Republic of Iraq
Ministry of Higher Education and
Scientific Research
Diyala University
Faculty of Education for Pure Science
Computer Science Department

Recognition of squared handwritten digits by heuristic search

A Research Project Submitted in Partial Fulfillment of the Requirement for the Degree of Bachelor IN Education of computer science

By

Ali Hussein Abd

Under The Supervision Of

Dr. Adel A. Abd AL Wahab

1440 A.H.   2019 A.D
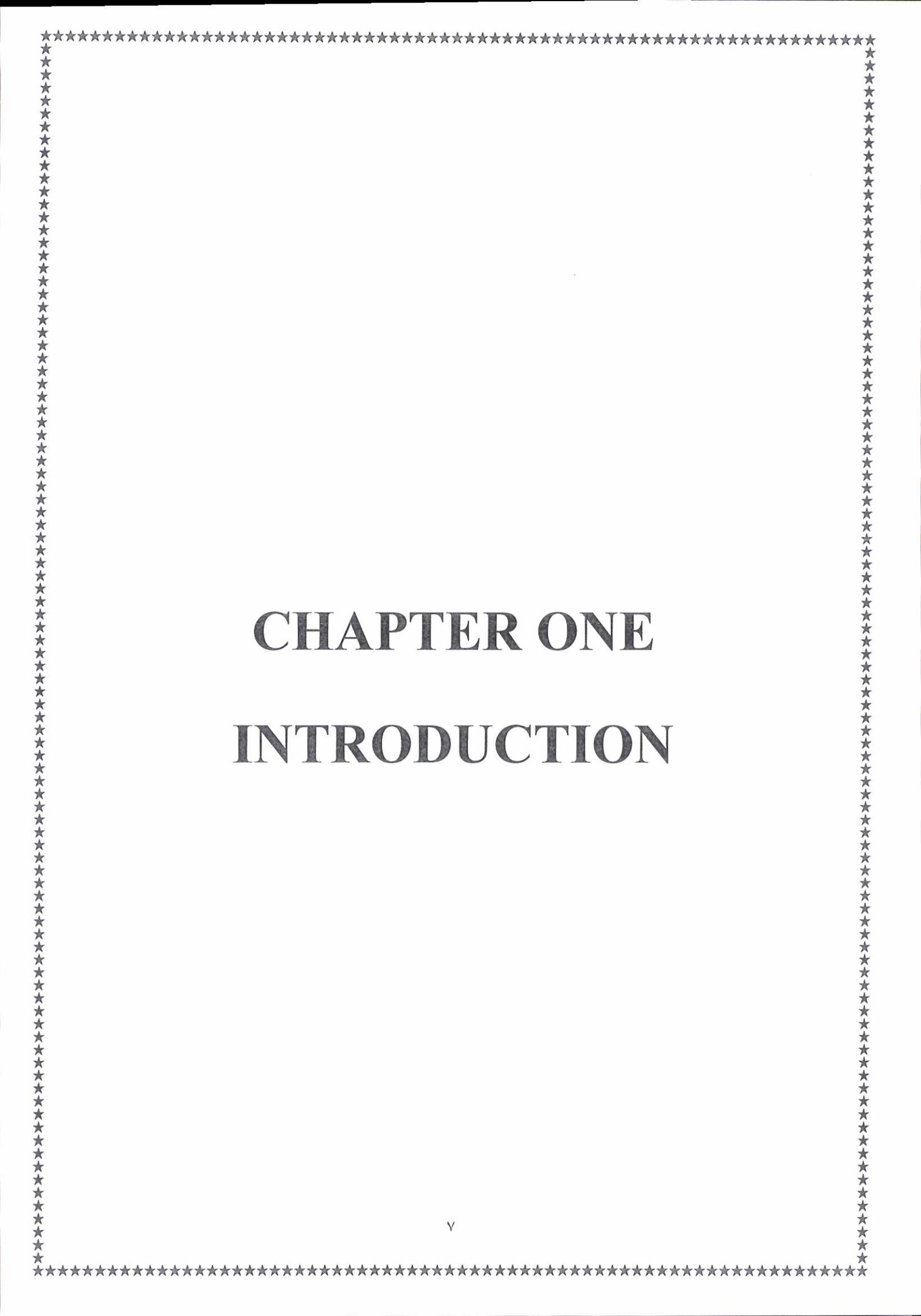
CHAPTER ONE

INTRODUCTION

1.1 <u>Introduction</u>

Human being is gifted with natural intelligence to recognize letters, objects, numbers, voices, and the like. However, making a machine solve these types of problems is a very difficult task. Pattern recognition is an essential component of artificial intelligence and computer vision. Interest in pattern recognition is fast growing aligned with the prohibitive amount of information that we encounter in our daily life. Consequently, computerization is franticly needed to handle this information explosion [2]. One of the difficult problems in the field of pattern recognition is Handwritten Digits Recognition (HDR), since the variation of the objects within each class is high, and simultaneously, objects from different classes may be quite analogous [3]. Furthermore, the ideas and methodologies to solve this problem would be very useful in many of pattern recognition problems that include large volumes of real-world data. In handwritten digits recognition tasks, formerly a digit is scanned, other preprocessing tasks come to pass,

