ALGERIAN PEOPLE'S DEMOCRATIC REPUBLIC

Ministry of Higher Education and Scientific Research

BLIDA 1 UNIVERSITY

Faculty of Technology

Department of Process Engineering

Dissertation

In view of obtaining the degree of

MASTER IN PROCESS ENGINEERING

Specialization: Industrial Pharmacy

Title of the dissertation

Machine Learning to Automatic CAPA

Studied by: ABD EL LAOUI Fethia

Supervised by: Dr CHERIET Nabil

Academic year 2023/2024

ملخص

لإدارة الفعالة لعمليات الإجراءات التصحيحية والوقائية ضرورية للغاية للحفاظ على معاييرالجودة في الصناعات. ومع ذلك،غالباً ما تعاني الطرق التقليدية من عدم الكفاءة والتأخير. تستكشف هذه الدراسة تطبيق تقنيات التعلم الآلي لتحسين عمليات من خلال دمج الخوارزميات و التحليلات البيانية .تكمن الأهداف الرئيسية في هذه الدراسة في النمذجة التنبؤية للكشف المبكر عن المشكلات والرصد في الوقت الحقيقي و الدعم الذكي للقرارات .يعتمد البحث نهجا متعدد التخصصات .تشمل منهجية البحث على معالجة البيانات المسبقة و تدريب النماذج ثم التحقق .تتضمن النتائج تطوير نموذج أولي لنظام CAPA مدفوع بتقنية تعلم الآلة

Abstract:

Efficient management of Corrective and Preventive Action (CAPA) processes is crucial for maintaining quality standards in industries. However, traditional methods often suffer from inefficiencies and delays. This dissertation explores the application of machine learning (ML) techniques to automate CAPA workflows. By integrating ML algorithms and data analytics, the research aims to revolutionize CAPA management. Key objectives include predictive modeling for early issue detection, real-time monitoring, and intelligent decision support. The study adopts a multidisciplinary approach, drawing insights from computer science and quality engineering. The research methodology encompasses data preprocessing, model training, and validation. Outcomes include the development of a ML-driven CAPA system prototype and insights for organizations seeking to streamline quality management processes.

Résumé :

La gestion efficace des processus d'action corrective et préventive (CAPA) est cruciale pour maintenir les normes de qualité dans les industries. Cependant, les méthodes traditionnelles souffrent souvent d'inefficacités et de retards. Ce mémoire explore l'application des techniques d'apprentissage automatique (ML) pour automatiser les flux de travail CAPA. En intégrant des algorithmes ML et des analyses de données, la recherche vise à révolutionner la gestion des CAPA. Les objectifs clés incluent la modélisation prédictive pour la détection précoce des problèmes, la surveillance en temps réel et le soutien à la décision intelligent. L'étude adopte une approche pluridisciplinaire, tirant des enseignements de l'informatique et de l'ingénierie de la qualité. La méthodologie de recherche englobe le prétraitement des données, la formation et la validation des modèles. Les résultats comprennent le développement d'un prototype de système CAPA piloté par ML et des perspectives pour les organisations cherchant à rationaliser les processus de gestion de la qualité.

ACKNOWLEDGEMENTS

To the Almighty Allah, who granted me the strength and perseverance to embark on and complete this humble endeavor, and the courage to overcome every challenge.

Heartfelt thanks to my beloved parents, who have stood by my side, and to my siblings whose encouragement has been a constant source of inspiration.

Special gratitude to ***Dr. Nabil CHERIET****, my mentor, for his unwavering support, attentive guidance, boundless patience, unwavering encouragement, and constant motivation. His mentorship has enabled me to work under optimal conditions.*

I owe a great debt of gratitude to my supervisor, ***Mr. Fayçal ANDJECHAIRI****, for warmly welcoming me into the* ***HIKMA PHARMACEUTICALS laboratory****. His receptiveness, exchange of ideas, and invaluable advice have significantly contributed to my professional growth.*

I extend my sincere thanks to the jury members for their presence and humble willingness to assess my work.

I express my heartfelt gratitude to the faculty team of the chemical engineering department.

A special mention of gratitude to ***Mme.Manel BOUZIT****, for her support and encouragement throughout this journey.*

Lastly, I extend my deepest appreciation and sincere thanks to my parents, whose unwavering support has been our strength throughout my academic journey.

Thank you!

DEDICATION

To my beloved parents, brother, sister, and family

Your unwavering love, encouragement, and sacrifices have been the cornerstone of my journey. Your belief in me has been a constant source of strength, and I am so grateful for your boundless support and understanding.

To my dear friends

Your camaraderie, laughter, and understanding have illuminated even the darkest of days. Thank you for standing by me, for cheering me on, and a special mention to my beloved friend NIVIN, your loyalty, empathy, and companionship have been my guiding light. you have been my confidant and source of endless inspiration. Thank you for being there for me.

This dissertation is dedicated to each of you, with heartfelt gratitude for your love, encouragement, and unwavering belief in me. You have made this journey worthwhile.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ASQ: American Society for Quality

QC: Quality control

QA: Quality Assurance

TQM: Total Quality Management

BSI: British Standards Institution

ISO: International Organization for Standardization

GLP: Good Laboratory Practices

GCP: Good Clinical Practices

API : Active Pharmaceutical Ingredients

QRM: Quality Risk Management

QbD : Quality by Design

FDA:Food and Drug Administration

ICH: International Council for Harmonisation

cGMP: Current Good Manufacturing Practices

API : Active Pharmaceutical Ingredients

QMS: quality management system

WHO : World Health Organization

SOP: Standard operating procedure

FTA : Fault Tree Analysis

CAPA : Corrective and Preventive Action

AI: Artificial intelligence

t-SNE: t-distributed stochastic neighbor embedding

PCA: principal component analysis

LOF: local outlier factor

LDA: latent Dirichlet allocation

GANs : Generative Adversarial Networks

RNNs: Recurrent Neural Networks

CNNs : Convolutional Neural Networks

LSTMs: Long Short-Term Memory Networks

RL: Reinforcement Learning

DQNs: Deep Q-Networks

GNNs: Graph Neural Networks

SAR: Structure-Activity Relationship

GENERALE INTRODUCTION:

In the areas of quality management and regulatory compliance, implementing effective corrective and preventive actions (CAPA) is critical to ensuring product quality, compliance with standards and the continuous improvement of organizational processes. The CAPA framework provides a systematic approach to identifying, addressing, and mitigating nonconformities, deviations, and potential risks in the quality management system.

This work examines how machine learning techniques can be used to automate and optimize CAPA processes, with a particular focus on improving the efficiency, accuracy, and responsiveness of corrective and preventive actions in regulated industries.

Ensures the products are safe, effective, and reliable within the regulated industries like pharmaceutical, biotechnologies, and medical devices among others in the aerospace field. Properly developed CAPA processes are part and parcel of quality management systems that provide organizations with proactiveness on finding out the root causes of their problems, taking corrective actions on immediate concerns, coming up with the preventive measure against recurrence, and improvement on long-term performance.

Although CAPA processes have been recognized to have very high importance, most of them are likely to face challenges in implementing and thus managing the corrective and preventive actions correspondingly due to the barrier of a wide variety of challenges, including those related with data complexity, process variability, and organizational silos. The practices related with the manual CAPA procedures are likely to create the major constraints of lags, inconsistency, weakness, and delays to give an affect to quality shortfalls in a timely and thus preventive manner. This is an expression of the need to bring to bear innovative solutions, applying machine learning to capture the workflow of CAPA in an automated and optimally correct way that would aid in driving the organization toward achieving higher and improved quality of service, compliance, and operational excellence.

This dissertation focuses on the application of machine learning techniques to automate and optimize corrective and preventive actions within CAPA systems. We will consider a range of industries and quality management frameworks. The research will encompass both theoretical investigations and practical implementations, with an emphasis on addressing real-world challenges and industry-specific requirements.

Automating CAPA processes with machine learning has a significant impact on business performance, regulatory compliance, and customer satisfaction. By leveraging data-driven insights and predictive analytics, companies can proactively identify and mitigate quality issues, reduce the risk of non-compliance, and drive continuous improvement across operations. This research aims to advance the state of the art in CAPA management and contribute to the development of innovative solutions that enable organizations to achieve excellence in quality and operational efficiency.

In the first chapter, we explore the foundational concepts of quality management and deviation management. We delve into the importance of maintaining product quality and the consequences of deviations in manufacturing processes, emphasizing the need for timely corrective and preventive actions (CAPA). Additionally, we discuss common methodologies used in deviation management, such as root cause analysis ect.

The second chapter provides an in-depth understanding of artificial intelligence (AI) and its relevance to the pharmaceutical industry. We explore various AI techniques, including machine learning, and their applications in drug discovery, development, and manufacturing.

In third chapter, we focus on the design and implementation of machine learning (ML) algorithms for automatic corrective action and preventive action (CAPA) in the pharmaceutical manufacturing process. We discuss the selection of appropriate ML models based on the nature of data and desired outcomes, as well as data preprocessing techniques to ensure data quality and integrity.

In the final chapter, we present the results of implementing ML-based automatic CAPA in the pharmaceutical manufacturing setting. We evaluate the performance of the developed ML models in terms of accuracy, efficiency, and scalability, comparing them with traditional deviation management approaches.

CHAPTER I:

FUNDAMENTALS OF QUALITY AND DEVIATION MANAGEMENT IN THE PHARMACEUTICAL INDUSTRY

The pharmaceutical industry operates within a highly regulated environment where ensuring product quality is paramount.

I.1 Quality management in the pharmaceutical industry:

I.1.1 General Definitions:

I.1.1.1 The Quality:

Quality could be a rather more complicated term than it appears. Each quality professional defines quality could be a rather completely different approach. There are various views which will be taken in shaping quality (e.g. clientele perspective, specification-based outlook).

The contemporary definition of quality traces its origins to Jordan's "fitness for supposed use." The essence of quality is therefore to "meet or exceed consumer or client expectations. [1]

I.1.1.2 Quality Control:

Quality control is concerned with the production process of the product or inspect with the aim of getting rid of evils which might result in defects. According to ASQ, QC contain the organized techniques and hence the activities which maintain a quality of product or service which will satisfy given needs, also the use of such techniques and activities. [1]

I.1.1.3 Quality Assurance:

Maybe a grand theory that focuses on the whole system of quality including suppliers and ultimate consumers of the merchandise or check. It includes all the tricks designed to supply products and services of appropriate quality. Consistent with ASQ, QA includes all those meant or methodical actions necessary to supply sufficient confidence that a manufactured goods or service will satisfy given needs. [1]

I.1.1.4 Conformance quality:

It is a measure of degree to which product or service was produced correctly.[2]

As per pharmaceutical industry, quality is defined as conformance to specifications of the product. The specifications are different for different products depending upon its therapeutic effect, potency etc.

