A FUSION-BASED FRAMEWORK FOR EXPLAINABLE SUICIDE ATTEMPT PREDICTION

NORATIKAH BINTI NORDIN

UNIVERSITI SAINS MALAYSIA

2024

A FUSION-BASED FRAMEWORK FOR EXPLAINABLE SUICIDE ATTEMPT PREDICTION

by

NORATIKAH BINTI NORDIN

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

May 2024

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to Allah SWT for giving me the opportunity and helping me endlessly in completing this thesis. I am very grateful to my supervisors, Assoc. Prof. Dr. Zurinahni Zainol and Dr. Mohd Halim Mohd Noor, School of Computer Sciences, Universiti Sains Malaysia, for their unconditional support, positive encouragement, constructive ideas, invaluable advice, and unequivocal time, that they gave me throughout my studies to complete this thesis. I would also like to express my gratitude to my co-supervisor, Assoc. Prof. Dr. Chan Lai Fong, Faculty of Medicine, Department of Psychiatry, Universiti Kebangsaan Malaysia, for providing useful clinical datasets, assistance, and support throughout my research journey. Under their supervision and guidance, I have been able to learn and grow very much over the progressive years. And I am most grateful for their motivation, professional and immense knowledge, and endless comments and editing of my English writing. I would also like to acknowledge Dr. Ryna Imma Buji for her support and help in the real practices of treating individuals with suicidal behaviour. In addition, I would like to thank my fellow colleagues and friends in the School of Computer Sciences for their candid opinions, inputs, and suggestions. Last but not least, and most importantly, I thank my mother Fatimah Hassan, my father Nordin Ibrahim, my sister Noradawiyah Nordin and my brother Muhammad Anas Nordin for their continuous love and support throughout this journey. Without their understanding, it would have been difficult to embark on this journey. Lastly, I would also like to thank the Ministry of Higher Education Malaysia for supporting this work and my PhD journey under the Fundamental Research Grant Scheme (FRGS) 2020/2022.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

BD	Breakdown Plot
CP	Ceteris-paribus Profiles
DL	Description Logics
DT	Decision Trees
EPM	Explainable Predictive Model
EPM-OM	Fusion Model
GB	Gradient Boosting
MDM	Medical Decision-Making
OM	Ontology Model
OWL	Web Ontology Language
SDG	Sustainable Development Goals
SHAP	Shapley Additive Explanations
SVM	Support Vector Machine
XAI	Explainable Artificial Intelligence

LIST OF APPENDICES

KERANGKA KERJA BERASASKAN PENGGABUNGAN BAGI RAMALAN PERCUBAAN BUNUH DIRI YANG DAPAT DIJELASKAN

ABSTRAK

Bunuh diri kekal sebagai masalah kesihatan awam utama dan salah satu punca utama kematian di seluruh dunia. Pencegahan bunuh diri diperlukan untuk mengurangkan kematian bunuh diri secara global, seperti yang diketengahkan dalam Matlamat Pembangunan Lestari Ketiga (SDG) Pertubuhan Bangsa-Bangsa Bersatu. Percubaan bunuh diri adalah tingkah laku yang paling kompleks dan dinamik, yang mana penting untuk strategi pencegahan bunuh diri. Walau bagaimanapun, membuat keputusan dalam mengklasifikasikan individu yang berisiko tinggi untuk percubaan bunuh diri adalah subjektif dan tidak pasti. Kajian sedia ada mengenai rangka kerja ramalan percubaan bunuh diri menggunakan kecerdasan buatan tidak dijelaskan dengan secukupnya dan tidak dapat memberikan ramalan yang boleh difahami tentang percubaan bunuh diri secara sistematik. Oleh itu, kajian ini membentangkan rangka kerja berasaskan gabungan untuk ramalan percubaan bunuh diri yang boleh dijelaskan menggunakan pendekatan pembelajaran dan pendekatan pengetahuan yang dapat membantu membuat keputusan oleh pakar perubatan. Kerja yang dicadangkan bertujuan untuk menganalisa algorithma pembelajaran yang boleh dijelaskan untuk meramalkan percubaan bunuh diri, model ontologi dicadangkan untuk membina konsep risiko klasifikasi secara semantik dan kaedah penjanaan teks berasaskan gabungan maklumat dicadangkan dengan menggabungkan ramalan untuk menghasilkan huraian ramalan untuk menyokong membuat keputusan. Model gabungan menunjukkan bahawa rangka kerja yang dicadangkan mencapai ketepatan 92%, kekhususan 88%, dan kepekaan 100%. Penjelasan dalam klasifikasi risiko

