reels.

## 3.2 Component Design

The overall system architecture of the project is shown in Figure 3.1. The framework consists of five major components or modules: Article Segmenter, OCR Module, Summarizer, Classifier, and Video Generator. Each of these 5 modules are detailed in this section.

### 3.2.1 Article Segmentation Module

The Segmentation module separates each article into different images for OCR. It takes a newspaper image as input, then performs preprocessing to improve accuracy. Conversion to grayscale is done and a Gaussian blur is applied to reduce noise. Thresholding is applied to create a binary image where text areas are in white. Separation of foreground and background is carried out using the available data. Detecting Contours which creates segments.[2][5] The segmented images are saved for the OCR module to access. Figure 3.2 represents the architecture of this module.

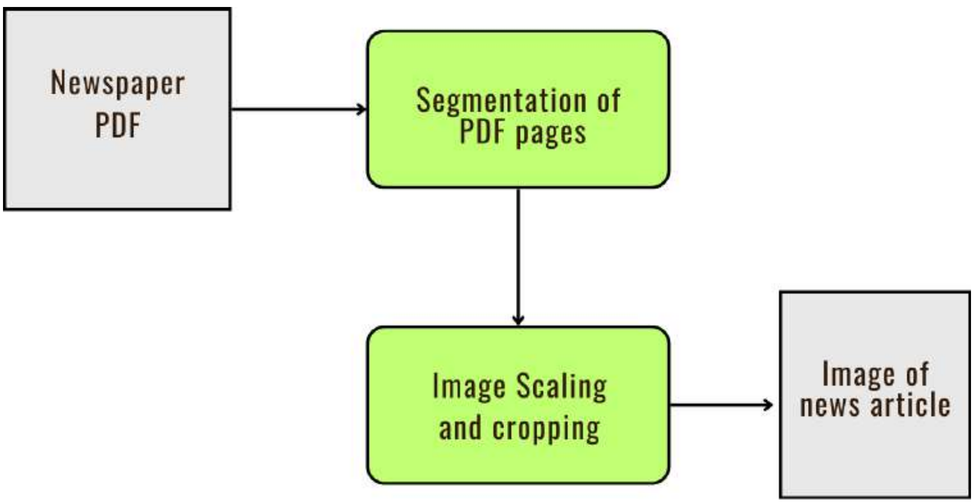



Figure 3.2: Article Segmentation Module.

### 3.2.2 OCR Module

The OCR module consists of a computer vision based tool to extract the text present in images. The input to the module will be an image of a news article. This image is provided to this module by the Article Segmenter module. The text in the image is recognized and extracted. Hence, the output produced by this module is a string comprised of the text of a news article.[6]

Steps involved in a typical ML based OCR model is as follows:
First line segmentation is done using baseline and capline of individual text lines. Character segmentation is carried out using a language independent CNN. Next, feature extraction from segmented characters is done. Then, character recognition is done using a separate CNN. Finally, characters are concatenated to get the output text.[7]

The project leveraged the OCR capabilities of modern LLMs, since it involves images having complex text layouts. Images may contain text divided into several vertical paragraphs, different font sizes and even pictures of the news event. AI models like Gemini and 4o are able to handle them effectively. The output text is used by the Summarization module. Figure 3.3 represents the architecture of a typical OCR tool.

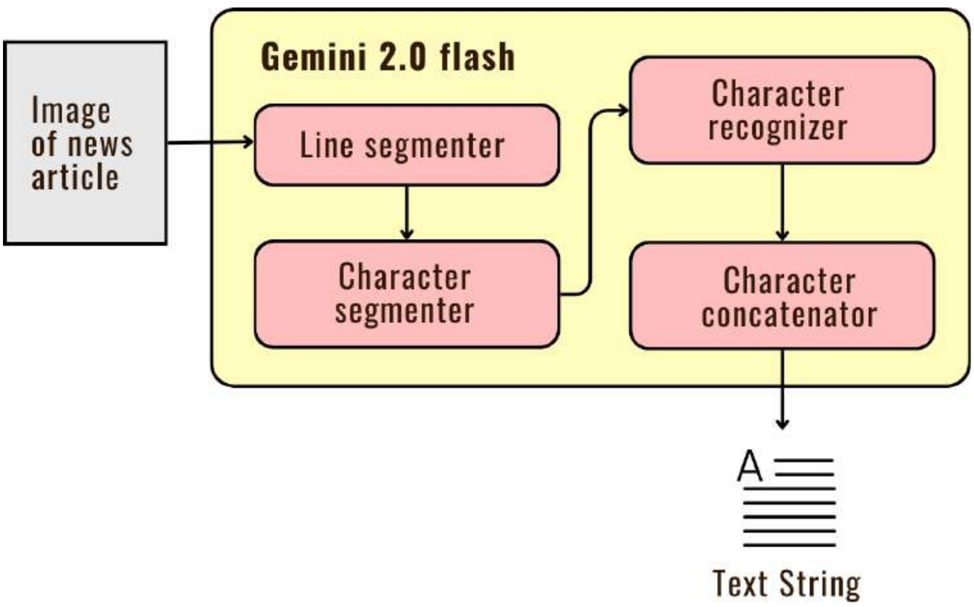



Figure 3.3: OCR Module.

### 3.2.3 Summarization Module

The module deals with the task of summarizing news articles into short paragraphs of around 75 words. This module takes the text string received from the OCR module as input. Abstractive summarization is performed to generate a summary using word embeddings from a pre-trained BART model. It is a fine-tuned version of Facebook's BART (Bidirectional and Auto-Regressive Transformer) model. This model is designed for text summarization tasks and fine-tuned on the SAMSum dataset, which is a dataset of dialogues.[8][9]

Performance metrics such as ROUGE-1, ROUGE-2, and ROUGE-L can be used to evaluate the quality of text summarization by measuring the overlap of unigrams, bigrams, and longest common subsequences between the generated and reference summaries. These metrics help assess how well the summarization model captures key information, balancing precision and recall to determine overall effectiveness. Finally, the summary text is passed to the Classifier module.[10] Figure 3.4 shows the architecture diagram of this module.

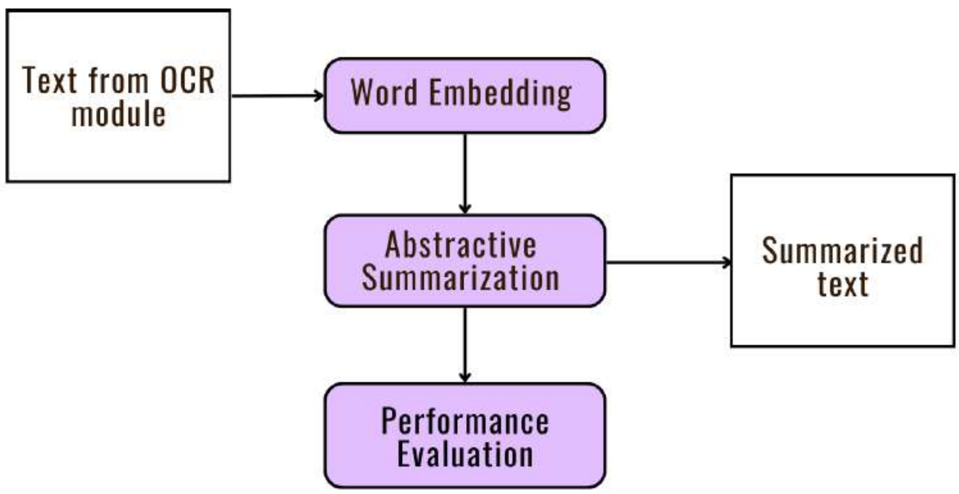



Figure 3.4: Summarization Module.

### 3.2.4 Classification Module

The module deals with the task of classifying news articles into the class that they belong to, according to the matter they address. This module takes the summarized article as input here to work as a piece of raw text for the classification problem and converts it into a representational format that can be fed into the modeling domain using a word embedding mechanism named BERT. In relation to BERT, the term ”pre-trained language model” means that it can, in essence, characterize context relationships between words in a sentence, and gives each word a vector representation with high-dimensionality encoding its meaning in a particular context relative to both the left and right sides.[11]

The extracted features and word embeddings are then concatenated to form a one-dimensional feature vector. Such a multi-model representation allows the classifier to enjoy both benefits, extracting one hundred probable semantic contexts from BERT and retaining extracted feature information that is structured towards improving overall classification accuracy. Softmax provides the classifier with a multi-class capability whereby an article may be assigned to one out of several mutually exclusive classes, transforming predictions into probabilities summing to 1.[12] Figure 3.5 shows the architecture diagram of this module.

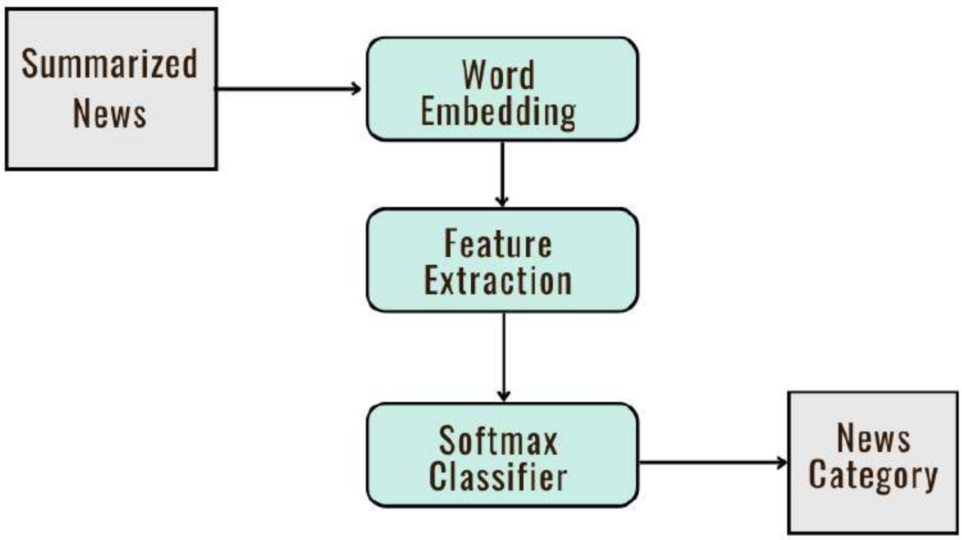



Figure 3.5: Classification Module.

### 3.2.5 Video Generation Module

The Video Generator is the final module of the project. This is most crucial component of the project, since it directly affects the quality of the reel produced. The Summarized text and Category of news article are taken as input by this module.[13]

Footage and Images for the video are retrieved using pexels API based on the category. From summarized text, voice and captions are generated and synced using gTTS and moviepy library. All the obtained outputs are assembled and synchronized to obtain the final video. Figure 3.6 shows the architecture diagram of this module.

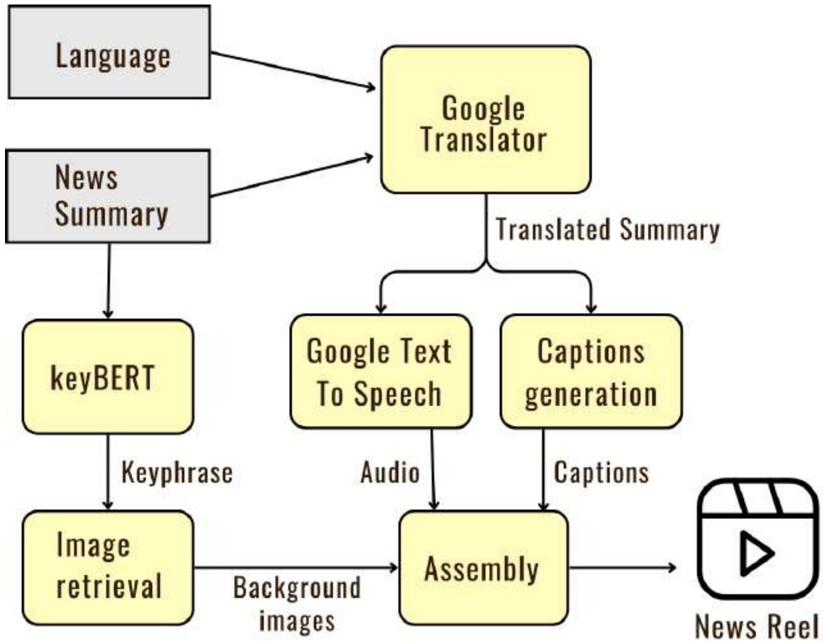



Figure 3.6: Video Generation Module.

