DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
UNIVERSITY OF MOHAMED BOUDIAF - M'SILA

FACULTY OF MATHEMATICS
AND INFORMATICS

DEPARTMENT OF COMPUTER SCIENCE

A dissertation in fulfillment
for the requirement of the degree of master in computer science

DOMAINE: Mathematics and Computer science.

FILIERE: Computer science**.**

OPTION: Business intelligence and optimization.

By: Ahlem YOUCEFI **and** Rim BEZAZ

Entitled

OCR detection of checkboxes from PDF files.

Presented publicly to the jury:

Dr. BOUGUERRA Abdelbaki	**University of M'sila**	**President**
Dr. LOUNNAS Bilal	**University of M'sila**	**Supervised**
Dr. THERAFI Abdellah	**University of M'sila**	**Examiner**

Academic Year 2021/2022

Aknowledgment:

In The Name of ALLAH, The Most Beneficent, The Most Merciful.

First of all, we want to pass a big thanks to our dears mothers and fathers for all those efforts they make to make us feel comfortable.

And we would like to thank my supervisor dr.LOUNNAS Bilal, for their advices and for being patient with us all the time.

We would also thank dr.BENOUIS mohamed and our colleague Makri Aymen Abderraouf for thier help.

DEDICATE:

I dedicate this dissertation to the soul of my grand-father Youcefi Omar may god bless him.

Youcefi Ahlem

I dedicate this dissertation to my lovely mother with my prayers for her healing.

Bezaz Rim

Contents

List of Figures

1

General introduction

Since the dawn of the technologie, human have been obsessed with automating everything. Artificial Intelligence (AI) is a broad area of computer science, it can be defines as making the machine able to act like human an this include thinking, learning, interacting with the external environment without any human intervention. [54]. These machines simulate natural intelligence and human behavior to solve complex problems.

Machine Learning (ML) is a branch of artificial intelligence that uses statistical methods and computer algorithms to build predictive models. The ML models can automatically learn from experience and improves its performance without explicitly being programmed.

Machine learning algorithms could be considered as a mathematical model containing many parameters that map inputs (or features) into one or more outputs (or targets).The broad categories of machine learning algorithms are supervised and unsupervised learning, The Supervised learning involves training a model by providing training labels .Meanwhile, unsupervised learning does not involves training labels [38].

Deep learning (DL) is a subset of machine learning where artificial neural networks algorithms inspired by the way a human brain filters information, learn from enormous amount of both structured and unstructured data. The algorithms usually have many (deep) layers [76]

Pattern Recognition is an important and fast evolving field. it refers to data analysis method that automatically recognize patterns in any data form such as text, images and sounds, this process done by using a machine learning algorithms. Pattern recognition systems can find regularities and similarities in data and recognize shapes from different angles to identify and classify unfamiliar objects [66].
Pattern recognition has many applications, including image processing, fingerprint recognition, aerial photo interpretation, optical character recognition and other many applications. The Optical Character Recognition (OCR) is a methodology used to convert text images into editable text documents. OCR is basically the electronic interpretation of manually written pictures into machine-editable content [42].
The ability to convert a document to an electronic format is a challenging task and growing need around the world, many techniques have been proposed in the literature for finding letters and numbers, however checkboxes are often overlooked and very difficult to detect. Determining whether checkboxes are checked or unchecked is a very useful tool on automatic reading of forms, including medical and QCM forms.
This dissertation will present an algorithm for checkboxes detecting by using optical character recognition algorithm with a deep learning model that uses the Convolutional Neural Network CNN. Our algorithm process start by converting the scanned PDF file into image,then applying morphological image processing for solving various image problems such as the noises and broken shapes , after that the algorithm detected the checkboxes and their position by using OCR algorithm . The process continuous by cropping the checkboxes parts from the image based on their location. Finally, the algorithm used the Conventional Neural Network (CNN) to classify the cropped checkboxes to either checked or unchecked.

Our dissertation consists of three chapters, the details of these chapters are as follows: The first chapter provides an overview of pattern recognition . First, we defined the pattern recognition and explained how it works, then we presented various pattern recognition applications. The first chapter focused on Optical Character Recognition by mentioning its definition, characters recognition's Branches, its Principle, types of Ocr and Technique and applications of OCR. Finally, we discussed the image processing and its techniques and applications. The second chapter offers a summary of Artificial intelligence . We have defined Artificial intelligence , then we have discussed its Types and technologies .
Furthermore, the second chapter gives a detailed presentation of artificial intelligence, Machine learning and deep learning. In the third and the last chapter, we presented our contribution in checkbox detection. We have proposed a new method based on using Optical Character Recognition (OCR) and the Convolutional neural network CNN. The third chapter discusses also the experimental results on our method.

2

Pattern recognition

In this chapter we are going to describe information about pattern recognition its definition and principle,OCR and how it works and the area of use and finally Image Processing there types and technics and its applications.

2.1 Pattern Recognition

2.1.1 Definition

Pattern recognition is often considered part of artificial intelligence [49] it is the scientific discipline whose goal is the classification of objects into a number of categories or classes. Depending on the application, these objects can be images or signal waveforms or any type of measurements that need to be classified [83]. Typical inputs to a PR system are images or sound signals, out of which the relevant objects have to be found and identified . The pattern recognition process can be viewed as a twofold task, namely , developing decision rules based on human knowledge (learning) and using them for decision making regarding an unknown pattern (classification) . The problem of pattern recognition is divided into two parts. The first part is concerned with the study of recognition mechanism of patterns by human and other living organisms. This part is related to the disciplines such as physiology, psychology, biology, and so on. The second part deals with the development of theory and techniques for designing a device that can perform the recognition task automatically [49].

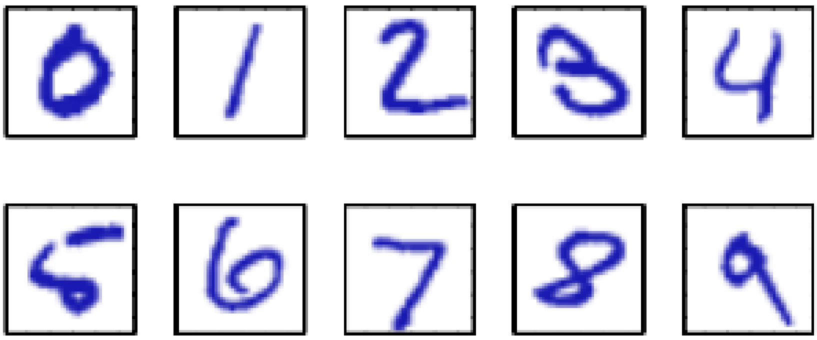

Figure 2.1: Example of hand writing digits taken from us zip codes [37].

2.1.2 How pattern recognition works

Pattern recognition involves making sense or identifying the objects we see, this technique is known as the template matching hypothesis and the feature detection model [36]. A template is a pattern used to produce items of the same proportions. The template-matching hypothesis suggests that incoming stimuli are compared with templates in the long term memory. If There is a match, the stimulus is identified [15]. In addition to this, the process of pattern matching depends on the type of the output, on whether learning is supervised or unsupervised [36].

2.1.3 Applications of pattern recognition

Because of the increased demand, information handling and retrieval are becoming increasingly important, propelling pattern recognition to the forefront of today's engineering applications and research. Pattern recognition has become an essential component of most machine intelligence systems as a result of this evolution. As a result, there are numerous patterns that can be identified using the appropriate techniques [83]

- Machine vision : Machine vision is an area in which pattern recognition is of importance [83]. Its system deals with the complete solution from illuminating the object, capturing the image, processing analysis of the image, making a decision and passing the results on to the automation controller [18]. A typical application of a machine vision system is in the manufacturing industry, either for automated visual inspection or for automation in the assembly line [83].

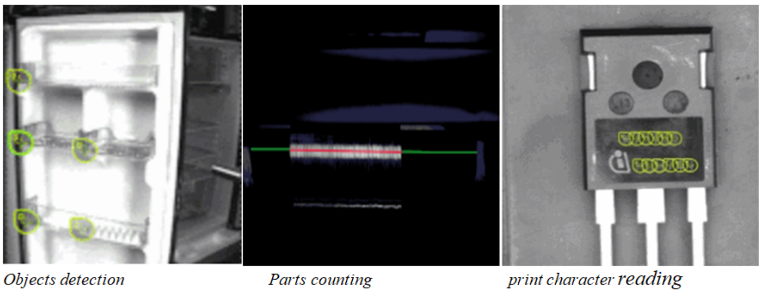

Figure 2.2: Objects detection,parts counting, print character reading Examples [20]

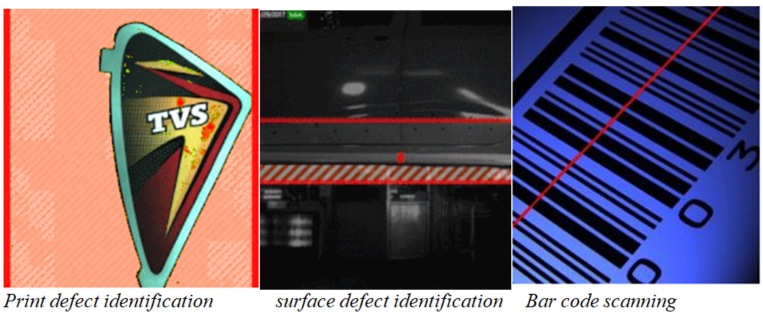

Figure 2.3: Print defect identification,surface defect identification,bar code scanning [20]

- Computer-aided diagnosis : Computer-aided diagnosis is an important pattern recognition application that aims to assist doctors in making diagnostic decisions. Computer-assisted diagnosis has been applied to and is of interest for a wide range of medical data, including X-rays and computed tomographic images [83].

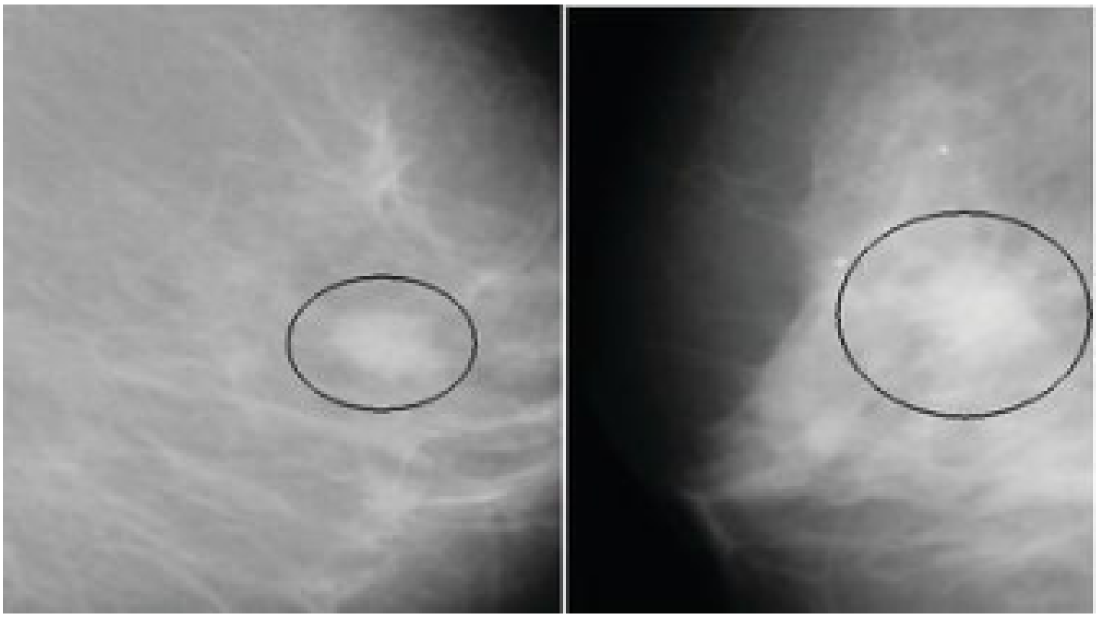

Figure 2.4: mass in mammography detection [21]

- Speech recognition : Speech recognition is a field in which a significant amount of research and development has been invested. The most natural way for humans to communicate and exchange information is through speech. As a result, scientists, engineers, and science fiction writers have long sought to create intelligent machines that recognize spoken information [36]

Figure 2.5: speech recognition [36]

- Data mining and knowledge discovery in databases :

  Is another key application area of pattern recognition. Data mining is gaining popularity in a variety of fields, including medicine and biology, market and financial analysis, business management, science exploration, and image and music retrieval. Its popularity stems from the fact that in the age of information and knowledge society there is an ever increasing demand for retrieving information and turning it into knowledge [83].

- Character (letter or number) recognition : Optical character recognition (OCR) technology is a business solution for automating data extraction from printed or written text from a scanned document or image file and then converting the text into a machine-readable form to be used for data processing like editing or searching [19].

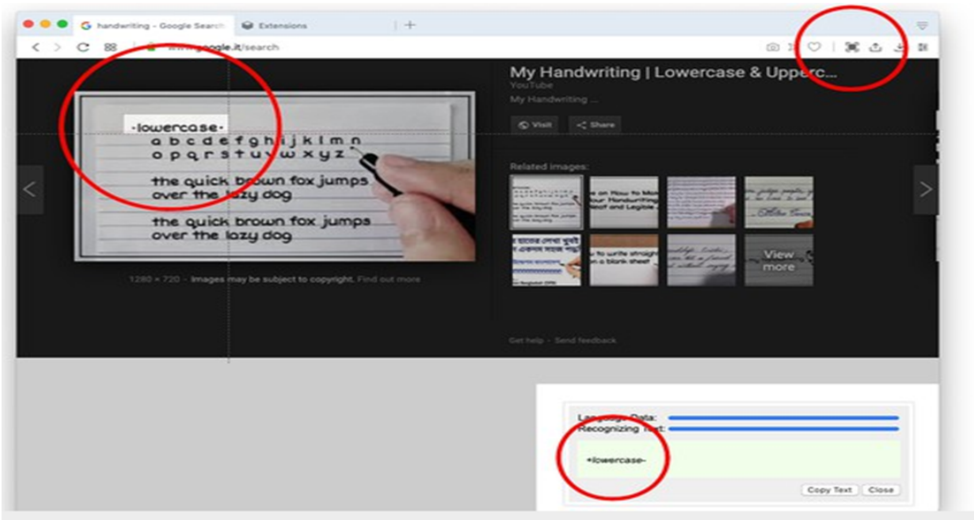

Figure 2.6: Ocr example [22]

2.2 Optical Carecter Recognition OCR

2.2.1 Definition

OCR is an input technology that aims to recognize characters such as letters, numbers, and special characters, whether handwritten or typed or written down. The recognition result is returned as a coded representation of the input characters. These character codes may be used for all kinds of computer processing, from storage in databases to machine translation [94].

2.2.2 Branches of characters recognition

1. **Online Recognition:** data is acquired in real-time such as digitizer tablets [63]. This kind of data is known as digital ink and can be regarded as a digital representation of handwriting [26].

2. **Offline Recognition:** systems collect data from static devices such as scanners and cameras. off-line recognition systems require specific methodologies to prepare the image for recognition process and to eliminate damage (noise and errors) of the input image caused by collection process [63].

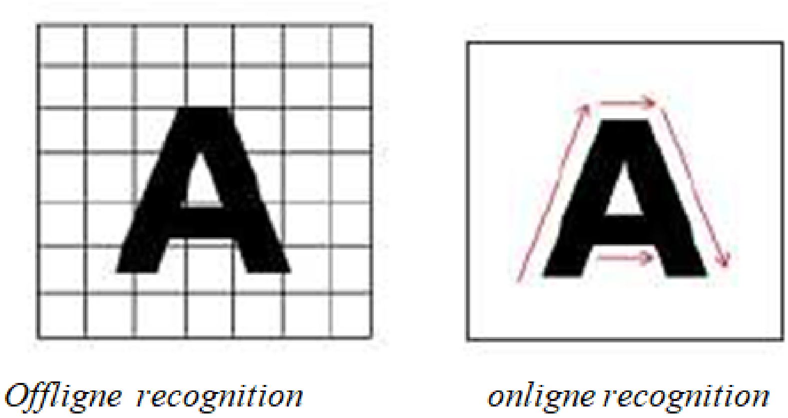

Figure 2.7: online/offline recognition [34].

2.2.3 Principle

The first step in OCR is to use an optical scanner to digitize an analog document. When text-containing regions are found, each symbol is extracted using a segmentation process. To facilitate feature extraction, the extracted symbols are pre-processed to remove noise. Each symbol's identity is determined by comparing extracted features to descriptions of symbol classes obtained during a previous learning phase. Finally, contextual information is used to reconstruct the original text's words and numbers [42].

2.2.4 Types of Ocr

1. Intelligent Word Recognition: IWR captures cursive text or handwritten texts. Their algorithm works by recognizing an entire unconstrained handwritten word rather than picking up individual characters.

2. Intelligent Character Recognition: ICR captures handwritten or cursive text. The engine works by identifying a single character at a time and evolves with its embedded machine learning.

3. Optical Word Recognition: OWR Targets typewritten text wordwise and is sometimes referred to as OCR

4. Optical Character Recognition: OCR captures typewritten text and goes one character at a time [17].

5. **Optical Mark Recognition**

   a) Definition: OMR Optical mark recognition (also called optical mark reading and OMR) is the process of reading information that people mark on surveys, tests and other paper documents. OMR is used to read questionnaires, multiple choice examination papers in the form of shaded areas.

   b) Field types of OMR [14]:
   OMR has different fields to provide the format the questioner desires. These fields include:

      - Multiple, where there are several options but only one is chosen. For example, the form might ask for one of the options ABCDE; 12345; completely disagree, disagree, indifferent, agree, completely agree; or similar.

      - Grid: the lines are set up in a grid format for the user to fill in a phone number, name, ID number and so on.

      - Add, total the answers to a single value.

      - Boolean, answering yes or no to all that apply.