Over the past three to four decades, many different methods have been explored and used in this field, including statistical, structural and syntactic methods, mathematical transforms, template (or model) matching, neural network and expert systems.

In general, algorithms with good performance have either large descriptive complexity or are computationally heavy (in training and/or classification) [1]. However, more work is still required before human performance is matched.

1.2 Problem of the Current System

Due to the manual means been used, a lot problems are encountered which includes:

1. Low accuracy: Man by nature is not very accurate.

2. Solve the problem of delay :in the manual system, for example, we find that the teacher needs a lot of time in the writing degrees process.

3. Lose of vital document as the filing system is manual.

4. Low security: An unauthorized person may can access resources in the paper system more easily than the Recognition of squared handwritten digits system which save degrees and students information into the access database.

5. Take a lot of time to retrieve: in the manual system it takes a lot of time to retrieve the documents or papers required in the process.

1.3 Research Objective

The aim of this study is to develop the work and use the software systems to make accuracy high and solve problems of delay and make the security high to increase efficiency of the work.

1.4 Research Scope

The research work focuses on the use of heuristic search algorithm for the development of work to recognition of squared handwritten digits to increase efficiency of the work.
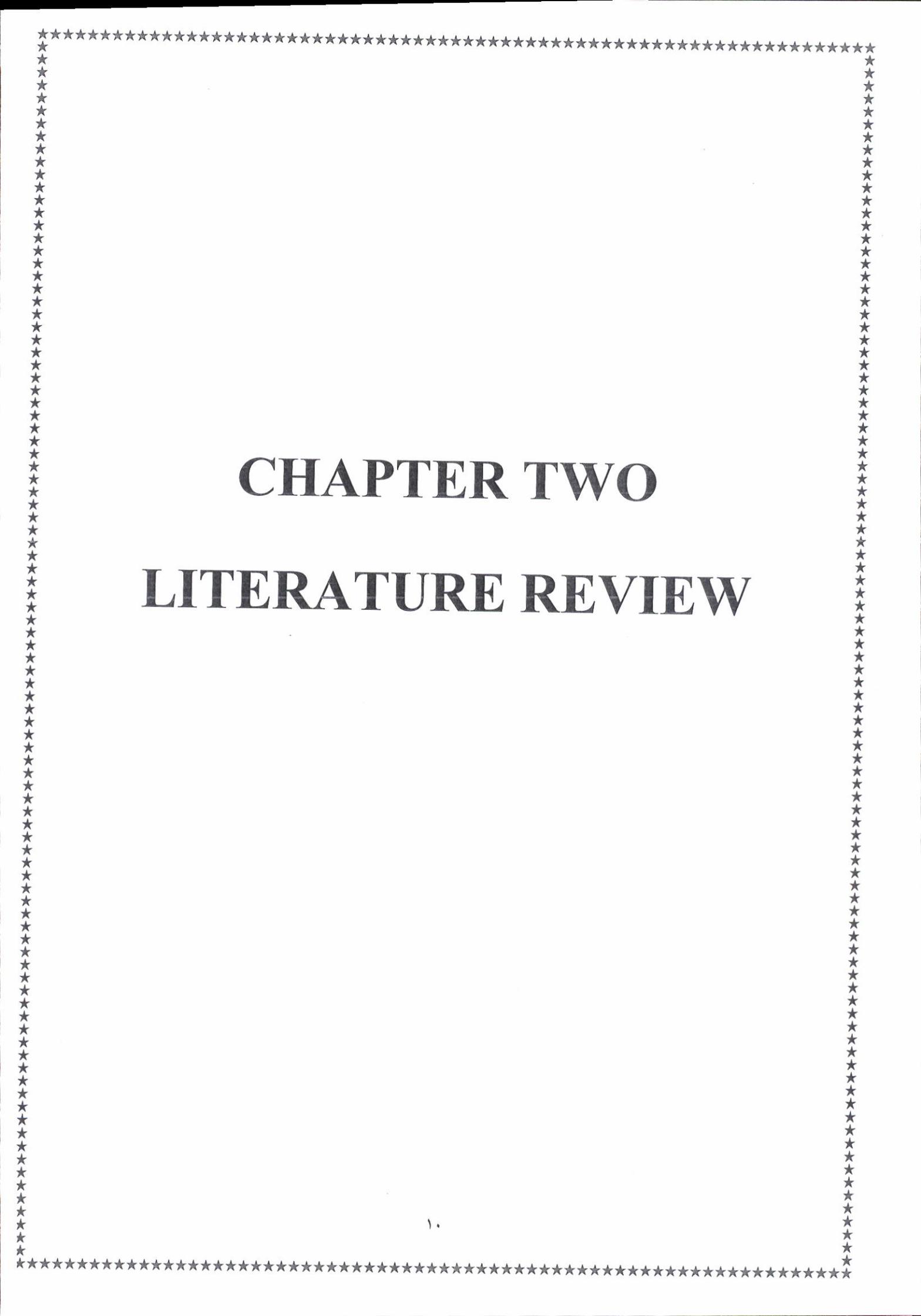
CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction [1]