## 3.3 Algorithm Design

### 3.3.1 Video Generation Algorithm

A novel video assembly solution was designed for this project. It was coded in python, with the help of several built in libraries as well as 3rd party services. The algorithm is roughly as follows:

**Algorithm 1** News Reel Generation

**Input:** *Summary* and *Category* of news article, Required output *language*

Extract a *keyphrase* from the *Summary*.<br>
**if** *language* is not english **then**<br>
    Translate the *Summary* to *language*.<br>
**end if**<br>
Convert *Summary* text to audio using gTTS, save it as reel.mp3.<br>
Speed up reel.mp3 by 10% to reduce duration.<br>
Declare *duration* as duration of reel.mp3.<br>
Initialize *video* with 1080x1920p resolution.<br>
Calculate no.of images, $\mathbf{n} \leftarrow \lceil \mathbf{duration/7} \rceil$<br>
**for** $i = 1$ to $n$ **do**<br>
    Fetch **i**th image under *keyphrase* using Pexels API.<br>
    Download the image as i.jpg.<br>
    Append i.jpg as background of *video* for 7 seconds.<br>
**end for**<br>
Transcribe reel.mp3 using Whisper to get synchronized captions.<br>
Add captions to *video* using the font suitable for *language*.<br>
Set reel.mp3 as the audio for *video*.<br>
Write *video* at 1 fps and save it as reel.mp4.

**Output:** News video file reel.mp4

Video assembly module is the most crucial part of the project, since it directly affects the quality of the reel produced. Several video creation techniques were explored to find the option which is both feasible, reliable and produces good output. The latest LLM based video generation tools produce visually appealing content, but is extremely computationally expensive. The custom algorithm for the project turned out to be both lightweight and produces good quality reels.

### 3.4 Sequence Diagram

The sequence diagram shows a graphical representation that illustrates the interactions between various components in our project. The sequence of interactions are arranged in a chronological order. It shows how each components and objects pertaining to the project communicate with each other through a series of messages to accomplish a specific task or workflow. It gives an overall picture from uploading of a newspaper image to the generation of news reel.

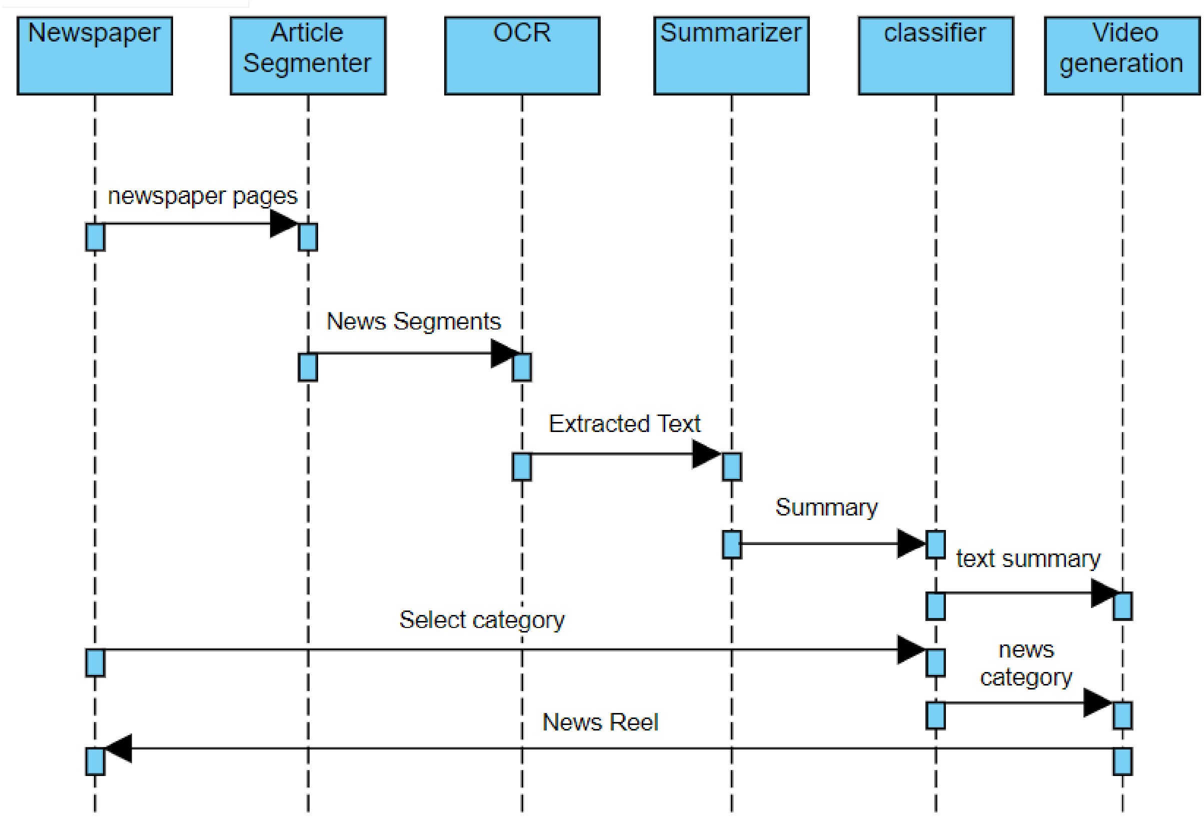



Figure 3.7: Sequence Diagram

## 3.5 Hardware and Software Requirements

The software requirements include Visual Studio Code, a versatile source code editor with debugging and version control; Firebase, a platform offering backend services such as authentication and hosting; PgAdmin 4, a web-based tool for managing PostgreSQL databases; Postman, an API testing and development utility; and a modern web browser like Chrome or Firefox for running and testing web applications.

On the hardware side, the system needs an Intel Core i5 processor or higher for efficient multitasking, an Nvidia GTX 1650 graphics card for handling visual tasks, 8 GB of RAM to ensure smooth operation, a screen with a resolution of at least 1280x720 for clear display, and stable network connectivity for accessing online resources and deploying applications.

## 3.6 Datasets Identified

Three of the five main components of the project involve ML models at their core - Classification module, Optical Character Recognition and Summarization Module. The datasets used in the training and testing of these models are as follows:

### 3.6.1 Datasets for Classifier

Dataset used is "Multilingual IPTC Media Topic dataset EMMediaTopic 1.0". It is a collection of news articles in Catalan, Croatian, Greek, and Slovenian, automatically annotated with the 17 top-level topic labels from the IPTC NewsCodes Media Topic hierarchical schema. The texts were annotated by the GPT-4o LLM.

### 3.6.2 Datasets for OCR model

Datasets used are those typically used in the training of LLMs. They include diverse datasets, like, synthetic text datasets, printed text scans, handwritten text samples, and multilingual text datasets. They are likely derived from publicly available OCR benchmarks such as ICDAR datasets and other open-license sources. These datasets ensure the model can generalize across various fonts, layouts, languages, and scripts, enabling accurate text extraction from a wide range of image inputs.

### 3.6.3 Datasets for Summarizer

Dataset used is "SAMSum", prepared by Samsung R&D Institute Poland and is distributed for research purposes. It contains about 16k life-like conversations with summaries. The style and register are diversified - conversations could be informal, semi-formal or formal, they may contain slang words, emoticons and typos. Then, the conversations were annotated with summaries. It was assumed that summaries should be a concise brief of what people talked about in the conversation in third person.

## 3.7 Work Division

All four team members contributed equally to our project. Each member had an equal share of responsibilities and tasks, to make sure that the workload was evenly distributed among us. This collaborative approach helped maintain balance & fairness and fostered teamwork & accountability.

The Table 3.1 details the assignment of modules to the team members. Team members worked not only on designing and developing their respective modules, but also on testing the model's accuracy, creating documentation, reports, presentations, etc.

| Sl. no. | Team Member | Assigned Module |
|---|---|---|
| 1 | Abhinav Sobi | News Classification Module |
| 2 | Alan Joseph | Optical Character Recognition |
| 3 | Basil Eldho Joseph | Article Segmentation |
| 4 | Daniel Robin | News Summarization Module |
| 5 | All members | Video Assembly Module |

Table 3.1: Assignment of Modules to Team Members.

## 3.8 Expected Outputs

The application should create an interactive short video summary of news articles. Each of the short films is expected to be interlaced with very appealing audio narration summarizing the storyline, along with subtitling for better accessibility and viewer engagement. The project should develop an intuitive yet eye-catching and attractive user-friendly platform for users to consume briefs, tutorials, and news as a small capsule of concise, interactive, and impactful way.

## 3.9 Project Timeline

The project timeline is represented by the Gantt chart below:

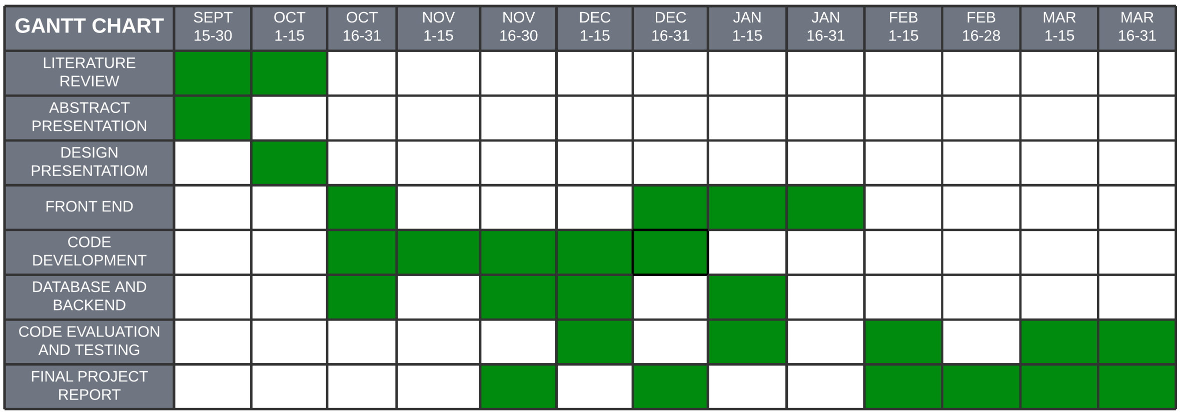



Figure 3.8: Gantt Chart.

## 3.10 Summary

This chapter encompasses both the comprehensive architectural framework and the component designs of the project. Additionally, the chapter provides an in-depth description of the division of tasks and the project's timeline hence offering a detailed overview of how the work is structured and scheduled.

# Chapter 4

# Results and Discussion

## 4.1 Introduction

This chapter presents the results of the project and analyzes the outcomes. It includes a performance evaluation and highlights the main findings. The chapter also discusses the significance of these results in relation to the project's goals.

## 4.2 Quantitative Results

Quantitative results of OCR, Summarization, Classification, and Video Generation are detailed in the experiments below.

### 4.2.1 Experiment 1: OCR

Gemini-2.0-Flash is an advanced Optical Character Recognition (OCR) system that utilizes deep learning-based text extraction. It processes scanned document images as input and outputs machine-readable text. The system applies image preprocessing, character segmentation, and neural network-based recognition to ensure high accuracy in printed and handwritten text recognition.

