Faculty of Computer Science
and Information Technology

INTELLIGENT IMAGE ENHANCEMENT SYSTEM

Under Supervision of:

Dr.

Mohamed Mamdouh Eissa

Submitted by:

1	**Aya Mostafa Badr**
2	**Eslam Ehab Fouad**
3	**Loay Hassan Ali**
4	**Mustafa Muhamed El-Sherbiny**
5	**Mohamed Alaa El-dien Motaman**
6	**Mark Emad Mamlouk**

A dissertation submitted in partial fulfillment of the requirements for the degree of Bachelor of computer science and information technology

Egypt 2023

Committee Report

We certify we have read this graduation project report as examining committee, examine the student in its content and that in our opinion it its adequate as a project document for “ ”.

Chairman:	***Supervisor:***
Name:	Name:
Signature:	Signature:
Date: / /202	Date: / /202

Examiner:

Name:

Signature:

Date: / /202

Intellectual Property Right Declaration

This is to declare that the work under the supervision of Dr. Mohamed Mamdouh having title “Intelligent Image Enhancement System” carried out in partial fulfillment of the requirements of Bachelor of Science in Computer Science is the sole property of Ahram Canadian University and the respective supervisor. It is protected under the intellectual property right laws and conventions. It can only be considered used for purposes like extension for further enhancement, product development, adoption for commercial/organizational usage, etc. with the permission of the University and respective supervisor. This above statement applies to all students and faculty members.

Names:

Supervisor:

Anti-Plagiarism Declaration

This is to declare that the above publication produced under the supervision of Dr. Mohamed Mamdouh having title " Intelligent Image Enhancement System " is the sole contribution of the author(s) and no part hereof has been reproduced illegally (cut and paste) which can be considered as Plagiarism. All referenced parts have been used to argue the idea and have been cited properly. We will be responsible and liable for any consequence if violation of this declaration is proven.

Names:

Acknowledgment

First and foremost, gratitude and appreciation to Allah for the blessing of intellect and health to accomplish this job. I do this to acknowledge the numerous contributions of my superiors and mentors, as well as my family and friends' help.

To my supervisor, Dr. Mohamed Mamdouh: Thank you for your persistent direction and unending patience in bringing this project to conclusion.
This project would not have been possible without your ideas and enthusiasm.

To my graduation team: Thank you for your hard work unwavering patience in making this project a reality.

To my parents: I owe you for your love. affection and respect.
Thank you for allowing me the freedom and chance to pursue my education.

Abstract

Most of businesses and companies need to archive and save their data for a lot of purposes in the format of documents and with the current technological evolution it was mandatory to develop a tool as a system to keep track of these documents and handling the various problems of storing and retrieving the data saving effort and time.

From this idea we got motivated to build and Document Management System to manage, store, retrieve and arrange the different documents.

In this project, we will propose secure and intelligent document management and enhancement system based on the concept of optical character recognition for documents captured by cameras using AI and machine learning, a secure and fast database.
The proposed system; called Intelligent Image Enhancement System, is a service-oriented intelligent document management system through which business owners and their employees can view, save, share, scan and retrieve documents faster and easier as possible.

Document management should be a simple and straightforward task with no time-consuming or difficult stages. This is what we are attempting to construct.

Table of contents

List of Figures

List of Tables

List of Equations

List of abbreviations and acronyms

Term	Definition
DMS	Document Management System
EDMS	Electronic Document Management System
PSF	Point Spread Function
API	Application Programming Interface
RNN	Recurrent Neural Network
CNN	Convolutional Neural Network
FC	Fully Connected
LSTM	Long short-term memory
CTC	Connectionist Temporal Classification
BRNN	Bidirectional Recurrent Neural Network
CRNN	Convolutional-Recurrent Neural Network
RGB	Red, Green and Blue
MVC	Model View Controller
OCR	Optical Character Recognition
BPM	Business Process Management

Chapter 1

1.1 Overview:

The Document Management System (DMS) is a software that stores documents and allows organization to easily access, sort and manage documents, DMS is highly essential for any association to manage and sort their documents.

One of the important features that any modern DMS has is Optical Character Recognition (OCR), which enables the system to convert scanned paper documents into digital documents, which makes access to data in these documents, sorting them and searching through them much easier, which saves time, effort and money.

The system we propose will allow for simpler and easier use of main DMS and OCR features.

1.2 General Constraints:

Despite their potential to streamline document management and increase productivity, Document Management Systems can still encounter various issues. These issues may include:

1. Mobile Camera Scan: The DMS system will incorporate mobile camera scanning to improve usability and user experience.

2. Complexity of use: Some DMS software can be difficult to use, especially for users with limited technical experience, which can impede productivity and make it difficult to take full advantage of the system.

3. Document Retrieval: DMS software may not have the proper search mechanism in place, which can lead to difficulty in finding the needed documents and increase the time to retrieve them.

4. Integration issues: DMS software may not integrate well with other systems or workflows, which can make it difficult to use effectively and may impede productivity.

5. Data Loss: DMS software may not have proper backup mechanisms in place, which can lead to data loss in case of system failure.

1.3 Motivation:

Our main purpose is based on 2 points:

The first one is that most DMS systems are overly complicated and have many unnecessary features and bloated with complexities, and requires specialists to operate it on the company systems.

The second point is making the use of OCR much broader by enabling people to scan documents from mobile cameras and not necessarily use scanners to upload documents in their systems to be scanned by OCR.

1.4 Objective:

Our main objective is establishing simpler system containing main DMS features with support of powerful, simple, OCR technology, which will focus on the previously mentioned points that allows easier interaction with the DMS system and easier upload and control over documents.

Also, the procedure to install and use the program would be easier than most systems.

1.5 Scope:

Our scope applies to associations that want to use main document management features or any group that want a simple system to store their documents.

Any organization can download our software and use it on their local network, with most operations done on local machines.

1.6 Project timeline:

The following graph represents the timeline of our project starting from the 1st of October 2022.

Type	Task	Team Member	Status	Days	Start	End	Progress
S	**Introduction**						
T	Project Overview	Mohamed Alaa El Dien Motamein	Complete	1	1-Oct	1-Oct	100%
T	Aim of the project	Islam Ehab Foad	Complete	1	1-Oct	1-Oct	100%
T	Determining project Scope	Aya Mostafa Badr Ibrahim	Complete	1	2-Oct	2-Oct	100%
T	Motivation for the project	Aya Mostafa Badr Ibrahim	Complete	1	2-Oct	2-Oct	100%
T	Defining General constraints	Loay Hassan Ali	Complete	1	2-Oct	2-Oct	100%
S	**Background and Previous work**						
T	Background	Aya Mostafa Badr Ibrahim	Complete	2	4-Oct	5-Oct	100%
T	Previous work	Loay Hassan Ali	Complete	3	6-Oct	8-Oct	100%
S	**Project planning and analysis**						
T	Feasibility study and estimated cost	Mustafa Mohamed El Sherbiny	Complete	3	8-Oct	10-Oct	100%
T	Gantt chart (Project Plan)	Mark Emad Mamlouk	Complete	1	8-Oct	8-Oct	100%
T	Analysis and limitations of existing system	Aya Mostafa Badr Ibrahim	Complete	5	10-Oct	14-Oct	100%
T	Need for new system	Mustafa Mohamed El Sherbiny	Complete	2	12-Oct	13-Oct	100%
T	Advantage of new system	Mohamed Alaa El Dien Motamein	Complete	4	15-Oct	18-Oct	100%
S	**System Features and Requirements**						
T	User requirements	Islam Ehab Foad	Complete	2	18-Oct	19-Oct	100%
T	System requirements	Mustafa Mohamed El Sherbiny	Complete	2	18-Oct	19-Oct	100%
T	Functional requirements	Loay Hassan Ali	Complete	2	19-Oct	20-Oct	100%
T	Non-functional requirements	Mohamed Alaa El Dien Motamein	Complete	2	19-Oct	20-Oct	100%
T	Tools to develop the system	Loay Hassan Ali	Complete	2	21-Oct	22-Oct	100%
T	User characteristics	Aya Mostafa Badr Ibrahim	Complete	1	21-Oct	21-Oct	100%
S	**Documentation 1.0**						
T	Writing chapter 1,2,3 and 4 in the document	Loay Hassan, Aya Mostafa	Complete	4	23-Oct	26-Oct	100%
S	**Design**						
T	Design and Implementation Constraints	Aya Mostafa Badr Ibrahim	Complete	4	26-Oct	29-Oct	100%
T	Assumptions and dependencies	Loay Hassan Ali	Complete	3	28-Oct	30-Oct	100%
T	Risks and risk management	Mohamed Alaa El Dien Motamein	Complete	3	31-Oct	2-Nov	100%
T	Design of database (ERD - Mapping)	Eslam Ehab, Mustafa Muhamed, Mark	Complete	5	1-Nov	5-Nov	100%
T	Class diagram		In Progress	3	15-Nov	17-Nov	
T	Use case diagram	Mustafa Mohamed El Sherbiny	Complete	3	6-Nov	8-Nov	100%
T	Activity diagram	Mohamed Alaa El Dien Motamein	Complete	3	6-Nov	8-Nov	100%
T	Sequence diagram		In Progress	3	15-Nov	17-Nov	
T	State diagram	Mustafa Mohamed El Sherbiny	Complete	3	9-Nov	11-Nov	100%
S	**Implementation**						
T	Software Architecture	Islam Ehab Foad	In Progress	9	12-Nov	20-Nov	25%
T	User Interface		Not Started	5	15-Feb	19-Feb	
T	Results and Discussion		Not Started	1	20-Feb	20-Feb	
S	**Testing**						
T	Unit Testing		Not Started	3	23-Feb	25-Feb	
T	Integration Testing		Not Started	3	23-Feb	25-Feb	
S	**Documentation 2.0**						
T	Writing chapter 4,5 and 6 in the documentation		Not Started	4	1-Mar	4-Mar	

Fig. 1. Gannt chart of our project.