      - Binary, answering yes or no to only one.

      - Dotted lines fields, developed by Smartshoot OMR, allow border dropping like traditional color dropping.

2.2.5 Technique of OCR

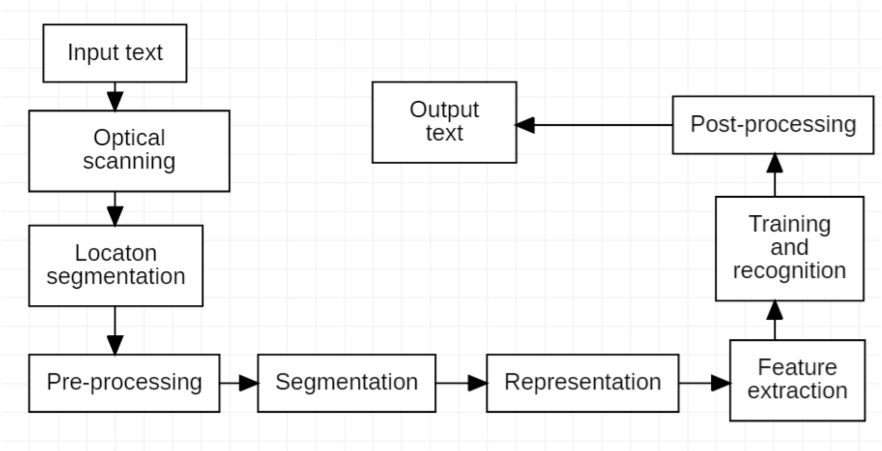

Figure 2.8: Techniques of ocr [42].

2.2.5.1 Optical Scanning :

The first component in OCR is optical scanning .Through scanning process digital image of original document is captured .In OCR optical scanners are used which consist of transport mechanism and sensing device that converts light intensity into grey levels. Printed documents consist of black print on white background.When performing OCR multilevel image is converted into bi-level black and white image [42].

2.2.5.2 Location Segmentation :

The next OCR component is location segmentation. segmentation is isolation of characters or words. Most of OCR algorithms segment words into isolated characters which are recognized individually. Usually segmentation is performed by isolating each connected component [42]. The constituents of an image are determined by segmentation. It is necessary to identify document regions that contain printed data and are distinct from figures and graphics [42].

2.2.5.3 Pre-processing :

The third OCR component is pre-processing. The image that is created as a result of the scanning process may contain some noise. Characters may be smeared or broken depending on scanner resolution and inherent thresholding. Some of these flaws, which can lead to low recognition rates, are removed by smoothing digitized characters in the pre-processor. Smoothing implies both filling and thinning. Filling eliminates small breaks, gaps and holes in digitized characters while thinning reduces width of line [42]. Pre-processing also includes normalization along with smoothing. The normalization is applied to obtain characters of uniform size, slant and rotation [42]. The main objectives of pre-processing can be pointed as:

a) Noise reduction: The noise introduced by the optical scanning device or the writing instrument causes disconnected line segments, bumps and gaps in filled loops, etc [42]. The noise reduction techniques can be categorized in three major groups:

   - The basic idea behind filtering is to convolute a predefined mask with the image to assign a value to a pixel as a function of the gray values of its neighboring pixels [42].

   - morphological operations :The basic idea behind morphological operations is to filter the character image using logical operations instead of convolution operations.

   - noise modeling : There exists some available literature on noise modeling introduced by optical distortion such as speckle, skew and blur. It is also possible to assess the quality of the character images and remove the noise to a certain degree [42].

b) Normalization: The normalization methods aim to remove the variations of thewriting and obtain standardized data [42]. Some of the commonly used methods for normalization are :

   - Skew normalization and baseline extraction:Due to inaccuracies in the scanning process and writing style the writing may be slightly tilted or curved within the image. After skew detection the character or word is translated to the origin, rotated or stretched until the baseline is horizontal and retranslated back into the display screen space [42].

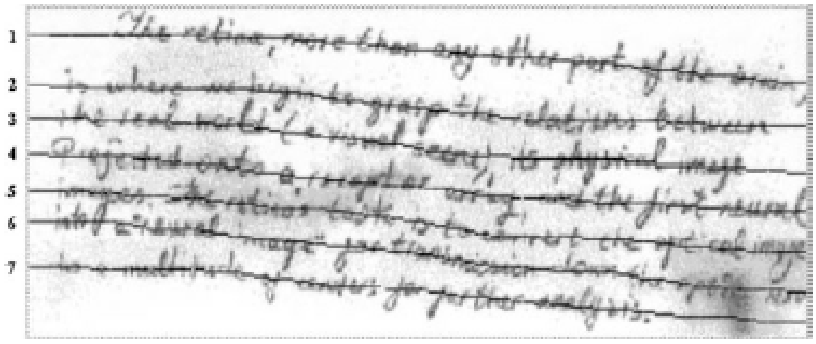

Figure 2.9: The baseline extraction using attractive and repulsive network [42]

- Slant normalization : Slant normalization is used to normalize all characters to a standard form. The most common method for slant estimation is the calculation of the average angle of near vertical elements [42].

Figure 2.10: Slant normalization example [42]

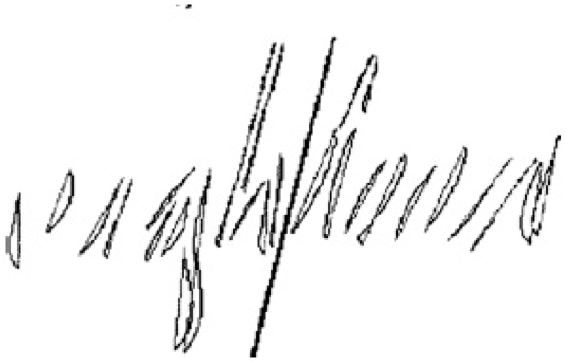

Figure 2.11: Slant normalization example [42]

- Size Normalization :is used to adjust the character size to a certain standard. The OCR methods may apply for both horizontal and vertical size normalizations. The character is divided into number of zones and each of these zones is separately scaled [42].

- Contour smoothing : Contour smoothing eliminates the errors due to the erratic hand motion during the writing. It generally reduces the number of sample points needed to represent the script and thus improves efficiency in remaining pre-processing steps [42].

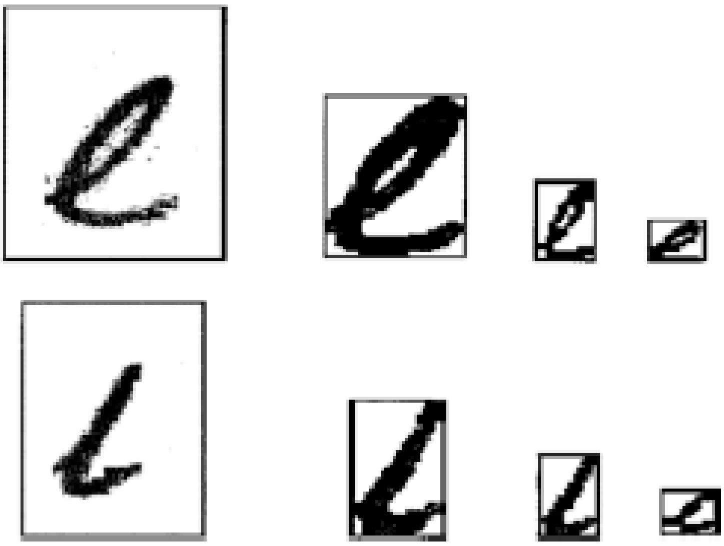

Figure 2.12: The normalization of characters [42]

c) Compression:The compression for OCR requires space domain techniques for preserving the shape information. The two popular compression techniques used are: [42]

- Thresholding: It is often preferable to represent gray scale or color images as binary images by selecting a threshold value in order to reduce storage requirements and increase processing speed Č.

- Thinning:While it provides a tremendous reduction in data size, thinning extracts the shape information of the characters. Thinning can be considered as conversion of offline handwriting to almost online like data with spurious branches and artifacts [42].

2.2.5.4 Segmentation

Here the character image is segmented into its subcomponents. Segmentation is important because the extent one can reach in separation of the various lines in the characters directly affects the recognition rate. Internal segmentation is used here which isolates lines and curves in the cursively written characters [42].

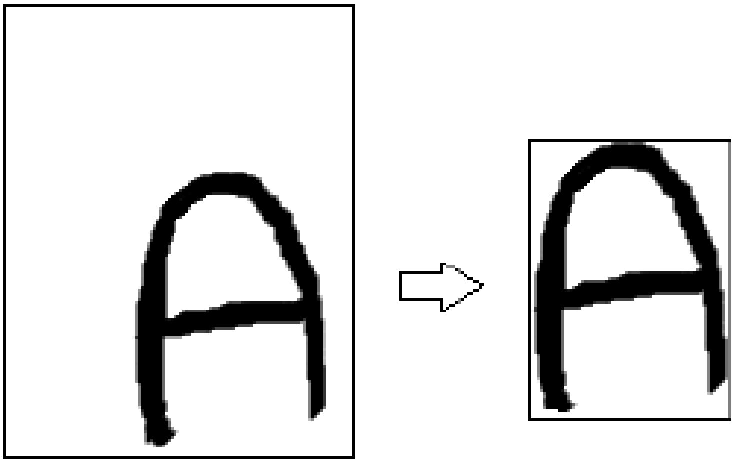

Figure 2.13: Segmented image [41]

2.2.5.5 Representation

The image representation plays one of the most important roles in any recognition system. In the simplest case, gray level or binary images are fed to a recognizer. the algorithms, a more compact and characteristic representation is required. For this purpose, a set of features is extracted for each class that helps distinguish it from other classes while remaining invariant to characteristic differences within the class [42].

2.2.5.6 feature extraction

The sixth OCR component is feature extraction. The objective of feature extraction is to capture essential characteristics of symbols. Feature extraction is accepted as one of the most difficult problems of pattern recognition [42].

2.2.5.7 Training and Recognition :

The seventh OCR component is training and recognition. OCR systems extensively use the methodologies of pattern recognition which assigns an unknown sample into a predefined class [42]. The OCR are investigated in four general approaches of pattern recognition as suggested in:

a) Template matching:The simplest way of OCR is based on matching the stored prototypes against the character to be recognized[42]. Template matching is the process of finding the location of sub image called a template inside an image. Once a number of corresponding templates is found their centers are used as corresponding points to determine the registration parameters [71]. Generally speaking, matching operation determines the degree of similarity between two vectors such as group of pixels, shapes, curvature etc [42]

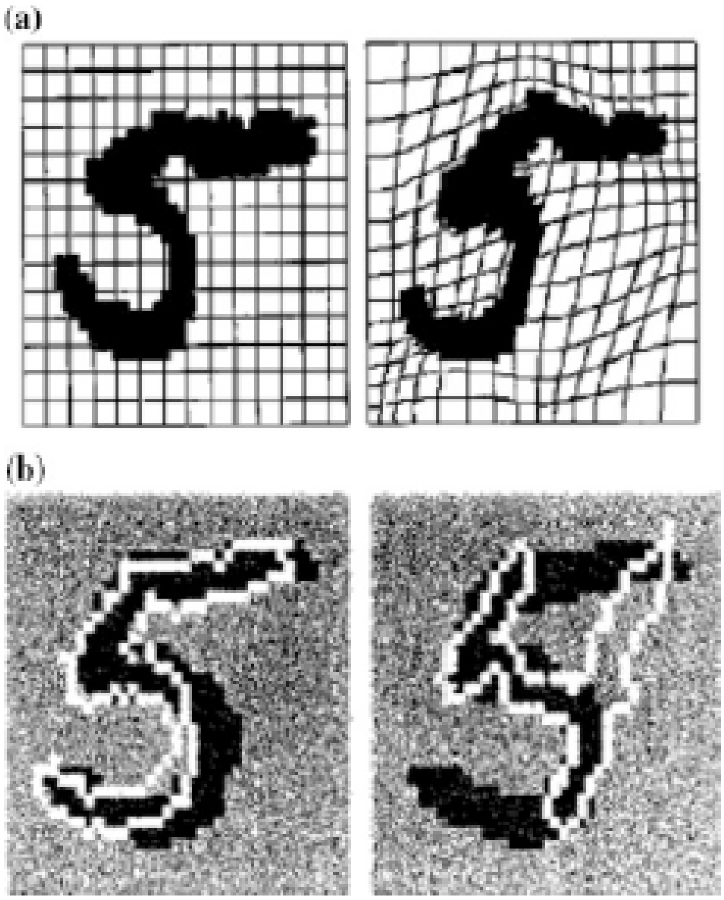

Figure 2.14: A The deformable templates: deformations of a sample digit [42].

b) Statistical techniques: is to determine to which category the given pattern belongs. By making observations and measurement processes, a set of numbers is prepared, which is used to prepare a measurement vector [31]. The statistical decision theory is concerned with statistical decision functions and a set of optimality criteria which maximizes the probability of the observed pattern given the model of a certain class [42].

c) Structural techniques: The recursive description of a complex pattern in terms of simpler patterns based on the shape of the object was the initial idea behind the creation of the structural pattern recognition [42].Structural algorihtm classifies the input patterns on the basis of components of the characters and the relationship among these components. Firstly the primitives of the character are identified and then strings of the primitives are checked on the basis of pre-decided rules [31].

d) Artificial Neural Network (ANN): The ANN possess a massively parallel architecture such that it performs computation at a higher rate compared to the classical techniques. It adapts to the changes in data and learns the characteristics of input signal [42].

2.2.5.8 Post-processing :

Post-processing approaches are divided into two tasks: error detection and correction. The purpose of post-OCR error detection is to identify incorrect tokens in the input text. and it allows for the flagging of noisy data for reprocessing if necessary. Furthermore, error detection generates a list of detected errors for use as the input of post-OCR error correction, which is intended to correct invalid tokens. Typically, given an error, a list of word candidates is generated and scored based on the available information. In automatic correction approaches, the highest-ranked candidate is then chosen to correct the error [16].

2.2.6 Applications of OCR

Over the last few decades, commercial OCR products that meet the needs of various users have proliferated.

a) Aid for blind :Prior to the advent of digital computers and the requirement for This was an imagined application area that involved the input of large amounts of data. in order to read machines Such a reader, in conjunction with a speech synthesis system, enables unable to comprehend printed documents [42].

Figure 2.15: Text to speech in nepali software [23].

b) Automatic number plate readers: is systems combine high-speed cameras and sophisticated software to capture and convert license plate images into data that can be compared with information in other databases [52].

Figure 2.16: Automatic plate number recognition example [80].

c) Automatic cartography:Character recognition from maps presents unique challenges in the field of character recognition. The symbols are interspersed with graphics, the text is printed at various angles, and the characters are of various fonts.or even written by hand [42].

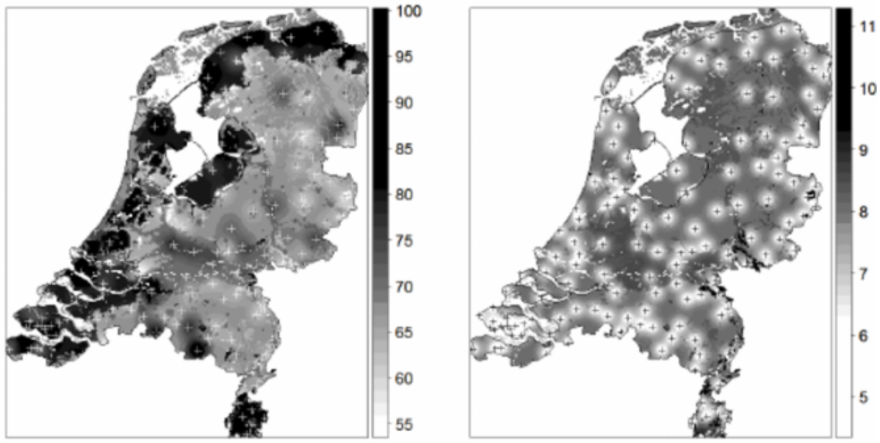

Figure 2.17: Example of two types of Automatic cartography [24].

d) Form readers: Such systems are able to read specially designed forms. In such forms all irrelevant information to reading machine is printed in color invisible to scanning device [42].

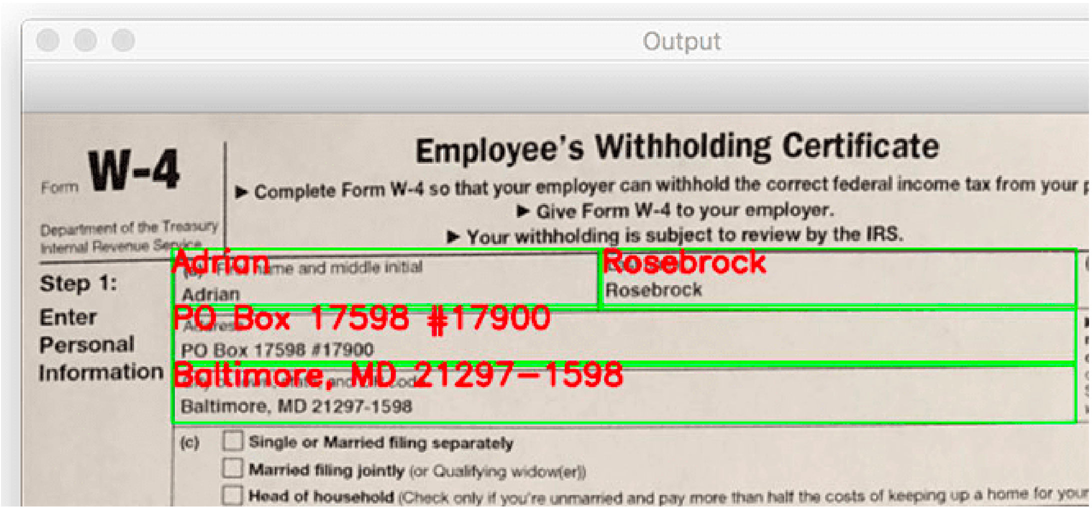

Figure 2.18: Form employees withholding certificate reader [25]

e) Signature verification and identification:This application is useful for banking environment. Such system establishes the identity of writer without attempting to Read handwriting [42].