The performance of Gemini-2.0-Flash is evaluated using three key metrics:

- **Precision Character Recognition (PCR)**: The percentage of correctly recognized characters, calculated as:

$$PCR = \left(\frac{\text{Correctly Recognized Characters}}{\text{Total Characters in Ground Truth}}\right) \times 100 \tag{4.1}$$

- **Character Error Rate (CER)**: The percentage of misrecognized, inserted, or deleted characters, calculated using the Levenshtein distance ($D$) as follows:

$$CER = \left( \frac{D}{\text{Total Characters in Ground Truth}} \right) \times 100 \quad (4.2)$$

where $D$ represents the number of character edits (insertions, deletions, and substitutions) needed to transform the recognized text into the ground truth.

- **Word Error Rate (WER)**: The percentage of incorrectly recognized words, computed similarly to CER but at the word level:

$$WER = \left( \frac{D_w}{\text{Total Words in Ground Truth}} \right) \times 100 \quad (4.3)$$

where $D_w$ represents the number of word-level edits required to match the ground truth text.

| **Model / System** | **PCR %** | **CER %** | **WER %** |
|---|---|---|---|
| Gemini-2.0-Flash | 96.21 | 3.79 | 6.63 |

Table 4.1: OCR Performance Metrics for Gemini-2.0-Flash

The results indicate that Gemini-2.0-Flash achieves a high PCR of 96.21%, with a low CER of 3.79% and a WER of 6.63%. These values demonstrate its strong ability to accurately recognize characters and words. Additionally, Gemini-2.0-Flash outperforms similar OCR systems in recognition accuracy, making it one of the most reliable options for OCR-based text extraction.

#### 4.2.2 Experiment 2: Summarization

The summarization module utilizes BART, a transformer-based neural network, to perform abstractive summarization. Unlike extractive methods that copy sentences directly from the input, BART generates new text by first encoding the input document and then decoding it into a concise, human-like summary. The model is trained using a denoising autoencoder approach, where corrupted input sequences are reconstructed into coherent summaries. Given a news article as input, BART produces a summary that aims to capture the key points in a more natural and readable format.

The performance of the summarization system is evaluated using ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metrics. The system takes a reference summary and a generated summary as input and computes similarity scores based on n-gram overlap and the longest common subsequence (LCS).

The ROUGE metrics used are:

- **ROUGE-1**: Measures unigram (word-level) overlap between the generated and reference summaries, calculated as:

$$\text{ROUGE-1} = \frac{\text{Number of overlapping unigrams}}{\text{Total unigrams in reference summary}} \tag{4.4}$$

- **ROUGE-2**: Measures bigram (two-word sequence) overlap, computed as:

$$\text{ROUGE-2} = \frac{\text{Number of overlapping bigrams}}{\text{Total bigrams in reference summary}} \tag{4.5}$$

- **ROUGE-L**: Measures the longest common subsequence (LCS) similarity between summaries, which considers sentence structure preservation. It is given by:

$$\text{ROUGE-L} = \frac{\text{Length of LCS}}{\text{Total words in reference summary}} \tag{4.6}$$

| **ROUGE** | **Precision** | **Recall** | **F1-score** |
|---|---|---|---|
| ROUGE-1 | 0.9651 | 0.1593 | 0.2735 |
| ROUGE-2 | 0.8353 | 0.1365 | 0.2347 |
| ROUGE-L | 0.9419 | 0.1555 | 0.2669 |

Table 4.2: ROUGE Scores for Summarization Module

The evaluation results indicate that ROUGE-1 achieves a high precision of 0.9651 but a relatively low recall of 0.1593, meaning that while the generated summaries contain accurate words from the reference, they fail to capture the full content. ROUGE-2 follows a similar pattern, with precision at 0.8353 and recall at 0.1365, reflecting limited bigram overlap.

ROUGE-L evaluates sentence structure similarity, yielding a precision of 0.9419 and a recall of 0.1555, further supporting that individual words and phrases match well, but full sentence structures may differ. The low recall scores across all metrics indicate that key details from the reference summaries are not fully retained, which may impact informativeness.

Overall, the F1-scores (ROUGE-1: 0.2735, ROUGE-2: 0.2347, ROUGE-L: 0.2669) suggest that the model generates summaries that are concise but may omit some important details from the reference. This indicates a precision-biased generation, where the summarization system prioritizes fluency and readability while potentially missing key information from the original text.

### 4.2.3 Experiment 3: Classifier

The classification model processes news articles and assigns them to one of 16 predefined categories based on textual features. Given an input news article, the model outputs a predicted category, which is compared against the ground truth labels to evaluate performance.

The classification performance is assessed using three key metrics:

- **Accuracy**: The proportion of correctly classified instances out of the total instances, calculated as:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Instances}} \tag{4.7}$$

- **Micro-F1 Score**: A harmonic mean of precision and recall that gives equal weight to each instance, computed as:

$$\text{Micro-F1} = \frac{2 \times \sum \text{True Positives}}{2 \times \sum \text{True Positives} + \sum \text{False Positives} + \sum \text{False Negatives}} \tag{4.8}$$

- **Macro-F1 Score**: Computes the F1-score independently for each category and then averages them, calculated as:

$$\text{Macro-F1} = \frac{1}{N} \sum_{i=1}^{N} \text{F1-score}_i \tag{4.9}$$

  where $N$ is the total number of categories (16 in this case).

| | Micro-F1 | Macro-F1 | Accuracy | No. of Instances |
|---|---|---|---|---|
| All (combined) | 0.734 | 0.746 | 0.894 | 1129 |

Table 4.3: Overall Classification Performance

A category-wise performance breakdown is presented in Table 4.4. Categories such as Weather and Sport demonstrate high precision and recall, leading to F1-scores above 0.89. However, categories like Politics and Lifestyle and Leisure show relatively lower F1-scores, indicating potential misclassifications or challenges in distinguishing them from similar categories.

| Category | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Arts, Culture, Entertainment | 0.602 | 0.875 | 0.713 | 64 |
| Conflict, War, and Peace | 0.611 | 0.917 | 0.733 | 36 |
| Crime, Law, and Justice | 0.862 | 0.812 | 0.836 | 69 |
| Disaster, Accident, Emergency | 0.691 | 0.887 | 0.777 | 53 |
| Economy, Business, Finance | 0.779 | 0.508 | 0.615 | 118 |
| Education | 0.847 | 0.735 | 0.787 | 68 |
| Environment | 0.589 | 0.754 | 0.662 | 57 |
| Health | 0.797 | 0.797 | 0.797 | 59 |
| Human Interest | 0.552 | 0.673 | 0.607 | 55 |
| Labour | 0.855 | 0.831 | 0.843 | 71 |
| Lifestyle and Leisure | 0.774 | 0.477 | 0.590 | 86 |
| Politics | 0.568 | 0.735 | 0.641 | 68 |
| Religion | 0.842 | 0.941 | 0.889 | 51 |
| Science and Technology | 0.638 | 0.800 | 0.710 | 55 |
| Society | 0.918 | 0.500 | 0.647 | 112 |
| Sport | 0.823 | 0.968 | 0.891 | 63 |
| Weather | 0.953 | 0.932 | 0.943 | 44 |

Table 4.4: Classification Performance by Category

The overall classification performance, summarized in Table 4.3, shows that the model achieves an accuracy of 89.43%, with a Micro-F1 score of 0.734 and a Macro-F1 score of 0.746, indicating strong predictive capabilities across all categories. The model performs particularly well in categories with well-defined textual patterns, such as Weather and Sport, but exhibits lower recall and F1-scores in areas like Politics and Society, highlighting opportunities for improvement. Enhancing classification accuracy could involve refined feature selection, dataset balancing, or incorporating additional contextual embeddings. Future work should focus on improving recall in underperforming categories to ensure balanced performance across all labels.

#### 4.2.4 Experiment 4: Video Generation

The video generation module compiles a summary from the Summarization Module, extracts keywords using KeyBERT to fetch relevant stock images from the Pexels API, and generates speech from the summary using gTTS. Captions are created to align with the narration. All these components are combined in a custom algorithm to produce a cohesive video reel.

Video reel quality is assessed using key metrics: Context Similarity, measuring alignment between visuals, audio, and subtitles; Audio Clarity, evaluating speech intelligibility; Subtitle Readability, ensuring well-timed and legible captions; Background Image Suitability, assessing image relevance to the news category; Attention Capturing Capability, gauging viewer engagement; and Overall Quality, providing a holistic evaluation of the final product.

Each metric is rated on a scale of 1 to 5, where 5 signifies optimal performance. A high Context Similarity score means all elements align seamlessly. Audio Clarity ensures clear, natural speech. Subtitle Readability reflects accurate synchronization and legibility. Background Image Suitability captures how well visuals enhance comprehension. Attention Capturing Capability measures viewer engagement, while Overall Quality reflects the integration of all elements. Lower scores indicate areas needing improvement, such as misaligned subtitles, unclear audio, or weak engagement. This structured evaluation framework helps identify strengths and areas for refinement.

### Context Similarity

Context Similarity assesses how well the visuals, audio, and subtitles align with the summarized content. A high rating indicates strong coherence, where all elements accurately reflect the news topic, while a low rating suggests inconsistencies or weak contextual relevance.

Based on the human evaluation of 1,000 people, 77% of the votes indicate high satisfaction (5-star rating), while 23% represent unfavorable conditions, including any rating lower than 5 stars.

### Audio Clarity

Audio Clarity measures the intelligibility and naturalness of the generated speech. A high score signifies clear, well-pronounced, and noise-free audio, whereas a low rating indicates distortion, robotic speech, or background interference affecting comprehension.

Based on the human evaluation of 1,000 people, 73% of the votes indicate high satisfaction (5-star rating), while 27% represent unfavorable conditions, including any rating lower than 5 stars.

### Subtitle Readablility

Subtitle Readability evaluates the clarity, size, and synchronization of captions with the spoken content. A high score means well-timed, easily readable text, while a lower rating suggests misalignment, difficult-to-read fonts, or poor contrast.

Based on the human evaluation of 1,000 people, 63% of the votes indicate high satisfaction (5-star rating), while 47% represent unfavorable conditions, including any rating lower than 5 stars.

### Background Image Suitability

Background Image Suitability examines the relevance and visual appeal of selected images in relation to the news category. A high rating indicates appropriate, engaging visuals, whereas a lower score reflects mismatched or low-quality imagery that reduces contextual accuracy.

Based on the human evaluation of 1,000 people, 45% of the votes indicate high satisfaction (5-star rating), while 55% represent unfavorable conditions, including any rating lower than 5 stars.

**Attention Capturing Capability**

Attention Capturing Capability measures the video's ability to engage viewers through dynamic presentation, pacing, and visual appeal. A high score indicates a compelling and visually stimulating reel, while a lower rating suggests a dull, unengaged, or poorly structured video.

Based on the human evaluation of 1,000 people, 51% of the votes indicate high satisfaction (5-star rating), while 49% represent unfavorable conditions, including any rating lower than 5 stars.

**Overall Quality Of The Reel**

Overall Quality of the Reel provides a comprehensive assessment of all factors combined. A high rating signifies a well-executed, engaging, and high-quality output, while lower scores indicate areas needing improvement, such as weak synchronization, poor visuals, or low engagement.

The Overall Reel Quality Score is calculated using a weighted sum approach. Each evaluation criterion's favorable percentage is multiplied by its assigned weight, and the results are summed to derive the final score:

$$\text{Overall Quality Score} = \sum \left( \frac{\text{Weight}}{100} \times \text{Favorable Percentage} \right) \tag{4.10}$$

This formula ensures that more critical criteria contribute proportionally to the final assessment, providing a balanced evaluation of the video quality.

| **Criteria** | **Weight** | **Favorable** | **Unfavorable** | **Favorable Score** | **Unfavorable Score** |
|---|---|---|---|---|---|
| Context Similarity | 25% | 77% | 23% | 19.25 | 5.75 |
| Audio Clarity | 25% | 73% | 27% | 18.25 | 6.75 |
| Subtitle Readability | 25% | 63% | 47% | 15.75 | 11.75 |
| Background Image Suitability | 15% | 45% | 55% | 6.75 | 8.25 |
| Attention Capturing Capability | 10% | 51% | 49% | 5.1 | 4.9 |
| **Total Quality Score** | 100% | | | 65.1 | 34.9 |

Table 4.5: Overall Reel Quality Matrix

Based on the previous human evaluation results of 1,000 people, The overall reel quality is calculated as 65.1% of the votes indicate high satisfaction (5-star rating), while 34.9% represent unfavorable conditions, including any rating lower than 5 stars.