Handwriting number recognition is a challenging problem researchers had been research into this area for so long especially in the recent years. In our research there are many fields concern with numbers, for example, checks in banks or recognizing numbers in car plates, the subject of digit recognition appears. A system for recognizing isolated digits may be as an approach for dealing with such application. In other words, to let the computer understand the numbers that is written manually by users and views them according to the computer process. Scientists and engineers with interests in image processing and pattern recognition have developed various approaches to deal with handwriting number recognition problems such as, minimum distance, decision tree and statistics.

Recently, a lot of works was done by depending on the computer; In order to let the processing time to be reduced and to provide more results that are accurate, for example, depending on different types of data, such as characters and digits and the numbers are used frequently in normal life operation. In order to automate systems that deal with numbers such as postal code, banking account numbers and numbers on car plates. And a Recognition of squared handwritten digits by heuristic search is proposed in this research.

2.2 Heuristic Algorithms [4]

Operations research is the whole set of methods involving finding the most appropriate solution for a given problem. Mathematical modeling of the real complex problems of the world is a scientific field that uses scientific methods such as algorithms and statistics. It optimizes the optimal solution for the query. Optimization Problems are the most desirable solutions. In general, NP-Complete is used to solve problems such as Decision Problems.

Optimization problems are making your fitness function minimum or maximum. In other words, increasing productivity in a company is maximizing optimization while reducing the duration of a business in the company is a minimum optimization. In fact, we face many times in our daily life with optimism. For example, when I write this article, I try to set up shorter sentences that will give the same clarification, while optimizing to the minimum, optimizing to maximize when searching for more and more efficient how to work today. Optimization problems are classified as discrete or continuous optimization according to the structure of decision variables in the mathematical model. In some places you can see intermittent optimization as a combinatorial. Interrupted optimization variables include a restricted set. For example, the famous Traveling Salesman Problem is solved with intermittent optimization. Continuous optimization problems are problems that do not limit the variables.

Intuitive Optimization Algorithms find a place to find solutions to their Optimization Problems in operational researches. A real life question that needs to be solved must first be cast into a mathematical model. Thus, using this mathematical formula as a criterion aims to find the best solution in a short time. Heuristic Optimization Algorithms do not guarantee the best solution. The algorithm is considered to be so effective as to how quickly it reaches a good solution.

EXAMPLE ALGORITHMS

1. Genetic Algorithms

2. Swarm Intelligence

3. Tabu Search

4. Simulated Annealing

5. Artificial Neural Networks [ANNs]

6. Support Vector Machines

2.3 Pattern recognition [7]

Pattern recognition is an area of study that is well-established and known through years of research, especially in the field of digit recognition which is considered one of the obvious challenges and one of the significant contributors to digit recognition.

2.4 Heuristics [6]

- Heuristics are rules to search to find optimal or near-optimal solutions. Examples are FIFO, LIFO, earliest due date first, largest processing time first, shortest distance first, etc.
- Heuristics can be *constructive* (build a solution piece by piece) or *improvement* (take a solution and alter it to find a better solution).

2.5 Optical characters recognition systems [5]

Today, the OCR (Optical Characters Recognition) systems are only able to recognition high quality printed or neatly handwritten documents. The current research is now basing on documents that are not well handled and including severely degraded, omnifont machine printed text, and unconstrained handwritten text. A wide variety of techniques are used to perform handwriting recognition. A general model for handwriting recognition is used to highlight the many components of a handwriting recognition system. The model begins with an unknown handwritten character that is presented at the input of the recognition system as an image. Firstly, to convert this image into information understandable by computers.