I.1.1.5 Quality Management:

Quality management is management of different sectors to assure good quality of the product. Quality management consist of four main components i.e, Quality planning, Quality control, Quality assurance, Quality improvement. [3]

The outline of quality management system is given in Figure 1.

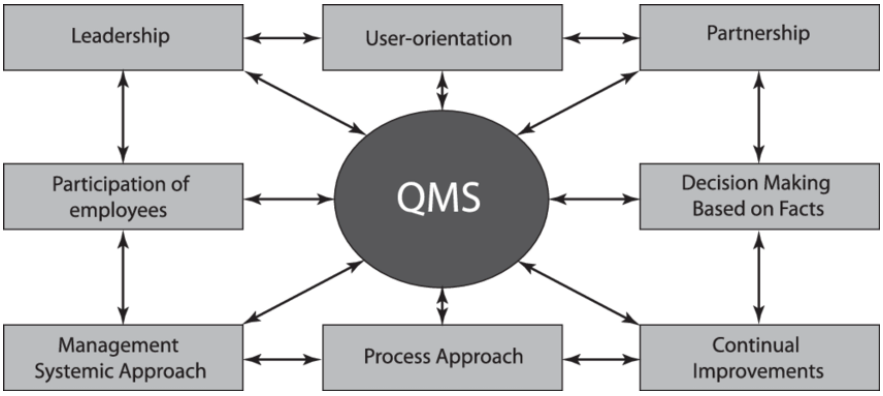

Figure1. 1: The outline of quality management system

I.1.2 Total Quality Management:

I.1.2.1 Definition:

Total Quality Management (TQM) is an integrated effort within firms in raising the level of quality within its different levels. To this end, TQM focuses on compliance with the quality standards that are set by customers, which is referred to as customer-defined quality. [4]

It includes organizational efforts to build a culture supportive to continuous improvement in quality delivery to consumers. Hence, quality control techniques developed and defined over the years are an important source of TQM initiatives. [5]

Organizations all over the world have given different definitions to TQM.

- TQM is a management philosophy and set of company practices for optimizing the human and material resources of an organization so that its purposes, missions, and objectives are achieved effectively according to British Standards Institution (BSI) standard BS 7850-1:1992. [6]

-Total quality management (TQM) is a management approach of an organization, centered on quality, based on participation of all its members, and aiming at long-term success through customer satisfaction and benefits brought to all interested stakeholders together with society" -

such is defined in the ISO 8402:1994 by the International Organization for Standardization (ISO). [7]

- The American Society for Quality (ASQ) first coined the term in reference to the management approach for quality improvement purposes. In other words, TQM is a management philosophy toward long-term success through customer satisfaction. It is the full involvement of every single member in an organization that works towards improving processes, products, services, and organizational culture. [5]

I.1.2.2 Requirement for implementation of TQM: [8] [9]

The key requirements needed in implementing the concept of Total Quality Management (TQM) successfully are:

1. **Top management support:**

   If the culture of continuous improvement in quality is not built through the organization, then without the support from the top management, it cannot be taken into practice. Hence the management should provide an active initiative in supporting TQM by providing resources and giving its quality mission.

2. **Comprehensive training and motivation of employees:**

Training of the company's employees in order to provide them with all necessary skills and knowledge to participate in efforts aimed at the quality improvement. Through recognition, rewards, and many opportunities that the management can offer in growth in engaging the people in and giving them a share of ownership of the TQM initiatives.

3. **Understanding deep process relationship cause and effect:**

For an effective implementation of TQM, the organization needs to understand a number of the factors that add towards processes and their results. The application involves finding the root cause and also implementing corrective actions along with continuous monitoring and measurement of the performance metrics that drive improvement.

I.1.2.3 TQM: A multifaceted approach:

Total quality management (TQM) is a comprehensive approach to quality control in all areas of the pharmaceutical industry, including research and development, production and marketing.

1. **Research and Development:** Apart from the quality of management in the actual manufacturing process, the quality of management also ensures the quality of the product in the research and development (R&D) process. Research and development include:

a. Good Laboratory Practice (GLP): 'Good Laboratory Practices' or Good Laboratory Practice, concerns rigorous controls on the use of animals in laboratory experimentation. TQM implementation in GLP incorporates the under-mentioned steps :

- Preparing the protocols or master schedule sheets for the study.
- Maintaining the protocol copy in the laboratory room where the study will be carried out.
- Conducting periodic inspections of the facility where the study is to be carried out.
- It also requires documentation and approvals for any changes to the approved study protocol, which should include the reasons for the change.
- Adequate well-documented procedures through the course of the study.

b. Good Clinical Practices (GCP): GCP involves stringent control over the utilization of human beings in clinical trials. The regulations for GCP are akin to those of GLP, with a major distinction lying in the requirement for obtaining informed consent from subjects before commencing the study or involving them in clinical trials. A comprehensive informed consent form, duly filled and signed by the

2. **Manufacturing:** This part of production involves the manufacturing of raw material, manufacturing of the Active Pharmaceutical Ingredients (API), and formulation and packaging of dosage forms.
3. **post-marketing surveillance**: Quality management is extended into the post-marketing surveillance which bases on market surveys for product performance and safety in a real environment. This calls for the implementation of the change control procedures where documentation for any required changes shall be captured as per the approved processes as dictated by the surveillance findings. subjects, is necessary to ensure their understanding and agreement to participate in the trials. These records must be meticulously maintained. Additionally, in cases where patients drop out during the study, the number of dropouts and the reasons for their withdrawal should be documented.

I.1.2.4 TQM Approaches:

The industries use the following methodologies in the implementation of Total Quality Management (TQM) in the pharmaceutical process:

- **Six Sigma Techniques:**

This involves the set of tools and techniques applied to process improvement and quality management. Through the identification and addressing of root causes for defects (errors) and the reduction in process variability in manufacturing and business processes, the techniques under Six Sigma then lead to the enhancement of the process' quality outputs. [10][11]

The two core methods in Six Sigma are:

1. **DMAIC**: It is an acronym for Define, Measure, Analyze, Improve, and Control that defines the cycle of structured improvement, preferably containing five distinct phases. Its application, in most cases, is in the improvement of the existing business processes. The Five Phases of DMAIC are illustrated in Figure 2.

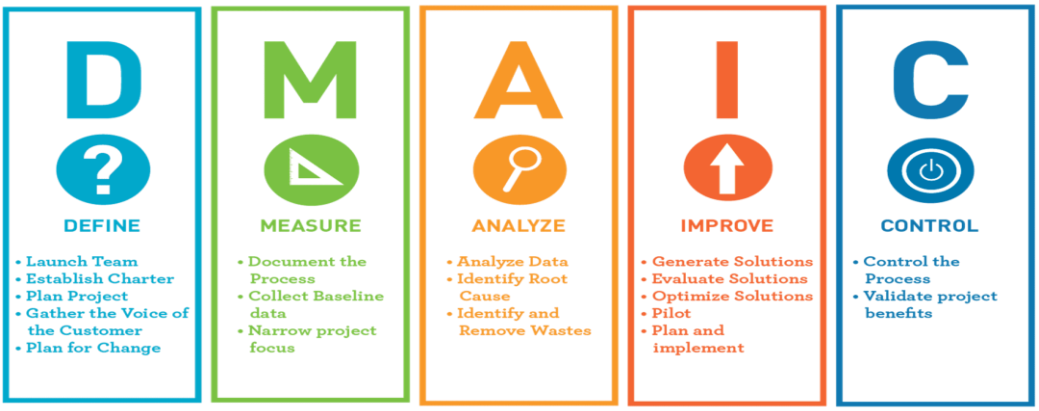

Figure1. 2: Various phases of the improvement system

2. : **DMADV** stands for Define, Measure, Analyze, Design, and Verify. The method is used for the development of new business processes or the introduction of a new product. This follows a development cycle, which is shown in Figure 3

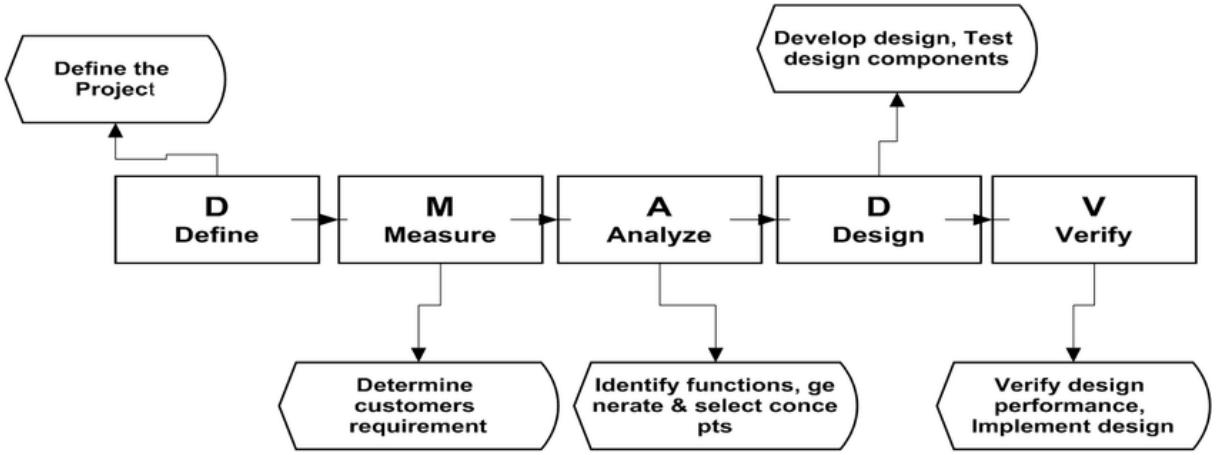

Figure1. 3: Various phases involved in creating new business

I.1.3 Quality Risk Management (QRM):

Quality Risk Management is the systematic process of identifying, assessing, controlling, communicating, and review risks to the quality of medicinal products, over their lifecycle, with decisions being made about each step in the procedure. Broadly, it is realized that risk management is an integral active management tool widely adopted in the pharmaceutical industry that involves the systematic process of identification, analysis, and control of risks associated with ongoing processes within the industry.[12][13]

The various components of quality risk management are depicted in Figure 4.

Figure1. 4: Parts involved in quality risk management

The quality risk management include:

- **Identification of risk**: This involves the issue of identifying those risks that are potentially there before they mature to become big problems calling for the solutions.
- **Data analysis**: This is the process of looking at data on risk and giving it a classification on the impact of the risk and the priority of the risk.

