

الجمهورية الشعبية الديمقراطية الجزائرية  
République Algérienne Démocratique et Populaire  
وزارة التعليم العالي و البحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
المدرسة العليا للإعلام الآلي 08 ماي 1945. بسيدي بلعباس  
École Supérieure en Informatique  
-08 Mai 1945- Sidi Bel Abbès



## THESIS

To obtain the diploma of **Engineer**  
Field: **Computer Science**  
Specialty: **Artificial Intelligence and data Science**

### Theme

---

**Data-driven prediction of blood transfusion  
performance**

---

Presented by:  
**BOUCHOUKA Zohra Hidaya**

Submission Date: **Sept, 2024**  
In front of the jury composed of:

<b>Mr. KHALDI Belkacem</b>	President
<b>Ms. DIF Nassima</b>	Examiner
<b>Ms. Julia FLECK</b>	Supervisor
<b>Mr. Sidi Mohamed BENSLIMANE</b>	Co-Supervisor
<b>Mr. Xie XIAOLAN</b>	Co-Supervisor
<b>Mr. Fabrice COGNASSE</b>	Co-Supervisor
<b>Ms. Hind HAMZEH</b>	Co-Supervisor

*Academic Year : 2023/2024*

# Abstract

In the era of data-driven healthcare, blood transfusion management has become crucial, with machine learning offering significant enhancements to decision-making and patient care. This project utilizes a dataset from CHU Saint-Étienne, comprising over 50 patient records, which presents challenges such as limited sample size, missing values, and longitudinal complexity. To address these issues, we will evaluate various data imputation methods and benchmark synthetic data generation techniques to improve transfusion outcome predictions. This study aims to identify key predictive features that support more efficient and effective clinical decision-making, ultimately contributing to improved patient safety and resource utilization in blood transfusion management.

**Keywords**— Healthcare, Clinical Decision-Making, Tabular Data, Small Datasets, Missing Data, Data Processing, Supervised Learning, Blood Transfusion Prediction

## ملخص

في عصر الرعاية الصحية المبنية على البيانات، أصبحت إدارة عمليات نقل الدم أمرًا بالغ الأهمية، حيث يقدم التعلم الآلي تحسينات كبيرة في اتخاذ القرار ورعاية المرضى. يستخدم هذا المشروع مجموعة بيانات من المستشفى الجامعي لسانت ايتيان، تتضمن أكثر من 50 سجلًا للمرضى، والتي تواجه تحديات مثل صغر حجم العينة، القيم المفقودة، والتعقيد الطولي. لمعالجة هذه القضايا، سنقوم بتقييم طرق استكمال البيانات المختلفة ومقارنة تقنيات توليد البيانات الاصطناعية لتحسين توقعات نتائج نقل الدم. تهدف هذه الدراسة إلى تحديد الميزات التنبؤية الرئيسية التي تدعم اتخاذ القرارات السريرية بشكل أكثر كفاءة وفعالية، مما يساهم في تحسين سلامة المرضى واستخدام الموارد بشكل أفضل في إدارة عمليات نقل الدم.

**الكلمات المفتاحية:** الرعاية الصحية، اتخاذ القرار السريري، البيانات الجدولية، مجموعات البيانات الصغيرة، القيم المفقودة، معالجة البيانات، تعلم الآلة، التعلم الموجه، توقعات نقل الدم

# Résumé

À l'ère des soins de santé basés sur les données, la gestion des transfusions sanguines est devenue cruciale, avec l'apprentissage automatique apportant des améliorations significatives à la prise de décision et aux soins des patients. Ce projet utilise un ensemble de données du CHU Saint-Étienne, comprenant plus de 50 dossiers de patients, qui présente des défis tels qu'une taille d'échantillon limitée, des valeurs manquantes et une complexité longitudinale. Pour relever ces défis, nous évaluerons diverses méthodes d'imputation de données et comparerons des techniques de génération de données synthétiques afin d'améliorer la prise de décision. Cette étude vise à identifier les caractéristiques prédictives clés qui soutiennent une prise de décision clinique plus efficace et efficiente, contribuant ainsi à une meilleure sécurité des patients et à une utilisation optimisée des ressources dans la gestion des transfusions sanguines.

**Keywords**— Soins de santé, Prise de décision clinique, Données tabulaires, Petits ensembles de données, Données manquantes, Traitement des données, Apprentissage supervisé, Prédiction des transfusions sanguines