Significant progress has been made in these classification methods but more work is required to match human performance.

2.6 Working Background

Vb programming language, **access** Database, Pattern recognition, handwritten digits recognition problem, Heuristic search algorithm, Classification, Machine learning.

2.7 Data Bases

In the early days of computerization, it was normal to maintain specific files for individual application. Data where processes centrally in batches and there was little or no online interrogation of data. This approach is wholly inefficient for most of today's data processing systems. Supporting this enumerated the problems that result from organizing the data using the file system [8].

a. There exist high redundancies between files which result from the fact that the information is replicated in different places, and that these replications are not controlled by a central monitor

b. Inconsistencies might result from the possibilities that a program makes changes on the files it uses without these changes being made (at the same time) by all other programs that uses the files.

c. There exist in flexibility against changes in the application: if new actions or event arise in the cause of time, these can be realize at a substantial expense of time.

d. The work of many programmers involved is characterize by low productivity, seems program maintenance is expensive: if the structure of an existing file has to be modify during it life time, then all application program has to be modify correspondently.

e. Finally, there is the problem of adopting and maintaining standard (with respect to coding data format etc.), which is important for exchanging data or for migration to new operating system
Released, or even to a new computer system.

To overcome these problems, data bases where developed. It is now common for large organization to organize their operational data using the data base technology. The subject of data is adequately covered in many works in data base technology.

Data base as a collection of data supporting the operation of an organization. Quoting CIMA, provide a more detailed definition [8].
A data base is a file of data structured in such a way that it may serve a number of application without it structure being dictated by any one of those application, the concept being that programs are written round the data base rather than files being structure to meet the need of particular programs[9].

2.7.1 Access Database: [14]

Microsoft Access is a Database Management System (DBMS) from Microsoft that combines the relational Microsoft Jet Database Engine with a graphical user interface and softwaredevelopment tools. It is a member of the Microsoft Office suite of applications, included in the professional and higher editions.

- Microsoft Access is just one part of Microsoft's overall data management product strategy.

- It stores data in its own format based on the Access Jet Database Engine.

- Like relational databases, Microsoft Access also allows you to link related information easily. For example, customer and order data. However, Access 2013 also complements other database products because it has several powerful connectivity features.

- It can also import or link directly to data stored in other applications and databases.

- As its name implies, Access can work directly with data from other sources, including many popular PC database programs, with many SQL (Structured Query Language) databases on the desktop, on servers, on minicomputers, or on mainframes, and with data stored on Internet or intranet web servers.

- Access can also understand and use a wide variety of other data formats, including many other database file structures.

- You can export data to and import data from word processing files, spreadsheets, or database files directly.

- Access can work with most popular databases that support the Open Database Connectivity (ODBC) standard, including SQL Server, Oracle, and DB2.

- Software developers can use Microsoft Access to develop application software.

2.8 Introduction to VB

Is a multi-paradigm, object oriented programming language, implemented on the .NET Framework. Microsoft launched VB.NET in 2002 as the

successor to its original Visual Basic language. Although the ".NET" portion of the name was dropped in 2005, this article uses "Visual Basic [.NET]" to refer to all Visual Basic languages releases since 2002, in order to distinguish between them and the classic Visual Basic.

Along with Visual C#, it is one of the two main languages targeting the .NET framework. Microsoft's integrated development environment (IDE) for developing in Visual Basic .NET language is Visual Studio.

Most of Visual Studio editions are commercial; the only exceptions are Visual Studio Express and Visual Studio Community, which are freeware. In addition, .NET Framework SDK includes a freeware command-line compiler called vbc.exe. Mono also includes a command-line VB.NET compiler its advantages are:[1]

1. Easy and fast language for creating Windows applications.

2. It supports object oriented programming, but not fully.

3. Visual Basic is the object language of the curve.

4. Easy to learn and understand.

5. Easily detectable errors.

6. When you write correct commands, they give you examples to confirm that the code is correct.

7. Allows you to skip some errors when typing a specific code.

CHAPTER THREE

PRACTICAL PART

3.1 Introduction:

The system was created using Visual Basic programming language and an access database was used.

