People's Democratic Republic of Algeria

Ministry of Higher Education and Scientific Research

UNIVERSITY OF SAIDA DR MOULAY TAHAR

FACULTY OF TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE

ORDER N °: .....

THESIS

Presented by

BOUKHALFA Siham

For the obtain of the

«L. M. D» DOCTORAL degree in COMPUTER SCIENCE

Specialty : Computer Science

OPTION : Computer Science

Big Data and Internet of Things

Defends publicly, in ../../2021

In front of a jury composed of :

HAMOU Reda Mohamed	Professor	University of saida Dr MOULAY Tahar	President
AMINE Abdelmalek	Professor	University of saida Dr MOULAY Tahar	Supervisor
RAHMOUN Abdelatif	Professor	Higher School of Computer Science of Sidi-Bel-Abbes	Examiner
BENDAOUD Zakaria	MCA	University of saida Dr MOULAY Tahar	Examiner
MEKKAOUI Kheireddine	MCA	University of saida Dr MOULAY Tahar	Examiner

Academic Year : 2021-2022

Knowledge management and complex data laboratory (GeCoDe Lab), University of saida

Acknowledgements

In the Name of **Allah**, the Merciful, the Beneficent. Prayers and peace be upon our Prophet, **Mohamed** aalayh Elssalet wa Elssalem , his family, and all of his companions.

First and foremost, I would like to praise Allah the Almighty, for his help and protection, and for giving me the patience, ability, will and courage to do this work.

The completion of this thesis would not have been possible without the guidance and support of a number of people. This is the opportunity to thank them.

First and foremost, I would like to give my special gratitude and appreciation to my thesis supervisor, Professor **Abdelmalek AMINE**, professor and head of **GECODE** laboratory in university of saida Dr MOULAY Tahar, for his continuous academic and emotional support, his patience, his motivation, and his immense knowledge. I thank him for having shared with me all his know-how and for being available, for his guidance helped me all the time of the research and writing of this thesis. He gave me his valuable suggestions and corrections made; I hope to continue working with his noble thoughts.

My sincere gratitude goes to Professor **Reda Mohamed HAMOU**, for his support and encouragement during my period in this doctoral training. His enthusiasm for research has inspired me to continuously improve as a scientist. I also like to thank him, for doing me the honor of chairing the jury.

My special thanks goes to all the members of the jury who took the time to read this modest work and to evaluate it, represented by:Prof **RAHMOUN Abdelatif**, Professor at Higher School of Computer Science of Sidi-Bel-Abbes , Dr **BENDAOUD Zakaria**, Senior Lecturer at University of saida Dr MOULAY Tahar, Dr **MEKKAOUI Kheireddine**, Senior Lecturer at University of saida Dr MOULAY Tahar;I am honored and delighted that they have accepted to read, examine, and enlighten this work and me with their valuable expertise.

Next, I also thank my friends who will recognize themselves here, for their valuable help and especially for their availability which allowed me to carry out this work with their advice and guidance.

Finally, I would like to thank all those who supported me in the realization of this thesis.

Dedication

To my beloved mother and father for their endless Love, Support and Encouragement.

To my brother and sisters.

To my husband who inspired me all the time.

To my sweet daughter Fatima zahra.

And to my dear Supervisor Pr.Amine Abdelmalek.

As well as to all my family, friends and everyone who motivated me to complete this research.

- Abstract -

Nowadays, in a world covered by networks, there are more smart devices than peoples, since a person owns different smart devices in different forms. These devices, which interconnect and exchange a very large flow of data, perform several functions including monitoring, data collection, and data evaluation.
In this thesis, we will focus on this new trend of interconnected objects used to improve the daily life of individuals. For this, the exploitation of the Internet of Things in the field of monitoring and control is a recent research axis that helps human beings to ensure this task based on the data captured by the intelligent devices that will be subsequently analyzed and processed by different methods. It is in this context that we orient our research on the concept of linking objects to the Internet, known today as the Internet of Things. Our work is articulated around two issues, physical activity and fall prevention in the elderly and the security of international borders.
In our first work, we proposed an approach based on metaheuristics for real-time security and boundary protection. This technique is inspired by the behavior of natural cockroaches and the phenomenon of seeking the most attractive and secure place to hide.
In our second work, we used classification algorithms to combat the risk of falls in the elderly and enable these individuals to continue their lives in the best possible condition. We examine the applicability of three data mining algorithms for real-world IoT datasets. These include K-nn, Naive Bayes, and Decision Tree. The main contribution of this work is the analysis of the efficiency of three data mining algorithms.
All the experiments carried out and the results obtained have shown the benefits derived from the use of our system.

Keywords: Big Data, Internet of Things, Data mining, metaheuristic, RFID, Classification, Human activity recognition, Batteryless wearable sensor, cockroach, older people, Border Surveillance, Security, Border Patrol, Video Surveillance, Security Alarm, Remote Surveillance System.

- ملخص -

اليوم، في عالم تغطيه الشبكات، يوجد عدد من الأجهزة الذكية أكثر من الأشخاص لأن الشخص الواحد لديه أجهزة ذكية مختلفة، بأشكال مختلفة. تقوم هذه الأجهزة التي تربط وتتبادل تدفقًا كبيرًا جدًا من البيانات بأداء العديد من الوظائف بما في ذلك المراقبة وجمع البيانات وتقييم تلك البيانات.

في هذه الأطروحة سنهتم بهذا الاتجاه الجديد للأشياء المترابطة المستخدمة لتحسين الحياة اليومية للأفراد. لهذا فإن استغلال إنترنت الأشياء في مجال المراقبة والتحكم هو محور بحث حديث يساعد البشر على أداء هذه المهمة بناءً على البيانات التي تلتقطها الأجهزة الذكية والتي سيتم من خلالها تحليلها ومعالجتها لاحقًا بطرق مختلفة.

في هذا السياق نوجه بحثنا إلى مفهوم ربط الأشياء بالإنترنت، والمعروف اليوم باسم "إنترنت الأشياء تمحور عملنا حول قضيتين، النشاط البدني والوقاية من سقوط كبار السن وحماية وأمن الحدود الدولية.

في عملنا الاول، اقترحنا نموذجا استرشاديا مستوحى من الطبيعة للمساعدة في مراقبة وحماية الحدود في الوقت الفعلي. هذه التقنية مستوحاة من السلوك الطبيعي للصراصير وظاهرة البحث عن أكثر الأماكن جاذبية وأمانًا للاختباء.

في عملنا الثاني، استخدمنا خوارزميات التصنيف لمحاربة مخاطر السقوط لكبار السن والسماح لهؤلاء الأشخاص بمواصلة حياتهم في أفضل الظروف الممكنة. ندرس إمكانية تطبيق ثلاث خوارزميات لاستخراج البيانات لمجموعات بيانات إنترنت الأشياء الحقيقية. المساهمة الرئيسية لهذا العمل هي تحليل كفاءة ثلاث خوارزميات لاستخراج البيانات.

أظهرت جميع التجارب التي تم إجراؤها والنتائج التي تم الحصول عليها الفائدة المستمدة من استخدام نظامنا.

الكلمات المفتاحية : البيانات الضخمة ، إنترنت الأشياء ، التنقيب عن البيانات ، الأدلة العليا ، تحديد الهوية بموجات الراديو، التصنيف ، التعرف على النشاط البشري ، مستشعر يمكن ارتداؤه بدون بطارية ، الصراصير ، كبار السن ، مراقبة الحدود ، الأمن ، حرس الحدود ، المراقبة بالفيديو ، إنذار الأمن ، نظام المراقبة عن بعد.

- RÉSUMÉ -

Aujourdh́ui, dans un monde couvert par des réseaux, il y a plus dáppareils intelligents que de personnes puisquúne personne possède différents appareils intelligents, dans différentes formes. Ces appareils qui sínterconnectent et échangent un flux de données très important assurent plusieurs fonctions dont la surveillance, la collecte des données et l évaluation de ces données.
Dans cette thèse on va síntéresser à cette nouvelle tendance des objets interconnectés utilisés dans le but dáméliorer la vie quotidienne des individus. Pour cela, Léxploitation de línternet des objets dans le domaine de la surveillance et le contrôle est un axe de recherche récent qui aide l'être humain à assurer cette tâche en basant sur les données capturées par les appareils intelligents qui seront par la suite analysés et traités par différentes méthodes.Cést dans ce contexte que nous orientons notre recherche sur le concept de lier les objets lÍnternet, connu de nos jours comme Internet of Things. Nos travaux se sont articulés autour de deux problématiques, láctivit physique et prévention des chutes chez les personnes âgées et la sécurité des frontière internationale.
Dans notre premier travail nous avons proposé une approche basée sur les métaheuristiques dáide à la sécurité et la protection des frontières en temps réel. Cette technique est inspirée par le comportement des cafards naturel et le phénomène de chercher léndroit le plus attrayant et sécurisé pour se cacher.
Dans notre deuxième travail nous avons utilisé des algorithmes de classification pour lutter contre les risques de chute des personnes âgées et permettre à ces personnes de continuer leur vie dans les meilleures conditions possibles. Nous examinons lápplicabilité de trois algorithmes déxploration de données pour des ensembles de données IoT réels. Ceux-ci incluent K-nn, Naive Bayes, Arbre de décision. La principale contribution de ce travail est lánalyse de léfficience de trois algorithmes de léxploration de données.
Lénsemble des expérimentations réalisées et les résultats obtenus ont montré le bénéfice dérivé de lútilisation de notre système.

Mots clés: Big Data, Internet des objets, fouille de données, métaheuristique, RFID, Classification, Reconnaissance de l'activit humaine, Capteur portable sans pile, Cafard, Personnes âgées, Surveillance des frontières, Sécurité, Patrouille des frontières, Vidéosurveillance, Alarme de sécurité, Système de surveillance à distance.

Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

ABA	**A**rtificial **B**at **A**lgorithm
ABC	**A**rtificial **B**ee **C**olony Algorithm
ACA	**A**nt **C**olony **A**lgorithm
AI	**A**rtificial **i**ntelligence
AIS	**A**rtificial **I**mmune **S**ystems
AMQP	**A**dvanced **m**essage **Q**ueue **p**rotocol
BAT	**Bat** Optimization Technique
BD	**B**ig **D**ata
B&B	**B**ranch **and** **B**ound method
BFO	**B**acterial **F**oraging **O**ptimization
BUTLER	u**B**iquitous, sec**U**rein**T**ernet-of-things with **L**ocation and cont**E**x-awa**R**eness
CRM	**C**ustomer **R**elationship **M**anagement
CoAP	**Co**nstrained **A**pplication **P**rotocol
CONV	**CONV**olution layer
CNNs	**C**onvolutional **N**eural **N**etworks
CCTV	**C**losed-**C**ircuit **T**ele**V**ision
COP	**C**ombinatorial **O**ptimization **P**roblem
DARPA	**D**efense **A**dvanced **R**esearch **P**rojects **A**gency
DL	**D**eep **L**earning

DM	**D**ata **M**ining
EAs	**E**volutionary **A**lgorithms
ES	**E**volution **S**trategies
FC	**F**ully **C**onnected layer
GA	**G**enetic **A**lgorithms
GPS	**G**lobal **P**ositioning **S**ystem
GP	**G**enetic **P**rogramming
GUI	**G**raphical **U**ser **I**nterface
H2T	**H**uman to **T**hing
HTTP	**H**yper **T**ext **T**ransfer **P**rotocol
IEEE	**I**nstitute of **E**lectrical and **E**lectronics **E**ngineers
IM	**I**nstant **M**essaging
IoE	**I**nternet of **E**verything
Ipv6	**I**nternet **P**rotocol **V**ersion **6**
IETF	**I**nternet **E**ngineering **T**ask **F**orce
IERC	**I**oT **E**uropean **R**esearch **C**luster
IoT	**I**nternet of Things
ILP	**I**nteger **L**inear **P**rogramming
JSON	**J**ava**S**cript **O**bject **N**otation
K-NN	**K**-**N**earest **N**eighbor
KDD	**K**nowledge **D**iscovery in **D**ata
ML	**M**achine **L**earning
M2M	**M**achine to **M**achine
MQTT	**M**essage **Q**ueue **T**elemetry **T**ransport
NASA	**N**ational **A**eronautics and **S**pace **A**dministration
NITRD	**N**etworking and **I**nformation **T**echnology **R**esearch and **D**evelopment
NB	**N**aive **B**ayes
NN	**N**eural **N**etworks

OCU	**O**perator **C**ontrol **U**nit
PSO	**P**article **S**warm **O**ptimization
PSOA	**P**article **S**warm **O**ptimization **A**lgorithm
POOL	**Poo**ing **L**ayer
RNNs	**R**ecurrent **N**eural **N**etworks
REST	**R**epresentational **S**tate **T**ransfer
RPL	**R**outing **P**rotocol for **L**ow power and lossy networks " SVM Support vector machine
SPUs	**S**ensor **P**rocessing **U**nits
SOA	**S**ervice-**O**riented **A**rchitecture
T2T	**T**hing-to-**T**hing
TCP	**T**ransmission **C**ontrol **P**rotocol
T2H	**T**hing to **H**uman
UDP	**U**ser **D**atagram **P**rotocol
URI	**U**niform **R**esource **I**dentifier
VSAM	**V**ideo **S**urveillance and **M**onitoring System
WSN	**W**ireless **S**ensor **N**etwork
XML	e**X**tensible **M**arkup **L**anguage
XMPP	e**X**tensible **M**essaging and **P**resence **P**rotocol
6LoWPAN	IPv6 over **Lo**w power **W**ireless **P**ersonal **A**rea **N**etworks

INTRODUCTION

Context

Today, in a world covered by networks, there are more smart devices than people, since a person has different smart devices in different forms. Connecting all These devices and forming a network of devices is the basic idea of IoT. A major objective of IoT is to make the environment around us smarter, by giving the environment the information it needs, the IoT uses the internet to connect devices that can be easily monitored and controlled, also the same things can be automatically detected by other things, further communicate with each other through the internet, and can even make decisions themselves.
Over time, various sensory data are collected and generated by an enormous amount of sensing devices. This will result to generate in a big amount of data from the sensors used for collecting the data. To prevail over these applications some meaningful information must be deduced out of the collected data to make decisions. Applying analytics over such data streams to make control decisions, discover new information and foresee future insights is a pivotal procedure that makes IoT a worthy paradigm for businesses and a quality-of-life improving technology. Among the most extremely useful technologies are Data Mining and metaheuristic. A major challenge in these settings is the timely analysis of large amounts of data (big data) to produce decisions and highly reliable and accurate insights so that IoT could satisfy its guarantee. The aim of this work is to check whether the traditional data mining and meta-heuristic algorithms would likewise work for the IoT datasets, or new families of data mining and meta-heuristic algorithms are required.

Problematic

The great power of the IoT lies in the fact that its objects communicate, analyze, process, and manage data autonomously and without any human intervention. However, a major challenge in these settings is the timely analysis of large amounts of data (big data) to produce decisions and highly reliable and accurate insights so that IoT could satisfy its guarantee. The aim of this work is to check whether

the traditional data mining and meta-heuristic algorithms would likewise work for the IoT datasets, or new families of data mining and meta-heuristic algorithms are required.
In our research we have touched on two major problems, the first problemis the border and sensitive site surveillance. We applied a new meta-heuristic to the IoT data. The second contribution of this research is the introduce a new meta-heuristic inspired by the natural behavior of cockroaches for the surveillance of borders through videos captured by one of the drones via sensors.
Security along the international border is a critical process in security assessment due to increased terrorism risks, illegal immigration, and menaces. Border surveillance is one of the fundamental parts of the nations security because it is the system that maintains security along the border. It must be exercised 247 , and no break and no fall of alertness can be tolerated. The illegal border crossing activities have been raised due to less secure fence border security systems. The traditional border surveillance systems consist of borders soldiers, which cannot offer surveillance all the time. As technology advancement is increasing, automatic ways of increasing border security along international borders are needed. On the other hand, the implementation of security and surveillance systems is significant over borders for the prevention of illegal immigration, smuggling, and terrorism dependable detection and at long distance and the identification of the possible threats of the day as night and in all weathers and to spot and tell possible menaces as fast as possible This is why we will present techniques to overcome this problem by applying a metaheuristic method to analyze data (images and videos) emitted by IP cameras and sensors placed in drones along the borders and make decisions in real-time by detecting suspicious acts.
The second problem is the problem of falls in the elderly; we compare three data mining algorithms (Naive Bayesian, KNN, and Decision Tree) for IoT data The contribution of this work is to focus on analyzing techniques that are used for prediction purposes of falls in the elderly; in which Traditional care of an elderly person is a difficult, costly and complex task. The need to have a caregiver with the elderly person almost all the time drains the human and financial resources of the health care system. In addition, the relationship is often complicated between the caregiver and the elderly who wishes to preserve his or her privacy. The emergence of the field of Artificial intelligence has allowed the conception of technological assistance where an artificial agent, also called an ambient agent, helps and reduces the time spent by the caregiver with the elderly person. This work presents a methodology based on multimodal sensors to configure a simple, comfortable and fast fall detection and human activity recognition system that can be easily implemented and adopted. This work aims to focus on analyzing techniques that are used for prediction purposes of falls in the elderly. The parameters detection is gotten by the utilization of sensors and is kept up as datasets. Internet of Things (IoT) is used primarily to gather data from the user. We examine the applicability of three data mining algorithms (Naive Bayesian, KNN, and Decision Tree) for IoT data. These algorithms are analyzed and a comparative study is undertaken to find the classifier that performs the best analysis on the dataset obtained, using a set of predefined performance metrics to compare the

results of each classification.

Contributions

The aim of this thesis is to check whether the traditional data mining and meta-heuristic algorithms would likewise work for the IoT datasets, or new families of data mining and meta-heuristic algorithms are required. To this end, first we present an approach to assist in border and sensitive site security through the push the limits of facial identification in the context of video surveillance for responding to suspicious activities automatically. For this, an intelligent border and sensitive site surveillance system must:

1. Prediction and Prevention of falls among Elderly People.
2. Follow older persons and human activity recognition.
3. Ensuring the privacy of elderly.
4. Assist caregivers or family members to enable appropriate interventions.
5. Protecting the privacy of the elderly.

Secondly, we compare three data mining algorithms (Naive Bayesian, KNN, and Decision Tree) for IoT data The contribution of this work is to focus on analyzing techniques that are used for prediction purposes of falls in the elderly;
For this, a system for recognising activities in healthy older people must:

1. Identify malevolent people (thieves, terrorists, criminals. etc.) without facial recognition.
2. Follow suspect persons.
3. Ensuring the privacy of individuals.
4. Assist authorities in their investigations.
5. Preventing crime before it is committed.

Our contribution is to focus on problems 1, 2, 3, 4, and 5 by the use of different recent technologies to perform this task;

- The Internet of things
- Classification algorithms
- Metaheuristics
- The big data

Thesis Organization

In this thesis, we mainly dealt with two problems, the first problem is the use of Artificial Intelligent to combat falls risks of older folks. We examine the applicability of three data mining algorithms for real IoT datasets. These include K-nn, Naive Bayes, Decision tree the main contribution of this work is the analysis of the efficiency and effectiveness of three of the data mining. The secon problem is real-time border and sensitive site protection assistance, which we have completed with a proposal for an approach based on IP cameras, sensors, and metaheuristics.
The plan of this thesis is organized around a progression of the reflection, starting from the problematic as treated in the state of the art (see part 1) towards the operational solutions that have been designed, through the description of our contribution, the experiments, and the results. (See part 2).
The first part presents an overview of the different research fields concerned by our problematic, starting with a state of the art on Big Data, IoT, Data Mining, and Metaheuristic (see Chapter 1, 2).
This first study allows us to expose the essential concepts for understanding the field of research and to emphasize the key points related to our problematic. Then, we approach the contribution part.
In the second part, we present our contribution. In chapter 3 and 4,
In chater 3, the approach 'Border security and surveillance System using IoT' is detailed and, the approach 'Recognizing Physical Activity of hospitalized Older People from Wearable Sensors Data using IoT' is detailed in chapter 4.
We began this chapter with a detailed description of our system, which will be followed by a presentation of the experiment, in the end, results and analysis are reported, and finally, Conclusions are drawn.

PART ONE: Theoretical background

1 BIG DATA AND IOT

Contents

1.1 Introduction

Nowadays, enormous amounts of data circulate massively, from various sources of social networks, the internet, Google, mobile devices, GPS, etc. This data flow rapidly and in real-time and on a large scale of different types and structures (texts, audio, images, and video...), big data is a term that has appeared with the increase in these data. The challenge faced in this context traditional system and data warehouses encounters the problem of performance degradation in the face of such a large amount of data in terms of analysis and processing. Big data solve this problem and several fields use big data as an optimal solution for data management.

The term Internet of Things (IoT) emerged into public view only more recently, but it is not a new concept. In 1999, Kevin Ashton the executive director of the Auto-ID Center was laying the groundwork for what would become the Internet of Things that is a technological revolution that represents the future of computing and communications, and its development depends on dynamic technical innovation in several important fields [184]. IoT is a paradigm that consolidates technologies and aspects coming from various methodologies. Internet Proto-

col, communication technologies (Machine-to-Machine Communication (M2M)), RFID Radio Frequency Identification, Ubiquitous computing, sensing technologies, Smart Services, and embedded devices are merged to form a system where the real and digital worlds meet and are continuously in symbiotic interaction [26].

1.2 Data Science

Data Science brings together all the techniques used to extract information from data. It is an area that consists of all that concerns the collection, cleaning, preparation, analysis, and processing of structured and unstructured data.
Data Science is a multi-disciplinary approach to the intersection of mathematics, statistics, data analysis, information theory, and computer programming that aims to extract knowledge from data.
It can be defined simply as the discipline specialized in the study of information. More specifically, it is interested in their source, what they represent and the methods to be used to transform them into useful resources.

1.3 Artificial Intelligence (AI)

Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term can also be applied to any machine that exhibits traits associated with a human mind such as learning and problem solving.
The term Artificial Intelligence (AI) was introduced by McCarthy in 1956 [106] during a conference at Dartmouth College, and the term has since been adopted to represent the field. AI is directly related to the concepts of knowledge-based systems, expert systems, intelligent systems, knowledge acquisition, machine learning among other subjects of study and practical application.
Artificial intelligence is based on the principle that human intelligence can be defined so that a machine can easily imitate it and perform tasks, from the simplest to the more complex ones. The goals of artificial intelligence include learning, reasoning, and perception. With technological advancements, previous criteria that have defined artificial intelligence becomes obsolete. For example, machines that calculate basic functions or recognize text through optimal character recognition are no longer considered embodying artificial intelligence since this function is now taken for granted as an inherent computing function. AI is continuously evolving for the benefit of many different industries. The machines are linked according to an interdisciplinary approach based on mathematics, computer science, linguistics, psychology.
AI is widely used by large organizations to simplify the life of an end-user. The uses of artificial intelligence would largely fall under the category of data processing, which includes the following:

- Data research and search optimization to yield the most relevant results.
- Logic chains for if-then reasoning, which can be applied to execute a series of commands based on parameters.

- Pattern detection to identify significant patterns in a large data set for a unique perspective.
- Probabilistic models applied to predict future results.

1.4 Knowledge Discovery Data

Knowledge discovery in data (KDD) is an iterative and interactive process of analyzing a large set of raw data in order to extract knowledge that can be used by an analyst user who plays a central role in the process.
The Knowledge Discovery from Data (KDD) process [4] involves browsing the big data in a database, looking for knowledge.
This process includes steps of problem definition (domain definition, end-user purpose), data preparation (selection, preparation, transformation), data mining (Selection, appropriate data mining tools, pattern matching), and evaluation of results to arrive at new knowledge.
The process presented is iterative, and several backtracking in the different steps may be necessary to refine the results.

1.5 Data mining

First, certain distinctions between the following concepts: data, information, and knowledge are to be discerned. Several definitions can be found in the literature [23] [117]. These three notions are defined as follows:
Data: is a basic element that represents information in a database, namely a measure or a characteristic. It describes specific examples or events. It can be collected automatically or in writing. The notion of data is therefore perceived as the lowest level layer in the conceptual hierarchy of knowledge.
Information Information is an element of knowledge capable of being coded or represented using conventions to be stored, processed, or communicated, when one gives meaning to a given through an interpretative framework, it becomes information, information is organized data presented in context, the notion of information, therefore, has a higher conceptual degree than the notion of data in value and meaning. [29]
The knowledge Knowledge is interpretable and exploitable information, having a meaning. So it is defined either as refined, synthesized, systematized information or as information associated with a context of use. [167].

1.5.1 Data Mining Definitions

Data Mining (DM) is the set of methods and techniques intended for the exploration and analysis of computer databases (often large) automatically or semi-automatically, to detect in these data of rules, associations, unknown or hidden trends, particular structures restoring most of the useful information while reducing the quantity of data [104].
Frawley and Piatesky-Shapiro define data mining as the extraction of original, previously unknown that potentially useful information from data [57].

According to [87], the most commonly accepted definition of DM is that of [56]: Data Mining is a non-trivial process that consists of identifying, in data, valid, potentially useful, and above all understandable and usable.
In short, we can say that Data Mining is a set of techniques and methods in the field of statistics, mathematics, and computer science, allowing the extraction from a large volume of raw data of previously unknown original knowledge.
It appeared in the mid-1990s in the United States, as a new discipline at the interface of statistics and information technologies (databases, artificial intelligence, machine learning) [36].
The data mining community initiated its first conference in 1995 following numerous workshops on KDD between 1989 and 1994. The first journal in the field Data mining and knowledge discovery journal published by Kluwers was launched in 1997. [36]
There are a number of data mining tasks such as
Classification: Consists of examining the characteristics of a new object and assigning it to a predefined and well-characterized class.

Estimation: The classification relates to discrete events (for example the patient was hospitalized or not). The estimate is based on continuous variables (for example the length of hospitalization).

Clustering: We segment a heterogeneous population into a number of more homogeneous subgroups (clusters). In this case, the classes are not predefined.

Prediction: This function is close to classification or estimation, but the observations are classified according to future behavior or estimated value. The model, built on sample data and applied to new data, predicts future behavior.

Association Rules: The Association Rules consist of grouping the elements that are naturally together. The most appropriate technique for grouping these similarities is the analysis of the kitchen basket [22].

Optimization: Optimization is the problem of optimizing one or more system parameters according to a set of constraints. To solve many problems, it is common for each potential solution to include an evaluation function. The goal of optimization is to maximize or minimize this function.

Association: Association is the task of looking for relationships or dependencies between several characteristics of an individual.

Description: Sometimes the goal of the excavation is simply to describe what is happening on a complicated database by explaining the existing relationships in the data in order to first understand as best as possible the individuals, products, and processes present in this base. The most suitable technique for the description is the association rules.

1.5.2 Data Mining Process

It is very important to understand that data mining is not just about discovering patterns in a dataset. It is only one-step in a whole process followed by scientists, engineers, or anyone else who seeks to extract knowledge from data. In 1996, a group of analysts defined data mining as a process made up of five steps under the CRISP-DM standard (Cross-Industry Standard Process for Data Mining) as shown schematically below (Figure 1.1):

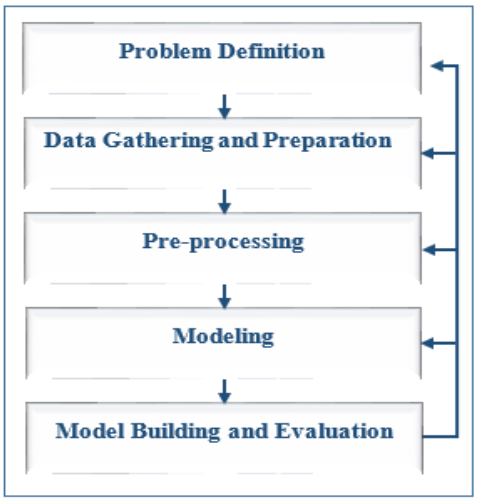

Figure 1.1: The Data Mining Process

Problem Definition: This phase focuses on understanding the problem. Once the problem is specified, we can prepare the data necessary for the exploration, and develop a preliminary implementation plan.

Data Gathering and Preparation : From the definition of the problem, and the objectives of the data mining, we can get an idea of which data should be used and which does not always have the same format and structure.

Pre-processing : The preprocessing should prepare aggregate information about the data for the following steps ; because the data can contain several types of anomalies or errors in this case it is necessary to replace these data or to eliminate their records or they can be inconsistent, that is outside the allowed intervals, they must be excluded or normalized. Preprocessing also includes data reduction, which helps reduce the number of attributes to speed up calculations and represent the data in an optimal format for exploration. Once the data has been collected, cleaned, and preprocessed, it is called a data warehouse.

Modeling : In this step, one must choose the right technique to extract knowledge from the data. Techniques such as neural networks, decision trees, Bayesian networks, clustering, etc... are used.

Model Building and Evaluation : Usually, the goal of data mining is to aid decision making by providing understandable models to users. This is the final phase of the knowledge discovery process where the discovered knowledge is presented to users in visual forms.

1.5.3 Data Mining Techniques

For database discovery, many data mining techniques can use to turn raw data into actionable insights. Such as classification, clustering, prediction, regression, time-series analysis, association, summarization neural networks, association rules, decision trees, genetic algorithm, nearest neighbor method, etc.

Associations Rules Mining: Association generally involves finding results of frequent sets of items among large data sets. This type of discovery helps companies make certain decisions, such as catalog design, cross marketing, and customer behavior analysis. Association rule algorithms must be able to generate rules with confidence values less than one. However, the number of possible association rules for a given dataset is usually very large and a large part of the rules are generally low (or even zero). [88].

Classification: Classification is the most commonly used data mining technique, which uses a set of pre-classified examples to build a model for classifying the population of records in general. This approach frequently uses classification algorithms based on a decision tree or a neural network. The process of classifying data involves learning and classification. In training, the training data is analyzed by a classification algorithm. In Classification, test data is used to estimate the accuracy of classification rules. If the precision is acceptable, the rules can be applied to the new data tuples. The classifier training algorithm uses these pre-classified examples to determine the set of parameters required for appropriate discrimination. The algorithm then encodes these parameters in a model called a classifier. [88]

Decision trees: The decision tree is a tree-like structure that represents sets of decisions. These decisions generate rules for classifying a data set. Specific decision tree methods include Classification and Regression Trees and Automatic Chi Square Interaction Detection. [88]

Clustering Techniques: Clustering can be thought of as identifying classes of similar objects. Using Clustering techniques, we can further identify dense and sparse regions in model space and discover the overall distribution pattern and correlations between data attributes. The classification approach can also be used for efficient means of distinguishing groups or classes of objects, but it becomes expensive and Clustering can be used as a preprocessing approach for the selection and classification of subsets of attributes. [88]

Regression: The Regression analysis can be used to model the relationship be-

tween one or more independent variables and dependent variables. In data mining, independent variables are already known attributes, and response variables are what we want to predict[88].

Neuron networks: The method of neural networks (NN) is one of the most widely used and evaluated tools from biology. A neural network is an abstract model, composed of neurons (elementary units) organized according to architecture and an intercommunication.
Genetic algorithms: Genetic algorithms (GA) are based on the mechanisms of genetics and their function is similar to that of the human genome.
A genetic algorithm is an iterative algorithm. We start from a population of initial potential solutions, arbitrarily chosen. The relative performance of the initial potential solutions is evaluated and based on these performances; a new population of potential solutions is created which will be evaluated. This cycle is repeated until a satisfactory solution is found.

1.5.4 Data mining methods

The methods of data mining can be classified as follows [5] :
Depending on the type of learning used in the excavation methods :

Supervised data mining : Predict a hidden function using training data which are pairs of input/output variables (labels or classes). The output of the method provides for the label of the class of input variables. Classification and prediction are examples of supervised data mining.

Unsupervised data mining : This is the identification of hidden patterns of data without the introduction of training data (ie, input pairs and class labels). Typical examples of unsupervised data mining are clustering and associative rule mining.

Data mining methods can be classified into two main families : Predictive methods and descriptive methods. [56]

Descriptive methods : They allow the current situation to be described, they characterize the general properties of the data in the database, and emphasize the understanding and interpretation of the latter.

Predictive methods : Which, by learning about the past, simulate the future. They use the data with known results to develop models to predict the values of other data.

1.5.5 Data Mining Application

With the performance of current computer systems and the maturity of machine learning methods, Data Mining has become very attractive in many fields of application : medicine, genetics, astronomy, industrial processes, agriculture, or even

customer relationship management. , and industrial production, etc. Companies have implemented these tools to improve their knowledge in order to better serve them and increase their satisfaction and loyalty, to finally increase their profitability. The main economic sectors using these techniques are the financial sector (banks and insurance), telecommunications as well as large distribution companies. In these sectors, which have long been massively computerized, data is available in data warehouses.
We can summarize the most important fields of application of Data Mining in the following areas :

- **Customer Relationship Management (CRM Customer Relationship Management) :** This is the main area where Data Mining has proven its effectiveness. Indeed, in this case, Data Mining makes it possible to increase sales through better knowledge of the customer base. In an increasingly competitive environment, the ability to win over and retain customers relies on detailed knowledge of their needs and their behavior. The objectives of Data Mining analyzes are multiple, such as loyalty, additional and cross-selling, efficiency of the sales force, personalization of the offer, customer contact, customer satisfaction survey, etc.

- **Decision support (Business intelligence)** : This is one of the best factors in increasing productivity. Data Mining is incorporated into this activity to better analyze data, look for factors explaining production defects and their quality, and anticipate possible reactions.

- **Manufacturers:** particularly in production units, control and surveillance, have always used statistical methods and modeling.

- **Scientific and medical research :** Data Mining provides hospitals with solutions and services to enable them to better understand health behavior and the pathologies encountered, namely : Medical diagnosis, inventory of health behaviors, the analysis of health risks, the study of therapeutic treatments, as well as the various studies in hospitals such as genomics, the genetic code, etc...

- **The financial sector (Banks and Insurance) :** Thanks to Data Mining, a financial organization can determine the exact profile of these customers in order to target those with the same profile (mailing). Other applications may be such as the management and calculation of customer risk, claims analysis, assistance with recovery by directing the right approach, the search for Fraud, the search for correlations between financial indicators, the return on investment of equity portfolios.

- **Pharmaceutical and cosmetic laboratories :** Data Mining provides pharmaceutical and cosmetology laboratories with solutions and services to enable them both to better understand their core target and to improve manufacturing processes, to ensure the quality of their products, and assess their marketing potential.

- **Internet :** The use of Data Mining is envisaged in many concrete applications at different stages of the Internet relationship and in particular:
    - Optimize the Internet relationship, by making it personalized, even adaptive.
    - Identify market segments and the attributes of value-added prospects.
    - Identify the key attributes of Internet customers for each product.
    - Select promotion strategies best suited to a consumer segment.
    - Improve targeting of Internet consumers, and those likely to be interested in new products.
    - Test and identify the marketing actions that have the most impact.
    - Identify the best online prospects for a service.
    - Reduce costs and improve the quality of contacts with consumers.
    - Maximize the impact of online advertising. Etc.

1.6 Evaluation measures

Evaluation measures are numerous, and most of them are based on recall and accuracy between 0 and 1. To evaluate the performance of our algorithm, we first need to calculate the Confusion Matrix for each class that provides four essential pieces of information:

The Confusion Matrix: For each test, four essential pieces of information are generated.

True positive (VP): the number of instances assigned to a category properly (instances assigned to their true categories). The number of suspicious gestures and finds it by our system as suspicious.

False positive (FP): the number of instances assigned to a category incorrectly (instances assigned to wrong categories). The number of non-suspicious gestures and finds it by our system as suspect.

False Negative (FN): the number of inappropriately unassigned instances (that should have been assigned to a category but were not). The number of suspicious gestures and finds it by our system as not suspicious.

True Negative (VN): the number of instances not inappropriately assigned to a category (that have not been assigned to a category, and have not been assigned to a category). The number of non-suspicious gestures and finding it by our system as non-suspicious.

Confusion Matrix		Actual Value (as confirmed by experiment)	
		Positive	Negative
Predicted Value (as confirmed by our system)	Positive	TP	FP
	Negative	FN	TN

Table 1.1: Confusion Matrix

Recall (R): Recall (called sensitive in psychology) measures the ability of our system to detect well-classified instances. As shown in the equation (1.6.1).

$$R = \frac{TP}{TP + FN} \tag{1.6.1}$$

Precision (P): Precision is a measure of the ability of an algorithm to return only well-ranked instances. As shown in the equation (2).(1.6.2).

$$P = \frac{TP}{TP + FP} \tag{1.6.2}$$

Accuracy: It enables the accuracy of each algorithm to be calculated. It represents the percentage of instances correctly classified.

$$accuracy = \frac{TP + TN}{TP + TN + FN + FP} 100 \tag{1.6.3}$$

ErrorRate: The ErrorRate represents the percentage of instances incorrectly classified by the algorithm. It is the opposite of the Accuracy.

$$ErrorRate = \frac{FN + FP}{TP + TN + FN + FP} 100 \tag{1.6.4}$$

Entropy(E): Entropy allows us to measure the loss of information from our system during its treatment. The result of this measurement is directly related to the accuracy of our system.

$$E = -P(\log(p)) \tag{1.6.5}$$

F-measure (F): As shown in the equation 6 the f-measure allows to calculate the quality of classification of an algorithm from the recall and the precision.

$$F = \frac{2RP}{R + P} \tag{1.6.6}$$

1.7 Distance measures

Cosine similarity measures: The Cosine Similarity (or cosine measure) allows calculating the similarity between two n-dimensional vectors by determining the

cosine of the angle between them. Let two vectors A and B, the angle $\cos\theta$ is obtained by the scalar product and the norm of the vectors :

$$\cos\theta = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

Euclidean Distance: This is a special case of Minkowski distances that examines the square root of the difference between the square of two pairs of objects.

$$d = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Manhattan distance: This distance is also known as the City block. It is the distance between two points X and Y and is defined as follows:

$$d(x; y) = \sum_{i=1}^{n} |x_i - y_i|$$

Chebyshev Distance: Chebyshev Distance or Tchebychev distance, also called infinie distance is the distance between two points given by the maximum difference between their coordinates on a dimension.

$$d(A; B) = \max_{i \in [0,n]}(|A_i - B_i|)$$

1.8 Big Data

1.8.1 Data Evolution

From the advent of computing to the ubiquity of today's web in everyday life, data has been produced in ever-increasing amounts. (Texts, logs, photos, sounds, videos) of all kinds daily by individuals, businesses, and now also connected objects [112].
Many studies predict exponential growth in data by 2020 and beyond. However, they all agree that the size of the digital universe will double at least every two years, a 50-fold increase between 2010 and 2020. Human and machine-generated data is experiencing a rate growth 10 times greater than that of traditional professional data, while only the data captured has a growth rate of around 50x [63]. Now, the term Big Data refers to this explosion of data.