Chapter 2

Previous work

2.1 Introduction:

2.1.1 History of DMS:

Several businesses started creating software solutions to handle paper-based documents in the 1980s. These systems dealt with printed and published papers, pictures, prints, and other paper-based materials. Later, developers began to design a second sort of system capable of managing electronic documents, i.e., all those papers, or files, generated on computers and often saved on users' local file-systems. The first Electronic document management (EDM) systems only supported proprietary file types or a small number of file formats. Since they concentrated on the capture, storage, indexing, and retrieval of image file formats, several of these systems later came to be known as document imaging systems. EDM systems have progressed to the point that they can manage any file format that can be stored on a network. Electronic documents, collaboration tools, security, workflow, and auditing features have all been added[1].

2.1.2 History of OCR:

Kurzweil Computer Products first sold a commercial version of the optical character recognition computer application in 1978. LexisNexis was among the earliest clients, using the application to upload legal and news materials to their fledgling online databases. Kurzweil sold his firm two years later to Xerox, which was interested in commercializing paper-to-computer text conversion. Scansoft was set out by Xerox and then merged with Nuance Communications. OCR was made available as a service (WebOCR) in the 2000s, as well as in cloud computing environments and mobile applications such as real-time translation of foreign-language signs on a smartphone. With the introduction of smart phones, OCR may now be utilized in internet-connected mobile device apps that extract text recorded with the device's camera. These devices will normally utilize an OCR API to extract the text from the picture file recorded and delivered by the device if the operating system does not have OCR capabilities. The OCR API delivers the extracted text to the device app, together with information about the identified text's placement in the original picture, for additional processing (such as text-to-speech) or display[2].

2.2 Previous work:

Text Recognition of Low-resolution Document Images, Charles Jacobs, Patrice Y. Simard, Paul Viola, James Rinker in 2005[3]: This paper tries to extract text from low quality images by using CNN for character recognition, then adding a model to recognize words.

Two approaches were used for word recognition language-neutral model and dictionary model which gave better results. The latter model tries to find out which word in a dictionary is the most likely match for a given input image. First a version of the dictionary-based recognizer that simply scans linearly through the entire lexicon is described, evaluating the probability for each word, and outputting the word with the highest score.

Disadvantages: System is quite slow. It takes a total of 2 minutes and 40 seconds to produce the output. Of that time, 2 minutes and 20 seconds are taken up by the character recognizer.

Optimization of the Business Processes Via Automatic Integration with the Document Management System, Almir Djedović, Emir Žunić, Dino Alić, Samir Omanović, and Almir Karabegović in 2016[4]: This Paper Discusses how activities inside the BPM(Business Process Management) while integrated with an EDMS can be decreased duration wise based on a data warehouse model, based on an analytical approach, it is faster to establish a connection to the DMS with the user and await important fields to be filled rather than accepting filled documents to be uploaded, also, the paper advises that the database becomes united between the BPM and EDMS rather than having separate Database (DB) for each one, and discusses that integration process is what bottlenecks the performance of the DMS, using automatic integration aids the processes performance.

Results: Automatic integration reduces the time needed from checking the account status to submission of the request (the entire process) by a noticeable amount as shown in fig 2.

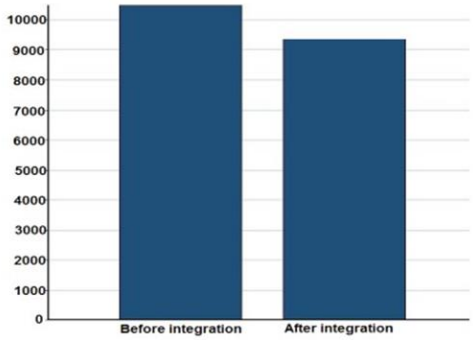

Fig. 2. Time execution in the process are minimized after automatic integration between BPM and EDMS systems.

Image Processing based Degraded Camera Captured Document Enhancement for Improved OCR Accuracy, Pooja Sharma, Shanu Sharma in 2016[5]: This paper evaluates the performance of various de-blurring techniques for noisy and blurred camera captured documents. It compares different results of enhancing text pictures by de-noising then de-blurring of images.

It focuses on removing these types on noises: Gaussian, salt and pepper, by using Weiner filter for Gaussian blur and median filter for salt and pepper. Then removing blur which can be classed as motion and out of focus blur, de-blurring has 2 types blind and non-blind that the PSF (Point Spread Function) is known. After distortions have been minimized contrast enhancement for enhancing grayscale values can be used.

Results: Regularized filter de-blurring method provides the best results in comparison to Lucy Richardson and Weiner methods.
Blind Deconvolution method helps in guessing the PSF and returns a good de-blurred image. Imadjust() contrast enhancement is fit for use in our case as shown in fig 3.

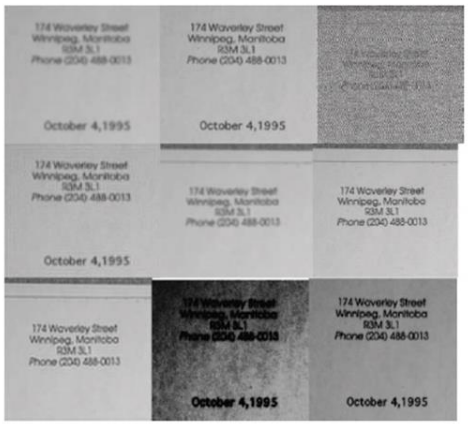

Fig. 3. Results of various de-blurring techniques on camera captured documents

Electronic document management systems and distributed large-scale systems, V. L. Orlov, E. A. Kurako in 2017[6]: In this paper direct object management in object-oriented document flow is discussed, as it is to modify and check documents when in an object-oriented EDMS,

when a document is modified, object connected to said document is also modified which makes it easier to check details about the document, the object contains data selected from the database and are located in an electronic image of the document.

Results: The processing of these document's objects can be tiresome, as it requires more time to be process data and its objects rather than documents only, however, SQL scripts can be used to adjust new documents, but with the cost of complicated monitoring, as SQL scripts can be written in different ways, thus, the paradoxical conclusion, object-oriented EDMS will lead to less volume being used but more complicated methods of exchange data in large scale businesses.

Text Extraction and Categorization from Watermark Scientific Document in Bulk, Wai Chong Chia, Phoey Lee Teh, Colin Mathew Hew D Gill, in 2018[7]: This paper tries an approach that makes use of direct text recognition from PDF and optical character recognition (OCR) to produce two version of digital text that can be combined for better accuracy. It compares the results of OCR algorithm tesseract and PDFminer to get the ultimate result in case of a watermark. As OCR can't handle text recognition with watermarks it's used for proper text formatting and handling lines order (using text images) In the other hand, PDFminer spots the watermark and includes it in the text extracted from the PDF.

Results: In case of watermark in front of the text OCR miss some words up, but PDF extracted text don't but includes watermark in text. Together they result the text wanted well formatted and correctly spelled. In case of watermark in the background OCR extract text fine.

Disadvantages: This method requires the document to be in PDF format to extract text and separate watermark.

Text recognition in document images obtained by a smartphone based on deep convolutional and recurrent neural network, Hassan El Bahi, Abdelkarim Zatni in 2019[8]: This paper proposes a method to recognize text in document images caught by a mobile camera.

It discusses a detailed way to preprocess the images, extract the features, and perform the classification on the text. Preprocessing phase includes text detection, noise removal, text segmentation, baseline correction, and normalization, which outputs text-line images to be entered as an input to the feature extraction phase, which uses a CNN (Convolutional Neural Network) to automatically recognize and extract features, then outputs feature vectors to be the input of the classification phase which uses a bidirectional recurrent neural network (BRNN) with the gated recurrent units (GRU) block and a Connectionist temporal classification (CTC) layer to classify the characters in the text.

Results: Experiments on the ICDAR2015 Smartphone document OCR dataset were performed with several feature extraction and classification methods, and these were the results in table 1.

Table 1: ICDAR2015 Smartphone document OCR dataset classification results.

Features extraction	Classification	Training (%)	Test (%)
CNN_{4x32}	RNN	86.40	78.16
	BRNN	90.73	84.96
CNN_{8x32}	RNN	92.32	88.03
	BRNN	95.80	94.49
CNN_{16x32}	RNN	90.59	87.21
	BRNN	95.18	92.13

Analysis Of Main Requirements For Electronic Document Management Systems, Nazila Ali Ragimova, Vugar Hajimahmud Abdullayev, Vasila Soltanaga Abbasova in 2020[9]: The aim of this paper is to analyze the basic requirements for electronic document management systems, and determine its importance in the organization of documents.