- **Planning**: Risk data analysis for strategic planning of the risk mitigation, hence, in effect making decisions on how to treat identified risks.
- **Tracking:** Monitoring how the identified mitigation plans of the risk are implemented and indicators of the risk.
- **Control:** Tight control of the mitigation plans of risk to keep them in their adherence and the lack of deviation
- **Communication**: Communication has been one of the major purposes of informing the stakeholders on the quality of the implemented risk management plans and assessing their achievement in the dimension of the risk mitigation and informing about the identified risks.

I.1.4 Quality by Design (QbD):

Joseph M. Juran first explained the concept of building quality into products by planning. [14]. He highlighted that the quality of an It uses the US FDA approved technique of statistical methods for optimization of composition of ingredients of the formulation.

QbD for pharmaceutical quality: The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) defines Pharmaceutical Quality by Design (QbD) as "a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control based on sound science and quality risk management ."Pharmaceutical QbD is a systematic, scientific, risk-based approach to the development of a pharmaceutical product beginning with pre-defined objectives. It is needed for the understanding of the product and on-going processes that compel a good depth of knowledge in relation to critical process parameters and critical quality attributes. [15]

Quality by Design is a systematic approach to pharmaceutical development that involves the development of a drug or drug product to provide quality assurance. It involves the design of experiments to identify CQAs and process parameters, establishing the relationship besides, it involves coming up with a control strategy that will result in the formation of the product uniformly, well-embodied with a strong control measure in a bid to ensure the quality standards are maintained accordingly.[16][17]

I.1.4 Current Good Manufacturing Practices (cGMP):

This technique also involves quality management in the pharmaceutical industry. A lot of regulatory agencies including US FDA, WHO, European Medicines Agency, India's Schedule M, to mention a few have laid down the guidelines on Good Manufacturing Practices (GMP). The site selection, facility design, attire requirements, waste disposal procedures, sanitation practices, testing protocols, documentation of analyses, procedures for reprocessing or recalls, documentation of any process changes by change controls should be covered by those rules.

Good Manufacturing Practices relevant to this need to be properly understood and adhered to in the process to get quality pharmaceutical products.[18][19]

I.1.6 International Conference on Harmonization (ICH):

The International Committee for Harmonization (ICH) has developed guidelines for the quality risk management of pharmaceutical products, known as the Q9 guidelines. In addition, Q8(R2)[15][17] guidance has been developed for drug development. These guidelines, especially Q9, provide valuable insights into risk assessment and risk control. They also provide guidance on various approaches to quality risk management, including Failure Mode and Effects Analysis (FMEA), Failure Mode, Effects, and Severity Analysis (FMECA), Hazard Analysis and Critical Control Points (HACCP), and Preliminary Hazard Analysis (PHA). and risk ranking and filtering. These methods are effective tools for identifying, assessing and mitigating risks throughout the life cycle of pharmaceutical products and ensuring their quality, safety and effectiveness.

Total Quality Management (TQM) of the pharmaceutical processes involves a number of approaches to the use of Six Sigma, Lean Manufacturing, Quality Risk Management, Quality by Design, ISO standards, cGMP, and ICH guidelines. This definition of Total Quality Management (TQM) of the pharmaceutical processes from various viewpoints makes it far more difficult than elsewhere. [20]

One of the important aspects is the sound system of documentation It is always said, "anything not documented means not done," so documentation is a must for TQM implementation. Any change or deviation in the processes is supposed to pass through the formal change or deviation control which ensures proper documentation and approval of each alteration from validated procedures. This meticulous approach guarantees transparency and consistency in maintaining quality standards.

I.2 Handling deviation:

Handling deviations is a critical aspect of a quality management system (QMS), ensuring product quality by continuously improving it. As part of Corrective and Preventive Action (CAPA), immediate action (corrections) is required upon detecting a deviation. Root cause analysis should follow, leading to systemic actions (corrective actions) to prevent future non-conformances.

In recent times, Quality Risk Management (QRM) has emerged as a preventive measure against deviations in the pharmaceutical industry. International standards such as the International Conference on Harmonization (ICH) guideline Q9 (ICH Q9) and recommendations from the World Health Organization (WHO) advocate for implementing QRM systems in pharmaceutical industries.

The personnel's response to deviations presents a primary challenge to the system, largely influenced by their level of training, qualification, commitment, and support from higher authorities within the company. [21]

To record, classify, and investigate events based on their risk levels, a decision tree is employed to guide individuals in making informed decisions. This decision tree provides a simplified risk assessment, addressing the following questions (figure 5).

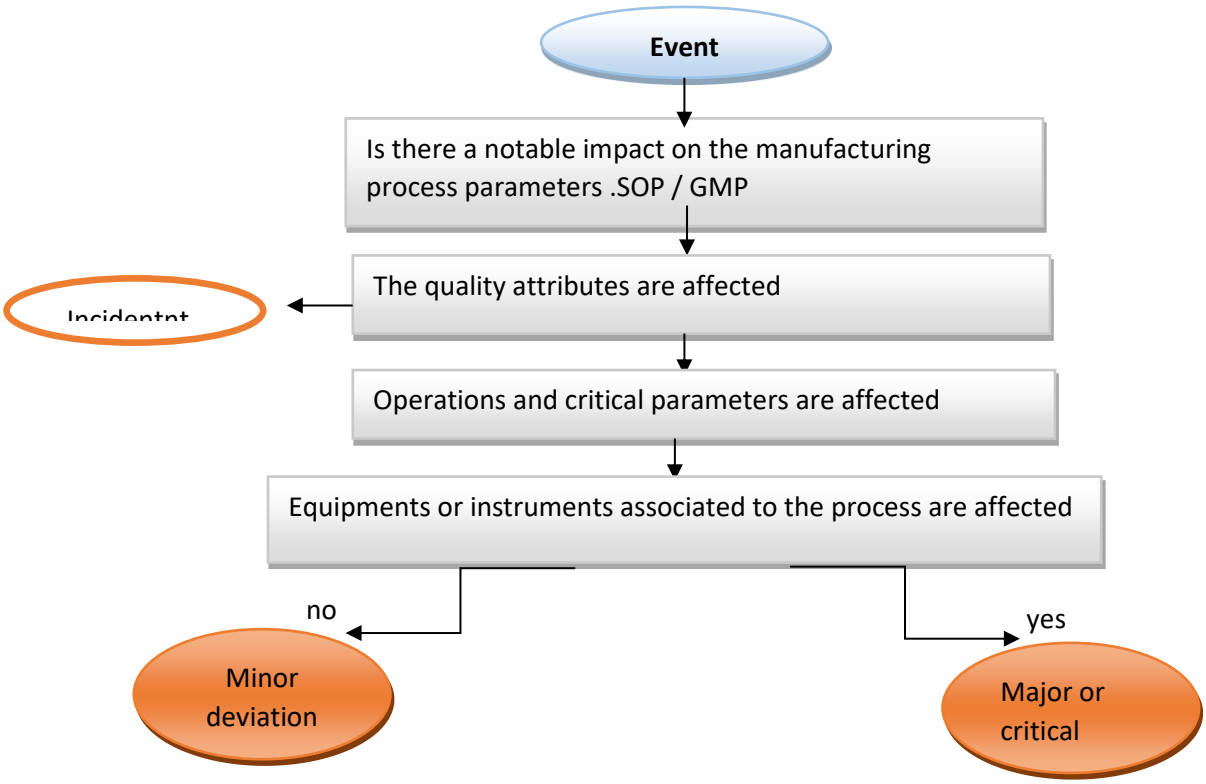

Figure1. 5: Classification of deviation

- **Whether the events affect the quality of the product ?**
- **Do the approved specifications or written procedure which is examined, conflict with a requirement?** [21]

I.2.1 Types of deviations:

Deviations are of two types: Planned deviations, unplanned deviations [22]

I.2.1.1 Planned Deviations:

Any deviation from a standard procedure selected intentionally for a short period to avoid undesirable situation without affecting the safety and quality of the product or procedure.

E.g.: Batch executed with lower input due to nonavailability of raw materials [23]

I.2.1.2 Unplanned Deviations:

Unplanned deviations are the accidental nonconformance observed after or during the implementation of an activity. Unplanned deviations may occur due to the following reasons: Equipment breakdown, Interruption of Power supply, Site Accidents, Utility Breakdown errors during documentation.

Deviations may be further categorized into 3 types based on the impact of the deviation on the product quality, safety and validation state of the facility and process: Critical, major, minor [24]

- **Critical deviation:**

The deviation will have a notable impact on the critical attributes of the product.

For Example: Usage of contaminated raw materials and solvents. Integrity failure of high efficiency particulate air filters [24]

- **Major Deviation:**

The deviation will or may have a notable impact on critical attributes of the product.

For Example: Critical process and in process parameter failure. Significant variation from standard output range.

- **Minor Deviation:**

The deviation will not have any direct impact on the quality of the product.

For Example: Weights not replaced properly after use. Equipment and measuring device malfunction. [24]

I.2.2 Procedure for handling of Deviation

I.2.2.1 When to raise a deviation:

As soon a deviation occurs it should be immediately reported to Quality assurance (QA) within one working day from the time when the deviation has occurred or as per the Standard operating procedure (SOP).

Record the issue in the deviation record, the time of occurrence and name of the department and the person who observed it should be documented. Standard against which the deviation occurred should be recorded.[25]

I.2.2.2 Initial details:

All the initial details regarding the deviation should be mentioned. All technical details and critical process parameters should be recorded. [25]

I.2.2.3 investigating deviations:

Establish what happened, understand the events, Use investigation tools, Identify the causes and check for any common causes. Check if any other materials, components, batches and equipment are affected.

I.2.2.4 Root causes analysis:

Root cause analysis is a systematic process for identifying "root causes" of problems or events and an approach for responding to them. Root cause analysis can be identified and classified as errors caused by Men, Material, Machine, Method and Mother Nature.

A /Tools for identifying the Root cause

- ✓ **Fishbone analysis:**

A fishbone diagram, also called a cause-and-effect diagram, is a visualization tool for categorizing the potential causes of a problem to identify its root causes. A fishbone diagram is useful in product development and troubleshooting processes to focus the conversation. After the group has brainstormed all the possible causes for a problem, the facilitator helps the group to rate the potential causes according to their level of importance and diagram a hierarchy. The design of the diagram looks much like a skeleton of a fish. Fishbone diagrams are typically worked right to left, with each large "bone" of the fish branching out to include smaller bones containing more detail.[25]

- ✓ **5 whys analysis**

5 Why analysis is used as a tool in root cause analysis. It is a set of five questions to find out the base of the problem. Sometimes, it is necessary to find out by asking more than 5 questions. Ask a question 'why' repeatedly to know the root of the problem until you find out the correct root cause [26]

For Example: Why the machine was stopped suddenly?

Answer: Due to human error, Why did the human error occur?

Answer: Human suddenly pressed the stop button. Why did the human press it?