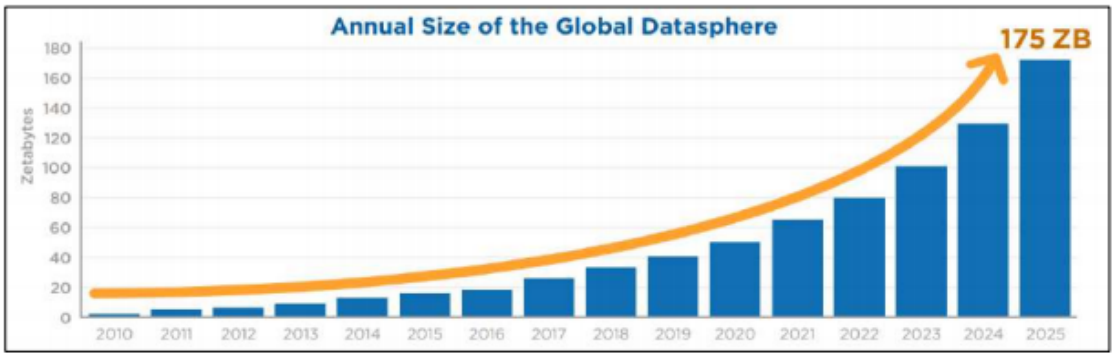

Figure 1.2: Annual evolution of Universal Data [149] [150]

1.8.2 Big Data definitions

There are several definitions of big data (BD) from different perspectives. For example,

according to [116] Big Data is a term used to describe large data at high speed and/or large variety, it requires new technologies and techniques to capture, store and analyze it; it is also used to improve decision-making, provide information and insight, and support and optimize processes.

According to [132] Big Data is a term where the volume of data, the speed of processing, or the representation of the data determine the ability to perform effective analysis using traditional approaches, Big Data requires significant scaling (more nodes) for efficient processing.

On the other hand, [181] defines big data as a term describing the storage and analysis of large and/or complex data sets using a series of techniques, including NoSQL, MapReduce, and Machine Learning.

In short, the term Big Data is a very large set of data, so classic tools (relational or analytical data management engines) or data processing tools (data extraction and transformation) cannot process. This amount of data in the shortest possible time.

The era of Big Data is in effect today because the world is going through a digital and information revolution. Big data researchers, however, remain puzzled as to how to effectively use all of this data. They seek to find a balance between the two equations for the analysis of Big Data; the first equation, if the volume of data increases, then the machine learning algorithms give very precise results, while the second equation, it is hoped that these algorithms will be able to give the results within acceptable time frames. Perhaps because of this inherent conflict, many experts in the field considered that Big Data not only presents one of the greatest challenges, but also one of the most exciting opportunities over the past decade. [58].

1.8.3 Historique

- From 1940 to 1970: (hOctets) the use of Statistics in data analysis: a question is associated with an experimentally refutable hypothesis, a planned experiment with $n \leq 30$ individuals observed on $p \leq 10$ variables, a linear model assumed to be true given rising to a test, a decision, a response.

- During the 1970s : (kBytes) The first computer tools become widespread, data analysis explores, supposedly without a model, larger data. [114]

- During the 1980s : (Mbytes) In Artificial Intelligence, expert systems expire, supplanted by machine learning from neural networks. Statistics deals with non-parametric or functional models.

- During the 1990s : (Gbytes) First paradigm shift. Data is no longer planned, it is previously acquired and based in warehouses for the usual purposes of the business. Decision support enhances them. The mining software brings together in a single environment data management tools, exploratory techniques, and statistical modeling). It is the advent of quantitative marketing and customer relationship management. [114]

- During the 2000s : (TOctets) Second paradigm shift. The number p of variables is exploding (in the order of 104 to 106), especially with biotechnologies. The objective of forecasting quality takes precedence over the reality of the model which has become a "black box". Faced with the scourge of dimension, Machine Learning and Statistics unite in Statistical Learning : selecting models by balancing bias vs. Variance ; jointly minimize approximation errors (bias) and estimation error (variance).

- During the 2010s : (POctets) Third paradigm shift. In industrial applications, e-commerce, geolocation ... it is the number n of individuals that explodes, the databases are overflowing, are structured in clouds (cloud), the means of calculations are grouped (cluster), but the raw power is no longer enough for the greed of the algorithms. A third error term should be taken into account : that of optimization, induced by the limitation of the calculation time or that of the volume/data flow considered. The decision becomes adaptive or sequential [114]. Big Data came as a solution to database overflows.

1.8.4 Sources of Big Data

The data collected, stored, and processed in Big Data can come from different domains and create by several heterogeneous data sources, which generates a mass of data of different types structured and unstructured [86]. Data is everywhere in PDF, doc, email, or post in a social network. This data can be captured and analyzed to better understand people. Here are some different types of data produced in different sectors [115]

- Sensor data : GPS data, Radiofrequency, smart meters, medical devices, telephone call logs, etc.

- Web data : The operation of servers, applications, and networks can capture all kinds of data such as text, images, videos, and anything that may appear on web pages.

- Commercial data : When the cashier swipes the barcode of a purchased product, all data associated with the product is generated. It is enough to think of all the products acquired by all the people, to imagine the volume of data created.

- Financial data : Many financial systems are automated ; they are operated based on predefined rules which automate the processes.

- Input data : This is any data that a human being can enter into a computer, such as name, age, salary, surveys, and so on. This data can be analyzed to understand customer behavior.
- Click Flow Data : Data is generated every time a link is clicked on a website. This data can be analyzed to determine customer behavior and purchasing patterns.
- Game related data : Each movement made in a game can be recorded. This can be useful in understanding user behavior.
- Satellite imagery : Include weather data or satellite surveillance imagery data captured by government departments. A typical example of these systems is Google Earth.
- Scientific data : This includes seismic imagery, atmospheric data, astronomical data, environment, genomics, subatomic physics, and high-energy physics.
- Photographs and video : Relates to security, surveillance, and traffic video.
- Corporate Texts and Internal Mail : Includes all text contained in documents, journals, reports, minutes, balance sheets, survey results, polls, and letters. Business information makes up a large percentage of textual data in the world today.
- Social media data : This data is generate from social media platforms such as YouTube, Facebook, Twitter, Linked In, and Flickr.
- Mobile data : This includes data such as text messages and location information.
- Website Content : This is from any site offering unstructured content, such as YouTube, Flickr, or Instagram [86] [59].

Note at the end of this section that there is another category of data qualified as semi-structured which falls between the two categories structured and unstructured. Semi-structured data does not necessarily conform to a predefined fixed structure but can be self-descriptive and defined by simple mark/value pairs.

1.8.5 The context of big data: Volume or Technology

We sometimes talk about big data, but we also speak of concept, phenomenon, or even big data discipline. This term refers to the automated processing of large amounts of data to extract information. For this, new transfer, storage and analysis procedures are used. The term "Big data" has also become synonymous with "data analysis". Big data therefore corresponds not only to considerable mass of data, but also to the technologies, processes and to techniques implemented to manage data on a large scale in order to extract knowledge [55], [118].

We can speak of "Big Data" when:
The volumes to be treated reach sizes "larger" than the current problems : bytes (10^{15} bytes), exabytes (10^{18} bytes) and zetta-bytes (10^{21} bytes)
Existing tools cannot deal with the problem : relational DBMS, search engines [7].
So it becomes necessary to use new procedures and means to facilitate the management of these data and we chose the "BIG DATA"

1.8.6 Data Structure

There are several types of data in Big Data:

Structured data

Data that adheres to a predefined data model and is therefore easy to analyze, access, and capture like relational databases and spreadsheets. For this reason, structured data provides inherent advantages when processing large volumes of information

Semi-Structured data

They are mainly characterized by the fact that they have a non-uniform and implicit structure, which can evolve rapidly like electronic messaging, XML (eXtensible Markup Language). These data are express in formats such as XML and JSON (JavaScript Object Notation), which allow the representation of information in a hierarchical form (i.e. a tree structure) using tags or symbols as separating elements. [151]

Unstructured data

Any data whose shape or structure is unknown is classify as unstructured data such as data from social networks, videos, and images. Besides the huge size, unstructured data poses a lot of processing challenges [34].

1.8.7 Big Data Characteristics

Since the advent of the Internet to this day, we have seen explosive growth in the volume, speed, and variety of data created daily. This data comes from many sources including mobile devices, sensors, personal records, the Internet of Things, government databases, software logs, public social media profiles, business data sets, etc.

In 2001, Gartner offered a three-dimensional view (volume, variety, and velocity) of the challenges and opportunities associated with data growth [35]. In 2012, Gartner updated this report as follows: Big data is high volume, high speed, and/or a wide variety of information resources that require new forms of processing to improve decision-making [54].

The characteristics that define Big Data often called the three Vs: Volume, Variety,

and Velocity, so that:

- Volume: How much data is there?
- Variety: How diverse are the different types of data?
- Velocity: How fast is new data generated?

1.8.7.1 Volume

The first thing everyone thinks about Big Data is its size [6] [193]. In the age of the Internet, especially social networks producing streaming data whose volumes are increasing exponentially [85] [46]. In 2000, eight hundred thousand (800,000) petabytes of data were stored worldwide [194]. We expect this number to grow to thirty to forty (30-40) zettabytes (Zo) by 2020. For example, on Facebook, more than five hundred (500) terabytes (TB) of data are created each day [70]. Twitter alone generates more than seven (7) terabytes (TB) of data every day, and some companies generate terabytes of data every hour of every day [71].

1.8.7.2 Variety

Previously, all data needed by an organization to run its operations were structured data generated within the organization, such as customer transaction data, etc. Today, companies are looking to leverage much more data from a wider variety of sources, both inside and outside the company, such as documents, contracts, and machine data. , sensor data, social media, medical records, emails, etc. However, the problem is that many of this data is unstructured or has a complex structure that is difficult to represent in rows and columns in structured or semi-structured databases [138][166].

1.8.7.3 Velocity

Just like the volume and variety of data, we collect and store, velocity refers to the speed at which data generated and the time required to process it. Alternatively, in another way, it refers to the increasing speed of data generation, processing, and use of that data [141].
Often these characteristics are supplement by a fourth V, veracity: How accurate is the data?

1.8.7.4 Veracity

Refers to the fact that the data must be credible, accurate, complete, and fit for the task. Since big data comes from various sources beyond the control of organizations such as social media. Veracity has become a real problem. False messages or spam are very common, they make trust a major challenge [33] [194].
We can extend this model to the Big Data dimensions over ten Vs: volume, variety, velocity, veracity, value, variability, validity, volatility, viability, and viscosity [100].

1.8.7.5 Value

Is a major factor that all organizations should consider when implementing big data, because the other characteristics of big data don't make sense if you don't

derive business value from that data So, we can say that value helps companies better understand their customers [35] [101].

1.8.7.6 Variability

Variability in the Big Data context refers to different things. One is the number of inconsistencies in the data. Is the data consistent in terms of availability or reporting interval? Does it accurately represent the reported event? [102] Data should be detect by anomalies and outliers detection methods to allow meaningful analysis. Therefore, proper treatment of the variability property increases the utility of Big Data systems [141].

1.8.7.7 Validity

The term refers to the validation of data. That is, check whether the data used is correct and accurate for the intended use so that this data is therefore use to assess the performance of the forecast [60]. Correct input data followed by proper data processing should yield accurate results. With Big Data, you need to be more vigilant about validity.

1.8.7.8 Volatility

Volatility is the nature of sudden, unstable changes, changed inadvertently or anonymously. The volatility of big data refers to the validity period of data and its retention [140].

1.8.7.9 Viability

Viability means that big data has to be active for a very long time. It must be able to grow, evolve and produce more data when needed. We can identify the characteristics and factors most likely to predict outcomes so that the most important point for companies is to generate additive value [100].

1.8.7.10 Viscosity

Viscosity refers to the stability and resistance of the large data stream. Big Data offers a limited perspective by telling certain storytelling. Viscosity measures the resistance to flow in the volume of data. This resistance can come from different data sources, frictions resulting from integration rates and processing required to transform data into information. Technologies for dealing with viscosity include event handling.

1.8.8 Big Data Management

Big Data Management is a new discipline in which data management techniques, tools, and platforms, including storage, pre-processing, processing, and security, can be applied. [192]
The role of data management is to ensure a high level of data quality and help companies cope with the growing amount of data.

1.8.8.1 Data Storage Phase

Storing data in petabytes in a distributed way uses Cloud services, storage consists of three main operations (Clustering, Replication, indexing).

1.8.8.2 Data pre-processing

Before big data analysis, we need to check the quality of the data and repair the data during processing by applying the following steps (data cleaning, transformation, integration, transmission, reduction, and discretization).

1.8.8.3 Data processing

It is the ability to process a large volume of data regardless of the type of structure and location of such data, this processing can be classification or prediction.

1.8.8.4 Security

To secure a large volume of data, several security algorithms have appeared for confidentiality, integrity, availability.

Figure 1.3: Big Data Management [31]

1.8.9 Big Data Architecture

The architecture of a Big Data system consists of the following layers:[168]:

- **Hardware layer (Infrastructure Layer):** perhaps VMware virtual servers, or blade servers ;
- **Storage layer:** data will be stored either in a NoSQL database or directly in the distributed file system or Data warehouse;
- **Management and processing layer:** his layer contains data processing and analysis tools such as MapReduce or Pig;
- **Visualization layer:** for visualization of the treatment result.

1.8.10 Related Technologies

In order to better understand big data, this section will present several fundamental technologies closely related to big data, including cloud computing, IoT, and the Hadoop ecosystem.

1.8.10.1 Cloud Computing

One of the important factors in the emergence of Big Data is cloud computing which has greatly facilitated access to infrastructure. Based on adjustable resources, by identified duration and at a more suitable cost, cloud computing has opened many doors to innovative projects by considerably lowering the cost of the entry ticket to these solutions [112].

1.8.10.2 Internet Of Thing

The IoT is closely linked to the notion of big data, the objects of which generate a large quantity and a variety of data in real-time, known as big data [145]. So, each needs the other to make it useful. There is no IoT without big data, and big data reaches the highest position when used for IoT.

1.8.10.3 Hadoop

Hadoop is an open-source framework providing tools to operate a cluster in order to store and manipulate large volumes of data quickly and optimally, and to harness the computing power of the machines present in this cluster. Hadoop is managed by the Apache Foundation and is licensed under the Apache License 2.0. [95]

1.8.10.4 Nosql

The absence of an obvious relation or possible classification is an important point that announces the difficulty of resorting to well-defined structures such as those of relational databases. This is why the NoSQL category is created, it designates all the data that can be queried thanks to the queries or structured data, and those that are not structured. According to this definition, NoSQL databases differ from SQL databases in that there is no predefined static schema (relational) structuring the data; the qualification of semi-structured which can be perceived as a lack can thus be perceived as a richer or less much flexible functionality [34].

1.8.10.5 MapReduce

MapReduce is a distributed programming model (called a framework) widely used in NoSQL systems and on very large volumes of data. It was invented by "Google" and revealed to the public through the publication MapReduce: 'Simplified Data Processing on Large Clusters' (2004), to respond to its problem of web indexing (billions of pages). The Framework allows processing to be distributed over clusters (clusters) of servers. It is thus faulted tolerant and automates the parallelization of calculations on several machines, the scheduling of program execution, load balancing, and the management of communications between the machines in the cluster. [30]

1.8.11 Big Data challenges and issues

The growth of data produced by businesses, individuals, smartphones, or social networks poses a huge challenge in terms of acquisition, storage, and processing. This voluminous mass of information is sometimes difficult to manage and process, so it becomes necessary to use new technologies such as Cloud Computing and

NoSQL databases.
We list below the different challenges facing Big Data technology:

- Reduction of redundancy and data compression: Data compression after minimizing redundancies, which will reduce the cost without negatively impacting the data value.
- Data life cycle management: In order to avoid system saturation, it is best to delete unnecessary data and keep only what is important.
- Analytical mechanism: process massive heterogeneous data in a limited time.
- Data privacy and security: since Big Data encompasses large amounts of data, it is very difficult for companies to enforce and ensure the security of this data, so they will need the help of professionals in the field which also presents a potential security risk.
- Energy management: control and optimize energy consumption, as the latter continues to improve in storage systems.
- Scalability: Big data analysis system must support current and future data sets. Algorithms must be able to handle ever-expanding data sets. [35]

1.8.12 Big Data Applications

Big Data's fields of application are vast; any industry that generates and processes a large volume of data can be a big data target. These can include web server logs, Internet click stream data, and social media content, text from email, phone call logs, and IoT data.

Organizations from different fields are investing in big data applications to examine large data sets to discover all market trends, customer preferences, and other useful business information. Among these areas, we find:

- The commercial sector: Big Data is processed to extract value-offering opportunities for innovation and competitiveness. This value is achieved by improving decision-making processes, by precisely studying customer satisfaction and product performance, or by personalizing products and services more than ever [113]. The commercial sector is also entrenched in the Internet through e-commerce and online sales sites such as eBay and Amazon which must handle millions of transactions and track user clicks to offer them the best products.
- Web behavior: Web giants such as social networks face an enormous amount of data that they must store, organize and transfer. This data can also be used for analysis to gather user preferences and trends.
- The healthcare sector: By mapping healthcare data to geographic datasets, it is possible to predict a disease that may escalate in specific areas. Based on forecasts, it is easier to develop diagnostic strategies and plan the storage of sera and vaccines.

- Media and Entertainment: Big Data provides actionable information points on millions of viewers. Today's publishing environments are tailoring ads and content to appeal to consumers.
- The government sector: since the government acts in all areas, it, therefore, plays an important role in the innovation of big data applications in all areas. Among the main government areas are:
  - Cybersecurity and intelligence
  - Crime prediction and prevention
  - Weather forecast
  - Tax compliance
  - Traffic optimization
  - Scientific research

1.9 Internet of Things

1.9.1 Ubiquitous computing

Ubiquitous computing (or ubicomp) is described as pervasive computing. Often it is referred to as a new era of computing and is expected to revolutionize the way we live. Ubiquitous computing is the method of enhancing computer use by making many computers available throughout the physical environment, but become less visible to the user. Ubiquitous computing is the core of the Internet of Things, which means incorporating computing and connectivity in all the things around us.

1.9.2 Machine-to-Machine (M2M)

Machine to Machine (M2M) designates the set of Solutions and Technologies allowing tools, machines, automatons, systems, to communicate between them automatically. This expression used more in the professional field and in particular, that of telecommunications generally refers to the interaction of objects between them using technology. M2M communication offers various ubiquitous services and is one of the main enablers of the vision inspired by the Internet of Things (IoT).

1.9.3 Definitions of Internet of Things

There is not yet a standard unified and shared definition of the Internet of Things because the sector is still emerging. We have chosen a few definitions:

"The Internet of Things is a network of networks that enables, through standardized and unified electronic identification systems, and wireless mobile devices, to directly and unambiguously identify digital entities and physical objects and thus power recover, store, transfer, and process, without discontinuity between the physical and virtual worlds, the related data." [19]

The CERP-IoT "Cluster of European Research Projects on the Internet of Things" defines the Internet of Things as:
"A dynamic infrastructure of a global network. This global network has auto-configuration capabilities based on interoperable communication standards and protocols. In this network, physical and virtual objects have identities, physical attributes, virtual personalities, and intelligent interfaces, and they are integrated into the network transparently "[169].

"The Internet of Things is an extension of the current Internet to all objects that can communicate, directly or indirectly, with electronic equipment that is itself connected to the Internet". [182]

With the advent of the Internet of Things, a new dimension has been added to the Internet connection; as shown in Figure 1.4, in addition to the ability to connect anytime and anywhere as it already was; it is now possible to be connected with any object.

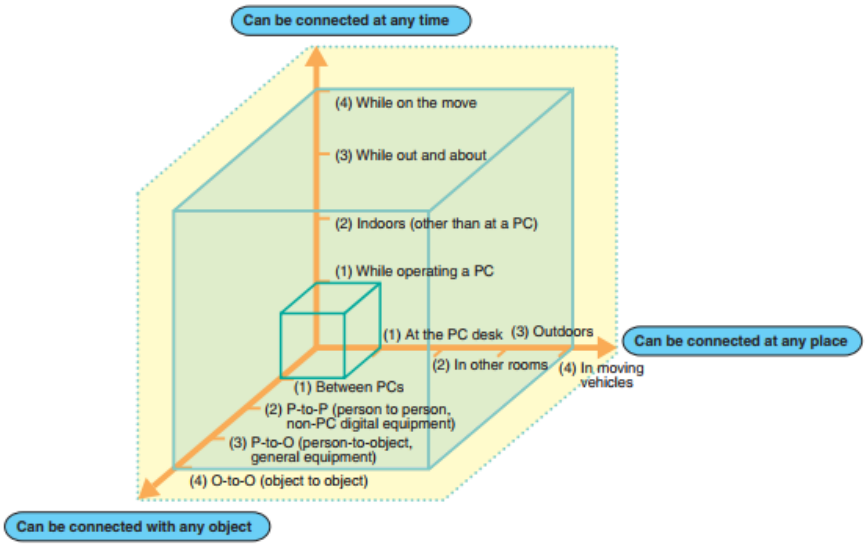

Figure 1.4: The new dimension introduced by IoT [120]

1.9.4 History of The Internet Of Things of Things

The Internet of Things (IoT) was born in 1999 as a result of the convergence of multiple technologies, namely the Internet, wireless communication, embedded systems, microelectronic systems, and nanotechnology.

In this section, we list the most important events on the way to make IoT a reality

- In 1989, Tim Berners-Lee laid the foundations for the World-Wide-Web, the

universal Internet that we know today.

- In 1991, Tim Berners-Lee created The First Web Page
- In 1991, Mark Weiser envisioned the 21st-century computer and introduced ubiquitous computing through his paper titled: The Computer for the 21st Century and advanced the contemporary vision of the Internet of Things.
- In 1994, Steve Mann created WearCam.
- In 1995, the concept of the Internet of Things first appeared in the book by Microsoft founder Bill Gates, The Road Ahead
- In 1998, the creation of Google on September 4
- In 1999, Kevin Ashton (MIT Auto-ID Center Co-Founder and CEO) used the designation Internet of Things for the first time in a presentation at Proctor and Gamble.
- In 2000, the Korean manufacturer LG was the first manufacturer to talk seriously about a household appliance connected to the Internet.
- In 2003, the expression "Internet of Things" appeared at MIT, as part of the work of the Auto-ID lab.
- In 2005, the first report of the world of telecommunications by the ITU, concerning the Internet of Things.
- Since 2011, IPV6 has offered new possibilities for connected objects that have new ranges of available and assignable IP addresses.
- At the end of 2012, there were approximately 8.7 billion connected objects in the world.

Cisco estimates that this number will easily reach 50 billion connected objects in 2020

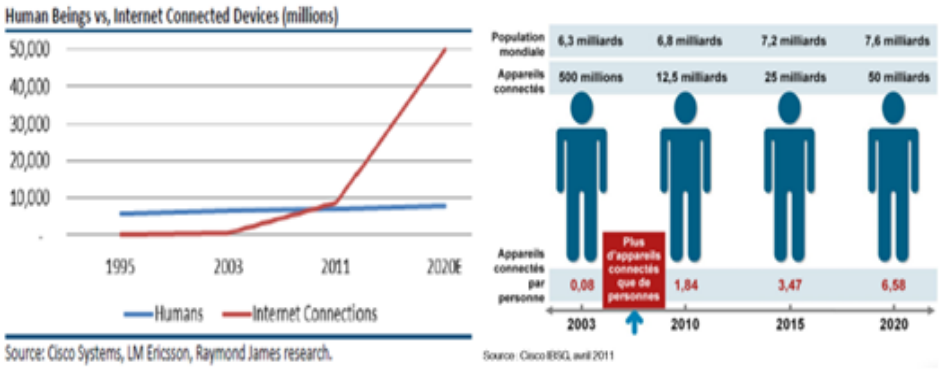

Figure 1.5: Number of connected devices in 2020

1.9.5 Things in The IoT

An object is, above all, a physical entity. In the concept of the Internet of Things, an object refers to any object that can be connected. An object from the physical world (physical object) or from the information world (virtual object), which can be identified and integrated into communication networks.

- **Physical objects:** exist in the physical world and are capable of being detected, actuated and connected. (Examples: industrial robots, goods and electrical equipment, etc.)
- **Virtual objects:** exist in the world of information and can be stored, processed and consulted. (Examples: multimedia content and application software.)

1.9.6 Lifecycle of connected object in the IoT:

In the IoT, smart objects go through three stages: the preparatory phase, the operational phase and the maintenance phase [164]

Figure 1.6: Object life cycle. [164]

The preparatory phase:

deployment of objects (sensors, tags), their configuration with the necessary information, for example identifiers, security keys, etc. [164]

The operational phase:

in the operational phase, the connected object begins to carry out its mission, which differs from one application to another. [164] The maintenance phase: perform updates, solve problems by making possible repairs to objects in the event of failures, for example. It is even possible to outright replace objects and restart again from the preparatory phase. [164]

The maintenance phase:

perform updates, solve problems by making possible repairs to objects in the event of failures, for example. It is even possible to outright replace objects and restart again from the preparatory phase. [164]

1.9.7 Visions of internet of things:

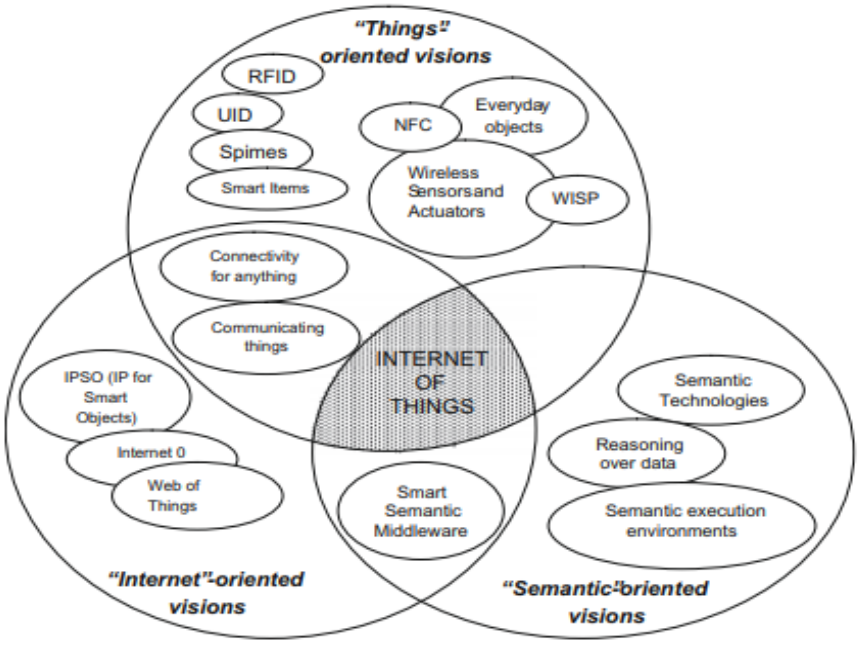

Figure 1.7: Internet of Things" paradigm as a result of the convergence of different visions. [8]

As shown in Figure 1.7, the Internet of Objects emerges from the convergence of three major visions an object-oriented vision, an Internet-oriented vision and a semantic-oriented vision [8] and each of these visions has its key concepts and technologies.

Object-oriented vision

Focuses on physical objects and the systems on board them, with the identification of objects in a unique way and it allows objects to obtain information about their environment by using sensors, for tracking change and process this information. This vision stems mainly from work on networks of objects identified by radio frequency (RFID), near field communication (NFC) and networks of wireless sensors and actuators.

Internet-oriented vision

Suggests that this interconnection should be done specifically through the Internet network, considering that this is already a "network of networks". This vision therefore focuses on studying how the Internet Protocol (IP) can be adapted for embedded systems characterized by low material resources, in particular through new standards.

The semantic-oriented vision :

Breaks away from technical issues specific to physical objects and networks allowing them to communicate, to focus on the representation, organization and storage of data relating to the Internet of Things.

1.9.8 IoT elements:

There are three components of the IoT [72]:

- **Hardware:** composed of sensors, actuators and integrated communication hardware. [72]
- **Middleware:** IT tools for data analysis. [72]
- **Presentation:** understand the visualization and interpretation tools available on different platforms. [72]

In this section, we discuss a few enabling technologies in the previous three components:

- **Radio frequency identification (RFID):** designating technologies that use radio waves to uniquely recognize or identify each object at a greater or lesser distance and collect the data stored at the object level.
- **Wireless Sensor Networks (WSN):** is a sensor network that consists of a large number of intelligent sensors, enabling the collection, processing and analysis of valuable information, gathered in a variety of environments. Sensor data is shared between sensor nodes and sent to a distributed or centralized system for analysis.
- **Addressing schemes:** The ability to uniquely identify "Objects" is critical to the success of IoT. Each element already connected and those which will be connected, must be identified by their identification, location and unique functionalities

- **Data Storage and Analytics:** One of the most important results in this field is the creation of a huge amount of data. Data storage, ownership and expiration become critical issues. It is important to develop artificial intelligence algorithms that could be centralized or distributed as needed

- **Visualization:** It is essential for an IoT application because it allows user interaction with the environment. With recent advancements in touch screen technologies, the use of tablets and smartphones has become very intuitive. So that people can enjoy the IoT revolution, Attractive and easy to understand. The visualization should be created.

1.9.9 Enabling technologies

The Internet of Things (IoT) enables the interconnection of different smart objects via the Internet. Thus, for its operation, several technological systems are necessary. Let us cite a few examples of these technologies. The Internet of Things (IoT) enables the interconnection of different smart objects via the Internet. Thus, for its operation, several technological systems are necessary. Let us cite a few examples of these technologies.

The IoT refers to various technical solutions (RFID, TCP / IP, mobile technologies, etc.) that allow objects to be identified, captured, stored, processed, and transferred in physical environments, but also between physical contexts and virtual universes. [19] Indeed, although there are several technologies used in the operation of the IoT, we focus only on a few that are, according to Han and Zhanghang [89], the key technologies of the IoT. These technologies are: RFID and WSN and are defined below.

Radio Frequency Identification (RFID):

Uses radiofrequency radiation to identify objects and people. RFID technology plays an important role in IoT in solving identification problems. The RFID system is made up of one or more readers and multiple RFID tags that use radiofrequency electromagnetic fields to transfer data attached to an object. Labels contain information stored electronically. Passive tags collect energy from radio waves from a nearby RFID reader. The RFID device serves the same purpose as a bar code or magnetic stripe on the back of a credit card or ATM card; it provides a unique identifier for that object. And, just as a barcode or a magnetic stripe must be scanned to obtain the information, the RFID device must be scanned to retrieve the identifying information. [39]

Sensors

Sensors allow us to learn more about the environment around us. They are important for the Internet of Things because they provide essential information. Sensors are present everywhere around us: in cell phones, in GPS, on traffic lights to make traffic more fluid, etc.

SensorsM2M (Machine to Machine)

Refers to communications between computers, integrated processors, smart sensors, actuators, and mobile devices. The use of M2M communication is increasing rapidly in the scenario M2M has multiple applications in various fields such as healthcare, intelligent robots, cyber transport systems (CTS), manufacturing systems, smart home technologies, and smart grids. Example area network M2M typically includes personal area network technologies, such as ultra-wideband and Bluetooth or local area networks [39].

1.9.10 IoT lifecycle

Connected objects are multiplying and diversifying both on the consumer and professional market. This has generated a new need: that of creating interactions between these objects, beyond their manufacturers or sectors of activity, in order to provide new services and thus break down silos. Automating certain tasks of daily life (e.g, in-home automation) or professional (e.g, in industry, transport, health) will thus become possibly. The IoT platforms are intended to connect these heterogeneous objects and make them communicate with each other.

The IoT ecosystem is quite complex, as it integrates several technologies and areas of expertise. An IoT system generally encompasses both hardware communication protocols, software, cloud, and mobile. Thus, an IoT project requires having a multidisciplinary team.

We can break down an IoT system into 4 distinct functionalities :

1. Collect / Activate: At this stage, we are at the level of the connected object. We are talking about sensors that make it possible to take measurements of the physical environment (e.g. temperature, humidity, noise) and actuators that can act on the environment (e.g. motors to close or open a door). Some objects may have electronic, computer, and network capabilities that allow them to connect directly to the Internet. But generally, having hardware and software constraints (limited autonomy, limited processing capacity, no network stack, etc.), the objects implement low energy / low-speed communication protocols and communicate with the Internet network through a gateway "Gateway".

2. Communicate: This is the step that sends data from the local network to the cloud. We are essentially talking about protocols for transporting data, and we can distinguish two models: The Publish / Subscribe model with MQTT type protocols and the REST model with protocols such as HTTP or CoAP.

3. Execute: This is the data storage and processing stage. At this stage, we often speak of "IoT Platform", which is often a cloud solution capable of connecting several connected objects, storing their data, processing them, analyzing them, and exposing them through various applications. IoT platforms also allow heterogeneous objects to communicate. These platforms are

multiplying nowadays (Amazon, Google, Microsoft, etc.) and there is even talked of "war of IoT platforms".

4. Visualize: This is the step that allows you to expose the services of connected objects through various dedicated applications. A user, through a mobile application, can for example communicate with his objects by consulting their data or by sending actions to his objects.

1.9.11 Protocols in IoT

There are many IoT standards available to make it easier and simpler for application programmers and service providers to work. Various groups have been established to provide protocols, including efforts by the W3C, IETF, EPCglobal, IEEE, and ETSI [10].

The IoT aims to make each system communicate with all the others using common protocols. The large-scale implementation of the IoT concept appears to depend largely on the standardization of communication between objects known as M2M [61].

- At the link-layer level, the IEEE 802.15.4 standard is more suitable than Ethernet for difficult industrial environments.
- At the network level, the 6loWPan standard has succeeded in adapting the Ipv6 protocol to wireless communications between nodes at very low consumption.
- At the routing level, in 2011, the IETF published the RPL standard.
- At the level of the application layer, the CoAP protocol attempts to adapt HTTP, which is much too greedy to the constraints of low-power communication between nodes.
- **Constrained Application Protocol:** Constrained Application Protocol (CoAP) is an application layer protocol for IoT applications, It defines a web transfer protocol based on the HTTP functionality is bound to UDP (not TCP) by default, which makes it more suitable for IoT applications. In addition, CoAP is modifying some HTTP functionality to meet IoT requirements such as low power consumption and operation in the presence of loss and noisy links. CoAP was designed based on REST, which represents an easier way to exchange data between clients and servers over HTTP [10]. REST can be thought of as a cache able connection protocol that is based on the stateless server less architecture. It is used in social and mobile networking applications and eliminates ambiguity using the HTTP get, post, put, and delete methods. It allows clients and servers to expose and consume web services like Simple Object Access Protocol (SOAP) but more simply using Uniform Resource Identifiers (URIs). CoAP aims to enable tiny devices with low power, computing, and communication capabilities to

use RESTful interactions. With CoAP, the interactions between web services of the Internet of PCs and the Internet of Things become much easier to achieve, a fairly light application gateway (correspondence between the REST and CoAP commands) takes care of the adaptation of one world to another [10] [61].

- **Message Queue Telemetry Transport:** Message Queue Telemetry Transport (MQTT)represents an ideal messaging protocol for IoT and M2M communications. It aims to connect devices and integrated networks to applications and middle ware. MQTT uses to publish/subscribe model to provide transition flexibility and ease of implementation. It is suitable for devices with limited resources that use weak links or low bandwidth. MQTT is built on top of the TCP protocol. It consists of three components; subscribers, publishers, and brokers. Many applications use MQTT such as healthcare, monitoring, energy meter, and Facebook notification. Therefore, the MQTT protocol allows small, low power, low memory devices to be routed in vulnerable areas and low bandwidth networks [10].

- **Extensible messaging and presence protocol:** Extensible messaging and presence protocol (XMPP)is an IETF instant messaging (IM) standard that is used for multiparty conversations, voice and video calls, and telepresence. It allows users to communicate with each other by sending instant messages over the Internet regardless of the operating system they are using. XMPP allows instant messaging applications to access authentication, access control, privacy metering, hop-by-hop encryption, and compatibility with other protocols. XMPP's many features make it one of the preferred protocols for most instant messaging applications and relevant to the IoT. It operates on a variety of internet-based platforms in a decentralized fashion. XMPP is secure and allows new applications to be added on top of basic protocols [10].

- **Advanced message Queue protocol:** Advanced message Queue protocol (AMQP)is an open standard application layer protocol for IoT focusing on message-driven environments. It requires a secure transport protocol. Like TCP to exchange messages. It supports reliable communication via message delivery guarantee primitives, by defining a protocol at the wire level, AMQP implementations can interoperate with each other. Communications are handled by two main components: Message exchanges and queues. Exchanges are used to route messages to the appropriate queues.
  Routing between exchanges and message queues is based on certain predefined rules and conditions. Messages can be stored in the queues and then sent to the receiver later. AMQP also supports the publish/subscribe communication model [10].

1.9.12 Communication Paradigms For IoT

With the advent of the Internet of Things, new paradigms of communication are emerging, such as human-human, human-thing, and thing-thing (also known as

machine-to-machine (M2M)) [174]

Human-to-Things: Human-to-Things (H2T) communications are very common in some Internet of Things applications. The user can query objects connected to the Internet at any time via their smartphone (or other connected devices). This type of interaction is characterized by a strong material and technological heterogeneity because on the user side we generally use equipment much more powerful than the sensors constrained on the side of the object requested in the IoT. However, heterogeneity in all its forms must be addressed effectively.

Human-to-Human: Human-to-Human (H2H) Communication: humans communicate with each other using attached devices.

