بسم الله الرحمن الرحيم

MONA MAGHRABY

شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم

MONA MAGHRABY

شبكة المعلومات الجامعية

التوثيق الإلكتروني والميكروفيلم

جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأقراص المدمجة قد أعدت دون أية تغيرات

يجب أن

تحفظ هذه الأقراص المدمجة بعيدا عن الغبار

MONA MAGHRABY

قسم علوم الحاسب
كلية الحاسبات و المعلومات
جامعة عين شمس

تطوير نظام ذكي لمتابعه الحالات الصحية باستخدام تقنيات التعلم الآلي

رسالة مقدمة إلى قسم علوم الحاسب بكلية الحاسبات والمعلومات جامعة عين شمس
كجزء من متطلبات الحصول على درجة الماجستير في علوم الحاسب

اعداد:

زينه عمرو عصام الدين على ريان

معيدة، قسم علوم الحاسب،
كلية الحاسبات و المعلومات،
جامعة عين شمس

تحت اشراف:

أ.د. عبد البديع محمد سالم
استاذ علوم الحاسب المتفرغ، قسم علوم الحاسب،
كلية الحاسبات و المعلومات،
جامعة عين شمس

د. ماركو الفونس توفيق
مدرس، قسم علوم الحاسب
كلية الحاسبات و المعلومات،
جامعة عين شمس

ديسمبر 2020

ملخص الرسالة

يعيش اليوم أكثر من نصف سكان العالم في المدن مع أكثر من ستة أجهزة لكل شخص متصل بالإنترنت. وهذا يعني أن مليارات الأجهزة والنظم مدمجة في البنية التحتية للمدينة، والتي تسمى اليوم مدينة ذكية. يواجه نظام الرعاية الصحية الحالي تحديات كبيرة في توفير الخدمات الصحية ذات الجودة العالية والمنخفضة التكلفة. وتتفاقم هذه التحديات أيضا بسبب تزايد عدد السكان المسنين، مما يترجم إلى ارتفاع الطلب على خدمات الرعاية الصحية. وعلاوة على ذلك، من الصعب في بعض المدن الحصول على خدمات الرعاية الصحية المناسبة بسبب محدودية الموارد. ونتيجة لذلك، يحتاج نظام الرعاية الصحية الحالي إلى أن يتطور إلى نظام رعاية صحية ذكية. الرعاية الصحية الذكية هي مفهوم يشمل مختلف الكيانات والتكنولوجيات، بما في ذلك: أجهزة الاستشعار، وأجهزة يمكن ارتداؤها، وتكنولوجيا المعلومات والاتصالات وأكثر من ذلك بكثير.

أحد هذه الاجهزة هي وحدة العناية المركزة والتى تشمل مراقبة المرضى والدعم التنفسي والقلب، وأجهزة الإنعاش في حالات الطوارئ، وغيرها من معدات دعم الحياة المصممة لرعاية المرضى الذين يعانون من إصابات خطيرة، او لديهم مرض حرج أو مهدد للحياة، أو خضعوا لعملية جراحية كبرى، وبالتالي هذا يتطلب رعاية ومتابعة على مدار 24 ساعة. يُعد تعفن الدم هو سبب رئيسي للوفيات في وحدات العناية المركزة. لا يزال تعفن الدم مشكلة صحية كبيرة لدى مرضى وحدة العناية المركزة في جميع أنحاء العالم، ويرتبط بارتفاع معدلات الوفيات. مع ذلك ، هناك تباين واسع في معدل تعفن الدم ونتائجه لدى مرضى وحدة العناية المركزة حول العالم. يُعد التنبؤ بمرض تعفن الدم من أصعب التحديات ولا تزال كذلك على الرغم من سنوات عديدة من البحث لأن مظاهره غالبًا ما تكون غير واضحة حتى مراحل متأخرة.

التعلم الآلي هو وسيلة لتحليل البيانات لبناء نموذج تحليلي، وهو فرع من الذكاء الاصطناعي، المعني بتصميم وتطوير الخوارزميات التي تسمح للحواسيب بتطوير السلوكيات على أساس البيانات التجريبية. التعلم الالى يمكن الآلات من فهم و تحليل البيانات من مختلف المجالات و التى تمتد من الحكومة إلى التعليم إلى الرعاية الصحية، وهو يلعب دورا أساسيا في الصحة الذكية من خلال التعامل مع البيانات الطبية لاتخاذ القرارات. يمكن أن تساعد طرق التعلم الآلي في تحسين التنبؤ بتعفن الدم (Sepsis).