The system is able to read the examination books for the students who are examiners, extract the sequence and the degree, and then retrieve the student information based on the sequence of the student that was previously stored in the database and then the storage of the student's exam score in the database.

```
                    ┌──────────────┐
                    │    Start     │
                    └──────┬───────┘
                           │
         ┌─────────────────▼───────┐
         │      Input image        │
         └─────────────┬───────────┘
                       │
             ┌─────────▼─────────────┐
             │ Extract number of degree
             │ and sequence from image │
             └─────────┬─────────────┘
                       │
    No         ◇ Get name of student ◇
               ◇   from database     ◇
               ◇ depending on sequence ◇
                       │
                      Yes
             ┌─────────▼─────────────┐
             │ Save degree to database │
             └─────────┬─────────────┘
                       │
                    ┌──▼───┐
                    │ End  │
                    └──────┘
```
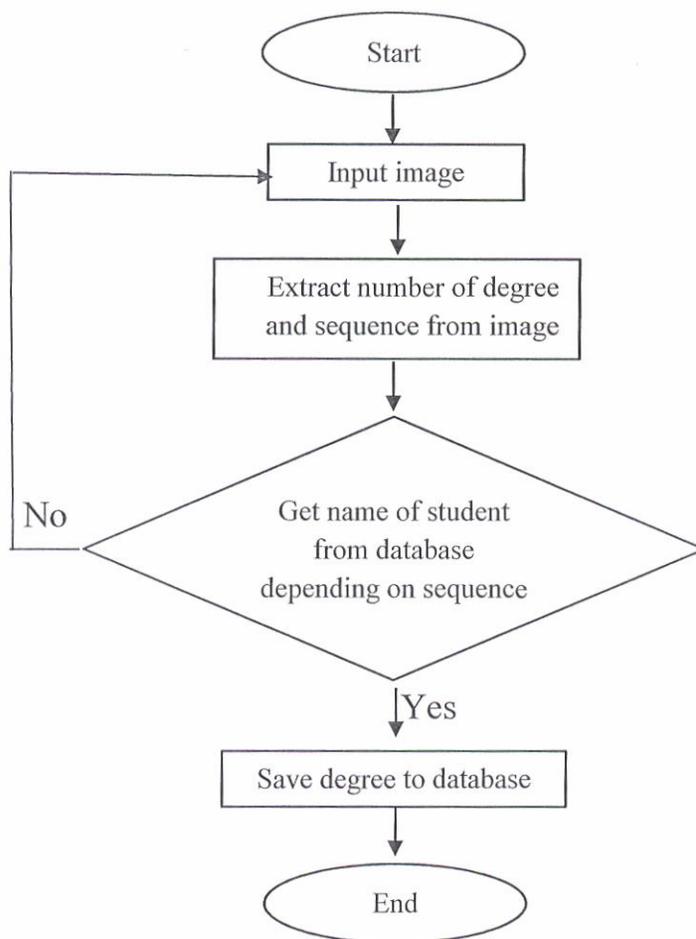
The flow diagram of the implemented system

In the beginning we have to have a dataset for the images that were taken in advance in order to extract the numbers from them using our algorithm:
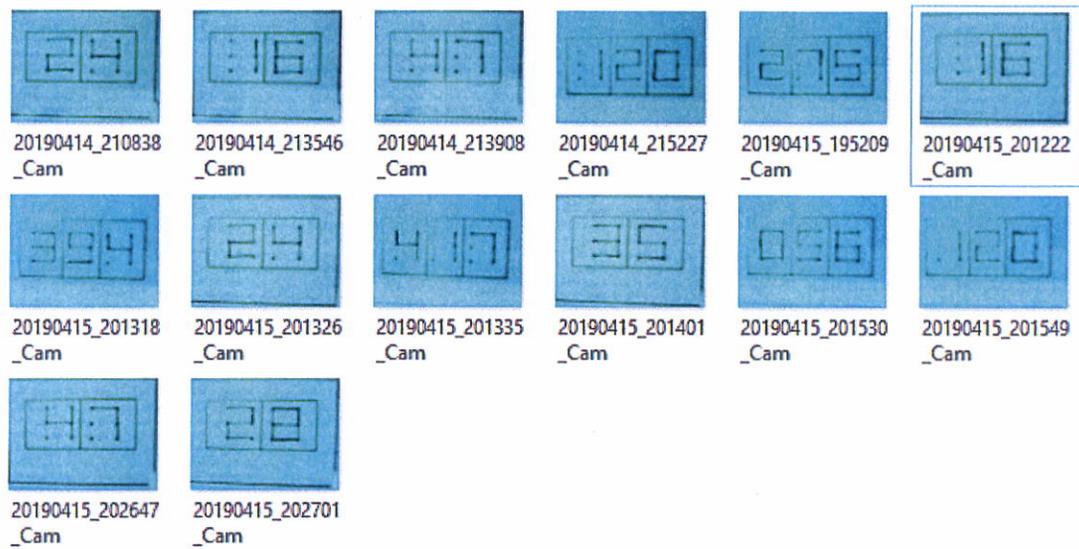
Figure (1) dataset contents