The evaluation highlights strong performance in Audio Clarity and Context Similarity, with most ratings concentrated at 5, indicating high coherence and intelligibility. Background Image Suitability and Subtitle Readability also received positive feedback, though some minor issues persist. Attention Capturing Capability shows a wider distribution, suggesting room for improvement in engagement. Overall, while the reel maintains a high-quality standard, enhancing engagement strategies and refining visual elements could further optimize viewer experience.

## 4.3 Outputs


Figure 4.1: Sign In Page.


Figure 4.2: Home Page.


Figure 4.3: PDF Processing.

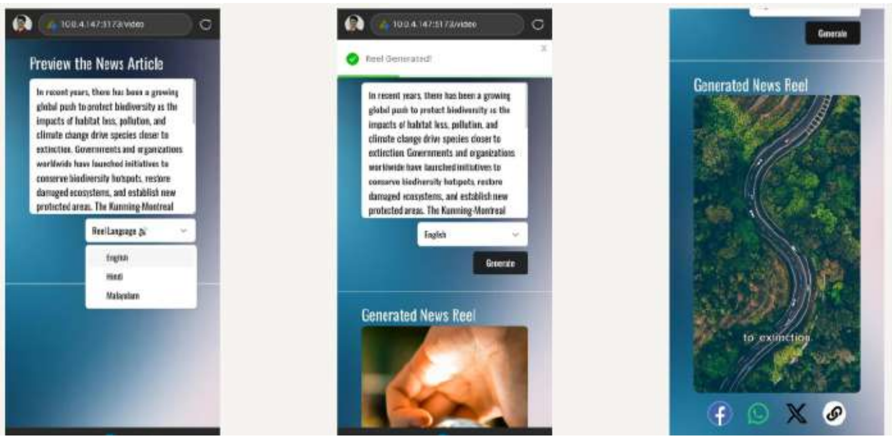



Figure 4.4: Reel Generation Page.

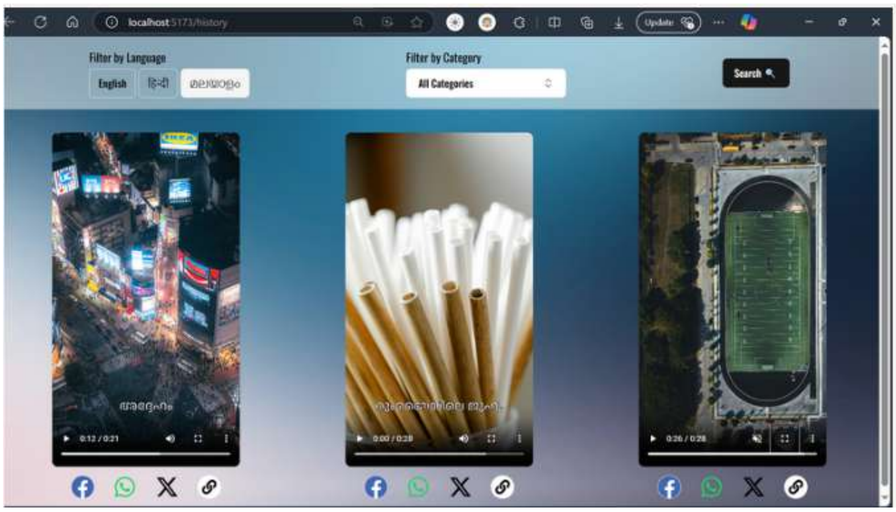



Figure 4.5: Reel History Page.

## 4.4 Summary

This chapter discussed the analysis of the OCR, summarization, classification, and video generation modules developed in the project. For the OCR module, we found that Gemini-2.0-Flash outperformed the other systems we tested in terms of accuracy, and significantly reduced the number of character and word errors. BART trained on SamSUM was used in our summarization module, which showed strong fluency and structure, and high precision. Although, in terms of recall, lower values mean some relevant information was sometimes left out, and as a result, the module could improve its ability to extract all important content. BERT trained on IPTC News was used for our classification module, which showed strong performance with defined categories like Weather and Sport, but recall values were not as strong for subjective categories like Politics and Society.

The video generation module worked well by merging the generated summaries with appropriate stock images found via the Pexels API by using keywords extracted by KeyBERT. Speech synthesis was accomplished using gTTS, and subtitles were auto-generated. The videos were assessed across several dimensions, including context matching, audio clarity, subtitle readability, background image matching, attention-capturing capability, and overall quality. Overall, the system produced visually-engaging and coherent videos; however, there continues to be areas for potential improvement, such as enhancing image selection relevance and subtitle synchronization.

The strengths of the system are showcased by these results, while also revealing opportunities for enhancements, including improving classification recall, improving summarization while still maintaining strengths of key details, and improving video generation for better continuity. Future work should focus on these optimizations in all modules to substantially boost performance on all fronts.

# Chapter 5

# Conclusions

The "NewsFlix" initiative provides a mobile solution to the traditional print media industry with respect to keeping up with modern-day audiences. As discussed in Chapter 1, the system has addressed lower interest in newspapers with younger generations, the visually impaired, and audiences with a short attention span. This initiative creates impressive multimedia videos of articles, to maximize audience reach and retention in relation to reading the articles

In chapter 2, technological frameworks and previous research will be discussed that support the ideas that contributed to the development of this project. An example is the incorporation of the modern text summarization models like BERT alongside newer video synthesizing methods. These advancements will support contextual accuracy and greatly engaging output to unify the traditional vs. digital media consumption sphere.

Chapter 3 addresses the system's architecture in which the overall system, including its distinct components i.e. Article Segmentation Module, OCR Module, Summarization Module, Classifier, and Video Generator are introduced. Each module is intended to serve the purpose of automating each of its converting process, ensuring precision in the extraction of the civic text, orderly organization, and motivated video outputs. Through the use of APIs, such as Pexels, for the multimedia sources, and useful approaches in the project it is shown to build previous promise of utility in mass applications. The features of; live streaming, multi-lingual, and personalized recommendations will give weight to validating "NewsFlix" as a significant game changer to the future of news delivery systems.

Chapter 4 presents the outcome of the examination related to the project and explores elements of performance and key findings obtained from this project. It discusses how effective the system was with OCR, summarization, and classification, and highlights both the system's merits and areas for future improvements. The Gemini-2.0-Flash system achieved better results using the OCR process than conventional systems, operating with higher confidence and lower mistake rates. The summarization module built using BART, and based on training data from the SamSUM corpus, achieved high fluency and coherence attributes in its processing, with even higher relevance and accuracy attributes in the summaries it produced. Although the recall was lower, the BART model was successful at producing coherent summaries that were easily read, improving the content generation task in short formats of content. The classification model built using BERT and based on IPTC News training data, demonstrated structure across clear categories like Weather and Sport, but was inconsistent in recall with less obvious categories, like Politics and Society. There are opportunities to improve recall, improve categorization, and expand the diversity of training data to improve accuracy and robustness in future work. Improving accuracy, robustness, and recall can be part of future work.

The "NewsFlix" initiative aims to improve accessibility to news for people by redesigning articles into more dynamic forms in the form of a video. It uses cutting edge Optical Character Recognition (OCR), summarization, and classification to extract text to video without losing the original article's content. The performance evaluations demonstrate good accuracy but could possibly be better through improved recall, classification, and multilingual support.

# References

[1] T. Ma, Q. Pan, H. Rong, Y. Qian, Y. Tian, and N. Al-Nabhan, "T-bertsum: Topic-aware text summarization based on bert," *IEEE Transactions on Computational Social Systems*, vol. 9, 2022.

[2] Y. Liu, Z. Xie, and H. Liu, "An adaptive and robust edge detection method based on edge proportion statistics," *IEEE Transactions on Image Processing*, vol. 29, 2020.

[3] Y. S. Chernyshova, A. V. Sheshkus, and V. V. Arlazarov, "Two-step cnn framework for text line recognition in camera-captured images," *IEEE Access*, vol. 8, 2020.

[4] X. Chen, P. Cong, and S. Lv, "A long-text classification method of chinese news based on bert and cnn," *IEEE Access*, vol. 10, 2022.

[5] J. Xu, W. Ding, and H. Zhao, "Based on improved edge detection algorithm for english text extraction and restoration from color images," *IEEE Sensors Journal*, vol. 20, no. 20, pp. 11 951–11 958, 2020.

[6] J. Park, E. Lee, Y. Kim, I. Kang, H. I. Koo, and N. I. Cho, "Multi-lingual optical character recognition system using the reinforcement learning of character segmenter," *IEEE Access*, vol. 8, pp. 174 437–174 448, 2020.

[7] D. Coquenet, C. Chatelain, and T. Paquet, "End-to-end handwritten paragraph text recognition using a vertical attention network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 508–524, 2023.

[8] W. Liu, Y. Gao, J. Li, and Y. Yang, "A combined extractive with abstractive model for summarization," *IEEE Access*, vol. 9, pp. 43 970–43 980, 2021.

[9] S. Ghodratnama, A. Beheshti, M. Zakershahrak, and F. Sobhanmanesh, "Extractive document summarization based on dynamic feature space mapping," *IEEE Access*, vol. 8, pp. 139 084–139 095, 2020.

[10] A. Dilawari, M. U. G. Khan, S. Saleem, Zahoor-Ur-Rehman, and F. S. Shaikh, "Neural attention model for abstractive text summarization using linguistic feature space," *IEEE Access*, vol. 11, pp. 23 557–23 564, 2023.

[11] J. Zheng and L. Zheng, "A hybrid bidirectional recurrent convolutional neural network attention-based model for text classification," *IEEE Access*, vol. 7, pp. 106 673–106 685, 2019.

[12] F. M. A. S. Ahmad, M. Z. Asghar, and S. Khan, "Classification of poetry text into the emotional states using deep learning technique," *IEEE Access*, vol. 8, pp. 73 865–73 878, 2020.

[13] Q. Chen, Q. Wu, J. Chen, Q. Wu, A. van den Hengel, and M. Tan, "Scripted video generation with a bottom-up generative adversarial network," *IEEE Transactions on Image Processing*, vol. 29, 2020.

# Appendix A: Presentation

External Evaluation | 08/04/2025

# NEWSFLIX

**PRESENTED BY TEAM 1**


## THE TEAM

| | |
|---|---|
| **Abhinav Sobi** | **U2103008** |
| **Alan Joseph** | **U2103021** |
| **Basil Eldho Joseph** | **U2103057** |
| **Daniel Robin** | **U2103072** |

**Guide**
**Dr. Mary Priya Sebastian**

THE
# PRESENTATION *FOCUS*



01

# PROBLEM DEFINITION

To develop a system that can automatically transform news articles into short, engaging videos by incorporating visuals and captions.


02

# PURPOSE & NEED

Traditional newspaper formats are unfavorable for elderly and for those with visual impairments, limiting accessibility and reach of the news.

News providers can't satisfy the growing demand for digital news content, creating engaging & quick news reels is labor-intensive and time consuming.

# OBJECTIVE

Develop a fully automated system using NLP and CV:

- Transform scanned newspaper articles into engaging, summarized videos.
- From the article extract text, summarize & classify the news.
- Generate video with relevant footage, audio of news summary, and synchronized subtitles.
- Implement a multi-view layout for consuming multiple news videos.