The main stages of the document lifecycle, include creation, management/storage, access, retrieval, administration, reassignment, collaboration, distribution, preservation, disposal, storage. The main purpose of the EDMs should be to assist in the recovery and retrieval of user files.

Results: To reduce costs, customers can only take the necessary functionality of electronic document management systems. Like other information systems, customers have their own requirements.

– Make documents available to all employees of the corporation.

– Ensure document integrity throughout the lifecycle.

– Ensure confidentiality of documents, including ensuring the protection of documents from unauthorized access.

– Reliable archiving tools available.

– Ensure system modularity.

– A powerful document search tool.

– Availability of tools for implementing business processes and graphical user interface.

Chapter 3

Background

3 Background:

3.1 Document management:

Document management is how an organization uses, manages, and stores documents in the business. These documents may come in different formats, such as PDFs, images, videos, audio, spreadsheets, code packages, etc.

Using document management software helps simplify the document life cycle management process to a large extent. You can set up automated workflows and processes to manage, update, and store documents shared and used in the company. Document management also helps maintain:

- Confidentiality with teams, customers, and partners
- Efficiency of workflows and systems
- Compliance with industry standards and regulatory authorities

3.1.1 The mechanism of document management system:

A document management system organizes your documents in many ways, e.g., file type, data type, security, or priority level. When new documents are received or created, they're captured and moved into the system. The document management system may automatically add more information to the file at this point, e.g., internal tags, labels, and metadata such as who created or uploaded a document and when. DMSs then sort and store the new document or data based on previously set rules as shown in fig 4.

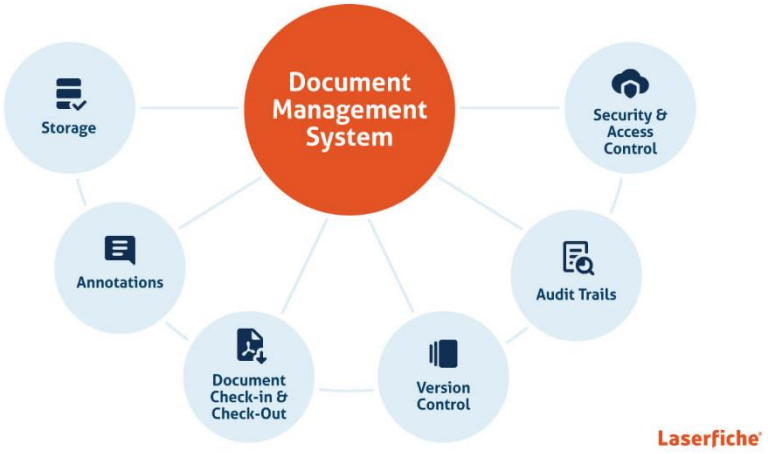

Fig. 4. Document Management System architecture

3.1.2 Benefits of document management systems:

- Improved collaboration and teamwork: Document management systems ensure team members always have access to the most recent document version.

- Enhanced security: Document management systems ensure documents entering and leaving your organization are safe and stored securely in a central, accessible platform for all teams to access.
- Compliance and audits: Document management systems help track a document's history from creation to edits and downloads. This information is critical for industry-compliance practices and businesses being audited.
- Ease of access to information: Document management systems usually have robust search functionality, making it easy for teams to find the right documents when they need them.
- Improved productivity: Using a document management system means employees don't waste time looking for files in multiple computers, apps, or cloud-based platforms. Anyone with the right access can find the documents they need quickly and easily by using the search bar[10].

3.2 The .NET platform:

.NET is a developer platform made up of tools, programming languages, and libraries for building many different types of applications[11].

Fig. 5. A representation of .NET platform.

3.2.1 ASP.NET and C#:

ASP.NET is an open source web framework, created by Microsoft, for building modern web apps and services with .NET, ASP.NET is cross platform and runs on Windows, Linux, macOS, and Docker.

ASP.NET extends the .NET platform with tools and libraries specifically for building web apps.

These are some things that ASP.NET adds to the .NET platform:

1 Base framework for processing web requests in C# or F#
2 Web-page templating syntax, known as Razor, for building dynamic web pages using C#
3 Libraries for common web patterns, such as Model View Controller (MVC)
4 Authentication system that includes libraries, a database, and template pages for handling logins, including multi-factor authentication and external authentication with Google, Twitter, and more.
5 Editor extensions to provide syntax highlighting, code completion, and other functionality specifically for developing web pages[12].

C# is a simple, modern, general-purpose, object-oriented programming language developed by Microsoft within its .NET initiative. C# is used for building interactive web pages with the framework Razor.

3.2.2 React.js:

The React.js framework is an open-source JavaScript framework and library developed by Facebook. It's used for building interactive user interfaces and web applications quickly and efficiently with significantly less code than you would with vanilla JavaScript.
In React, you develop your applications by creating reusable components that you can think of as independent Lego blocks. These components are individual pieces of a final interface, which, when assembled, form the application's entire user interface.

React's primary role in an application is to handle the view layer of that application just like the V in a model-view-controller (MVC) pattern by providing the best and most efficient rendering execution. Rather than dealing with the whole user interface as a single unit, React.js encourages developers to separate these complex UIs into individual reusable components that form the building blocks of the whole UI.

3.2.3 Microsoft SQL Server:

SQL Server is a relational database management system (RDBMS) developed by Microsoft. It is primarily designed and developed to compete with MySQL and Oracle database. SQL Server supports ANSI SQL, which is the standard SQL (Structured Query Language) language. However, SQL Server comes with its own implementation of the SQL language, T-SQL (Transact-SQL)[13].

3.3 Computer vision:

Computer Vision is an interdisciplinary field that has been gaining huge amounts of traction in the recent years(since CNN) and self-driving cars have taken center stage. Another integral part of computer vision is object detection. Object detection aids in pose estimation, vehicle detection, surveillance etc. The difference between object detection algorithms and classification algorithms is that in detection algorithms, a bounding box around the object of interest is drawn to locate it within the image.

3.3.1 Gray scale of an image:

It is an image conversion technique in digital photography. It eliminates every form of color information and only leaves different shades of gray; the brightest being white and the darkest of it being black. Its intermediate shades usually have an equal level of brightness for the primary colors (red, blue and green). Alternatively it uses equal amounts of cyan, yellow and magenta which are the primary pigments. Each pixel is a representation of the luminous intensity of the image.

3.3.2 RGB and Gray image:

Color images in deep learning are represented in RGB format. At a high level, RGB is an additive color model where each color is represented by a combination of red, green and blue values; these are usually stored as separate ‘channels’, such that an RGB image as shown in fig 6 is often referred to as a 3 channel image.

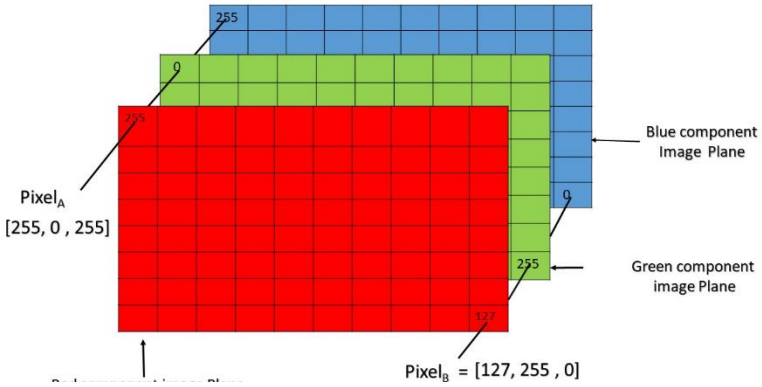

Fig. 6. RGB 3-channel matrices of an image.

A greyscale image is simply one in which the only colors represented are different shades of grey. A truly "black and white image" would consist of only these two distinct colors, which is very rarely the case; making 'greyscale' the more accurate term.

As there is no color information to represent for a greyscale image, less information needs to be stored for each pixel and an additive color model is not required! For greyscale images, the only information required is a single value to represent the intensity of each pixel; the higher this value, the lighter the shade of grey. As such, greyscale images usually consist of a single channel, where each pixel intensity is just a single number ranging from 0 to 255[14].

3.3.3 Image Enhancement:

Image enhancement is the process that improves the quality of the image for a specific application.

Image Enhancement Methods includes:

- Spatial Domain methods: The term spatial domain refers to the 2-D image plane and the approaches manipulates the pixel of a given image for enhancement.
- Frequency Domain methods: The methods in frequency domain manipulate the Fourier transform of a given image for enhancement.

Image Enhancement in Spatial Domain:

If a 1x1 neighborhood is used for the transformation function, then the function is called the gray-level transformation function.

$s = T(r)$, where, $s = g(x, y)$ and $r = f(x, y)$ (equ.1)

Point Processing:

•Refers to the processing techniques where the enhancement at

any point in an image depends only the gray-level at that point.

Mask Processing / Filtering:

•Mask (also referred to as filter/kernel/template/window) is a 2-D

array (i.e. 3x3) defined around a pixel, where the values of the

mask coefficients determine the nature of the process.

Image enhancement methods using mask approach is called Mask Processing / Filtering[15].

3.3.4 Thresholding:

Thresholding is a type of image segmentation, where the pixels of an image is changed to make the image easier to analyze. In thresholding, as shown in fig 7. An image from color or grayscale is converted into a binary image. This is typically done in order to separate "object" or foreground pixels from background pixels to aid in image processing[16].