percubaan bunuh diri dengan tahap faktor risiko dihasilkan dengan prestasi keputusan untuk menyokong membuat keputusan. Rangka kerja berasaskan gabungan untuk penjelasan ramalan percubaan bunuh diri telah dinilai oleh pakar-pakar perubatan dan menunjukkan bahawa rangka kerja berasaskan gabungan berjaya mengklasifikasikan dan menerangkan ramalan percubaan bunuh diri untuk menyokong membuat keputusan klinikal.

A FUSION-BASED FRAMEWORK FOR EXPLAINABLE SUICIDE ATTEMPT PREDICTION

ABSTRACT

Suicide remains a major public health problem and one of the leading causes of death worldwide. Suicide prevention is needed to reduce global suicide mortality, as highlighted in the United Nations Third Sustainable Development Goals (SDGs). A suicide attempt is the most complex and dynamic suicidal behaviour, which is important for suicide prevention strategies. However, decision-making in classifying individuals at higher risk of suicide attempts is subjective and uncertain. Existing studies on the framework for predictive models using data-driven and knowledge-driven approaches are insufficiently explained and unable to provide an understandable prediction of suicide attempts for suicide prevention in a systematic way. Therefore, this study presents a fusion-based framework for explainable suicide attempt prediction using explainable data-driven and knowledge-driven approaches to classify and explain individuals with suicide attempts to support decision-making by medical experts. The proposed work aims to analyse an explainable learning algorithms for predicting suicide attempts, propose an ontology model for semantically representing the classification risk of suicide attempts and propose an explanation generation algorithm by combining predictions from explainable machine learning and ontology models. An information fusion-based explanation generation method is proposed by integrating predictions to generate a prediction description to support decision-making. The fusion model shows that the proposed framework achieves 92% accuracy, 88% specificity, and 100% sensitivity. The explanation in the classification risk of suicide attempts with the level and contribution of risk factors is produced with

the results performance to support decision-making. The fusion-based framework for explanation suicide attempts prediction was evaluated by medical experts, and shows that the fusion-based framework successfully classifies and explains the suicide attempts prediction to support clinical decision-making.

CHAPTER 1

INTRODUCTION

1.1 Background

Physicians and clinicians in the medical field must make calculated decisions every day that have a significant impact on patients at the individual, community, and national levels. Medical decisions are vital and highly important to physicians and patients as they provide results and outcomes for future interventions and treatments (Masic, 2022). Therefore, medical decision-making is defined as the process of using clinical knowledge and expertise to diagnose and manage a patient's health condition based on patient preferences, available clinical information, and available resources (Kattan & Cowen, 2009).

Psychiatry is one of the medical specialties that require demanding and complex decision-making processes. Psychiatry is a medical field that deals with the diagnosis, treatment, and prevention of mental health conditions, including anxiety disorders, personality disorders, schizophrenia, depressive disorder, bipolar disorder, sleep disorders, and suicidal behaviour. Psychiatrists and clinicians are constantly involved in decision-making for patients with these disorders. However, these decisions were mostly influenced by their clinical experience and uncontrollable factors such as time pressure, cost, and availability of resources (Bhugra et al., 2011).

Currently, the decision-making process in psychiatry, particularly in the area of suicidal behaviour, involves manual decision-making based on clinical practise guidelines. Indeed, psychiatrists often have to make decisions under uncertain conditions, relying heavily on patients' subjective self-assessment, which leads to errors of judgement and bias (Large et al., 2018). Therefore, it is important to minimise

the errors in practise to reduce the cases of suicide attempts, and providing a meaningful framework for decision-making in psychiatry is a good prevention strategy.