تتناول هذه الرسالة دراسة موجزة عن استخدام تقنيات التعلم الآلي في كلا من الصحة الذكية و وحدة العناية المركزة و الامراض المرتبطة بها. كما تقدم ايضا الخلفية الطبية الخاصة بمرض تعفن الدم و مقارنة بين استخدام تقنيات التعلم الآلي في التشخيص المبكر للمرض. كما تحتوي على نتائج تطبيق خوارزميات آلية دعم المتجه (SVM) على مجموعة بيانات من متعدد المتغيرات للمراقبة الذكية فى العناية المركزة (MIMIC) للتنبؤ بمرض تعفن الدم.

تتناول هذه الرسالة ايضا نظام صحى ذكى للتشخيص المبكر لمرض تعفن الدم. يعتمد هذا النظام على عدة مراحل بدءا من توحيد نطاق البيانات، مرورا بمعالجة البيانات غير المتوازنة و القيم المفقودة، ثم اختيار القياسات السريرية المناسبة لاستخدامها بواسطة خوارزميات التعلم الآلي. لقد أظهرت المنهجية المقترحة في هذه الرسالة للتشخيص المبكر لمرض تعفن الدم في الوقت الفعلي لمرضى وحدة العناية المركزة، أداءا متتميزا و ذلك بنسبة 98 ٪ كدقة. بالنظر إلى حقيقة أن جميع المتغيرات جاءت من البيانات السريرية التي يتم جمعها بانتظام في وحدات العناية المركزة ، يمكن بسهولة تطبيق

المنهجية المقترحة فى مراقبة مرضى وحدة العناية المركزة لتحسين اتخاذ القرارات السريرية.

تتكون الرسالة من سبع فصول على النحو التالي:

الفصل الاول: يتناول المقدمة و الغرض من البحث العلمى.

الفصل الثانى: يقدم دراسة موجزة عن الصحة الذكية و علاقتها بانواع اخرى للصحة و احدث تقنيات التعلم الآلي المستخدمة فيها.

الفصل الثالث: يحتوى على الخلفية الطبية الخاصة بوحدة العناية المركزة و الامراض المرتبطة بها، كما يتضمن تحليلات بيانات وحدة العناية المركزة باستخدام تقنيات التعلم الالى.

الفصل الرابع: يحتوى على الخلفية الطبية الخاصة بمرض تعفن الدم و احدث تقنيات التعلم الآلي المستخدمة في التشخيص المبكر للمرض.

الفصل الخامس: يحتوي على نتائج تطبيق خوارزميات SVM على مجموعة بيانات من MIMIC للتنبؤ بمرض تعفن الدم.

الفصل السادس: يقدم نتائج تطبيق تقنيات التعلم الآلي المختلفة من أجل التنبؤ بمرض تعفن الدم.

الفصل السابع: يحتوي على الملخص والاستنتاج والعمل المستقبلي.

وقد تم نشر ستة ابحاث من الرسالة فى المجلات والمؤتمرات العالمية المتخصصة فى هذا المجال على النحو التالي:

1. Rayan, Z., Alfonse, M. and Salem, A.B.M., 2019. “Machine learning approaches in smart health”. Procedia Computer Science, ElSevier, vol. 154, pp. 361-368, ISSN: 1877-0509, E-ISSN:1877-0509 (Scopus - CiteScore: 2.5, SJR: 0.342).

2. Rayan, Z., Alfonse, M. and Salem, A.B.M., 2019. “Intensive Care Unit (ICU) Data Analytics using Machine Learning Techniques”. International Journal of “Information Theories and Applications”, vol. 26, no. 1, pp. 69-82, ISSN: 1310-0513, E-ISSN: 1313-0463.

3. Rayan, Z., Alfonse, M. and Salem, A.B.M., 2020. “Sepsis Prediction Model in the Intensive Care Unit (ICU) using Support Vector Machine (SVM)”. Accepted for publication in Proceedings of Internet of Things: Applications & Future (ITAF 2020), Lecture Notes in Networks and Systems, Springer Nature, ISSN:2367-3370, E-ISSN:2367-3389 (Scopus - CiteScore: 0.4, SJR: 0.125 Q4).

4. Rayan, Z., Alfonse, M. and Salem, A.B.M., 2020. “Comparative Study of Machine Learning Models for Onset Sepsis Prediction” Accepted for publication in Proceedings of Internet of Things: Applications & Future (ITAF 2020), Lecture Notes in Networks and Systems, Springer Nature, ISSN:2367-3370, E-ISSN:2367-3389 (Scopus - CiteScore: 0.4, SJR: 0.125 Q4).