Answer: The label on the button was not visible. Why was the label not clearly?

Answer: Because it was covered with dirt. Why was it covered with dirt?

Answer: It was not properly cleaned. Therefore, here we understand that the root cause is no proper cleaning practice. [28]

- ✓ **Fault Tree Analysis (FTA)**

Fault tree analysis is a tool to find out the root cause analysis for the deviations. This helps to evaluate the failure of system one at a time and sometimes, by identifying the casual chain of events, multiple causes can be combined [27].

The results of these events are represented pictorially in the tree form. FTA is used to investigate the deviations and complaints to understand the root cause and to make improvements so that it does not lead to further problems. [28] Few other tools for identifying the causes are: Pareto charts, brainstorming, flowcharting, change analysis.

I.3 CAPA (CORRECTIVE ACTION /PREVENTIVE ACTION)

I.3.1 definitions:

I.3.1.1 correction:

An action taken to rectify non-conformities or undesirable situations; correction primarily deals with symptoms rather than addressing root causes. [29]

I.3.1.2 corrective action:

A corrective action is a term to refer to the response and resolution process to a product, customer, or non-conformance issue. Root Cause Analysis means the application of problem-solving tools and techniques towards defining and resolving the unwanted situation in such a manner that they do not repeat in the future. [30]

I.3.1.3 Preventive action:

Preventive action means processes aimed at identification of a possible problem or non-conformance and its removal in advance. The purpose of preventive actions is to prevent nonconformities or undesirable situations from occurring. They anticipate and avert problems, which are bound to occur at some time.[31]

I.3.1.4 Difference between corrective and preventive action:

The following table outlines the difference between corrective and preventive action [32]

Table I. 1:Difference between Corrective and Preventive Action

Corrective Action	Preventive Action
Reactive	Proactive
Initiated from customer complaint	Initiated from customer suggestion
Existing non-conformity	Opportunity for improvement
The issues is currently affecting the process	No immediate issue affecting the process
A solution is requires now (at least a temporary one)	No need to do something right away
Changes must be made	Changes might not be implemented
A root-cause analysis is required	A root-cause analysis is not required

I.3.2 Objective:

The Corrective and Preventive Action Subsystem is developed to collect and analyze data, identify and research problems of product and quality, and bring about the proper and effective corrective and/or preventive action to prevent reoccurrence.

Key points in managing products effectively to avoid recurrence, minimize quality problems, and device failures lie in ensuring the verification or validation of corrective and preventive actions, in effective communication of such actions to responsible persons, provision of relevant information for management review, and thorough documentation of these activities. Among the most critical elements of a quality system is the corrective and preventive action subsystem.[33]

I.3.3 Why CAPA:

Corrective action and preventive action, many times referred to simply as corrective action, are the improvements imposed on processes of organizations to bring about the elimination of undesirable situations and non-conformities.

Corrective and preventive action is instrumental in resolving issues that take place in the course of manufacture. It is aimed at making such issues do not recur in cases of manufacture and analysis**:**

- The estimate the effectiveness of a given action plan.
- The CAPA also certify that information associated with nonconforming products and quality problems emanate from those to blame for the prevention of such problems or assuring the standard of such products.
- The corrective action and preventive action is accountable with for information collection, analysis, identification and analysis of quality and merchandise problems and taking effective and suitable preventive and/or corrective actions in an a trial to stop their occurrence or re-occurrence.
- Providing necessary information for management review, and documenting the activities that are vital in handling with quality and merchandise problems, preventing problem re-occurrence and minimizing or preventing device failure.
- CAPA is a necessary tool for improving organizations processes.
- CAPA is chargeable for validating and verifying preventive and corrective actions, communicating preventive action and corrective action activities to responsible people. [34]

I.3.4 Detection:

Proper documentation of the problem at hand is a necessity in problem identification and CAPA detection. This means that the outline should be comprehensive, elaborating on the who, what, when, where, why, and how many aspects of the problem.

Secondly, there has to be a risk assessment against compliance risk. The output of the risk analysis has to clearly communicate the timeframe of the CAPA. There has to be a clear understanding that low-risk issues are not going to be any kind of an issue, compared to high-risk issues.[34]

I.3.5 Investigation and root cause determination:

The quality management teams should prioritize swift investigation and root cause determination. There are various methods available for conducting analysis, including：

- Brainstorming;
- Flowcharting;
- Fishbone Diagrams;
- Affinity Diagrams;
- Physics of Failure;

Normally, quality management systems enable root cause determination. The system of end-to-end traceability enables the easy following of all changes and action sequences from beginning to end, therefore fully integrating quality processes in a closed loop.

I.3.6 Proposed correction:

In this next phase, corrections and adjustments should be completed as much as possible to stop any further disruption. Additionally, organizations should consider proactively process and procedure reviews to address broader issues. In the case of a product-related issue, these also may require field corrections and/or recalls.

I.3.7 Implementation:

At this point, long-term corrective and preventive action works to observe or get rid of the explanation for nonconformity. A corrective action is the erasing of the cause of nonconformity. On the other hand, a preventative action is an action to delete the explanation for potential nonconformity.

I.3.8 Verification of effectiveness:

Last but not least, it is very important to check or verify the effectiveness of the CAPA. After an investigation for CAPA has been Also, any change in the process that has been undertaken to address an issue is a potential source of another problem. [35]

I.3.9 Stages of CAPA:

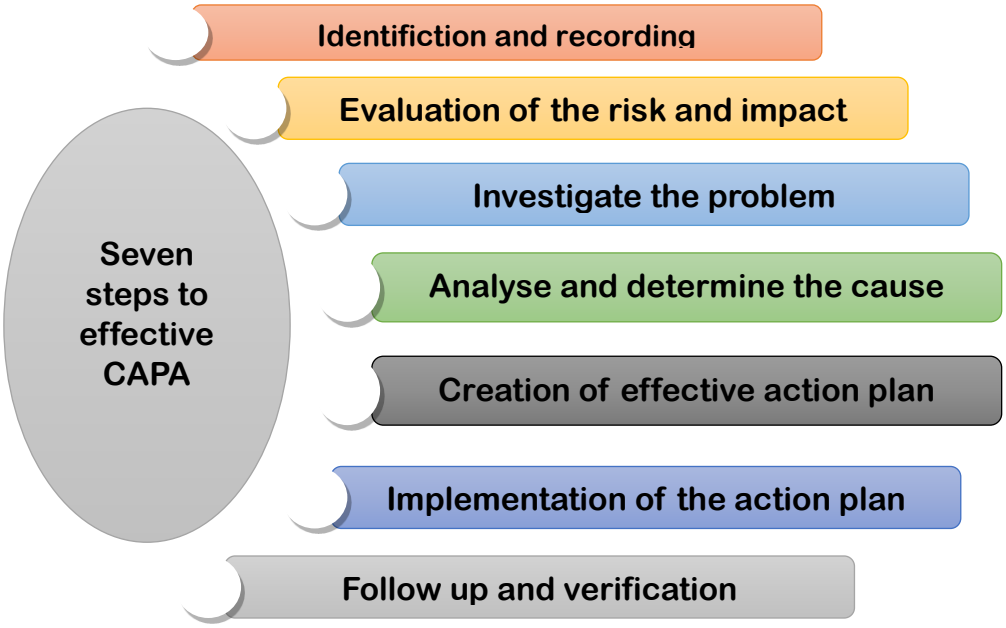

Figure1. 6: Stages of CAPA

I.3.9.1 Identification:

Identification defines an existing problem or a possible problem. It includes:

- Explanation of problems;
- Documentation of accessible evidence;
- Internal audits, process monitoring, data trends, customer complaints and QA inspection;

I.3.9.2 Evaluation:

The objective is to see the necessity for action and also the level of action required

- Potential impact – concerns the impact on company and clients in terms of cost, product quality, safety and customer satisfaction
- Risk – level of risk related to the matter
- Remedial action – potential impact and risk assessment to assist decide the desired remedial action

I.3.9.3 Investigation

- Review various parameters associated with a controversy e.g. equipment, materials, procedures, analyst training, software capabilities and environmental parameters
- Fix individual responsibilities and resources requirement such as finances, manpower and equipment. These requirements are worked out and documented

I.3.9.4 Analysis:

- Analysis leads to the root cause of the problem
- Collect data and try to list out the possible sources so as to arrive at the root cause. Data may come from such sources as records, processes, service information, operations, etc.
- Root cause will not simply deal with symptoms but help uproot the main contributing factors.

I.3.9.5 Action plan:

- Action plan is aimed at correcting and preventing future occurrence of failure. Plan involve to changes to be made in existing procedures and assignment of responsibilities
- All modifications and changes must be tell too concerned personnel, departments and suppliers
- Employee training is an essential part of any change and is always part of an action plan

I.3.9.6 Implementation:

Implementation stage involves:

- Execution of identified tasks
- Modification in documents
- Modification in processes
- Modification in environmental conditions
- Provision of training on modifications

NB: All stages in implementation stage need to be correctly documented

I.3.9.7 Follow up:

The follow up confirms completion of the identified tasks and also assesses the appropriateness and Effectiveness of the action taken. The validatory report must record:

- If the main cause of the problem is solved
- Any resulting secondary situations have been corrected
- Proper controls established to prevent future recurrence
- Actions taken that have no other adverse effects
- Adequate monitoring arrangements that are in place which assigned the responsibilities

I.3.10 What is CAPA Report:

A CAPA (Corrective and Preventive Action) Report is a tool, which, at the perspective of the regulator or the organization, would be used to identify, address, and prevent non-conformity. Compliance officers capture the key points of a problem or an incident on a CAPA Report form with mostly having captured a summary of the event, date of occurrence, items and other people involved, corrective actions taken, and preventive action established to avoid future recurrence. [36]

I.3.11 When to put in writing CAPA:

A CAPA report is flexible and can be used for all sorts of issues and incidents. But not every event needs a CAPA report. Quality teams need to use risk management techniques to look at the severity of an event and then decide if a CAPA report is needed.

CHAPITRE II:

ARTIFICIEL INTELLIGENCE IN PHARMACEUTICAL INDASTRY

Artificial intelligence (AI) has emerged as a transformative technology with vast potential to revolutionize various industries, including pharmaceuticals.

II.1 Artificiel intelligence:

II.1.1 Definition:

Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing and speech recognition and machine vision. [37]

Artificial intelligence focuses on simulating cognitive processes, which include the acquisition, storage, and application of knowledge. This knowledge is often divided into two types: procedural knowledge, which is generic and applies to multiple domains, and domain-specific knowledge. The goal of intelligent systems is to explicitly separate these two types of knowledge to facilitate the transfer of procedural knowledge from one domain to another.