Things-to-Things: Things to Things communications (T2T) or Object-to-Object are also called machine-to-machine or M2M (Machine-to-Machine). It means automatic and autonomous inter-machine communications without human intervention. Recall that M2M communications form the basis of pervasive computing which is part of the set of principles and concepts of the Internet of the future. In fact, the intelligent inter-object interactions in the IoT are often homogeneous, at least in terms of the constraints where we find sensors that can use different transmission technologies but which observe the same limitations in terms of resources and which have the same vulnerabilities.

1.9.13 Architecture of The Internet of Things

Internet of things could define as the interconnection between heterogeneous and ubiquitous objects through the Internet, the network characteristics on the Internet of Things (large-scale networks, more dynamic and more heterogeneous) increase the complexity in the design and provision of advanced services, making the traditional approaches inefficient or even inappropriate, so there is a critical need for a flexible layered architecture. There is no single consensus on architecture for IoT, which is agreed upon universally [135]

application layer	IoT application
	Application support
Transmission layer	Local and wide area network.
	core network
	access network
perception layer	network of perception
	nodes of perception

Figure 1.8: IoT Architecture [135]

The perception layer: The perception layer (also called the 'device layer', 'sensory layer', 'recognition layer') which is the lowest layer of the IoT architecture is responsible for capturing real-world information and representing it in a digital format. It includes the technologies used for detection (collection of environmental data), identification (identification of objects), activation (realization of detected data), and communication (establishment of connectivity between heterogeneous smart devices) with minimal human interaction. Depending on the functionalities it provides, this layer can be divided into two sub-layers: the nodes of perception (or sensory nodes) and the network of perception (as a network of sensors).

Transmission layer: The transmission layer (also called the "transport layer" or "network layer") is responsible for transmitting the data collected by the perception nodes to the information processing unit (or high-level decision-making units) through a network or an interconnection of networks. This layer then allows the integration of a variety of heterogeneous networks, technologies, and protocols.
This layer can be divided into three sub-layers: access network, core network, and local and wide area network.

The application layer: It is the highest layer of the IoT architecture visible to the end-user. The purpose of the application layer is to manage and deliver global applications based on the information collected by the perception layer. It provides end-users with access to personalized services over the network, according to their needs, using various mobile devices and terminal equipment.

This layer can be divided into two sub-layers: Application Support Layer and IoT Applications.

The architecture can be extended to a fourth layer called the middleware layer [74] between the application layer and the other two layers. This layer serves as an interface between the hardware layer and the applications. It includes quite complicated features for device management and deals with data aggregation, analysis and filtering, and service access control. The middleware layer also helps conceal the complexity of network operating mechanisms and makes it easier for designers to develop applications.

1.9.14 Iot Security

The IoT is a technology characterized by high ubiquity in the physical world and omnipresence around its users. The various potential applications of IoT, the heterogeneity of its enabling technologies, and its strong human and socio-economic dimension make its security a difficult and complex subject. In addition to the security concerns of the technologies that will constitute it, the IoT accentuates the security concerns of the people who will use it and gives rise to new problems related to the security of the systems under its control. Security and privacy in the

IoT can be approached from three complementary angles that reflect its technological, human, and systemic dimensions. The protection of technology primarily concerns the security of data, communications, and network infrastructures. This protection is necessary to thwart classic and future attacks on the integrity, authenticity, and confidentiality of data, as well as attacks on network infrastructures and their functionalities. The protection of individuals will concern the protection of the privacy of users that requires, in addition to technological solutions, appropriate regulation that establishes responsibilities in the event of disputes. The protection of interconnected systems and hosting IoT objects will concern the protection of the objects themselves delivered to these systems and the processes they will control [38].

1.9.15 Challenges In the IoT

Developing a successful IoT application is not yet an easy task, due to the multiple challenges. These issues include : data security, protection of user privacy, resource limitations, heterogeneity, virtualization, mobility, interoperability, and reliability

- **Data security:** Data security is a fundamental issue when talking about the Internet in general and the Internet of Things in particular since all these objects operate in a network so there are significant security risks, especially about confidentiality, authenticity, and data integrity. So security is an important point that must be addressed so that the Internet of Things can develop Without fear of the confidentiality of personal data

- **Protection of user privacy :** With the development of the Internet of Things, more and more objects of daily use will be connected. Many of these objects have owners and users and the data detected relating to the objects may contain confidential information concerning these objects. owners and users. The IoT must ensure the protection of confidentiality during the transmission, aggregation, storage, exploration, and processing of data.

  Manufacturers of connected objects could use the data emitted by these objects to analyze what people do with them without them being notified. Some experts believe that personal data could be exploited by other people or companies to understand consumer behavior without their knowledge. This is an intrusion into people's privacy because it reveals our habits, our health, our geographic location, and other types of information that are private to us. So privacy must be protected in the IoT.

- **Resource management :** Managing devices and tracking failures, configurations, and performance of a large number of devices are some of the IoT's biggest challenges. Vendors must manage the errors, configuration, performance, and security of interconnected devices.

- **The heterogeneity of the Internet of Things :** the devices used in the IoT are heterogeneous since they have varied capacities and belong to networks of different nature. They can interact with other devices or service platforms via different networks.

- **Interoperability :** This is one of the biggest challenges in the making of the Internet of Things. Interoperability means that heterogeneous devices and protocols must be able to interconnect with each other. This is difficult due to the large number of different platforms used in IoT systems. It is necessary to ensure the interoperability of heterogeneous and distributed systems to allow the provision and use of a wide range of information and services and the possibility of making them cooperate and interact flexibly.

- **Transparency :** Transparency is essential because users need to know which entities are managing their data and how and when those entities are using it.

- Stakeholders such that service providers should be part of this equation,

- **Mobility:** IoT devices must move freely and change their IP address and networks according to their location.

- **Reliability:** the system should function perfectly and provide all of its specifications correctly. In IoT applications, the system must be extremely reliable and fast to collect data and make decisions, and subsequently, bad decisions can lead to catastrophic scenarios.

1.9.16 Internet of Things: applications and future

Connected objects produce large amounts of data, the storage, and processing of which fall within the scope of what is called big data. In logistics, it can be sensors that are used for the traceability of goods for the management of stocks and transport. In the environmental field, it is a question of sensors monitoring air quality, temperature, sound level, condition of a building, etc. In-home automation, IoT covers all communicating household appliances, sensors (thermostat, smoke detectors, presence detectors, etc.), smart meters, and security systems connected to home automation box-type devices. The IoT phenomenon is also very visible in the field of health and well-being with the development of connected watches, connected bracelets, and other sensors monitoring vital constants. According to various projections (cf. Cisco and the Gartner firm), the number of connected objects should greatly increase over the years.

1.9.17 The advantages of the internet of things

We have cited scattered around in different parts of this chapter some advantages of the Internet of Things. In this section, we summarize the main benefits of IoT.

- Ubiquitous access to information for a smarter world and a sophisticated and comfortable lifestyle.

- Improvement of the quality of service and of remote monitoring in various fields of application, namely industrial, medical, etc.

- Improve productivity and customer experience: Connected objects send reports to their manufacturers indicating customer preferences and habits, further helping businesses to act in a proactive and responsive manner that meets customer demand and demands.

- Saving time is another benefit of IoT. Unnecessary trips are therefore replaced by simply browsing the web to order products, check the condition of connected objects and/or places.

- In some applications, the IoT even allows us to rationalize our expenses and save money because we only consume when needed, whether for purchases or energy consumption (necessary for lighting or air conditioning) Or other.

Ability to leverage giant Internet resources for storing and processing data streamed from the IoT.

1.9.18 The Internet of Everything (IoE)

After Cisco Cisco2 [40], the convergence between networks of people, processes, data, and objects, the IoT goes to the Internet of Everything (IoE), (Figure 1.9). It is a multidimensional Internet that combines the fields of IoT and Big data [125].

Figure 1.9: Internet of Everything

1.9.19 Marketing and research projects

Many companies find the Internet of Things market a fertile field to invest in. Google, IBM, and Intel are the three main companies in the field. Each of them is content to adapt quickly to the evolution of the IoT by developing innovative solutions, based on cloud facilities, for the connectivity of objects to the Internet, while ensuring good security at the different levels.

On the road to realizing the vision of the Internet of Things, research groups

bringing together academic or institutional researchers have been created. This to develop advanced products and solutions that meet the needs of the IoT as a global project that has just become a reality, the benefits and returns of which would be just as expected. In this context, Promising projects have been, Butler (uBiquitous, secUre inTernet-of-things with Location and contEx-awaReness) [32] is the first European project, its goal is the development of secure and intelligent applications based on ubiquitous and contextual information systems.

Bulter is interested in scenarios such as smart cities, smart homes, ubiquitous computing-assisted health applications, and smart business applications. Regarding security requirements, the project aims to enable users to manage their distributed profiles, which involves control of data duplication and identities used by distributed applications. The ultimate goal is to implement a system capable of integrating dynamic user data (eg location, behavior) into security protocols. NITRD (Networking and Information Technology Research and Development) [133] started in 2012. This project brings together around ten federal agencies, such as NASA (National Aeronautics and Space Administration) and DARPA (Defense Advanced Research Projects Agency). The objective is to develop intelligent infrastructures for the efficient realization of the different application scenarios of the IoT. IoT European Research Cluster (IERC) [94] represents a wide range of research projects concerning the application of the Internet of Things with European dimensions. Ensuring collaboration and communication between these projects is an essential prerequisite for a competitive industry and a secure and secure deployment of IoT in Europe.A promising research project called, Hydra project [91], co-funded by the European Commission, is used to develop middleware for networked embedded systems that allow developers to create ambient intelligence (AmI) applications based on wireless devices and sensors. Through its unique combination of service-oriented architecture (Service-Oriented Architecture (SOA)). This project considers issues of distributed security and social trust. Such middleware allows developers to incorporate physical heterogeneous devices into their applications by providing easy to use web service interfaces to control any type of device without worrying about the different transmission technologies adopted in the network, such as Bluetooth, ZigBee, and Wifi. Hydra incorporates mechanisms for device and service discovery, a semantic model-oriented architecture, and even P2P (Peer to Peer) communications.

1.10 Conclusion

The Internet of Things as an evolution of the current Internet allows a dramatic improvement in our way of life and how smart objects in our surroundings interact with each other and with their users so that our activities, our goods, our state of health, our expenses, can be controlled effectively and in a ubiquitous manner.

The growing flow of data from connected objects supports the growth of Big Data, which, in turn, facilitates the explosion of uses. By realizing the growing importance that Big Data would be brought to play, companies found themselves

confronted with a host of major concepts, with blurred outlines, which it is now necessary to take advantage of. Algorithms, Smart Data, real-time, connected objects Mastering these new fields rich in promise first requires understanding what Big Data involves from a business point of view. At this stage, we can say that Big Data is a large and complex ecosystem. It requires mastery of various hardware and software technologies (storage, parallelization of processing, virtualization, etc.). Big Data requires skills and expertise in mastering and analyzing data. In this chapter, we have mainly discussed the operation, core technologies as well as featured applications of big data andIoT . We also highlighted the constraints linked to the deployment of the IoT and which should be carefully processed to achieve predefined objectives.

The next chapter offers a journey into the theory and applications of bio-inspired meta-heuristic algorithms where we have compiled a non-exhaustive literature review on a range of natural algorithms.

2 META-HEURISTICS

Contents

2.1 Introduction

Nature is a powerful source of inspiration for solving complex computer problems, as it shows extremely diverse, dynamic, robust, complex, and fascinating phenomena. It always finds the optimal solution to solve its problem and maintains the perfect balance between its components. A new era is opened to the nature-inspired algorithms (Bio-inspired) which are metaheuristics imitating nature to solve optimization problems. In the past decades, many research efforts have been concentrated in this particular area. This chapter presents an overview of nature-inspired algorithms.

2.2 Classification of Resolution Methods

The resolution of different kinds of problems encountered in our daily lives has led researchers to propose methods of resolution and to make great efforts to improve their performance in terms of the computing time required and/or the quality of the proposed solution. Over the years, many methods for solving problems of different complexity have been proposed. Thus, a great variety and remarkable differences in principle, strategy, and performance have been discerned. This variety and difference have made it possible to group the different methods of solving different problems into two main classes: the class of exact methods and the class of approximate methods.

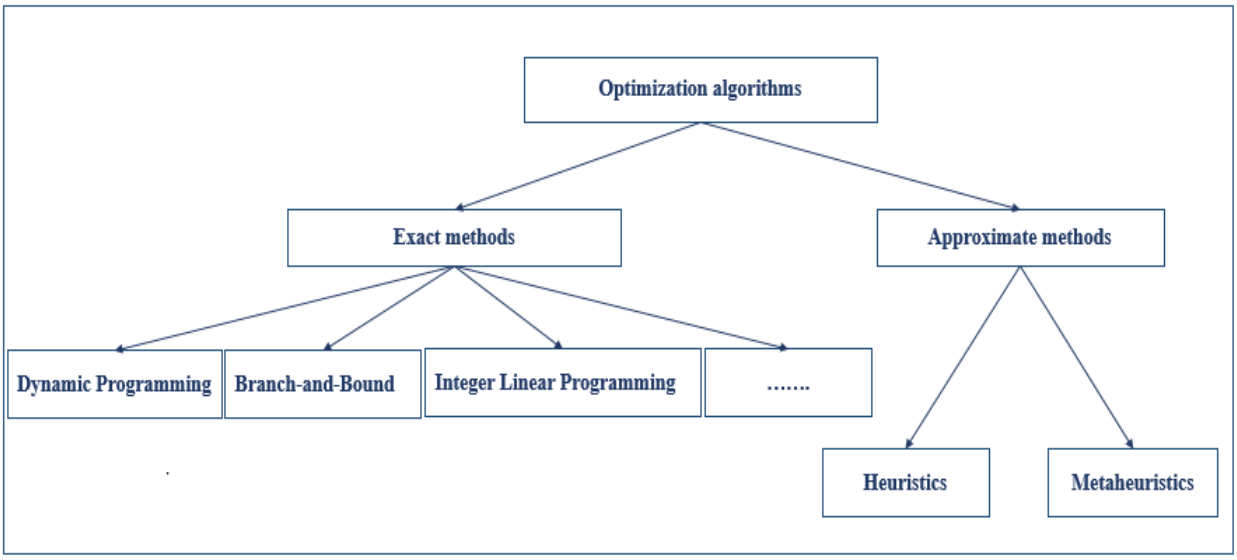

Figure 2.1: Resolution methods classification.

2.2.1 Exact Methods

Usually based on a complete search of the combination space to find an optimal solution.
The principle of exact methods generally consists of enumerating, often implicitly, all the solutions to find the optimal solution.

Advantage Certainty of finding the optimal solution.

Disadvantage Prohibitive execution time The most successful exact algorithms in the literature belong to the four paradigms:

- Separation and evaluation methods
- Methods with backtracking
- Dynamic programming
- Linear programming

2.2.1.1 Branch and Bound method (B-B)

These are exact methods that practice a complete and improved enumeration of solutions. They divide the space of solutions into smaller and smaller subsets, most of which are eliminated by boundary calculations. Applied to NP-complete problems, these methods are of course exponential, but their complexity is much lower than for exhaustive enumeration. They can compensate for the lack of polynomial algorithms for medium-sized problems. For a Combinatorial Optimization Problem (COP), several methods can be invented by separation and evaluation. However, they will have three common components:

- A separation rule, allowing a set of solutions to be partitioned into subsets.
- An evaluation function, allowing the calculation of a bound for a set of solutions.
- A strategy for exploring the search tree.

2.2.1.2 Dynamic programming

Dynamic programming is a general methodology for designing algorithms to efficiently solve certain optimization problems. An optimization problem consists in searching, among a set of solutions to a problem, those that optimize a certain criterion. For example, finding a shorter path to get from one point to another in a transportation network is an optimization problem [11].

2.2.1.3 Linear programming

linear programming is a mathematical technique for optimizing (maximizing or minimizing) a linear objective function under constraints in the form of linear inequalities.It aims at selecting among different actions the one that will most probably achieve the desired objective [49].The term linear programming implies that the solutions to be found must be represented in real variables. If it is necessary to use discrete variables in the modeling of the problem, then it is called integer linear programming (ILP). It is important to know that the latter is much more difficult to solve than linear programming with continuous variables [122].

2.2.2 Approximate methods

The approximate methods aim to find an admissible solution in a reasonable time, but without guaranteeing the optimality of this solution. The main advantage of these methods is that they can be applied to any class of problems, easy or very difficult. Moreover, they have demonstrated their robustness and efficiency in the face of several combinatorial optimization problems.
They encompass two classes: Heuristics and Metaheuristics.

2.2.2.1 Heuristics

Heuristic methods are methods specific to a particular problem. They require knowledge of the domain of the problem being treated. They are rules of thumb that are based on experience and results to improve future research. Several definitions of heuristics have been proposed by different researchers in the literature, among which:

Definition 1 : "A heuristic (heuristic rule, heuristic method) is an estimation rule, a strategy, a trick, a simplification, or any other type of device that considerably limits the search for solutions in important problematic spaces. Heuristics do not guarantee optimal solutions. They do not guarantee a solution at all. All that can be said about a useful heuristic is that it offers solutions that are quite well most of the time." [62].

Definition 2 : "A heuristic method (or just a heuristic) is a method that helps discover the solution to a problem by making plausible, but fallible conjectures of what is the best thing to do. "[62].

Definition 3 : "A heuristic is an estimation rule, strategy, method, or trick used to improve the efficiency of a system that attempts to discover solutions to complex problems." [172].

Definition 4 : "Heuristics are rules of thumb and pieces of knowledge, useful (but not guaranteed) for performing different selections and evaluations." [134].

Definition 5 : "Heuristics are sets of rules of thumb or strategies that work, in effect, like rules of estimation." [173].

Definition 6: "Heuristics are criteria, methods, or principles for deciding which among several other courses of action, promises to be the most effective in achieving a certain goal." [142].

2.2.2.2 Metaheuristic

Metaheuristic methods are general methods, versatile heuristics applicable to a wide range of problems and more particularly the problems of optimization known to be difficult for which there is no efficient classical method, they can build an alternative to heuristic methods when we do not know the heuristic specific to a given problem. Metaheuristics aim to achieve a global optimum generally buried in the middle of many local optimums. They provide good quality solutions in a reasonable time.

In general, metaheuristics are iterative stochastic algorithms, which progress towards a global optimum by evaluating some objective function. There is a growing interest in metaheuristics, justified by the development of machines with enormous computational capacities [21].

Researchers in the literature have proposed several definitions of metaheuristic, among which: " A metaheuristic is an iterative process that subordinates and guides a heuristic, intelligently combining several concepts to explore and exploit the entire search space. Learning strategies are used to structure information to effectively find optimal, or near-optimal, solutions. " [137].

2.3 Classification of metaheuristics

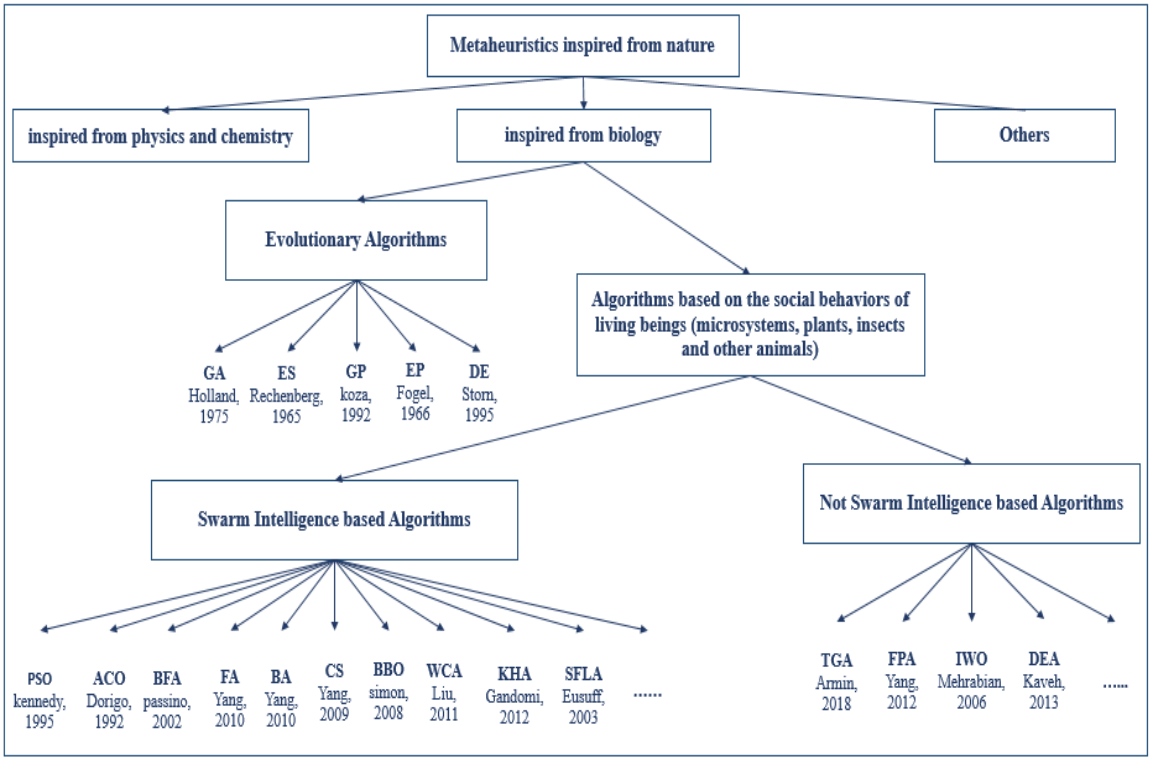

Figure 2.2: Metaheuristic classifications by inspiration sources [15]

2.3.1 Evolutionary algorithms

2.3.1.1 The theory of evolution

During the 1960s and 1970s, with the advent of computers, many attempts to model evolution were made. Several researchers in computer science have independently studied scalable systems as global optimization tools for complex problems where the classical optimization methods learning algorithms or Automatic Algorithm Design are unable to produce satisfactory results. Evolutionary algorithms (EAs) developed in the 1950s, from a family of stochastic and probabilistic algorithms that were inspired by the biological evolution of species to solve NP-complete problems. They are inspired by the theory of the natural selection of species proposed by Darwin in 1860 in his book "the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life".

The vocabulary used is directly modelled on that of the theory of evolution and genetics. We will therefore speak of individuals (potential solutions), population, genes (variables), chromosomes, parents, descendants,

reproduction, crossbreeding, mutations, etc.. And we will constantly rely on analogies with biological phenomena.

2.3.1.2 The general process of evolution algorithms

The majority of algorithms belonging to the EA family follow the same process (problem/solution) detailed in Figure 2.3 The only difference between them lies in the type of solution coding used and the reproduction steps applied. The evolutionary research of the EA consists of the following 5 steps:

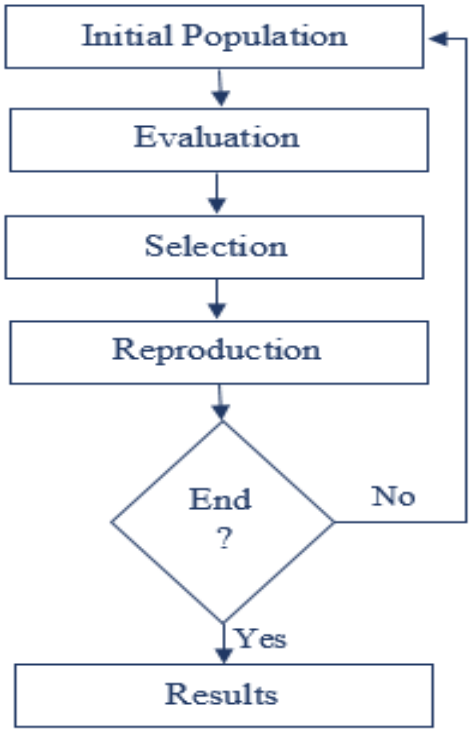

Figure 2.3: The general process of evolutionary algorithms (EA).

Population initialization: This is a first phase of the algorithm, in which the initial population is constructed randomly or through results from other optimization techniques.

Population evaluation: It consists in calculating the cost function value (fitness) of each individual in the population.

The selection of individuals: It consists in choosing the most suitable individuals from the current population in order to train the new generation.

Reproduction of individuals: This is the phase in which a new population is built from the selected individuals, via reproduction operators (such as

Crossover and mutation for GA). Noting that this stage is often stochastic.

AE stopping criteria: This is a test of the efficiency of the algorithm, through a value of the objective function to be reached, the number of iterations or the execution time. The current solution is taken when this test is verified; otherwise, the algorithm goes to the next iteration, starting from the evaluation step.

2.3.1.3 The General Strcture of Evolutionary Algorithms

The General Strcture of EA for a global optimization problem is grouped in the pseudo code :

Algorithm1 Evolutionary algorithms

Input: CP: Crossover probability; MP: Mutation probability; BI: Best individual; SC: stopping criterion.
Code (P (i))
i ← 0
Generate an initial random poplation (P (i))
Evaluate (P (i))
while Ø SC **do**
begin
i ← i + 1
select P (i) from P (i-1)
Crossover(P(i))
Mutation (P(i))
BI ← evaluate(P(i))
P(i) ← Replace(P(i))
end while
end
Output: BI

Figure 2.4: Algorithm1 Evolutionary algorithms.

These algorithms are expensive in computation, difficult to implement and do not guarantee optimal results. In the rest of this section we will discuss each evolutionary algorithm independently in more detail:These algorithms are expensive in computation, difficult to implement and do not guarantee optimal results. In the rest of this section we will discuss each evolutionary algorithm independently in more detail:

2.3.1.4 Genetic Algorithms

The origin of GA
The beginnings of genetic algorithms (GA) can be traced back to the early 1950s when several biologists used computers to simulate living systems [76].

However, in the 1970s John Holland professor at the University of Michigan USA studied evolutionary systems during his PhD. Together with his students, he undertook a vast study that established the foundations of evolutionary systems. In 1975, he was the first to introduce the term "Genetic Algorithm" with a formal model (the canonical genetic algorithm CGA) in his book "Adaptation in Natural and Artificial Systems". In his work, the notion of crossbreeding was paramount. He proposed to add intelligence in a computer program with Crossovers (exchanging genetic material) and mutation (source of genetic diversity). His work was not centered on optimization problems. It is his student De Jong K.A. who in his publications (in 1975, 1976, 1980 and 1992) focused his work on the use of GA in the solution of optimization problems[154].

The inspiration source of GA

Initially created by John Holland [84] at the University of Michigan, USA. Holland began by simulating the adaptation processes of natural systems. Then he came to explore them in optimization and machine learning in the 1980s. Then developed by other researchers such as De Jong [52], Goldberg [73], Davis [53] and Michalewicz [128], GA owe their name to the analogy of the mechanisms of natural evolution of living species and use the two principles neo-Darwinism (the survival of Darwin's best-adapted individuals) and Mendel's genetic recombination. The basic principle is to mimic these two mechanisms to develop a population of solutions (representing individuals) in order to obtain solutions of increasingly better quality.

Summary of GA

GA is iterative population-based stochastic metaheuristics that simulate the natural evolution of individuals in a population.

The objective of GA is to guide individuals within a population to obtain an approximate solution to a given optimization problem.

These GA are applied when a solution is impossible or difficult to approach in a reasonable amount of time. GA uses the principles of natural selection, crossover, and mutation by applying them to a population of potential solutions to a given problem. The solution is approached iteratively.

Traditionally, GA is associated with the use of binary representation. One can find GA that uses other types of representation (Scatter search). A GA applies crossover and mutation operators to promote diversity. GA uses probabilistic selection [21].

To implement a GA, it is necessary to have:

- A genetic representation of the problem, i.e. coding of solutions used in the form of chromosomes.
- A mechanism of generation of the initial population. This mechanism is essential to build a non-homogenous population of individuals.
- A function that evaluates the adaptation of a chromosome to its environment, offering the possibility of comparing individuals. This function is constructed from the criterion that one wishes to optimize. The application of this function to an element of the population gives its

fitness.

- A mode of selection of the chromosomes to be reproduced. This selection is based on reproduction and genetic coding, which stores information describing the individual in the form of genes.
- Crossover and mutation operators allowing to diversify the population over generations and to explore the state space.
- Parameters used by the algorithm: population size, probability of crossover and mutation, the total number of generations.

The general process of Genetic Algorithm

Figure 2.5 is a flow diagram of the Genetic Algorithm, where each step walkthrough of this process will be detailed in the following:

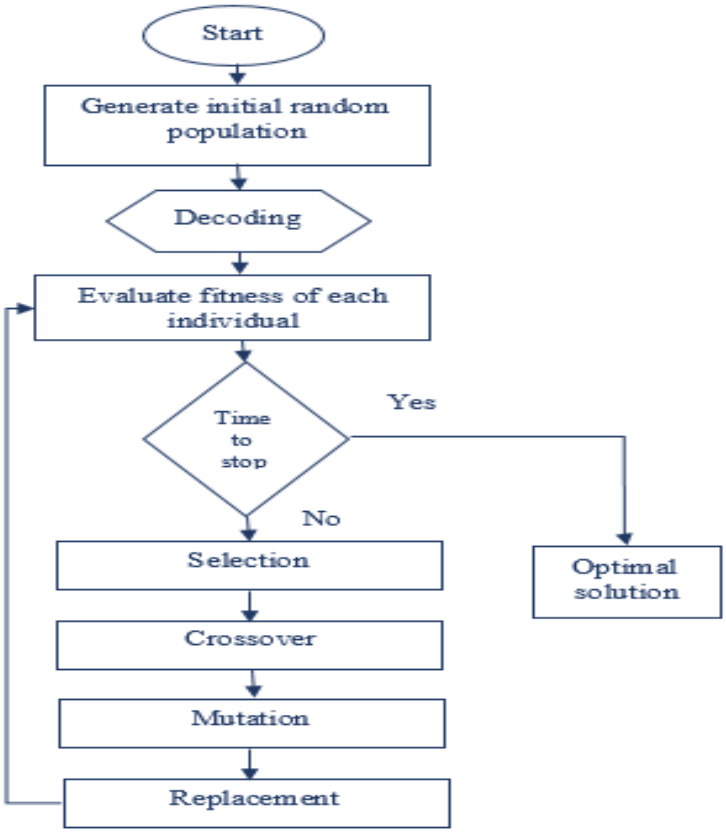

Figure 2.5: General flowchart of the Genetic Algorithm

Initialization:

The first step in the implementation of GA is the random generation of a population of individuals called chromosomes distributed in the research space to provide the GA with varied genetic material. This step may condition the speed and efficiency of the GA.

Coding:
The coding of individuals is a delicate and determining step to facilitate the implementation of the reproduction operators. In the literature there are three types of coding: binary coding, real coding and grey coding [123].

The evaluation function:
The evaluation function called fitness is used to measure the performance of each individual. It quantifies the quality of each chromosome in relation to the problem. It is used to select chromosomes for reproduction. Chromosomes with a good quality are more likely to be selected for reproduction and therefore more likely that the next population will inherit their genetic material. The evaluation function produces the pressure to move the population from the genetic algorithm to better quality individuals. Clearly, the choice of the evaluation function will strongly influence the success of the genetic algorithm.

This function, specific to the problem, is often simple to formulate when there are few parameters to optimize. On the other hand, when there are many parameters, it becomes more difficult to define. In this case, the function becomes a weighted sum of several functions (multi-objective function).

Reproduction operators
These operators play a major role in the success of GA, they define how individuals recombine and arrange themselves during the reproduction phase. There are different operators: Selection operator (Os), Crossover operator (Oc), Mutation operator (Om) and Replacement operator (Or). Ensuring the passage from one generation to another while improving the solutions carried by the individuals in the population.

1. **Selection Operator:** Selection This operation is based on the principle of adaptation of each individual in a population to its environment, following the theory of natural selection introduced by Charles Darwin in 1859. Thus, only individuals with high performance costs will be selected to survive and multiply. There are several methods of selection. The best known are rank selection, roulette wheel selection, and tournament selection.

   Roulette wheel selection: it is inspired by lottery wheels. Each individual in the population is associated with a sector of a wheel and a wheel throw is simulated by a random draw. The angle of the sector is proportional to the quality of the individual it represents. Thus the best individuals have more chance to be selected and to participate in the improvement of the population.

   - Tournament selection: A subset of all individuals suitable for reproduction is selected. These individuals are selected in the following way: a number k of individuals are drawn at random from the

population of individuals (k is a parameter called the tournament size) and to select the one who has the best performance or by the confrontation, each time, of two individuals so that the best one wins. There are different selections per tournament: deterministic or probabilistic. In the case of the deterministic tournament, the best of the K individuals wins the tournament. In the probabilistic case, each individual can be chosen as a winner with a probability proportional to his evaluation function. Some individuals may participate in several tournaments: if they win several times, they will therefore be entitled to be copied several times in the intermediate generation, which will favor the durability of their genes.

- Rank selection: This selection method is divided into two steps[185] [48]. First, one must rank the individuals in ascending (or descending for a maximization problem) order of their quality. Then, a selection procedure similar to the selection on fitness but applied to rank is used.This procedure assigns a probability of selection according to their rank ("Ranking selection").

2. **Crossover operator:** The crossover operator aims to enrich the diversity of the population by combining information between chromosomes. It consists of combining any two individuals (called parents) with a probability (Px) in order to obtain two other individuals (called children), not necessarily better than the parents. The accomplishment of this task needs a Crossover probability (CP) fixed in advance. Afterward, the selection of two-parent individuals a random number $A \in [0; 1]$ will be generated and if $A \leq CP$ a cross will be applied on these two individuals. The Crossover area is chosen randomly but there are several strategies to perform the Crossover.
   - One-point crossover: It divides two selected parent individuals into two parts, with a single cut-off point. The cut-off position is chosen randomly and will be the same for both parents in order to exchange information.
   - Two-point crossover: It divides the two selected parent individuals into 3 parts with two cut-off points.
   - Uniform Crossover: It is based on the same principle of the previous crossover but the crossover point can be multiple and randomly chosen.
3. **Mutation operator:** The role of mutation is to make new genes appear. This operator introduces the diversity necessary to explore the research space and guarantees the diversity of results randomly changing a gene within a chromosome with another. The simplest mutation on a chromosome randomly changes one bit. A chromosome has a probability of mutation of one rate.
4. **Replacement:** This last step of the iterative process is the incorporation of the new solutions into the current population. The new solutions are added to the current population by replacing (total or partial) the

old solutions. Generally, the best solutions replace the worst; the result is an improvement in the population.

The general structure of GA

For a global optimization problem the general structure of GA is grouped together in pseudo code 2:

Algorithm2 Genetic Algorithm

CP: Crossover probability; MP: Mutation probability; SC: stopping criterion; N: P Populatin size; A: real variable; BS: Best solution.
Input: CP, MB, SC, N.
begin
Initialization and Coding(P,N)
Evaluate (P)
while Ø SC **do**
P'← Selection(P)
If Generate-random (A) <CP **then**
P'← Crossover(P)
End if
If Generate-random (A) <MP **then**
P'←Mutation(P)
End if
P ← Replacement (P')
BS← Evaluate (P)
Output: BS

2.3.1.5 Genetic programming

The origin of the GP

The GP was proposed by John Koza [105] for the evolution of programs based on studies carried out by Cramer in 1985 on tree structure for program representation and crossbreeding of parent subtrees [44].

The inspiration source of the GP

Genetic programming is a paradigm allowing the automatic programming of computers by heuristics based on the same principles as genetic algorithms: operations of genetic variation and operations of natural selection. The difference between genetic programming and genetic algorithms lies essentially in the representation of individuals. Indeed, genetic programming consists of making individuals evolve whose structure is similar to computer programs. Genetic programming represents individuals in the form of trees, that is to say, oriented graphs without cycles, in which each node is associated with an elementary operation related to the domain of the problem. Several other representations, such as linear programs and cyclic graphs, have since been used. Genetic programming is particularly adapted to the evolution of complex structures of variable dimensions.

Summary of the GP

The GP allows problems to be solved automatically without the user specifying the form or structure of the solution. At the most abstract level, the GP is a systematic, domain-independent method since it provides computers that can solve problems automatically based on a high-level statement of what needs to be done. Since its inception, GP has attracted the interest of myriads of people around the world.

This technique seeks to use induction to design a program with the principle of evolving a set of computer programs from one generation to another, The specificity of the GP is its research space, a space of programs most often represented in the form of trees structure for coding programs (individuals) to improve the fit between candidate programs and an objective function.

The general process of GP

The basic steps of the GP are shown in Figure 2.6 In a first step, we need a basis to be able to start generating programs (initialization phase). We will thus obtain a certain number of individuals that will allow us to generate future generations. At this moment, we verify if one of the solutions offered by these individuals is satisfactory (Evaluation Block). If no solution is suitable, we will then proceed to a selection of the best ones in order to generate the descendants by different techniques. (Selection and crossover/mutation phase). Finally, these descendants will replace the previous generation by being in turn the parents, and the cycle starts again with the Evaluation block.