1.1.1 Suicide Attempt Prediction

According to the World Health Organization (WHO), there were more than 700,000 deaths by suicide worldwide in 2019, with an estimated suicide rate of 9.0 per 100,000 per year (World Health Organization, 2021). Reducing the global suicide mortality rate by one-third by 2030 is a goal of the United Nations Sustainable Development Goals (SDGs) (World Health Organization, 2013). Although suicide is not a disease, suicidal behaviour, which includes suicide attempts, suicidal ideation, and suicide planning, is an important health issue that is receiving increasing attention in research and public awareness campaigns around the world (Belsher et al., 2019).

The general term suicidal behaviour is defined as thoughts and behaviours related to a person intentionally taking his or her own life (De Leo et al., 2021). Suicidal behaviours are most often related to suicide, suicide attempt, suicidal ideation, suicide plan, and self-harm. A suicide attempt is described as a non-fatal and self-inflicted destructive act with an explicit or inferred intent to die (Chan et al., 2011). The simple definition of the term suicide attempt therefore refers to someone who harms themselves with the intention of ending their life but they do not die because of their actions. A suicide attempt is a complex phenomenon that is dependent contextually and changes rapidly from one day to the next. Thus, suicide attempt prediction is defined as the classification of individuals with suicide attempts into suicide attempters and non-suicide attempters. However, a major challenge in clinical

psychiatry is accurately predicting who is at risk for future suicide attempts (Belsher et al., 2019).

Identifying and classifying individuals at risk of attempting suicide in order to take preventive measures can potentially reduce suicide rates. Chan et al. (2011) highlighted that accurately identifying individuals at the highest risk for suicide attempts is one of the most important processes in suicide prevention. Strategies to accurately predict which individuals will attempt suicide or die from it remain inadequate, although existing literature identifies common risk factors such as family history of suicide attempts, child abuse, substance abuse, alcohol dependence, and severity of mental illness (Belsher et al., 2019; Franklin et al., 2017).

Strategies to predict suicide attempts also require accurate and efficient decision-making. However, decision-making in predicting suicide attempts is highly subjective and uncertain, due to judgment depends on clinicians' knowledge, skills, and experience (Bernert et al., 2020; Bhugra et al., 2011). In fact, individuals with suicide attempts have different risk factors and vary from person to person, making decision-making difficult and challenging. Therefore, identifying and understanding an individual with suicide attempts are two important elements in decision-making for predicting suicide attempts, which require computational intelligence with good modeling and advanced measures (Boudreaux et al., 2021).

In addition, understanding the prediction of suicide attempts requires an explanation for decision-making. An explanation is defined as information that makes a decision clear and easy for clinicians to understand (Kessler et al., 2020). In the medical field, explanations are crucial and critical for effective decision-making, which consists of understanding the context of the prediction. In the prediction of suicide attempts, explanations provide valuable feedback that can help to learn from

the patterns of risk factors in individuals with suicide attempts and make better decisions in the future (Burke et al., 2019). In fact, the explanation increases the level of confidence of the clinician and decreases the risk of making wrong decisions.

There are three main methods for predicting suicide attempts, namely suicide risk assessment tools, data-driven approaches, and knowledge-driven approaches (Velupillai et al., 2019). Recently, a framework to support decision-making in predicting suicide attempts using data-driven approaches and knowledge-driven approaches has been developed. Data-driven approaches have been introduced to support decision-making and reduce the problems in conventional medical decision-making using machine learning (Montani & Striani, 2019). Machine learning is an artificial intelligence technique that allows computers to learn from a large amount of data without having to be explicitly programmed to make effective predictions (Dwyer et al., 2018).