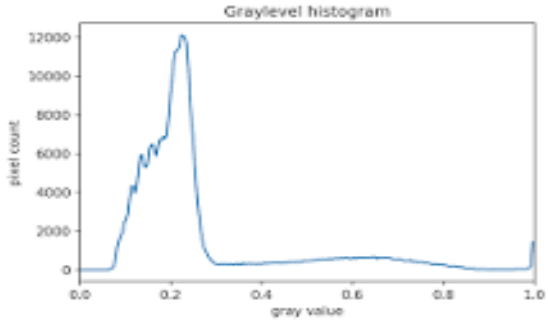

Fig. 7. An example of image threshold.

3.3.5 Local thresholding:

If the image background is relatively uniform, then you can use a global threshold. However, if there is large variation in the background intensity, adaptive thresholding (a.k.a. local or dynamic thresholding) may produce better results as shown in fig 8. Note that local is much slower than global thresholding[17].

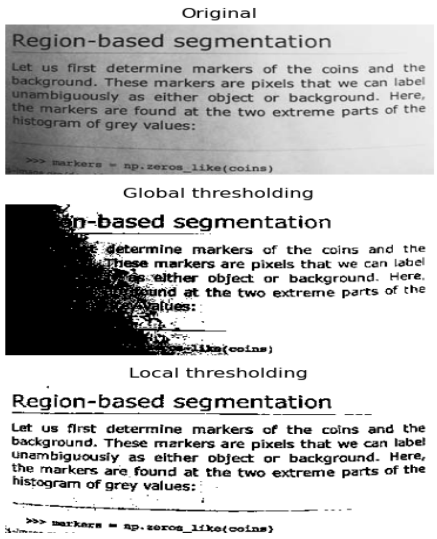

Fig. 8. Difference between Global and Local thresholding.

3.3.6 Segmentation:

Segmentation is nothing but breaking the whole image into subparts to process them further.

Segmentation of image is done in the following sequence :
1. Line level Segmentation
2. Word level Segmentation
3. Character level Segmentation

1) **Line Level Segmentation:** In this level of segmentation, skew corrected image containing text written in the form of lines is provided. The objective of Line Level Segmentation is to segment the image into lines as shown in fig 9.

   The idea is, If the binary image horizontally projected,

   - Rows that represent the text in a line have high No. of foreground pixels, which correspond to higher peaks in the histogram.
   - Rows that represent the gaps in-between the lines have high number of background pixels, which correspond to lower peaks in the histogram.

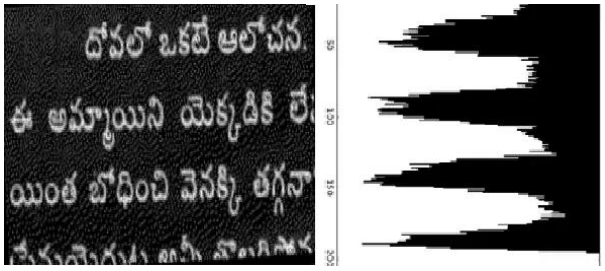

Fig. 9. Segmented lines threshold.

2) **Word Level Segmentation:** At this level of segmentation, an image containing a single line (segmented in the previous step) is provided with which consists of a sequence of words. The objective of Word Level Segmentation is to segment the image into words as shown in fig 10..

   The idea is similar to the previous step, but the only change is, here it is a must to project the image vertically (Sum is taken along columns) because it is a must to segment words vertically.

   - Columns that represent the text have high number of foreground pixels, which correspond to higher peaks in the histogram.

- Columns that represent the gaps in-between the words have high number of background pixels, which correspond to lower peaks in the histogram.

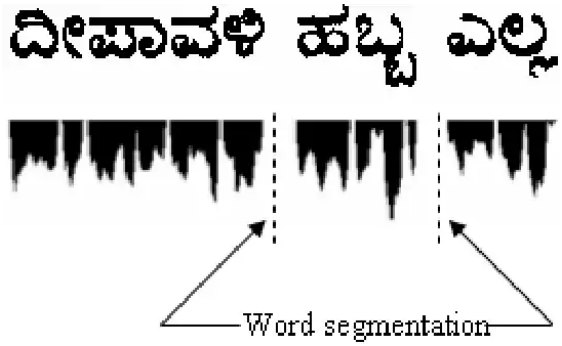

Fig. 10. Segmented words threshold.

3) **Character Level Segmentation:** At this level of segmentation, an image containing a single word (segmented in the previous step) is provided with which consists of a sequence of characters. The objective of Character Level Segmentation is to segment the image into individual characters.

   the two main approaches to image segmentation are based on convolutional neural networks (CNN) and super pixels. Super pixel is an approach that divides an image into regions (called super pixels) with similar properties, such as color, texture, and brightness[18].

3.4 Artificial Neural Networks (ANN):

Artificial neural networks enable computer programmers to identify patterns and resolve common issues by mimicking the behavior of the human brain. A node layer of an artificial neural network consists of an input layer, one or more hidden layers and output layer.

Each node, or artificial neuron, is connected to others and has a weight and threshold that go along with it. Any node whose output exceeds the defined threshold value is activated and begins providing data to the network's uppermost layer. Otherwise, no data is transmitted to the network's next tier[19].

The neural network equation looks like this:

$$Z = Bias + W1X1 + W2X2 + \ldots + WnXn \quad \text{(equ 2)}$$

Neural networks classified into different types:

1-The perceptron is the oldest neural network, created by Frank Rosenblatt in 1958. It has a single neuron and is the simplest form of a neural network

2-Convolutional neural networks (CNN)

3-Recurrent neural networks (RNN)

3.4.1 Convolutional Neural Network (CNN):

Convolution neural network is a Deep Learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

There are three types of layers that make up the CNN architecture as shown in Fig 11 which are the convolutional layers, pooling layers, and Fully-Connected (FC) layers[19].

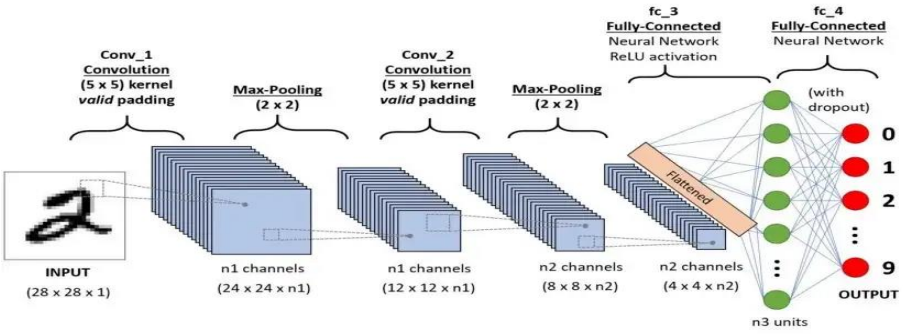

Fig. 11. CNN architecture.

3.4.2 Convolution Layer:

The convolutional layer is the core building block of a CNN, and it is where the majority of computation occurs. It requires a few components, which are input data, a filter, and a feature map.

The feature detector is a two-dimensional (2-D) array of weights, which represents part of the image. While they can vary in size, the filter size is typically a 3x3 matrix; this also determines the

size of the receptive field. The filter is then applied to an area of the image, and a dot product is calculated between the input pixels and the filter.

This dot product is then fed into an output array. Afterwards, the filter shifts by a stride, repeating the process until the kernel has swept across the entire image. The final output from the input and the filter as shown in fig 12, is known as a feature map.

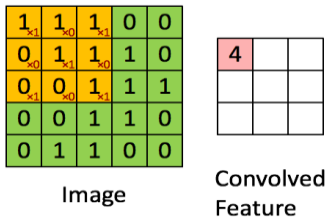

Fig. 12. Feature map after applying a convolution layer.

After each convolution operation, a CNN applies an activation function transformation to the feature map, introducing nonlinearity to the model.

As mentioned earlier, another convolution layer can follow the initial convolution layer. When this happens, the structure of the CNN can become hierarchical as the later layers can see the pixels within the receptive fields of prior layers. [20]

3.4.3 Activation Function:

The activation function is an important parameter in the CNN model.

They are used to learn and approximate any kind of continuous and complex connection between network variables. In other words, it decides whether information in the model should and should not be conveyed at the network's end. It adds nonlinearity to the network. The ReLU, SoftMax, tanH, and Sigmoid functions as shown in fig 13 are among the most often utilized activation functions. Each of these functions has a distinct purpose. For a binary classification CNN model, the sigmoid and SoftMax functions are favored, while SoftMax is typically utilized for multi-class

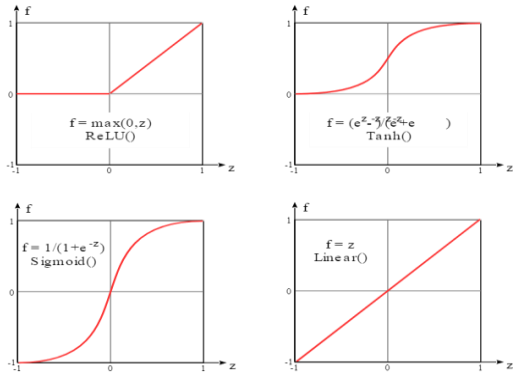

Fig. 13. Examples of activation functions.

classification. In a nutshell, activation functions in a CNN model determine whether or not to activate a neuron. It determines if the input is significant enough to be predicted using mathematical processes[21].