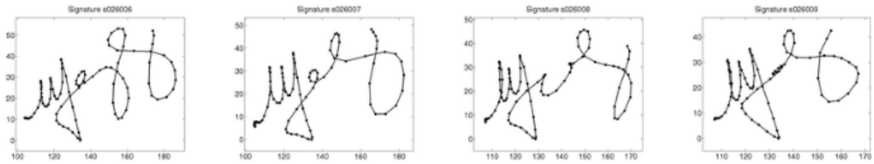

Figure 2.19: Four types of signature [72].

2.3 Image Processing

2.3.1 Definition

An image is a subset of a signal. A signal is a function that conveys information generally about the behavior of a physical system or attributes of some phenomenon. Signal processing is the process of extracting information from the signal. Digital signal processing (DSP) is concerned with the representation of signals bysequences of numbers or symbols and processing of these sequences. The purpose of such processing is to estimate characteristic parameters of a signal or to transform a signal into a form that is more sensible to human beings. DSP includes subfields such as digital image processing, video processing, statistical signal processing, signal processing for communications, biomedical signal processing, audio and speech signal processing, sonar and radar signal processing, sensor array processing, spectral estimation, and so on.... [49]. Image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too. Image processing basically includes the following three steps:

- Importing the image via image acquisition tools;
- Analyzing and manipulating the image;
- Output in which result can be altered image or report that is based on image analysis. [1]

2.3.2 Techniques of Image processing

2.3.2.1 Image Enhancement:

Image enhancement is the processing of images to increase their usefulness [58]. which improves the quality of an image for a specific purpose. The process depends up on the characteristics of the image and whether it is required for human perception or machine vision. Some features are enhanced to suit human or machine vision. For example, the spot noise is reduced in median filtering so that a better viewing of the original image is obtained. Edges are enhanced by highpass filtering and the output image is a step in computer vision [90].

Figure 2.20: Image enhanced [61].

Categories on image enhancement :

a) Spatial domain : Spatial domain technique enhances an image by direct manipulation of pixels in an image [82].

b) Spatial filtering :Spatial filtering process consists of moving the filter mask from point to point in the given image. Each point in the resultant image is the response of filter, calculated using a predefined relationship [82].

2.3.2.2 Image restoration :

Image restoration is a technique that allows you to recover the clear image from a degraded or corrupted image. Corrupted/Blur images are due to noisy, blur images or camera misfocus [61]. A restoration filter is designed based on the nature of the degradation process and the noiseDigital [90]. As a result, the images will be restored to their original quality by reducing physical degradation [61].

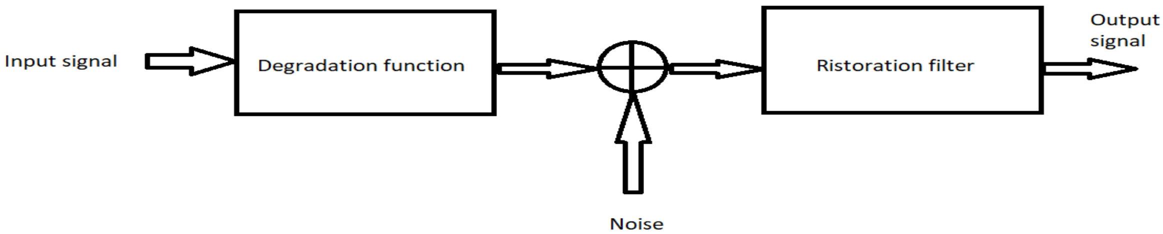

Figure 2.21: Image Restoration Method [61]

2.3.2.3 Image Compression :

Image compression is minimizing the size of bytes of a image file without degrading the quality of the image in order to obtain the image in more clarity[61]. Image compression reduces the storage space required to save an image by retaining the same physical size as that of the original image. The use of the compression techniques helps in the less usage of the space of hard disk [82].

Figure 2.22: Compressed image from 2.1mb into 306kb [6]

types of image compression :there are two types of image compression

- lossless compression: the original image can be reconstructed exactly from its compressed version [90].

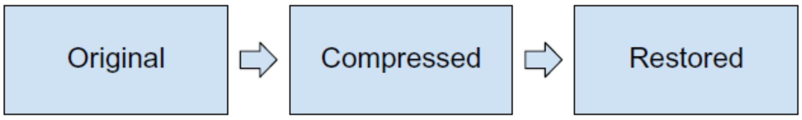

Figure 2.23: Lossless compression process [87].

- Lossy compression : is based on the fact that the magnitude of the frequency components of typical images decreases with increasing frequency [90].

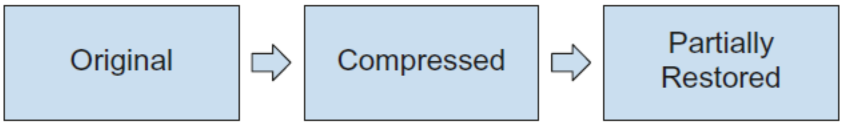

Figure 2.24: Lossy compression process [87].

2.3.2.4 Image segmentation :

Segmenting or partitioning the original image with some defined pixels into number of regions for the purpose of image analysis [61]. Image segmentation tries to identify and label the different objects and regions, the image is composed of. For example, the characters in a word have to be segmented before their identification [90].

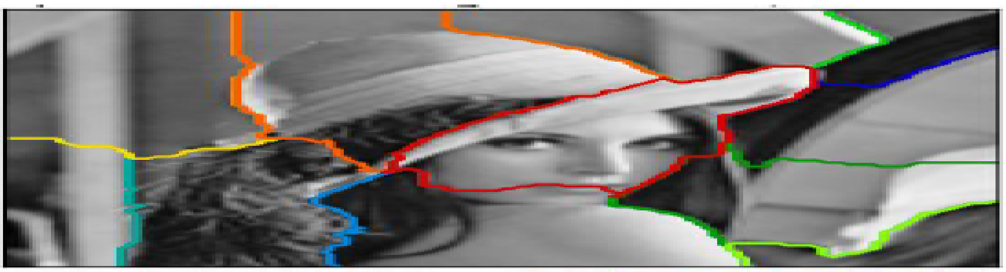

Figure 2.25: Image segmentation example [61].

2.3.2.5 Image recognizing

Image recognition technique involves in recognizing / identifying and detecting features such as objects in video or images. During the recognition mechanism, images from the database are compared with the current image, if the match is found then further execution of process will be carried out in real time application [61].

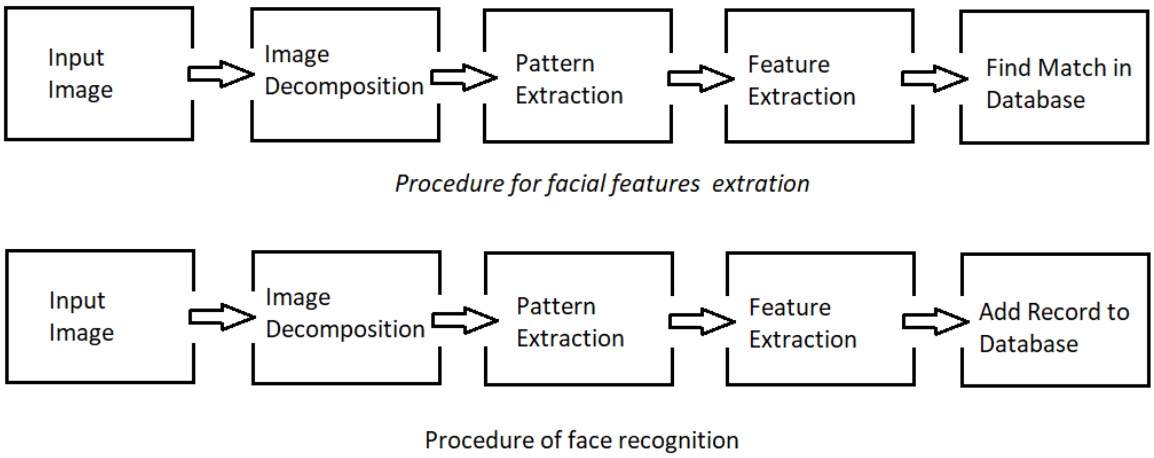

Figure 2.26: Procedure for facial featurs extration and face recognition [61].

2.3.2.6 Image smoothing :

Image smoothing is an operation thats used to remove noise, sharpness and clutter in the image to give you much more smoother and blended effect.site web evergreen technologies [27]. Image may contain noisy data such as dots, blur, speckles, stains, using this smoothing technique that acts as filter to remove the noisy data. It works Based on the low pass filter, which helps in decreasing the great difference between pixel values by averaging nearby pixel value. Considering single value calculated for an image such as median and average value [61].

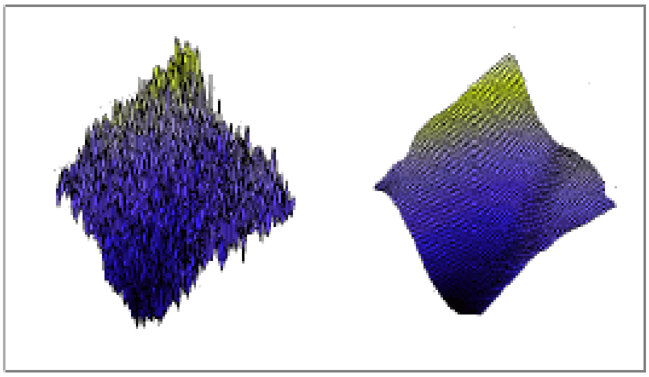

Figure 2.27: Examples for Smoothing Technique [61]

2.3.3 Application of image processing

- Remote sensing : Remote sensing technology is now used routinely to obtain accurate, timely information for a significant variety of applications, including: the study of daily weather and longterm climate change; urban-suburban land-use/land cover monitoring [57].

- Image transmission and storage for business: Its applications include broadcast television, teleconferencing, transmission of facsimile images for office automation, communication over computer networks, security monitoring systems, and military communications [49].

- Medical processing : Its applications include X-ray, cineangiogram , transaxial tomography, and nuclear magnetic resonance These images may be used for patient screening and monitoring or for detection of tumors or other diseases in patients [49].

- Radar, sonar, and acoustic image processing: For example, the detection and recognition of various types of targets and the maneuvering of aircraft [49].

- Robot/machine vision: Its applications include the identification or description of objects or industrial parts in 3D scenes [49].

2.4 Conclusion

We gave an overview of pattern recognition and its main applications in this chapter, then we focused on optical character and image processing and the techniques used to process them.

3

Artificial Intelligence

In this chapter we are going to talk about artificial intelligence and its technics and technologies, machine learning and its technics, area of use, and the deep Learning and the neural networks and the application of deep learning.

3.1 Artificial Intelligence

3.1.1 Definition

When we are about to define artificial intelligence, we find a lot of definitions, because it makes much discussion and conflit. So we collect some definitions of the artificial intelligence:

- An area of study in the field of computer science. Artificial intelligence is concerned with the development of computers able to engage in human-like thought processes such as learning, reasoning, and self-correction [60].

- The concept that machines can be improved to assume some capabilities normally thought to be like human intelligence such as learning, adapting, self-correction, etc [60].

- In a restricted sense, the study of techniques to use computers more effectively by improved programming techniques [60].

3.1.2 Types of AI

There are 2 principal types of AI based on AI capabilities and features.

3.1.2.1 Type 1: AI-Based Abilities

- **Weak or Artificial Narrow Intelligence (ANI):** A weak or narrow AI is a type of AI that performs assigned tasks using intelligence. It is the most common form of AI available in industries today. Narrow AI cannot operate beyond what is assigned to the system. Indeed, it is trained to perform a single specific task.ANI stands for all artificial intelligence machines created and deployed to date. All artificially intelligent systems that can perform a dedicated task autonomously using human-like capabilities fall into this category. As the name suggests, these machines have very limited responsibilities [10].
  The assertion that machines could act as if they were intelligent weak AI is called the weak AI or narrow AI hypothesis [89]. Narrow artificial intelligence (narrow AI) is a specific type of artificial intelligence in which a learning algorithm is designed to perform a single task, and any

knowledge gained from performing that task will not automatically be applied to other tasks. Most AI applications in use today can be categorized as being narrow AI [30].

- **Artificial General Intelligence (AGI):** Artificial general intelligence is a type of AI capable of performing any intellectual task as a human being. AGI machines are designed to perceive, learn and function entirely like humans. AGI systems are machines capable of reproducing human multifunction capabilities. AGI are designed to create multiple skills that can significantly reduce the time needed to train these machines [10]. Many philosophers have claimed that a machine that passes the Turing Test would still not be actually thinking, but would be only a simulation of thinking. The assertion that machines that could act as if they were intelligent called STRONG AI are actually thinking (not just simulating thinking) is called the strong AI hypothesis [89].The concept of strong artificial intelligence refers to a machine capable not only of producing intelligent behavior, in particular of modeling abstract ideas, but also of experiencing an impression of real consciousness, of "real feelings" (although one can put behind these words), and understanding of ones own reasonings [50].

Narrow AI	General AI
○ Application specific/ task limited	○ Perform general (human) intelligent action
○ Fixed domain models provided by programmers	○ Self-learns and reasons with its operating environment
○ Learns from thousands of labeled examples	○ Learns from few examples and/or from unstructured data
○ Reflexive tasks with no understanding	○ Full range of human cognitive abilities
○ Knowledge does not transfer to other domains or tasks	○ Leverages knowledge transfer to new domains and tasks
○ Today's AI	○ Future AI?

Figure 3.1: the difference between weak and string AI [30]

- **Super artificial intelligence (ASI):** Artificial super-intelligent systems can be described as the zenith of AI research. ASI is intended not only to reproduce multi-faceted human intelligence, but also to have faster memory, data processing and analysis capabilities. It is a hypothetical concept of artificial intelligence in which researchers try to develop machines capable of surpassing humans. This is the result of General AI [10].

3.1.2.2 Type 2: AI-based features

- **Reactive AI:** Reactive machines directly perceive the real world and react according to the environment. The intelligence of reactive machines consists in perceiving the real world directly and reacting to it. An example of a responsive machine is Google's AlphaGo [10].

- **Limited memory AI:** Limited-memory machines are those that can retain memory for a short period of time. These machines have the same capabilities as those of purely reactive machines. Moreover, machines with limited memory can learn from previous experiences to make decisions. For example, self-driving cars are limited-memory machines that can store data such as the distance the car has traveled with nearby cars, their recent speed, speed limit, lane [10].

- **Theory of mind AI:** Theory of mind can be defined as a simulation. Let's be clear, when a person thinks of them selves as another person, their brain tends to simulate that other person's mind. Theory of mind is essential to human cognition. Moreover, it is also crucial for social interactions. A break in the concept of theory of mind, for example, can be illustrated as a case of autism [10].

- **Self-aware AI:** Self-awareness machines hypothetically exist today. As the name suggests, these machines are supposed to be aware of themselves, like the human brain. Machines can be described as the ultimate goal of AI scientists. The goal of developing self-awareness machines is to make them capable of generating emotions and needs identical to those of humans [10].

3.1.3 Technologies

- **Speech recognition:** Speech recognition, or speech-to-text, is the ability of a machine to identify words spoken aloud and convert them into readable text [67]. And it is also the process of enabling a computer to identify and respond to the sounds produced in human speech [5].

Figure 3.2: speech recognition [53]

- **Natural language processing:** Natural Language Processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things [44].

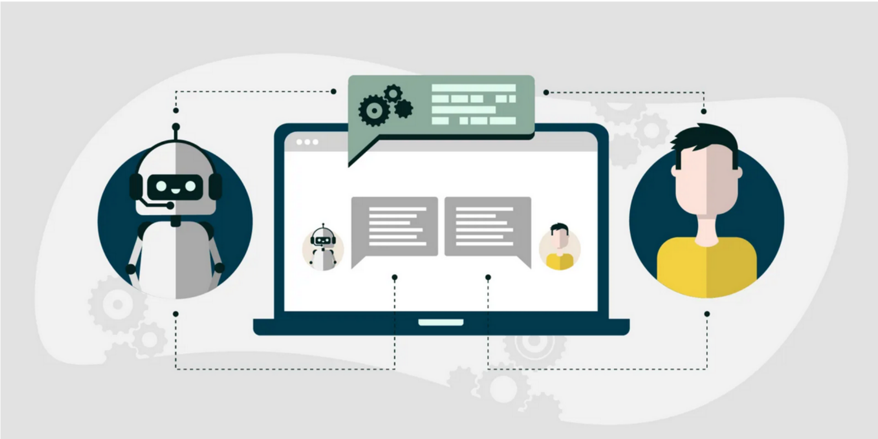

Figure 3.3: Natural language processing [33]

- **Computer vision:** Computer vision is a field of artificial intelligence that trains computers to interpret and understand the visual world. Using digital images from cameras and videos and deep learning models, machines can accurately identify and classify objects and then react to what they see [88].

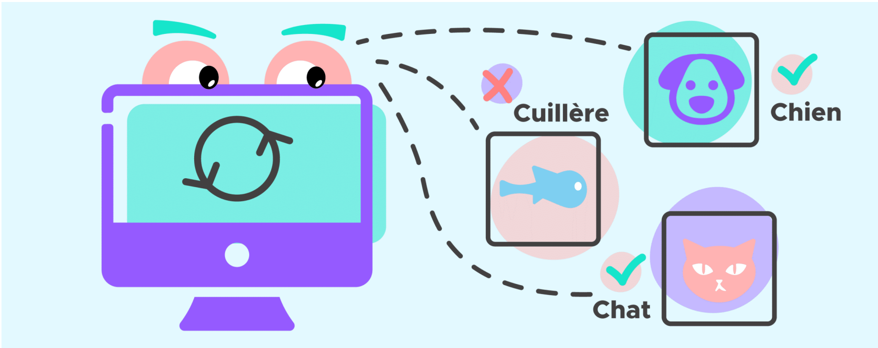

Figure 3.4: computer vision [91]

- **Machine learning:** It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention, is a method of data analysis that automates analytical model building [8]. Machine learning is programming computers to optimize a performance criterion using example data or past experience [32].