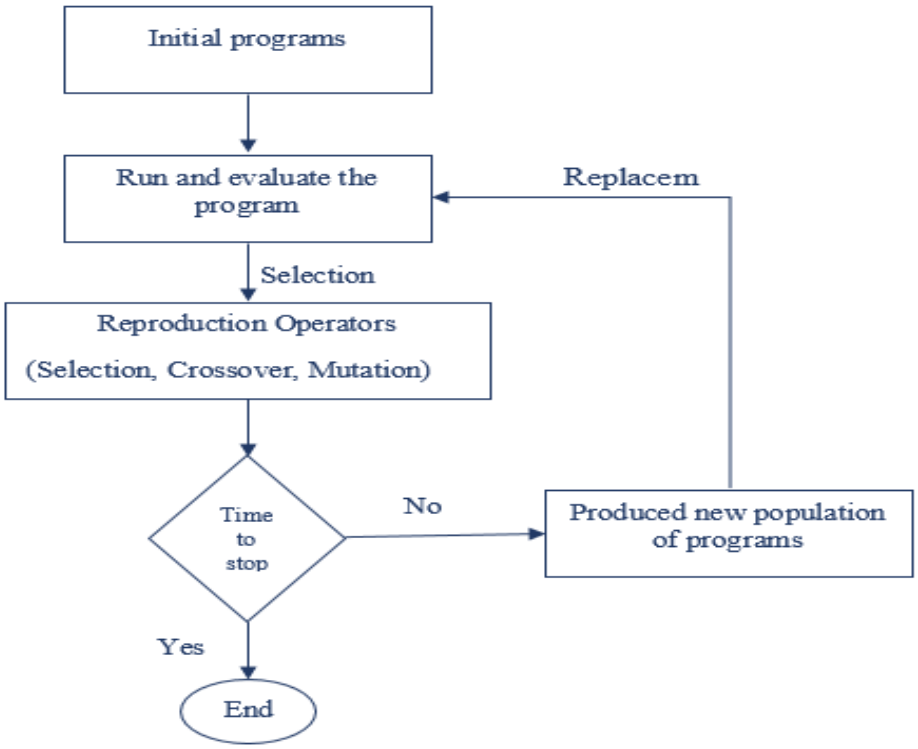

Figure 2.6: General architecture of genetic programming (GP)

1. Initialization phase: individuals are generated randomly: the individuals of the first population. This population is evaluated in order to estimate its adequacy in relation to the problem to be solved.
2. Calculation of the quality of individuals (fitness), then replacement of the old generation by the new generation of individuals.
3. Selection of the best individuals and application of genetic operators (mutation, crossover and mutation, copying...).
4. Return to step two if the stopping criterion is not met (satisfactory solution found, or stopping after a criterion such as the maximum number of generations).

The General Strcture of Genetic programming

The general structure of the GP for a global optimization problem is grouped together in pseudo code 3:

Algorithm3 Genetic programming

BP: Best program, CP: Crossover probability ; MP: Mutation probability; SC: stopping criterion; RM : Random Mutation variable [0-1] ;NG: set of programs (parent and child) which will be built after each generation ; RC: Random Crossover variable [0-1] ; Pi: list of size Npg of the programs initially generated from the set of available primitives ; Npg : size of list Pi and NG ; N : P Populatin size, A: real variable ; PGi: program number i of the program population; Sp: the list of programs to select; SP Gi: program number i from among the programs selected.

Input: CP, MB, Pi.
for each PGi from the Pi **do**
Run and evaluate the program i.
end for
while Ø SC **do**
SP ← selection - by - tournament (PGi)
for each PGi and P Gj randomly selected from Sp **do**
Generate-Randomly (RC).
if RC <CP **then**
NG ← Crossover (SPGi, SPGj)
else NG ← not − Crossover(SPGi , SPGj)
end if
end for
for each P Gi from NG **do**
Generate-Randomly (RM).
if RM <CP **then**
NG ← Mutation (SPGi)
else not- Mutation (SPG)
end if
end for
for each PGi of NG **do**
Evaluate (PGi)
BP ← best (PGi)
end for
end while
Output: BS

2.3.1.6 Evolution Strategies

The origin of ES

Have been developed independently in Germany and the first work was presented in [64] as a method for solving practical optimization problems. It

was later developed by Ingo Rechenberg, Hans-Paul Schwefel and their colleagues in 1964 to solve an aerodynamic design problem [155].

The inspiration source of ES

ES takes into consideration the macro level of the evolutionary species process (phenotype, hereditary, variation) unlike GA which deals with the micro level of genomics (genomics, chromosomes, genes, alleles). This algorithm is commonly applied as a black box for optimization problems with continuous and discrete search spaces.

Summary of the ES

ES is a paradigm particularly well suited for non-linear optimization tasks. Similar to other EA, genetic operators in ES are applied in a loop until a stopping criterion is reached. An important feature of ES is the use of self-adaptation mechanisms to optimize not only the solutions of the studied problem but also some parameters to mutate these solutions [64].

The general process of Evolutionary Strategy

Evolution strategies represent individuals as a set of characteristics of the potential solution. In general, this set takes the form of a vector of real numbers of fixed dimensions. Devolution strategies apply to a population of parents from which individuals are randomly selected to generate a population of descendants. These are then modified by mutations that consist of adding a randomly generated value according to a parameterized probability density function. The parameters of the probability density function, called the parameters of the strategy, also evolve over time according to the same principles as the parameters characterizing the individuals. To form the new population, individuals are selected from among the best individuals of the offspring, or from among the best individuals of the parents and offspring. Modern devolution strategies can also integrate a multi-scale approach or evolutionary strategies are intertwined.

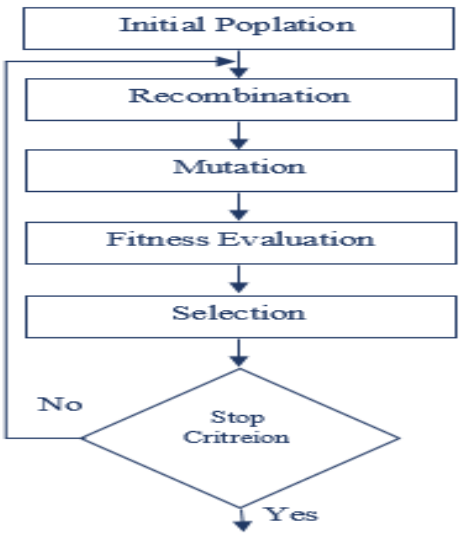

Figure 2.7: Evolutionary Strategies flow chart

The different Evolution Strategies can be introduced by the Schwefel notation:

- $SE(\mu+\lambda)$: designates an Evolution Strategy which generates λ descendants of μ parents and selects the μ best individuals from all descendants and parents to form the new generation.
- The $SE(\mu, \lambda)$ generates λ descendants of μ parents but selects the best μ individuals from the population of descendants only.
- $ES(\mu + 1)$ is similar to Steady State Genetic Algorithms in that only one individual is created in each generation.

The General Strcture of Evolutionary Strategy
The general structure of the ES algorithm for a global optimization problem is detailed in pseudo-code 4:

Algorithm4 Evolutionary Strategy algorithms
Bi: best individual; NG: the list of child individuals ; CP: Crossover probability; MP: Mutation probability; P: population of individuals ;
Inputs: μ: the number of candidate solutions in the parents' generation.
λ: the number of candidate solutions generated from the parents' generation
P ← gnrer - randomly(μpotential solution)
for each individual Ni ∈ P **do**
Evaluate(Ni)
end for
Bi ← referrals-best-individual(P)
while Ø SC **do**
NG ← Ø
for i=0 **to** λ **do**
PRi ← references-parent(P, i)
if A < MP **then**
NG← Crossover (Crossover-discrete, Crossover-intermediate)
end if
if A < CP **then**
NG ← mutation (mutation (strategy-parametes), mutation (dcision--parametes))
end if
end for
Evaluate(NG)
Bi ← referrals-best-individual(NG, M i)
P ← select(P, NG, μ)
end while
Output: Bi

Differential Evolution (DE)

The origin of DE
The differential evolution algorithm was proposed by R. Storn and K. Price in 1997 [170] to solve the Chebyshev polynomial fitting problem.
The inspiration source of Differential Evolution
Differential evolution is a stochastic optimization metaheuristic that was

inspired by genetic algorithms and evolutionary strategies combined with a geometric search technique. Genetic algorithms change the structure of individuals using mutation and Crossover, while evolutionary strategies achieve self-adaptation by geometric manipulation of individuals.

Summary of the DE

Differential evolution is a stochastic optimization metaheuristic that was inspired by genetic algorithms and evolutionary strategies combined with a geometric search technique. Genetic algorithms change the structure of individuals using mutation and crossover, while evolutionary strategies achieve self-adaptation by geometric manipulation of individuals. These ideas were implemented through a simple but powerful vector mutation operation proposed in 1995 by K. Price and R. Storn [170]. Although the differential evolution method was originally designed for continuous and constraint-free optimization problems, its current extensions can handle mixed-variable problems and handle non-linear constraints.

The general process of Differential Evolution

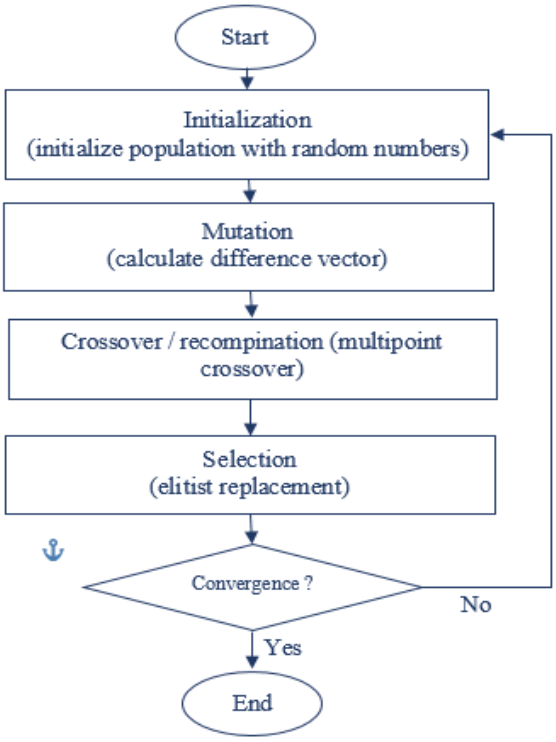

Figure 2.8: Differential Evolution flow chart.

In the DE method, the initial population is generated by uniform random selection over the set of possible values for each variable. The upper and lower bounds of the variables are specified by the user according to the nature of the problem. After initialization, the algorithm performs a series of transformations on the individuals, in a process called evolution

The DE standard uses three techniques (mutation, Crossover and selection) like genetic algorithms. At each generation, the algorithm successively applies these three operations on each vector to produce a test vector.

The General Strcture of Differential Evolution

The general structure of the Differential Evolution algorithm for a global optimization problem is detailed in pseudo-code 5:

Algorithm 5 Differential evolution
Input Po : Population ; Pr : Problrm ; W : Weihting ; C : Crossover
P← initialize_population (Po, Pr)
Evaluate_Population (P)
Bs←Get_Best_Solution(P)
While Ø SC **do**
NP ← Ø
for each individual Pi ∈ P **do**
Ns ← New_Sample(Pi, P,Pr,W,C)
If Crossover (Ns)<= Crossover (Pi) **then** NP←NS
Else NP←Pi
End
End
p←NP
Evaluate_Population (P)
Bs← Get_Best_Solution(P)
End
Output: Bs

2.3.1.7 Evolutionary Programming

Evolutionary programming was introduced by Laurence Fogel in 1966 [66] with the aim of creating Finite State Machines to predict future events on the basis of past observations. Evolutionary programming follows the classical scheme of evolutionary algorithms as follows:

1. A population of n individuals is randomly generated and evaluated;
2. Each individual produces a son by the application of a mutation operator following a normal distribution;
3. The new individuals are evaluated and a new population of size n (best adapted) is stochastically selected among the 2m individuals of the current population ($parents + children$);
4. From the second step on, the selection is repeated until the chosen stopping criterion is valid.

Evolutionary programming shares many similarities with evolutionary strategies: individuals are, a priori, real multidimensional variables and there is no recombination operator. The selection follows a strategy of type $(i + A)$.

2.3.1.8 Synthesis of evolutionary algorithms (EA)

This section presents a comparative analysis of evolutionary algorithms (EA) previously detailed in terms of coding, evolution operators, application domains, decisions and adjustment parameters.

	Steps and parameters	Decisions and remarks	Application areas
GA	– Reproduction Population initialization – operators (selection, Crossover (One-Point Crossover, Two-point crossover, Uniform Crossover), and mutation). – Replacement. – Number of iterations. – Population size – Crossover Probability and Mutation Probability – Chromosome size	– When parameter tuning is static, GA fail to converge to the desired solution within a set number of iterations or fixed execution time. – Contrary to GP and ES it is expensive in terms of computation time. – The operation on dynamic data sets is difficult. – For specific optimization problems, and given the same amount of computing time, simpler optimization algorithms may find better solutions than GA. – GA are not directly suitable for solving constrained optimization problems.	– Workshop production scheduling – Air traffic management – COLORING Geographical MAPS with a minimum number of colors – Knowledge extraction – Commercial traveler problem – Network planning – Image segmentation
GP	– population of initial programs – Tournament selection – Healthy Tree crossing – Mutation (change under a tree of a program) – Replacement – The same parameters as the GA	– Unlike GA and DE, GP avoids premature convergence – Gives good results in the face of global and nonlinear optimization problems. – Contrary to GA the population in GP generates diversity not only in the values of the genes but also in the structure of the individual.	– Automatic Discovery of Reusable subPrograms – classification – Association Rules – Medical knowledge extraction
ES	– Mutation Parameter – Selection (deterministic process). – Discreet Crossover strategy parameters – Intermediate crossover for decision parameters – Probability of discrete crossover – The sampling used.	– Use a lot of parameters compared to GA and GP – Gives a better convergence speed than GA and GP. A very important characteristic of ES is the use of self-adaptive mechanisms to control the application of the mutation	– Image segmentation – decision support system – scheduling problem problems of Mobile Wireless Networks
DE	– Selection. – Crossover. – Mutation (arithmetic combination between individuals). – Replacement – The same parameters as the GA	– - The DE is easy to implement and requires little parameter setting. – - The DE causes rapid convergence. – - DE is generally regarded as a reliable, accurate, robust and fast optimization technique. – - Noise can hinder performance DE due to its greedy nature. – The user has to find the best values for the control parameters for the DE depending on the problem at hand and this is a tedious task.	– Design of Digital Filters – Global optimization – nonlinear problem optimization – Multi-objective optimization

Table 2.1: The analytic comparison between Evolutionary Algorithms (EA).

2.3.2 Swarm intelligence

Swarm intelligence-based algorithms are a branch of algorithms inspired by natural phenomena. These algorithms are generally inspired by the collective behaviors of certain species in solving their problems, to develop metaheuristics allowing the resolution of various optimization problems. According to the authors (Bonabeau et al., 1999) , any attempt to design distributed algorithms or problem-solving devices inspired by the collective behaviour of social insects or other animal societies is part of swarm intelligence .The word "swarm" is generally used to denote a finite set of particles or interactive agents. The birds evolving in groups form swarms whose particles are birds, schools of fish form swarms whose particles are fish, colonies of ants form swarms whose particles are ants, colonies of bees form swarms whose particles are bees, the immune system forms a swarm of particles represented by cells of recognition and protection. Thus, by mimicking the social behavior of particles forming swarms capable of self-organizing, several algorithms have been proposed in recent decades such as Particle swarm optimization, artificial immune systems, ant colony algorithm, Artificial bee colony algorithm, ... etc. [28]

Four principles governing swarm intelligence :

1. Positive feedback: it reinforces the best choices in the system
2. Negative feedback: it allows to ignore and remove bad choices in the system.
3. The random aspect: it allows a good exploration of the solution space, independently of the quality, favoring the principle of diversification. The multiple interactions that allows the construction of the best solutions and choices.

2.3.2.1 Particue Swarm Optemisation (PSO)

The Origin of PSO
Particle Swarm Optimization (PSO) is a stochastic SI algorithm, invented by Russel Eberhart (electrical engineer) and James Kennedy (sociopsychologist) in 1995 [152] for solving difficult continuous optimization problems. This algorithm was originally inspired by the living world. It is based on a model developed by the biologist Craig Reynolds on the social behavior of a group of birds, in the late 1987's.
The algorithm is based on a model developed by the biologist Craig Reynolds on the social behavior of a group of birds, at the end of 1987.
The inspiration source of PSO
PSO comes from an analogy with the collective movements of certain animals evolving in swarms. These animals are characterized by intelligence and limited memory, but despite this handicap, they manage to move around creating complex movements so we speak of collective intelligence or even

a swarm. [24][25] This collective behavior is generated from simple rules of local behavior, where each element of the swarm moves using its personal experience as well as information obtained by its closest neighbors. Besides the spectacular aspect of their movements, this socio-psychological behavior allows them to achieve common objectives such as [37] [50] [65]

- food-searching: In a flock of birds, the food search is carried out, knowing the distance between each element of the flock and the target. The closest bird will be followed by its fellows.
- Energy minimization: Wild geese minimize the energy of movement through a V-shaped formation. The latter allows each bird to take the suction of its predecessor and thus extend the flight distance.
- Defense against predators: A school of fish can avoid a predator by dividing into two groups and then reforming the original school. As the school changes direction, each fish turns and follows the fish in front.

Summary of the PSO
Particle swarm optimization is an evolutionary "meta-heuristic" method that uses a population of candidate solutions to determine an optimal solution to the problem posed. The degree of optimality is measured by a fitness function defined by the user. PSO differs from other methods of evolutionary computation by the fact that members of the population called "particles" are dispersed throughout the problem space. The behavior of the particle swarm must therefore be described from the point of view of a particle. Each of the particles is equipped with : [159]

- A position, that is to say, its coordinates in the definition set.
- A velocity that allows a particle to move. In this way, during the iterations, each particle changes position. It evolves according to its best neighbor, its best position, and its previous position. It is this evolution which makes it possible to approach an optimal particle.
- A neighborhood, that is to say, a set of particles that interact directly on the particle, especially the one with the best criterion. At each instant t, any particle therefore, knows :
- Its best position visited. The value of the calculated criterion and its coordinates are essentially retained.
- the position of the best neighbor of the swarm that corresponds to the optimal scheduling.
- The value of the objective functions (fitness) because it is necessary to compare the value of the criterion given by the current particle with the optimal value.

The general process of PSO
An initial population (swarm) is randomly distributed over the search space, each particle is characterized by a position x and a velocity v. Each particle

is evaluated and saved in memory its best past position and informed its neighbors of its best position and from this, each particle chooses the best of the best positions of which it is aware and modifies its velocity and displacement according to this information.
The general architecture of PSO is shown in Figure 2.9 :

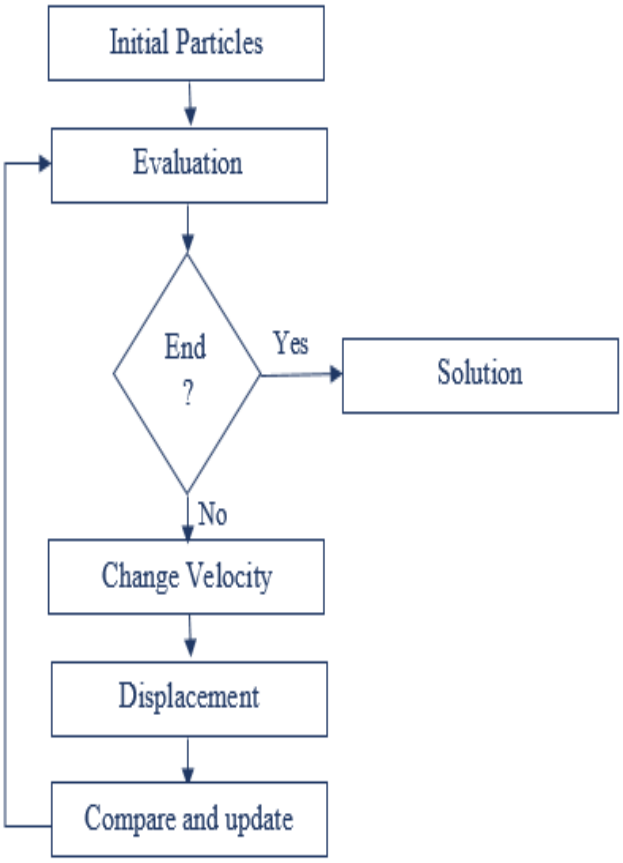

Figure 2.9: General architecture of particle swarm optimization (PSO) algorithm

Each step of the PSO operation is detailed below:

- Initial particles: The initialization of the particles is done based on the random generator of the speed and location of every particle.
- Evaluation: This step allows to evaluate the quality of each solution, by assessing the position of each particle.
- Change Velocity : Iteratively, each particle varies its velocity according to its actual position and the best position of its neighborhood.
- Displacement: The displacement of each particle from one location to an other is done according to its actual position and its velocity of movement.

- We compared the current position of each particle with its better location obtained in the preceding phases.

The General Strcture of Particle swarm optimization algorithm
The general structure of POS is detailed in the pseudo code 6 :

Algorithm6 Particle swarm optimization algorithm (PSO)
P_m : best solution
Inputs : N :the number of particles of the initial population
P ← randomly_generate (N, Vi, Xi)
for each Pi **do**
Pg← Evaluer(pi)
end for
While Not termination-condition **do**
for each particle pi of the population **do**
ChangeVelocity (pi)
Displacement (pi)
Pi_{K+1} ← Compare (Pi_K, Xi_{K+1})
Update (P_{K+1})
end for
Pg_{K+1} ← evaluat(Pi)
P_m← Pg_{K+1}
K ← K + 1
end while
Output: P_i

Ant Colony Optimisation (ACO)

The origin of ACO
The ACO is an Evolutionary Algorithm that was introduced in 1991 by Colorni, Dorigoin his doctoral thesis, and Maniezzo to solve the Travelling salesman problem.
Several extensions of the ACO algorithm such as [75] introduce a reinforcement learning approach, [160] proposes the MAX-MIN Ant System introduces minimum and maximum pheromone trace thresholds, and [47] describe the Ant Colony System (ACS) algorithm.
The inspiration source of ACO
The ACO was inspired by the behavior of ants seeking a path between their colony and a food source. The original idea has since diversified to solve a broader class of problems and several algorithms have emerged, inspired by various aspects of ant behavior. These insects are characterized by the phenomenon of stigmergy introduced in 1959 by Grasse in [123] which refers to the indirect communication between ants through traces of pheromone. Ants first begin by randomly exploring the environment. As they move, ants leave a trace of chemical pheromone on their path. The movement of ants is guided by the smell of pheromone. Ants tend to choose the paths marked out by the strongest pheromone concentration. On the way back,

the amount of pheromone an ant leaves on its path depends on the amount of food it is carrying.

Summary of the ACO

+. The Ant Colony Algorithm (ACO) is one of several swarm intelligence-based algorithms it was introduced in the early 1990s by Dorigo and was inspired by studies of real ant behavior to naturally solve relatively complex problems.

A colony of ants communicate indirectly via dynamic changes in pheromone tracks and construct a solution to a given problem.

The ant colony algorithm is well suited for discrete optimization problems such as quadratic assignment problems, job scheduling, network routing, graph coloring, bioinformatics, and data mining [51] .

The general process of ACO

The general Process of ACO can be summarized in the following steps Figure 2.10 :

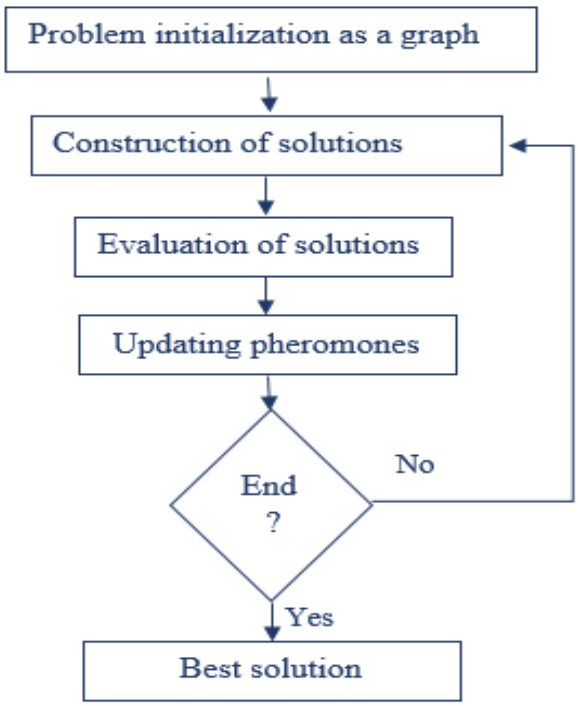

Figure 2.10: the process of the ACO algorithm.

Each step of the ACO is detailed below:

- Algorithm initialization: The solution search space is represented as a graph containing N initial solutions and each solution is a path such that the values of the arcs connecting the nodes are generated initially.
- Construction of solutions: in this step a solution construction process is constructed such that the ants artificial move through the different

nodes of a graph that represents the problem, according to a probabilistic rule in iterative solution construction The probability of movement of each ant artificial.

- Solution evaluation: Solutions are evaluated using a fitness function to designate the amount of pheromone to be produced on each pathway.
- Update pheromones: Updates are made on the pheromone traces once complete solutions have been built. This step also includes evaporation of the pheromone traces to help ants "forget" the bad solutions learned at the beginning of the algorithm initialization.

Procedure of ACO

The pseudo code 7 resumes the global structure of the ACO

Algorithm7 Ant Colony Algorithm (ACO)

```
HS : heuristicSolution
Inputs : S :Size of population,m,P, α, β
ProblemCoding ( representing the solutions as a weighted graph)
HS← CreatingHeuristicSolutions
F(Bbest) ← Evaluate(HS)
Bbest ← HS
Ph ←InitialiserPhromone
while Ø SC do
S←Ø
for i=1 to i=m do
Si←probability( Ph, α, β)
F(Si) ← Evaluate(Si)
If F(Si)<= F(Bbest) then
F(Bbest) ←F(Si)
Bbest←Si
End If
S←Si
End for
PheromoneEvaporation (Ph, P)
for each Si ∈ S do
UpdatePheromone (h, Si, H(Si))
End for
End while
Output: Bbes
```

Bacterial Foraging Optimization (BFO)

The origin of BFO

The Bacterial Foraging Optimization Algorithm was proposed in 2002 by Passino [161]. This algorithm is based on the foraging behavior of groups of bacteria such as E. coli and M. xanthus. More precisely, the bacterial feeding optimization algorithm is inspired by the feeding behavior of Escherichia coli (E. coli) bacteria present in the human intestine. [121]

The inspiration source of BFO

Optimization by bacterial foraging is inspired by the hoof exposed by bacterial foraging behaviors [180]. More precisely, the bacterial foraging optimization algorithm is inspired by the chemotactic behavior of bacteria that tend to gather in nutrient-rich areas through an activity called "Chemotaxis" and represents the key idea of this algorithm.

Summary of the BFO

The Bacterial Foraging Optimization BFO algorithm proposed in 2002 by Passino [143] is based on the feeding behavior of Escherichia coli (E. coli) bacteria present in the human intestine [144]. The theory of exploration for food is based on the hypothesis that the bacteria obtain the maximum energy constituents in a supposedly short time. The basic BFO consists of three main mechanisms: chemotaxis, reproduction and elimination-dispersion[143] [144].

The general process of BFO

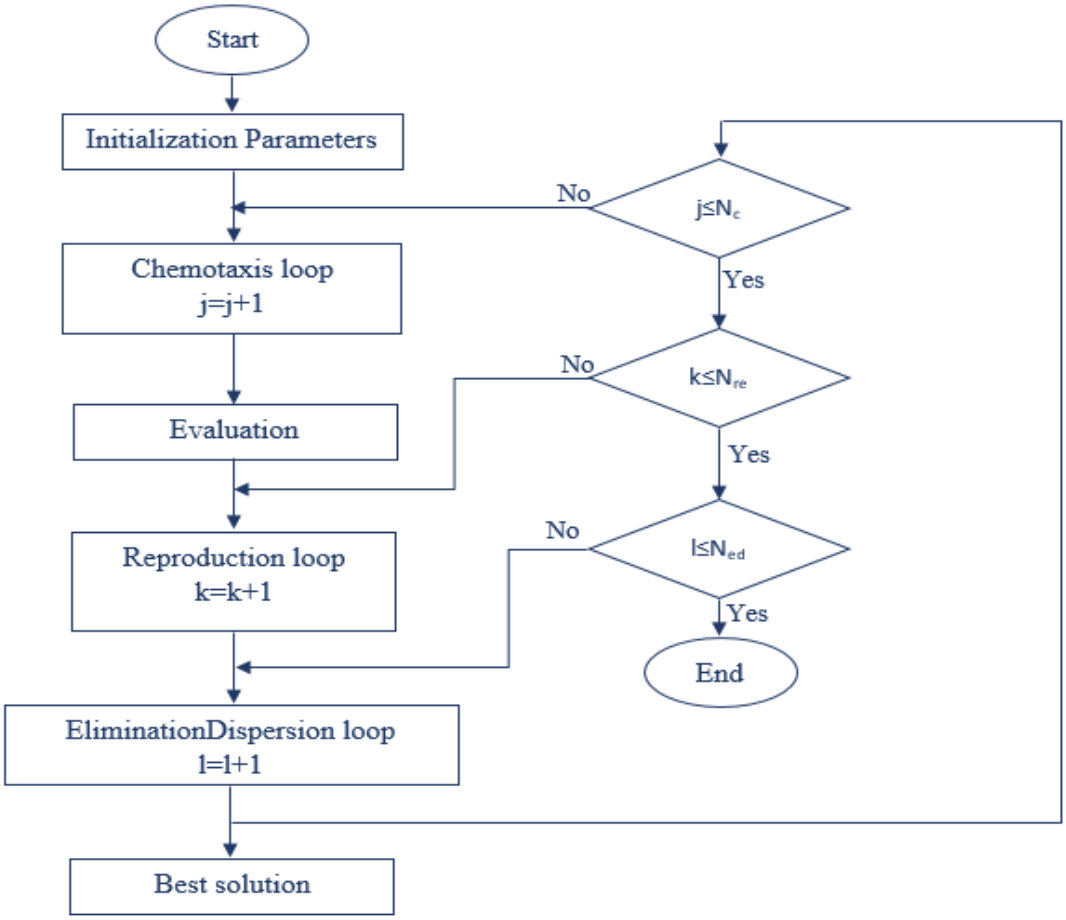

Figure 2.11: the process of the BFO algorithm

- P: Dimension of the search space,
- N: number of bacteria in the population,
- N_c: chemotaxis steps,
- N_{re}: number of reproduction steps,
- N_{ed}: number of events elimination dispersion,

Bacterial foraging optimization algorithm (BFO) is a stochastic optimization algorithm proposed in 2002 by Passino. The algorithm is inspired by the behavior of a certain type of bacteria (Escherichia, coli) in its search for the most nutritious areas.
The evolution of a group of bacteria consists of four phases

1. Chemotaxis, which is the advancement or tumbling phase of the bacteria to the most promising areas.
2. Grouping : the bacterium that finds the best promising area, informs the other bacteria to group together.
3. Reproduction : of the most adapted bacteria (better fitness).
4. Eelimination or dispersal: of all individuals in a bad area, and new random samples are inserted with low probability

Procedure of BFO
The pseudo code 8 resumes the global structure of the BFO

Algorithm8 Bacterial Foraging Optimization Algorithm (BFO)

P_s : best solution ; PS : The population size; N_{ed} : number of elimination and dispersal stages ; N_{re}: number of reproduction steps ; N_c: chemotaxis steps; P_{ed} : probability of a cell being subjected to elimination and dispersion
Inputs : PS, N_{ed}, N_{re}, N_c, P_{ed},
P ← Initiate the position of each bacterium
for i=0 to N_{ed} **do**
for K=0 to N_{re} **do**
for j=0 to N_c **do**
P ← chemotaxis(P)
for i=1 to PS **do**
calculate the function of fitness for each bacterium f(i)
if $P_s \leq$ f(i) **then**
P_s ← f(i)
end if
end for
end for
P ← Reproduction(P)
end for
for each bacterium i in the population P **do**
if A ≤ P_{ed} **then**
P(i)←EliminationDispersal(i)
end if
end for
end for
Output: P_s

2.3.2.2 Bat algorithm

The origin of BA
The Bat Algorithm (BA) is a metaheuristic developed by Xin-She Yang in 2010 [189]. It was based on the echolocation characteristics of bats.

The source of inspiration for BA
The Bat Algorithm is inspired by the echolocation behavior of micro bats, which is based on their ability to imitate sounds that bounce off objects in their path. These echoes will be used to locate and identify prey. The principle of the bat algorithm is to use a swarm of artificial bats to fly randomly through the search space with speed and position with a fixed frequency, to find an optimal solution.
Summary of the BA
A bio-inspired optimization algorithm developed by Xin-She Yang in 2010 [189], the Bat Optimization Technique (BAT) is an approach inspired by the hunting behavior of bats. During their hunt, to avoid obstacles and target their prey, each individual (bat) emits a bi sonar [79] through its environment; the return of the echo allows the identification of various objects in its environment. The studies developed in this field [171] have shown that the resonance of the emitted wave varies from a high value during a prospecting flight to a fairly low value with an increase in frequency when the bat detects and oblique towards a prey. It is intuitively assumed that these animals can differentiate their prey from other nearby obstacles, including nearby bats. Similar to PSOA, the BAT algorithm can be implemented for solving continuous optimization problems where possible solutions can be represented by the geographical positions of the bats.

The general process of the BA

- Initialization: we have as input N bats randomly positioned with a random speed.

- Bat displacement: the random movement of each bat is associated with their speeds and locations by adjusting their first frequencies and updating their speeds and positions the movement of the bats will be performed.

- Intensity representation: in this step Sound value plays an important role in obtaining the solution. So for each bat, a new solution will be generated at a given moment in the local search.

- Evaluation: the position of each artificial bat will be evaluated with the use of a fitness function to reach the best optimal solution.

- Updating the echolocation parameters: if the bat has started to approach the best Solution the intensity value will be decreased.

Procedure of BA
Pseudo code 9 summarizes the general structure of the bat algorithm (BA)

Algorithm9 Bat Algorithm (BA)

PAB : population of artificial bats ;
Inputs : A_i, F_i, N: the number of bats generated, P_s : best solution
PAB ← initialize (N, V_i, X_i)
Evaluate(PAB)
while Ø SC **do**
PAB← displacement (PAB)
PAB← local_search (PAB)
Evaluate(PAB)
Update_echolocation_parameters (A_i,f_i)
P_S← Evaluate(PAB)
end while
Output: P_s

2.3.2.3 Artificial immune systems

Artificial immune systems (AIS) [90] are models of metaheuristics inspired by natural immune systems. Based on these, several models of artificial immune systems have been proposed to deal with many problems: network security [45], character recognition, image alignment, multiple sequence alignment [20] [110]. Thus, we distinguish the clonal selection algorithm [175], the negative selection algorithm, and the immune network algorithm [175].

2.3.2.4 Grouping Cockroaches Classifier (GCC)

GCC is inspired by the natural behavior of cockroaches and the phenomenon of seeking the most attractive and secure place (shelter) for hiding. We used the studies carried out by Bell on the social life and behavior of cockroaches in [17]. We can identify different types of cockroaches in our work, we are interested in the cockroaches that live in apartments, which are fertile and are never isolated. This phenomenon is well detailed in an experiment conducted by French biologists when they met a group of cockroaches in a basin where there's light everywhere, and they built two artificial shelters (shelter is a place with less brightness as shown in the Figure 2.12) using two red circles because cockroaches do not observe the color red as shown in the following Figure (Figures 2.12 and 2.13).

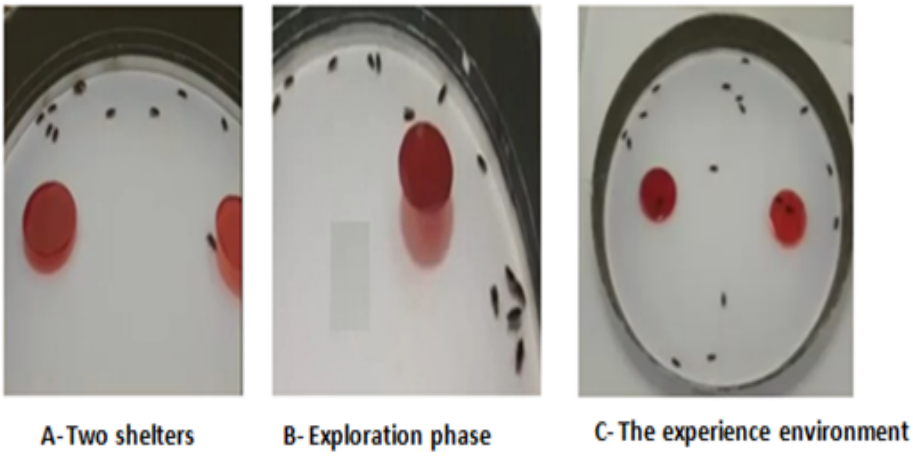

Figure 2.12: Description of the cockroach grouping experiment.[17]

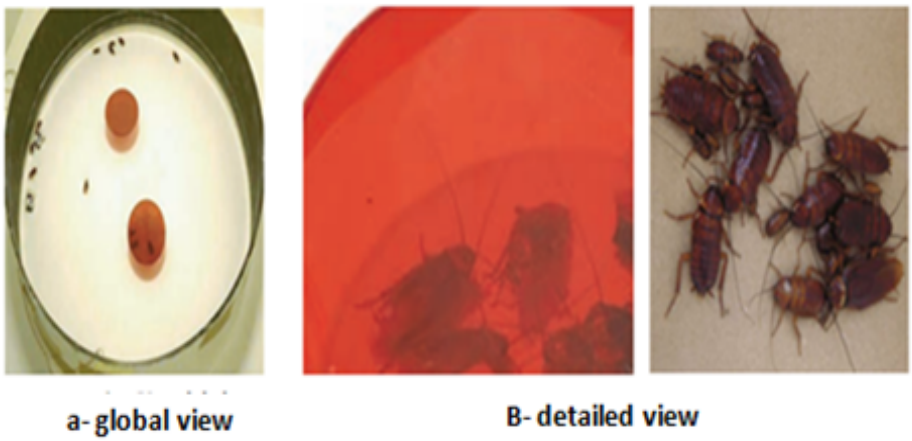

Figure 2.13: The groping of cockroaches under the same place [17]

2.3.2.5 Artificial bee colony algorithm (ABC)

Artificial Bee Colony (ABC) algorithm is a new metaheuristic that has enriched the number of optimization methods based on swarm intelligence. It was proposed in 2005 by Karaboga [108]
The ABC algorithm is based on the natural pattern of honeybee behavior when searching for food. The process of foraging in bees is based on a very efficient movement mechanism. It allows them to draw the attention of other bees in the colony to the food sources found in order to collect various resources. In fact, bees use a set of wriggling dances as a means of communication between them. These dances allow bees to share information about the direction, distance, and amount of nectar with their fellows. Collaboration and collective knowledge of bees from the same colony is based on the exchange of information about the amount of nectar in the food source

found by different members. Studies on bee wriggling dance behavior have shown [146]

- The direction of the bees indicates the direction of the food source relative to the sun
- The intensity of the dance indicates the distance from the food source.
- The duration of the dance indicates the amount of nectar in the food source found.
- In a bee colony optimization algorithm, a source of nectar corresponds to a possible solution to the problem at hand. The artificial bee colony is made up of three types of bees : workers, spectators, and scouts.
- The worker exploits the food source found. She relies on her memory and tries to make changes to her current position (solution) to find a new position (i.e. food source).
- The bee spectator awaits the return of the workers to the dance floor to observe their dances and gather information on the sources of nectar they have found.
- The Scout Bee exploits the research space by initiating a random search for a new food source. A worker bee is assigned to each food source. The size of the colony population is equal to the number of worker bees and the amount of nectar in a food source corresponds to the quality (fitness) of the proposed solution.