3.4.4 Pooling Layer:

Pooling layers, also known as down sampling, conducts dimensionality reduction, reducing the number of parameters in the input. Similar to the convolutional layer, the pooling operation sweeps a filter across the entire input, but the difference is that this filter does not have any weights. Instead, the kernel applies an aggregation function to the values within the receptive field, populating the output array. There are two main types of pooling:

1. Max pooling: As the filter moves across the input, it selects the pixel with the maximum value to send to the output array.
2. Average pooling: As the filter moves across the input, it calculates the average value within the receptive field to send to the output array as shown in fig 14.

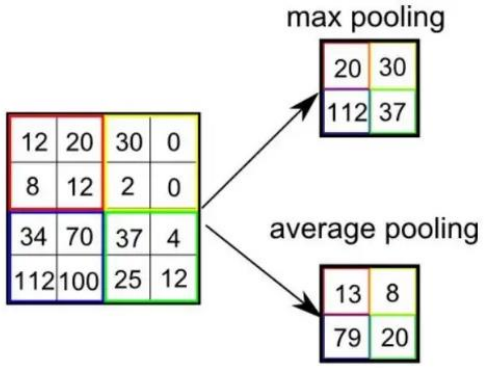

Fig. 14. The two types of pooling layers.

While a lot of information is lost in the pooling layer, it also has a number of benefits to the CNN. They help to reduce complexity, improve efficiency, and limit risk of overfitting.

3.4.5 Fully-Connected Layer:

The name of the full-connected layer aptly describes itself. As mentioned earlier, the pixel values of the input image are not directly connected to the output layer in partially connected layers. However, in the fully-connected layer, each node in the output layer connects directly to a node in the previous layer.

This layer performs the task of classification based on the features extracted through the previous layers and their different filters. While convolutional and pooling layers tend to use ReLu functions, FC layers usually leverage a softmax activation function to classify inputs appropriately, producing a probability from 0 to 1 as shown in fig 15[22].

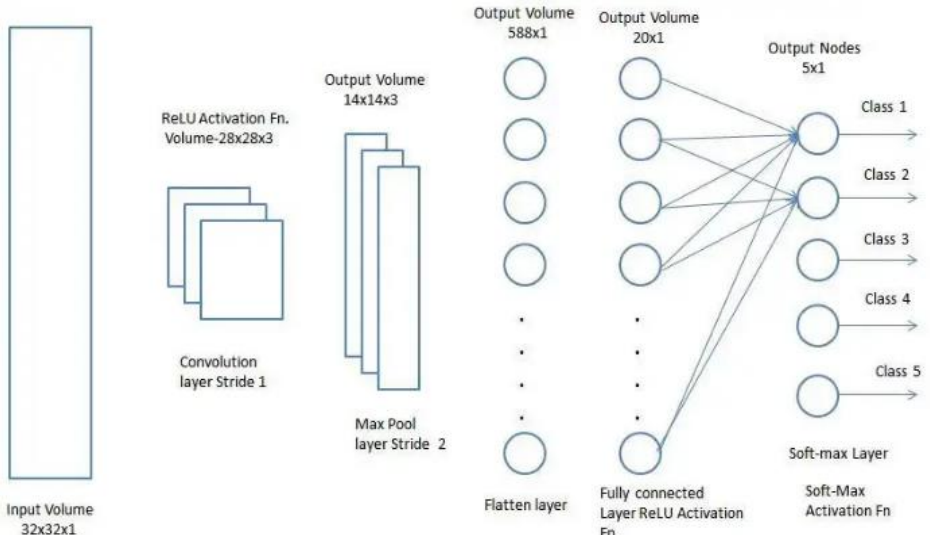

Fig. 15. Fully Connected layer.

3.4.6 Data Augmentation:

Data augmentation is a set of techniques used to increase the amount of data in a machine learning model by adding slightly modified copies of already existing data or newly created synthetic data from existing data as shown in fig 16. Data augmentation can be useful for a model to learn better and to avoid overfitting of a model.

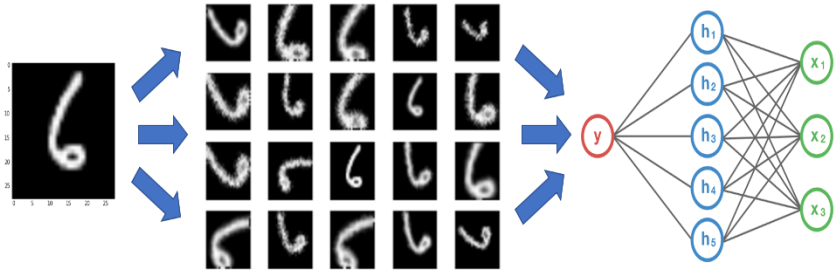

Fig. 16. Data Augmentation.

3.4.7 Dropout:

Overfitting in the training dataset is common when all features are connected to the FC layer. Overfitting occurs when a model performs so well on training data that it degrades the model's performance when applied to new data. To overcome this issue, a dropout layer is used, in which a few neurons are removed from the neural network during the training process as shown in fig 17, resulting in a smaller model.

When a dropout of 0.3 is reached, 30% of the nodes in the neural network are randomly removed. Dropout improves the performance of a machine learning model by preventing overfitting by simplifying the network. During training, it removes neurons from neural networks[23].

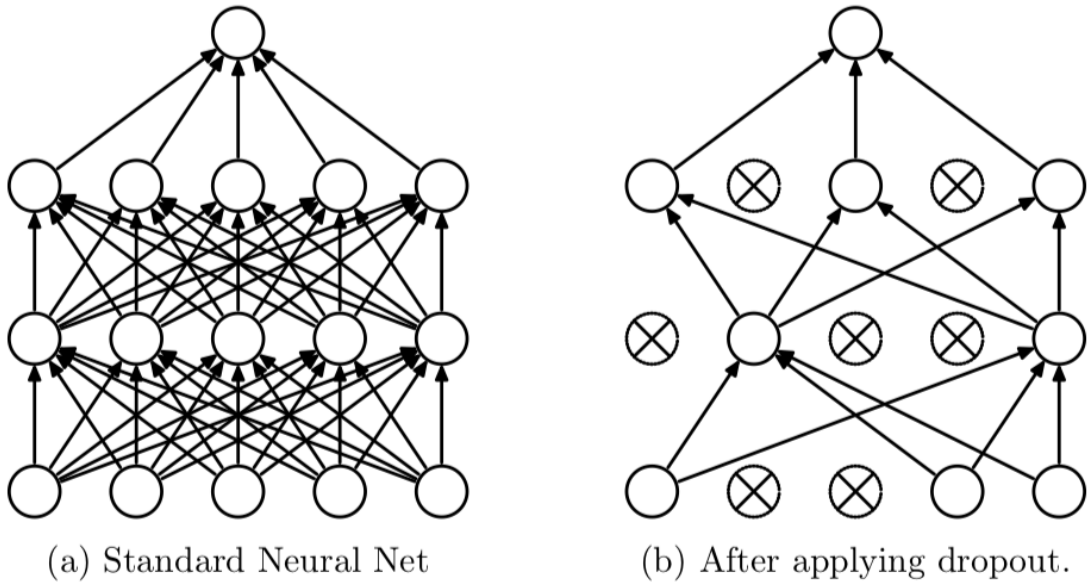

Fig. 17. Drop-out

3.4.8 Recurrent Neural Network:

Recurrent Neural Network is a generalization of feedforward neural network that has an internal memory. RNN is recurrent in nature as it performs the same function for every input of data while the output of the current input depends on the past one computation. After producing the output, it is copied and sent back into the recurrent network. For making a decision, it considers the current input and the output that it has learned from the previous input. RNNs can use their internal state (memory) to process sequences of inputs as shown in fig 18.

This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition. In other neural networks, all the inputs are independent of each other. But in RNN, all the inputs are related to each other. Recurrent neural network are even used with convolutional layers to extend the effective pixel neighborhood[24].

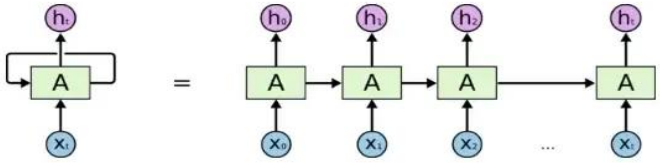

Fig. 18. RNN architecture.

3.4.9 Long Short-Term Memory:

Long short-term memory (LSTM) networks are a modified version of recurrent neural networks, which makes it easier to remember past data in memory. The vanishing gradient problem of RNN is resolved here. LSTM is well-suited to classify, process and predict time series given time lags of unknown duration. It trains the model by using back-propagation. In an LSTM network, three gates are present as shown in Fig 19:

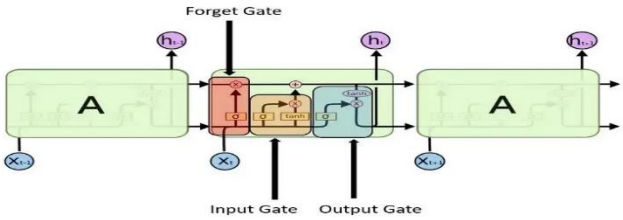

Fig. 19. LSTM architecture

- Input gate: This gate is used to add information to neuron cell. It is responsible of what values should be added to cell by using activation function like sigmoid. It creates an array of information that has to be added. This is done by using another activation function called tanh. It generates a values between -1 and 1. The sigmoid function act as a filter and regulate what information has to be added in cell.
- Forget gate: This gate decide which information is important and should be stored and which information to forget. It removes the non-important information from neuron cell. This results in optimization of performance. This gate takes 2 input- one is the output generated by previous cell and other is input of current cell. Following required bias and weights are added and multiplied and sigmoid function is applied to the value. A value between 0 and 1 is generated and based on this the decision of which information to keep is done. If value is 0 the forget gate will remove that information and if value is 1 then information is important and has to be remembered.

