People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

Ferhat Abbas University – Sétif 1
Faculty of Sciences

MASTER'S DEGREE

Specialization: F3I – Fondements et Ingénierie de l'Information et de l'Image

---

Personalization of Dynamic Drug Dosing in Intensive Care Using Offline Reinforcement Learning on Real Hospital Data

---

Presented by:

Sebia Kods Chellal Abir

Before the Jury Composed of:

Supervisor: Dr. Ferhat Hamida Zineb
President of the jury: Dr. Zenbout Imene
Examiner: Dr. Laouarem Ayoub

Defended on June 2026:

Academic Year: 2025–2026

Abstract

Personalized insulin dosing in Intensive Care Units (ICUs) requires continuous adaptation to rapidly changing patient conditions, making it a suitable application for data-driven decision support. This thesis explores the use of Offline Reinforcement Learning (Offline RL) to learn insulin dosing policies from historical clinical data without interacting with patients during training.

Using the MIMIC-III critical care database, a complete pipeline was developed, including cohort selection, data preprocessing, missing-value handling, and the construction of a Markov Decision Process (MDP) representing the insulin dosing problem. Two state-of-the-art Offline RL algorithms, Batch-Constrained Q-Learning (BCQ) and Conservative Q-Learning (CQL), were implemented and trained on real ICU trajectories.

The learned policies were evaluated using multiple complementary approaches, including Q-value analysis, Fitted Q Evaluation (FQE), clinician agreement rates, and glycemic safety metrics. Experimental results indicate that CQL provides more stable value estimation and better overall safety characteristics than BCQ while maintaining a high level of agreement with historical clinical decisions.

To demonstrate practical applicability, the best-performing policy was integrated into *IntelliGlu*, a prototype clinical decision support platform combining policy inference, safety mechanisms, and physiological simulation. The proposed framework illustrates the feasibility of applying Offline RL to personalized insulin dosing and provides a reproducible foundation for future research on adaptive treatment optimization in critical care.

Keywords: Offline Reinforcement Learning, Intensive Care Unit, Insulin Dosing, Conservative Q-Learning, Batch-Constrained Q-Learning, MIMIC-III, Clinical Decision Support Systems.

Résumé

Le dosage personnalisé de l'insuline en réanimation nécessite une adaptation continue aux conditions changeantes du patient, ce qui en fait une application idéale pour les systèmes d'aide à la décision fondés sur les données. Cette thèse explore l'utilisation de l'apprentissage par renforcement hors-ligne pour apprendre des politiques de dosage à partir de données cliniques historiques, sans interaction directe avec les patients. À partir de la base de données MIMIC-III, un pipeline complet a été développé, comprenant la sélection de cohorte, le prétraitement des données et la construction d'un processus de décision de Markov. Deux algorithmes, BCQ et CQL, ont été entraînés et évalués via l'analyse des Q-valeurs, l'évaluation FQE, le taux d'accord avec les cliniciens et les métriques de sécurité glycémique. CQL a démontré une estimation plus stable et de meilleures caractéristiques de sécurité. La meilleure politique a été intégrée dans *IntelliGlu*, une plateforme prototype d'aide à la décision clinique.

Mots-clés : Apprentissage par renforcement hors-ligne, Dosage de l'insuline, Réanimation, BCQ, CQL, Aide à la décision, MIMIC-III.

ملخص

يتطلب جرعات الأنسولين الشخصية في وحدات العناية المركزة تكيفًا مستمرًا مع الحالات المتغيرة للمريض، مما يجعلها تطبيقًا مناسبًا لأنظمة دعم القرار المبنية على البيانات. تستكشف هذه الأطروحة استخدام التعلم التعزيزي غير المتصل
لتعلم سياسات جرعات الأنسولين من البيانات السريرية التاريخية دون التفاعل مع المرضى أثناء التدريب. باستخدام
قاعدة بيانات III-CIMIM، تم تطوير خط أنابيب متكامل يشمل اختيار المجموعة، والمعالجة المسبقة للبيانات، وبناء
عملية قرار ماركوف. تم تدريب خوارزميتين QCB وLQC وتقييمهما من خلال تحليل seulav-Q وتقييم EQF
ومعدلات الاتفاق مع الأطباء ومقاييس السلامة الجلوكوزية. أظهر LQC تقديرًا أكثر استقرارًا وخصائص أمان
أفضل. تم دمج أفضل سياسة في منصة ulGilletnI، وهي منصة نموذجية لدعم القرار السريري.

الكلمات المفتاحية: التعلم التعزيزي غير المتصل، جرعات الأنسولين، العناية المركزة، QCB، LQC، دعم القرار السريري،
III-CIMIM.