2.3.2.6 Synthesis of swarm intelligence algorithms

This section presents a comparative analysis of swarm intelligence algorithms previously detailed in terms of coding, operating steps, application domains, decisions and control parameters.

	Steps and parameters	Decisions and remarks	Application areas
PSO	– Initializing the particles (velocity and location of each particles) – Shifting the speed – Particle movement. – Evaluation. – Comparison and update. – Parameters : – Better positions of every particle. – Better neighborhood position. – Two factors of training.	– Allows you to find the optimal overall solution with a much better calculation efficiency than GA. – PSO is easy to implement and requires the adjustment of a few parameters. – It has a more efficient memory capacity. – PSO has proven stability and convergence in a complex multidimensional or continuous space.	– Monitoring of dynamic systems, – Control of collective power, – Composing music for wireless sensor networks – Unsupervised classification. – Discrete optimization problem.
ACO	– Initializing the Parameter. – Encoding of the problem in the form of a graph. – Construction of heuristic solutions. – Evaluation. – Initializing the Pheromone. – Pheromone update and evaporation. – Parameters – Probability of movement of each ant at each node. – The masses of the arches linking the nodes. – Evaporation rates.	– Solve all types of linear and non-linear problems. – Strong tendency to premature convergence. – One of the most popular algorithms with simple implementation.	– Trade traveler problem, – Vehicle routing problem, – railway scheduling, – quadratic assignment and scheduling problems, – backpack problem.
ABC	– Random initialization of solutions. – Evaluation. – Selection and search of the neighborhood. – Recruitment of bees – Choose the strongest bees from each path. – Construction of the new population. – Parameters : – number of scout bees (n), – the number of sites selected (m), – the number of the most selected sites (e), – the number of bees used for the best e (CIP) sites, – number of bees used for the other sites (me) selected sites (nsp), and the initial size of the room (NGH)	– Performs the neighborhood search and the global search. – Can be used for combinatorial and continuous research. – It is one of the best bio-inspired methods. – It surpasses GA. – Ensures the ranking of solutions	– Multi-objective optimization. – Unsupervised classification. – Bio-informatics. – Ad hoc vehicle network problem.
BFO	– Initialization of positions of bacteria. – Chemotaxis. – Oriented and random movement. – Trialiming. – Reproduction, dispersion and elimination. – Parameters : – Number of bacterias. – Length of the walking unit. Number of chemotactic, elimination and dispersion steps. – Probability of elimination.	– It uses the notion of parallelism and distributive processing. – Ensures global optimization. – Sensitive to parameter initialization. – The reproduction operator is used to improve the convergence speed of the algorithm.	– Extraction of the rules floue, – travelling salesman problem, – backpack problem classification, – scheduling problems, – Multi-objective optimization problem.
BA	– Initialization of the echolocation parameters. – Initial move. – Movement and local search. - Evaluation and Update of echolocation parameters. – Parameters : – A minimum frequency fixe. – A wavelength variable. – Minimum intensity – Average intensity of all bats.	– It can provide very rapid convergence at a very early stage from exploration to exploitation. – It is very effective in dealing with classification problems. – If we allow the algorithm to move to the exploitation stage too quickly, it can lead to stagnation after a certain initial stage.	– Scheduling Problem, Wireless Sensor Networks – unsupervised data classification, – problem of traveler of trade, – backpack problem. – Combinatorial and continuous optimization problem. – Image processing.

Table 2.2: The analytic comparison between Swarm Intelligence algorithms.

2.4 Conclusion

Bio-inspired meta-heuristic algorithms are a new revolution in computer science, The scope of this field is vast; compared to nature, the problems of computer science are only a subset. Unfortunately, we must keep in mind that imitating a biological mechanism does not necessarily bring an advantage, either because the technology cannot match the biology or because the desired goal may be different from that of biological mechanisms. The correct practice of bio-mimicry requires an effort by scientists to extract the principles of biological intelligence from the data and theories provided by biologists in order to translate these principles into functional algorithms and technologies. This chapter offers a journey into the theory and applications of bio-inspired meta-heuristic algorithms where we have compiled a non-exhaustive literature review on a range of natural algorithms.

PART TWO: Contributions

3 BORDER SECURITY AND SURVEILLANCE SYSTEM USING IOT

Contents

3.1 Introduction

The security of borders plays a key role in the assertion of national security, management of lawful immigration, prevention of smuggling, and defense against hostile threats. It is necessary to avoid hostile intrusions, the fluxes of underground immigrants, and the traffic linked by the conductors.

Indeed, borders remain the most visible mark of a state's sovereignty over a territory, and their management of its involvement in protecting its people from threats it defines as such: international terrorism, smuggling, organized crime, irregular migration, and multifaceted trafficking (human beings, drugs, raw materials or ALPC). Border threats differ from country to country, so each of the neighboring countries has developed its own structure for border guard units.

However, they notice that a typical resolution for the surveillance of borders consists in putting towers of observation, of post offices of security and organizing patrols of surveillance to discern possible illegal movements of

persons or vehicles in the area around border, to accomplish a big variety of missions: observation, detection and tells real-time about the slightest changeability the centers of command and control For this we propose a surveillance system to combat terrorism, smuggling, organized crime, and irregular migration, it is designed to ensure the missions of permanent control or temporary border or camp protection, bivouac, sensitive site, convoy route. The system allows continuous operation under "complex and demanding" conditions, without putting lives in danger, and which helps armies and governments to manage changes at the level of threats.

The Internet of Things (IoT) is a developing worldview that enables communication among sensors and electronic gadgets through the Internet to facilitate our lives. IoT utilize smart devices and internet to give creative answers for different difficulties and issues [97]. The structure of our system depends on connections between objects (sensors, drones, and surveillance cameras) that they give capacities to control by the center of command (which can be far from borders) before it is too late. Nature is vast, it is a powerful source of inspiration for solving complex computer problems, she always finds the optimal solution to solve her problem and maintains the perfect balance between its components, an interesting new paradigm known as the bio-inspired consists in analyzing the living world in order to translate biological models of all forms (animals, plants, micro-organisms, ecosystems... etc.)

into technical and algorithmic concepts, many works have been done in the field of bio-inspiration to solve different problems and others are still in progress, the main issue in this work is the creation of a new bio-inspired technique that can enhance security while respecting the privacy of human beings represents. In this work is we use a bio-inspired model based on the style of life of the Cockroaches for the purpose of detect terrorists and non-soldier people by the characteristics of gestures which are intruders of being dangerous (criminals, terrorists etc.) instead of the faces.

The technique is based on the connections between smart objects in which is based on picking up pictures through drones equipped with cameras that are able to connect with smartphones So that it can monitor borders from any place and the use of the characteristics of gestures that are suspected of being dangerous instead of the faces. We apply the classification of gestures human being by the Bio-Inspired technique of Grouping Cockroaches Classifier (GCC) based on the style of life Cockroaches and operate on the phenomenon of seeking the most attractive and secure place (shelter) for hiding for a good detected the gesture of unwanted individuals this algorithm is based on a learning base and classify the gestures of the test base and labels them, each gesture take one of two classes (gestures of border soldiers and gesture of terrorists and non-soldier people), and we apply also a new technology for the presentation of picture (n-grams pixels) to con-

struct a system of control of borders. Our objective is to use drones instead of normal soldiers to cover the space of the borders, detect terrorists hiding their faces, detect people in disguise; react effectively and faster, react at night, or even when the climate is difficult.

3.2 Literature review

In recent years, the security of borders and sensitive sites such as tourist sites has undeniably become a major global issue. The massive influx of refugees, the fight against terrorism, illegal immigration and smuggling, the strengthening of internal security, and the rise of cybercrime are prompting governments to strengthen their border and sensitive site surveillance systems. Governments are investing more and more to ensure their protection through the use of electronic surveillance technology and video surveillance systems. Certain events such as September 11th or the attacks in the London Underground on Thursday 7 July 2005 urged public and political authorities to act and to tighten surveillance in sensitive sites in order to secure humans.

In the conventional solution of controlling border traveling (that the person is eligible to enter the country and does not pose a threat to its citizens or institutions), the border guards have the responsibility for this monitoring takes place manually which are responsible for continuously keeping an eye on the borders. It takes a lot of manpower and assets as the borders are stretched across hundreds of miles and have extreme terrain as well as climatic conditions. With the improvement of document forging techniques, the uses of look-alikes and aliases, as well as the time pressure associated to border control processing, it is not surprising that border control authorities are revising the traditional manual approach and considering the deployment of the most advanced surveillance technologies to facilitate a more efficient and reliable controlling of cross-border travels.

Nowadays, surveillance cameras are ubiquitous everywhere (streets, stores, museums, subway stations, ticket machines, airports, and banks ... etc.) in order to detect malicious and delinquent people (thieves, criminals, and terrorists). Overall, the use of this technology has led to a decrease in crime both both in public and private places, an improvement in incident management as well as an increase in efficiency and speed of intervention [81]. Several works have been done in the field of security surveillance for the border, military, and academic purposes.

In this section, we have defined the different video surveillance systems, then we have presented some advantages and disadvantages and the different components of its systems.

3.2.1 The various threats to borders and sensitive site

With the turbulence and socio-political and institutional instability in countries in the world, insecurity is a real source of concern at borders. the resurgence of armed rebellions across the broad border and the presence of terrorist groups in some countries raise concerns about the possible spread of the extremist threat to borders. In particular, through the trafficking of weapons of war, drug trafficking, and the movement of illegal migrants.

- Drug trafficking and terrorist financing

  Some areas have become assimilated to a so-called lawless zone, have become the hub of transnational trafficking with its various organized crime networks and mainly arms and drug trafficking.

  This drug trafficking finances both organized crime and terrorist groups present in the region as a complement to other sources of funding (hostage-taking, ransoms, migrants, weapons, cigarettes, fuel, livestock, etc.). The violent extremist groups in the border region are increasingly experiencing the adhesion of several drug traffickers who have seen this alliance as a means of protection. Drug trafficking is thus aggravated by the links between drug traffickers and these terrorist groups, which constitute narco-terrorism networks. The financial manna available to these terrorist groups and generated by the trafficking of all kinds is, according to many experts, very difficult to quantify. Jihadist fighters, even if they are not directly involved in trafficking, are often in charge of protecting convoys, securing airstrips, and ensuring fuel supplies.

- Illegal immigration

  Amid the war against banditry and terrorism, there is above all the fight against illegal immigration and the networks feeding the vast traffic, very lucrative for the new barons, that is human trafficking. The organized passage of people is developing exponentially in the illegal border areas. Over time, the care of illegal emigrants has changed hands and has shifted to jihadist smugglers. The evolution of the phenomenon of illegal immigration has become alarming.

- Arms trafficking

  In the last few years, arms trafficking has increased, with the multiplication of so-called low intensity conflicts and wars fuelling the influx of arms, the world has become an uncontrollable and prosperous space for all kinds of trafficking. The circulation of arms is both a consequence and a factor in the development of other trafficking (drugs, migrants, etc.). The exponential development of trafficking mechanically

increases the level of violence in the border region and vice versa.

3.2.2 border and sensitive site video surveillance systems

Video surveillance technology or CCTV (closed-circuit television) has proven its essential role in security operations as it is the witness to everything that happens in a place where a CCTV camera is in operation day and night. CCTV is a television system in which signals are not distributed publicly, but are monitored, primarily for surveillance and security purposes. It is based on one or more video cameras connected in a closed circuit to transmit the images captured by the camera to one or more television monitors or video recorders through coaxial cable trays or wireless communication links [9]. The implementation of a video surveillance system requires the presence of different tools such as monitors, video recorders, cameras, and management and coordination equipment.

CCTV (Closed-Circuit Television) systems meet the need for surveillance and control of the physical perimeter. Historically reserved for the most sensitive sites, they have been democratized and are now part of the urban landscape.

3.2.3 The advantages of video surveillance systems

The majority of people use video surveillance technology to keep their relatives and belongings safe and secure. This technology can have several advantages such as :

1. The deterrence of criminals and feeling of security : Once people place video surveillance systems in their homes or workplaces, they begin to feel safe because offenders are less likely to commit acts of violence or crime if they are aware that they are being monitored. For example, customers entering a store who know they are being videotaped are much less likely to steal. Members of the public who know that CCTV is constantly monitoring a hospital waiting room are less likely to assault staff. This technology can also protect employees. In many jobs, employees are vulnerable to either physical or verbal attacks from the public. Often, in stressful situations, customers may take a company representative to the task. If the employee knows that he/she is being watched by surveillance cameras, he/she will try to remain calm in front of these people. To ensure the process of deterring offenders two elements must exist :
    - The offender must be aware of the presence of surveillance systems and security cameras. For example, in Germany 31 video surveil-

lance cameras have been placed in the city center, only 28% of the respondents are aware of the existence of surveillance cameras.

- The offender must believe that the surveillance systems present are likely to capture him or her in order to nullify the intended act of crime or theft.

2. To assist police officers in their investigations : Having cameras installed in strategic locations is very practical when you need to follow the actions and words of people or during an event. Modern security cameras are not only equipped with high-quality video capabilities but also audio. Clear images coupled with impeccable sound make them more effective than ever at recording a series of events. There are many examples of video footage being recorded to help convict an offender. With a security camera, judicial authorities can see the series of events that actually took place at the crime scene. For example, as shown in figure 3.1, a surveillance camera placed in portland airport made it possible to film the two Al-Qaeda terrorists Mohamed Atta and Abdulaziz Al Omari on September 11, 2011.

Figure 3.1: Image captured by a surveillance camera shows two terrorists who carried out the attacks of September 11, 2011 in the United States. [41]

3. Medical assistance: Video surveillance camera operators can contact emergency and medical services if they see people on the street suffering from illness or injury as a result of criminal activity (such as robbery and assault).
4. Monitor Activities: Monitor the activities of people visiting our homes, as well as the comings and goings at these locations. This is a great way to detect suspicious people and keep an eye on their activities.
5. Decision Support: Images from security cameras can help us to make correct and fair decisions when resolving disputes, both in domestic

and professional scenarios. Whether faced with a situation involving disagreements within a family, or among employees.

6. Event recording : Video surveillance systems allow you to save everything the camera has captured in a specific folder with the date and time of each recording. When people want to know something that happened outside of their homes two weeks ago, then they just need to consult their security files.

3.2.4 Disadvantages of video surveillance systems

Simply, the use of video surveillance in social security, and the application of the policy of identifying suspicious persons through facial recognition can collide with several disadvantages:

1. The inability to identify the identity of offenders without facial recognition : The inability to detect offenders who disguise themselves or hide their faces with masks. For example, we have a surveillance camera located in a house and there is a thief with a hidden face who enters that house to steal, as shown in Figure 3.2. Traditional video surveillance systems cannot recognize the identity of this person because they are based on facial recognition.

Figure 3.2: A thief hiding his face. [77]

Figure 3.2 was taken in a jewelry store in the town of Messina on the Italian island of Sicily, showing a thief wearing a mask who enters the store and steals all the jewelry. The policeman viewed the burglary and analyzed the video recording by a surveillance camera located in

this store. The problem is, it is impossible to detect the identity of this thief because he hides his face

2. The inability to prevent crimes:
   Conventional video surveillance systems are unable to warn people before the crime is committed. For example, terrorists or wanted persons may travel on streets and airports with a false identity by altering their appearance. In this case, these dangerous people cannot be distinguished by surveillance systems based on facial recognition. In another real-life example, as part of the investigation into the double terrorist attack on the Booston area in the United States during the 117th edition of the city marathon on Monday, April 15, the FBI released videos and photos captured by surveillance cameras placed in that city showing two suspicious persons as shown in figure 3.3. The way authorities identified these two suspects shows that we should think about how cameras can help prevent crimes rather than just solve them after the crime is committed.

Figure 3.3: the two suspects named by the FBI as responsible for the terrorist attacks in Boston. [128]

3. The assault on citizens privacy :
   The use of a conventional video surveillance system in offices, streets, commercial establishments, city centers and in front of buildings does not respect the freedom of people who pass nearer the cameras. For example, there are people who use these systems to film reality TV that shows us people's daily lives. There are also people who could use images obtained from these systems to blackmail or for their own personal benefit. There have also been a few cases in the past in companies where employees have objected to being under constant surveillance without their permission. A few have taken legal action against their employ-

ers in relation to this. According to statistics in England every day a person walking in the city center is filmed by 300 cameras. Therefore, in order to protect the privacy of citizens, we have to find a technique to satisfy the interests of all parties, including the person filming and the person being filmed

4. Human problem:
   Conventional video surveillance systems consist of cameras to give a general view of the areas and surfaces monitored to a limited number of operators (the agents who manage the systems and view the events captured by the cameras). These officers have the role of constantly monitoring the monitors and if any suspicious actions occur, they notify security or the police. However, operators cannot stand 24 hours a day in front of the monitors and cannot monitor all the events captured by all the cameras at once. In this case, the agents are not able to prevent incidents or limit the damage. These CCTV systems cannot stop a crime while it is being carried out and they do not alert neighbors or the police like an alarm system does.

3.2.5 State of the Art on border and sensitive site Video Surveillance Systems

Video surveillance has received a lot of attention as a very active research area, this section contains a detailed description of several video surveillance systems that exist in the literature.

3.2.5.1 The intelligent video surveillance system :

Typically, video surveillance systems have two main purposes: to provide a human operator with images to detect and respond to potential threats, and to record evidence for investigative purposes. Unfortunately, the majority of them are unable to support both real-time threat detection and evidence for forensic investigations.

- Real-time threat detection: Human visual attention falls below acceptable levels, even when trained agents (operators) are assigned to the visual inspection task.
- forensic investigation: the problem of searching large collections of video recordings is even more tedious and error-prone for a human investigator.

In order to address the two previous issues, Hampapur and al in [80] proposed an intelligent video surveillance system that can help the human operator in real-time threat detection. This multi-scale analysis system uses a combination of active cameras and a set of models to address the problem of scale variations in visual monitoring applications. It is composed of :

1. Static camera : To cover the entire area of interest and provide a global view. Static camera video is used to detect, track multiple objects in two or three dimensions as well as to extract additional information about the objects e.g. the class of an object (person, car, truck).

2. Dynamic camera: To obtain detailed or small-scale information about objects of interest in the scene. The warning information from this camera is used for multi-scale analysis. For example, if the camera is pointed at a person, the analysis should include face detection.

3. Object detection : To identify all occurrences of objects from a known class, such as people, cars, or faces in an image. For this two approaches have been used. The first approach, called background subtraction, takes a still background and treats all changes in the scene as objects of interest. The second approach, called motion detection, assumes that a scene will have many different types of motion, only some of which are of interest from a surveillance point of view.

4. Object classification : This step is based on image segmentation to particular objects in order to assign each object a class and detect alarm situations in the filmed scene. For example, detecting an animal at a fence line at the edge of the woods cannot be an alarm condition, whereas spotting a person will certainly be an alarm condition.

5. Object tracking: To track the progress of objects in a visual scene, this section aims to develop object trajectories over time using a combination of object appearance and movement characteristics. For the realization of this task two approaches, 2D and 3D have been applied.

6. Analysis of the structure of the object : Often the object is present in the scene is not enough, and it becomes necessary to analyze and locate other parts of an object. For example, finding a person's head or the license plate of a car is important from an identification point of view. The purpose of this component is to extract meaningful information from images.

3.2.5.2 Video Surveillance and Monitoring System (VSAM)

Keeping track of individuals, traffic and their communications in an metropolitan environment or border region is a difficult undertaking. The role of VSAM is to permit a solitary human operator to screen a huge territory through a comprehension of video. To accomplish this objective, it is important to automatically analyze individuals and traffic in a video, determine their geolocalizations, and insert them in a dynamic scene visualization [42]. As shown in figure 3.4, this system is made up of 3 main components which are detailed below:

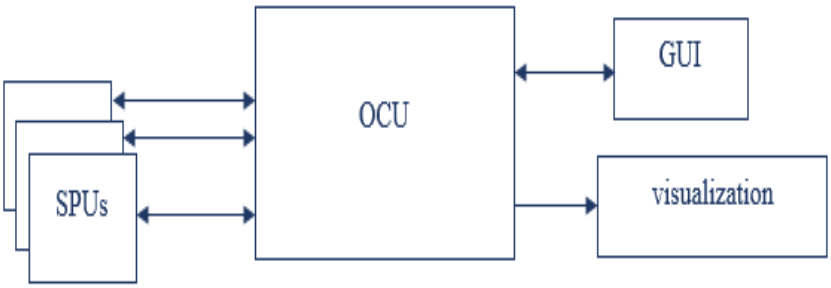

Figure 3.4: The components of the video surveillance and monitoring system (VSAM).

Sensor processing units (SPUs):
The VSAM consists of a set of processing units linked with sensors whose role is to analyze video images to determine the presence of significant entities or events, and to transmit this information symbolically to the OCU. In each unit, robust routines for detecting and tracking moving objects have been developed. The detected objects are classified into semantic categories such as human, human group, car, or truck using shape and color analysis. These labels are used to improve tracking using temporal consistency constraints. The classification of human activity, such as walking and running, has also been achieved. The geolocalization of labeled objects is determined from the coordination of their images using two or more overlapping camera views and a terrain model from monocular views. The calculated locations feed a tracking module at a higher level than the sensor spots with dynamic camera zoom and tilt variables to cooperatively and continuously track an object through the scene.

The Central Operator Control Unit (OCU):
All object hypotheses obtained from all sensor units are transmitted as symbolic data packets to the Central Operator Control Unit (OCU). The OCU is responsible for integrating the symbolic object trajectory information accumulated by the SPUs into a 3D geometric site model, and presenting the results to the user on a map via a graphical user interface (GUI).

Graphical User Interface (GUI):
One of the objectives of the VSAM project is to demonstrate that a single human operator can effectively monitor an important area of interest where he does not have to follow the tracks of a set of objects of people, vehicles, and their interactions in about twenty screens (showing raw video output) at a time. For this reason, this system uses a principle to automatically place dynamic agents (sensors) representing people and vehicles in a synthetic view of the environment as shown in figure 3.5 to the user through an interactive tool (GUI).

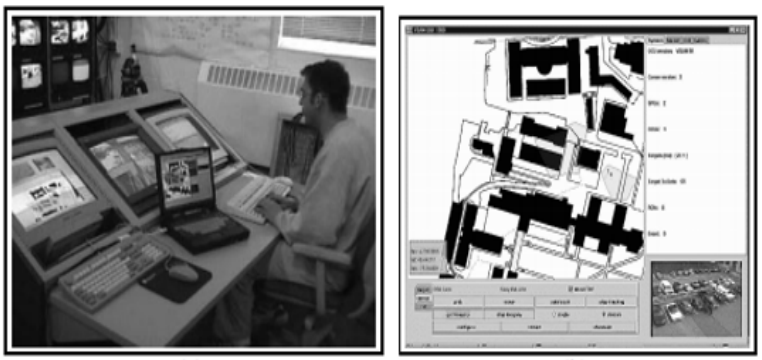

Figure 3.5: a) A laptop that represents the operator's console located in the control room b) Close-up view of the display node display screen [42].

3.2.5.3 Detection of suspicious behavior in shopping malls

For the purpose of having real-time monitoring of potentially suspicious behavior in shopping centers, an idea has been proposed by Arroyo and al [9] to extract useful data from captured videos in order to automate several video surveillance tasks by giving alarms when risk events are detected. It is intended to help human security officers make appropriate decisions to improve the protection of property.
In this scenario, there are particular situations that need to be analyzed, such as entering or leaving the store, loitering events that can lead to theft, as well as situations where checkout is unattended, as shown in Figure 3.7 The different components of this system are :

1. **Image segmentation :** First an image segmentation is applied to locate the foreground objects in the scene. Unfortunately, the detected objects may represent only partial parts of people's bodies. However, the leading, biased parts are filtered by size and position factors and grouped into final human targets by a new merging technique that reduces typical segmentation errors.

2. **Follow the people in the video :** This step is based on people's trajectories and visual appearance information of people in occlusion situations.

3. **Behavior analysis and alarm:** Finally, the resulting trajectories of people obtained from the previous step are processed to analyze human behavior and identify potential shopping center alarm situations such as people entering and leaving, loitering situations and unattended checkout situations(as shown in Figure 3.6).

(a) Entry and exit (b) Loitering event (c) Unattended cash desk

Figure 3.6: Alarm situations detected by the expert video surveillance system in a shopping mall [9].

(a) background subtraction (b) Blob fusion (c) Tracking

Figure 3.7: The operating steps of an expert video surveillance system in a shopping mall [9].

3.2.5.4 W(4) video surveillance system

The W4 is a real-time video surveillance system proposed by Haritaoglu and its employees in 2000 for the detection and tracking of several individuals and to supervise their actions in an outside environment. This system is the combination of a set of techniques to simultaneously track persons and their body parts in the monochromatic video. The W4 was developed to operate only with static and monochromatic video sources, either visible or infrared. It is designed for outdoor surveillance tasks, particularly at night or in other low-light situations. In such cases, the color will not be available, and people must be detected and monitored based on the faintest appearances and clues to their movements [81]. The main characteristics of W4 are as follows:

- Detect foreground objects using statistical background subtraction models even when the background is not stationary (e.g. movement of tree branches).
- Distinguish people from other objects (e.g. cars) using shape and periodic motion cues.
- Follow the tracks of several people simultaneously, even when they move together, or interact with each other.
- Construct an appearance model for each person during tracking that can be used to identify people after occlusion

- Detect and track the main body parts (head, hands, feet, and torso) of each person.
- Determine if a person is carrying an object, and the segments of the object so that it can be tracked during exchanges.

The general operation of the W4 is detailed in the following steps:

1. **Object detection :** The purpose of this step is to detect all occurrences of objects from a particular class known in advance, such as people, cars or faces in an image. This step detects the silhouette of foreground objects in an image
2. **Silhouette analysis :** All objects are categorized according to one of three pre-determined classes (one individual, persons in a group, and other objects) using the fixed form of the silhouette and regular dynamic analysis.
3. **Posture analysis:** The operation of this analysis depends directly on the class of object to detect :
   - A single person: A pose analysis based on the silhouette is used to evaluate this individual's pose. If a person is in the vertical stand-up position, then a new dynamic movement analysis and periodic symmetry analysis are applied to determine if the person is wearing an object. If the person is not moving an object or is in a generic pose that is different from the standing posture, then W4 detects the body parts with the use of shape analysis based on the silhouette borders.
   - Group of persons: In this case, W4 cannot detect individual body parts, postures, or transported objects. For this, it assumes that all members of the group are in a standing posture and treats each individual separately.
   - An object other than a person: In this situation, the analysis of the posture is not carried out; it simply tries to follow the object through the video.
4. **Object Tracking** This step calculates the correspondence between previously tracked objects and currently detected objects in order to build models of appearance and movement, and retrieves the trajectories of the tracked object.

3.2.5.5 A video surveillance system with unrecorded privacy protection

Today, the proper use of video surveillance cameras helps expose wrongdoing, but usually comes at the expense of the privacy of those who are not involved in any wrongdoing. For example, employees who work long hours in more confined areas such as businesses and hospitals may feel that their privacy is being violated since they are constantly monitored. A fundamental challenge

is to design surveillance systems that meet security needs and at the same time protect the privacy of individuals. For this reason, Wickramasuriya and al in [183] have proposed a system that is more selective in the video it captures, and it focuses only on abnormal events while protecting the privacy of authorized persons who may appear in these videos streams. The main idea was to combine information from different location sensors, in particular RFID technology, with information from traditional video surveillance to build a framework for privacy protection of data collections in media spaces. This system consists of a set of tags, a reader/receiver that can read and write data to these tags, and eventually, each region is equipped with a field generator (to relay information from a region). This system consists of three main components :

1. Detection of surveillance modes: This component is based on communication between two types of beam sensors to detect and locate the location of each person entering a monitored area and determine whether that person is authorized or unauthorized.
2. RFID (Radio Frequency Identification) sensors: RFID technology offers the ability to query data content without contact. It represents a means of storing and retrieving data via electromagnetic transmission at a compatible integrated circuit radiofrequency. The role of this sensor is to automatically identify and track tags attached to individuals. The labels contain information stored electronically. When a person entering a region is detected, the tag information is read and transmitted to the RFID Control Module to query an access rights database to verify whether that person's tag is authorized to enter that region or not. The access decision for that person is then transmitted to the video processing module.
3. Motion Sensor: Since we are interested in the input (and output) of a region, each field generator is equipped with a motion sensor that triggers a reading of the region when motion is detected. If there is no tag information associated with the person entering, the signal sent to the player is classified as unauthorized surveillance and the recording of video in the region is triggered.
4. Access Control Specification: This step is essential to this system, allowing the specification of spatial access constraints (i.e. what areas a person has access to) using XML to specify the policies that dictate how video surveillance is conducted in a physical space. Tags are distributed to staff, and a database stores the access rights associated with each tag. For example, a member of a company's cleaning staff enters to clean offices at 2 a.m. could be considered normal. However, if company policy prohibits entry to certain parts of the building after midnight, this event may be considered a potential security breach.
5. Video processing: This component uses several video signal processing techniques for motion detection and tracking of people, as well as masking of persons authorized to enter an area (with access rights).

3.2.5.6 A video surveillance system with recorded privacy protection

We have noted that the system proposed in [183] meets the privacy objective, but it has a potential security gap because it has not kept a record of privacy information. If authorized personnel have also been involved in malfeasance or even criminal behavior, the system should have the capacity to provide the original surveillance images if necessary. It is for this reason that Zhang et al in [195] has come up with the idea of proposing a detailed privacy information storage framework for signature video surveillance. Authorized personnel are not only removed from the surveillance video as in [183] but are also embedded in the video itself, which can only be retrieved with a secret key. In this sense, the system should provide multiple levels of access to the surveillance video: the general public can only see the unauthorized behavior, while special authorities that possess a secret key, such as the police, can observe all the information produced in the monitored space. The simplest solution to this problem is to separately store a copy of the original video surveillance. The goal was to monitor only unauthorized persons in a restricted environment and protect the privacy of authorized persons, but at the same time allow private information to be revealed securely and reliably. This system assumes that all persons captured by the camera in a confined space are categorized into authorized and unauthorized groups. A video surveillance system with privacy protection must meet the following requirements:

1. Register the entire region.
2. A person with a lower level of security clearance can only see the behavior of unauthorized persons, and all privacy information must be completely imperceptible to him/her.
3. Under special conditions, a person with a higher level of security clearance may use a secret key to view the original video, including privacy information if applicable.
4. Without the correct secret key, it is not possible to reconstruct the privacy information from the video, even in extreme cases.
5. The system should provide a reliable means of proving the authenticity and integrity of the resulting video in which the private information has been embedded.

3.2.5.7 Other Related works

Palak et al [139] provided a survey of different Methods in Border Security and Surveillance, The aim is to compare different researches in border security. Arfaoui et al [1] developed a model that estimates the crossing time of the monitored area taking into account the characteristics of the area and the behavior of the intruders crossing this area. Then they proposed a deployment method based on the intruders crossing paths that optimize the number

of deployed sensors while ensuring an early and high detection level of the intruders. Laura et al [109] proposed a multilayer hybrid architecture based on cameras, scalar sensors, radars, and UAVs to design a border surveillance system. Bhadwal et al [16] proposed a smart border surveillance system that can provide round the clock video surveillance at the places where human deployment is not possible. Al Abkal et al [2] investigates the use of drones, in border security and their ability to enhance existing security measures in Kuwait's ports and borders and also along borders of the United States. The study contributes to practice by introducing the use of UAVs to enhance port security, especially for monitoring and surveillance purposes. Segireddy et al [163] developed a light detection and ranging (LIDAR) sensor for the acquisition of distance with a range of 40 m from the position where an object resides. Data collected by the sensor is monitored and administered in a server. Software required for the analysis of data and generation of alert notifications is deployed in the server which is an added feature to the system and assists the concerned security personnel to respond quickly and engage the safety. Ayush et al [69] employ machine learning techniques in Remote Video Surveillance for real-time threat level detection and classification of targets crossing borders. The algorithm used for the machine learning-based detection of objects in the videos in this research is the Viola-Jones algorithm. A threat level classifier and alert warning system were also added to classify and annotate the videos in real-time for each frame. The threat level classifier performs four-fold categorization of the real-time video into safe, low, medium, and high (danger). The alert warning system specifies the type of warning based on the type of intrusion (human, vehicle, or weapon) detected. Kim et al [98] propose to develop a drone-aided border surveillance system with electrification line battery charging systems (DABS-E). This paper proposes an optimization model and algorithm to schedule drone flights for a DABS-E. Through a numerical example. Karthick et al [99] proposed an architecture that involves a low energy intrusion detection system on the first level. If the system detects any unusual event, it initiates a secondary authentication unit. This is again a sensor that detects the traces of the event. If the second sensor detects the same, it authenticates the event and switches ON the wireless camera. This system has multiple advantages like reduced power consumption, improved event detection accuracy, longer life span, and enhanced information clarity. D. Arjun et al [3] describes in his paper the current Wireless Sensor Network (WSN) techniques related to border surveillance and intruder detection. Harish Bhaskar [18] proposed integration of simultaneous detection, following, and face-acknowledgment based identification of human targets from a static camera is proposed. The precision, effectiveness, and heartiness of the proposed work are assessed and illustrated over different standard datasets over a wide scope of scenarios utilizing appropriate performance metrics. Jun He et al [82] demonstrate an ad-hoc WSN system for border surveillance. The network consists of heterogeneously autonomous sensor nodes that distributively cooperate with each other to enable a smart border in remote areas. This work also presents al-

gorithms designed to maximize the operating lifetime of the deployed sensor network.

3.3 Border and sensitive site surveillance

To meet the requirements of a border control policy that meets the needs of territorial security and enhanced deterrence against potential irregular migrants and to strengthen internal security and the fight against terrorism and organized crime and other illegal activities such as trafficking, illicit trafficking of migrants at the same time to facilitate the legitimate movement of persons and goods, while maintaining border security and protecting the privacy of individuals, we had the idea to model a system that will make it possible to better control movements at borders and to better manage migrations, which autonomously detects unwanted individuals through gestures and not their facial recognition while protecting the privacy of individuals. In our contribution, we assume that all the people that the drones equipped with cameras capture in the filmed area can be classified into two classes: gestures of border soldiers and gesture of terrorists and non-soldier people, our system is characterized by the following properties:

- The ability to protect the privacy of individuals to the general public. A person who does not have access can only see the private information (face and body) of the people. Initially, all the people filmed are considered "in good faith" and they are masked. Once a person is detected as unwanted individuals then that person's private information is automatically unmasked.
- The original video can be retrieved by persons with authorized access who are usually the authorities with special security clearances. For example, border guards can have access to data and with a private key can retrieve original videos to solve an investigation.
- Detect unwanted individuals through their gestures even if they hide their faces or change their look by disguising themselves.
- The automatization of the detection of the different situations of risk and to help the border guards to make decisions appropriated to ameliorate the control of the border

In the literature, the systems proposed in section 2 have a flaw is that they do not ensure the privacy of all people since they allow hidden only privacy information for certain authorized persons and known in advance. These systems cannot be used in public places where they can be placed only in restricted and refined areas. The objective of our offered system is to assure private life not only for some allowed person but for all persons. For it, we used human gestures (instead of facial recognition) to discern if a person is

undesirable or not to conceal all persons. The general architecture of the proposed system is shown in the following Figure

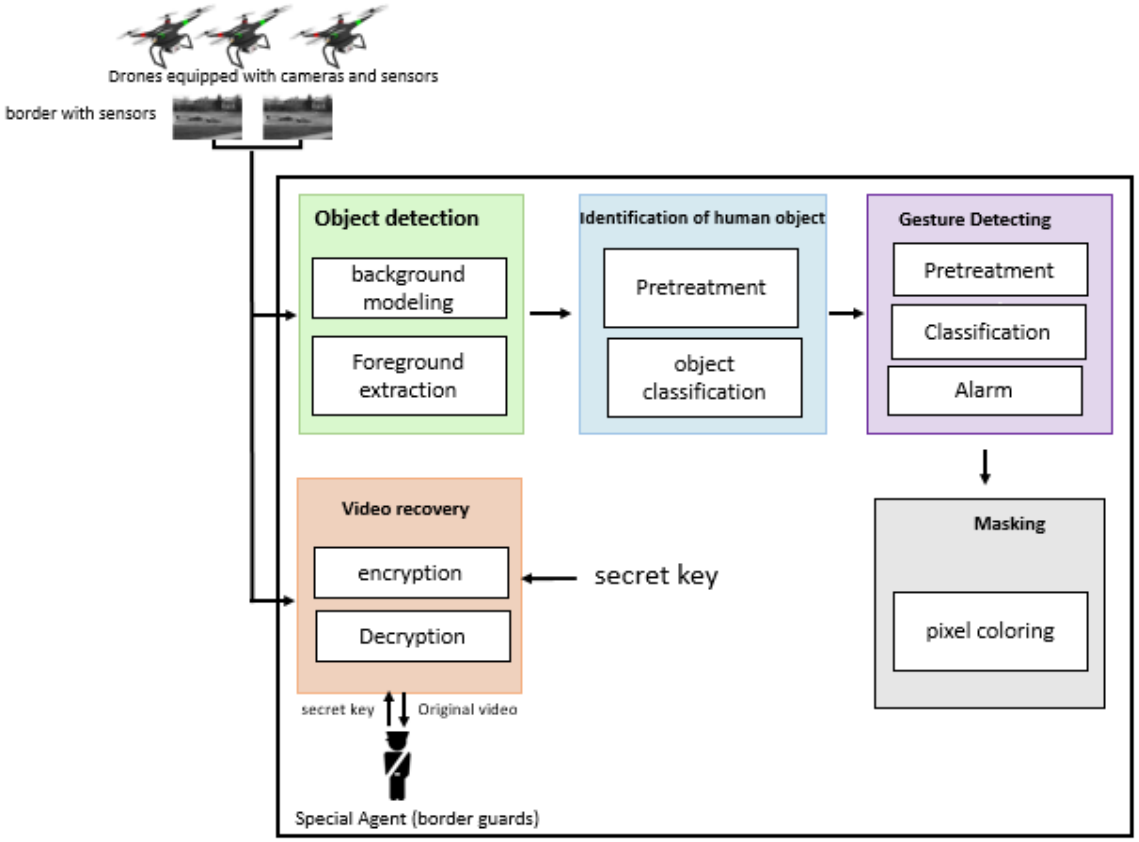

Figure 3.8: General architecture of the system for Surveillance of Borders and sensitive sites based on gestures.