We have created an Access database and stored student data in order to call the student information based on the sequence entered from dataset and then store the student's score in the database:
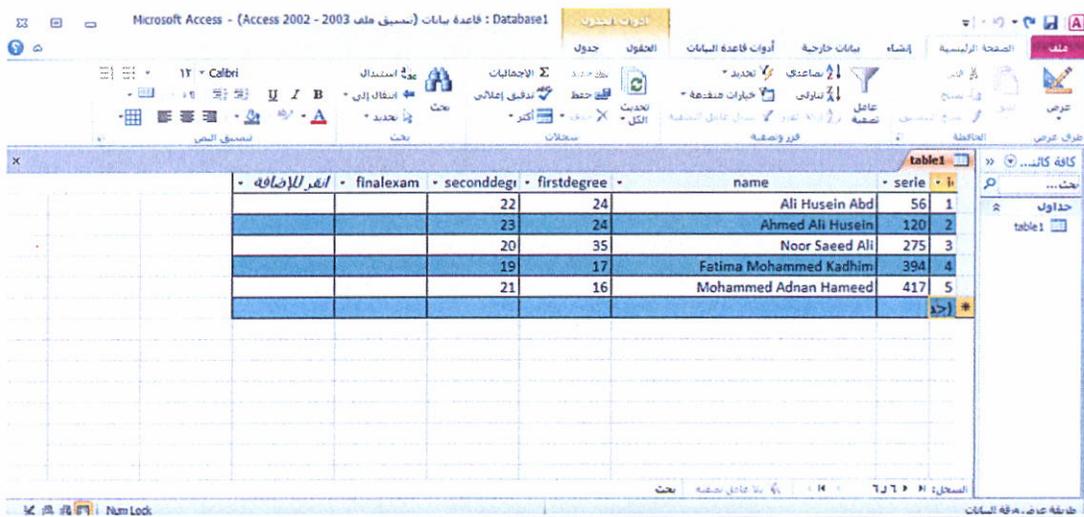

Figure (2) access database file contents

The main interface that appears in the first operation of the system is:

Recognition of squared handwritten digits by heuristic search

 وم بمعالجة الصور الملتقطة لدفتر امتحاني ومعرفة الأرقام داخل الصور التي تمثل تسلسل ودرجه الطالب لمادة معينة واضافة الدرجة الى قاعدة بيانات بالاكسس
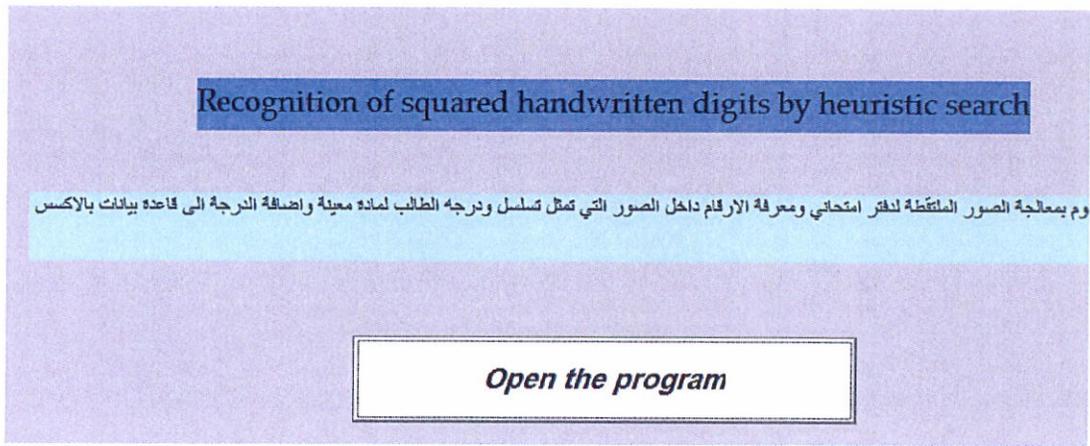
Open the program

Figure (٣) Main interface of the system

Where it shows a simple form that includes a definition of the system and below it a button to enter the interface of the implementation of the system.