- Output gate: This gate is responsible for selecting important information from current cell and show it as output. It creates a vector of values using tanh function which ranges from -1 to 1. It uses previous output and current input as a regulator which also includes sigmoid function and decides which values should be shown as output.

The units of an LSTM are used as building units for the layers of a RNN, often called an LSTM network. LSTMs enable RNNs to remember inputs over a long period of time.

This is because LSTMs contain information in a memory, much like the memory of a computer. The LSTM can read, write and delete information from its memory. This memory can be seen as a gated cell, with gated meaning the cell decides whether or not to store or delete information, based on the importance it assigns to the information[25].

3.4.10 The Convolutional Recurrent Neural Network:

The Convolutional Recurrent Neural Networks is the combination of two of the most prominent neural networks. The CRNN (convolutional recurrent neural network) involves CNN(convolutional neural network) followed by the RNN(Recurrent neural networks). In the end, There is a CTC layer that predicts the text given the image. The convolutional layers extract the visual information while the recurrent layers capture the temporal information of the text in the image. The CTC layer does the prediction of the characters[26].

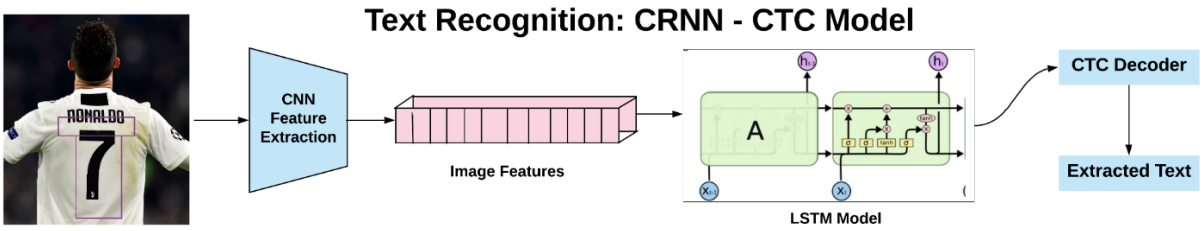

Fig. 20. CRNN-CTC architecture.

After the text detection step, regions, where the text is present, are cropped and sent through convolutional layers to get the features from the image. Later these features are fed to many-to-many LSTM architecture, which outputs softmax probabilities over the vocabulary. These outputs from different time steps are fed to the CTC decoder to finally get the raw text from images[27].

Chapter 4

Planning and Analysis and Design

4.1 Planning:

4.1.1 Feasibility study and estimated cost:

A feasibility study is an analysis of the potential of a proposed project or system to determine if it is technically and economically feasible.

- Our product has two main problems that we aim to solve, first is to convert all the paper document into digital document, second to manage this huge number of documents in a systemic way.
- All the other DMS have subscription which will be a lot of cost for the huge organization, so our plan is reduce the cost.
- Also, to reduce the cost we build our OCR.
- Like any DMS our product needs scanners.
- Our product needs local network server or internet server.
- Because the usability of our product, Employees training on our product will be just a few days.

4.2 Analysis and limitations of Existing systems:

Most of the currently existing DMS systems in the market have poor UI/UX and lack usability, which requires IT expertise to be able to install, operate and deal with the system. Also requires the employees to be trained in order to get used to it and be able to use it on daily basis.

As for the OCR systems, the existing systems are very much adequate, providing image scanning, text extraction with good enough accuracy for photos captured in proper conditions. However, we believe there is a room for improvement. Using the latest AI and Computer Vision technologies with OCR systems can be helpful in improving image quality and making it more suitable to be processed by the OCR system, preventing text detection mistakes, and increasing accuracy.

4.3 Need for new system:

Being bloated with loads of "useful" features, that probably aren't the main concern of most of the users, the current DMSs in the market offer a poor UI/UX for the users which might make all these features inaccessible or hard to use, might even make doing the basic DMS operations harder, and leads to the system having a difficult learning curve.

As for these reasons, the market is thirsty for an easy-to-use DMS product providing the basic functionality of archiving and retrieving documents with a nice UI/UX and an easy installation process that requires no technical expertise.

We Claim that the system we are proposing will quench the market's thirst and satisfy a lot of users, especially users who have got startups that need no more than the basic functionality, by providing the above mentioned requirements. In addition we will involve Artificial Intelligence and Computer Vision technologies to improve the system's OCR accuracy in detecting text. And an easy-to-use interface to deal with the OCR system as the user is intended to be able to scan the documents using any mobile camera.

4.4 Analysis of New System:

4.4.1 User Requirements:

1- System to manage all the electronic documents.

2- Support multiple users, each user has their private and shared data.

3- Each user can upload, delete, read, edit documents or folders according to their permissions.

4- Administrator can control other users' permissions to view and edit the files.

5- An administrator can create multiple folders and give other users permissions to view them.

6- Administrator can control other members' permissions.

7- An OCR System to convert paper documents captured by camera to electronic document.

8- Users can search to find files.

9- Search should not take a lot of time to find a document.

10- The system should not require a lot of training hours to be able to use.

11- System should have a file viewer if possible.

12- Users can view files metadata including author, last edit date, last editor and a short description if possible.

4.4.2 System Requirements:

1- Client application will operate on a web browser.

2- System will have a logging in functionality using a username and a password.

3- System will provide a functionality for a user to create a folder with certain permissions.

4- User shall be able to read/add/edit/delete documents in a folder depending on their permission level.

5- System will provide a functionality for an administrator to add, remove and manipulate users' and their permissions in the group.

6- User's accessible folders will be listed in a docked bar.

7- System will provide a view to list the file hierarchy.

8- System will provide an interface to input a captured document photo into the OCR system and show the OCR's output and store it.

9- System will have a view to search for files and preview them if they are in a supported format or download them otherwise.

10- System will provide a view to show documents metadata which contains name, last edit date, edited by whom and a short description if possible.

4.4.3 Functional Requirements:

1. **User shall be able to create account.**
2. **User shall have a private space to upload files.**
3. **User shall have the ability to share the private files he uploads with other users.**
4. **User shall be able to create Folders, upload documents to them, invite members and control members' permissions to manipulate these documents.**
5. **User shall be able to search for any file.**
6. **System shall have an OCR module that is responsible for converting photos of paper documents to electronic documents.**
7. **Users shall be able to add files either by uploading an electronic document or converting a paper document into electronic document.**
8. **System should have a panel to display selected documents or download the documents if not possible.**

4.4.4 Non-Functional Requirements:

1- **The system should have a modern simple UI to provide an easy UX so the users don't need a lot of training.**

2- **Users have access to documents that they are only permitted to.**

3- **Search functionality should be adequately fast.**

4.5 User Characteristics:

- **User characteristics Requirements: For Company employee**

In our product there is two types of users, administrator who has control over the system and its users, and company employee who is under control by administrators.

- **Table 2: Administrator user requirements:**

Requirements	**Needed**
Role:	Administrator
Age:	More than 21
Gender:	Both
School degree:	N/A
Previous experience with similar products:	Requires some experience with other products.
Special skills:	Requires Knowledge of using Computer and dealing with databases.

- **Table 3: Employee user requirements:**

Requirements	**Needed**
Role:	Employee
Age:	More than 21
Gender:	Both
School degree:	N/A
Previous experience with similar products:	Doesn't require any experience with other products.
Special skills:	Requires Basic Knowledge of using websites.

4.6 Design and implementation constraints:

4.6.1 Programming languages and Technologies:

Our project is meant to be based on a server-client architecture in which the server side will be responsible for doing all the computations related to the OCR AI model of text extraction, in addition to the logic responsible for storing and retrieving the files in the database. The client side on the other hand is responsible for providing a UI to the user displaying the files in the database and allowing him to manipulate it according to his permission.

The communication between the backend and the frontend will be done by a RESTful API as it's very suitable for our case of a web application sending requests to retrieve data from the server.

Our project will be more UI intensive. Meaning it a lot of workload will be done on the client side in comparison to the server side so we decided to make it an SPA (Single Page Application) web app.

4.6.2 For the backend (server side):

we chose to use ASP.NET core as it provides adequate functionality and class libraries to create the RESTful API that will be responsible for handling the client-side requests in addition to its great community as its very popular. It also supports integration with popular web frameworks such as React and Angular which fits perfect for our SPA client-side.

The .Net 7.0 framework supports a lot of programming languages. We chose to use C# because in addition to being the most popular. It's syntax and paradigms are familiar to us as we are used to OOP and procedural programming paradigms with languages such as C++ and JAVA.

4.6.3 For the frontend (client-side):

All popular web frameworks are enough for the job, but we chose React.js as part of our team has prior experience with it and it fits perfect with our ASP.Net backend.