II.1.2 Domain of application of AI:

Artificial intelligence is applied in various fields, including:

- Computer Vision: AI is used for image recognition and object detection in images. It is also applied in security surveillance and facial recognition. These applications are utilized in sectors such as security, medicine, industry, and logistics.[38]

- Natural Language Processing: It is also used for text generation and speech recognition. These applications are employed in sectors such as communication, research, healthcare, and finance.[39]

- Robotics: For motion planning, environment recognition, and decision-making in autonomous robots. These applications are used in sectors such as manufacturing, industry, healthcare, and scientific research.[40]

- Expert Systems: An expert system is software capable of simulating the behavior of a human expert performing a specific task. The success of artificial intelligence in this field is undeniable, owing to the targeted nature of the activity it is asked to simulate.[41]

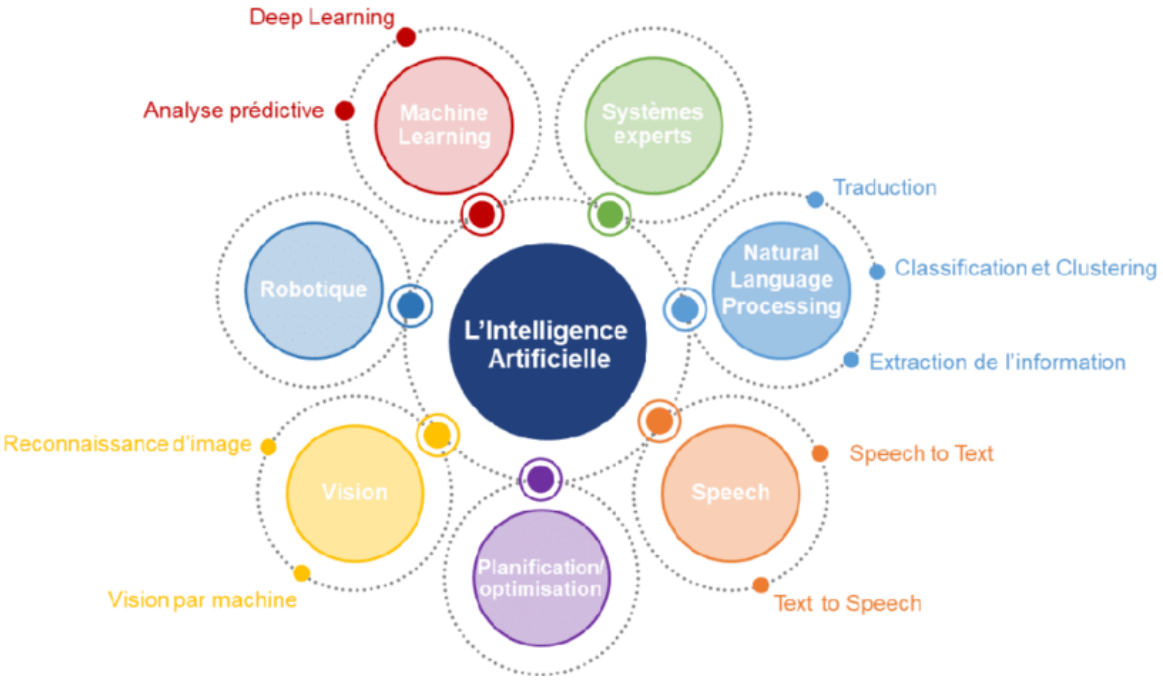

Figure II. 1: Overview of AI Domains

II.2 Machine Learning:

II.2.1 Definition:

Machine learning is a branch of computer science that broadly aims to enable computers to "learn" without being directly programmed. [42]

It has origins in the artificial intelligence movement of the 1950s and emphasizes practical objectives and applications, particularly prediction and optimization. Computers "learn" in machine learning by improving their performance at tasks through "experience". [43]

II.2.2 The Goals of Machine Learning:

The goal of ML, in simples words, is to understand the nature of (human and other forms of) learning, and to build learning capability in computers. To be more specific, there are three aspects of the goals of ML:

- To make the computers smarter, more intelligent. The more direct objective in this aspect is to develop systems (programs) for specific practical learning tasks in application domains.

- To develop computational models of human learning process and perform computer simulations. The study in this aspect is also called cognitive modeling.
- To explore new learning methods and develop general learning algorithms independent of applications.[44]

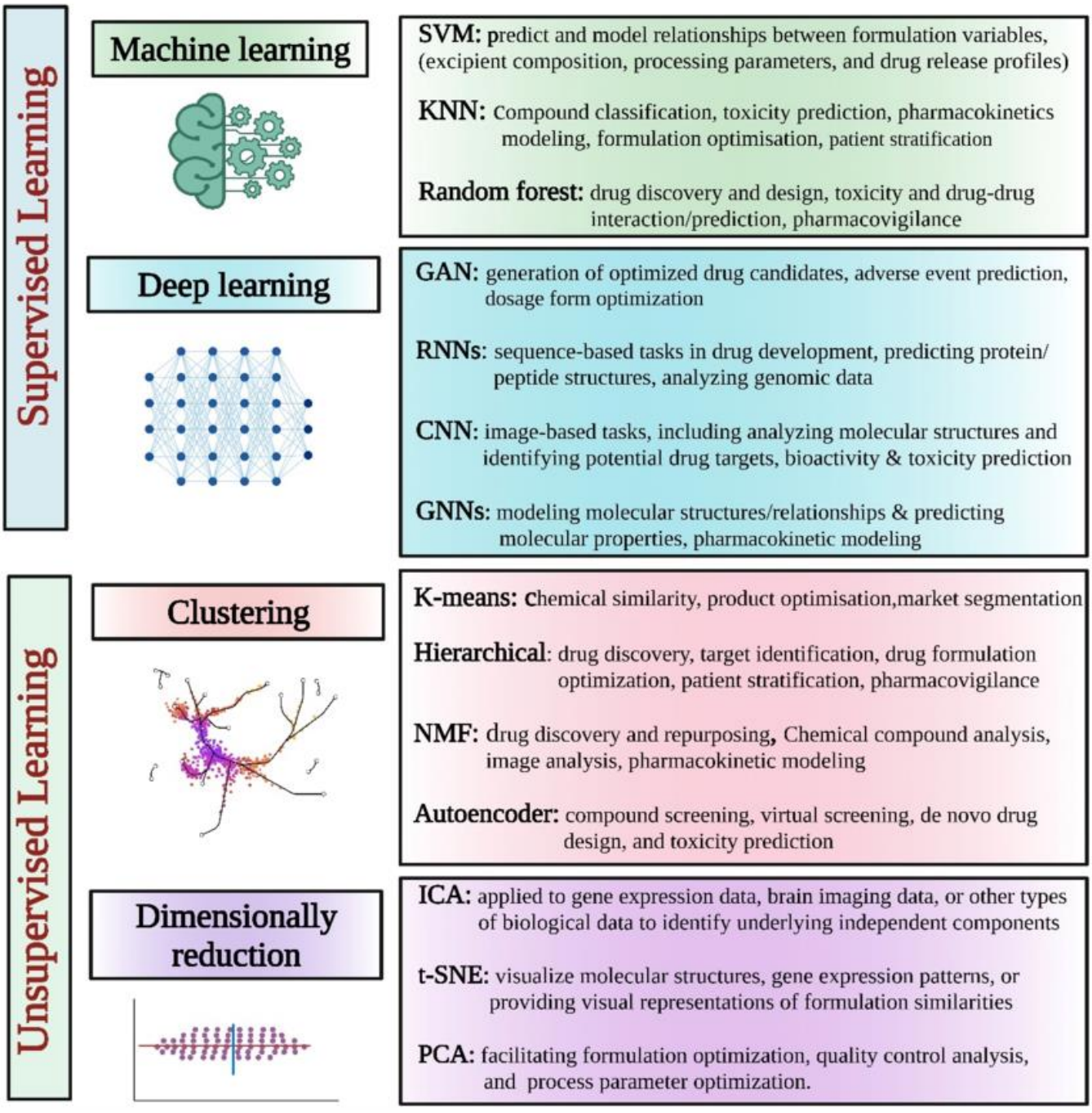

Figure II. 2: Supervised and Unsupervised AI models

II.3. Types of Machine Learning:

There are two types of machine learning

II.3.1 supervised AI learning:

Supervised learning refers to a type of machine learning in which an algorithm is trained on a labeled dataset, where the desired output is already known. The algorithm learns to map input data to the correct output by analyzing the patterns and relationships within the labeled data. This approach is commonly used in various applications, such as image recognition, natural language processing, and predictive modeling. Task-driven strategies involve setting specific goals for achieving desired outcomes from a given set of inputs. This approach utilizes labeled data to train algorithms for tasks such as data classification or outcome forecasting. The predominant supervised learning tasks are classification, which involves predicting a label, and regression, which involves predicting a quantity. These techniques include Naïve Bayes, K-nearest neighbors, support vector machines, ensemble learning, random forest, linear regression, support vector regression, and others [45]. It has several applications in the pharmaceutical industry, as described below

- Drug Discovery and Design: Supervised learning algorithms can be used to predict the activity or properties of new drug candidates. By training on a dataset of known compounds and their associated activities, the model can learn patterns and relationships between molecular features and desired outcomes. This enables the prediction of the activity, potency, or toxicity of novel compounds, aiding in drug discovery and design [46].
- Predictive Maintenance and Quality Control: In pharmaceutical manufacturing, supervised learning can be utilized for predictive maintenance and quality control. By training on data from manufacturing processes, equipment sensor data, or quality testing results, the model can learn to predict equipment failure, product quality deviations, or process abnormalities, allowing for proactive maintenance and quality assurance [47].
- Drug Target Identification: Supervised learning techniques can help identify potential drug targets by analyzing biological data. By training on data that include information about genetic, proteomic, or transcriptomic features and their relationship to drug response or disease progression, the model can learn patterns and identify potential targets for further investigation [48].

- Disease Diagnosis and Prognosis: Supervised learning models can be used to diagnose diseases or predict patient outcomes based on medical data. By training on labeled datasets containing patient characteristics, clinical data, and disease outcomes, the model can learn to classify patients into different disease categories or predict disease progression or treatment response [49].
- Adverse Event Detection: Supervised learning algorithms can be applied to pharmaco-vigilance data to identify and classify adverse events associated with drugs. By training on labeled adverse event reports, the model can learn to recognize patterns and identify potential safety signals, helping in the detection and characterization of adverse events [49].
- Predictive Modeling for Clinical Trials: Supervised learning can be used to predict outcomes in clinical trials. By training on historical clinical trial data, including patient characteristics, treatment interventions, and trial outcomes, the model can learn to predict patient response, treatment efficacy, or safety outcomes. This information can guide trial design and optimize patient selection [50].