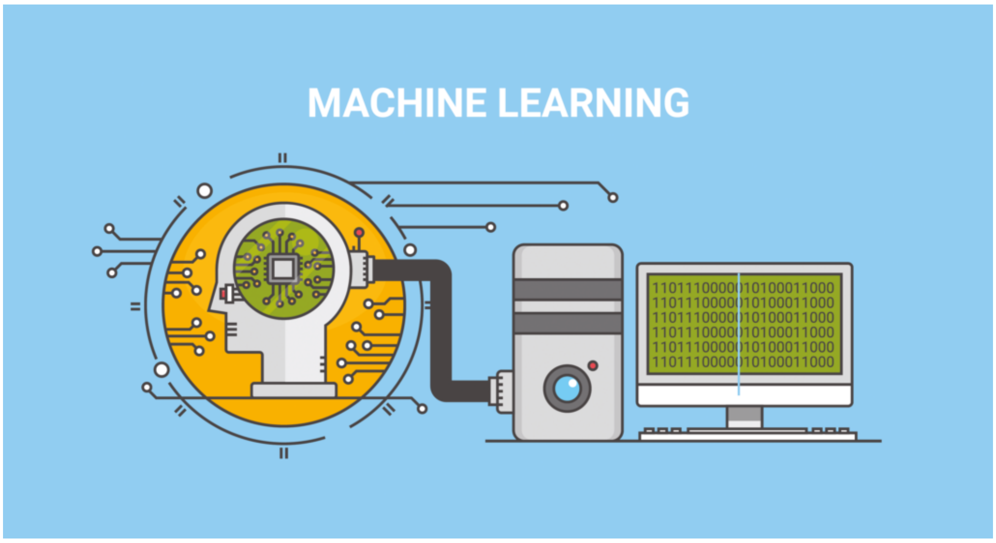

Figure 3.5: machine learning [55]

- **Expert systems:** Expert systems (ES) are a branch of applied artificial intelligence (AI), and were developed by the AI community in the mid-1960s. The basic idea behind ES is simply that expertise, which is the vast body of task-specific knowledge, is transferred from a human to a computer. This knowledge is then stored in the computer and user [85].

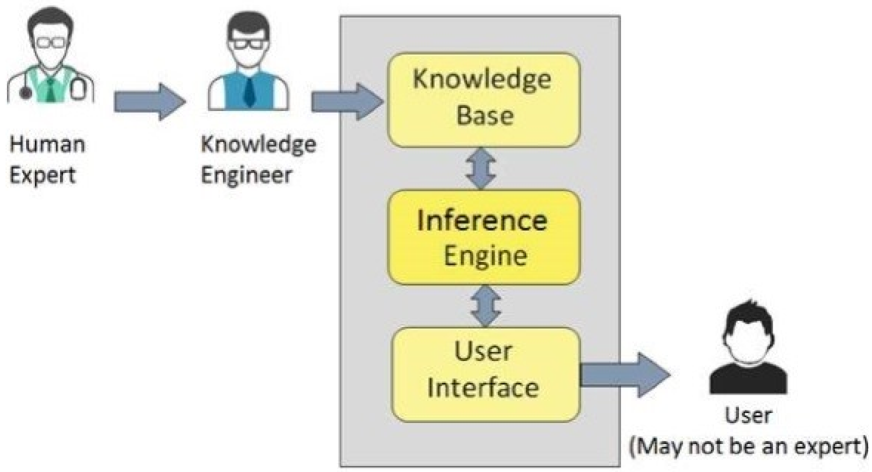

Figure 3.6: expert system [2]

3.2 Machine Learning

3.2.1 Definition

Machine learning is a branch of computer science which deals with system programming in order to automatically learn and improve with experience [59]. machine learning involves adaptive mechanisms that enable computers to learn from experience, learn by example and learn by analogy [75].

3.2.2 Technics of machine learning:

a) **Supervised learning:** supervised or active learning is learning with an external teacher or a supervisor who presents a training set to the network [75]. The standard approach to supervised learning is to split the set of examples into the training set and the test [59]. supervised learning is based on training. During its training phase, the system is fed with labeled data sets, which instruct the system what output is related to each specific input value. The trained model is then presented with test data: This is data that has been labeled, but the labels have not been revealed to the algorithm. The aim of the testing data is to measure how accurately the algorithm will perform on unlabeled data [39].

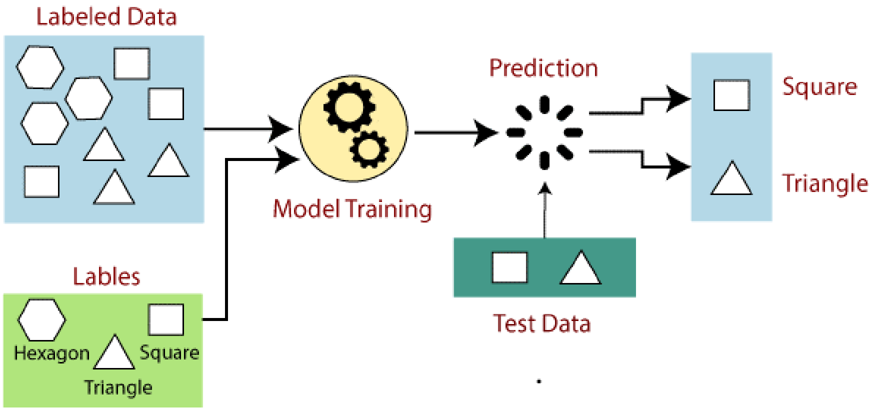

Figure 3.7: functioning of supervised learning [?].

b) **Unsupervised learning:** The function of the unsupervised learning is to find clusters, low-dimensional representations, interesting directions in data and interesting coordinates and correlations and find novel observations or database cleaning [59]. The unsupervised learning is often used in the more advanced applications of artificial intelligence. It involves giving unlabeled training data to an algorithm and asking it to pick up whatever associations it can on its own. Unsupervised learning is popular in applications of clustering (the act of uncovering groups within data) and association (predicting rules that describe data)[39].

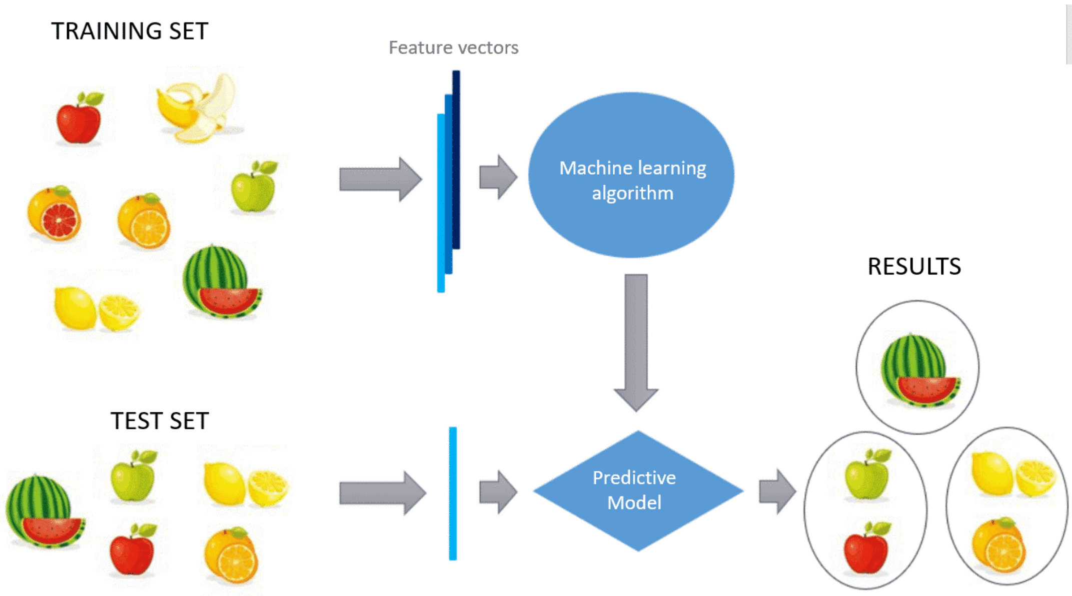

Figure 3.8: unsupervised learning [40]

c) **Semi-supervised learning:** Supervised learning is a set of techniques that allow you to infer models to extract knowledge from datasets where a priori is unknown. Unsupervised learning techniques can be applied without having the data labeled for training [40].In semi-supervised learning, algorithms train on small sets of labeled data and then, as in unsupervised learning, apply their learnings to unlabeled data. This approach is often used when there is a lack of quality data [39].

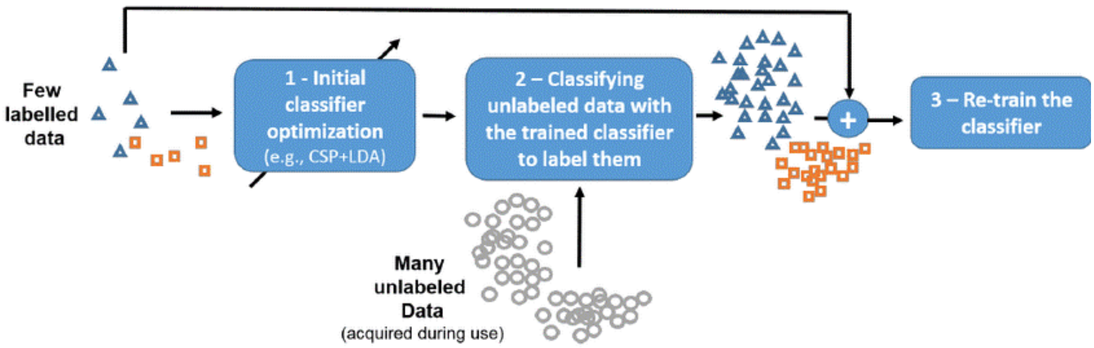

Figure 3.9: technic of semi-supervised learning [40]

d) **Reinforcement learning:** In some applications, the output of the system is a sequence of actions. In such a case, a single action is not important; what is important is the policy that is the sequence of correct actions to reach the goal. There is no such thing as the best action in any intermediate state; an action is good if it is part of a good policy. In such a case, the machine learning program should be able to assess the goodness of policies and learn from past good action sequences to be able to generate a policy. Such Learning reinforcement Methods are called reinforcement learning algorithms [32]. Reinforcement Learning designates the set of methods that allow an agent to learn to choose which action to take, and this in an autonomous way [47]. algorithms receive a set of instructions and guidelines and then make their own decisions about how to handle a task through a process of trial and error. Decisions are either rewarded or punished as a means of guiding the AI to the optimal solution to the problem [39].

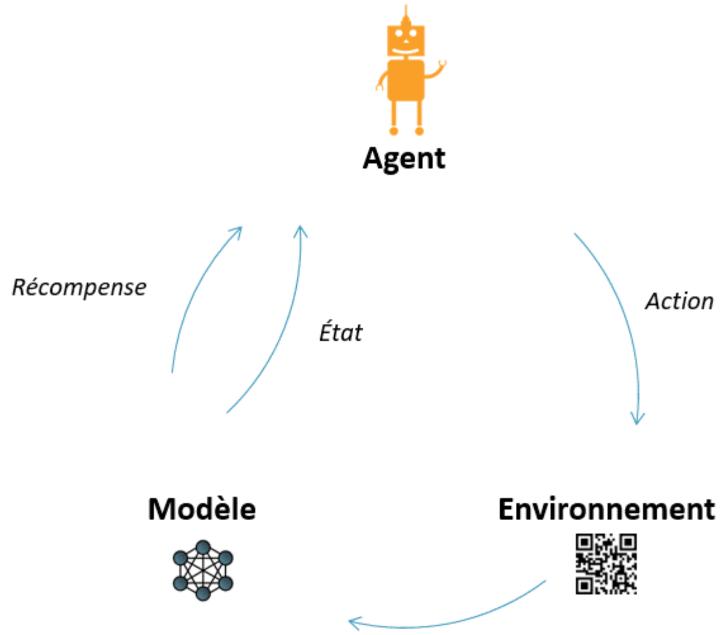

Figure 3.10: Reinforcement learning schema [47]

3.2.3 Area of use of machine learning:

a) **Virtual Personal Assistants:** When asked over the phone, the personal assistant can assist in discovering information. All you have to do is turn them on and tell them what you want them to do. They search for information, recollect your previous searches, or send a command to other resources (such as phone apps) to gather data. Machine learning is a key component of these personal assistants, as it collects and refines data based on your previous interactions with them. This information is then used to generate results that are personalized to your preferences [88].

Figure 3.11: virtual personal assistant [3]

b) **Predictions while Commuting:**

- Traffic Predictions: GPS navigation services have been used by all of us. While we are doing that, our current locations and speeds are kept on a central server for traffic management. This information is then used to create a traffic map. While this aids in traffic prevention and congestion analyses. In such cases, machine learning assists in estimating the places where congestion can be found based on daily encounters [88].

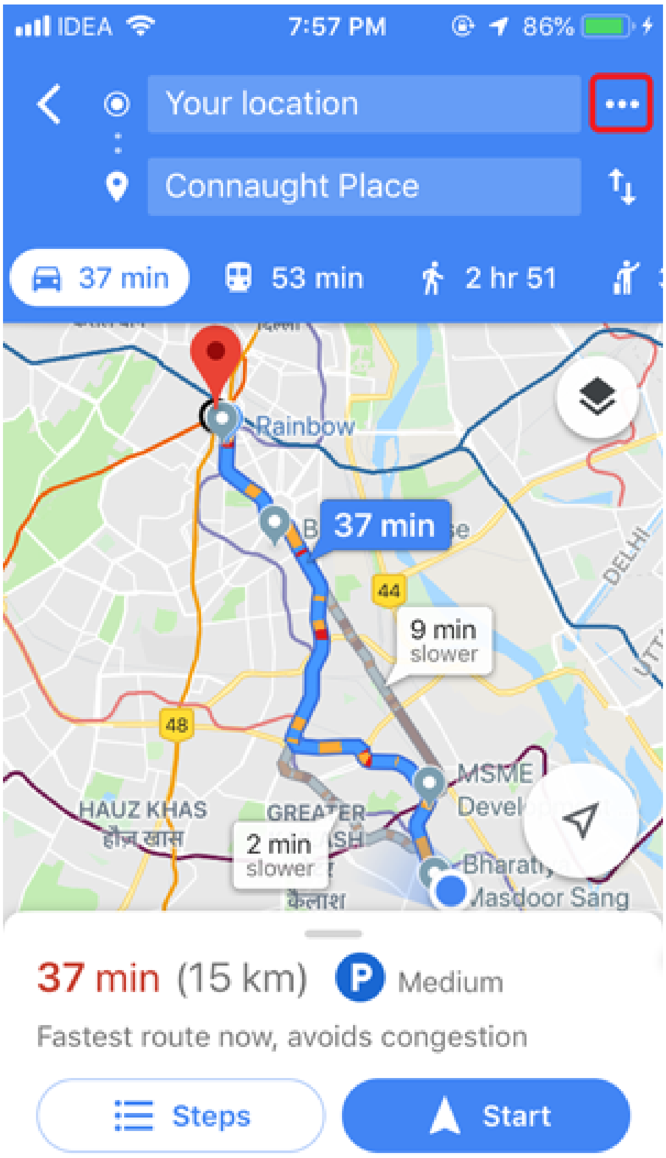

Figure 3.12: traffic prediction with google maps [62]

- Online Transportation Networks: The app predicts the cost of a cab ride when you book one. How do they reduce detours when sharing these services? Machine learning is the answer. In an interview, Jeff Schneider, the technical head at Uber ATC, explains that they employ machine learning to determine price surge hours by estimating rider demand. ML plays a significant part in the entire service cycle [88].

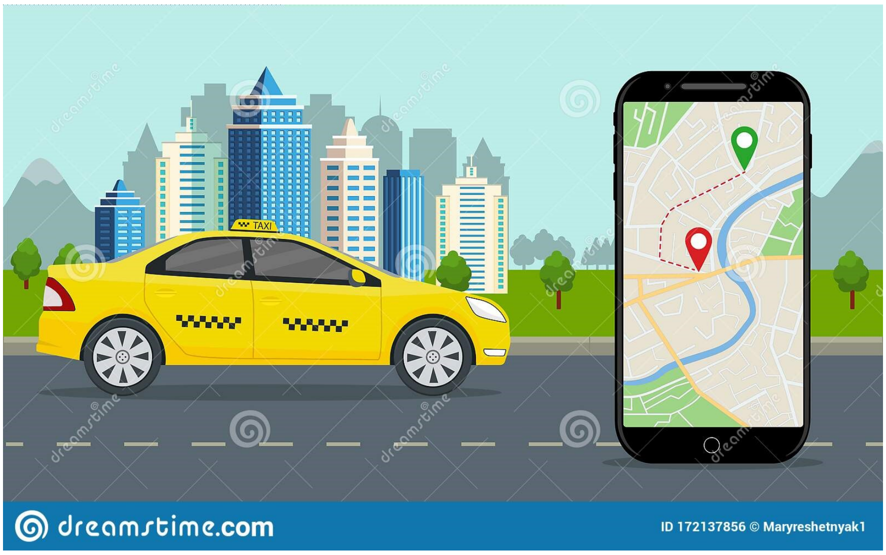

Figure 3.13: calling taxi cab with mobile app [78]

c) **Videos Surveillance :** Today's video surveillance systems use artificial intelligence (AI) to detect crimes before they occur. They monitor strange behavior such as people standing immobile for lengthy periods of time, stumbling, or napping on benches, among other things. As a result, the technology can inform human attendants, potentially preventing accidents. When such activities are reported and verified, they aid in the improvement of surveillance services. This occurs as a result of machine learning working in the background [88].

d) **Social Media Services :** From personalizing your news feed to better ads targeting, social media platforms are utilizing machine learning for their own and user benefits.