03

# LITERATURE SURVEY

| PAPER | ADVANTAGES | DISADVANTAGES |
|---|---|---|
| Y. Liu et al.[1] An Adaptive and Robust Edge Detection Method Based on Edge Proportion Statistic (2020) | • Clear single pixel edges<br>• Edges without breakages | • Time consuming<br>• Needs high computational power |
| Y. S. Chernyshova et al.[2] Two-Step CNN Framework for Text Line Recognition in Camera-Captured Images (2020) | • Multilingual segmentation<br>• Light-weight and fast | • Single-Lang recognition<br>• Work only for line texts |

04

05

| PAPER | ADVANTAGES | DISADVANTAGES |
|---|---|---|
| X. Chen et al.[3] A Long-Text Classification Method of Chinese News Based on BERT and CNN (2022) | • Classifies multilingual text<br>• Identify local features | • Complete input not considered<br>• Full potential of BERT not explored |
| T. Ma et al.[4] T-BERTSum: Topic-Aware Text Summarization Based on BERT (2022) | • High Accuracy<br>• Pre-Trained<br>• Understands Context | • Complex architecture<br>• Sensitive to input quality |
| Q. Chen et al.[5] Scripted Video Generation With a Bottom-Up Generative Adversarial Network (2020) | • Good performance across datasets | • Training instability<br>• Complex architecture |

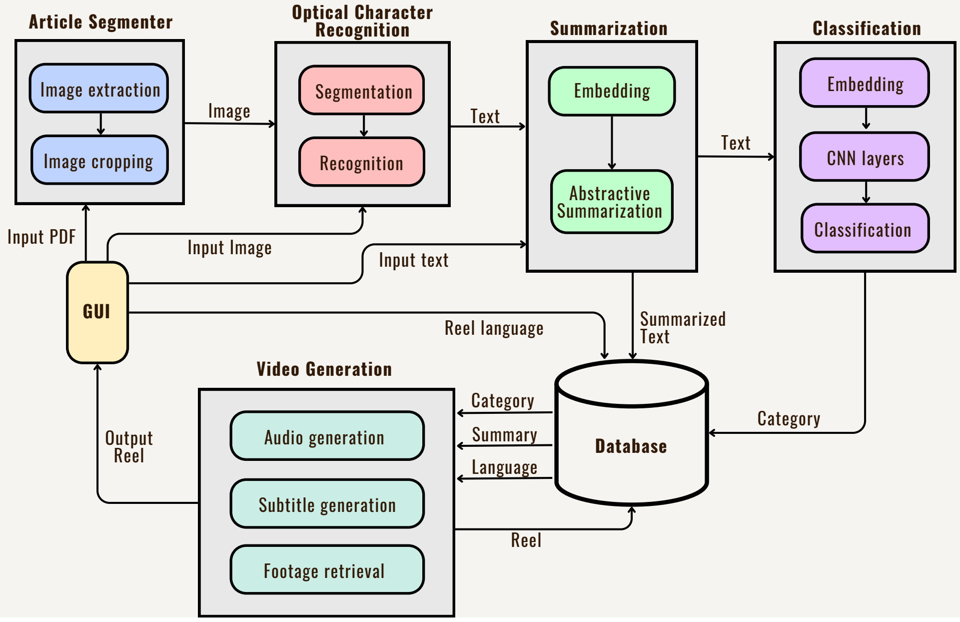
# PROPOSED METHOD

A SaaS that automates the creation of engaging short videos from newspaper articles, making news more accessible and appealing to the modern audience using NLP and AI technologies.

06

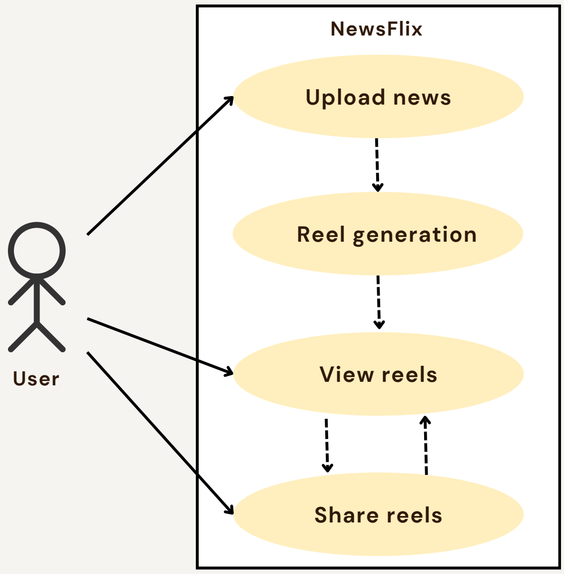
07

# ARCHITECTURE DIAGRAM



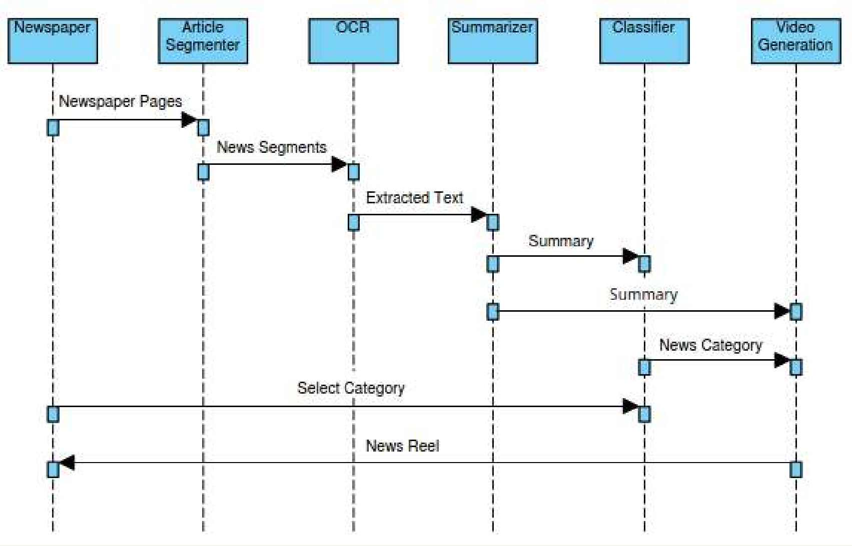
08

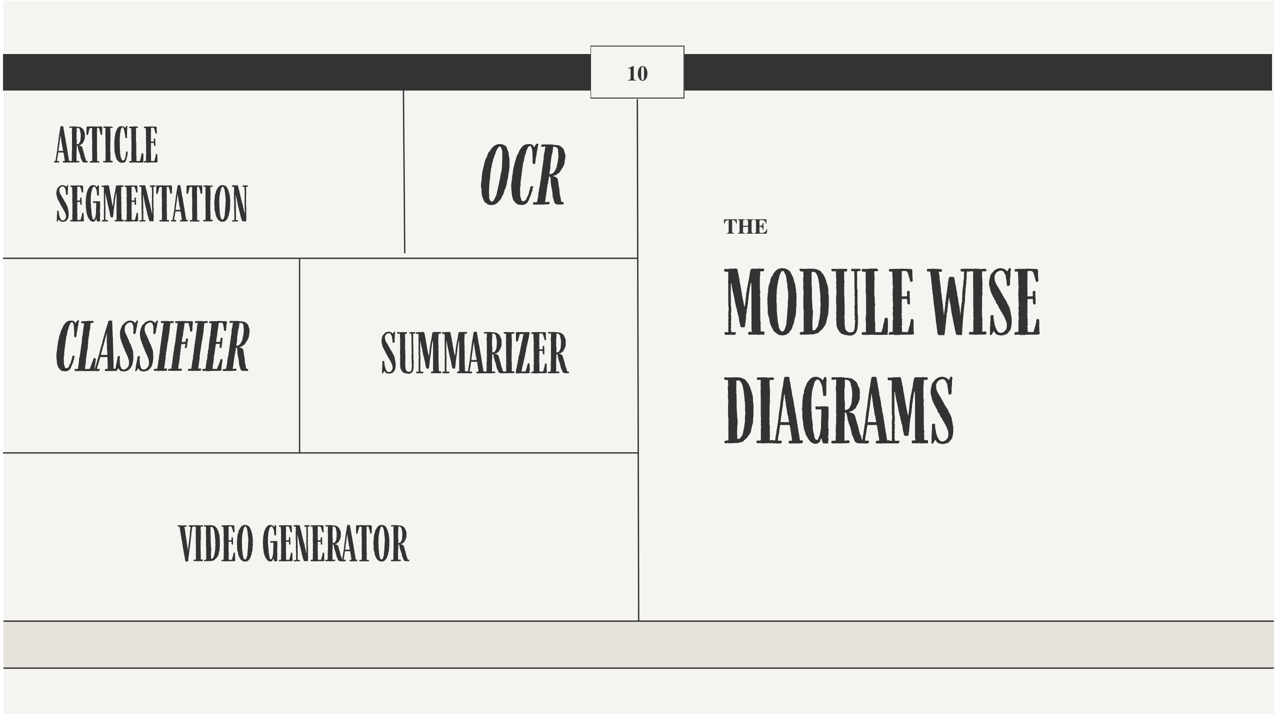
# USE CASE DIAGRAM

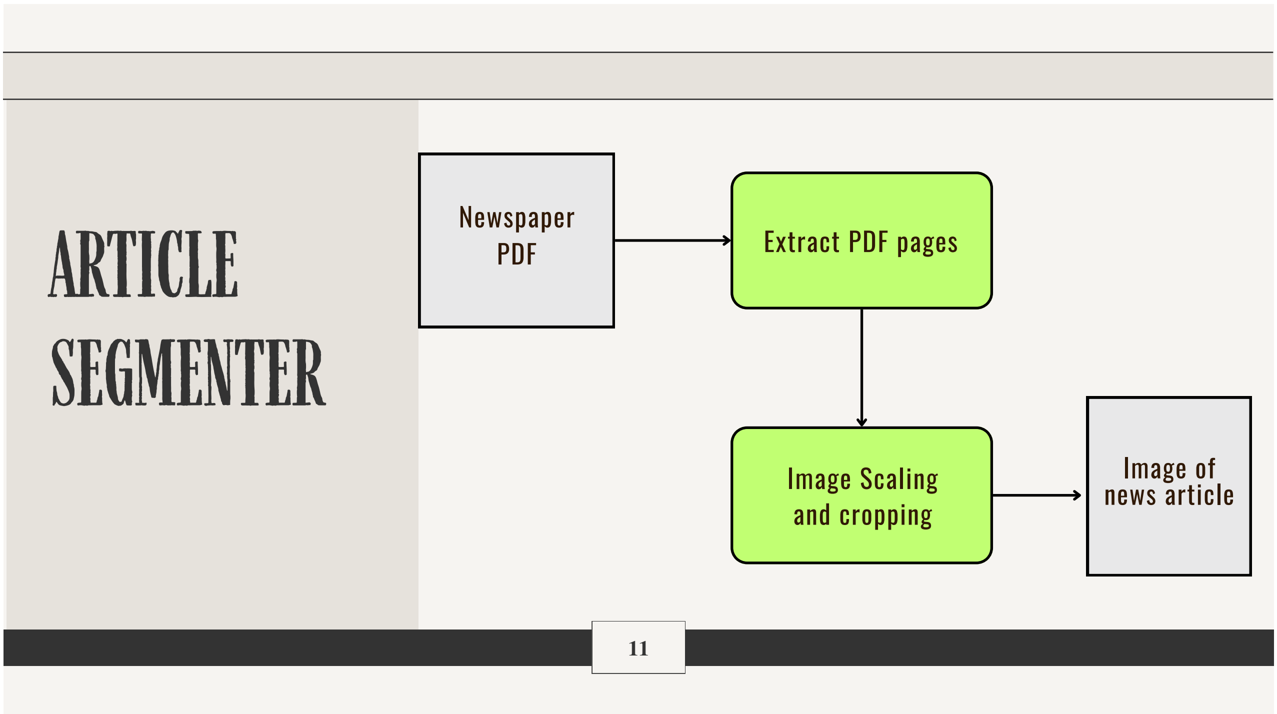



09

# SEQUENCE DIAGRAM




10
ARTICLE SEGMENTATION
OCR
CLASSIFIER
SUMMARIZER
VIDEO GENERATOR
THE
MODULE WISE DIAGRAMS


ARTICLE SEGMENTER
Newspaper PDF
Extract PDF pages
Image Scaling and cropping
Image of news article
11

## ARTICLE SEGMENTER

Input: PDF file uploaded by the user

Processing the PDF

- Extracts images from PDF pages using PDF.js and displays them in a carousel slider.