Machine learning such as decision trees, logistic regression, random forest, support vector machine, and gradient boosting (Hettige et al., 2017; Navarro et al., 2021; Walsh et al., 2017) has been used to develop a predictive model for suicide attempts. However, predictive models using machine learning lack explanatory power, because the inner workings of the approaches behave like a 'black box', such as an ensemble learning model, which makes it difficult to explain the predictive models created (Jung et al., 2019; Abdullah et al., 2021). The studies by Oh et al. (2017) and Zheng et al. (2020) showed good classification results, but a lack of explanation for the prediction of suicide attempts. Therefore, explainable learning algorithms have been introduced to overcome the lack of explanation in predictive modeling.

Explainable learning approaches also known as explainable machine learning (XAI) models provide a clear explanation for their decision that can contribute to the

understanding of the predictive model (Abdullah et al., 2021; Belle & Papantonis, 2021). Two explainable approaches are able to support the explainability of predictive models, namely are model-specific and model-agnostic. Model-specific methods are only applicable to the specific model for which they were developed and use knowledge of the architecture and inner workings of the specific model to explain its predictions, whereas model-agnostic methods do not rely on the internal structure or specific features of the underlying model and often work by approximating the behaviour of the model as a simpler and more interpretable model that is easier to understand (Sahakyan et al., 2021). Although explainable learning approaches have recently been introduced in medical applications, the potential use of these approaches in predictive models of suicide attempts is limited (Abdullah et al., 2021).

In addition, the importance of risk factors that influence the predictive model to support decision-making is necessary in the framework of suicide attempt prediction. Specific risk factors can signal vulnerability, and offer the opportunity to intervene early and provide support before suicide occurs (Belsher et al., 2019). By knowing and understanding the specific risk factors for suicide attempts, prevention measures, and resources can be tailored to the needs of individuals at high risk of suicide attempts. Therefore, knowledge-driven approaches have been introduced to present and capture clinical knowledge and medical literature on suicide attempts (Larsen & Hastings, 2018; Yamada et al., 2020).

Knowledge-driven approaches for the prediction of suicide attempts have been proposed to support clinical decision-making through rule-based, case-based, and ontology-based modeling (Chang et al., 2013; Yamada et al., 2020). The complex correlation between risk factors and the classification of individual with at risk of suicide attempts makes the representation of knowledge and the inference process

challenging. Therefore, the collection and creation of a concept of medical knowledge using existing medical literature and clinical practice guidelines for suicide attempts is able to support additional knowledge in predicting suicide attempts. Concepts and relationships with knowledge-driven approaches for predicting suicide attempts to clearly understand risk factors are presented in a meaningful way for clinicians to make a decision accurately. Therefore, an explanation of clinical outcomes is important in predicting suicide attempts.

1.2 Motivation

The decision-making process is important in predicting suicide attempts, as clinicians and psychiatrists often have to make decisions based on patient information and rely heavily on their clinical knowledge. In recent years, detection and risk assessment tools have been considered essential for suicide prediction strategies (O'Connor & Portzky, 2018). Suicide risk assessment tools are widely used worldwide to predict suicide attempts, but existing suicide risk assessment tools have low to moderate predictive value and an inconsistent validated scale to classify between patients at low and high risk for suicide attempts (Velupillai et al., 2019). This has been supported by interviewing psychiatrists from Hospital Universiti Sains Malaysia (HUSM), and Advanced Medical and Dental Institute (AMDI), Universiti Sains Malaysia, that identifying and classifying an individual with suicide attempts is based on manual risk assessment tools such as the Beck Scale for Suicide Ideation (BSI), Beck Hopelessness Scale (BHS), While Reasons for Living Inventory, Columbia-Suicide Severity Rating Scale (C-SSRS), and the SAD Person Scale (Baek et al., 2021). Also, according to Runeson et al. (2017), none of the risk assessment tools showed sufficient diagnostic ability, which means low specificity and low sensitivity, the description of risk factors is unclear, and limited interpretation of results.