II.3.2 Unsupervised Ai learning:

Unsupervised learning refers to a type of machine learning where the algorithm is not provided with labeled data. Instead, it is tasked with identifying patterns and relationships within the data on its own. This approach is often used in exploratory data analysis and can be useful for discovering hidden structures or clusters within a dataset. The approach being described is commonly known as a "data-driven methodology," which aims to extract patterns, structures, or insights from unannotated data. There are several prevalent unsupervised tasks, including clustering, dimensionality reduction, visualization, finding association rules, and anomaly detection. Various unsupervised learning tasks can be addressed using popular techniques such as clustering algorithms (e.g., hierarchical clustering, K-means, K-medoids, single linkage, complete linkage, BOTS), association learning algorithms, and feature selection and extraction techniques (e.g., Pearson correlation, principal component analysis) based on the data's characteristics [51-52].

Unsupervised learning techniques in AI can be valuable for pharmaceutical applications, particularly for exploratory analysis, pattern recognition, and data visualization, as described below:

- Clustering: Clustering algorithms group data points based on their similarities, allowing the identification of natural groupings or clusters within the data. In pharmaceutical applications, clustering can be applied to various datasets, such as gene expression profiles, chemical structures, or patient data, to uncover subgroups with similar characteristics. This can aid in target identification, patient stratification, and identifying distinct classes of compounds or diseases [53].
- Dimensionality Reduction: Dimensionality reduction techniques, such as principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE), are used to reduce the complexity of high-dimensional datasets while preserving meaningful information. These methods can help visualize and explore complex datasets, identify key variables or features, and support decision-making processes. Dimensionality reduction can be applied to various types of pharmaceutical data, including gene expression data, drug activity profiles, or imaging data [54].
- Anomaly Detection: Anomaly detection algorithms identify rare or unusual data points that deviate significantly from the expected patterns. In the pharmaceutical industry, anomaly detection can be useful for detecting adverse events, identifying potential safety concerns, and uncovering data quality issues. Unsupervised anomaly detection techniques, such as the local outlier factor (LOF) or isolation forest, can help highlight abnormal patterns or data points that warrant further investigation [54].
- Association Rule Mining: Association rule mining techniques, such as the Apriori algorithm, aim to discover interesting relationships or associations between items in a dataset. In the pharmaceutical context, association rule mining can be applied to drug– drug interactions, adverse event data, or co-occurrence patterns between medical conditions and medications. These techniques can provide insights into potential drug interactions, identify medication patterns, or support pharmacovigilance activities [55].
- Topic Modeling: Topic modeling algorithms, such as latent Dirichlet allocation (LDA), extract latent topics or themes from large text datasets. In the pharmaceutical industry, topic modeling can be used to analyze the scientific literature, clinical trial reports, or social media data to identify key research themes, emerging trends, or patient sentiments. This can aid in literature mining, competitive intelligence, or understanding patient perspectives [55-56].

II.4 Ai for drug discovery:

AI has revolutionized drug research and discovery in numerous ways. Some of the key contributions of AI in this domain include the following:

II.4.1 Target Identification:

AI systems can analyze diverse data types, such as genetic, proteomic, and clinical data, to identify potential therapeutic targets. By uncovering disease-associated targets and molecular pathways, AI assists in the design of medications that can modulate biological processes.

II.4.2 Virtual Screening:

AI enables the efficient screening of vast chemical libraries to identify drug candidates that have a high likelihood of binding to a specific target. By simulating chemical interactions and predicting binding affinities, AI helps researchers prioritize and select compounds for experimental testing, saving time and resources.

II.4.3 Structure-Activity Relationship (SAR):

Modeling AI models can establish links between the chemical structure of compounds and their biological activity. This allows researchers to optimize drug candidates by designing molecules with desirable features, such as high potency, selectivity, and favorable pharmacokinetic profiles.

II. 4.4 De Novo Drug Design:

Using reinforcement learning and generative models, AI algorithms can propose novel drug-like chemical structures. By learning from chemical libraries and experimental data, AI expands the chemical space and aids in the development of innovative drug candidates.

II.4.5 Optimization of Drug Candidates:

AI algorithms can analyze and optimize drug candidates by considering various factors, including efficacy, safety, and pharmacokinetics. This helps researchers fine-tune therapeutic molecules to enhance their effectiveness while minimizing potential side effects.

II.4.6 Drug Repurposing:

AI techniques can analyze large-scale biomedical data to identify existing drugs

That may have therapeutic potential for different diseases. By repurposing approved drugs for new indications, AI accelerates the drug discovery process and reduces costs.

II.4.7 Toxicity Prediction

AI systems can predict drug toxicity by analyzing the chemical structure and characteristics of compounds. Machine learning algorithms trained on toxicology databases can anticipate harmful effects or identify hazardous structural properties. This helps researchers prioritize safer chemicals and mitigate potential adverse responses in clinical trials.

Overall, AI-driven approaches in drug research and development offer the potential to streamline and expedite the identification, optimization, and design of novel therapeutic candidates, ultimately leading to more efficient and effective medications [57].

CHAPTER III:

DESIGN AND IMPLEMENTATION OF MACHINE LEARNING

In this chapter, we delve into the intricacies of designing and implementing ML algorithms tailored to the unique requirements of the pharmaceutical industry. Through a step-by-step approach, we elucidate the process of developing ML models for identifying and addressing deviations

III.1 System Design:

III.1.1 Presentation of python:

Python is an interpreted, high-level programming language. Its design philosophy is intended to be readable and expressive. Conceived by Guido van Rossum and released in 1991, it has since widely found adoption not only in web development but also in scientific computation, data analysis, artificial intelligence, and machine learning. [58]

- **Key Features of Python:**

1. Readability and Simplicity: Python's syntax is designed to be simple and intuitive, making it easy to read and write code. This characteristic not only reduces the cost of program maintenance but also facilitates collaboration among developers.

2. Interpretation: Python is an interpreted language, which means that code is executed line by line. This feature simplifies debugging and testing, as developers can quickly identify and rectify errors.

3. Cross-Platform Compatibility: Python is supported on all major operating systems, including Windows, macOS, and Linux. This cross-platform compatibility ensures that Python applications can run seamlessly across different environments.

4. Rich Standard Library: Python comes with an extensive standard library that provides support for various programming tasks, such as file I/O, networking, and data manipulation. This built-in functionality eliminates the need for developers to write code from scratch for common tasks, thereby enhancing productivity.

5. Vast Ecosystem of Third-Party Libraries: Python boasts a thriving ecosystem of third-party libraries and frameworks that extend its functionality for specific use cases. These libraries cover a wide range of domains, including web development (e.g., Django, Flask), data analysis (e.g., NumPy, pandas), scientific computing (e.g., SciPy), machine learning (e.g., TensorFlow, PyTorch), and more. The availability of these libraries allows developers to leverage existing solutions and accelerate the development process.

- **Applications of Python**:

1. Web Development: Python is widely used for web development, thanks to frameworks like Django and Flask. These frameworks provide developers with the tools needed to build scalable and maintainable web applications.

2. Data Analysis and Visualization: Python is the language of choice for data analysis and visualization, with libraries like NumPy, pandas, and Matplotlib offering powerful tools for working with data.

3. Artificial Intelligence and Machine Learning: Python has become the de facto language for artificial intelligence and machine learning projects. Frameworks like TensorFlow, PyTorch, and scikit-learn provide developers with the tools needed to build and train machine learning models efficiently.

4. Scientific Computing: Python is widely used in the scientific computing community, thanks to libraries like SciPy and SymPy, which provide tools for numerical computing and symbolic mathematics, respectively.

III.2 Implementation:

III.2.1 Core Concepts:

- **Data:** The foundation of ML. It includes the input features (variables) and the target variable (what you want to predict).
- **Features:** Individual measurable properties or characteristics of the data. In a dataset, each feature represents a column.
- **Model:** A mathematical representation of a real-world process. In ML, models are trained using algorithms to learn patterns from data.
- **Algorithm:** A set of rules or instructions given to an ML model to help it learn from the data. Different algorithms are used for different types of tasks (e.g., regression, classification, clustering).

III.2.2 Internal ML architecture:

The internal architecture of machine learning typically involves layers of algorithms and data processing stages. At its core, it involves data input, feature extraction, model training, and evaluation.

- **Data Input:**
  This is the initial stage where raw data is collected or generated. This could be structured data (like tables in a database) or unstructured data (like text, images, or audio).
- **Preprocessing:**
  Raw data often needs to be cleaned and preprocessed before it can be used for training a model. This involves tasks such as removing noise, handling missing values, normalization, and transforming data into a suitable format for analysis.
- **Feature Extraction/Selection:**
  In this stage, relevant features are extracted from the preprocessed data. Feature extraction involves identifying important characteristics or patterns in the data that can help the model make predictions. Feature selection is the process of choosing a subset of the most relevant features to reduce dimensionality and improve model performance.
- **Model Selection/Training:**
  This is where the actual machine learning algorithm comes into play. Depending on the problem and the type of data, different algorithms can be chosen, such as decision trees, neural networks, support vector machines, etc. The selected model is trained on the preprocessed data, which involves adjusting its parameters to minimize the error between predicted and actual outcomes.
- **Evaluation:**
  Once the model is trained, it needs to be evaluated to assess its performance. This is typically done using a separate dataset called the validation or test set, which the model hasn't seen during training. Evaluation metrics such as accuracy, precision, recall, F1-score, etc., are calculated to measure how well the model generalizes to new data.
- **Hyperparameter Tuning**:
  Many machine learning algorithms have hyperparameters that need to be tuned to optimize model performance. This involves selecting the best combination of hyperparameters through techniques like grid search, random search, or Bayesian optimization.

- **Deployment:**

  After the model has been trained and evaluated, it can be deployed into production to make predictions on new, unseen data. Deployment involves integrating the model into existing systems or applications, often through APIs or other interfaces.

III.2.3 Steps in the Machine Learning Process:

- **Problem Definition**:

  Clearly define the problem we aim to solve. Identify the inputs (features) and the desired output (target variable).

- **Data Collection**:

  Gather data relevant to the problem. We were able to collect the data from the compliance office. the data contains “database for mena corporate audit observation 11/09/2023”.

➢ After gathering the necessary data, now we can start coding, we used google colab (a short from google collaborator) it is a cloud-based platform provided by google that allows us to write and execute python codes in a browser, particularly suited for machine learning and data analysis tasks .

Figure III. 1 Google Colab

- **Data Preprocessing**:
  - **Cleaning**: we removed noise and errors from the data,we handled the missing values and removed duplicates.
  - **Transformation**: Convert data into a format suitable for analysis
  - **Splitting**: Divide the data into training and testing sets to evaluate the model's performance.

```python
# import libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.linear_model import SGDClassifier
from sklearn.metrics import classification_report
from google.colab import drive
```

Figure III. 2: Python libraries

```python
data = pd.read_excel(dataset_path)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Figure III. 3: Data preprocessing

```python
X = data_combined["Observation Description"].astype(str)
y = data_combined["Site Response"].astype(str)
```

Figure III. 4: Variable identification

- **Exploratory Data Analysis (EDA):**

  Analyze the data to understand its structure and underlying patterns.