Cropping an Image

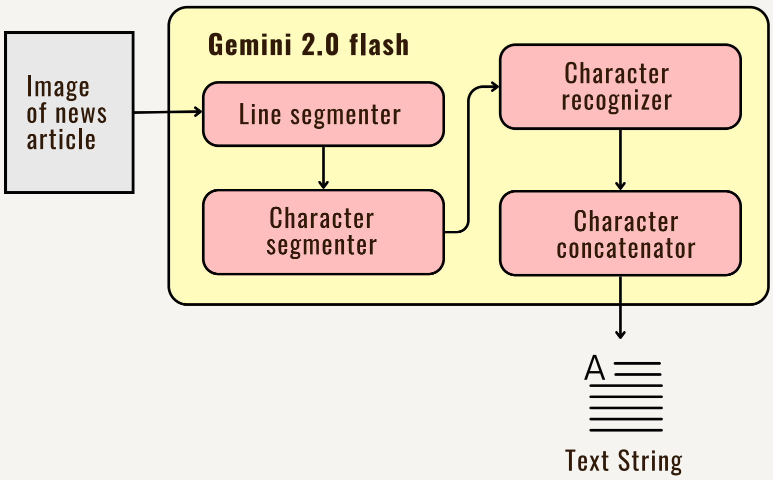
- User selects an image, opens a modal, and draws a crop box.
- Cropped image is extracted using a canvas and displayed.

Sending to OCR Module

- Cropped image is converted to a Blob, then a File object which is then sent to the OCR module

12

## OCR MODULE



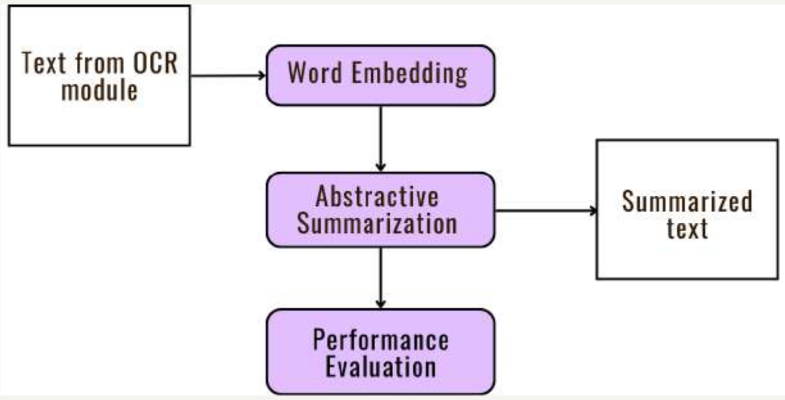

13

# OCR MODULE

- Input: Output sub-image from the segmenter module.
- Gemini 2.0 flash is used for text extraction.
- Line segmentation using baseline and capline.
- Character segmentation using a language independent CNN.
- Character recognition using a separate CNN.
- All characters are concatenated to get the output text.

14

# SUMMARIZER



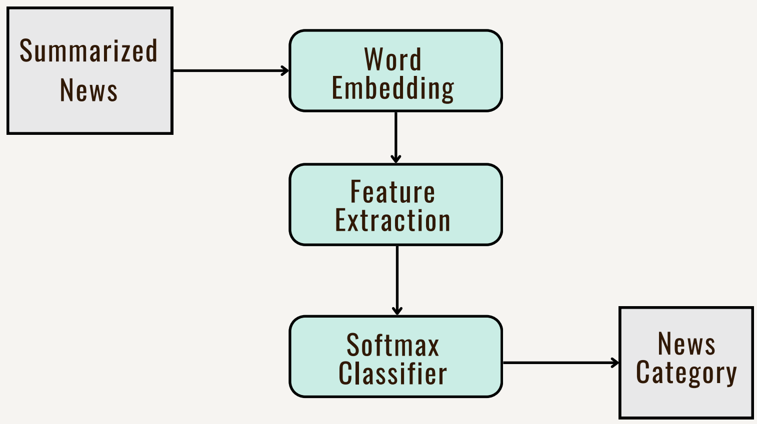

15

## SUMMARIZER

- Input: Text recognized by OCR module
- BART tokenizes input text and converts the tokens into word embeddings.
- Uses a bi-directional encoder (like BART) and a left-to-right decoder (like GPT).
- Implements Abstractive Summarization. Instead of copying sentences, BART rephrases and compresses the text into a coherent summary.
- Fine-tuned on SAMSum dataset.
- Output: Summarized text of each article

16

## CLASSIFIER



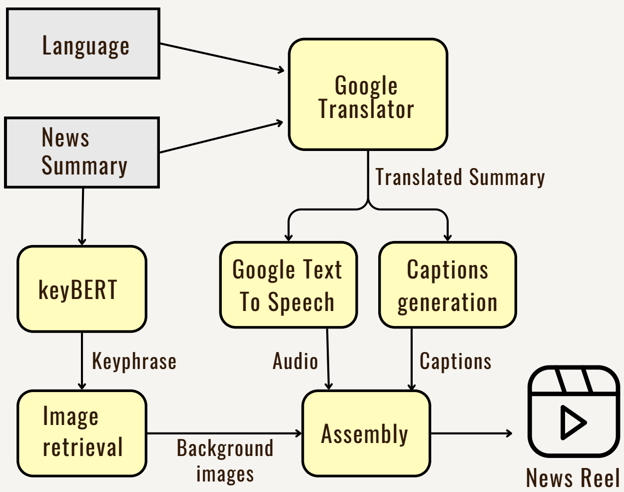

17

# CLASSIFIER

- Input: Summarized text of each article
- Using BERT model, word embedding is done. Feature extraction is also done and its information is combined with the word embedding.
- The feature vector obtained is now classified using a classifier( Softmax or Sigmoid ).
- Output: All articles are classified into their respective classes

18

# VIDEO GENERATOR




19

# VIDEO GENERATION

- Input: Summarized text & category of news article.
- Footage & Images for the video are retrieved using pexels API based on the category.
- From summarized text, voice & subtitles are generated & synced using gTTS & moviepy.
- All the obtained outputs are assembled and synchronized to obtain the final video.
- Output: An engaging, short video about the news article.

20

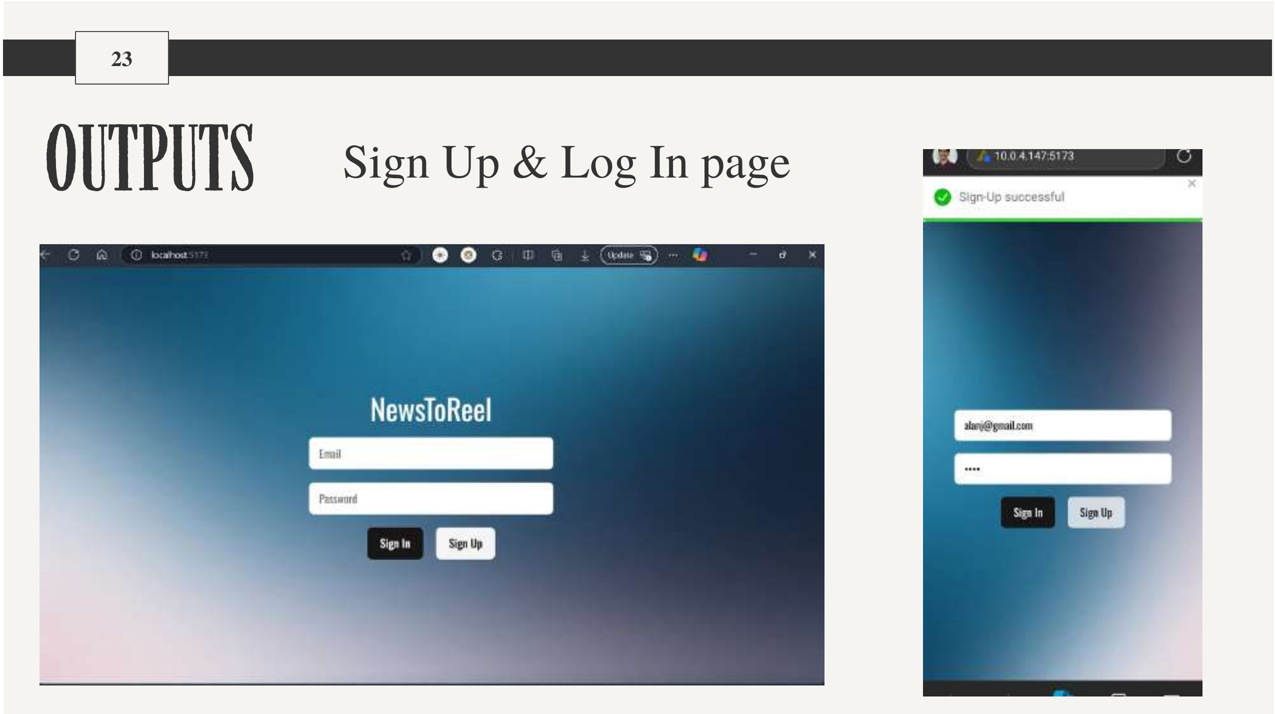
# VIDEO GENERATION

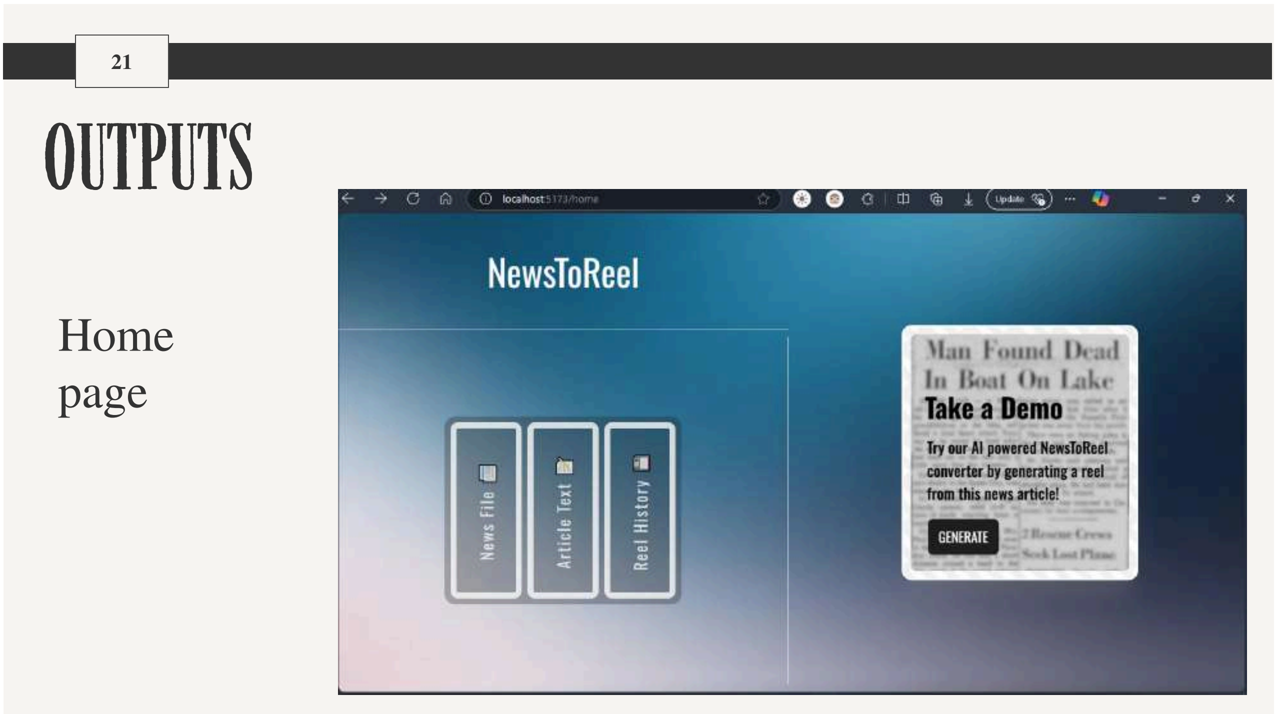
**Algorithm 1** News Reel Generation

**Input:** *Summary* and *Category* of news article, Required output *language*

```
Extract a keyphrase from the Summary.
if language is not english then
    Translate the Summary to language.
end if
Convert Summary text to audio using gTTS, save it as reel.mp3.
Speed up reel.mp3 by 10% to reduce duration.
Declare duration as duration of reel.mp3.
Initialize video with 1080x1920p resolution.
Calculate no.of images, n ← ⌈duration/7⌉
for i = 1 to n do
    Fetch ith image under keyphrase using Pexels API.
    Download the image as i.jpg.
    Append i.jpg as background of video for 7 seconds.
end for
Transcribe reel.mp3 using Whisper to get synchronized captions.
Add captions to video using the font suitable for language.
Set reel.mp3 as the audio for video.
Write video at 1 fps and save it as reel.mp4.
```