- People You May Know: Machine learning works on a simple concept: understanding with experiences. Facebook continuously notices the friends that you connect with, the profiles that you visit very often, your interests, workplace, or a group that you share with someone etc. On the basis of continuous learning, a list of Facebook users is suggested that you can become friends with [88].

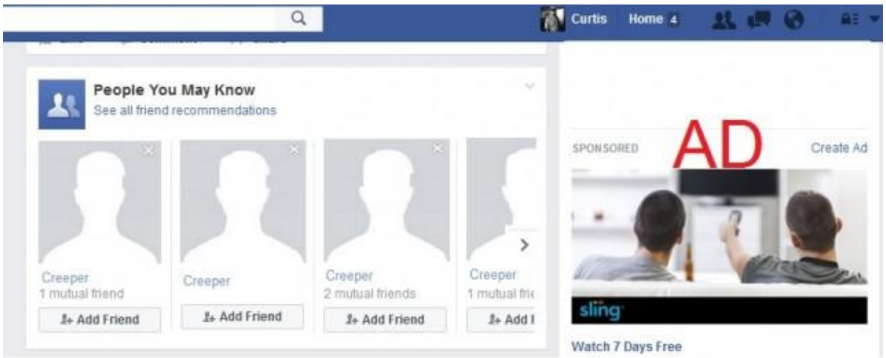

Figure 3.14: Facebook people you may know [86]

- Face Recognition: When you post a photo of yourself with a buddy, Facebook recognizes that person right away. Facebook analyzes the stances and projections in the photo, identifies distinguishing qualities, and matches them to people in your friend list. The entire backend procedure is sophisticated and takes into account the precision factor, yet the front end appears to be a simple application of machine learning [88].

e) **Email Spam and Malware Filtering:** Different researchers have made various approaches to prevent spam using machine learning techniques [46]. When rule-based spam filtering is used, it fails to keep track of spammers' latest techniques. Some of the spam filtering algorithms powered by ML include Multi-Layer Perceptron and C 4.5 Decision Tree Induction [88].

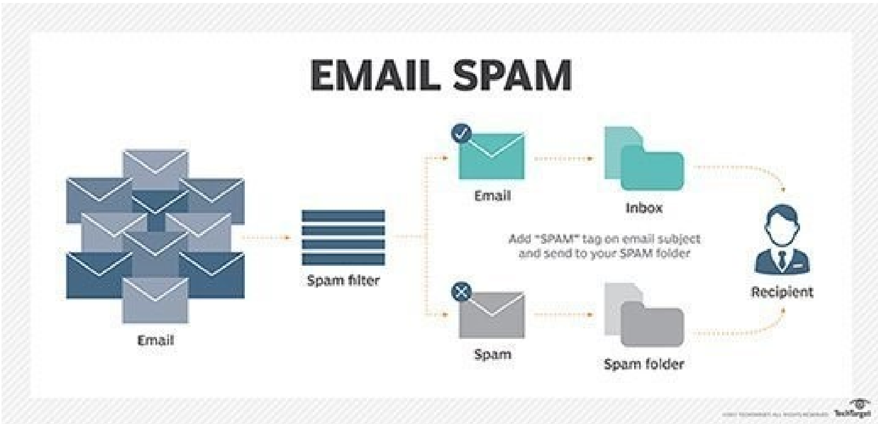

Figure 3.15: how spam filter works [95]

3.3 Deep Learning

3.3.1 Definition

Deep learning is a new emerging area which exploits artificial intelligence and machine learning to learn features directly from the data, using multiple nonlinear processing layers and trains a computer to perform human-like tasks [46] [7].
DL is making major advances in solving problems that have resisted the best attempts of the artificial intelligence community for many years [35].Deep learning method is based more specifically on the concept of an artificial neural network [45].
deep learning uses multi-layered artificial neural networks to deliver state-of-the-art accuracy in tasks such as object detection, speech recognition, language translation, and others [4].

3.3.2 Neural networks

A neural network is a series of algorithms that endeavors to recognize underlying relationships in a set of data through a process that mimics the way the human brain operates. In this sense, neural networks refer to systems of neurons, either organic or artificial in nature [43].

Neural networks just use big activity vectors, big weight matrices and scalar non-linearities to perform the type of fast intuitive, inference that underpins effortless commonsense reasoning [64].

3.3.2.1 Artificial neural networks:

Artificial Neural Networks are interconnected clusters of simple processing units that communicate with one another via a large number of weighted connections [46]. Artificial neural networks are based on early brain models of sensory processing. By replicating a network of model neurons in a computer, an artificial neural network can be created. We can make the network 'learn' to tackle a variety of issues by using algorithms that replicate the activities of real neurons. Artificial neural networks have been used to solve a variety of problems, including voice recognition, protein secondary structure prediction, cancer categorization, and gene prediction [64].

There are two conventional types of neural networks that are usually implied whenever ANN is used. They are the perceptron and the multi-layer perceptron [46].

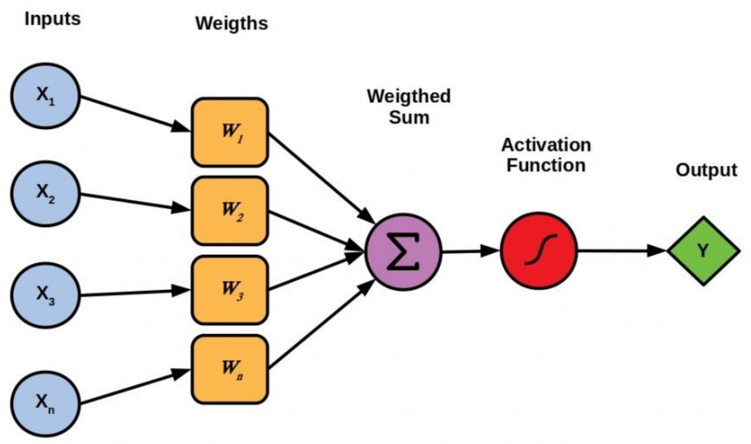

Figure 3.16: perceptron learning process [12].

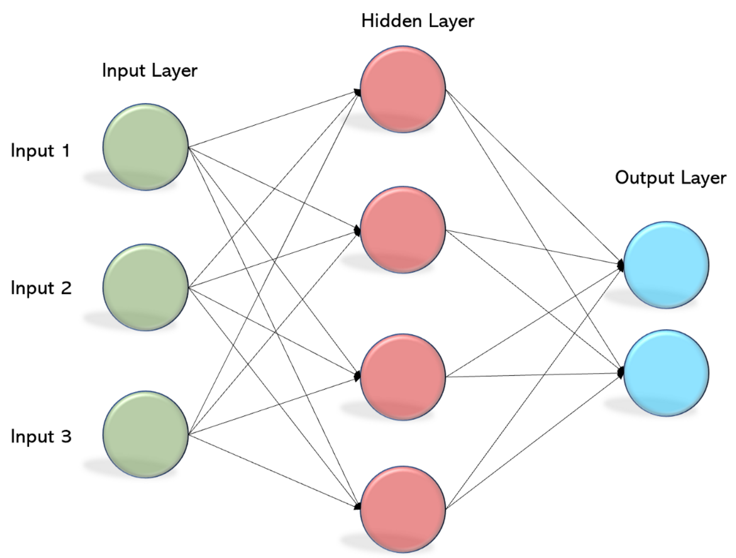

Figure 3.17: multi-layered perceptron process [69]

3.3.2.2 Deep neural networks:

A deep neural network is a neural network with a certain level of complexity, a neural network with more than two layers. Deep neural networks use sophisticated mathematical modeling to process data in complex ways.
Many experts define deep neural networks as networks that have an input layer, an output layer and at least one hidden layer in between. Each layer performs specific types of sorting and ordering in a process that some refer to as feature hierarchy. One of the key uses of these sophisticated neural networks is dealing with unlabeled or unstructured data. The phrase deep learning is also used to describe these deep neural networks, as deep learning represents a specific form of machine learning where technologies using aspects of artificial intelligence seek to classify and order information in ways that go beyond simple input/output protocols [11].
Deep neural networks exploit the property that many natural signals are compositional hierarchies, in which higher-level features are obtained by composing lower-level ones [35].

3.3.3 Application of deep learning :

a) **Image recognition:** Due to its advantages, such as strong feature extraction ability and high recognition accuracy, deep learning is commonly utilized in image recognition. Convolutional Neural Networks (CNN) have a noticeable effect on image recognition when compared to some common networks, such as RNN. Conventional neural networks are a type of deep learning model established on the basis of multi-layer Neural Networks for image classification and recognition [65].

b) **Speech recognition:** Deep learning algorithms have primarily been used to improve the skills of computers so that they can grasp what humans can accomplish, such as speech recognition. Speech, in particular, has been the primary mode of human communication for the past five decades, dating back to the development of artificial intelligence. As a result, it is only natural that speech was one of the first applications of deep learning, and a large number of research papers on the use of deep learning for speech-related applications, notably speech recognition, have been published to this day [74].

c) **Robotics:** Over the last decade, the application of deep learning to robotics has sparked a flurry of research into deep artificial neural networks, as well as a slew of new challenges and questions that the computer vision and machine learning communities have yet to tackle. Robotics is a branch of technology that deals with the design, manufacture, operation, and application of robots, as well as the use of computer systems for control, sensory feedback, and data processing. These technologies deal with automated devices that can replace humans in hazardous environments or production processes, or that look, behave, and think like humans. The application of deep learning in robotics is primarily motivated by the fact that it is more general than any other learning method. Deep networks have been shown to be capable of high-level thinking and abstraction [70].

d) **Cyber Security :** Cybersecurity is a set of technologies and processes designed to protect computers, networks, programs and data from attacks and unauthorized access, alteration, or destruction [93]. There are many applications of deep learning in the cybersecurity such as [48]: 1.Intrusion Detection and Prevention Systems (IDS/IPS).
2. Dealing with Malware.
3. Spam and Social Engineering Detection
4. Network Traffic Analysis.
5. User Behavior Analytics.
Deep Instinct uses deep learning to anticipate problems before they occur. Caspi explained how their system examines files and vectors before executing them, protecting clients in "zero-time." He noted that in today's danger scenario, speed is critical, and many solutions claim real time but real time is too late [48].

e) **Bioinformatics :** The transformation of massive amounts of data into useful knowledge has become increasingly vital in the era of 'big data,' and bioinformatics is no exception. There has been a significant accumulation of biomedical data, including omics, image, and signal data, and the potential for applications in biological and healthcare research has piqued the interest of both industry and academics. Deep learning is used in bioinformatics to locate junctions in DNA sequences, identify finger joints in X-ray pictures, and detect gaps in electroencephalogram (EEG) signals, among other things [68].

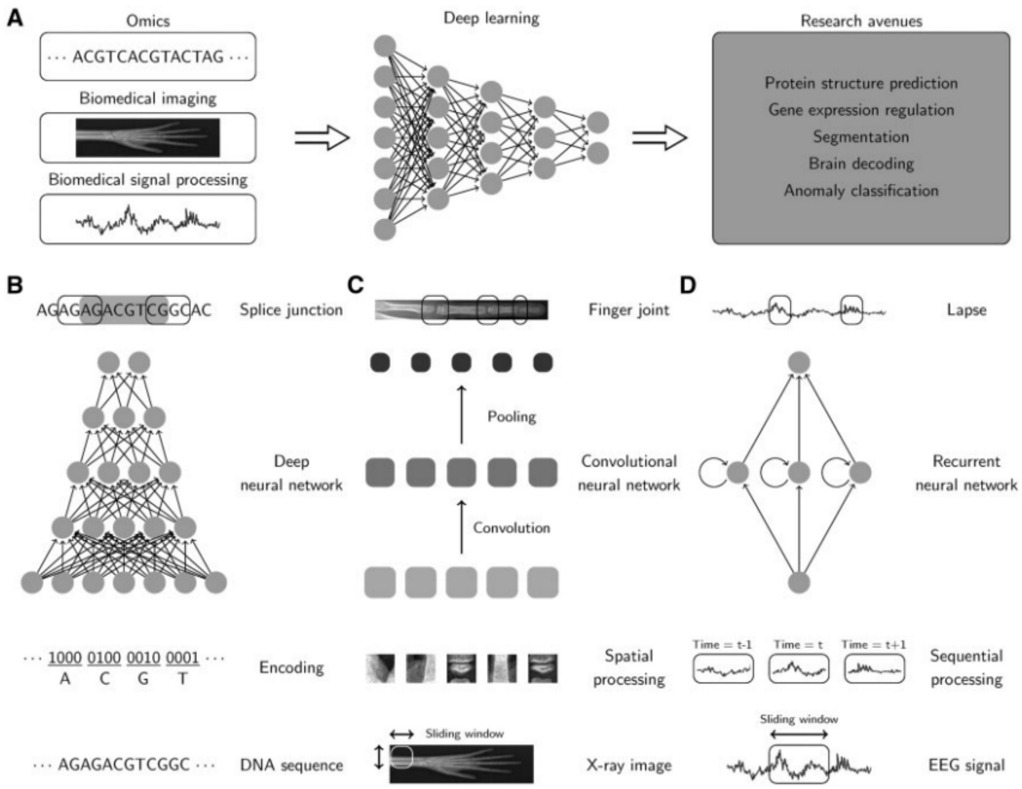

Figure 3.18: Application of deep learning in bioinformatics research. (A) Overview diagram with input data and research objectives. (B) A research example in the omics domain. Prediction of splice junctions in DNA sequence data with a deep neural network. (C) A research example in biomedical imaging. Finger joint detection from X-ray images with a convolutional neural network. (D) A research example in biomedical signal processing. Lapse detection from EEG signal with a recurrent neural network [68].

3.4 Conclusion

in this chapter we made an overview on Arifitial intelligence , machine learning, its technics, area of use and applications, then we talk about the deep learning and neural networks.

4

Our contribution

In this chapter we are going to describe information about our desertation and the algorihm we propose to solve the problem of checkbox detection in pdf files and images using deep learning and OCR .

4.1 Flowchart

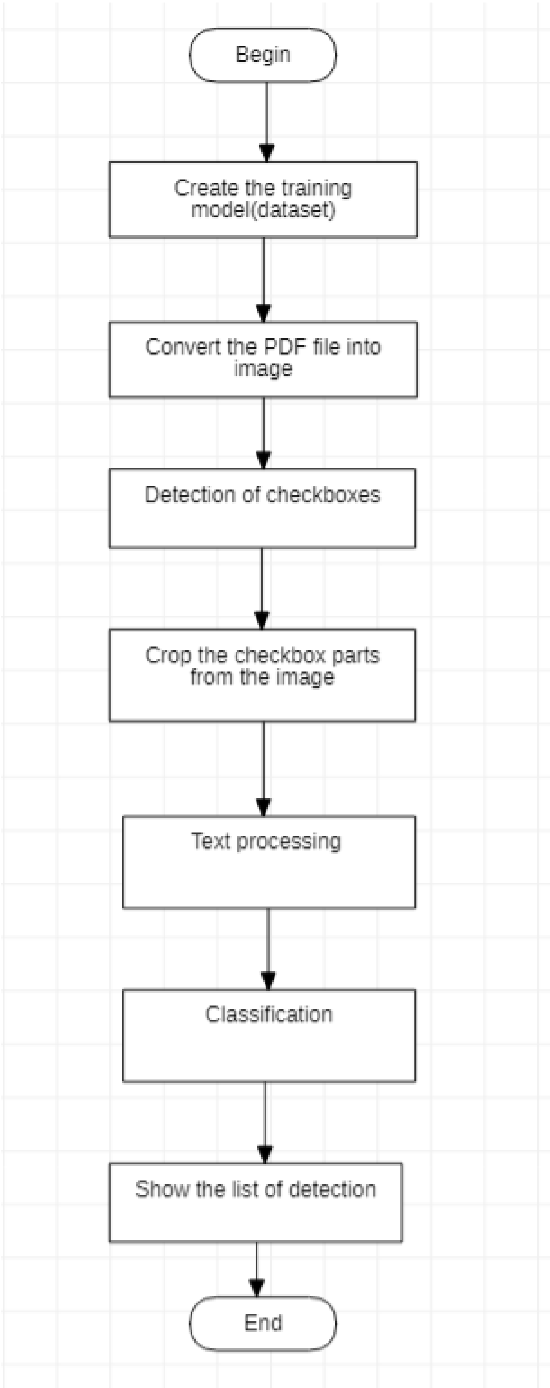

Figure 4.1: Flowchart

a) **Creating the training model:**

- Collecting the dataset: In any case, before we train a model, we need a data set, which is a collection of data [51].
  In Machine Learning projects, we need a training data set. It is the actual data set used to train the model for performing various actions [51].
  Our model is about classifying checkboxes to checked or unchecked, we prepared the dataset by collecting the images from many files and we find that the checked ones can be marked by cross, big dot, true sign or full color etc...