After opening the program, the main implementation interface will appear as follows:
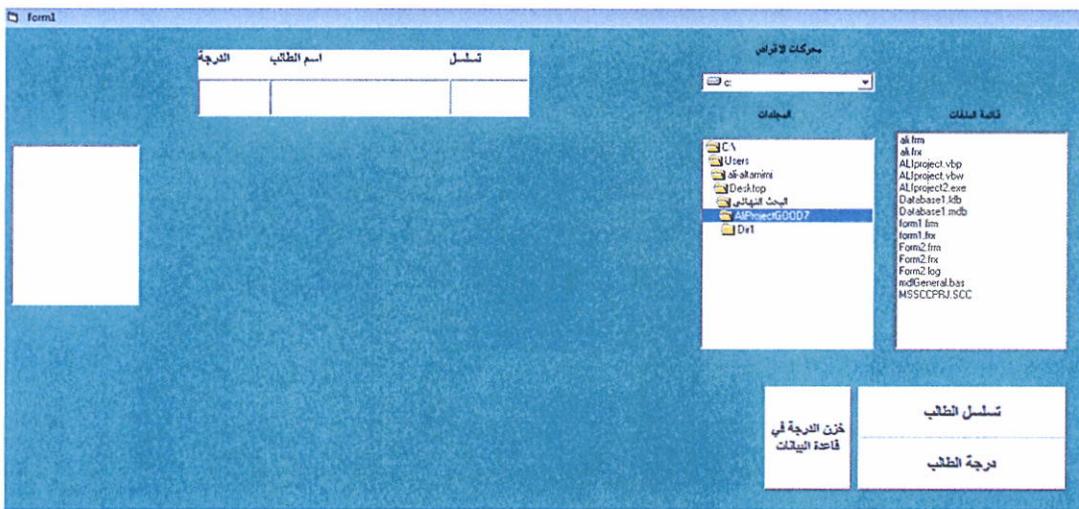
After we select an image of the grading images or sequences of the examiners as in Figure (4) before the process begins:
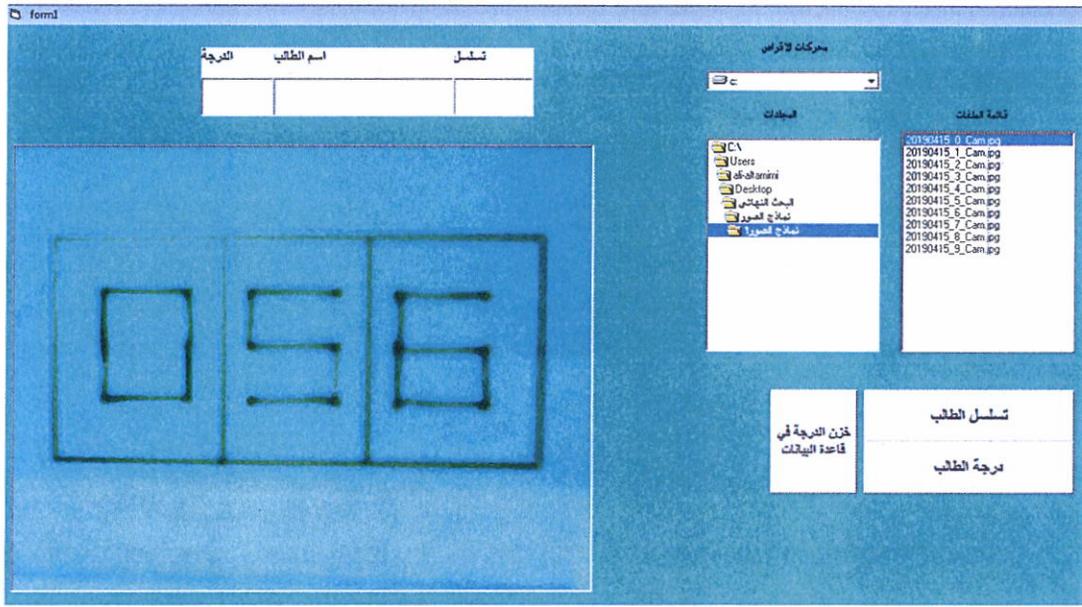
Figure (٤) Implementation interface before processing