**Output:** News video file reel.mp4

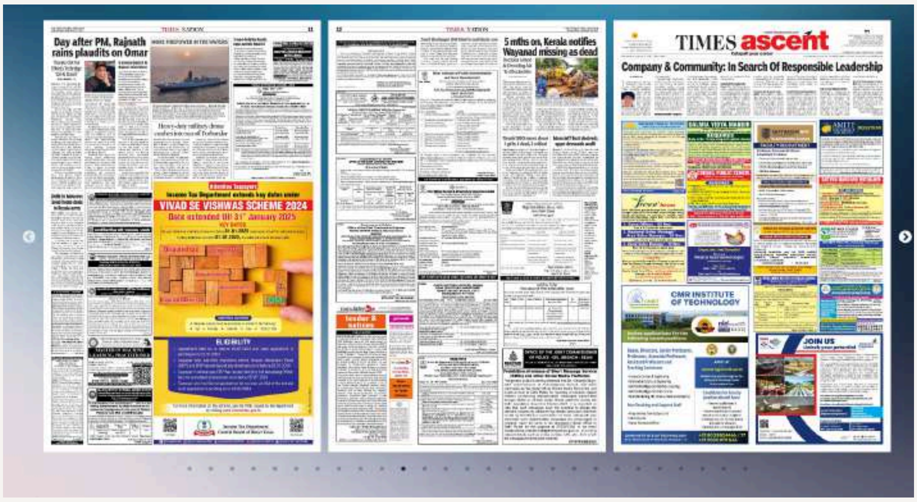

23
OUTPUTS
Sign Up & Log In page
NewsToReel
Email
Password
Sign In
Sign Up
10.0.4.147:5173
Sign-Up successful
alanj@gmail.com
Sign In
Sign Up


21
OUTPUTS
Home page
localhost:5173/home
NewsToReel
News File
Article Text
Reel History
Man Found Dead In Boat On Lake
Take a Demo
Try our AI powered NewsToReel converter by generating a reel from this news article!
GENERATE

21

# OUTPUTS

Pages for user selection




21

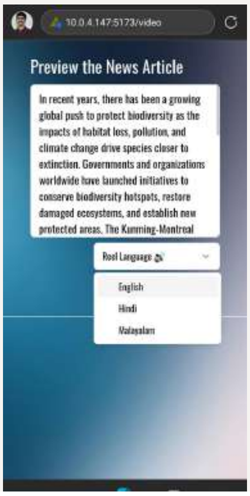
# OUTPUTS

PDF processing page

22

# OUTPUTS

## Reel generation page

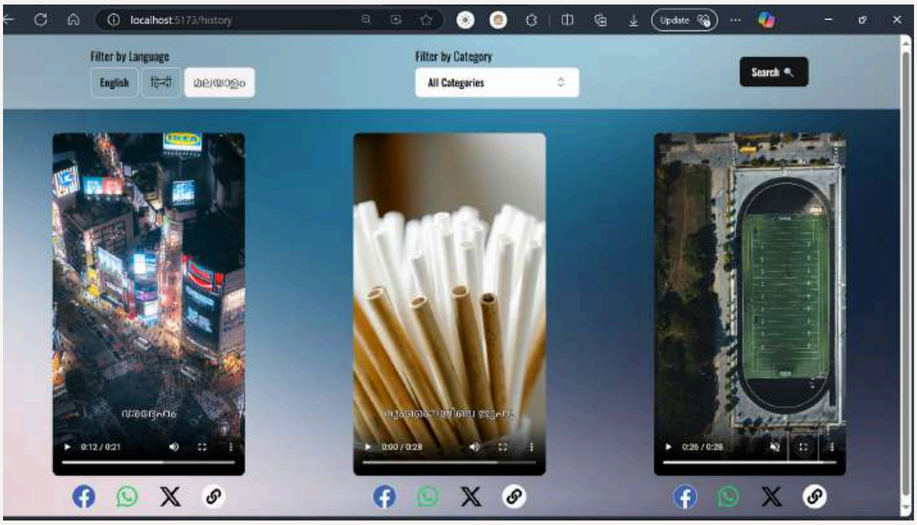





22

# OUTPUTS

Reel history page



27

# EVALUATION RESULTS

## OCR

| Model / System | PCR % | CER % | WER % |
|---|---|---|---|
| Tesseract 3.04 | 90.08 | 9.92 | 17.36 |
| Tesseract 4.00 | 91.13 | 8.87 | 15.52 |
| FineReader Engine 11 | 88.40 | 11.60 | 20.30 |
| Gemini-2.0-Flash | 96.21 | 3.79 | 6.63 |

TABLE I: OCR Performance Comparison

27

# EVALUATION RESULTS

## Summarizer

```
ROUGE scores for Summary 2: {'rouge1': Score(precision=0.9651162790697675, recall=0.15930902111324377, f
measure=0.27347611202635913), 'rouge2': Score(precision=0.8352941176470589, recall=0.13653846153846153,
fmeasure=0.23471074380165288), 'rougeL': Score(precision=0.9418604651162791, recall=0.15547024952015356,
fmeasure=0.26688632619439867)}
```

27

# EVALUATION RESULTS

## Classifier

| | Micro-F1 | Macro-F1 | Accuracy | No. of instances |
|---|---|---|---|---|
| All (combined) | 0.734278 | 0.745864 | 0.734278 | 1129 |

Performance per label:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| arts, culture, entertainment and media | 0.602151 | 0.875 | 0.713376 | 64 |
| conflict, war and peace | 0.611111 | 0.916667 | 0.733333 | 36 |
| crime, law and justice | 0.861538 | 0.811594 | 0.835821 | 69 |
| disaster, accident and emergency incident | 0.691176 | 0.886792 | 0.77686 | 53 |
| economy, business and finance | 0.779221 | 0.508475 | 0.615385 | 118 |
| education | 0.847458 | 0.735294 | 0.787402 | 68 |
| environment | 0.589041 | 0.754386 | 0.661538 | 57 |
| health | 0.79661 | 0.79661 | 0.79661 | 59 |
| human interest | 0.552239 | 0.672727 | 0.606557 | 55 |
| labour | 0.855072 | 0.830986 | 0.842857 | 71 |
| lifestyle and leisure | 0.773585 | 0.476744 | 0.589928 | 86 |
| politics | 0.568182 | 0.735294 | 0.641026 | 68 |
| religion | 0.842105 | 0.941176 | 0.888889 | 51 |
| science and technology | 0.637681 | 0.8 | 0.709677 | 55 |
| society | 0.918033 | 0.5 | 0.647399 | 112 |
| sport | 0.824324 | 0.968254 | 0.890511 | 63 |
| weather | 0.953488 | 0.931818 | 0.942529 | 44 |

28

# WORK DONE

Article Segmentation - Basil Eldho Joseph

News Summarizer - Daniel Robin

News Classifier - Abhinav Sobi

Video Generation - Alan Joseph

User Interface - Alan Joseph, Basil Eldho Joseph

29

# ASSUMPTIONS

- **Clear Input Image:** Effectiveness depends on clarity and quality of the input images. Distorted or low-resolution images may lead to inaccurate results.
- **Genuineness of Provided News:** Authenticity of the input news content is not validated by the system. Misinformation or fabricated news can skew the results.
- **No Copyright Issues:** We assume that all materials used in this news article or text data are free from copyright restrictions, ensuring that we can share and discuss the content without legal complications.

30

# REQUIREMENTS

| Software Requirements | Hardware Requirements |
|---|---|
| • Visual Studio Code<br>• FireBase<br>• PgAdmin 4<br>• Postman<br>• WebBrowser | • Intel Core i5 or higher<br>• Nvidia GTX 1650<br>• 8 GB RAM<br>• Screen with a resolution of 1280x720<br>• Network connectivity |

31

# GANTT CHART


# BUDGET

**CLOUD INFRASTRUCTURE**

- Cloud Hosting : For video storage, backend hosting, and user data storage.

RS.5000

**THIRD-PARTY TOOLS**

- Optical Character Recognition : Using APIs like Google Cloud Vision or AWS Textract.

RS.4500

32

# RISKS & CHALLENGES

- Segmenting complex article layouts
- Extracting paragraph texts
- Handling multiple languages
- Video Generation complexity

33

34

# OUTCOMES

- Segmenter outputs sub-images containing individual news articles.
- OCR module extracts news article text from the sub-images.
- Summarizer produces summarized news fit for reel content.
- Classifier assigns news category like sports, politics, etc. to the article text.
- Video assembler uses the outputs so far to generate the final reel.
- Final outcome: Users should be able to view and share the generated short & engaging news reels via a mobile friendly web app.
- Technical paper about the project published in a reputed journal.

# CONCLUSION

A software solution that converts newspaper articles into engaging short videos, making news more accessible to a wider audience.

Utilizes advanced NLP and Image Processing techniques to automatically generate audio summaries, select relevant visuals, and produce high-quality, captivating video content that brings news to life.

35

# REFERENCES


- Y. S. Chernyshova, A. V. Sheshkus and V. V. Arlazarov, "*Two-Step CNN Framework for Text Line Recognition in Camera-Captured Images*," in IEEE Access, vol. 8, pp. 32587-32600, February 2020
- Y. Liu, Z. Xie and H. Liu, "An Adaptive and Robust Edge Detection Method Based on Edge Proportion Statistics," in IEEE Transactions on Image Processing, vol. 29, pp. 5206-5215, March 2020
- T. Ma, Q. Pan, H. Rong, Y. Qian, Y. Tian and N. Al-Nabhan, "*T-BERTSum: Topic-Aware Text Summarization Based on BERT,*" in IEEE Transactions on Computational Social Systems, vol. 9, no. 3, pp. 879-890, June 2022
- 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, July 2020
- Q. Chen, Q. Wu, J. Chen, Q. Wu, A. van den Hengel and M. Tan, "Scripted Video Generation With a Bottom-Up Generative Adversarial Network," in IEEE Transactions on Image Processing, vol. 29, pp. 7454-7467, May 2020


36

# REFERENCES


- J. Zheng and L. Zheng, "A Hybrid Bidirectional Recurrent Convolutional Neural Network Attention-Based Model for Text Classification," in IEEE Access, vol. 7, pp. 106673-106685, March 2019
- K. M. Hasib et al., "MCNN-LSTM: Combining CNN and LSTM to Classify Multi-Class Text in Imbalanced News Data," in IEEE Access, vol. 11, pp. 93048-93063, June 2023
- J. Park, E. Lee, Y. Kim, I. Kang, H. I. Koo and N. I. Cho, "Multi-Lingual Optical Character Recognition System Using the Reinforcement Learning of Character Segmenter," in IEEE Access, vol. 8, pp. 174437-174448, Jan 2020
- 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, July 2020
- Q. Chen, Q. Wu, J. Chen, Q. Wu, A. van den Hengel and M. Tan, "Scripted Video Generation With a Bottom-Up Generative Adversarial Network," in IEEE Transactions on Image Processing, vol. 29, pp. 7454-7467, May 2020




# THANK YOU

# Appendix B: Vision, Mission, Programme Outcomes and Course Outcomes

# Vision, Mission, Programme Outcomes and Course Outcomes

## Institute Vision

To evolve into a premier technological institution, moulding eminent professionals with creative minds, innovative ideas and sound practical skill, and to shape a future where technology works for the enrichment of mankind.