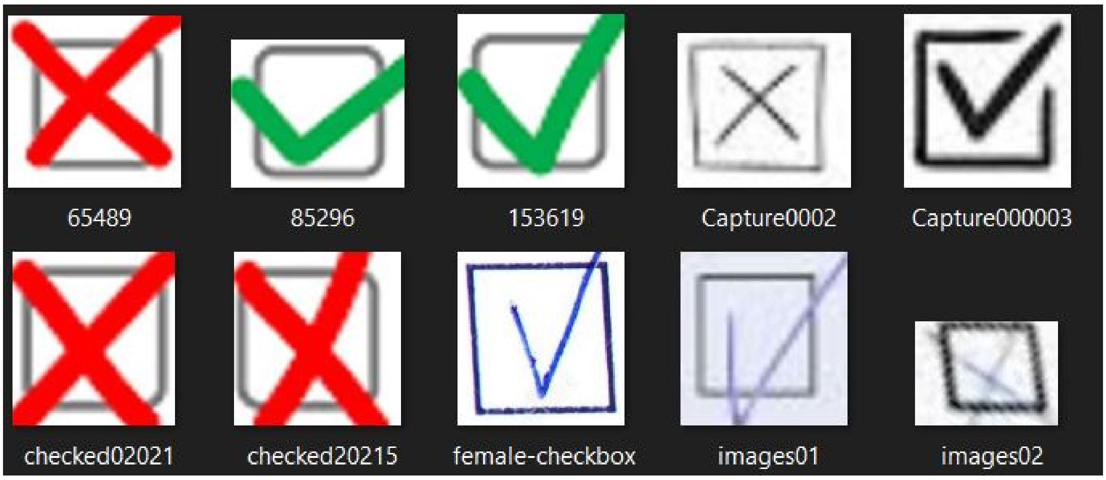

Figure 4.2: Checkboxes models

During an AI development, we always rely on data. From training, tuning, model selection to testing, we use three different data sets: the training set, the validation set, and the testing set. For your information, validation sets are used to select and tune the final ML model [51].
The training data set is the one used to train an algorithm to understand how to apply concepts such as neural networks, to learn and produce results. It includes both input data and the expected output [51]. The test data set is used to evaluate how well your algorithm was trained with the training data set.
In AI projects, we can't use the training data set in the testing stage because the algorithm will already know in advance the expected output which is not our goal [51].

In our training data set there are two folders: the cbimagstrn (the checkbox images trains)and chimagstst(checkbox images test) each one of them have two other folders inside checked witch have the checked checkbox images and unchecked witch have the unchecked checkbox images. The traine folder must have more images then the test folder.

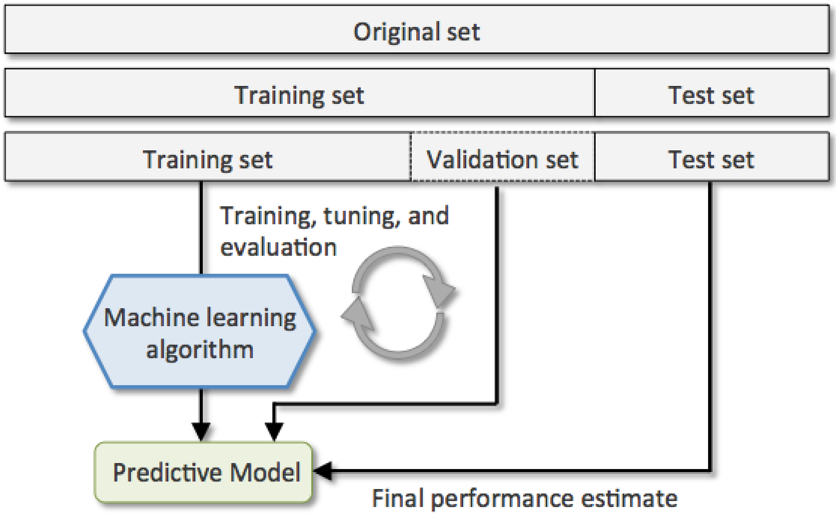

Figure 4.3: training model diagram [51].

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be the more and better data that we get, the better our model will perform [79].

- Training the model: Here we come to the model creation, the process has two parts; creation the conventional neural networks and the model fitting.
  - **1st part: conventional neural networks** This part it is all about preparing the model by creating the sequence and adding seven layers .

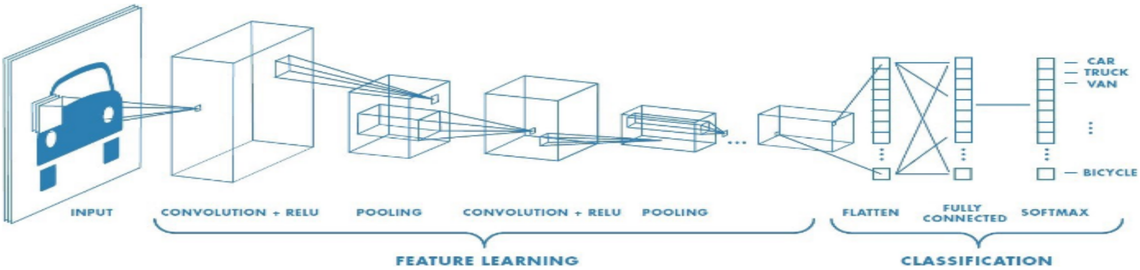

Figure 4.4: the model structure [81].

```
model = tf.keras.Sequential()
model.add(Conv2D(32, (3, 3), input_shape=(64, 64, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Conv2D(32, (3, 3), activation='relu'))
model.add(MaxPooling2D(pool_size=(2, 2)))
model.add(Flatten())
model.add(Dense(units=128, activation='relu'))
model.add(Dense(units=1, activation='sigmoid'))
```

Figure 4.5: code creation of layers.

– Then we compile the model that we define it for finalise and make it completely ready to use, with the optimizer **adam**, loss function **binaryCrossentropy** and metrics **accuracy** .

```
model.compile(optimizer='adam', loss='binary_crossentropy',
metrics=['accuracy'])
```

Figure 4.6: code compiling the model.

– **2nd part: fitting the model using our proper data** We execute the model on image data using **ImageDataGenerator** that rescales the image, applies shear in some range, zooms the image and does horizontal flipping with the image, this **ImageDataGenerator** includes all possible orientation of the image [28].
Generate training data using flowFromDirectory :Is the function that is used to prepare data from the **trainDataset directory** **Target size** specifies the target size of the image [28] .
Generate test data using flowFromDirectory: is used to prepare test datafrom **test datasetdirectory** for the model [28].

```
# Part 2
train_datagen = ImageDataGenerator(rescale=1.0/255, shear_range=0.2,
zoom_range=0.2, horizontal_flip= True)
test_datagen  = ImageDataGenerator(rescale=1.0/255)
training_set  = train_datagen.flow_from_directory('./cbimagestrn',
target_size=(64,64), batch_size=32, class_mode='binary')
test_set      = test_datagen.flow_from_directory('./cbimagestst',
target_size=(64,64), batch_size=32, class_mode='binary')
```

Figure 4.7: code03

We train the model by using the function **fitGenerator** that fit the data into our model and define the number of epochs (iterations).

```
model.fit(training_set, epochs=8, validation_data = test_set)
```

Figure 4.8: code04

b) **Convert the scanned PDF file into image:** This process is about converting the pages of PDF files which we are about to extract and analyse information into Jpeg images and store them in a folder for later work. In order to achieve this, we follow the following approach:

- <u>Import pdf2image module:</u> it is the library that allows us to do operations on pdf files (read, write, convert, transfer, etc...)

```
from pdf2image import convert_from_path
```

Figure 4.9: code05

- <u>import os:</u>The OS module in Python provides functions for interacting with the operating system(*os.path*)[29].

```
import os
```

Figure 4.10: code06

- <u>Create the function convert:</u>first it does a test if the output path exists, if no it will create it, and use the method **convertfrompath()** and convert all the pages of the file into image and store them in a variable and save the image in the folder.

```
def convert(file, outputdir):

    if not os.path.exists(outputdir):
        os.makedirs(outputdir)
    pages = convert_from_path(file, 500)
    counter = 0
    for page in pages:
        myfile = outputdir + 'output' + str(counter) + '.jpeg'
        counter = counter + 1
        page.save(myfile, 'JPEG')
```

Figure 4.11: code07

c) **Checkboxes detection:** The second phase after converting the PDF files into images is to detect the checkboxes positions. The algorithm of the detection of checkboxes is divided into four sections:

- **1st: Read the converted image:** In this phase we use Opencv library, or the PIL.Image, to load the image from her directory, and store it in a variable to do the next processes on it.
  *The code using **PIL.Image**:

```
from PIL import Image
img = Image.open('check.jpg')
```

Figure 4.12: code08

*The code using Opencv:

```
import cv2
image_array = cv2.imread('tester.JPG')
```

Figure 4.13: code09

The loaded image:

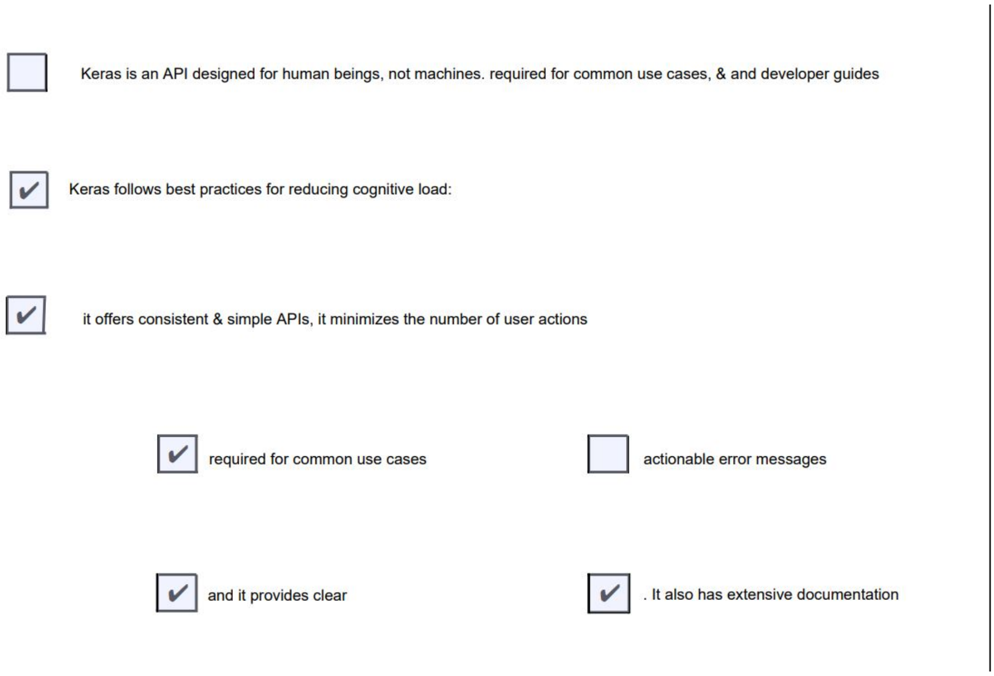

Figure 4.14: The original image

- **2nd: Binarization:** The type of the image that we are going to analyse is RGB, in another word it have much information that we do not need, so we convert it to grayscale, the 3 channels image are reduced to single-channel image and each pixel value is between 0 and 255, being black and is to make the pic in 2 pixels 0, 255 being white.
  Still, to reduce the complexity further, we do the Thresholding. It is nothing but making all the pixels either ON or OFF based on a threshold on the pixel values [77].

```
gray_scale_image = cv2.cvtColor(image_array, cv2.COLOR_BGR2GRAY)

_, img_bin = cv2.threshold(gray_scale_image, 0, 255, cv2.THRESH_BINARY |
cv2.THRESH_OTSU)
img_bin = 255 - img_bin
```

Figure 4.15: code10

And this is the image after the binarization:

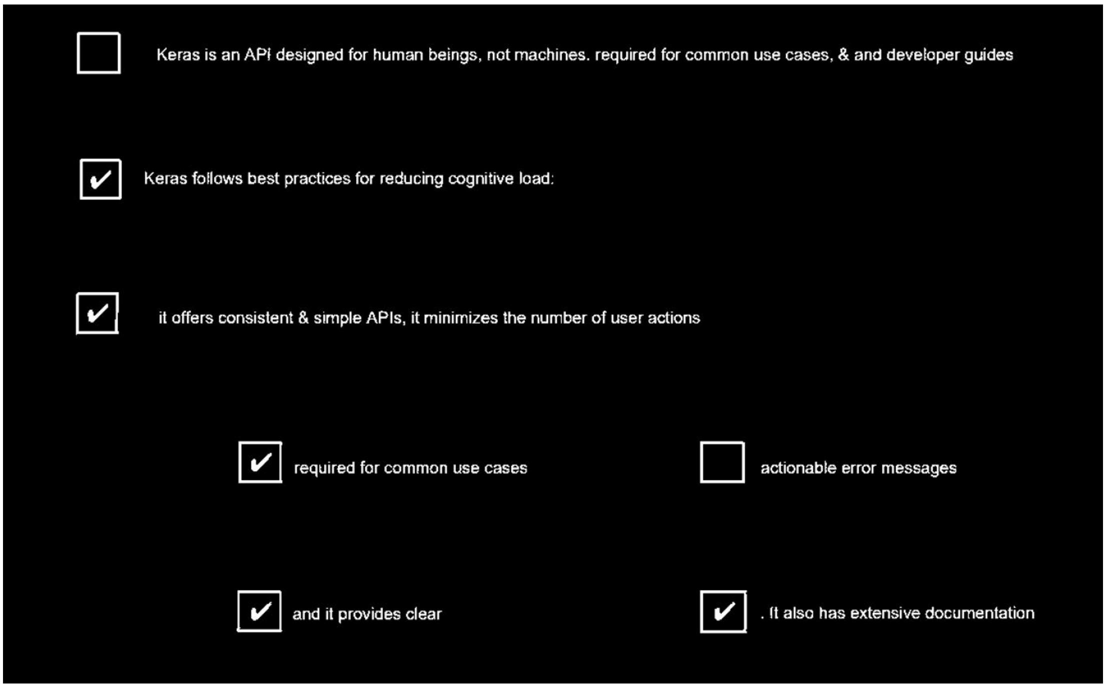

Figure 4.16: image after Binarization.

- **3rd: Applying morphological operations:** This is the most important part in the detection process, where the purpose here is to detect the horizontal and vertical kernels and merge them to have the boxes.
  When solving Image processing problems, Morphological Transformation is a term that you will come across a lot of times.
  We need to know these four terms; dilation, erosion, opening, closing.
  In simple terms, Dilation: Makes the image thicker -> Used to join broken parts.
  Erosion: Makes the image thinner -> Used to remove noise, separate nearby components.
  Opening: Erosion then Dilation -> Used in Noise removal.
  Closing: Dilation then Erosion -> filling small holes [77].
  *First of all we set the min size of the line of the boxes to not get the wrong positions, and it depends on our data.

```
# ----mophologie operation -------------
# -----set min width to detect horizontal line -----
line_min_width = 15
```

Figure 4.17: code11.

*Then we come to setting the kernel to detect the horizontal and vertical lines.

```
# -------kernal to detect horizontal & vertical lines-------
kernal_h = np.ones((1, line_min_width), np.uint8)
kernal_v = np.ones((line_min_width, 1), np.uint8)
```

Figure 4.18: code12.

*Then we apply the horizontal and vertical detection on our image.

```
# -------horizontal & vertical kernal on the image ------
img_bin_h = cv2.morphologyEx(img_bin, cv2.MORPH_OPEN, kernal_h)
img_bin_v = cv2.morphologyEx(img_bin, cv2.MORPH_OPEN, kernal_v)
```

Figure 4.19: code13.

*Now we combine them to get the boxes positions(x, y) and characteristics (height and width) and we get them after making call to the function from OpenCV; connectedComponentsWithStats.

```
# ------combinig the image----------
img_bin_final = img_bin_h | img_bin_v
# ---------contour filtering ----------
_, labels, stats, _ = cv2.connectedComponentsWithStats(~img_bin_final,
connectivity=8, ltype=cv2.CV_32S)
```

Figure 4.20: code14.

*After that, in a loop that manipulates all the boxes founded we draw bounds around all the boxes to make it visible and store the positions in a list to use them in the crop and classification parts.

```
ListPositions =[]
i = 0
for x, y, w, h, area in stats[2:]:
    i+=1
    hgh = h
    wdth = w
    cv2.rectangle(image_array, (x, y), (x + w, y + h), (0, 255, 0), 2)
    ListPositions.insert(i, (str(x) + "_" + str(y)))
```

Figure 4.21: code15.

This code will display the following image:

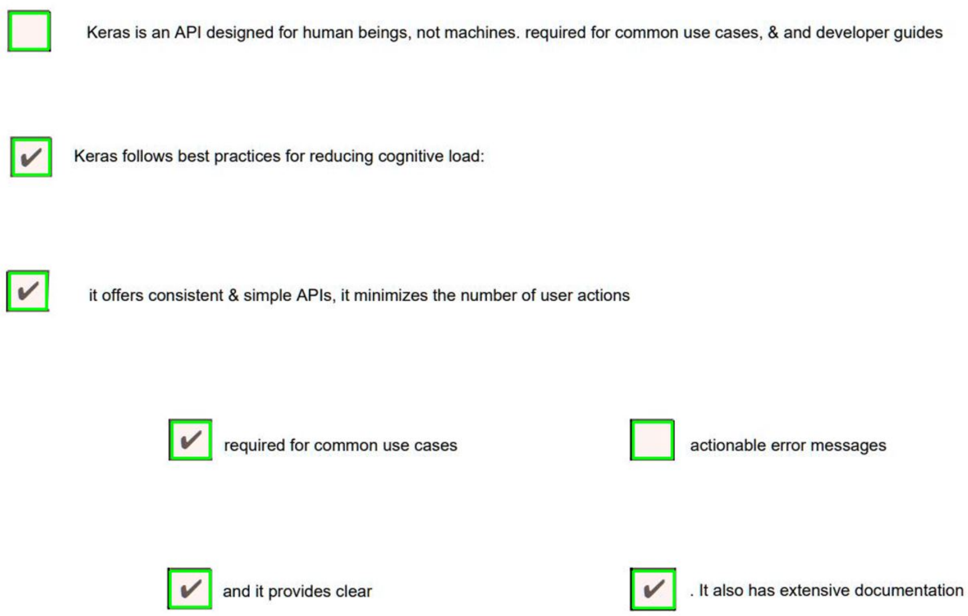

Figure 4.22: bounded checkboxes image

d) **Crop the checkboxes parts from the image:**

This part of the algorithm do the process of cut the regions in the image where the checkboxes are located and store them in a folder to use it in the next part. This section is based on this approach:

- loading the necessary package:

```
from PIL import Image
```

Figure 4.23: code16.