As shown in the Figure above, the cameras integrated or added to the drone represent the main elements of each surveillance system. They are used to cover the entire area of interest and provide a global and detailed view to track objects and extract additional information, for example, the class of an object (person, car, truck).

3.3.1 Object detection

This module consists of identifying objects in an image. This requires a segmentation step to partition a digital image into several groups (pixel set). Each group is supposed to correspond to an object in the image. In a video surveillance scenario, the goal is to separate the areas of the scene that belongs to the background from the regions belonging to the foreground namely the moving objects. For this step a background subtraction algorithm is used that can provide real-time results to automatically generate the silhouette of human actions presented in video image sequences, where the data from each camera is processed by the following two steps:

3.3.1.1 Background modeling

This step is implemented by creating a model that represents the regions of the scene that remain constant over time. For this, we propose the use of the Gaussian statistical model since it gives precise results in real-time compared to the other background subtraction methods [83].

3.3.1.2 Extracting objects from the foreground

Once the background pattern is calculated, the foreground objects are detected by calculating the difference between the original image and the background pattern as shown in the following Figure. The output of this operation is a binary mask called foreground image containing objects that move in the filmed area.

Figure 3.9: An example of background subtraction and extraction silhouettes of foreground objects moving in three different images.

3.3.2 Identification of human objects

This module aims to classify interesting objects in the field of vision of the camera (s) since in our work we are interested only by the movement of humans. For this reason, it is necessary to label each moving object to distinguish humans from other objects. The entry of this step is the silhouettes of moving objects extract from the images (video sequence) by the previous step. In this case, we are faced with a supervised classification problem since the purpose is to classify each moving object in one of two classes (person and other objects) using pre-classified images by an expert as learning data. For the realization of this stage, we can to complete this step we can follow a two-step process:

- Binary image preprocessing: This step detailed in section 3.3.1, is to transform each bit mask (foreground object image) to a vector.
- Object classification: We used the k-nearest neighbors (KNN) algorithm, which requires the presence of a learning base consisting of images pre-classified by an expert (each image is a binary mask containing

the silhouette of an object). Afterwards, a distance is calculated between each new image to be classified with each frame of the learning base.

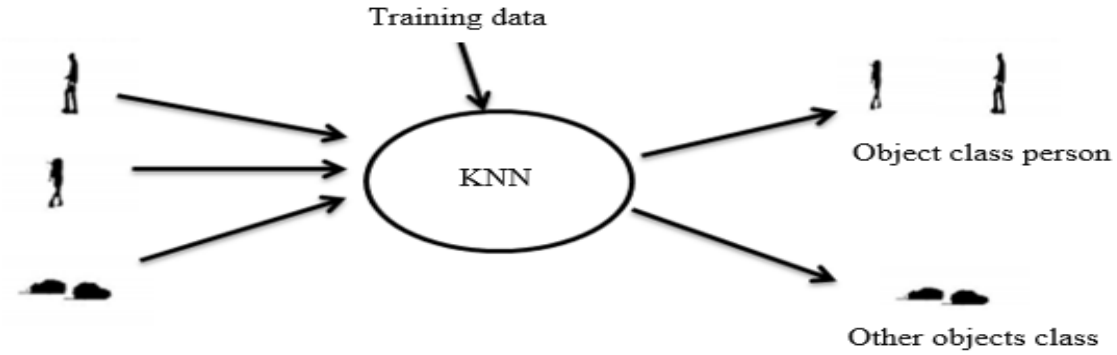

Figure 3.10: Identification of human objects.

3.3.3 Detection of undesirable persons

Once the silhouette of the object in an image is labeled as a person then we analyze this object in order to detect whether it is an undesirable persons or not through its gesture present by the silhouette. The classification of human gestures is a problem of binary classification (undesirable human gesture and border soldier's gesture). For the realization of this module the algorithms of supervised classification can be used where we applied the classifier of the artificial cockroaches detailed in 3.3.3.2 after a stage of vectorization of each gesture of a human realized by the pretreatment process. On the other hand if the object detected is not a human then nothing will be reported.

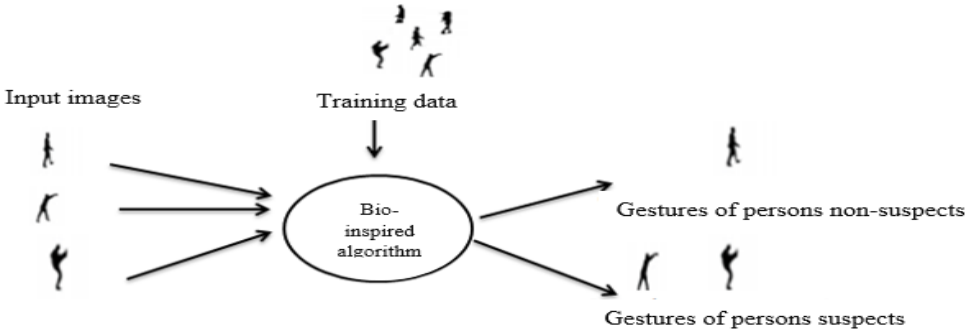

Figure 3.11: Process for detecting undesirable persons.

3.3.3.1 Preprocessing of binary images

When the data entered into our software are binary images then the preprocessing steps are as Follows:

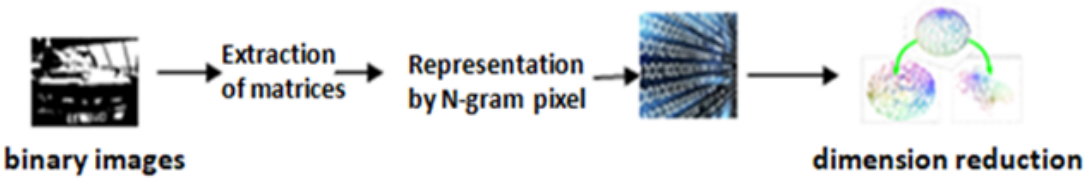

Figure 3.12: The preprocessing steps for binary images (black and white).

- **Extraction of matrices:**

  Pixels in our pictures are in color RBG, therefore for the extraction of the matrix we compare the stocks RBG of every pixel with the stocks RBG of color black ($R = 255, G = 255, B = 255$) and white ($R = 0, G = 0, B = 0$). If it am black we shall replace RGB with zero, otherwise we shall replace it with one. Finally, we shall stock every matrix in a text file. The pseudo encodes according to sums up this stage:

Algorithm10: Binary matrix extraction

```
Input:: Binary images
binary-matrix ⟵ chain of character «»
  Begin
      for i =1 at the width of the image make
      for i=0 at the width of the image make
      If  pixel-picture (i,j) is black then
       binary-matrix ⟵ binary-matrix + « 0 »
       If else binary-matrix ⟵ binary-matrix + « 1 »
  End
            binary-matrix ⟵ binary-matrix + new-line
  End
        Text file⟵ Save binary matrix
End
```

- **Representation by N-gram pixel:**
  We had the idea of representing N-Gramme pixels, trying to mimic the representation N-gram characters. Each binary matrix of an image (built from the previous phase) is considered a text and each pixel is taken as a character and we follow the same instructions of the N-gram character technique. The basic principle is that two images are similar if they carry the same elements (N-gram pixels). This step ensures the transition of each image to a set of small units called N-Gram pixels.

- **Tf-Idf :**
  TF-IDF (Term Frequency-Inverse Document Frequency) is a weighting method often used in information retrieval and particularly in text

mining [162]. This statistical measure makes it possible to evaluate the importance of a term contained in a document, relative to a collection or corpus. The weight increases proportionally to the number of occurrences of the word in the document. It also varies according to the overall frequency of the word in the corpus [43].
The term frequency is simply the number of occurrences of the term in the document under consideration. The principle is that the more frequent a term is in the document, the more important it is in the description of the document.

3.3.3.2 Grouping Cockroaches Classifier (GCC)

The Origin of the Algorithm

We used the studies carried out by Bell on the social life and behavior of cockroaches in [17]

The Inspiration Source

GCC is inspired by the natural behavior of cockroaches and the phenomenon of seeking the most attractive and secure place (shelter) for hiding. We can identify different types of cockroaches in our work, we are interested by the cockroaches that live in apartments, which are fertile and they are never isolated. This phenomenon is well detailed in an experiment conducted by French biologists when they met a group of cockroaches in a basin where there's light everywhere, and they built two artificial shelters (shelter is a place with less brightness as shown in the Figure 3.13) using two red circles because cockroaches do not observe the color red as shown in the following Figure (Figures 3.13 and 3.14).

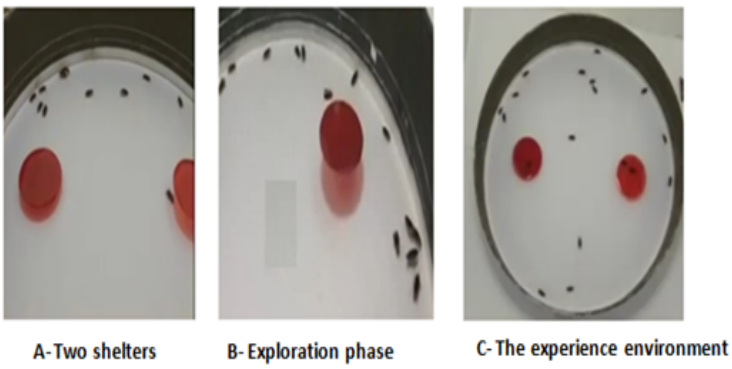

Figure 3.13: Description of the cockroach grouping experiment [17].

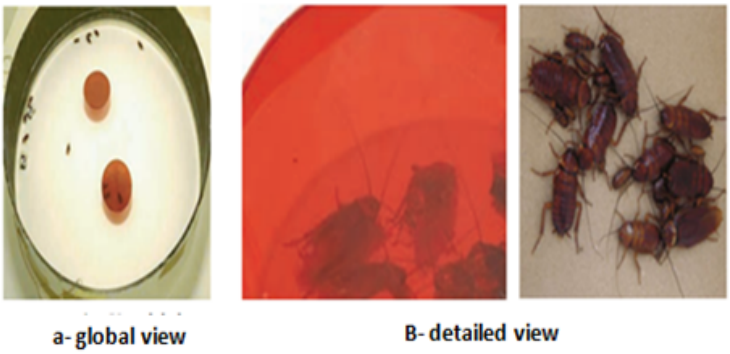

Figure 3.14: The groping of cockroaches under the same place [17].

From previous experience it was observed that cockroaches have a choice of two shelters to hide where they always choose the most secure shelter. A biological model explaining this phenomenon is presented by the following:

- **Random Displacement of Cockroaches**

  Initially cockroaches will move randomly in all directions (exploration phase) as demonstrated in Figure 3.14.b. When a cockroach finds an attractive shelter, it hides and sends pheromones as smell to its congeners. The movement of cockroaches is guided by a set of displacement rules:

  1. The Darkness Shelters: Cockroaches are attracted by the darkest places like corners and shelters with less brightness. The degree of darkness plays a very important role for the quality security of each shelter.
  2. The Congener's Attraction: Each cockroach seeks shelter where there are more of its congeners (cockroaches from the same colony) to hide it.
  3. The Security Quality: Cockroaches positioned in the middle of the shelter have a higher safety compared with cockroaches positioned at the border of the shelter.

General Processes

In the supervised classification problem the data set is divided into two bases (learning basis and test basis). Each new instance (cockroach) well be classified (hidden) in a class (shelter) using a security function that is based on the attractiveness of each class (calculated using the aggregation operators (shelter darkness, congeners attraction and the quality of security), and the displacement probability (calculated using the naive Bayes algorithm). the general process of the GCC is illustrated in the next Figure

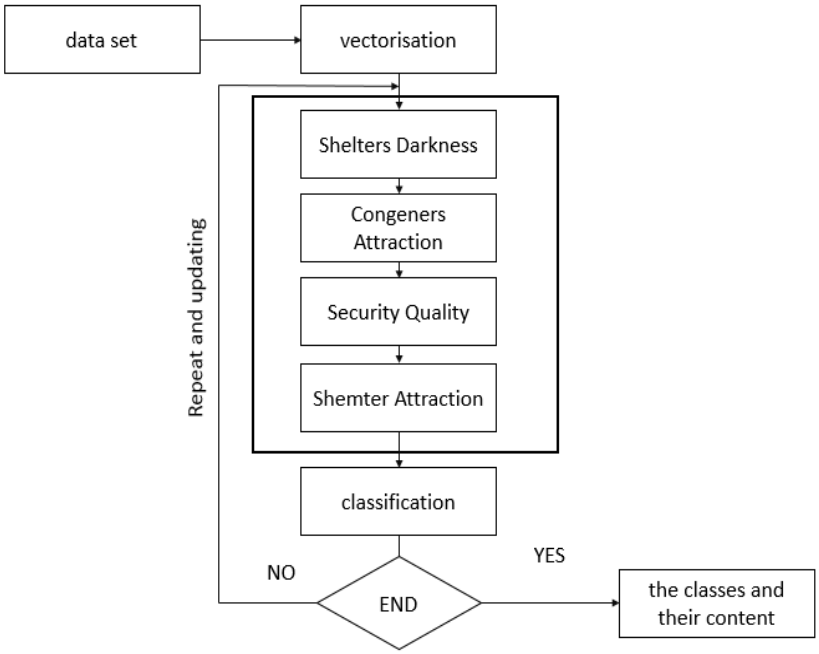

Figure 3.15: The general functioning of Grouping Cockroaches Classifier (GCC).

The general architecture of Grouping Cockroaches Classifier (GCC) is illustrated in Figure 3.15 and each stage of its operation is are described below.

1. **The Darkness Shelters :**

   The Darkness Shelters calculates the rate of darkness in each shelter that represents the number of instances belonging to each class (shelter) relater relative to the total number of instances in all classes. Initially, the cockroaches of the learning base are hidden in each corresponding shelter (we know the instances classes of the learning basis).

$$OA(S_i) = \frac{CS_i}{CL} \tag{3.3.1}$$

   - CS_i: The number of instances belonging to the class S_i.(the number of cockroaches in the shelter)
   - CL: The total number of instances in all classes (the total number of cockroaches in the shelter)
   - $OA(S_i)$: the darkness rate of the shelter S_i.

2. **The Congeners Attraction (CA):**

   As shown in equation(3.3.2), The CA is defined by a parameter K fixed in advance, and for each new instance classifying C_n, we randomly select k instances (designated as k congeners for the cockroach C_i in the shelter S_i) of each class. Then, the total number of instances

belonging to this class divides the sum of the distances between this instance and its K congeners.

$$CA(C_n, S_i) = \frac{\sum_{k=i}^{K} distance(C_n, C_k S_i)}{CS_i} \tag{3.3.2}$$

- Distance $(C_n, C_k S_i)$ The distance between the instances to be classified C_n and its k nearest neighbors in the class S_i.
- CL: The total number of instances in all classes (the total number of cockroaches in the shelter)
- K: the number of selected instances.
- CS_i: The total number of instances in the classes S_i.

3. The Security Quality:

A cockroach must be in good condition to stay in a shelter and it has a maximum quality of security when it is close to the middle of the shelter. The security quality of the instance C_n in a class S_i is calculated through equation 3:

$$QS(C_n, S_i) = distance(C_n, BS_i) \tag{3.3.3}$$

- BS_i: the centroid of the class S_i.

4. Shelter Attraction :

We use the results of the previous aggregation operators to calculate the attraction of each class for each instance as follows:

$$SA(C_n, S_i) = \frac{\alpha OA(S_i)}{\beta CA(C_n, S_i) + \lambda QS(C_n, S_i)} \tag{3.3.4}$$

- α, β and λ: the Adjustment coefficients to adjust the impact of each operator in calculating the attractiveness of each class.

5. Probability of displacement:

For this, to calculate this probability we used the naive Bayes algorithm. Bayes' theorem provides a way to assign each instance a probability for each possible class. He assumed that the effect of the value of a predictor x_n on a given class S_i is independent of the values of other predictors. the probability of each instance to be classified in a class S_i is calculated by the next equation:

$$P(S_i/C_n) = P(x_1, S_i)P(x_2, S_i) \cdots P(S_i) \tag{3.3.5}$$

- $P(S_i/C_n)$ =: The posterior probability is the probability that the instance C_n is classified in the class S_i .
- $P(S_i$: is the prior probability of the S_i .

- $P(x, S_i)$: is the probability that component x generates the class S_i .

6. The security function:

The cockroach always belongs to the most attractive shelter where it is more likely to reach it (each new instance will be classified in the most attractive class where it has more probability). For this, we used the security function $f(C_i, S_i)$ which allows us to find the most appropriate class for each instance (the most secure S_i shelter for each cockroach C_n). The final decision concerning the class of each instance is done following the value of the security function:

$$f(C_n, S_i) = SA(C_n, S_i) + P(S_i/C_n) \tag{3.3.6}$$

- $P(S_i/C_n) =$: The probability of the cockroach C_i to be classified in the class S_i. Each instance is classified in the shelter that has the highest value of the safety function.
- $SA(C_n, S_i)$: The attraction of the class S_i for the instance C_n.

7. Update:

After each iteration, we update the values of the aggregation rules and the probability of displacement for each instance .when a cockroach does not feel safe (instance is miss-classified), then it will look for another more secure shelter (we reclassify this instance again). The process is repeated until a stopping criterion.

8. stop criterion :

The stopping criterion GCC is the number of iterations fixed in advance, or if the number of instances in each class remains the same for the iteration i and iteration $i + 1$.

3.3.4 Masking normal people

Once a person's gesture in the binary mask is detected as border soldiers then that person's face and body will be automatically hidden following the pixel coloring technique [190] as shown in the next Figure in order to hide his privacy information. On the other hand, if the person is detected as undesirable persons then his or her privacy information will not be hidden.

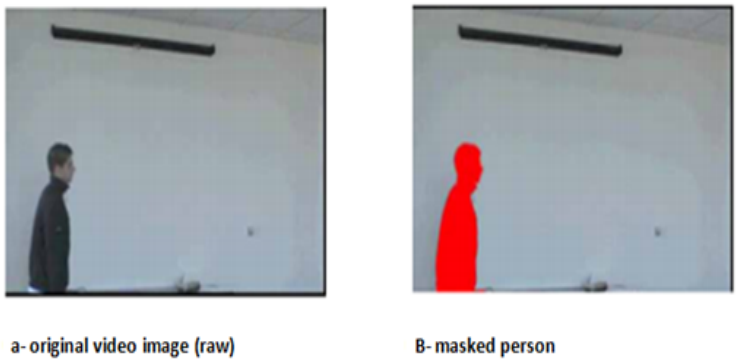

Figure 3.16: Example of masking of a person using the pixel coloring approach that hides privacy details such as the face and body.

3.3.5 Alarm

Our border surveillance system should be able to respond to specific events. Once a person is detected as unwanted then an alarm will be triggered. The goal is to tell the screening officers or users of this system that they are in a situation with an abnormal event and that you have to intervene by following in that person's footsteps or trying to arrest that person.

3.3.6 Original video recovery

In the event that people detect by our system as a border soldiers to whom we have hidden their privacy information (face and body) have been involved in some evil or criminal behaviors, our system has the ability to provide original surveillance images when necessary. In this sense, our system can provide access to special authorities who have a secret key, such as border guards, who can observe all the information that has arrived in the guarded space. The simplest solution to this problem is to store a copy of the original encrypted surveillance video separately.

3.4 Experimentation and results

3.4.1 Experiments

Given that our system is primarily based on detecting people who cross the border illegally even if they hide their faces based on human gestures instead of facial recognition, we will only present the results experimental purposes for this part of the system only. Before we begin our experimental protocol, we must first determine the baseline data set used.

3.4.1.1 The MuHaivi dataset

1. **Video clips:**
   A significant body of human action video data has been collected using 8 Schwan CCTV cameras in a site with challenging lighting conditions. The cameras are located at 4 sides and 4 corners of a rectangular platform (Figure 3.17 and Table 3.1). These cameras are not automatically synchronized, but the video segments for each action/actor combination have been manually synchronized. There are 17 human action classes $(C_j : C_1, C_2, \cdots, C_{17})$ as listed in Table 3.2 performed by 14 actors $(A_k : A_1, A_2, \cdots, A_{14})$. The video sequences contain a number of image frames before the action takes place so as to allow background estimation algorithms sufficient time to model the background, if necessary. [165]

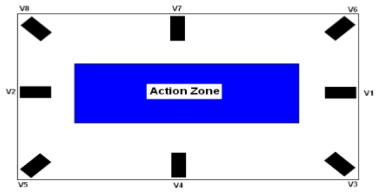

Figure 3.17: View of the configuration of eight cameras used to capture actions in the blue action area (marked with white bands on the floor of the stage). [165]

Silhouettes manually annotated:
The dataset provides a sub-set of data that has been (painstakingly) manually annotated. This of course reduces the size of the data available for "pure" action recognition. A detailed performance evaluation of state-of-the-art object detection algorithms using this small sub-set of data is currently underway with the view to select a robust method to compute these silhouettes automatically.

CAMERA SYMBOL V_i	CAMERA NAME
V1	Camera_1
V2	Camera_2
V3	Camera_3
V4	Camera_4
V5	Camera_5
V6	Camera_6
V7	Camera_7
V8	Camera_8

Table 3.1: The names of the camera views used in the data record and the corresponding symbols used in Figure 3.17. [165]

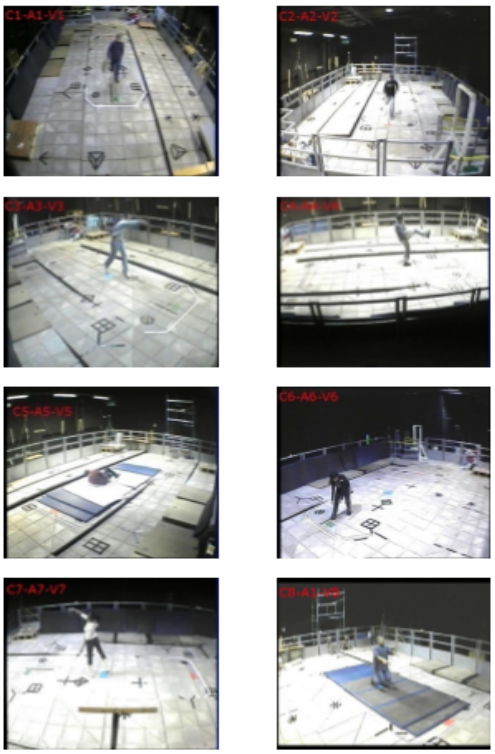

Figure 3.18: Views of all 8 cameras showing examples of measurements and actors sample camera symbols as in Figure 3.17.

This subset of manually annotated data consists of actions $C_1 \cdots C_5$, actors A_1 and A_4 and cameras V_3 and V_4, therefore a total of $522 = 20$ actions. Samples of the manually obtained silhouettes are shown in Figure 3.19. Although actions $C_1 \cdots C_5$ are relatively elemental from a human point of view, they can still be decomposed further into primitive actions. [165]

ACTION CLASS	ACTION NAME
C1	WalkTurnBack
C2	RunStop
C3	Punch
C4	Kick
C5	ShotGunCollapse
C6	PullHeavyObject
C7	PickupThrowObject
C8	WalkFall
C9	LookInCar
C10	CrawlOnKnees
C11	WaveArms
C12	DrawGraffiti
C13	JumpOverFence
C14	DrunkWalk
C15	ClimbLadder
C16	SmashObject
C17	JumpOverGap

Table 3.2: The action class names used in the data record and the corresponding symnoles used in Figure 3.19 [165].

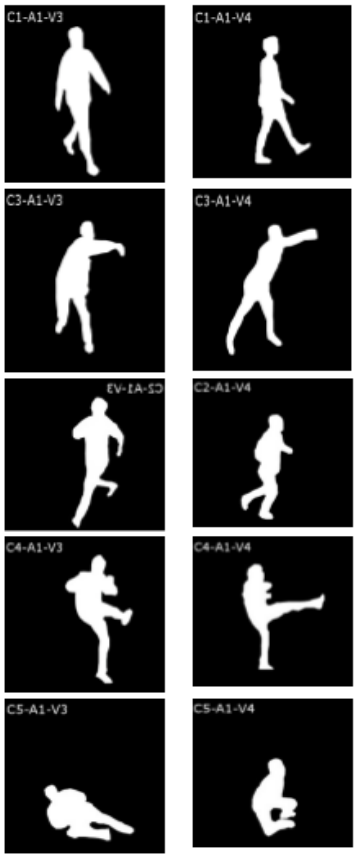

Figure 3.19: Examples of manually annotated silhouettes [165].

They grouped the images of the actions as a class of images of the gestures of terrorists and non-soldier people and on the other hand images of the actions of border soldiers gestures. The modified MUHAVI dataset is defined in Table 3.3. Each gesture-based unwanted person detection algorithm has as input learning data pre-classified by an expert (binary masks for gestures of unwanted persons and other for the actions of border soldiers). The following table represents the redistricting of the data used to conduct our tests (learning and test data):

	Number of learning images	Number of test images	Total
gestures of terrorists and non-soldier	200	314	514
Gestures of border soldiers	100	326	426

Table 3.3: Muhaivi dataset decomposition.

3.4.2 Results and analysis

To test the suspicious gesture detection module only, we used Grouping Cockroaches Classifier (GCC) with number of *iteration* = 1 and weights of aggregation rules ($\alpha = 1$, $\beta = 1$ and $\lambda = 1$) as well as $K = 1$. For this, we applied GCC on the Muhavi-suspicious-persons dataset detailed earlier since it consists of a set of manually annotated human silhouettes which does not require testing the object detection and human identification modules. For the validation of the obtained results, we used the supervised measures with the class of suspicious people as a positive class in the contingency matrix. We performed different tests in order to analyzing the performance of the GCC by studying the influence of each parameter.

1. **The influence of image representation and distance measurement**
   Before applying the GCC a process of vectoring the images is necessary. We have varied the value of the N parameter used by the N-grams pixel representation method in the pre-processing phase and each time we set a distance measurement to assess the quality of each output. The results are detailed in the following tables:

		Evaluation Measures							
		Recall	Precision	F-Measure	Entropy	Accuracy	Confusion matrix		
								U	N
							U	TP	FP
							N	FN	VN
N-Gram Pixel	2-gram	0.7229	0.8376	0.776	0.1484	79.53	227		44
							87		282
	3-gram	0.7675	0.8743	0.8124	0.1174	83.125	241		35
							73		291
	4-gram	0.8184	0.8862	0.8506	0.107	86.937	257		33
							57		293
	5-gram	0.8248	0.8961	0.859	0.0983	86.71	259		30
							55		296

Table 3.4: Undesirables person detection results based on human gestures using Euclidian distance and variation of the N parameter for N-grams pixel representation.

		Evaluation Measures							
		Recall	Precision	F-Measure	Entropy	Accuracy	Confusion matrix		
								U	N
							U	VP	FP
							N	FN	VN
N-Gram Pixel	2-gram	0.694	0.767	0.7289	0.2034	74.68	218		66
							96		260
	3-gram	0.7006	0.7885	0.745	0.1873	76.093	220		59
							94		267
	4-gram	0.707	0.8014	0.7511	0.1774	77.03	222		55
							92		271
	5-gram	0.7197	0.8071	0.7638	0.1729	77.81	226		54
							88		272

Table 3.5: The detection results of undesirables people based on human gestures using cosinus distance and variation of the parameter N for the N-grams pixel representation.

		Evaluation Measures							
		Recall	Precision	F-Measure	Entropy	Accuracy	Confusion matrix		
								U	N
							U	VP	FP
							N	FN	VN
N-gram pixel	2-gram	0.799	0.8655	0.8308	0.125	84.062	251		39
							63		287
	3-gram	0.8216	0.8896	0.8542	0.104	86.25	258		32
							56		294
	4-gram	0.8566	0.9057	0.8804	0.0897	88.59	269		28
							45		298
	5-gram	0.8503	0.89	0.8696	0.1037	87.5	267		33
							47		293

Table 3.6: The detection results of illegal migrants based on human gestures using the Manhattan distance and variation of the parameter N for the representation of images.

In our contribution, the main idea is that two images are identical if the number of occurrences of each N-pixel in these two images are the same.

After observing the results in Tables 3.4, 3.5 and 3.6 we noticed that:

- The Manhattan distance measurement gives the best results compared to the cosine and Euclidean distance validated by an $f - measure = 0.8804$ and $entropy = 0.0897$ (blue cases in Table- 3.6) because our goal is to find the exact difference between the vector components. In other words, two gestures are different if the occurrence values of their vector components are distant from each other. The distance between Euclidean and Manhattan give good results in relation to the cosine distance because we are interested in the magnitude of the image and not only by the relative frequencies of the N-pixels in the images.
- The recall is always less than the accuracy given that the majority of cases are classified as a border soldiers gesture validated by the Confusion matrix with $FN = 45$ and $VN = 298$ (the green cases of Table 3.6) because on the one hand malicious or criminal persons always try to hide their appearances and be as normal as possible and on the other hand the learning data we used does not aggregate all the gestures of terrorists and non-soldier people that may exist. We may also have a conflict in detection between border soldiers and undesirables person.
- Every time we increase the value of the N parameter, the results improve because the vectors will be made up of more components, making it possible to better differentiate between 2 images of any kind. For example the 2-gram pixels can generate only 4 components (00, 01, 10, and 11), the 3-gram pixels generates 8 components (001, 100, 010, 110, 111, 000,001, 011) and so on.
- In terms of Accuracys even though we will get a percentage of 88% but this is not enough as part of our goal since it means that it has a lot of false alarm report (border soldiers detect as undesirable people) and the private information of innocent people will be revealed to the normal public.
- In terms of entropy, the results are clearly performing as the accuracy is elevated because we did not use the normalization of images, which allowed getting less loss of information.

3.4.2.1 Comparative study

1. Statistical comparison

To give more reference to our results obtained, we have put in confrontation the best performance of our Classifier of Grouping Cockroaches (GCC) in the face of the problem of detection of illegal migrants based on gestures against the results of other algorithms that exist in the literature such as classical learning algorithms like the nearest Nearby K (KNN) with $K = 1$ and cosine distance and C4.5 decision trees that have been applied using the WEKA API that provides tools and libraries ready to be used directly. The results of this comparison are

presented in the following table:

		Evaluation Measures				
		Recall	Precision	F-Measure	Entropy	Accuracy
Classifiers	K nearest neighbors	0.8429	0.8681	0.853	0.1227	84
	Decision tree C4.5	0.6497	0.7329	0.695	0.2306	67.2
	Grouping Cockroaches Classifier	0.8503	0.89	0.8696	0.1037	87.5

Table 3.7: Comparative study in terms of the quality of results of different classifiers for the detection of unwanted persons based on gestures.

It should be noted in the table that the maximum value in the f-measure=0.8804 is obtained with the classifier of artificial cockroaches (blue cases) because it is based on different rule and property as (attraction of congeners, darkness of the shelter, the quality of safety, and the likelihood of travel). We also found that the convergence of this classifier takes a lot of time given the number of calculations and its complexity, which requires several tests and comparisons.
KNN classifiers give almost similar results to GCC (the yellow cases in the table) because they are based on a direct and nave operation using a distance measurement. On the other hand the bad results are obtained by the decision tree C4.5 method, because we are faced with binary images and the C4.5 is based on the gain of the ration and cannot identify the optimal root (prove in the literature).

Comparison in terms of services

What are the reasons for a good or poor performance of a video surveillance system for undesirable detection tasks? Table 3.8 compares our system with five other systems that exist in the literature Drone-Aided Border Surveillance with an Electrification Line Battery Charging System [98], WSN-based Border Surveillance Systems [1] , An efficient WSN based solution for border surveillance [109], Internet of Things based High Security Border Surveillance Strategy [99] ; Wireless IoT-Based Intrusion Detection Using LIDAR in the Context of Intelligent Border Surveillance System [163]) from several angles such as: The preservation of the privacy of all undesirable persons.

- Automatic detection
- The ability to detect undesirable people who hide their faces.
- The ability to retrieve original videos
- Detect and unmask undesirable individuals automatically.

- Location of use.

	Our proposed system	Drone-Aided Border Surveillance with an Electrification Line Battery Charging System [Kim, 2018]	WSN-based Border Surveillance Systems [Arfaoui, 2017]	An efficient WSN based solution for border surveillance[Laouira, 2019]	Internet of Things based High Security Border Surveillance Strategy [Karthick, 2019]	Wireless IoT-Based Intrusion Detection Using LIDAR in the Context of Intelligent Border Surveillance System [Segireddy, 2020]
Privacy Preservation	Yes	No	No	No	No	No
Automatic detection with alarm	Yes	No	No	Yes	Yes	Yes
Detection of undesirable people who hide their faces	Yes	No	No	No	No	Yes
Revelation of original videos for authorized persons	Yes	No	No	No	No	Yes
Unmask undesirable people automatically	Yes	NO	Yes	No	No	No
Place of use	international border and sensitive site	international border between countries	international border between countries	international border between countries	international border	international border

Table 3.8: Comparison in terms of services between our system and 4 other systems which exist in literature.

From the previous table, we note that our proposed system (the blue cases in Table 3.8) can be used in any location as it clearly meets all the requirements of a modern security policy by providing all services to ensure the safety of citizens and the government with the preservation of privacy. Unlike other systems that exist in the literature where each of them has shortcomings especially in terms of privacy as well as their inability to detect people who hide their faces.

3.4.3 Decisions

1. Every time we increase the N value of the N-gram pixel representation the results improve.
2. The adaptation of the N-gram technique for the representation of binary images was a very interesting experience since it does not require the normalization of images and it is tolerant to the problems of incomplete images.

3. The ideal configuration of the Grouping Cockroaches Classifier (GCC) is:
   - 4-gram pixel as a method of representation.
   - Manhattan as a measure of distance.
4. The GCC gives better results than classifiers like the KNN and C4.5 decision tree.
5. The GCC takes a lot of time to run compared to other conventional learning algorithms.

Our private detection system for undesirable's persons provides many advantages in terms of quality of services compared to other video surveillance systems that exist in the literature.

3.5 Conclusion

We introduced a new mta-heuristic inspired by the natural behavior of cockroaches for the surveillance of borders through videos captured by one of the drones via sensors; this algorithm is inspired of work of researcher's biologists who discovered the links of communication between the Cockroaches and their behavior. Acquired results are satisfactory and prove that algorithm is able of guaranteeing surveillance of borders. It gives better results in comparison with other algorithms existing in literature (k-means, tree of decision, C4.5), Validated by the measurements of valuation (recall, precision, Fr - Measure, entropy, rate of success, rate of error) We studied the impact of every parameter for the quality of performance of every algorithm to identify ideal .Finally, we can conclude that our contentment is full because targets fixed at the beginning were reached. For our future work, we will can extended this system for use on a larger scale then the system can be equipped with the mobile-based applications. Since IoT provides a global coverage, the data that is generated from the system can be accessed anywhere over the earth. Besides, we would like to propose to use an architecture based on deep learning in future work in order to improve our system.

4

RECOGNIZING PHYSICAL ACTIVITY OF HOSPITALIZED OLDER PEOPLE FROM WEARABLE SENSORS DATA USING IOT

Contents

4.1 Introduction

With the increase in life expectancy, it is now possible to age while remaining active. This is an opportunity, often a grace. However, old age nevertheless leads to physical and functional frailty. Thus, the elderly increased risk of falling. Falls of the elderly represent a major public health problem, both because of the seriousness of their often dramatic medical consequences and because of their undeniable social and economic impact. Therefore, Everything possible must be done to enable these people to continue their lives in the best possible conditions. For this, the idea of preventing falls is necessary to reduce their number and their undesirable consequences. challenges society and technology to find better ways to mitigate the occurrence of such costly and detrimental events as falls.

Today's Internet of Things (IoT) domain is increasing rapidly [12], It is "a global infrastructure for the data society, enabling advanced services by interconnecting (physical and virtual) things based on existing advancing

interoperable data and communication technologies" [156], IoT is one of the major technological developments of our times given its potential is fully realized [68]. A large amount of data are emerging every day to be part of the IoT infrastructure. According to Machina Research [78], 27 billion connected devices are expected by 2024, while according to Cisco's report [40], there will belong nearly 1.5 mobile devices per capita by 2020, and more than 601 million wearable devices will be in use. Connecting all the objects [96] and forming a network of devices is the basic idea of IoT. A major objective of IoT is to make the environment around us smarter, by giving the environment the information it needs, the IoT uses the internet to connected devices that can be easily monitored and controlled also the same things can be automatically detected by other things, further communicate with each other through the internet, and can even make decisions themselves [179]. Over time, various sensory data are collected and generated by an enormous amount of sensing devices. This Will result to generate in a big amount of data from the sensors used for collecting the data. To prevail over these applications some meaningful information must be deduced out of the collected data to make decisions. Applying analytics over such data streams to make control decisions, discover new information and foresee future insights is a pivotal procedure that makes IoT a worthy paradigm for businesses and a quality-of-life improving technology. Among the most extremely useful technologies is Data Mining. A major challenge in these settings is the timely analysis of large amounts of data (big data) to produce decisions and highly reliable and accurate insights so that IoT could satisfy its guarantee.