Existing suicide risk assessment tools are usually based on self-report and are time-consuming, which can lead to difficulties in classifying and measuring the patient's low and high-risk condition to make a good decision. Besides that, inaccurate decisions and misdiagnosis of individuals with suicide attempts may occur which may lead to suicide (Abu Bakar et al., 2023). Therefore, there is a need to develop a predictive model for suicide attempts instead of conducting manual suicide assessments based on risk scores, that are highly dependent on clinicians. The predictive model could provide a more effective method for routine clinical assessment, and help clinicians predict suicide attempts to enable accurate decision-making. The development of predictive models for suicide has the potential to improve the identification of individuals at heightened risk of suicide by utilizing predictive algorithms such as artificial intelligence (Belsher et al., 2019).

Artificial intelligence is currently being used in predicting suicide attempts accurately for decision-making, with a focus on data-driven and knowledge-driven approaches (Boudreaux et al., 2021; Dwyer et al., 2018). Artificial intelligence techniques for predicting suicide attempts offer new opportunities to significantly improve risk prediction and broaden the framework for suicide prevention (Bernert et al., 2020; Burke et al., 2019). This is because an individual with a suicide attempt has different risk factors that make accurate decision-making by clinicians difficult manually and time-consuming. Therefore, understanding the predictive model developed accurately and effectively by artificial intelligence, specifically machine learning is important to increase clinical applicability in suicide attempts (Kessler et al., 2020).

Accurate prediction of suicide attempts is important for clinicians and psychiatrists as it can influence the decisions they make. Identifying individuals at

higher risk for suicide attempts could allow for timely intervention and support, and potentially even save lives (Boudreaux et al., 2021). Understanding specific risk factors for suicide attempts could help to create individualised treatment plans and allocate resources more efficiently, which in turn contributes to better decision-making. In addition, improving the suicide attempt prediction model for accurate prediction is important for raising awareness, which contributes to a better public understanding of suicide risk and prevention strategies (Bernert et al., 2020).

1.3 Problem Statement

Machine learning models have been developed to identify individuals at risk of suicide and classify them into suicide attempters and non-suicide attempters (Walsh et al., 2017), and explainable learning methods have been leveraged to assist with prediction and explain why the model suggests a specific feature for a patient to support decision-making by clinicians and medical experts. Recent studies have developed a framework with an explainable predictive model for medical applications using data-driven approaches. The study by Weller et al. (2022) and Tang et al. (2021) used machine learning algorithms for predicting suicide attempts and integrated explainable learning algorithms to explain the prediction. However, the explanation of the predictions may be misinterpreted by clinicians and medical experts due to the lack of familiarity and complexity of the algorithms (Nguyen et al., 2022). In addition, the explanation of prediction in explainable data-driven approaches is not intuitive to medical experts because the explanations provided by explainable learning algorithms are in the form of graphs and numbers (Tang et al., 2023), and these factors may hinder the successful implementation and acceptance of the frameworks in clinical settings. Besides, the explainable machine learning models are not able to capture the complex interactions between various features or risk factors to represent clinical knowledge

and provide an explanation for predicting suicide attempts. This could lead to oversimplification and potentially incorrect predictions that require knowledge-driven approaches to support decision-making (Bahani et al., 2021).

The current frameworks with explainable data-driven and knowledge-driven approaches have provided information related to predictions in the context of the classification results of suicide attempts and non-suicide attempts (Tang et al., 2023). The most important features that influence the prediction of suicide attempts in the framework have also been highlighted in previous studies (Kim et al., 2021; Kirlic et al., 2021). However, the framework lacks intuitive explanations whereby the predictions and explanations generated from both approaches are only available in the form of feature importance and classification classes. Furthermore, the predictions produced by the data-driven and knowledge-driven approaches are not fused in the framework to generate the explanation (Weller et al., 2021; Uddin et al., 2022; Tang et al., 2023). Therefore, the explanation may be inaccurate, unreliable, and difficult to understand and interpret for clinical decision-making.