- Load the image in variable.

```
img = Image.open('tester.JPG')
```

Figure 4.24: code17.

- Determine the directory where the cropped images are going to be stored in.

```
outputdr =
"C://Users//hp//PycharmProjects//Cropping//dataset//single_prediction//"
```

Figure 4.25: code18.

- In a loop that manipulate every element in the list of coordinates :

  i. Store x and y values in variables using split method.

  ii. We deliberately add 4 pixels from all the sides in order to avoid clipping of parts of the checkboxes.

  iii. Finally, we crop the image and save it in the output directory after naming it by the position coordinates.

```
for element in liste:
    X = element.split("_", 1)
    Y = element.split("_", 2)
    x = X[0]
    y = Y[1]
    x1 = int(x) - 4
    x2 = int(x) + wdth + 4
    y1 = int(y) - 4
    y2 = int(y) + hgh + 4
    roi = img.crop((x1, y1, x2, y2))
    roi.save(outputdr+str(element)+".jpeg", 'JPEG')
```

Figure 4.26: code19.

Remark: the code of cropping can be included in the loop that displays the image with bounding checkboxes in the detection part.

e) **Text processing:** In this part we extract the text related to each checkbox to display it with the result of the classification.
The principle is to cut the part of the image that contains the text for each checkbox separately.

- First install and import the pytesseract package that allows us to extract the text from images.

```
Import pytesseract
pytesseract.pytesseract.tesseract_cmd = r"C:\Program Files\Tesseract-
OCR\tesseract.exe"
```

Figure 4.27: code20.

- We load the image in a variable and binarizate it and save the height and the width in variables.

```
image = cv2.imread('tester.jpg', 0)
thresh = 255 - cv2.threshold(image, 0, 255, cv2.THRESH_BINARY_INV +
cv2.THRESH_OTSU)[1]
hgh, wdth = image.shape
```

Figure 4.28: code21.

- Create list for x and y separately from 'ListePositions'.

```
x = []
y = []
for element in ListPositions:
    X = element.split("_", 1)
    x.insert(i, int(X[0]))
    y.insert(i, int(X[1]))
    i = i+1
```

Figure 4.29: code22.

- We sort the list of checkbox positions 'ListePositions' in ascending order by the Y.

```
for e in y:
    if( y[j] > y[j+1]):
        m = y[j]
        y[j] = y[j+1]
        y[j+1] = m
        n = x[j]
        x[j] = x[j + 1]
        x[j + 1] = n
        j = j+1
```

Figure 4.30: code23.

- The general situation is to find one checkbox in a line, so we crop the image from its position to the end of the line. But there is possibility to find two or three checkboxes in the same line.

- We crop the part of image from the position of the first checkbox to the position of the next one.

- Finaly, we extract the text in the image and save it to the according element in the 'ListePositions'.

```
if(y[i] == y[i+1]):
    r = y[i]
    roi = thresh[x[i]:x[i + 1], y[i]:y[i]]
    roi.save(path + "txt" + str(element) + ".jpeg", 'JPEG')
    roi2 = thresh[x[i + 1]: wdth, y[i]:y[i]]
    roi.save(path + "txt" + str(element) + ".jpeg", 'JPEG')
```

Figure 4.31: code24.

```
if((str(y[i]) + str(x[i])) == element):
    txt = pytesseract.image_to_string(roi)
    element.insert(l, element + "_" + str(txt))
    l = l + 1
```

Figure 4.32: code25.

f) **Classification:** For this part, we choose the deep learning algorithms to classify the checkboxes into two classes; checked and unchecked class.

In our model we import keras and tenserflow for deep learning which are libraries that contain a lot of functions that make the process of creating our model and training it easier.
The needed packages in this part of algorithm are the following:

```
import tensorflow as tf
import numpy as np
from keras.preprocessing.image import image
```

Figure 4.33: code26.

Then we do the test of the all the cropped images and the test must be image by image.

- Load the image in variable.
- Transform the image to an array of two dimensions [0, 1].
- Using the **expandDims** function from numpy to expend the shape of the array by inserting a new axis that will appear at the axis position in the expanded array shape [13].
- To predict the value of the loaded image we make a call to the function predict() from keras which is a function that enables to predict the labels the data values on the basis of the trained model [9], and store it in variable where if her value is [0, 0] the checkbox is checked otherwise is not.

```
for element in ListePositions:
    X = element.split('_', 1)
    Y = element.split('_', 2)
    T = element.split('_', 3)
    x = X[0]
    y = Y[1]
    t = Y[2]
    print(y)
    test_image = image.load_img('./single_prediction/' + str(x) + '_' +
str(y) + '.jpeg', target_size=(64, 64))
    test_image = image.img_to_array(test_image)
    test_image = np.expand_dims(test_image, axis=0)
    result = model.predict(test_image)

    if result[0][0] == 0:
        print('the checkbox in the position x=' + str(x) + '\t y=' + str(y)
+ ' \t with the text = ' + str(
            t) + ' \t the value= checked .....')
    else:
        print('the checkbox in the position x=' + str(x) + '\t y=' + str(y)
+ ' \t with the text = ' + str(
            t) + ' \t the value= unchecked ...')
```

Figure 4.34: code27.

4.2 Pseudo Code

Begin

a) **Creating the training model**

- Collecting the dataset.
- creating the conventionnal neural neworks.
- fitting the model on our propre data.

b) **Converting the scanned PDF into JPG images**

- convert the pdf file into images and store the in a specefic folder.

c) **Checkbox detection**

- load the converted images variables.
- binarizating the image.
- Apply the morphological operation on the image.

- Set the minimal size of the lines.

- detecte the horizental and vertical lines.

- combine the horizental and virtical lines.

- for each box
  *drow the bounds in the image.
  *save position in listePositions.

d) **Crop the checkboxes parts from the image**

- Determine the output directory.

- for each element in listePositions:
  *x=split (element,1).
  *y=split (element,2).
  *x'= x+width.
  *y'=y+height .
  *roi=crop(image, x,y,x',y').
  *store(roi,outputdir)

e) **Text processing**

- import the image in variable.

- separate the listePositions into X and Y ,then sort them in ascending order by y.

- for each element in listePositions:
  if there are 2 or more checkboxes in one line:
  *crop from the position of the first checkbox to the next one.
  *save the croped image.
  *extract the text from the image and save it in the related element.
  *else crop from the position of the checkbox to the end of the line.
  *save the croped image.
  *extract the text from the image and save it in the related element in ListePositions.

f) **Classification**

- for each image in outputdir:
  *testImg = image.
  *transform image into array.
  *expend the dimension of image.
  *result = predect(image,model).
  display result *if result =[0,0]
  print the image in position x,y with the text t is checked.
  else
  print the image in position x,y with the text t is unchecked.

end

4.3 Time complexity :

We anotate the time complexity of the training model by @.
The complexity in the other parts of the algorithm is like the following:
Convert the PDF files to images: O(n). n is the number of the pages of the PDF file.
Detecting the checkboxes positions: O(n).
Crop the checkboxes regions: O(n). n is the number of checkboxes in the image.
Text processing: O(n). n is the nomber of the checkboxes.
Classification of the checkboxes: (n*@). n is the number of the images of the checkboxes cropped in the previous part.
Displaying the results: O(n). n is the number of the images of the checkboxes cropped and classified.
So we can say that the total complexity is Max(@*n ,O(n))

4.4 Experimental and result:

We implement our algorithm using python programming language and we tested it in a computer with Core I7 processor and 8 GB of memory hard disk SSD 240 GB.
We execute the algorithm of detection and classification and evaluate its performance respecting the runtime and the accuracy.
We use 5 PDF files that contains varying numbers of checkboxes in different positions and different values to test the algorithm.
The files are presenting in the next figure:

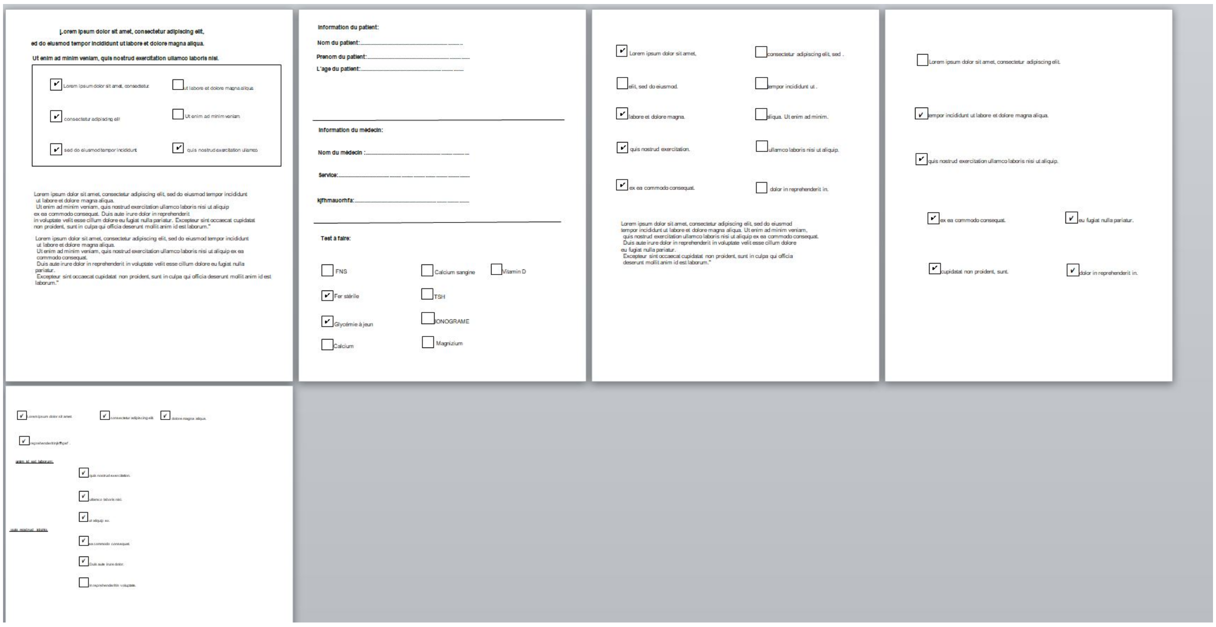

Figure 4.35: the test files.

The results of the execution of our algorithm are in the table and the diagrams bellow:

file	file01	file02	file03	file04	file05
Accuracy	100	100	100	71,40	40
runtime(ms)	12.02	14.37	15.63	11.85	15.72

Our algorithm gives us the high accuracy in the first three files with 100 pecent and in the fourth and the fifth files the accuracy descrease due to the number and the type of the checkboxes images in our dataset. And we can improve the accuracy by divercifying the dataset andthis by adding many different images of checkboxes from many sources.
The runtime is from 11 to 15 ms, so it doesnot take a lot of time and this is related to the number of checkboxes in the file or the image.

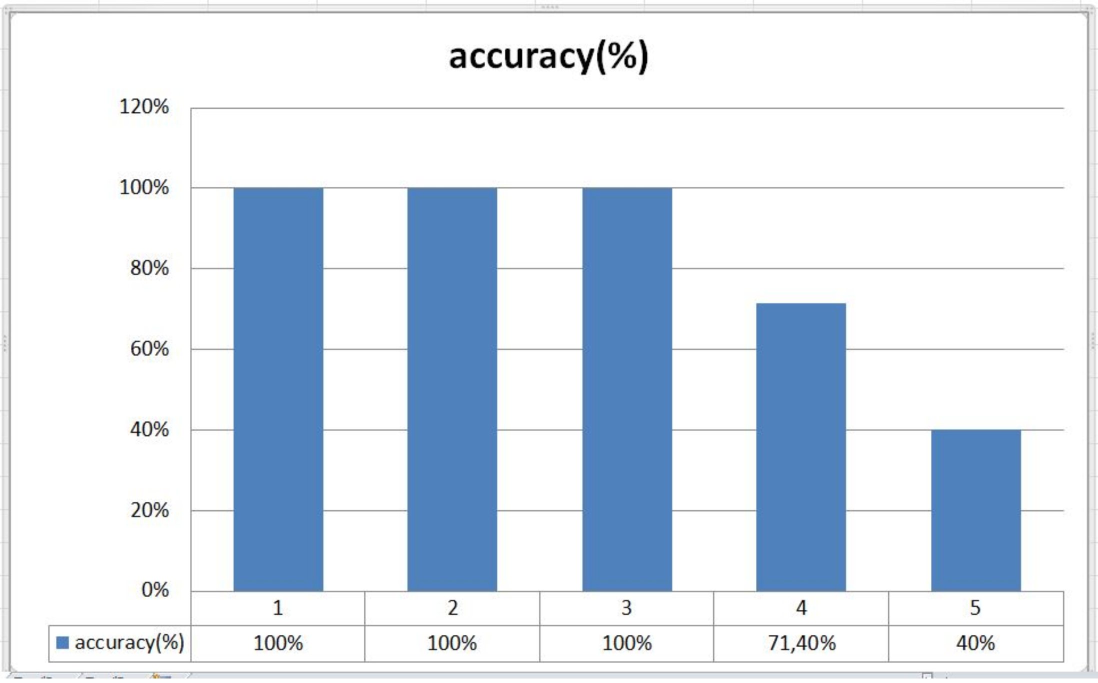

Figure 4.36: accuracy diagram.

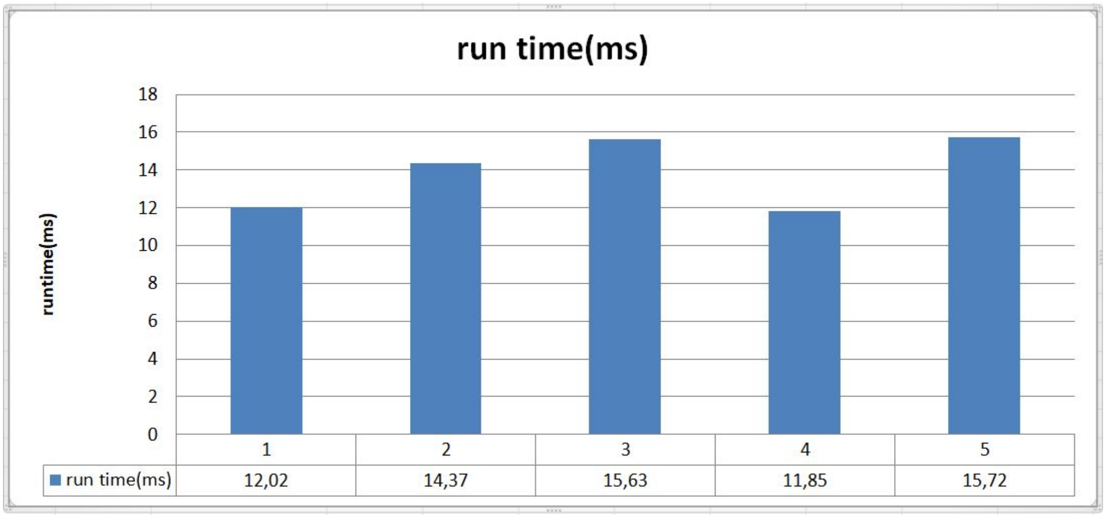

Figure 4.37: rumtime diagram.

4.5 Related work

Recently, several methods and techniques have been proposed for checkbox detection in the literature .
In order to automate checkbox detection for the digitization of surgical

flowsheet data, E. Murphy et al ,proposed image processing and machine learning techniques. A checkbox image is cropped based on its location with template matching and then processed through a trained convolutional neural network (CNN) to classify it as checked or unchecked. the template matching achieved an accuracy of 99.8 and 96.7 percent of the cropped images were correctly classified using the CNN model [73].
Z. Shengnan et al , introduced an automatic recognition method for checkbox, the authors uses the maximum tilt angle of table-run for fast skew detection and correction of form image, then adopts the connected domain labeling algorithm to rapidly locate the checkbox and extract the standard image of handwritten symbol.
Finally the structuhandwritten symbol recognitionre features of the above standard image are integrated to realize handwritten symbol recognition. The proposed method has shown its superiority in solving some problems, such as border rupture, adhesion of table lines with handwritten symbol. However ,because of the randomness of handwritten symbol, the situations of the stagger of table line with handwritten symbol are often occurred [84].
The thesis of john, M, Istle ,described a new algorithm for detecting checkboxes. First the proposed algorithm was run through a thinning algorithm and then an endpoint erosion algorithm to remove checkmarks that stretch beyond the borders. A contour following procedure with a prioritized search pattern is the main reason for the success to trace around a checkbox without accidentally tracing a checkmark [56].
Q, Tian et al. have developed an automated method of data verification with data from paper-based CRF and EMRs. the proposed method stared by analyzing the scanned images of paper-based CRFs with machine learning enhanced OCR, then, retrieving the related patient information from the EMRs using natural language processing (NLP) techniques, and finally , comparing the retrieved information in the previous two steps with the data in the registry, and synthesize the results accordingly. The proposed approach has been implemented in The Chinese Coronary Artery Disease Registry. For CRF data recognition, the accuracy of recognition for checkboxes marks and hand-writing are 0.93 and 0.74, respectively [92].

4.6 Conclusion

In this chapter we present our algorithm and their accuracy and runtime with review on some used functions.

5

General conclusion

Information extraction has been one of the most challenging technologies in the past few years. The main goal is to extract relevant information from unstructured data sources like scanned files, receipts, bills, etc., into structured data, using deep learning (DL) techniques. Some data forms, such as MCQs, bank bills, and medical insurance, usually contain the checkboxes fields that needs to fill in, and whether checkboxes are selected or not is the important statistical data, because of that, automatic recognition for checkbox is the key issue in the automatic processing of data forms.