When clicking on the processing button to read the sequence, the system will analyze the image and read it and extract the serial number from it and retrieve the student name from the database depending on the serial number of the student as in Figure (5):
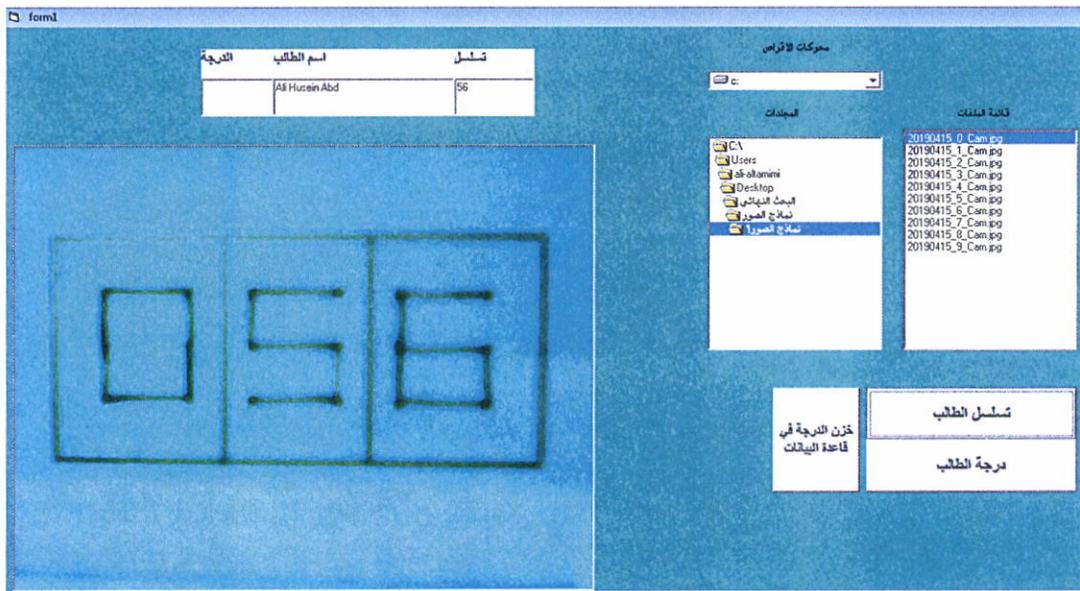

Figure (٥) Implementation interface after processing

In Figure 6, we will observe how the system reads student degree and how it is stored in the database:
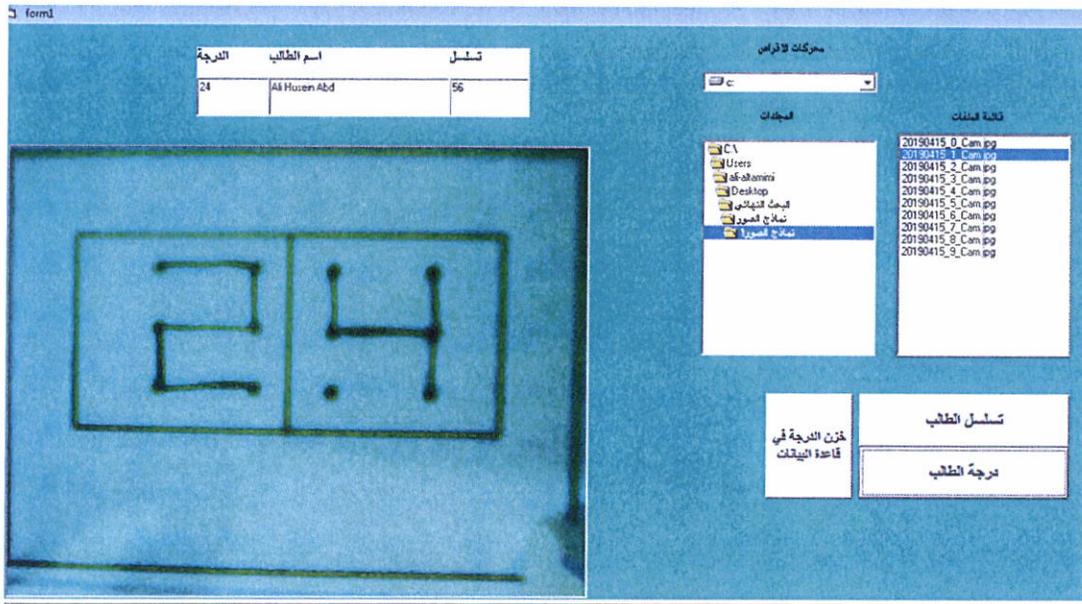
Figure (٦) read student degree and add to the database

After opening the database we will notice that the student's degree has been added to the database:

		24	22	24	Ali Husein Abd	56	1
			23	24	Ahmed Ali Husein	120	2
			20	35	Noor Saeed Ali	275	3
			19	17	Fatima Mohammed Kadhim	394	4
			21	16	Mohammed Adnan Hameed	417	5
						###	*

Finally, we created a system using the Visual Basic environment, which is capable of extracting the numbers from the exam books for the students who are tested and retrieving the student information from the database based on the student sequence that was extracted from the exam book and storing the student's score in the database.