```python
vectorizer = CountVectorizer(max_features=5000)
X_train_counts = vectorizer.fit_transform(X_train)
X_test_counts = vectorizer.transform(X_test)
```

Figure III. 5: Data Analysis

- **Feature Engineering:**

  Create new features or modify existing ones to improve the model's performance. Techniques include creating interaction terms, polynomial features, and aggregating features.

- **Model Selection:**

  Choose an appropriate algorithm based on the problem type (e.g., regression, classification) and the dataset characteristics.

- **Model Training :**

  Train the model using the training dataset. The model learns patterns in the data by adjusting its parameters to minimize the error between predictions and actual outcomes.

- **Model Evaluation :**

  Evaluate the model's performance on the test dataset using metrics such as accuracy, precision, recall, and F1-score.

```
y_pred = clf.predict(X_test_vec)
```

Figure III. 6: Model Evaluation

- **Model Interpretation :**

  Interpret the model's results to understand how it makes predictions.

```
[ ]  accuracy = accuracy_score(y_test, y_pred)
     print("Accuracy Score:", accuracy)
```

Figure III. 7: Model Accuracy

CHAPTER IV:

RESULTS AND DISCUSSION

In the final chapter of this dissertation, we present the outcomes of integrating machine learning into the pharmaceutical manufacturing process for automatic corrective and preventive actions (CAPA).

IV.1 Results:

The implementation of machine learning algorithms for CAPA yielded promising results. Initially, the system achieved an accuracy rate of 15%, which fell short of our expectations. Through rigorous analysis and experimentation, we identified several factors contributing to this limitation, including:

- **Data Quality:** The quality of the data collected from various sources posed a significant challenge. Inconsistent data formatting, missing values, and noise adversely impacted the performance of the machine learning models**.**
- **Feature Selection:** Identifying the most relevant features for training the models proved to be a complex task. The extensive range of variables and their interdependencies necessitated careful consideration to avoid overfitting or underfitting**.**
- **Model Complexity:** Balancing model complexity with interpretability was another challenge. Complex models tended to perform well on training data but struggled to generalize to unseen instances, while simpler models lacked the capacity to capture intricate patterns in the data.

Despite these challenges, the integration of historical data with new observations led to a significant improvement in the system's performance. By leveraging the insights gained from past incidents and their corresponding corrective actions, the machine learning models were able to enhance their predictive capabilities.

Upon incorporating the old data with the new ones, we observed a notable increase in accuracy, surpassing our initial target of 75%. The combination of historical knowledge with real-time data enabled the system to learn from past mistakes and adapt to evolving scenarios more effectively.

```
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
```

```
Accuracy: 0.1554959785522788
```

Figure IV. 1: Accuracy Score before fixing the issue

```
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy Score:", accuracy)