5. Rayan, Z., Alfonse, M. and Salem, A.B.M., 2020. “Predicting Sepsis in the Intensive Care Unit (ICU) through Vital Signs using Support Vector Machine (SVM)”. Submitted to The Open Bioinformatics Journal, ISSN:1875-0362 (Scopus - CiteScore: 0.8 , SJR: 0.323 Q2).

6. Rayan, Z., Alfonse, M. and Salem, A.B.M., 2020. “An Ensemble Model for Early Sepsis Prediction using Clinical Records from the Intensive Care Unit (ICU)”. Accepted for publication in the Egyptian Computer Science Journal, vol. 45, no. 1, January issue 2021, ISSN: 1110-2586.

Ain Shams University
Faculty of Computer & Information Sciences
Computer Science Department

Developing a Smart Health Monitoring System Using Machine Learning Techniques

Thesis submitted as a partial fulfillment of the requirements for the degree of master of science in computer and information sciences.

By

Zeina Amr EssamEldein Ali Rayan
Teacher Assistant, Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

Under Supervision of

Prof. Dr. Abdel-Badeeh Mohamed Salem
Professor of Computer Science, Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

Dr. Marco Alfonse Tawfik
Lecturer, Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University.

December 2020

Acknowledgements

All praise and thanks to ALLAH, who provided me the ability to complete this work. I hope to accept this work from me.

I am grateful to *my parents* and *my family* who are always providing love, help and support throughout the whole years of this study. I hope I can give that back to them.

I also offer my sincerest gratitude to my supervisors, *Prof. Dr. Abdel-Badeeh Mohamed Salem*, *Dr. Marco Alfonse* who have supported me throughout my thesis with their patience, guidance, knowledge, experience and constant support.

Finally, I would like to thank my friends and all people who gave me support and encouragement.

Abstract

Today more than half of the world's population lives in cities with more than six devices per person connected to the Internet. This implies that billions of devices and systems are embedded in a city's infrastructure, which is called today a smart city. The current healthcare system faces significant challenges in providing quality and low-cost healthcare services. These challenges are also aggravated by the increasing elderly population, which translates into a higher demand for healthcare services. Moreover, it is difficult in some cities to obtain a proper healthcare service due to limited resources. Due to this, the current healthcare system needs to evolve into a smart healthcare system. Smart healthcare is a concept involving various entities and technologies, including: sensors, wearable devices, Information Communication and Technology (ICT) and much more.

One of the ICT devices is the Intensive Care Unit (ICU) equipment. It includes patient monitoring, respiratory and cardiac support, pain management, emergency resuscitation devices, and other life support equipment. This equipment is designed to care for patients who are seriously injured, have a critical or life-threatening illness, or have undergone a major surgical procedure, thereby requiring 24-hour care and monitoring. Sepsis is a major cause of mortality and morbidity in ICUs. Sepsis remains a major health problem in ICU patients worldwide and is associated with high mortality rates. However, there is vast variability in the sepsis rates and outcomes for ICU patients around the world. Sepsis prediction is a challenging problem and remains so despite many years of research because its appearance is often unclear until later stages.

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence, concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data. It can help machines, millions of machines, get together to understand what people want from the data made by human beings. Machine learning touches industries spanning from government to education to health care. It plays an essential role in smart health by handling the medical data for decision-making. The machine learning techniques can help improve the sepsis prediction.

This thesis presents an important study of smart health and intensive care unit and the recent machine learning approaches used in each of them. It also includes the medical aspects of sepsis and a comparative study of recent machine learning approaches used for sepsis prediction. The results of applying the Support Vector Machine (SVM) algorithm to the MIMIC III dataset for predicting sepsis is 99% accuracy.

The thesis also presents a smart health monitoring system based on machine learning techniques for early prediction of sepsis. This system consists several components: patient monitoring equipment for data acquisition, sepsis prediction mode, and the prediction output. Sepsis prediction model consists of the following steps: The first step is data normalization; the second step is handling dataset imbalance; the third step is handling missing values; the fourth step is selection of clinical measurements (feature selection). The system proposed in this thesis shows a high prediction performance with accuracy of 98%. Considering the fact that all features came from clinical data that are regularly collected in ICUs, the proposed methodology can easily be applied to ICU patients monitoring to help in better clinical decision-making.