## Institute Mission

To impart state-of-the-art knowledge to individuals in various technological disciplines and to inculcate in them a high degree of social consciousness and human values, thereby enabling them to face the challenges of life with courage and conviction.

## Department Vision

To become a centre of excellence in Computer Science and Engineering, moulding professionals catering to the research and professional needs of national and international organizations.

## Department Mission

To inspire and nurture students, with up-to-date knowledge in Computer Science and Engineering, ethics, team spirit, leadership abilities, innovation and creativity to come out with solutions meeting societal needs.

## Programme Outcomes (PO)

Engineering Graduates will be able to:

**1. Engineering Knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

**2. Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

**3. Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

**4. Conduct investigations of complex problems**: Use research-based knowledge including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

**5. Modern Tool Usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

**6. The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

**7. Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

**8. Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

**9. Individual and Team work**: Function effectively as an individual, and as a member or leader in teams, and in multidisciplinary settings.

**10. Communication**: Communicate effectively with the engineering community and with society at large. Be able to comprehend and write effective reports documentation. Make effective presentations, and give and receive clear instructions.

**11. Project management and finance**: Demonstrate knowledge and understanding of engineering and management principles and apply these to one's own work, as a member and leader in a team. Manage projects in multidisciplinary environments.

**12. Life-long learning**: Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

### Programme Specific Outcomes (PSO)

A graduate of the Computer Science and Engineering Program will demonstrate:

**PSO1: Computer Science Specific Skills**
The ability to identify, analyze and design solutions for complex engineering problems

in multidisciplinary areas by understanding the core principles and concepts of computer science and thereby engage in national grand challenges.

**PSO2: Programming and Software Development Skills**
The ability to acquire programming efficiency by designing algorithms and applying standard practices in software project development to deliver quality software products meeting the demands of the industry.

**PSO3: Professional Skills**
The ability to apply the fundamentals of computer science in competitive research and to develop innovative products to meet the societal needs thereby evolving as an eminent researcher and entrepreneur.

**Course Outcomes (CO)**

After the completion of the course the student will be able to:

**Course Outcome 1:** Model and solve real world problems by applying knowledge across domains (Cognitive knowledge level: Apply).

**Course Outcome 2:** Develop products, processes or technologies for sustainable and socially relevant applications (Cognitive knowledge level: Apply).

**Course Outcome 3:** Function effectively as an individual and as a leader in diverse teams and to comprehend and execute designated tasks (Cognitive knowledge level: Apply).

**Course Outcome 4:** Plan and execute tasks utilizing available resources within timelines, following ethical and professional norms (Cognitive knowledge level: Apply).

**Course Outcome 5:** Identify technology/research gaps and propose innovative/creative solutions (Cognitive knowledge level: Analyze).

**Course Outcome 6:** Organize and communicate technical and scientific findings effectively in written and oral forms (Cognitive knowledge level: Apply).

# Appendix C:
# CO-PO-PSO Mapping

### Course Outcomes

After completion of the course the student will be able to:

| SL.NO | Description | Bloom's Taxonomy Level |
|---|---|---|
| CO1 | Model and solve real-world problems by applying knowledge across domains. | Level 3: Apply |
| CO2 | Develop products, processes, or technologies for sustainable and socially relevant applications. | Level 3: Apply |
| CO3 | Function effectively as an individual and as a leader in diverse teams to comprehend and execute designated tasks. | Level 3: Apply |
| CO4 | Plan and execute tasks utilizing available resources within timelines, following ethical and professional norms. | Level 3: Apply |
| CO5 | Identify technology/research gaps and propose innovative/creative solutions. | Level 4: Analyze |
| CO6 | Organize and communicate technical and scientific findings effectively in written and oral forms. | Level 3: Apply |

### CO-PO Mapping

| CO | PO 1 | PO 2 | PO 3 | PO 4 | PO 5 | PO 6 | PO 7 | PO 8 | PO 9 | PO 10 | PO 11 | PO 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 2 | 1 | 1 | - | 2 | 1 | - | - | - | - | 3 |
| 2 | 3 | 3 | 2 | 3 | - | 2 | 1 | - | - | - | - | 3 |
| 3 | 3 | 2 | - | - | 3 | - | - | 1 | - | 2 | - | 3 |
| 4 | 3 | - | - | - | 2 | - | - | 1 | - | 3 | - | 3 |
| 5 | 3 | 3 | 3 | 3 | 2 | 2 | - | 2 | - | 3 | - | 3 |

### CO-PSO Mapping

| CO | PSO 1 | PSO 2 | PSO 3 |
|---|---|---|---|
| 1 | 3 | 1 | 2 |
| 2 | 3 | 2 | 2 |
| 3 | 2 | 2 | - |
| 4 | 3 | - | 3 |
| 5 | 3 | - | - |

**Justification for CO-PO Mapping**

| Mapping | Level | Justification |
|---|---|---|
| 101003/CS722U.1-PO1 | M | Knowledge in the area of technology for project development using various tools results in better modeling. |
| 101003/CS722U.1-PO2 | M | Knowledge acquired in the selected area of project development can be used to identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions. |
| 101003/CS722U.1-PO3 | M | Can use the acquired knowledge in designing solutions to complex problems. |
| 101003/CS722U.1-PO4 | M | Can use the acquired knowledge in designing solutions to complex problems. |
| 101003/CS722U.1-PO5 | H | Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions. |
| 101003/CS722U.1-PO6 | M | Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices. |
| 101003/CS722U.1-PO7 | M | Project development based on societal and environmental context solution identification is the need for sustainable development. |
| 101003/CS722U.1-PO8 | L | Project development should be based on professional ethics and responsibilities. |
| 101003/CS722U.1-PO9 | L | Project development using a systematic approach based on well-defined principles will result in teamwork. |
| 101003/CS722U.1-PO10 | M | Project brings technological changes in society. |
| 101003/CS722U.1-PO11 | H | Acquiring knowledge for project development gathers skills in design, analysis, development and implementation of algorithms. |
| 101003/CS722U.1-PO12 | H | Knowledge for project development contributes engineering skills in computing and information gatherings. |
| 101003/CS722U.2-PO1 | H | Knowledge acquired for project development will also include systematic planning, developing, testing, and implementation in computer science solutions in various domains. |
| 101003/CS722U.2-PO2 | H | Project design and development using a systematic approach brings knowledge in mathematics and engineering fundamentals. |
| 101003/CS722U.2-PO3 | H | Identifying, formulating, and analyzing the project results in a systematic approach. |
| 101003/CS722U.2-PO5 | H | Systematic approach is the tip for solving complex problems in various domains. |

| Mapping | Level | Justification |
|---|---|---|
| 101003/CS722U.2-PO6 | H | Systematic approach in the technical and design aspects provides valid conclusions. |
| 101003/CS722U.2-PO7 | H | Systematic approach in the technical and design aspects demonstrates the knowledge of sustainable development. |
| 101003/CS722U.2-PO8 | M | Identification and justification of technical aspects of project development demonstrates the need for sustainable development. |
| 101003/CS722U.2-PO9 | H | Apply professional ethics and responsibilities in engineering practice of development. |
| 101003/CS722U.2-PO11 | H | Systematic approach also includes effective reporting and documentation which gives clear instructions. |
| 101003/CS722U.2-PO12 | M | Project development using a systematic approach based on well-defined principles will result in better teamwork. |
| 101003/CS722U.3-PO9 | H | Project development as a team brings the ability to engage in independent and lifelong learning. |
| 101003/CS722U.3-PO10 | H | Identification, formulation and justification in technical aspects will be based on acquiring skills in design and development of algorithms. |
| 101003/CS722U.3-PO11 | H | Identification, formulation and justification in technical aspects provides the betterment of life in various domains. |
| 101003/CS722U.3-PO12 | H | Students are able to interpret, improve and redefine technical aspects with mathematics, science and engineering fundamentals for the solutions of complex problems. |
| 101003/CS722U.4-PO5 | H | Students are able to interpret, improve and redefine technical aspects with identification, formulation and analysis of complex problems. |
| 101003/CS722U.4-PO8 | H | Students are able to interpret, improve and redefine technical aspects to meet the specified needs with appropriate consideration for public health and safety, and the cultural, societal, and environmental considerations. |
| 101003/CS722U.4-PO9 | H | Students are able to interpret, improve and redefine technical aspects for design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions. |
| 101003/CS722U.4-PO10 | H | Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools for better products. |
| 101003/CS722U.4-PO11 | M | Students are able to interpret, improve and redefine technical aspects by applying contextual knowledge to assess societal, health and consequential responsibilities relevant to professional engineering practices. |
| 101003/CS722U.4-PO12 | H | Students are able to interpret, improve and redefine technical aspects for demonstrating the knowledge of, and need for sustainable development. |

| Mapping | Level | Justification |
|---|---|---|
| 101003/CS722U.5-PO1 | H | Students are able to interpret, improve and redefine technical aspects, apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice. |
| 101003/CS722U.5-PO2 | M | Students are able to interpret, improve and redefine technical aspects, communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions. |
| 101003/CS722U.5-PO3 | H | Students are able to interpret, improve and redefine technical aspects to demonstrate knowledge and understanding of the engineering and management principle in multidisciplinary environments. |
| 101003/CS722U.5-PO4 | H | Students are able to interpret, improve and redefine technical aspects, recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change. |
| 101003/CS722U.5-PO5 | M | Students are able to interpret, improve and redefine technical aspects in acquiring skills to design, analyze and develop algorithms and implement those using high-level programming languages. |
| 101003/CS722U.5-PO12 | M | Students are able to interpret, improve and redefine technical aspects and contribute their engineering skills in computing and information engineering domains like network design and administration, database design and knowledge engineering. |
| 101003/CS722U.6-PO5 | M | Students are able to interpret, improve and redefine technical aspects and develop strong skills in systematic planning, developing, testing, implementing and providing IT solutions for different domains which helps in the betterment of life. |
| 101003/CS722U.6-PO8 | H | Students will be able to associate with a team as an effective team player for the development of technical projects by applying the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems. |
| 101003/CS722U.6-PO9 | H | Students will be able to associate with a team as an effective team player to identify, formulate, review research literature, and analyze complex engineering problems. |
| 101003/CS722U.6-PO10 | M | Students will be able to associate with a team as an effective team player for designing solutions to complex engineering problems and design system components. |
| 101003/CS722U.6-PO11 | M | Students will be able to associate with a team as an effective team player to use research-based knowledge and research methods including design of experiments, analysis and interpretation of data. |

| **Mapping** | **Level** | **Justification** |
|---|---|---|
| 101003/CS722U.6-PO12 | H | Students will be able to associate with a team as an effective team player, applying ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice. |
| 101003/CS722U.1-PSO1 | H | Students are able to develop Computer Science Specific Skills by modeling and solving problems. |
| 101003/CS722U.2-PSO2 | M | Developing products, processes or technologies for sustainable and socially relevant applications can promote Programming and Software Development Skills. |
| 101003/CS722U.3-PSO3 | H | Working in a team can result in the effective development of Professional Skills. |
| 101003/CS722U.4-PSO3 | H | Planning and scheduling can result in the effective development of Professional Skills. |
| 101003/CS722U.5-PSO1 | H | Students are able to develop Computer Science Specific Skills by creating innovative solutions to problems. |
| 101003/CS722U.6-PSO3 | H | Organizing and communicating technical and scientific findings can help in the effective development of Professional Skills. |