Contents

Abstract **2**

Résumé **3**

ملخص **4**

General Introduction **12**

1 Foundations of Reinforcement Learning **14**

1.1 Introduction 14

1.2 Artificial Intelligence in Healthcare 14

1.2.1 Applications of AI in Medicine 15

1.2.2 Challenges of AI in Healthcare 15

1.3 Reinforcement Learning Fundamentals 16

1.3.1 Limitations of Machine Learning 16

1.3.2 Agent and Environment 17

1.3.3 Reward Mechanism 17

1.3.4 Policy and Value Function 18

1.3.5 Exploration vs Exploitation 18

1.4 Markov Decision Process 19

1.4.1 States 19

1.4.2 Actions 19

1.4.3 Transition Probability 20

1.4.4 Reward Function 20

1.4.5 Discount Factor 20

1.5 Deep Reinforcement Learning 21

1.5.1 Q-Learning 22

1.5.2 Deep Q-Networks 22

1.5.3 Policy Gradient Methods....23
1.6 Offline Reinforcement Learning....24
1.6.1 Definition....24
1.6.2 Offline RL Algorithms....24
1.6.3 Extrapolation Error....24
1.6.4 Distribution Shift....26
1.6.5 Safe Learning from Logged Data....26
1.7 Conclusion....27

2 Clinical Background and Insulin Dosing in ICU 28

2.1 Introduction....28
2.2 ICU Environment and Patient Monitoring....28
2.2.1 Pharmacological Challenges in Critical Care....28
2.2.2 Alterations in Pharmacokinetics and Pharmacodynamics (PK/PD) 29
2.3 Pharmacological Challenges in Critical Care....30
2.3.1 Impact of Critical Illness on Drug Response and Disposition....30
2.3.2 Inter- and Intra-Patient Variability....30
2.3.3 The Narrow Therapeutic Index....32
2.4 Insulin Therapy in ICU....33
2.4.1 Role of insulin....33
2.4.2 Types of insulin....34
2.4.3 Principles of insulin dosing....35
2.4.4 Clinical Indications and Mechanism of Action....36
2.4.5 Current Standard Guidelines vs. Real-World Needs....37
2.5 Leveraging Real Hospital Data....38
2.5.1 Electronic Health Records (EHR) in the ICU....38
2.5.2 Challenges of Clinical Data....39

2.6 The Paradigm Shift: From Static Guidelines to Sequential Decision-Making 40

2.6.1 Limitations of Traditional Supervised Learning 40

2.6.2 The Clinical Need for Offline Reinforcement Learning 41

2.7 Conclusion 42

3 Data Pipeline and Algorithmic Framework 44

3.1 Introduction 44

3.2 Cohort Selection and Data Extraction 44

3.2.1 Inclusion and Exclusion Criteria 44

3.2.2 Cohort Profile and Data Properties 45

3.2.3 Feature Selection 46

3.3 Data Preprocessing and Imputation 47

3.3.1 Handling Missing Data and Irregular Sampling 47

3.3.2 Outlier Detection and Normalization 47

3.3.3 Data Integrity Verification 48

3.4 Mapping the Clinical MDP 48

3.4.1 Action Space Design 51

3.4.2 Reward Function Engineering 53

3.5 Problem Formulation (MDP Recap) 56

3.6 Offline Reinforcement Learning Algorithms 57

3.6.1 Batch-Constrained Q-Learning (BCQ) 58

3.6.2 Conservative Q-Learning (CQL) 59

3.7 Network Architecture 60

3.7.1 Shared Backbone 60

3.7.2 Parameter Counts 61

3.8 Training Procedure 62

3.8.1 Hyperparameter Summary 62

3.8.2 BCQ Training Procedure 62

3.8.3 CQL Training Procedure 63

3.8.4 Early Stopping and Reproducibility 64

3.9 Evaluation Framework 64

3.9.1 Evaluation Principle 65

3.9.2 Safety Definition 65

3.9.3 Policy Evaluation Metrics 66

3.10 Experimental Results 66

3.10.1 BCQ Training Dynamics 66

3.10.2 CQL Training Dynamics 67

3.10.3 Comparative Results 68

3.11 Conclusion 69

4 Policy Evaluation and Clinical Safety Assessment 71

4.1 Introduction 71

4.2 Experimental Evaluation Protocol 71

4.3 Q-Based Offline Policy Evaluation 72

4.3.1 Q-Value Analysis 72

4.4 Off-Policy Value Estimation with FQE 73

4.4.1 FQE Methodology 73

4.4.2 FQE Summary Statistics 74

4.5 Policy Agreement with Clinicians 76

4.5.1 Action Distribution Analysis 77

4.6 Glycaemic Outcome Analysis 78

4.6.1 Hypoglycaemia and Hyperglycaemia 78

4.7 Clinical Safety Assessment 79

4.8 Comparative Analysis of BCQ and CQL 80

4.9 Discussion 81

4.10 Limitations 82

4.11 Conclusion 83

5 IntelliGlu: Clinical Decision Support Platform 84

5.1 Introduction 84
5.2 System Architecture 84
5.3 Backend Architecture and CQL Inference Pipeline 85
5.3.1 API Endpoints 85
5.3.2 CQL Inference Pipeline 85
• State vector construction 86
• Model loading and fallback 86
• Safety rules 87
5.4 Frontend Modules 87
5.5 Physiological Simulator 89
5.5.1 Motivation 89
5.5.2 Glucose Dynamics Model 89
5.5.3 CQL Decision at Each Step 91
5.5.4 Critical Patient Presets 91
5.5.5 Correction of the CQL Heuristic for High Glucose Values 92
5.5.6 Stabilisation Detection and Handoff to the Recommendation Module 93
5.5.7 Simulation Example: DKA Preset 93
5.6 End-to-End Clinical Workflow 94
5.7 Discussion 97
5.7.1 Relationship to the Offline Evaluation Framework 97
5.7.2 Limitations 97
5.7.3 Contribution to the Thesis Objectives 98
5.8 Conclusion 98

6 General Conclusion and Future Work 99