List of Publications

1. Rayan, Z., Alfonse, M. and Salem, A.B.M., 2019. "Machine learning approaches in smart health". Procedia Computer Science, ElSevier, vol. 154, pp. 361-368, ISSN: 1877-0509, E-ISSN:1877-0509 (Scopus - CiteScore: 2.5, SJR: 0.342).

2. Rayan, Z., Alfonse, M. and Salem, A.B.M., 2019. "Intensive Care Unit (ICU) Data Analytics using Machine Learning Techniques". International Journal of "Information Theories and Applications", vol. 26, no. 1, pp. 69-82, ISSN: 1310-0513, E-ISSN: 1313-0463.

3. Rayan, Z, Alfonse, M. and Salem, A.B.M., 2020. "Sepsis Prediction Model in the Intensive Care Unit (ICU) using Support Vector Machine (SVM)". Accepted for publication in Proceedings of Internet of Things: Applications & Future (ITAF 2020), Lecture Notes in Networks and Systems, Springer Nature, ISSN:2367-3370, E-ISSN:2367-3389 (Scopus - CiteScore: 0.4, SJR: 0.125 Q4).

4. Rayan, Z., Alfonse, M. and Salem, A.B.M., 2020. "Comparative Study of Machine Learning Models for Onset Sepsis Prediction" Accepted for publication in Proceedings of Internet of Things: Applications & Future (ITAF 2020), Lecture Notes in Networks and Systems, Springer Nature, ISSN:2367-3370, E-ISSN:2367-3389 (Scopus - CiteScore: 0.4, SJR: 0.125 Q4).

5. Rayan, Z., Alfonse, M. and Salem, A.B.M., 2020. "Predicting Sepsis in the Intensive Care Unit (ICU) through Vital Signs using Support Vector Machine (SVM)". Submitted to The Open Bioinformatics Journal, ISSN:1875-0362 (Scopus - CiteScore: 0.8, SJR: 0.323 Q2).

6. Rayan, Z., Alfonse, M. and Salem, A.B.M., 2020. "An Ensemble Model for Early Sepsis Prediction using Clinical Records from the Intensive Care Unit (ICU)". Accepted for publication in the Egyptian Computer Science Journal, vol. 45, no. 1, January issue 2021, ISSN: 1110-2586.

Table of Contents

Acknowledgment 1
Abstract 2
List of Publications 4
List of Figures 8
List of Tables 10
List of Algorithms 11
List of Abbreviations 12
Chapter 1: Introduction 13
1.1-Thesis Motivation 14
1.2-Thesis Objectives 17
1.3-Thesis Importance 17
1.4-Thesis Methodology 17
1.5-Thesis Contribution 17
1.6-Thesis Organization 18
Chapter 2: Smart Healthcare paradigms 19
2.1- Healthcare 20
2.2- Smart Health 23
2.3- Analysis of recent Machine Learning (ML) Techniques used in Smart Health 27
2.4- A comparative study of ML in smart health 29
Chapter 3: Intensive Care Unit data analytics using ML Techniques ... 34
3.1- Intensive Care Unit (ICU) 35
3.2- Analysis of recent Machine Learning (ML) Techniques used in ICU data analytics 38
3.3- A comparative study of ML in ICU data analytics 40

Chapter 4: Medical Aspects of Sepsis and Analysis of its Prediction Models 47

4.1- Sepsis Aspects 48

4.1-1. Sepsis Statistics 50

4.1-2. Symptoms 53

4.1-3. Causes 55

4.1-4. Risk Factors 55

4.1-5. Complications 56

4.2- Analysis of recent Machine Learning (ML) techniques used in Sepsis Prediction 56

4.3- A comparative study of ML in Sepsis prediction 64

Chapter 5: Applying Support Vector Machine to the MIMIC dataset . 70

5.1- Dataset 71

5.2- Proposed Model for Sepsis Prediction 71

5.3- Support Vector Machine 74

5.4- Results and Discussion 78

Chapter 6: Developing a Smart Real-Time Monitoring System for Sepsis prediction 82

6.1- Real-Time Sepsis Prediction System 83

6.2- Dataset 84

6.3- Methodology 85

6.3-1. Data Normalization 86

6.3-2. Handling Data Unbalancing 86

6.3-3. Handling Missing Values 88

6.3-4. Feature Selection 89

6.3-5. Applying Machine Learning Approach 90

6.4- The used Machine Learning Algorithms 91

6.4-1. Naïve Bayes 91

6.4-2. Logistic Regression 92

6.4-3. Linear Discriminant Analysis 94