4.6.4 For the OCR AI model:

We will be using python to make our model.

Python is one of the most used languages in AI as it provides classes and libraries such as TensorFlow and PyTorch and a lot others, its powerful and easy syntax will be of a big help for us in making the model and the data preprocessing. In addition to that our team have a great experience in using python.

4.7 Assumptions and Dependencies:

4.7.1 Dependencies:

To develop the System:

- .Net 7.0 with ASP.Net
- Python with TensorFlow and Karas
- Node.js with React.js
- SQL Server relational database

To run the System:

- .Net 7.0 with ASP.Net on the server
- Python with TensorFlow and Karas on the server
- Any web browser on the client side.

4.7.2 Assumptions:

- The company has a local network with a server on it.
- The server hardware will have enough resources to run our backend and handle all requests.
- Each employee has access to a desktop device with a web browser installed, connected to the company's network.

4.8 Risk management:

Our project's development team are college juniors, there are no salaries and no budget to be calculated and no reputation to be hurt, which is why the risks for the project are low and minimal.

However, there are still some risks to consider and will be discussed.

4.8.1 Risk identification:

1- Scope creep: this risk was realized when we were presented with the project's idea, lack of clarity presented a lot of unnecessary features, and difficulty in prioritizing them.

2- Project's low performance: identified when the team were designing the system, some features of the project could hinder the overall performance, like search feature.

3- Delay in development: could occur when some features require some team members to learn new technology to develop them.

4.8.2 Risk analysis:

1- Scope creep:

-Likelihood: unlikely -Impact: High

Vision drift could lead the team to develop a completely different product, which results in redevelopment or removal of some parts of the product.

This is unlikely to occur because the team members were presented with many examples of how document management systems work and what is their goal.

2-Low performance:

-Likelihood: may occur -Impact: Medium

Performance is very important in our project, because part of our goal is to be efficient and simple, so features like search and OCR need to be quick.

However, project performance could be improved during testing and development, which is why it doesn't have high impact at this time.

3-Delay in development:

-Likelihood: may occur -Impact: Medium

Project's delay happens in all development processes to most teams for many reasons, but in our case it is likely to happen because some technologies need time to be trained on some team members, but it's still manageable.

As shown in fig 21, this is the relationship between impact and likelihood of the three risks types.

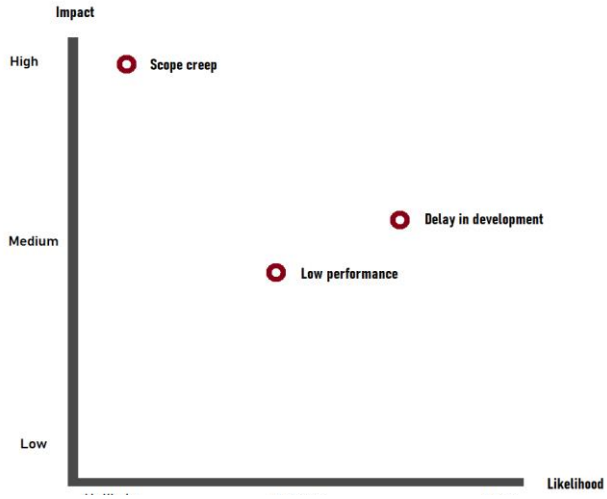

Fig. 21. Relationship between impact and likelihood.

4.8.3 Plan and action:

1- To prevent scope creep, a project roadmap was agreed on by team members, each team member understand the objective of this project and what the final product is expected to be and what the goal is.

2- Low performance risk could be reduced by continuous testing and improvement, also we may consider using third party tools (like an open source search engine) to use in our project, which may result in better performance.

3-Our team members are flexible, when delay occurs to a member, another member can help which should reduce the risk.

4.9 Design of Database ERD:

4.9.1 Entity Relationship Diagram (ERD):

This ERD design describes the way in which we are intending to store files and file hierarchy in database. In addition to admins, users, their roles and their relationships as shown in fig 22 below.

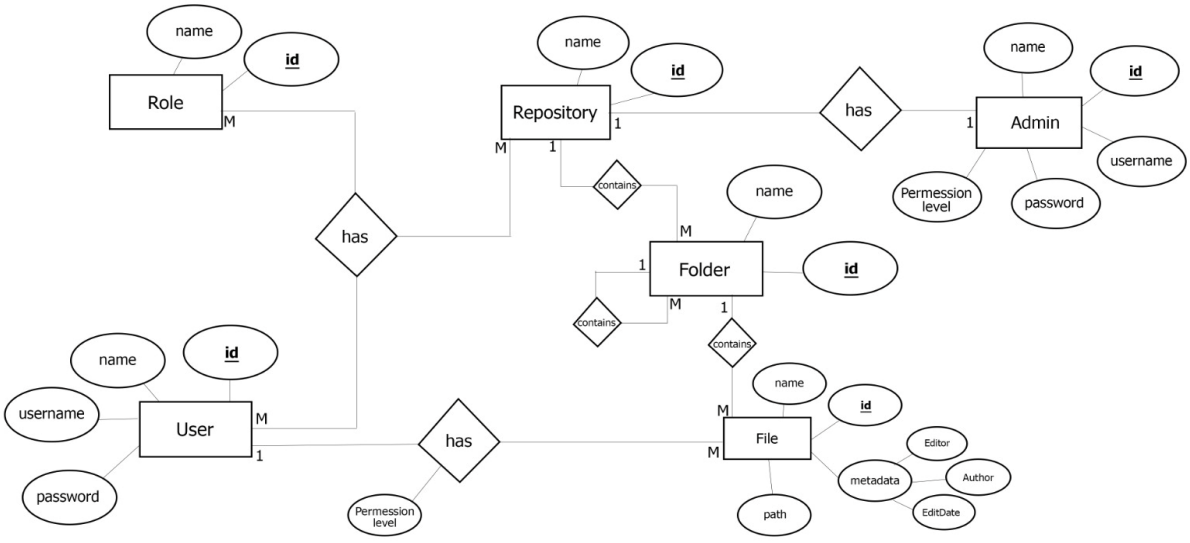

Fig. 22. Entity Relationship Diagram (ERD)

4.9.2 Mapping of Entity Relationship Diagram (ERD):

As shown in figure 23 this figure describes the mapping of the above-mentioned ERD as tables in the RDBMS.

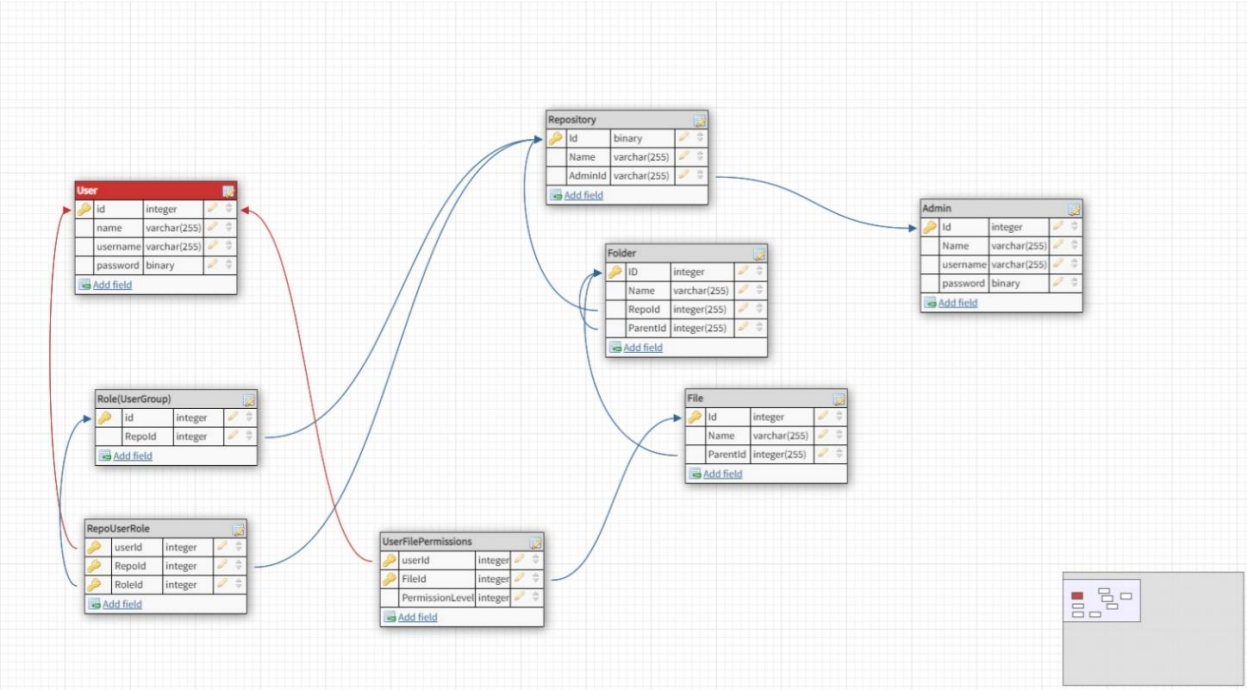

Fig. 23. Mapping of Entity Relationship Diagram.

4.10 Use Case Diagram:

4.10.1 Primary Use Case Diagram:

As shown in figure 24 below. The diagram describes the system use cases of the different types of users, it shows which use case is used by which actor.