The problem of automatic detecting checkboxes from PDF documents and images is a very common problem in pattern recognition field due to different reasons such as the many kind of handwriting marks, the difference in the checkboxes sizes and others. To solve this problems, there is many publication has been proposed depending on deep learning technologies and pattern recognition. Our work presented a new algorithm for checkboxes detecting by using Optical Character Recognition (OCR) algorithm with the Convolutional Neural Network (CNN) model.
OCR algorithm is very useful and efficient in Our process to make it able to detect the checkboxes and their position, and the Conventional Neural

Network (CNN) had a significant impact on classifying the checkboxes to either checked or unchecked. As future work, we have other ideas that can make extracting information easy and more efficiency such as UI web application that take the PDF files or scanned images and return the checkboxes and their value with extract the exact text related to the checkboxes in less time possible, and this will be achieved by reducing the complexity.

We think about solving the problems of the low quality images , because most of today's systems are computerized and to archive the old documents and files we have to reduce the quality most possible to be readable to the computer, we have to improve the data set to have an algorithm that gives the result to most scanned files and different types of checkboxes.

Bibliography

[1] Introduction to image processing. Technical report, university of tarto.

[2] Artificial intelligence - expert systems. Technical report, webopedia encyclopedia, (accessed April 02, 2022).

[3] personal virtual assistant. Technical report, PNG, (accessed April 02, 2022).

[4] Deep learning. Technical report, nvidia.developer, (accessed April 03, 2022).

[5] definition speech recognition. Technical report, Oxford Language, (accessed April 05,2022.

[6] Compression. Technical report, Adobe, (accessed June 07, 2022).

[7] Deep learning,what it is why it matters. Technical report, sas, (accessed Martch 07, 2022).

[8] Five ai technologies that you need to know. Technical report, sas, (accessed Martch 07, 2022).

[9] Python predict() function – all you need to know! Technical report, AskPython, (accessed May 20,2022).

[10] Intelligence artificielle: comprendre les différents types. Technical report, Blog ARC Optimizer, juillet 22, 2019(accessed Martch 08,2022).

[11] Deep neural network. Technical report, techopedia, last update April 13, 2018(accessed april 04 2022).

[12] Perceptrons – these artificial neurons are the fundamentals of neural

networks. Technical report, starship-knowledge, last update october 28,2018 (accessed april 03, 2022).

[13] numpy.expand$_{d}ims()function.Technicalreport, w3resource, LastupdateonMay282022(acc$

[14] optical mark recognition. Technical report, wikipidia, last update02/04/22 23:10 (accessed April 02, 2022).

[15] pattern recognition. Technical report, s-cool, last update02/04/22 23:10 (accessed April 02, 2022).

[16] Survey of post-ocr processing approaches. Technical report, acm dl, last update02/04/22 23:10 (accessed April 02, 2022).

[17] what is ocr. Technical report, docAcquire, last update02/04/22 23:10 (accessed April 02, 2022).

[18] machine vision. Technical report, stemmer, last update02/04/22 23:20 (accessed April 02, 2022).

[19] What is optical character recognition (ocr) technology? Technical report, terminology, last update03/04/22 20.15 (accessed April 03, 2022).

[20] 9 common machine vision applications in manufacturing. Technical report, Qualitas technologies, last update06/04/22 18:10 (accessed April 06, 2022).

[21] Computer aided diagnosis - medical image analysis techniques. Technical report, intechOpen, last update06/04/22 18:10 (accessed April 06, 2022).

[22] Ocr-image reader. Technical report, Operaaddons, last update06/04/22 20:30 (accessed April 06, 2022).

[23] text to speech in napalisoftware developed for visually challenged. Technical report, techlekh, last update06/04/22 20:30 (accessed April 06, 2022).

[24] Examples of the three types of output from the automatic mapping system. Technical report, researchgate, last update06/04/22 21:43 (accessed April 06, 2022).

[25] Ocr a document, form, or invoice with tesseract, opencv, and python. Technical report, laptrinhx, last update06/04/22 21:50 (accessed April 06, 2022).

[26] Handwriting recognition. Technical report, wikipedia, last update06/04/22 22:50 (accessed April 06, 2022).

[27] Image smoothing. Technical report, evergreen technologies, last update06/04/22 22:50 (accessed April 06, 2022).

[28] Python | image classification using keras. Technical report, Geeks For Geeks, Last Updated : 13 May,2022 (accessed May 18,2022).

[29] Os module in python with examples. Technical report, Geeks For Geeks, Last Updated : 25 Jan, 2022(accessed May 17,2022).

[30] Narrow artificial intelligence (narrow ai). Technical report, techopedia, Last updated: April 4, 2022 (accessed Martch 07, 2022).

[31] Dr. Jahid Ali and Rohit Verma. A-survey of feature extraction andclassification techniques in ocr systems. 2012.

[32] Ethem Alpaydın. *IntroductiontoMachineLearningSecondEdition*. Massachusetts Institute of Technology, 2010.

[33] Nora Ambroz. Natural language processing: A short introduction to get you started. Technical report, aliz, May 14, 2019(accessed April 02, 2022).

[34] N. VENKATA RAOand DR. A.S.C.S.SASTRY, A.S.N.CHAKRAVARTHY, and KALYANCHAKRAVARTHI P. Optical character recognition technique algorithms. 2016.

[35] Yoshua Bengio, Yann LeCun, and Geoffrey E. Hinton. Deep learning for ai. *Communications of the ACM*, 64:58 – 65, 2021.

[36] LOUNNAS BILAL. D i s c o v e ry a n d e x t r a c t i o n o fpat t e r n s i n b i o l o g i c a l s e q u e n c e s. Master's thesis, 2016.

[37] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. 2006.

[38] Jacob Bortnik and Enrico Camporeale. Ten ways to apply machine learning in the earth and space sciences. In *AGU Fall Meeting 2021*. AGU, 2021.

[39] Ed bruns. In-depth guide to machine learning in the enterprise. Technical report, techtarget, Published: 05 Apr 2021 (accessed Martch 07, 2022).

[40] Diego Calvo. Unsupervised learning. Technical report, diegocalvo, Mar 24, 2019(accessed April 02, 2022).

[41] Pranob K Charles1 and V.Harish znd M.Swathi znd CH.Deepthi. A review on the various techniques used for optical character recognition. 2012.

[42] Arindam Chaudhuri, Krupa Mandaviya, Pratixa Badelia, and Soumya K. Ghosh. *Optical Character Recognition Systems for Different Languages with Soft Computing*. 2017.

[43] James Chen. Neural network. Technical report, investopedia, Updated December 08, 2021(accessed April 14, 2022).

[44] Gobinda G. Chowdhury. Natural language processing. *Computer and Information SciencesUniversity of Strathclyde*, 2003.

[45] Antoine Crochet-Damais. Deep learning : définition et principes de l'apprentissage profond. Technical report, journaldunet, last update23/02/22 18:10 (accessed April 07, 2022).

[46] Emmanuel Gbenga Dada, Joseph Stephen Bassi, Haruna Chiroma, Shafi'i Muhammad Abdulhamid, Adebayo Olusola Adetunmbi, and Opeyemi Emmanuel Ajibuwa. Machine learning for email spam filtering: review, approaches and open research problems. *Heliyon*, 5(6):e01802, 2019.

[47] Lina F. Reinforcement learning : Définition et domaines d'application. Technical report, datascientest, last update 31 juillet 2020(accessed Martch 31, 2022).

[48] Elizabeth Fichtner. 5 amazing applications of deep learning in cybersecurity. Technical report, datto, last update February 21, 2022 (accessed April 05,2022).

[49] frank Y shih. *image processing and pattern recognition fundamentals and techniques*. 2010.

[50] André Le Garff. *le dictionnaire de l'informatique*. 1975.

[51] Alexandre Gonfalonieri. How to build a data set for your machine learning project. Technical report, Towards Data Science, Feb 14, 2019 (accessed May 16,2022).

[52] Pam Greenberg. Automated license plate readers. Technical report, NCSL, last update02/04/22 23:10 (accessed April 02, 2022).

[53] Behic Guven. Speech recognition in python- the complete beginner's guide. Technical report, Towards Data Science, Aug 15,2020 (accessed A$^{r}il02, 2022$).

[54] Pavel Hamet and Johanne Tremblay. Artificial intelligence in medicine. *Metabolism*, 69:S36–S40, 2017. Insights Into the Future of Medicine: Technologies, Concepts, and Integration.

[55] Georges Hebrail and Loïc Cantat. Le machine learning, un domaine de

l'intelligence artificielle en forte évolution. Technical report, IBlog, fabruary 12,2019 (accessed April 02, 2022).

[56] John Michael Istle. *Optical character recognition for checkbox detection*. University of Nevada, Las Vegas, 2004.

[57] John R. Jensen. *Introductry to digital image processing and Remote Sensing Perspective*. 2015.

[58] Bernd Jähne. *digital image processing*. 2005.

[59] Daniel Johnson. Top 50 machine learning interview questions answers. *Guru99*, 2022.

[60] Walter A. Kosters Peter van der Putten Joost N. Kok, Egbert J. W. Boers and Mannes Poel. Artificial intelligence: Definition, trends, techniques, and cases.

[61] R. Somasundaram K. Sumithra, S. Buvana. A survey on various types of image processingtechnique. 2015.

[62] Kaushal. How to predict traffic on google maps for android. Technical report, webopedia encyclopedia, January 30, 2019 (accessed April 02, 2022).

[63] Pelin Kocyigit. *Agent Based Optical Character Recognition*. 2012.

[64] Anders Krogh. What are artificial neural networks? *Nature biotechnology*, 26(2):195–197, 2008.

[65] Lingyun Li. Application of deep learning in image recognition. In *Journal of Physics: Conference Series*, volume 1693, page 012128. IOP Publishing, 2020.

[66] Jie Liu, Jigui Sun, and Shengsheng Wang. Pattern recognition: An overview. *IJCSNS International Journal of Computer Science and Network Security*, 6(6):57–61, 2006.

[67] Ben Lutkevich. speech recognition definition. Technical report, techtarget, (accessed April 03, 2022).

[68] Seonwoo Min, Byunghan Lee, and Sungroh Yoon. Deep learning in bioinformatics. *Briefings in bioinformatics*, 18(5):851–869, 2017.

[69] Awhan Mohanty. Multi layer perceptron (mlp) models on real world banking data. Technical report, becominghuman, last update May 15, 2019 (accessed April 03, 2015).

[70] Radouan Ait Mouha et al. Deep learning for robotics. *Journal of Data Analysis and Information Processing*, 9(02):63, 2021.

[71] Nadira Muda, Nik Kamariah Nik Ismail, Siti Azami Abu Bakar, and Jasni Mohamad Zain. Optical character recognition by using template matching (alphabet). In *National Conference on Software Engineering & Computer Systems 2007 (NACES 2007)*, 2007.

[72] Mario Munich and Pietro Perona. Visual identification by signature tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25:200– 217, 03 2003.

[73] Emily Murphy, Swathi Samuel, Joseph Cho, William Adorno, Marcel Durieux, Donald Brown, and Christian Ndaribitse. Checkbox detection on rwandan perioperative flowsheets using convolutional neural network. In *2021 Systems and Information Engineering Design Symposium (SIEDS)*, pages 1–6. IEEE, 2021.

[74] Ali Bou Nassif, Ismail Shahin, Imtinan Attili, Mohammad Azzeh, and Khaled Shaalan. Speech recognition using deep neural networks: A systematic review. *IEEE access*, 7:19143–19165, 2019.

[75] MICHAEL NEGNEVITSKY. *ArtificialIntelligenceA Guide to Intelligent Systems*. Pearson Education Limited 2002, 2005.

[76] Esuru Rita Okoroafor, Connor M. Smith, Karen Ifeoma Ochie, Chinedu Joseph Nwosu, Halldora Gudmundsdottir, and Mohammad (Jabs) Aljubran. Machine learning in subsurface geothermal energy: Two decades in review. *Geothermics*, 102:102401, 2022.

[77] Sreekiran A R. Checkbox/table cell detection using opencv-python. Technical report, Towards Data Science, (accessed Mar 16,2022).

[78] Mary reshetnyak. Taxi cab with mobile app. taxi service with online map on phone, city background. smartphone application with location ride on. Technical report, dreamstime, (accessed April 02, 2022).

[79] Victor Roman. How to develop a machine learning model from scratch. Technical report, Towards Data Science, Dec 23, 2018 (accessed May 16,2022).

[80] Adrian Rosebrock. Opencv: Automatic license/number plate recognition (anpr) with python. Technical report, pyimagesearch, last update06/04/22 21:43 (accessed April 06, 2022).

[81] Sumit Saha. A comprehensive guide to convolutional neural networks — the

eli5 way. Technical report, Towards Data Science, Dec 15, 2018 (accessed May 16,2022).

[82] D.J.Ravi2 Sanjana C.Shekar1. Image enhancement and compression using edge detection technique. 2017.

[83] sergios theodoridis and konstantinos koutroumbas. *pattern recognition*. 2009.

[84] Zhang Shengnan, Yuan Shanlei, and Niu Lianqiang. Automatic recognition method for checkbox in data form image. In *2014 Sixth International Conference on Measuring Technology and Mechatronics Automation*, pages 159–162. IEEE, 2014.

[85] Shu-Hsien Liao. Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert Systems with Applications*, 28(1):93–103, 2005.

[86] Curtis Silver. How facebook's 'people you may know' section just got creepier. Technical report, forbes, last update Jun 28, 2016,12:10pm (accessed April 03, 2022).

[87] Himanshu Singh. *Practical MachineLearning andImage ProcessingFor Facial Recognition, Object Detection,and Pattern Recognition Using Python*. 2019.

[88] Daffodil Software. 9 applications of machine learning from day-to-day life by daffodil software. Technical report, medium, jul 18,2017(accessed April 03, 2022).

[89] Peter Norvig Stuart Russell, editor. *Artificial IntelligenceA Modern ApproachThird Edition*. Pearson Education, 2010.

[90] D. Sundararajan. *Digital Image Processing A Signal Processingand Algorithmic Approach*. 2017.

[91] Alban T. Computer vision : tout savoir sur cette application du machine learning ! Technical report, datascientest, September 01,2020(accessed April 02, 2022).

[92] Qi Tian, Mengzhou Liu, Lingtong Min, Jiye An, Xudong Lu, and Huilong Duan. An automated data verification approach for improving data quality in a clinical registry. *Computer Methods and Programs in Biomedicine*, 181:104840, 2019.

[93] Yang Xin, Lingshuang Kong, Zhi Liu, Yuling Chen, Yanmiao Li, Hongliang Zhu, Mingcheng Gao, Haixia Hou, and Chunhua Wang. Machine learning and deep learning methods for cybersecurity. *Ieee access*, 6:35365–35381, 2018.

[94] Richard SuchenwirthJun Guolrmfried HartmannGeorg

HinchaManfred KrauseZheng Zhang. *optical recognition of chinese characteres*. 1989.

[95] Andrew Zola. spam filter. Technical report, techtarget, (accessed April 03, 2022).

ملخص:

يُعد الاستخراج التلقائي للمعلومات من الصور الممسوحة ضوئيًا مفيدًا جدًا للعديد من المجالات مثل الطب وعلوم الكمبيوتر والتي يمكن أن تكون أوراق الامتحان وبطاقات مرضى وما إلى ذلك ... في هذه المذكرة، نقترح خوارزمية لاكتشاف مواقع مربعات الاختيار وقيمها (محدد ، غير محدد) باستخدام التعلم العميق مع تقنيات أخرى مثل التعرف الضوئي على الحروف. أولا نقوم بتحويل ملف PDF إلى صور تمثل كل واحدة منها صفحة من الملف ثم نقوم بتحميل الصورة في الخوارزمية الخاصة بنا ونقوم بتحديد مواقع مربعات الاختيار باستخدام التعرف الضوئي على الحروف بعد ذلك نقوم بتقطيع مواقع مربعات الاختيار الى صور أصغر لاستخدامها في جزء التصنيف ، حيث نستخدم تقنيات التعلم العميق لتصنيف هذه الصور التي تم اقتصاصها إلى الفئات المناسبة ، ثم نقوم بتحميل الصورة في الخوارزمية الخاصة بنا

الكلمات المفتاحية: التعلم العميق ، الذكاء الاصطناعي ، الشبكة العصبية الإصطناعية التقليدية ، التعرف البصري على الأحرف ،كشف مربعات الاختيار ، نموذج التدريب

Abstract :

Automatic information extraction from scanned images is of great help for many fields such as medicine, computer science, which can be exam sheets, disease cards, etc... In this dissertation, we propose an algorithm to detect the positions of checkboxes and their values (checked, unchecked) using deep learning with other techniques such as OCR. First, we convert the pdf file into images representing each page of the pdf file, then load the image into our algorithm and detect the regions of the checkboxes using OCR. After that, we crop these regions into smaller images for use in the classification part, where we use deep learning techniques to classify these cropped images into the appropriate classes.

Key words: deep learning, Artificial intelligence, Conventional neural network, optical character recognition, checkbox detection, training model,

Résumé :

L’extraction automatique d’informations à partir d’images scannées est d’une grande aide pour de nombreux domaines tels que la médecine, l’informatique, qui peuvent être des feuilles d’examen, des cartes de maladie, etc... Dans cette thèse, nous proposons un algorithme pour détecter les positions des cases à cocher et leur valeur (coché, non coché) en utilisant l’apprentissage profond avec d’autres techniques telles que l’OCR. Dans un premier temps, nous convertissons le fichier PDF en images représentant chaque page du fichier PDF, puis nous chargeons l’image dans notre algorithme et détectons les régions des cases à cocher en utilisant l’OCR. Après cela, nous recadrons ces régions dans des images plus petites pour les utiliser dans la partie classification. Où nous utilisons des techniques d’apprentissage profond pour classer ces images recadrées dans les classes appropriées.

Mots clés : apprentissage en profondeur, intelligence artificielle, réseau neuronal conventionnel, reconnaissance optique de caractères, détection des cases à cocher, modèle de formation.