Accuracy Score: 0.7566448049706593
```

Figure IV. 2: Accuracy Score after fixing the issue

IV.2 Discussion:

The results obtained from our study underscore the importance of continuous learning and adaptation in the context of CAPA. While the initial accuracy fell short of expectations, the iterative process of refinement and augmentation proved instrumental in enhancing the system's performance.

The challenges encountered during the implementation phase highlighted the complexities inherent in leveraging machine learning for CAPA purposes. Addressing issues related to data quality, feature selection, and model complexity required a multidisciplinary approach encompassing domain expertise, data engineering, and machine learning techniques.

the success achieved through the integration of old and new data underscores the value of leveraging historical knowledge in enhancing predictive analytics capabilities. By harnessing the collective wisdom embedded within past incidents and corrective actions, organizations can proactively identify and mitigate risks before they escalate into critical issues.

IV.3 Web application for our machine learning:

IV.3.1 Presentation of our web application "MedAITron":

After editing the code and fixing all the previous problems we are now facing a new challenge which is transforming all the job done so far to a visual application which can be seen by others.

We were able to develop a web application to integrate and deploy our machine learning to streamline processes and enhance accessibility.

"MedAITron" interface is designed with simplicity .it features a clean layout with a single text input box where user can insert the problem. Adjacent to the input box is displayed button labeled "predict", and a layer with a text output box which provide the predicted solution.

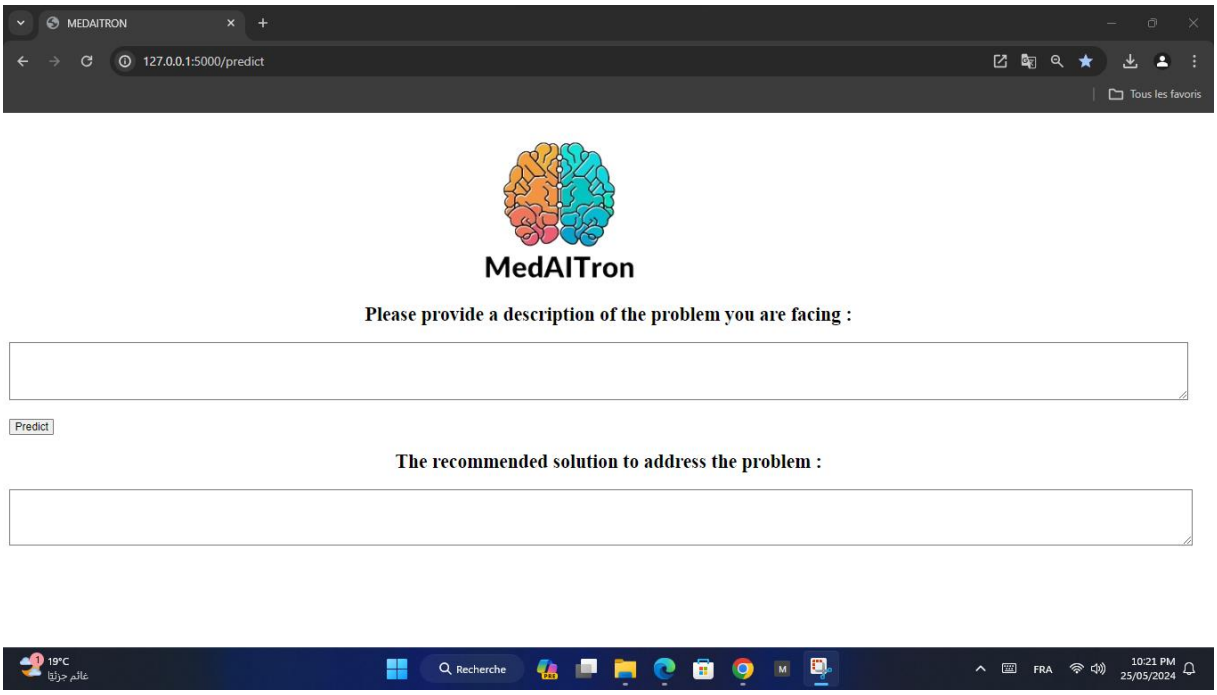

Figure IV. 3: The Inerface of the Web app

IV.3.2 An applicable example:

To understand how our web app operate here is an applicable example

Problem description: stability testing reveals that a batch of drug degrades faster than expected.

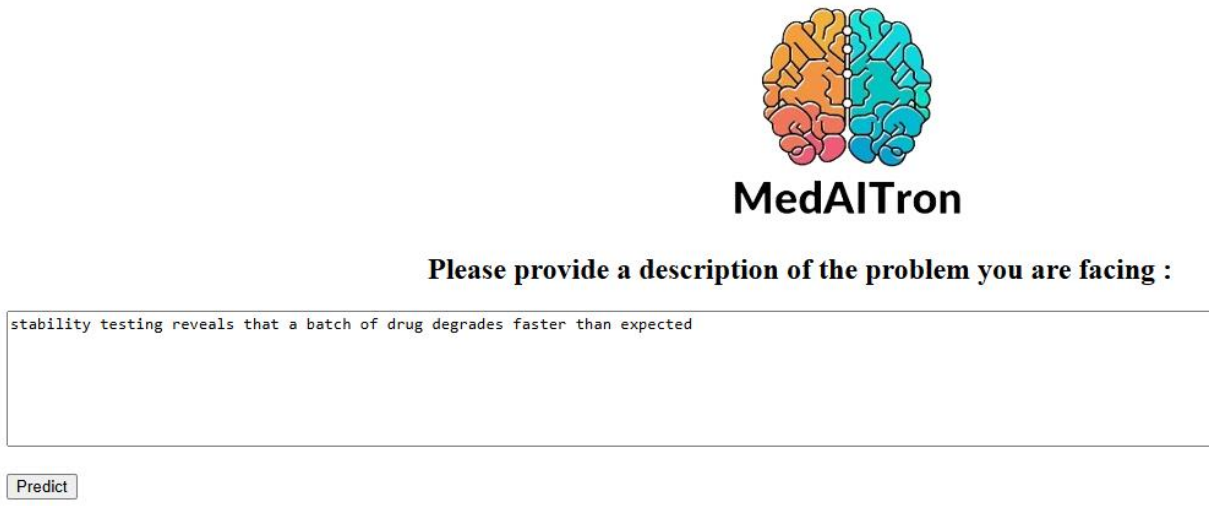

Figure IV. 4: Problem Discription

And then we press the “predict” bottom.

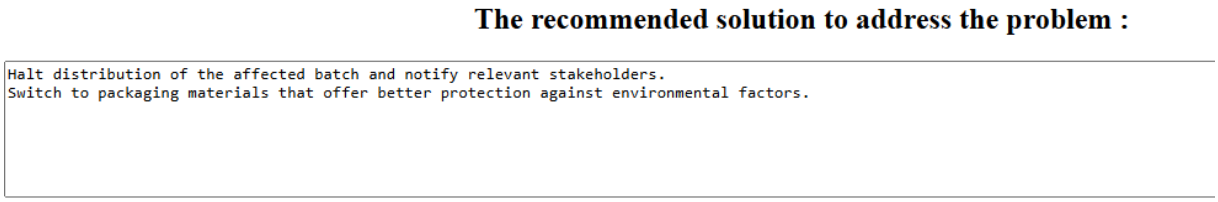

Figure IV. 5: The suggested solution

CONCLUSION:

In conclusion, the research conducted on integrating machine learning into automating Corrective Action and Preventive Action (CAPA) processes demonstrates its potential to significantly enhance quality management practices within organizations. Through comprehensive analysis and experimentation, it has been evidenced that machine learning algorithms can effectively streamline CAPA workflows, leading to faster issue resolution, proactive risk mitigation, and improved overall quality performance. However, while the findings highlight the promising benefits of ML-driven CAPA systems, several challenges and considerations have been identified.

These include the need for robust data infrastructure, algorithmic transparency, and ongoing human oversight to ensure the reliability and accountability of automated decision-making processes. Additionally, ethical concerns such as bias mitigation and fairness in algorithmic outcomes must be carefully addressed to uphold integrity and trust in the implementation of ML technologies. Despite these challenges, the results of this thesis underscore the transformative potential of integrating machine learning in CAPA processes, paving the way for future advancements in quality management practices and organizational excellence.

Moving forward, continued research efforts are imperative to optimize ML algorithms for specific manufacturing contexts, validate their performance, and ensure seamless integration into existing quality management frameworks. Moreover, fostering transparency and accountability in ML-driven decision-making processes will be pivotal in fostering industry-wide acceptance and adoption.

By embracing innovation and leveraging advanced technologies, the pharmaceutical sector stands poised to elevate its quality management standards, thereby advancing regulatory compliance and enhancing patient welfare in a dynamic and ever-evolving landscape.

REFRENCES:

[1] Sower, Victor E. Essential of Quality with cases and experiential exercises. New York; John Wiley and Sons, 2011.

[2] https://www.superfastcpa.com/what-is-quality-of-conformance

[3] Rose and H. Kenneth. Project Quality Management: Why, What and How, J. Ross Publishing, Fort Lauderdale, Florida, 2005, pp. 41.

[4] A. Lakshmana Rao, M. Bharani, Y. Madhu and G. G. Sankar. TQM: The Need of New Era, Indian J. Pharm. Educ. Res. 39(4): 203-206 (2005).

[5] Wikipedia the encyclopedia, http://en.wikipedia.org/wiki/Total_quality_manage ment accessed on 3/2/2014.

[6] Hoyle and David. Quality Management Essentials, Oxford, United Kingdom: Butterworth Heinemann, 2007, pp. 200.

[7]. Pfeifer and Tilo. Quality Management: Strategies, Methods, Techniques, Munich, Germany: Carl Hanser Verlag, 2002, pp. 5

[8] A. L. Rao. TQM: The need of new era, Indian Journal of Pharmaceutical Education 39(4): 203 (2005).

[9] M. Bhaskar, B. Sanjib and Y. Abhishek. TQM in pharmaceuticals: A review, Int. J. Pharm Tech Res. 3(1): 365-375 (2011).

[10] N. Case. Six Sigma Black Belts stamp out tough quality problems, Quality on line, 1999, available at http://qualitymag.com/articles/1999/aug99/0899 f3.html accessed on 25/2/2014.

[11] De Feo, A. Joseph and W. Barnard. Juran institute's six sigma breakthrough and beyond - quality performance breakthrough methods, Tata McGraw-Hill Publishing Company Limited, 2005

[12] E. Griffith. Risk management programs for the pharmaceutical industry, Fujitsu Consulting, white paper, pharmaceutical industry, 2004.

[13] “ICH Q9,” 2005. Quality risk management.

[14] Wikipedia the encyclopedia, http://en.wikipedia.org/wiki/Quality_by_Design accessed on 6/4/2014.

[15] “ICH Q8 (R2),” 2009. Pharmaceutical development

[16] S. Kumar and A. Baldi.Design of experiment based statistical optimization in life science research: An overview, Pharm Aspire 4: 35-43 (2013).

[17]. S. Kumar, S. Thakur, D. Tanwar and A. Baldi. A quantitative approach for pharmaceutical quality by design patterns, Inventi Rapid: Pharm Analysis & Quality Assurance 4: 1-8 (2012)

[18] “WHO,” guidelines on good manufacturing practices for herbal medicines, 2007, available at http://apps.who.int/medicinedocs/documents/s14 215e/s14215e.pdf accessed on 16/4/2014

[19] J. B. Karmacharya. Good Manufacturing Practices (GMP) for Medicinal Products, available at http://cdn.intechopen.com/pdfs-wm/37170.pdf accessed on 16/4/2014.

[20] M. G. De Marinis, M. Piredda, M. C. Pascarella, B. Vincenzi, F. Spiga, D. Tartaglini, R. Alvaro, M. Matarese. 'If it is not recorded, it has not been done!'? consistency between nursing records and observed nursing care in an Italian hospital, J Clin Nurs. 19(11-12): 1544-52

[21] https://www.who.int/immunization_standards/vaccine_ quality/risk_july_2013.pdf

[22] Difference between Deviation and Incident in Pharmaceutical. Pharma Pathway 2016. http://pharmapathway.com/differencebetween-deviation-and-incident-in-pharmaceutical/

[23]https://www.pharmout.net/wp-content/uploads/2016/07/2015_GMP_Validation_Forum_DeviationIncident-Non-conformance-Systems.pdf

[24] Appropriate Handling of Out-of-Calibration Equipment. Quality Digest 2009. https://www.qualitydigest.com/inside/metrology-article/appropriate-handling-out-calibrationequipment.html

[25] Compliance I. As-Found: Out-of-Tolerance – What to do next? [Internet]. In Compliance Magazine. 2019 [cited 6 August 2019]. Available from: https://incompliancemag.com/article/

[26] Measuring the performance of quality assurance processes: pharmaceutical industry deviation management case study. Academia 2010. https://www.academia.edu/16285499/Measuring_the_performance_of_quality_assurance_processes_pharmaceutical_industry_deviation_management_case_study

[27] Reporting Failures and Process Deviations: A Closed-Loop Approach. Pharma Manufacturing 2010.

[28] https://web.njit.edu/~armenant/PhEn602-102/PhEn602Spring09-Notes2.pdf

[29] JOURNAL OF ENGINEERING MANAGEMENT AND COMPETITIVENESS (JEMC) Vol. 1, No. 1/2, 2011, 16-20

[30] definition/abbreviation from procedure: N°FPG-043 version 11

[31] Andy Marker. The Beginner's Guide to CAPA. Definition. Smart sheet. 2018.

[32] CAPA|Beginner's Guide to Corrective Action Preventive Action. Difference between Corrective Action and Preventive Action.Softwaretestinghelp.com. 2021.

[33] Corrective and Preventive Actions (CAPA)| FDA. Objective.2014.

[34] Ankur Choudhary. Importance of Corrective and Preventive Action (CAPA). Importance. Pharmaceutical guidelines.

[35] Robert Fenton. 5 Fundamentals of the CAPA Quality Process.
Fundamental of CAPA. Qualio. 2020.

[36] Dr. Deepak. What are Different stages of CAPA in Quality Control Operations? Procedure of CAPA. Lab.Training.com.

[37] What is artificial intelligence (AI)? ByNicole Laskowski

[38] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).

[39] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. Open AI Blog, 1(8), 9.

[40] Kober, J., & Peters, J. (2010). Reinforcement learning in robotics: A survey. The International Journal of Robotics Research, 32(11), 1238-1274.

[41] Réseaux neuronaux : Une introduction accompagnée d'un modèle Java de Jean-Philippe Rennard, 2006;

[42] Samuel AL. Some studies in machine learning using the game of checkers. *IBM J ResDev*. 1959;3(3):210–229

[43] Mitchell TM. *Machine Learning*. 1st ed. New York, NY: McGraw-Hill Education; 1997

[44] https://csc.lsu.edu/~jianhua/csc7333-intro-0

[45] Sarker, I.H. AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. SN Compute. Sci. **2022**, 3, 158. [CrossRef]

[46] Dara, S.; Dhamercherla, S.; Jadav, S.S.; Babu, C.M.; Ahsan, M.J. Machine Learning in Drug Discovery: A Review. Artif. Intell. Rev.**2022**, 55, 1947–1999. [CrossRef]

[47] Kavasidis, I.; Lallas, E.; Gerogiannis, V.C.; Charitou, T.; Karageorgos, A. Predictive Maintenance in Pharmaceutical Manufacturing Lines Using Deep Transformers. Procedia Comput. Sci. **2023**, 220, 576–583. [CrossRef]
[48] Bagherian, M.; Sabeti, E.; Wang, K.; Sartor, M.A.; Nikolovska-Coleska, Z.; Najarian, K. Machine Learning Approaches and Databases for Prediction of Drug–Target Interaction: A Survey Paper. Brief. Bioinform. **2021**, 22, 247–269. [CrossRef]
[49] Kumar, Y.; Koul, A.; Singla, R.; Ijaz, M.F. Artificial Intelligence in Disease Diagnosis: A Systematic Literature Review, Synthesizing Framework and Future Research Agenda. J. Ambient. Intell. Humaniz. Comput. **2023**, 14, 8459–8486. [CrossRef]
[50] Chapman, A.B.; Peterson, K.S.; Alba, P.R.; DuVall, S.L.; Patterson, O.V. Detecting Adverse Drug Events with Rapidly Trained Classification Models. Drug Saf. **2019**, 42, 147–156. [CrossRef]
[51] Elkin, M.E.; Zhu, X. Predictive Modeling of Clinical Trial Terminations Using Feature Engineering and Embedding Learning. Sci. Rep. **2021**, 11, 3446. [CrossRef]
[52] Chavda, V.P.; Sorathiya, A.; Valu, D.; Marwadi, S. Role of Data Mining in Bioinformatics. In Computation in Bioinformatics: Multidisciplinary Applications; John Wiley & Sons: Hoboken, NJ, USA, 2021; pp. 69–84. [CrossRef]
[53] Parikh, S.; Patel, R.; Khunt, D.; Chavda, V.P.; Vora, L. Data Analytics and Data Visualization for the Pharmaceutical Industry. In Bioinformatics Tools for Pharmaceutical Drug Product Development; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2023; pp. 55–76. ISBN 978-1-119-86572-8.
[54] Karim, M.R.; Beyan, O.; Zappa, A.; Costa, I.G.; Rebholz-Schuhmann, D.; Cochez, M.; Decker, S. Deep Learning-Based Clustering Approaches for Bioinformatics. Brief. Bioinform. **2021**, 22, 393–415. [CrossRef]
[55] Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; et al. Applications of Machine Learning in Drug Discovery and Development. Nat. Rev. Drug Discov. **2019**, 18, 463–477. [CrossRef]
[56] Goldstein, M.; Uchida, S. A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. PLoS ONE **2016**, 11, e0152173. [CrossRef] [PubMed]
[57] Shah, H.; Chavda, V.; Soniwala, M.M. Applications of Bioinformatics Tools in Medicinal Biology and Biotechnology. In Bioinformatics Tools for Pharmaceutical Drug Product Development; Wiley: Hoboken, NJ, USA, 2023; pp. 95–116. ISBN 978-1-11986572-8.
[58] https://www.coursera.org/in/articles/what-is-python-used-for-a-beginners-guide-to-using-python