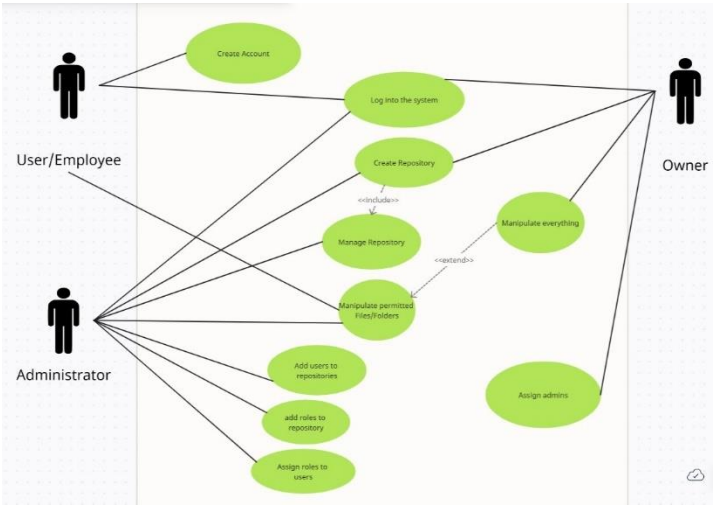

Fig. 24. Primary Use Case Diagram.

4.11 Activity Diagram:

As shown in figure 25 below. The activity diagram states who does what in a process.

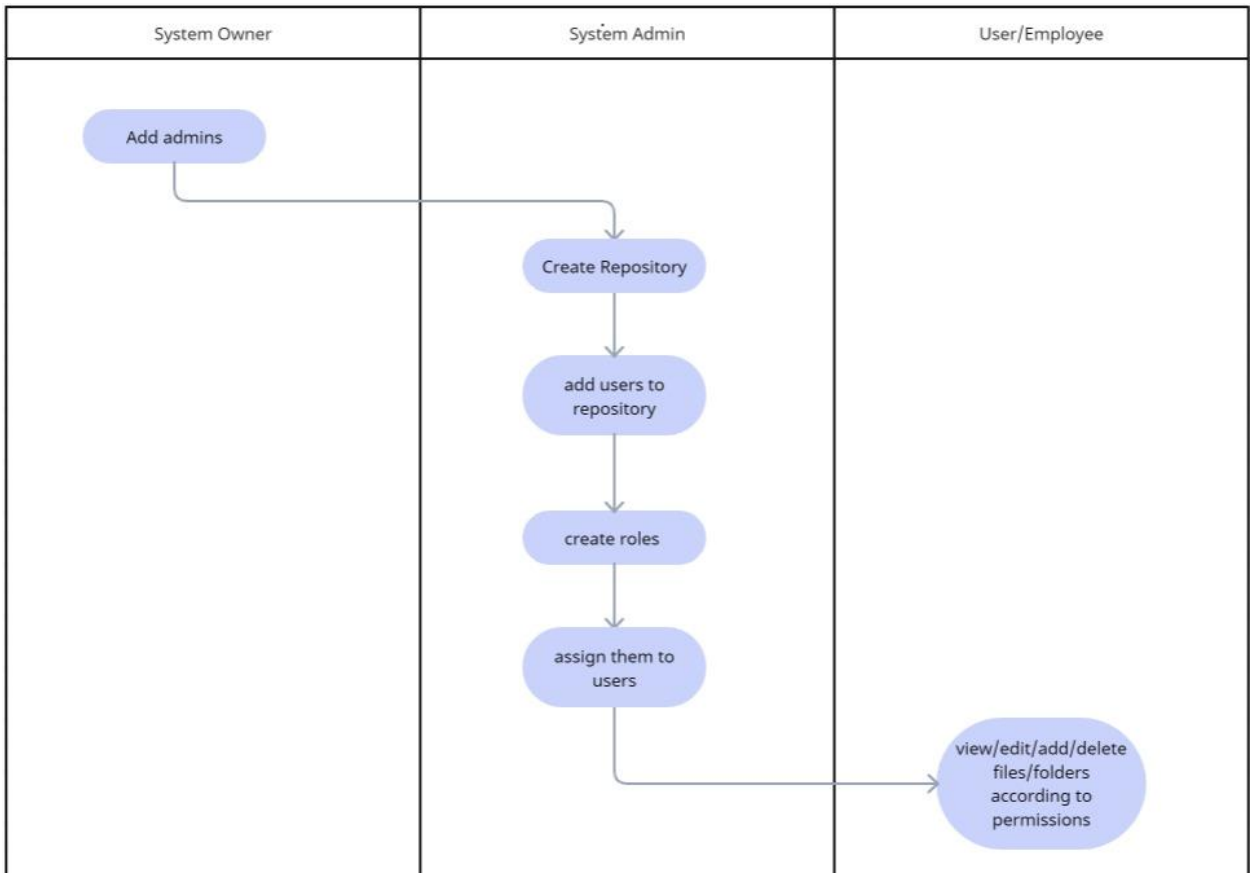

Fig. 25. Swimlane Activity Diagram.

4.12 State Diagram:

As shown in figure 26 below. State diagram describes the behavior of the system's states.

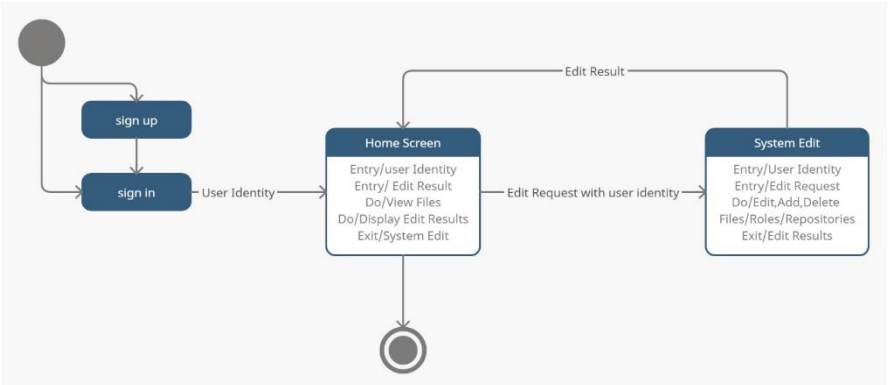

Fig. 26. State Diagram

References:

[1] https://www.folderit.com/blog/a-brief-history-of-document-management

[2] https://history-computer.com/optical-character-recognition/

[3] Jacobs, Charles, et al. "Text recognition of low-resolution document images." Eighth International Conference on Document Analysis and Recognition (ICDAR'05). IEEE, 2005.

[4] Djedović, Almir, et al. "Optimization of the business processes via automatic integration with the document management system." 2016 International Conference on Smart Systems and Technologies (SST). IEEE, 2016.

[5] Sharma, Pooja, and Shanu Sharma. "Image processing based degraded camera captured document enhancement for improved OCR accuracy." *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)*. IEEE, 2016.

[6] Orlov, V. L., and E. A. Kurako. "Electronic document management systems and distributed large-scale systems." 2017 Tenth International Conference Management of Large-Scale System Development (MLSD). IEEE, 2017.

[7] Chia, Wai Chong, Phoey Lee Teh, and Colin Mathew Hew D. Gill. "Text extraction and categorization from watermark scientific document in bulk." 2018 3rd International Conference on Computational Intelligence and Applications (ICCIA). IEEE, 2018.

[8] El Bahi, Hassan, and Abdelkarim Zatni. "Text recognition in document images obtained by a smartphone based on deep convolutional and recurrent neural network." Multimedia tools and applications 78.18 (2019): 26453-26481.

[9] Ali, Ragimova Nazila, Vugar Hajimahmud Abdullayev, and Vasila Soltanaga Abbasova. "Analysis of main requirements for electronic document management systems." ScienceRise 1 (2020): 28-31.

[10] https://www.wrike.com/blog/what-is-document-management-system/#What-is-document-management

[11] https://dotnet.microsoft.com/en-us/learn/dotnet/what-is-dotnet

[12] https://dotnet.microsoft.com/en-us/learn/aspnet/what-is-aspnet

[13] https://www.guru99.com/sql-server-introduction.html

[14] https://towardsdatascience.com/transfer-learning-on-greyscale-images-how-to-fine-tune-pretrained-models-on-black-and-white-9a5150755c7a
[15] EE 583: Digital Image Processing Course by Hasan Demirel
[16] https://datacarpentry.org/image-processing/07-thresholding/
[17] https://scikit-image.org/docs/stable/auto_examples/applications/plot_thresholding.html
[18] https://towardsdatascience.com/segmentation-in-ocr-10de176cf373
[19] https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-a-neural-network/
[20] https://towardsdatascience.com/convolution-neural-network-for-image-processing-using-keras-dc3429056306
[21] https://arxiv.org/abs/2109.14545
[22] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53
[23] https://towardsdatascience.com/dropout-in-neural-networks-47a162d621d9#:~:
[24] https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e
[25] https://medium.com/analytics-vidhya/undestanding-recurrent-neural-network-rnn-and-long-short-term-memory-lstm-30bc1221e80d
[26] https://towardsdatascience.com/build-a-handwritten-text-recognition-system-using-tensorflow-2326a3487cd5
[27] https://wandb.ai/authors/text-recognition-crnn-ctc/reports/Text-Recognition-With-CRNN-CTC-Network--VmlldzoxNTI5NDI