The aim of this work is the use of Artificial Intelligent to combat falls risks of older folks and to enable these people to continue their lives in the best possible conditions. In this paper, we present a methodology based on multimodal sensors to configure a simple, comfortable and fast fall detection and human activity recognition, and in cases of concern, alerts are sent to caregivers or family members to enable appropriate interventions. a system that can be easily implemented and adopted. In this work, we check whether the traditional data mining algorithms would likewise work for the IoT datasets, or new families of data mining are required. To this end, in this study, we examine the applicability of three data mining algorithms for real IoT datasets. These include K-nn, Naive Bayes, Decision tree The main contribution of this work is the analysis of the efficiency and effectiveness of three of the data mining. We began our work with some related works done in this field, after that in the third section we detailed a description of our system, which will be followed by a presentation of the experiment, these algorithms are analyzed and a comparative study is undertaken to find the classifier that performs the best analysis on the dataset obtained, using a set of predefined performance metrics to compare the results of each classifier and finally conclusions are given.

4.2 Literature review

Falls and fall-related injuries are common and serious problems affecting older adults hospitalized or housed in acute and long-term care facilities. As such, the reported rate of falls in hospitals ranges from approximately 1.3 to 9 per 1,000 beds per day [129] [136][93]. It is estimated that 30% to 50% of falls that occur in these facilities result in injury [129]. But this frequency is probably underestimated; these figures are undoubtedly only the tip of the iceberg. Indeed, elderly people do not always report all their falls to their doctor or their entourage, or even hide them voluntarily for fear of being considered unable to live at home, but also because elderly people tend to easily forget their past falls, especially if they did not have serious consequences. Among older adults, the frequency and severity of injuries that result from falls tend to increase with age. The consequences associated with falls (hip fracture, pain, anxiety, loss of independence, mortality, etc.) are significant and have significant impacts on health care costs [67] [93]. In the conventional solution for the mitigation of the risk of falls occurrences of older people the medical staff (the care and rehabilitation team, attending physicians, and family within the institutions) have the responsibility for this the permanence of care takes place manually which are continuously (at any time, day, night and weekend). The role of the Residential care homes for the elderly in the management of falls in the elderly begins during the admission visit to the institution. Indeed, during the pre-admission visit, the doctor must imperatively identify people with a high risk of falling, with the aim of better managing the resident by developing a specific individual care plan and ensuring the adequacy of the means implemented by the institution that takes a lot of manpower between the different members of the health care team, namely: the attending physician, physiotherapists, occupational therapists, nurses, care assistants, psychotherapists, and even the family. thus we tend to use Artificial Intelligent to the top falls risks of older folks. Nowadays, several works have been done in the field of prevention of falls in the elderly. Alwan et al [130] proposed to use ground vibrations to detect falls. This idea was successfully taken up by Zigel et al [196] by adding a sound detector (combination of an accelerometer and a microphone). However, they admit that their system is not sensitive enough for slow falls and falls from a chair. Furthermore, this type of method is dependent on the dynamics of the floor (different dynamics for a wooden floor or a carpet) and is still in its infancy. The detection of a person on the floor could also be detected with a floor pressure-sensitive paving [93], but this type of technology is difficult to implement in an apartment and is certainly expensive. Lara et al [111] provided a survey of different Methods in Human Activity Recognition using Wearable Sensors, The aim of this work is to compare Twenty-eight systems are qualitatively compared in regards to response time, flexibility, recognition accuracy, obtrusiveness, learning approach, and other important design issues, Also, it surveys the state of the art in on Human Activity Recognition using Wearable Sensors. Finally, it

discusses the main characteristics and challenges of the field and introduces some of the most relevant open problems in the field providing directions for future research. other surveys on Human Activity Recognition using Wearable Sensors provided by Turaga et al [178], Ronald [147], AGGARWAL et al [13], Shoaib et al [157], Wang et al [187], and Akin et al [14] to survey the different approaches for activity recognition using Wearable sensors, with a focus on applications in health care. Asanga et al [186] propose a dynamic sensor data augmentation algorithm that reduces online interpolation errors to facilitate interpolating sparse acceleration data streams from a passive RFID tag with an onboard accelerometer sensor to evaluate activity recognition performance and time taken to interpolate streaming sensor data collected from older volunteers (66-86 years) wearing a passive sensor-enabled RFID to tag embedded with an accelerometer over their attire in a clinical environment. this study provides a basis for many ubiquitous research and applications based on real-time activity recognition using wearable passive sensors, especially passive sensors enabled RFID tags. Roberto et al [153] present and evaluate a novel method for mitigating the high falls risk associated with bed exits based on using an inexpensive, privacy-preserving, and passive sensor-enabled RFID device. this approach is based on a classification system built upon conditional random fields that require no preprocessing of sensorial and RF metrics data extracted from an RFID platform. they evaluated his classification algorithm and the wearability of his sensor using elderly volunteers. The results demonstrate the validity of his approach and the performance is an improvement on previous bed exit classification studies. Aimilia et al [148] propose a system recognizing basic physical activities from wearable sensors, for activity recognition for elderly people. The classification was performed by standard machine learning, as well as deep learning techniques, exploiting feature extraction, along with heuristic computational solutions to address the challenges due to inconsistent measurements in non-standardized environments. In addition, they compare the customized pipeline with deep learning architectures, such as convolutional neural networks, applied to raw sensor data without any pre-or post-processing adjustments. The results o this work demonstrate that the generalizable deep architectures can compensate for inconsistencies during data acquisition providing a valuable alternative. Overall promising results support the use of the proposed activity recognition scheme for unobtrusive monitoring of elderly people. Mohammad et al [92] propose an effective end-to-end deep neural network model to recognize human activities from temporally sparse data signals of passive wearable sensors that improve the accuracy rate of human activity recognition. A dropout technique is used in the developed model to deal with the sparsity problem and avoiding the overfitting problem. In addition, optimization of the proposed deep neural network model was performed by evaluating a different number of hidden layers. Various experiments were conducted on a public clinical room dataset of sparse data signals to compare the performance of the proposed deep neural network model with the conventions and other deep learning approaches.

The experimental results demonstrate that the proposed deep neural network model outperforms the existing state-of-the-art methods in terms of lower inference delay and activity recognition accuracy. Roberto et al [153] develop a graphical model-based classification technique (conditional random field) to evaluate various sliding window-based techniques for the real-time prediction of activities in older subjects wearing a passive (batteryless) sensor-enabled RFID tag. The system achieved maximum overall real-time activity prediction accuracy of 95% using a time-weighted windowing technique to aggregate contextual information to input sensor data. present and evaluate a method to recognize the gesture of someone leaving bed using an RFID device. they use a classification approach in his system to conduct the experiment. The method that we are using is Multi-Layer Perceptron. Roberto et al [158] developed a movement monitoring sensor system that recognizes bed and chair exits. The system consists of a machine learning-based activity classifier and a bed and chair exit recognition process based on an activity score function. Saif et al [131] propose a self-attention-based neural network model that foregoes recurrent architectures and utilizes different types of attention mechanisms to generate higher-dimensional feature representation used for classification. they performed extensive experiments on four popular publicly available Human Activity Recognition datasets: PAMAP2, Opportunity, Skoda, and USC-HAD. his model achieves significant performance improvement over recent state-of-the-art models in both benchmark test subjects and Leave-one-subject-out evaluation. Huaijun et al [188] propose a deep learning-based scheme that can recognize both specific activities and the transitions between two different activities of short duration and low frequency for health care applications. In this work, they first build a deep convolutional neural network (CNN) for extracting features from the data collected by sensors. Then, the long short-term memory (LTSM) network is used to capture long-term dependencies between two actions to further improve the Human activity recognition identification rate. By combing CNN and LSTM, a wearable sensor-based model is proposed that can accurately recognize activities and their transitions. The experimental results show that the proposed approach can help improve the recognition rate up to 95.87% and the recognition rate for transitions higher than 80%, which are better than those of most existing similar models over the open HAPT dataset, Yang et al [191] designed an IoT-based medical cloud storage system with self-adaptive access control is proposed.

4.3 Proposed System

As shown in the Figure 4.1, we propose a real-time system for Recognising Activities in healthy older people. Here we gather raw data from wearable Body Worn Passive Sensors for Recognising Activities classification in the elderly. By using this model, we can be mitigating the risk of falls near beds. the radio frequency identification (RFID) technology with integrated kinematic sensors is able to Recognise Activities .the data set base is created

which includes the sequential motion data from 14 healthy older people aged 66 to 86 years old using a batteryless, wearable sensor on top of their clothing for the recognition of activities in clinical environments, To detect the risk of falls near beds we use the Data Mining and meta-heuristic technique, it examines the existing database in order to generate the required information for making the decision based on the majority results. Once the test data is given to the model and the model gives the Prediction of activities.

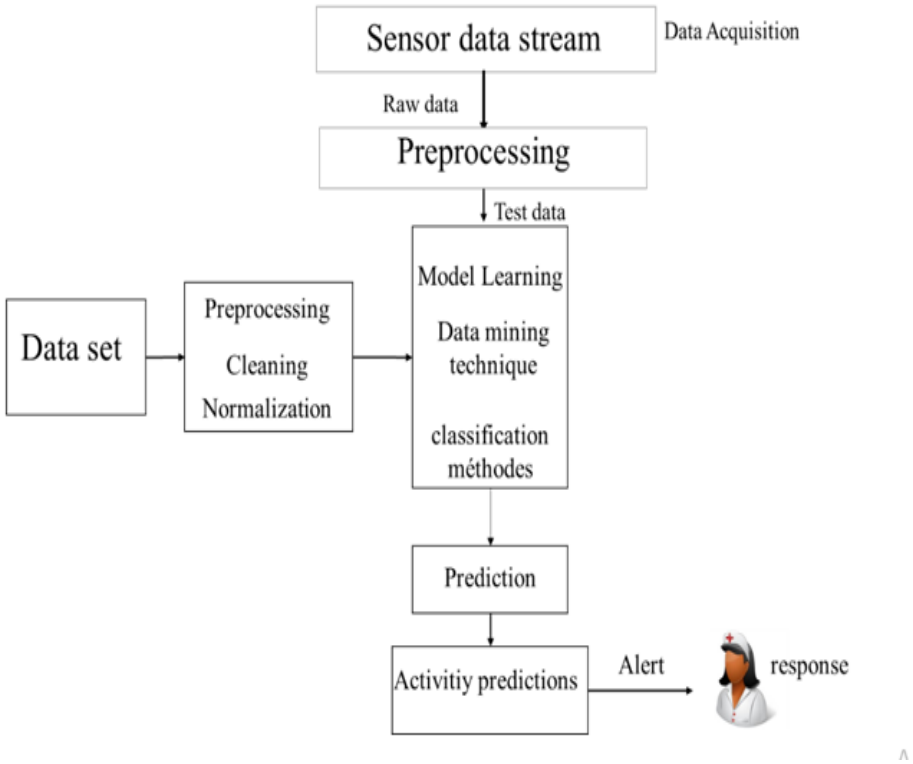

Figure 4.1: Proposed System for Recognising Activities in healthy older people.

4.3.1 Classification methods

Data classification is a dominant technology that is the foundation for the intelligent identification and management of huge information within the internet of things. The data mining tools predict future behaviors, knowledge-driven decisions, trends. Data classification is a well-known problem in the scientific world because it is at the origin of many applications, It is also known under various names (classification, clustering, segmentation) depending on the objects it deals with and the objectives it aims to achieve. Nowadays, it is used in many different fields, such as medical, industrial, security, etc. In most cases, except those for which classification is obvious and trivial, the classification task still depends on human expertise based on observations. For example, a physician diagnoses diseases based on years of medical experience and practice. Similarly, a botanist identifies plant species, a psychologist recognizes personality types. Classification

techniques in data mining are capable of processing a large amount of data. Data mining makes classification models by using already classified data and finds the predicted pattern. The classification problems are used to identify the features of the group in each case of class. Data mining can generate discover information and a large number of rules. Therefore can be used for classifying unseen available data [27].

Classification is the process of analyzing and organizing a set of data, according to their characteristics, into similarity classes. It is mainly based on classical representations of data whose processing limits are known and, in most cases, require a huge computation time. The classification methods can be grouped into two main families, this time, we take into consideration the intervention or not of a "class attribute" during the classification process, these two types are: "supervised (Classification)" and "unsupervised (Classification, Clustering)".

1. supervised (classification): fixed groups, examples of objects in each group.
2. unsupervised (classification): no group is known.

what concerns us is the supervised classification
supervised classification is a method that consists of defining a function that assigns one or more classes to each data. In this approach it is assumed that an expert provides labels for each data beforehand, the labels are classes of membership. among the algorithms of the supervised method.
Among the most popular supervised classification methods, we can mention for example example :

1. k-nearest neighbors
2. Neural networks
3. Decision trees
4. Genetic algorithms
5. Naive Bayes

K-NN (K-Nearest Neighbor)

The k-nearest neighbor algorithm is used in many computational problems including pattern recognition, multimedia data search, vector compression, computer statistics, and data mining. This type of method is widely used in algorithmics and several authors have developed efficient algorithms to solve it. It differs essentially from other methods in its simplicity and in the fact that no models are introduced from the examples during the classification process. The nearest neighbors method consists in determining for each new individual that we want to classify, the list of nearest neighbors among the individuals already classified. The individual is assigned to the class that contains the most individuals among its nearest neighbors. This

method requires choosing a distance, the most classical is the Euclidean distance, and the number of neighbors to take into account. This supervised and non-parametric method is often efficient. Moreover, its learning is quite simple, because it is of the rote learning type (we keep all the learning examples). However, the prediction time is very long, as it requires computing the distance to all examples, but there are heuristics to reduce the number of examples to be considered.

Naive Bayes

Naive Bayes Classification A simple linear probabilistic classifier based on Bayes' theorem assumes that the descriptors (attributes) that describe the objects in the training set are independent. Naive Bayes algorithms are often used in document categorization and classification. They estimate the probability of each class among the examples, given a document, and assign the most probable class to it. This process is called "Prior probabilities". . it is a classification algorithm that assigns class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from the same finite set.

Decision trees

Decision trees are considered among the most popular classification methods. Among the best-known algorithms are ID3 and C4.5. The operation of decision trees is mainly based on examples. Indeed, if we want to classify documents into categories, we have to build a decision tree by category. Generally speaking, a decision tree is a hierarchical sequence of logical rules that divide the database of examples into subgroups, depending on the value of the input variables. The tree is built by searching at each level for the most discriminating parameter to classify an example. Different algorithms can be used to develop the tree structure. Decision trees are very popular in data mining because they are simple and fast while restoring in a comprehensible way the existing relations between the input variables and the phenomenon to model.

4.3.2 Data acquisition

Person-worn sensors Typical methods of fall detection rely primarily on person-worn sensors or push buttons. Alert buttons allow the elderly person to call for help if they are in trouble. In the case of a fall, this type of technology is only effective if the person is conscious after the fall and is not immobilized or unable to operate a button. In our work, we proposed to use Automatic fall detection sensors because they do not require human intervention. Some are based on accelerometers that detect the magnitude and direction of the acceleration; we gather raw data from wearable Body Worn Passive Sensors for recognising activities in the elderly and to be able to detect a fall very early before they impact on the ground.

4.3.3 Data set

	Sitting on bed	Lying on bed	Ambulating,	standing, walking around the room
room (S1)	15162	30983	1956	4381
room (S2)	1253	20529	334	530

Table 4.1: Distribution of activities in data set [186].

We used two data sets described in [176][177] that are collected from 14 healthy older-aged between 66 and 86 years old, performed broadly scripted activities using a batteryless, wearable sensor on top of their clothing at sternum level. To conduct our experiments. These data sets include activities: i) walking to the chair; ii) sitting on the chair; iii) getting off the chair; iv) walking to bed; V) lying on a bed; VI) getting off the bed; VII) walking to the door.

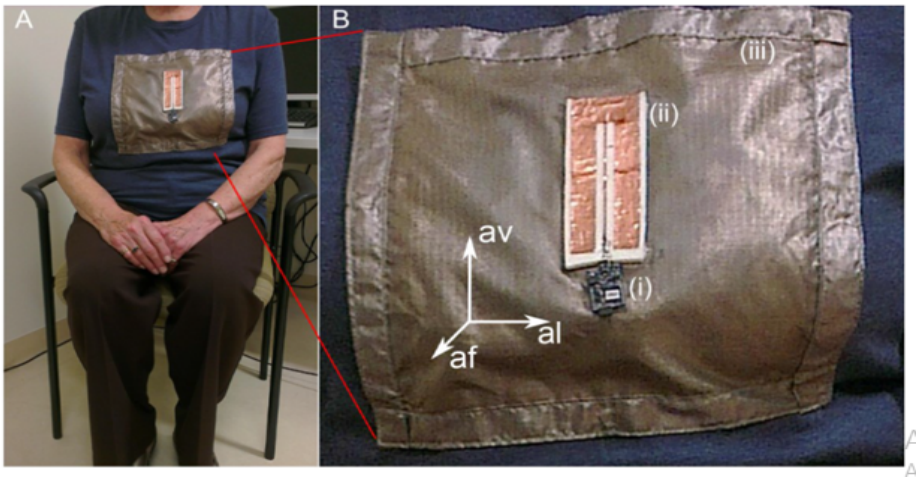

Figure 4.2: A participant wearing the W2ISP at the sternum level over the garment [158].

These data sets were collected in two clinical room settings (S1 and S2) (Figure 4.3). The setting of S1 (Room1) was equipped with four RFID reader antennas, around the room one is placed on top of the bed attached to the ceiling, and the rest are placed on the walls in a manner that they illuminate the area near the bed and the chair as shown in Figure 4.3. whereas the room setting S2 (Room2) uses 3 RFID reader antennas two of them were placed on top of the bed to provide better illumination of the sensor tag while the participant is getting out of the bed and the other was placed on the wall in front of the chair for the collection of motion data [186].

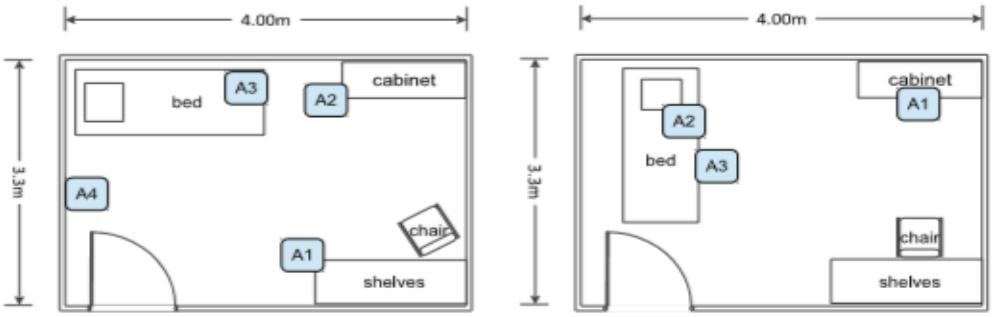

Figure 4.3: Configurations of the antenna placements for the two clinical rooms used to collect data [186]

4.3.4 Data preparation

In the classification technique, the training set is required to build a model. This training set situated holds a bunch of attributes with one quality being the characteristic of the class. Then, the built model is utilized to classify a case.

For this experiment, there are more than two hundred thousand raw data available. This data comes from combination of 9 building parameters with each parameter has 4 possible values (4^9 data).

The parameters and the qualities utilized in every parameter is as follows:

1. Time in seconds.
2. Acceleration reading in G for frontal axis.
3. Acceleration reading in G for vertical axis.
4. Acceleration reading in G for lateral axis.
5. Id of antenna reading sensor.
6. Received signal strength indicator (RSSI).
7. Phase.
8. Frequency.
9. Label of activity, 1: sit on bed, 2: sit on chair, 3: lying, 4: ambulating

Since the data is very large, an illustrative training set should be chosen. Other than that It is necessary to keep the training set as small as feasible. In light of the above contemplations, 4 candidate training sets made. They are with various numbers of data.

The candidate training sets are (Figure 4.4):

Training set 1: 20000 data

Training set 2: 5000 data

Training set 3 : 37500 data

Training set 4: 2500 data

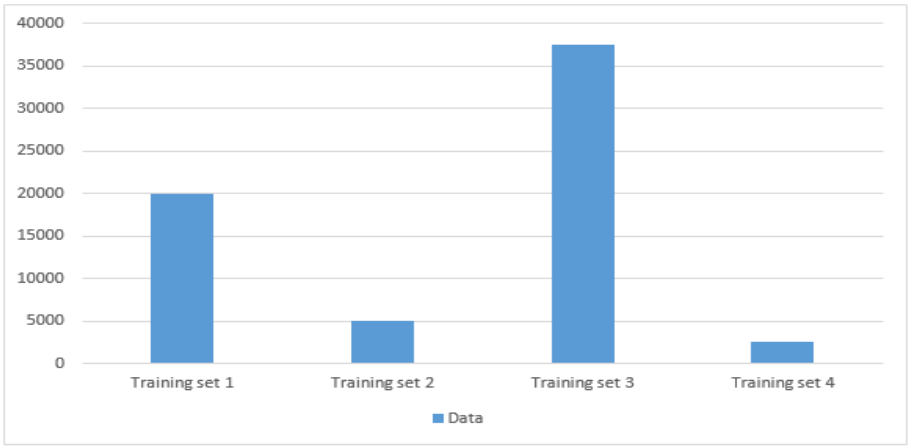

Figure 4.4: The candidate training sets.

An experiment using the three classifiers (Naive Bayesian algorithms, KNN and Decision Tree) is conducted to determine the optimal training set. The experiment is done by means of python 3.6. The results are depicted in Figure 4.5, 4.6 and 4.7

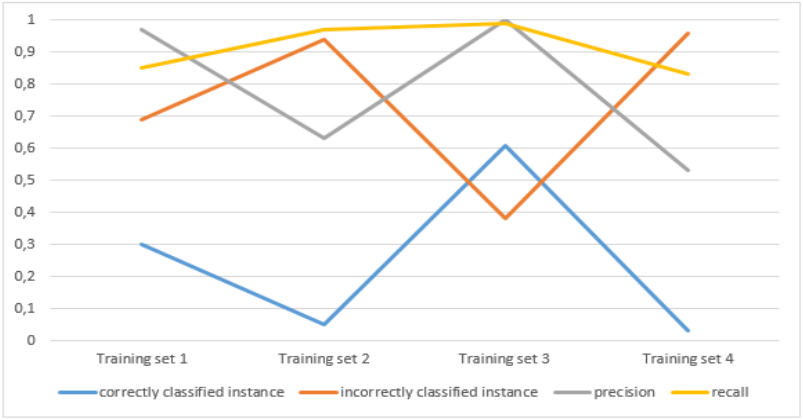

Figure 4.5: Nave Bayes performance on different training sets.

Figure 4.5 shows performance of Naive Bayes methods using the four training sets. The classifier shows the best performance when using training sets 3. Naive Bayes performance has better precision when using training set 3. Figure 4.6 shows performance of k-NN classifier using the same training

sets. k-NN performs best when using training set 3. This is shown by the highest correctly classified instance and precision, and the lowest incorrectly classified instance.
Meanwhile, Figure 4.7 shows The Decision Tree classifier shows the best performance when using training sets 3.
From this result, training set 3 is chosen as the working training set.

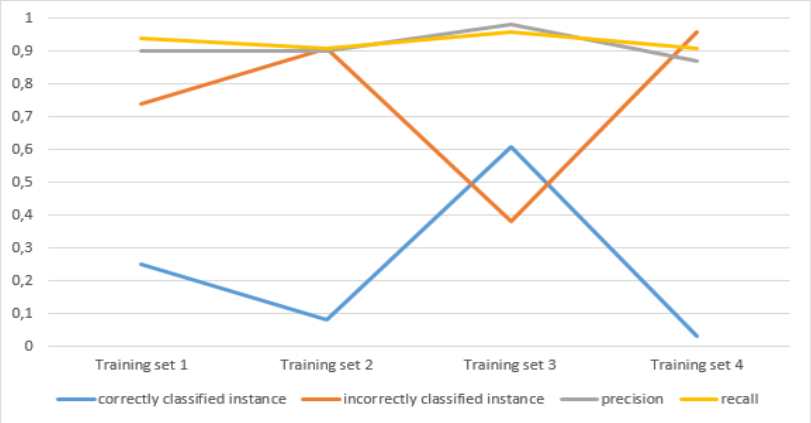

Figure 4.6: k-NN performance on different training sets.

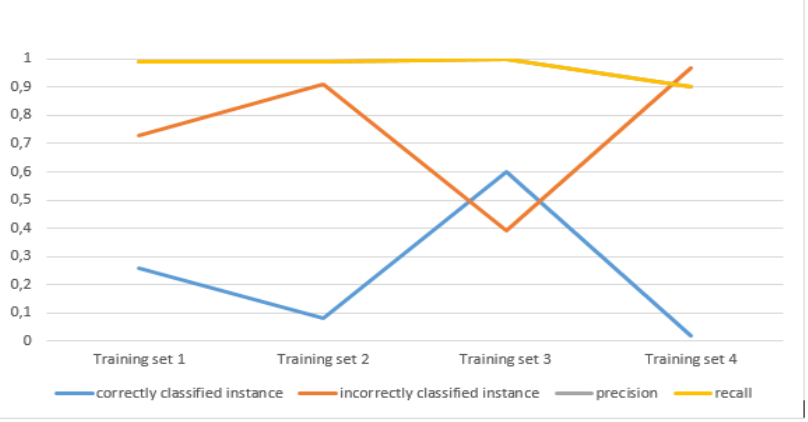

Figure 4.7: Decision Tree performance on different training sets.

4.3.5 Prediction

We will focus on classification models that use labeled data to predict to which class an object belongs. We will mainly talk about binary classification, where the goal is to distinguish whether an object belongs to a class or not.

4.3.6 Alarm

Our system for Recognising Activities in older people should be able to respond to specific events. Once a person is detected is at risk of falling

then an alarm will be triggered. The goal is to tell the care assistants or users of this system that they are in a situation with an abnormal event and that you have to intervene by following in that person's footsteps or trying to Help that person.

4.4 Exprimentation and Results

4.4.1 Expriment

The objective of our experiments is is the use of Artificial Intelligent To combat falls risks of older folks and to enable these people to continue their lives in the best possible conditions. we present a methodology based on multimodal sensors to configure a simple, comfortable and fast fall detection and human activity recognition. In this work, we check whether the traditional data mining algorithms (Naive Bayesian, KNN, Decision Tree,) would likewise work for the IoT datasets, or new families of data mining are required. To this end, We used real sensor datasets from the UCI data repository described in [176][177]. Datasets are collected by using sensors and accelerometers and are used to classify human activities. Before re-enacting the figures, we preprocessed the datasets to make them reasonable for the classifiers. This is a preliminary analysis and hence, we have only used partial datasets. Our experimental methodology is depicted in Figure 4.8.

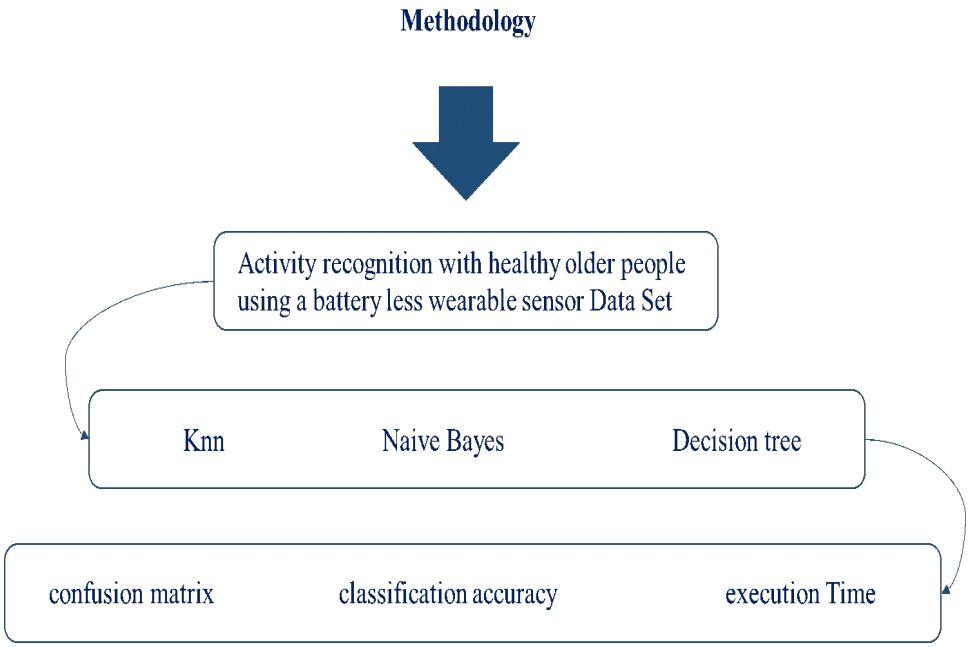

Figure 4.8: Experimental Methodology.

4.4.2 Results and Analysis

This section presents the experimental results and analysis of this study that utilize three classifiers Naive Bayesian, KNN Decision Tree. For the implementation and experimentation of different data mining algorithms to exploit them in the context of the IoT domain for the recognition physical activity of older people in real-time using sensors we made a comparative study in terms of the quality of the results of different classical learning algorithms that have been applied to human activity recognition dataset. A summary of our experiments on the three selected algorithms is given below. The F-Measure is one of the most robust metrics and most used for the evaluation of classification; The F-measure is a combination of Recall and precision. F-Measures an extrinsic evaluation in the beginning, and continue with an intrinsic evaluation: So this is a hybrid evaluation. The classification times of the three classifiers that are used to classify data are shown in Figure 4.9 This figure shows that k-Nearest Neighbor has the fastest classification time followed by Naive Bayes and Decision Tree.

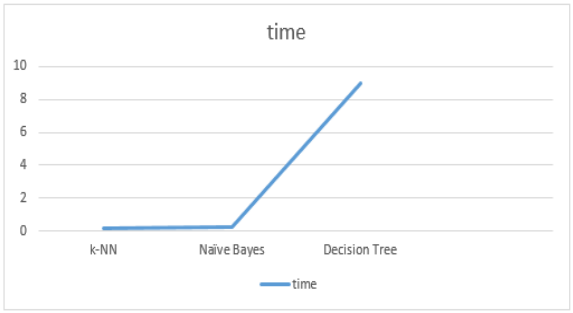

Figure 4.9: Classification times of Decision Tree, k-Nearest Neighbor and Naive Bayes.

The average precisions and recalls for Decision Tree, Nave Bayes, and k-Nearest Neighbor are: 0.92 and 0.78; 0.91 and 0.93; 0.91 and 0.93 respectively (Figure 4.10).

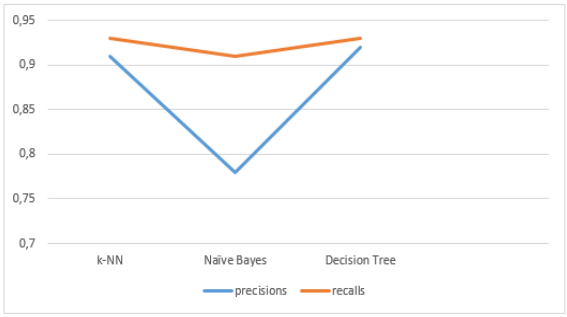

Figure 4.10: Classification precisions and recalls for Decision Tree, Naive Bayes, and k-Nearest Neighbor.

Since F-measure is the harmonic mean of precision and recall, hence to know which classifier is the best in terms of precision and recall, we can calculate the F-measure value (Figure 4.11).

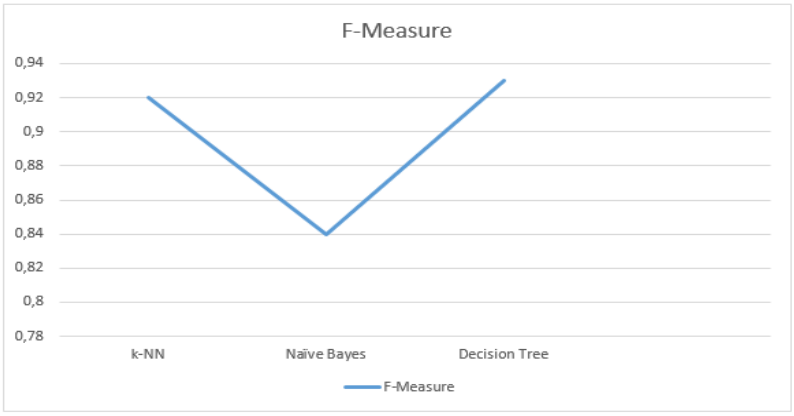

Figure 4.11: Classification F-measure for Decision Tree, Naive Bayes, and k-Nearest Neighbor.

The average F-measure value of Decision Tree is the biggest among the three, that is 0.93.k-NN has average F-measure of 0.92 and Naive Bayesian of 0.84. Therefore we can say that Decision Tree is the best in terms of precision and recall followed by Naive Bayesian and kNN.

Decision Tree is again the best in accuracy (Figure 4.12). Decision Tree is the most accurate classifier compared to Nave Bayes and k-NN with the average accuracy of 0.99 . Meanwhile the average accuracies of k-NN and Naive Bayesare 0.95 and 0.94, respectively.

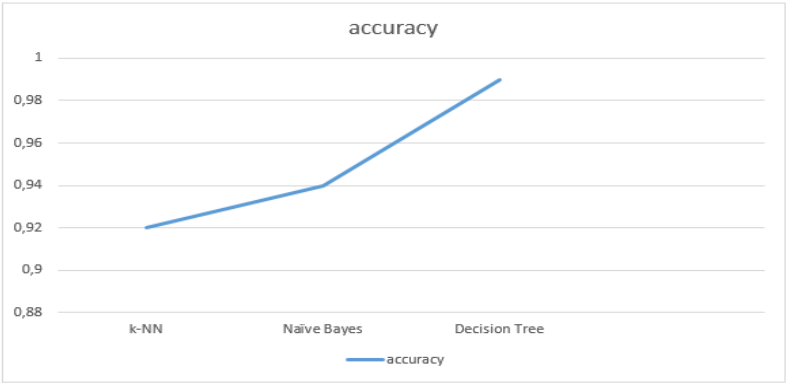

Figure 4.12: Classification accuracy for Decision Tree, Naive Bayes, and k-Nearest Neighbor.

4.4.3 Comparative study and discussion

		Recall	Precision	F-Measure	Accuracy	Confusion matrix		
							U	N
						U	VP	FP
						N	FN	VN
Classifiers	k-NN	093	0.91	0.92	0.95	10981		14
						35		148
	Naïve Bayesian	0.91	0.78	0.84	0.94	11177		0
						0		1
	Decision Tree	0.93	0.92	0.93	0.99	11172		1
						0		5

Table 4.2: Comparative study in terms of the quality of results of different classifiers (Naive Bayesian algorithms, KNN and Decision Tree algorithms) regarding Precision, Recall and F-Measure.

Table 4.2 presents the overall accuracy (Precision), detection rate (Recall), and F-Measure For the data mining algorithms (Naive Bayesian algorithms; KNN and Decision Tree algorithms), According to the table and research the execution time taken by the Naive Bayes is least with 0.02 seconds followed by Knn with 0.02 seconds, and Decision Tree algorithm took much more time for execution which is 8.97 seconds. The accuracy of the Decision Tree algorithm is 92%, KNN with 91%, and naive Bayes with 78%. the decision tree algorithm gave the best Prediction rate. It can be seen clearly that the decision tree algorithm performs better than Naive Bayes and KNN. It achieving a Prediction rate of 93%. KNN also had the best kappa statistic at 92%. From our experiments, the decision tree algorithm appears to be the best suited for real-time classification tasks due to its relatively fast classification speed and high detection rate.

4.5 Conclusion

The IoT paradigm brings new sets of useful and valuable data mainly collected from sensor devices. To capture this hidden knowledge from IoT data is a challenging task in data mining that the major challenge in these settings is the timely analysis of large amounts of data (big data) to produce decisions and highly reliable and accurate insights so that IoT could satisfy its guarantee. Hence needs data computation for the prediction of the unknown data. Data mining would play a critical role in creating smarter IoT; the various data-mining algorithm can be applied to IoT data Some researchers argue that a new family of data mining algorithms is needed to handle IoT data.

In our work, we examined the applicability of some of the well-established data mining algorithms including Naive Bayesian algorithms, KNN, and Decision Tree algorithms. With our preliminary analysis, we conclude that KNN and Decision Tree algorithms can give relatively higher accuracy results. We plan to conduct a detailed study on larger and diverse IoT datasets in the future.

CONCLUSION AND FUTURE PERSPECTIVES

The Internet of Things (IoT) is a new concept that combines autonomous and intelligent machines, machine-human collaboration to improve productivity and advanced predictive analytics, reliability, and efficiency. IoT provides a world where smart, connected, embedded systems and products operate as part of larger systems, giving rise to the amount of enormous data from the different sources device of the IoT are considered to have highly useful and valuable information. The Internet of Things (IoT) is helping create a safer world by enabling all organizations to securely and remotely monitor facilities and spaces in real-time with smart security and surveillance and protection solutions. The IoT is closely linked to the notion of big data, the objects of which generate a large quantity and a variety of data in real-time, known as big data. So, each needs the other to make it useful. There is no IoT without big data, and big data reaches the highest position when used for IoT. Data mining aims to uncover valuable information in large amounts of data that can help understand the data or predict the behavior of future data. The underlying idea of Data Mining is therefore to extract hidden knowledge from a pile of available data.

In this these: Data mining and meta-heuristic would play a critical role in creating smarter IoT; various data-mining algorithms and Metaheuristic Algorithms can be applied to IoT data. This work aims to focused on analyzing techniques that are used for prediction purposes. The parameters detection is gotten by the utilization of IP cameras and sensors and is kept up as datasets. Internet of Things (IoT) is used primarily to gather data from the user. The precision of data could be at stake since the sensor may glitch during the collection process.

In our work, we tried and succeed to deliver two systems, the first system is an Recognizing Physical Activity of hospitalized Older People from Wearable Sensors Data using IoT system. we examine the applicability of three data mining algorithms for real IoT datasets. These include K-nn, Naive Bayes,

Decision tree The main contribution of this work is the analysis of the efficiency and effectiveness of three of the data mining. the second system is Border and sensitive sites Surveillance Systems in real-time based on IoT, In recent years, the security of borders and sensitive sites such as tourist sites has undeniably become a major global issue. The massive influx of refugees, the fight against terrorism, illegal immigration and smuggling, the strengthening of internal security, and the rise of cybercrime are prompting governments to strengthen their border and sensitive site surveillance systems. Governments are investing more and more to ensure their protection through the use of electronic surveillance technology and video surveillance systems. Our system based on the classification of the human gestures drawn from videos envoy by Drones equipped with cameras and sensors that are deployed on the border by the use of a Bio-Inspired technique of Grouping Cockroaches Classifier (GCC) based on the style of life Cockroaches and operate on the phenomenon of seeking the most attractive and secure place (shelter) for hiding for a good detected the gesture of unwanted individuals this algorithm is based on a learning base and classify the gestures of the test base and labels them, each gesture take one of two classes (gestures of border soldiers and gesture of terrorists and non-soldier people), and we apply also a new technology for the presentation of picture (n-grams pixels) to construct a system of control of borders. Our objective is to use drones instead of normal soldiers to cover the space of the borders, detect terrorists hiding their faces, detect people in disguise; react effectively and faster, react at night, or even when the climate is difficult.

Experimental results prove the benefit derived from the use of our system and therefore it enables the border troops to surveillance the borders at each and every moment to effectively and at a low cost.

Finally, we are convinced that in a world where the security of citizens is the code of life, our proposed Border and sensitive sites Surveillance Systems represents a solution to design a modern camera surveillance policy since it meets the requirements of modern users by respecting people's image rights, it can report suspicious people even if they hide their faces, help security guards and police officers in their work by notification in real-time.

Future works

Our next work will be specifically related to the IoT industry, making our models more precise and having a scalable architecture , a few new research doors have been opened which are detailed below:

We are thinking of extending our results to other types of movements, and to scale our model for online learning, rather than an offline supervised learn-

ing, and that can increase the possibility for more insight making our model more precise and having a scalable architecture, as we know, there is a big need for surveillance systems. IoT is expected to be the main change in the field of security surveillance. The ongoing process is extending our results to other types of characteristics of suspicious people such as voice recognition of the suspect person, his favorite place. We can also store videos of malicious people to identify them by their gestures, even if they have changed their faces.

Mobile version of our work: We plan to develop a mobile version of Border and sensitive sites Surveillance Systems for operating systems such as Android and iOS, in order to make it easier to control at distance.

Improve the process of detecting suspicious people, by integrating the map to track the movement of dangerous people to stop them. We can integrate other characteristics of suspect people such as their friend's places and their clothes. The system can detect the voice and certain gestures and recognize them. Currently, our system can identify a single suspicious person, but what we want is to improve it so that it can identify suspicious gestures between a group of people.

adding a 2d and 3d visualization part of the results of each monitored area to help authorities place police officers in areas where there are more suspicious people.

LIST OF PUBLICATIONS

International publication

- Boukhalfa, S., Amine, A., Hamou, R. M. (2022). Border Security and Surveillance System Using IoT. International Journal of Information Retrieval Research (IJIRR), 12(1), 1-21. http://doi.org/10.4018/IJIRR.289953

- Boukhalfa, S., Amine, A., Hamou, R. M. (2022). Recognizing the Physical Activity of Hospitalized Older People From Wearable Sensors Data Using IoT. International Journal of Organizational and Collective Intelligence (IJOCI), 12(1), 1-19. http://doi.org/10.4018/IJOCI.2022010104

International communications

- Boukhalfa, S., Amine, A., Hamou, R. M. The Internet of Things: A survey, Doctorial Symposium of 6 the IFIP International Conference on Computational Intelligence and Its Applications (IFIP DS CIIA 2018).

- Boukhalfa, S., Amine, A., Hamou, R. M. A survey: Internet of Things (IOT) technologies, architectural and challenges, the 3 rd International Symposium on Informatics and its Applications (ISIA 2018).

- Boukhalfa, S., Amine, A., Hamou, R. M. A comparative study of classification algorithms and meta-heuristic in IoT, INTIS 2019 The 8th International Conference on Innovation and New Trends in Information Technology - December 20 - 21, 2019 - Tangier, Morocco.

National communications

- Boukhalfa, S., Amine, A., Hamou, R. M. Novel Driver chronic disease detection and prevention models using Internet of things and Data Mining Techniques accept to JERI'2020 (4th Edition of the National Study Days on Research on Computer Sciences.

BIBLIOGRAPHY

[1] Arfaoui, I., Boudriga, N., Trimeche, K., Abdallah, W. (2017, December). WSN-based Border Surveillance Systems Using Estimated Known Crossing Paths. In Proceedings of the 15th International Conference on Advances in Mobile Computing Multimedia (pp. 182-190).

[2] Al Abkal, S., Talas, R. H. A., Shaw, S., Ellis, T. (2020). The application of unmanned aerial vehicles in managing port and border security in the US and Kuwait: Reflections on best practice for the UK. International Journal of Maritime Crime and Security, 1(1).

[3] Arjun, D., Indukala, P. K., and Menon, K. U. (2017, April). Border surveillance and intruder detection using wireless sensor networks: A brief survey. In 2017 International Conference on Communication and Signal Processing (ICCSP) (pp. 1125-1130). IEEE.

[4] Agard, B., and Kusiak, A. (2005, April). Exploration des bases de donnes industrielles l'aide du datamining-Perspectives. In 9me colloque national AIP PRIMECA.

[5] Alaoui, A. (2012). Application des techniques des mtaheuristiques pour l'optimisation de la tche de la classification de la fouille de donnes (Doctoral dissertation, USTO).

[6] Adiba, M., Castrejon-Castillo, J. C., Oviedo, J. A. E., Vargas-Solar, G., and Zechinelli-Martini, J. L. (2016). Big data management challenges, approaches, tools and their limitations.

[7] AMRANE, A. (2015). Big Data: Concepts et Cas d'utilisation (No. CERIST-DSISM/RR-17-00000009–DZ). CERIST.

[8] Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: A survey. Computer networks, 54(15), 2787-2805.

[9] Arroyo, R., Yebes, J. J., Bergasa, L. M., Daza, I. G., and Almazn, J. (2015). Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls. Expert systems with Applications, 42(21), 7991-8005.

[10] Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., and Ayyash, M. (2015). Internet of things: A survey on enabling technologies, protocols, and applications. IEEE communications surveys and tutorials, 17(4), 2347-2376.

[11] abate, m. B. (2019). Application of a multi criteria decision model of entropy, topsis and linear programming techniques for building construction contractor selection in ethiopia.

[12] Abdel-Basset, M., Manogaran, G., Gamal, A., and Chang, V. (2019). A novel intelligent medical decision support model based on soft computing and IoT. IEEE Internet of Things Journal, 7(5), 4160-4170.

[13] Aggarwal, J. K., and Ryoo, M. S. (2011). Human activity analysis: A review. ACM Computing Surveys (CSUR), 43(3), 1-43.

[14] Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R., and Havinga, P. (2010, February). Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey. In 23th International conference on architecture of computing systems 2010 (pp. 1-10). VDE.

[15] Kaleche, R., Bendaoud, Z., and Bouamrane, K. (2020). Bio-Inspired Metaheuristics: A Comprehensive Survey. International Journal of Organizational and Collective Intelligence (IJOCI), 10(4), 1-18.

[16] Bhadwal, N., Madaan, V., Agrawal, P., Shukla, A., and Kakran, A. (2019). Smart Border Surveillance System using Wireless Sensor Network and Computer Vision. 2019 International Conference on Automation, Computational and Technology Management (ICACTM), 183-190.

[17] Bell, W. J., Roth, L. M., and Nalepa, C. A. (2007). Cockroaches: ecology, behavior, and natural history. JHU Press

[18] Bhaskar, H. (2012, September). Integrated human target detection, identification and tracking for surveillance applications. In 2012 6th IEEE International Conference Intelligent Systems (pp. 467-475). IEEE.

[19] Benghozi, P. J., Bureau, S., and Massit-Folea, F. (2008). L'Internet des objets. Quels enjeux pour les Europens?.

[20] Bendiab, E., Meshoul, S., and Batouche, M. (2003, September). An artificial immune system for multimodality image alignment. In International Conference on Artificial Immune Systems (pp. 11-21). Springer, Berlin, Heidelberg.

[21] Boudia mohamed amine. (2017). Optimisation, intgration des donnes et dcouverte de connaissances partir des donnes du web, methaheuristique, saida

[22] Berry, M. J., and Linoff, G. S. (2004). Data mining techniques: for marketing, sales, and customer relationship management. John Wiley and Sons.

[23] Bradji, L. (2012). Adaptation des techniques de l'Extraction des Connaissances partir des Donnes (ECD) pour prendre en charge la qualit des donnes. (Doctoral dissertation, Universit de Mentouri. Constantine).

[24] BONABEAU .E et THERAULAZ .G (1994), "Intelligence Collective". Edition Hermes Sciences, Paris, France.

[25] BONABEAU .E, DORIGO, M. and THERAULAZ .G (1999). "Swarm Intelligence: From Natural to Artificial Systems". Oxford University Press, New York.

[26] Borgia, E. (2014). The Internet of Things vision: Key features, applications and open issues. Computer Communications, 54, 1-31

[27] BOUKHALFA, S., AMINE, A., and HAMOU, R M. (2019). A comparative study of classification algorithms and meta-heuristic in IoT. In Proceedings of the 8th International Conference on Innovation and New Trends in Information Technology, INTIS 2019, Tangier, Morocco, Dec 20-21 pp. 280-281.

[28] Beni, G., and Wang, J. (1993). Swarm intelligence in cellular robotic systems. In Robots and biological systems: towards a new bionics? (pp. 703-712). Springer, Berlin, Heidelberg.

[29] Brooking, A. (1998). Corporate memory: Strategies for knowledge management. International Thomson Publishing.

[30] Bouquin, S. (2016). Mise en oeuvre d'un systme de publication/souscription bas sur les flux d'information de type RSS.

[31] Borgi, T., Zoghlami, N., and Abed, M. (2017, January). Big data for transport and logistics: A review. In 2017 International Conference on Advanced Systems and Electric Technologies (IC_ASET) (pp. 44-49). IEEE.

[32] http://www.tst-sistemas.es/en/rd/butler/ (Accessed February 2021).

[33] Chan, J. O. (2013). An architecture for big data analytics. Communications of the IIMA, 13(2), 1.

[34] Chesnot, G. (2012). Cloud computing, Big Data, paralllisme, Hadoop: stockage de donnes du futur. Vuibert.

[35] Chen, M., Mao, S., and Liu, Y. (2014). Big data: A survey. Mobile networks and applications, 19(2), 171-209.

[36] Chaabane, L. A. M. I. C. H. E. (2013). Fusion et fouille de donnes guides par les Connaissances: application l'analyse d'image. Doctorat, Universite Mohamedkhider-biskra.

[37] Collard, P., Clergue, M., Vrel, S. (2009). Introduction aux systmes complexes: Optimisation par essaims particulaires.

[38] Challal, Y. (2012). Scurit de l'Internet des Objets: vers une approche cognitive et systmique (Doctoral dissertation, Universit de Technologie de Compigne).

[39] Choudhari, S., Rasal, T., Suryawanshi, S., Mane, M., and Yedge, S. (2017). Survey paper on internet of things: IoT. International Journal of Engineering Science and Computing, 7(4), 10564-10567.

[40] Cisco: Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017-2022.

[41] CHRISTO. Article "point-cl du 11-Septembre" : Les pirates de l'air ... identifications douteuses.(May 18,2010) (Accessed March 4, 2021). URL : http://www.reopen911.info/11-septembre/les-pirates-section-points-cles/.

[42] Collins, R. T., Lipton, A. J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., ... Wixson, L. (2000). A system for video surveillance and monitoring. VSAM final report, 2000(1-68), 1.

[43] Chouaib, H. (2011). Slection de caractristiques: mthodes et applications. Paris Descartes University: Paris, France.

[44] Cramer, N. L. (1985, June). A representation for the adaptive generation of simple sequential programs. In proceedings of an International Conference on Genetic Algorithms and the Applications (pp. 183-187).

[45] Dasgupta, D., and Gonzlez, F. (2002). An immunity-based technique to characterize intrusions in computer networks. IEEE Transactions on evolutionary computation, 6(3), 281-291.

[46] DJAFRI, L. (2020). Analyse de donnes massives -Big Data- pour laprdictione (Doctoral dissertation, Universit Djillali Liabs de Sidi Bel Abbs).

[47] Dorigo, M., and Gambardella, L. M. (1997). Ant colonies for the travelling salesman problem. biosystems, 43(2), 73-81.

[48] Dumitrescu, D., Lazzerini, B., Jain, L. C., and Dumitrescu, A. (2000). Evolutionary computation. CRC press.

[49] Dantzig, G. B., and Thapa, M. N. (2006). Linear programming 1: introduction. Springer Science and Business Media.

[50] DREO .J, SIARRY .P,(2003) "Diverses Techniques D'optimisation Inspires Inspires de la Thorie de l'Auto-Organisation dans les Systmes Biologiques". Premier sminaire francophone sur le thme de l'Optimisation par Essaim Particulaire (OEP 03).Paris, France 2003

[51] Dorigo, M. (1992). Optimization, learning and natural algorithms. Ph. D. Thesis, Politecnico di Milano.

[52] De Jong., K. (1975) An analysis of the behavior of a class of genetic adaptive systems. PhD thesis, University of Michigan.

[53] L . Davis,(1991) The genetic algorithm handbook, chapter 17. Ed. New York : Van Nostrand Reinhold, 1991.

[54] Erl, T., Khattak, W., and Buhler, P. (2016). Big data fundamentals: concepts, drivers and techniques. Prentice Hall Press.

[55] Emilie, B. (2015). Vers une definition des Big data en sant base sur la littrature (Doctoral dissertation, Universit LILLE 2 DROIT ET SANT).

[56] Fayyad, U. M., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery: an overview. Advances in knowledge discovery and data mining, 1-34.

[57] Frawley, W. J., Piatetsky-Shapiro, G., and Matheus, C. J. (1992). Knowledge discovery in databases: An overview. AI magazine, 13(3), 57-57.

[58] Fan, J., Kalyanpur, A., Gondek, D. C., and Ferrucci, D. A. (2012). Automatic knowledge extraction from documents. IBM Journal of Research and Development, 56(3.4), 5-1.

[59] Fermigier, S. (2012). Big Data and Open Source: une convergence invitable?. Livre Blanc (http://fermigier. com/blog/2012/03/new-whitepaper-big-data-open-source/).

[60] Ferguson, M. (2013). Enterprise Information Protection-The Impact of Big Data. White Paper for IBM.

[61] S. Feng, J. Cerles, H. Dalmas, T. Do-Khac, and B. Paulin (2014). Scurit des objets. Connects. Institut national des hautes tudes de la scurit et de la justice.

[62] Feigenbaum, E. A., and Feldman, J. (1963). Computers and thought. McGraw-Hill: New York.

[63] FFoulkes, P. (2017). Inside BIG DATA guide to the intelligent use of big data on an industrial scale. InsideBIGDATA, Massachusetts.

[64] Fogel, L. J. (1962). Autonomous automata. Industrial research, 4, 14-19.

[65] FAREH .A,(2007). "Optimisation par Essaim de Particules une Mtaheuristique pour l'Optimisation Difficile". Premire Journe Nationale sur les Applications des Mtaheuristiques 2007 (JNAM 07). Alger, Algrie 2007

[66] Fogel, L. J., Owens, A. J., and Walsh, M. J. (1966). Artificial intelligence through simulated evolution.

[67] Fortin D, (2012). Chutes chez les personnes ages de 65 ans et plus vivant domicile. Qubec, Qc : Ministre de la Sant et des Services sociaux (MSSS); 2012. Disponible ?http://publications.msss.gouv.qc.ca/acrobat/f/documentation/2012/12-202-03F.pdf.

[68] Farahani, B., Firouzi, F., Chang, V., Badaroglu, M., Constant, N., and Mankodiya, K. (2018). Towards fog-driven IoT eHealth: Promises and challenges of IoT in medicine and healthcare. Future Generation Computer Systems, 78, 659-676.

[69] Goyal, A., Anandamurthy, S. B., Dash, P., Acharya, S., Bathla, D., Hicks, D., ... and Ranjan, P. (2020). Automatic Border Surveillance Using Machine Learning in Remote Video Surveillance Systems. In Emerging Trends in Electrical, Communications, and Information Technologies (pp. 751-760). Springer, Singapore.

[70] Grinter, B. (2013). A big data confession. Interactions, 20(4), 10-11.

[71] Gantz, J., and Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC Analyze the future, 2007(2012), 1-16.

[72] Gubbi, J., Buyya, R., Marusic, S., and Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. Future generation computer systems, 29(7), 1645-1660.

[73] D. Goldberg. (1989). Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, 1989.

[74] Granjal, J., Monteiro, E., and Silva, J. S. (2015). Security in the integration of low-power Wireless Sensor Networks with the Internet: A survey. Ad Hoc Networks, 24, 264-287.
.

[75] Gambardella, L. M., and Dorigo, M. (1995). Ant-Q: A reinforcement learning approach to the traveling salesman problem. In Machine learning proceedings 1995 (pp. 252-260). Morgan Kaufmann.

[76] Goldberg, D. E., and Holland, J. H. (1988). Genetic Algorithms and Machine Learning. Machine Learning, 3(2), 95-99.

[77] Sophie Bernard. (September 25, 2013). Jrmy GARANDEAU (accessed March 4, 2021). URL : https://www.gentside.com/cambriolage/regardez-comment-ce-bijoutier-sicilien-a-reussi-a-se-defendre-face-a-deux-voleurs-armes-art54903.html.

[78] Global M2 M market. https://machinaresearch.com/news/global-m2m-market-togrow-to-27- billion-devices-generating-usd16-trillionrevenue-in-2024/. Accessed 29 september 2019.

[79] Griffin, D. R. (1958). Listening in the dark Yale Univ. Press, New York.

[80] Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., ... and Pankanti, S. (2005). Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking. IEEE signal processing magazine, 22(2), 38-51.

[81] Haritaoglu, I., Harwood, D., and Davis, L. S. (2000). W/sup 4: real-time surveillance of people and their activities. IEEE Transactions on pattern analysis and machine intelligence, 22(8), 809-830.

[82] He, J., Fallahi, M., Norwood, R. A., and Peyghambarian, N. (2011, June). Smart border: ad-hoc wireless sensor networks for border surveillance. In Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X (Vol. 8019, p. 80190Z). International Society for Optics and Photonics.

[83] Hu, W., Tan, T., Wang, L., and Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(3), 334-352.

[84] Holland, J. H. (1975). Adaptation in natural and artificial systems, univ. of mich. press. Ann Arbor.

[85] Hota, A. K., and Prabhu, D. M. (2012). No problem with Big Data. What do you mean by Big. Journal of Informatics, 30-32.

[86] Hurwitz, J., Nugent, A., Halper, F., and Kaufman, M. (2013). Big data for dummies? Ebook.

[87] Haddad, M. H. (2002). Extraction et Impact des connaissances sur les performances des Systmes de Recherche d'Information (Doctoral dissertation, Universit Joseph-Fourier-Grenoble I).

[88] Hann, J., and Kamber, M. (2000). Data mining: concepts and techniques.

[89] HAN M., ZHANGHANG. (2013), "Business intelligence architecture based on internet of things ", Journal of Theoretical Applied Information Technology, Vol. 50 No.1.

[90] Hunt, J. E., and Cooke, D. E. (1996). Learning using an artificial immune system. Journal of network and computer applications, 19(2), 189-212.

[91] https://vicinity2020.eu/vicinity/content/hydra (Accessed February 2021).

[92] Hassan, M. M., Ullah, S., Hossain, M. S., and Alelaiwi, A. (2020). An end-to-end deep learning model for human activity recognition from highly sparse body sensor data in Internet of Medical Things environment. The Journal of Supercomputing, 1-14.

[93] Institut national d'excellence en sant et en services sociaux (INESSS),(2013). Chutes chez les patients hbergs dans les tablissements de sant (soins de courte et de longue dure). Synthse des connaissances et des rfrences rdige par Elene Morarescu. Qubec, Qc : INESSS; 2013. 23p

[94] http://www.internet-of-things-research.eu/ (Accessed February 2021).

[95] JLASSI, A., MARTINEAU, P., and CLOAREC, E.(2014). Hadoop: Optimisation et Ordonnancement. cole Polytechnique de l'Universit de Tours. Dpartement Informatique 5e anne.

[96] Jing, Q., Vasilakos, A. V., Wan, J., Lu, J., and Qiu, D. (2014). Security of the Internet of Things: perspectives and challenges. Wireless Networks, 20(8), 2481-2501.

[97] Kumar, S., Tiwari, P., and Zymbler, M. (2019). Internet of Things is a revolutionary approach for future technology enhancement: a review. Journal of Big Data, 6(1), 111.

[98] Kim, S. J., and Lim, G. J. (2018). Drone-aided border surveillance with an electrification line battery charging system. Journal of Intelligent and Robotic Systems, 92(3-4), 657-670.

[99] Karthick, R., Prabaharan, A. M., and Selvaprasanth, P. (2019). Internet of things based high security border surveillance strategy. Asian Journal of Applied Science and Technology (AJAST) Volume, 3, 94-100.

[100] Khan, N., Alsaqer, M., Shah, H., Badsha, G., Abbasi, A. A., and Salehian, S. (2018, March). The 10 Vs, issues and challenges of big data. In Proceedings of the 2018 international conference on big data and education (pp. 52-56).

[101] Kayyali, B., Knott, D., and Van Kuiken, S. (2013). The big-data revolution in US health care: Accelerating value and innovation. Mc Kinsey and Company, 2(8), 1-13.

[102] Katal, A., Wazid, M., and Goudar, R. H. (2013, August). Big data: issues, challenges, tools and good practices. In 2013 Sixth international conference on contemporary computing (IC3) (pp. 404-409). IEEE.

[103] Kennedy, J., and Eberhart, R. (1995, November). Particle sw-arm optimization. In Proceedings of ICNN'95-international conference on neural networks (Vol. 4, pp. 1942-1948). IEEE.

[104] Kantardzic, M. (2003). Data mining: concepts models methods and algorithms.

[105] Koza, J. R., and Koza, J. R. (1992). Genetic programming: on the programming of computers by means of natural selection (Vol. 1). MIT press.

[106] Kleene, S. C. (1956). Realization of nerve nets and finite automata. In Automata Studies (pp. 3-41). Princeton Univ. Press.

[107] Koza, J. R. (1994). Genetic programming II: automatic discovery of reusable programs. MIT press.

[108] Karaboga, D. (2005). An idea based on honey bee swarm for numerical optimization (Vol. 200, pp. 1-10). Technical report-tr06, Erciyes university, engineering faculty, computer engineering department.

[109] Laouira, M. L., Abdelli, A., Othman, J. B., and Kim, H. (2019). An efficient WSN based solution for border surveillance. IEEE Transactions on Sustainable Computing.

[110] Layeb, A., and Deneche, A. H. (2007, May). Multiple sequence alignment by immune artificial system. In 2007 IEEE/ACS International Conference on Computer Systems and Applications (pp. 336-342). IEEE.

[111] Lara, O. D., and Labrador, M. A. (2012). A survey on human activity recognition using wearable sensors. IEEE communications surveys and tutorials, 15(3), 1192-1209.

[112] Lemberger, P., Batty, M., Morel, M., and Raffalli, J. L. (2016). Big Data et Machine Learning-2e d. Dunod.

[113] J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

[114] Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). Multivariate Analysis Academic Press Inc. London) Ltd, 15, 518.

[115] MATALLAH, H. (2018). Vers un nouveau modle de stockage et d 'accs aux donnes dans les Big Data et les Cloud Computing (Doctoral dissertation, 19-05-2019).

[116] Mills, S., Lucas, S., Irakliotis, L., Rappa, M., Carlson, T., and Perlowitz, B. (2012). Demystifying big data: a practical guide to transforming the business of government. TechAmerica Foundation, Washington.

[117] Mehmood, Y., Grg, C., Muehleisen, M., and Timm-Giel, A. (2015). Mobile M2M communication architectures, upcoming challenges, applications, and future directions. EURASIP Journal on Wireless Communications and Networking, 2015(1), 1-37.

[118] Monino, J. L., and Sedkaoui, S. (2016). Big Data, Open Data et valorisation des donnes (Vol. 4). ISTE Group.

[119] Menaouer, M. B. (2014). Conception et Exprimentation d'une nouvelle mthode boolenne de cartographie des connaissances guide par data mining (Doctoral dissertation, Universit de Technologie de Troyes).

[120] Murakami, T. (2004). Ubiquitous networking: business opportunities and strategic issues. Nomura Research Institute Papers, (79), 1-12.

[121] Kamilia, M. M. (2015). Approches Bio-inspires pour la Slection d'Attributs (Doctoral dissertation, Universit Badji Mokhtar-Annaba).

[122] Margot, F. (2010). Symmetry in integer linear programming. 50 Years of Integer Programming 1958-2008, 647-686.

[123] Mitchell, M. (1998). An introduction to genetic algorithms. MIT press.

[124] Farhad MANJOO. April 23, 2013 (accessed March 4, 2021). URL : http://www.slate.fr/story/71235/attentats-boston-cameras-videosurveillance.

[125] Montaigne, I. (2015). Big data et objets connects Faire de la France un champion. Rapport, 200.

[126] Mitchell, T. M. (1997). Machine Learning, McGraw-Hill Higher Education. New York.

[127] Mittal, V. (2017). Top 15 deep learning applications that will rule the world in 2018 and beyond. URL: https://medium. com/@ vratulmittal/top-15-deeplearning-applications-that-will-rule-the-world-in-2018- andbeyond-7c6130c43b01.

[128] Z. Michalewicz. (1996). Genetic Algorithms + Data Structures = Evolution Programs. Springer Verlag, third - 1999 edition.

[129] Miake-Lye, I. M., Hempel, S., Ganz, D. A., Shekelle, P. G. (2013). Inpatient fall prevention programs as a patient safety strategy: a systematic review. Annals of internal medicine, $158(5_Part_2), 390 - 396$.

[130] Alwan, M., Rajendran, P. J., Kell, S., Mack, D., Dalal, S., Wolfe, M., Felder, R. (2006, April). A smart and passive floor-vibration based fall detector for elderly. In 2006 2nd International Conference on Information Communication Technologies (Vol. 1, pp. 1003-1007). IEEE.

[131] Mahmud, S., Tonmoy, M., Bhaumik, K. K., Rahman, A. K. M., Amin, M. A., Shoyaib, M., ... and Ali, A. A. (2020). Human activity recognition from wearable sensor data using self-attention. arXiv preprint arXiv:2003.09018.

[132] Interoperability, N. B. D. (2015). NIST Big Data Public Working Group Definitions and Taxonomies Subgroup In: Framework: Definitions. NIST Special Publication, 1500-1.

[133] https://www.nitrd.gov/ (Accessed February 2021).

[134] Newell, A. (1983). The heuristic of George Polya and its relation to artificial intelligence. Methods of heuristics, 195-243.

[135] Ovidiu, V., and Peter, F. (2014). Internet of Things-From Research and Innovation to Market Deployment. River Publishers Series in Communication.

[136] Oliver, D., Healey, F., Haines, T. P. (2010). Preventing falls and fall-related injuries in hospitals. Clinics in geriatric medicine, 26(4), 645-692.

[137] Osman, I. H., and Laporte, G. (1996). Metaheuristics: A bibliography.

[138] Pattnaik, K., and Mishra, B. S. P. (2016). Introduction to big data analysis. In Techniques and Environments for Big Data Analysis (pp. 1-20). Springer, Cham.

[139] Palak Sood, Himani Sharma, Sumeet Kaur Sehra, (2019). A Survey of Different Methods in Border Security and Surveillance. International Journal of Computer Sciences and Engineering, 7(10), 217-228.

[140] Patgiri, R., and Ahmed, A. (2016, December). Big data: The v's of the game changer paradigm. In 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (pp. 17-24). IEEE.

[141] Power, D. J. (2014). Using 'Big Data'for analytics and decision support. Journal of Decision Systems, 23(2), 222-228.

[142] Pearl, J. (1984). Heuristics: intelligent search strategies for computer problem solving.

[143] Passino, K. M. (2002). Biomimicry of bacterial foraging for distributed optimization and control. IEEE control systems magazine, 22(3), 52-67.

[144] Pattnaik, S. S., Bakwad, K. M., Devi, S., Panigrahi, B. K., and Das, S. (2011). Parallel bacterial foraging optimization. In Handbook of Swarm Intelligence (pp. 487-502). Springer, Berlin, Heidelberg.

[145] Poudel, S. (2016). Internet of Things: underlying technologies, interoperability, and threats to privacy and security. Berkeley Technology Law Journal, 31(2), 997-1022.

[146] Pham, D. T., Mahmuddin, M., Otri, S., and Al-Jabbouli, H. (2007). Application of the bees algorithm to the selection features for manufacturing data. In International Virtual Conference on Intelligent Production Machines and Systems (IPROMS 2007).

[147] Poppe, R. (2010). A survey on vision-based human action recognition. Image and vision computing, 28(6), 976-990.

[148] Papagiannaki, A., Zacharaki, E. I., Kalouris, G., Kalogiannis, S., Deltouzos, K., Ellul, J., and Megalooikonomou, V. (2019). Recognizing physical activity of older people from wearable sensors and inconsistent data. Sensors, 19(4), 880.

[149] Reinsel, D., Gantz, J., and Rydning, J. (2017). Data age 2025: The evolution of data to life-critical. Don't Focus on Big Data, 2.

[150] Rydning, D. R. J. G. J. (2018). The digitization of the world from edge to core. Framingham: International Data Corporation.

[151] Ruiz, D. S., Morales, S. F. and Molina, J. G. (2015). Inferring Versioned Schemas from NoSQL Databases and Its Applications.. In P. Johannesson, M.-L. Lee, S. W. Liddle, A. L. Opdahl and O. P. Lpez (eds.), ER (p./pp. 467-480), : Springer. ISBN: 978-3-319-25263-6

[152] Eberhart, R., Kennedy, J. (1995, October). A new optimizer using particle swarm theory. In MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science (pp. 39-43). Ieee.

[153] Torres, R. L. S., Ranasinghe, D. C., and Shi, Q. (2013, December). Evaluation of wearable sensor tag data segmentation approaches for real time activity classification in elderly. In International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services (pp. 384-395). Springer, Cham.

[154] Reyns, C. (2007). Etude des Algorithmes gntiques et application aux donnes de protomique (Doctoral dissertation, Universit Montpellier I).

[155] Rechenberg, I. (1989). Evolution strategy: Nature's way of optimization. In Optimization: Methods and applications, possibilities and limitations (pp. 106-126). Springer, Berlin, Heidelberg.

[156] Recommendation ITU-T Y.2060 (2012-06-15), availa- ble online: http://handle.itu.int/11.1002/1000/1155.

[157] Shoaib, M., Bosch, S., Incel, O. D., Scholten, H., and Havinga, P. J. (2015). A survey of online activity recognition using mobile phones. Sensors, 15(1), 2059-2085.

[158] Shinmoto Torres, R. L., Visvanathan, R., Abbott, D., Hill, K. D., and Ranasinghe, D. C. (2017). A battery-less and wireless wearable sensor system for identifying bed and chair exits in a pilot trial in hospitalized older people. PloS one, 12(10), e0185670.

[159] Serra, R., and Lopez, L. (2017). Damage detection methodology on beam-like structures based on combined modal Wavelet Transform strategy. Mechanics and Industry, 18(8), 807.

[160] Stutzle, T., and Hoos, H. (1997, April). MAX-MIN ant system and local search for the traveling salesman problem. In Proceedings of 1997 IEEE international conference on evolutionary computation (ICEC'97) (pp. 309-314). IEEE.

[161] Shi, Y. (Ed.). (2012). Innovations and Developments of Swarm Intelligence Applications. IGI Global.

[162] Salton G, MCGILL Michael. (1986). Introduction to Modern Information Retrieval. McGraw-Hill, Inc. New York, NY, USA.

[163] Segireddy, S., and Koneru, S. V. (2020). Wireless IoT-Based Intrusion Detection Using LIDAR in the Context of Intelligent Border Surveillance System. In Smart Intelligent Computing and Applications (pp. 455-463). Springer, Singapore.

[164] SAHRAOUI, S. (2016). Mcanismes de scurit pour l'intgration des RCSFs l'IoT (Internet of Things) (Doctoral dissertation, Universit de Batna 2).

[165] Singh, S., Velastin, S. A., and Ragheb, H. (2010, August). Muhavi: A multicamera human action video dataset for the evaluation of action recognition methods. In 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance (pp. 48-55). IEEE.

[166] Schmidt, S. (2012). Data is exploding: the 3 V's of big data. Business Computing World.

[167] Slimani, Y. (2018). Extraction et analyse de connaissances partir du Web (Doctoral dissertation).

[168] Sawant, N., and Shah, H. (2013). Big data application architecture. In Big data Application Architecture Q and A (pp. 9-28). Apress, Berkeley, CA.

[169] Sundmaeker, H., Guillemin, P., Friess, P., and Woelffl, S. (2010). Vision and challenges for realising the Internet of Things. Cluster of European research projects on the internet of things, European Commision, 3(3), 34-36.

[170] Storn, R., and Price, K. (1997). Differential evolution a simple and efficient heuristics for global optimization. J. of Global Opt, 11.

[171] Speakman, J. R., and Racey, P. A. (1991). No cost of echolocation for bats in flight. Nature, 350(6317), 421-423.

[172] Slagle, J. R. (1971). Artificial intelligence: The heuristic programming approach. McGraw-Hill.

[173] Solso, R. L. (1979). Cognitive psychology. New York: Harcourt Brace Jovanovich

[174] Tan, L., and Wang, N. (2010, August). Future internet: The internet of things. In 2010 3rd international conference on advanced computer theory and engineering (ICACTE) (Vol. 5, pp. V5-376). IEEE.

[175] De Castro, L. N., and Timmis, J. (2002, May). An artificial immune network for multimodal function optimization. In Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600) (Vol. 1, pp. 699-704). IEEE.

[176] Torres, R. L. S., Ranasinghe, D. C., Shi, Q., and Sample, A. P. (2013, April). Sensor enabled wearable RFID technology for mitigating the risk of falls near beds. In 2013 IEEE International Conference on RFID (RFID) (pp. 191-198). IEEE.

[177] Torres, R. L. S., Ranasinghe, D. C., and Shi, Q. (2013, December). Evaluation of wearable sensor tag data segmentation approaches for real time activity classification in elderly. In International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services (pp. 384-395). Springer, Cham.

[178] Turaga, P., Chellappa, R., Subrahmanian, V. S., and Udrea, O. (2008). Machine recognition of human activities: A survey. IEEE Transactions on Circuits and Systems for Video technology, 18(11), 1473-1488.

[179] Tsai, C. W., Lai, C. F., and Vasilakos, A. V. (2014). Future Internet of Things: open issues and challenges. Wireless Networks, 20(8), 2201-2217.

[180] Tang WJ, Wu QH, Saunders JR (2006) A novel model for bacteria foraging in varying environments. In: Proceedings of ICCSA, LNCS, vol 3980, pp 556-565

[181] Ward, J. S., and Barker, A. (2013). Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821.

[182] Weill, M., and Souissi, M. (2010). L'Internet des objets: concept ou ralit?. In Annales des Mines-Ralits industrielles (No. 4, pp. 90-96). Eska.

[183] Wickramasuriya, J., Datt, M., Mehrotra, S., and Venkatasubramanian, N. (2004, October). Privacy protecting data collection in media spaces. In Proceedings of the 12th annual ACM international conference on Multimedia (pp. 48-55).

[184] Wu, M., Lu, T. J., Ling, F. Y., Sun, J., and Du, H. Y. (2010, August). Research on the architecture of Internet of Things. In 2010 3rd international conference on advanced computer theory and engineering (ICACTE) (Vol. 5, pp. V5-484). IEEE.

[185] Whitley, L. D. (1989, June). The GENITOR algorithm and selection pressure: why rank-based allocation of reproductive trials is best. In Icga (Vol. 89, pp. 116-123).

[186] Wickramasinghe, A., and Ranasinghe, D. C. (2016, March). Recognising activities in real time using body worn passive sensors with sparse data streams: To interpolate or not to interpolate?. In Proceedings of the 12th EAI international conference on mobile and ubiquitous systems: computing, networking and services on 12th EAI international conference on mobile and ubiquitous systems: computing, networking and services (pp. 21-30).

[187] Wang, J., Chen, Y., Hao, S., Peng, X., and Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters, 119, 3-11.

[188] Wang, H., Zhao, J., Li, J., Tian, L., Tu, P., Cao, T., ... and Li, S. (2020). Wearable Sensor-Based Human Activity Recognition Using Hybrid Deep Learning Techniques. Security and Communication Networks, 2020.

[189] Yang, X. S. (2010). A new metaheuristic bat-inspired algorithm. In Nature inspired cooperative strategies for optimization (NICSO 2010) (pp. 65-74). Springer, Berlin, Heidelberg.

[190] Yang, R. (2003). View-dependent Pixel Coloring: A Physically-based Approach for 2D View Synthesis (Doctoral dissertation, University of North Carolina at Chapel Hill).

[191] Yang, Y., Zheng, X., Guo, W., Liu, X., and Chang, V. (2019). Privacy-preserving smart IoT-based healthcare big data storage and self-adaptive access control system. Information Sciences, 479, 567-592.

[192] Zhang, H., Chen, G., Ooi, B. C., Tan, K. L., and Zhang, M. (2015). In-memory big data management and processing: A survey. IEEE Transactions on Knowledge and Data Engineering, 27(7), 1920-1948.

[193] Zikopoulos, P., and Eaton, C. (2011). Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media.

[194] Zikopoulos, P. C., Eaton, C., and DeRoos, D. (2012). Understanding big data: analytics for enterprise class Hadoop and streaming data.

[195] Zhang, W., Cheung, S. C. S., and Chen, M. (2005, September). Hiding privacy information in video surveillance system. In IEEE International Conference on Image Processing 2005 (Vol. 3, pp. II-868). IEEE.

[196] Zigel, Y., Litvak, D., Gannot, I. (2009). A method for automatic fall detection of elderly people using floor vibrations and sound-Proof of concept on human mimicking doll falls. IEEE transactions on biomedical engineering, 56(12), 2858-2867.