From the understanding of the background problem discussed, it can be concluded that issues in the area of predictive modeling of suicide attempt studies need further investigation. The ability to classify individuals with suicide attempts and understand the risk factors that influence suicide attempts will focus on addressing the underlying problem of uncertainty and generate explanations for the classification of individuals with suicide attempts and further reduce discrepancies in decision-making by clinicians. In addition to developing new ideas for improvement, the limitations and gaps created by previous work raise some key questions that will be addressed in this study. The following research questions are posed here to be further investigated in this study for the framework of an explainable model for suicide attempt prediction.

1. Which explainable learning algorithms are most effective for predicting suicide attempts?
2. How can an ontology model be designed to semantically represent the classification risk of suicide attempts?
3. How can a text explanation algorithm be generated automatically to combine predictions from explainable machine learning and ontology models?

1.4 Research Objectives

Classifying patients at risk of suicide attempts and prevention measures can potentially reduce suicide rates. However, a major challenge in decision-making for clinical psychiatry is to accurately predict who is at risk of attempting suicide and to understand the risk factors that influence the behaviour of individuals with suicide attempts. An explanation for suicide attempts predictive model is crucial and important to provide clear and understandable insights for decision-making. Therefore, the main objective of this study is to design a fusion-based framework for explainable suicide attempt prediction to support decision-making by integrating information from explainable data-driven and knowledge-driven approaches. The main objective has been carried out through the following sub-research objectives.

1. To analyze explainable learning algorithms for suicide attempt prediction.

2. To propose an ontology model to semantically represent the classification risk of suicide attempts.

3. To propose an explanation generation algorithm by combining predictions from an explainable machine learning model and ontology model.

1.5 Scope and Limitation of Study

As mentioned earlier, the aim of this study was to design a framework for explainable suicide attempt prediction using explainable data-driven and knowledge-driven approaches. The framework focuses on two contexts, namely prediction, and explanation for decision-making, which is able to identify, classify, and explain an individual with suicide attempts by integrating information from data-driven and knowledge-driven approaches. Solving the classification problem in the predictive model of suicide attempts is critical and challenging due to the limited availability of the clinical dataset for legal and confidentiality issues and the highly subjective judgment of the clinician (Burke et al., 2019; Oh et al., 2017). Therefore, this framework for explainable suicide attempt prediction was designed to support decision-making using the Malaysian population with depressive disorders. The collection of patient information was supported by Universiti Kebangsaan Malaysia Medical Centre (UKMMC), Kuala Lumpur (Chan et al., 2011).

Suicidal behaviour refers to thoughts and behaviours related to an individual intentionally taking his or her own life, including suicide attempts, suicidal ideation, suicide plan, and self-harm (De Leo et al., 2021). However, in this study, the focus of suicidal behaviour is on suicide attempt, which refers to someone harming themselves with the intention of ending their life, but they do not die because of their actions. Therefore, the prediction of suicide attempts refers to the classification of individuals with suicide attempts into suicide attempter and non-suicide attempter. This study is important to improve understanding of individual situations and provides an opportunity to identify feasible preventive measures for clinicians and psychiatrists to reduce suicide rates worldwide.

Decision-making in the prediction of suicide attempts requires an explainable component to make a predictive model understandable and trustworthy. Explainable refers to the ability of a model to provide a clear and concise explanation of how the decision was made. Explainable learning approaches used in this study are limited to the model-agnostic method. Since the explanation of outcomes is important in predicting suicide attempts, the model-agnostic method is used to create an explanation that is able to analyse the patterns of risk factors that influence prediction (Sahakyan et al., 2021). Model-agnostic methods are preferred due to the flexibility and applicability of the models as they can be applied to any machine learning model and are not dependent on the internal workings of the model. Model-agnostic explanations are less prone to overfitting and generalize well to different datasets, making them more reliable and trustworthy, and facilitating the understanding of the explanation (Linardatos et al., 2020).

1.6 Outline of the Thesis

This thesis is divided into seven chapters. Figure 1.3 shows the structure of the thesis. The contents of each chapter are briefly described below:

I. Chapter 1 provides a general introduction to the topic of the research work. This chapter also provides a brief overview of some issues concerning the research work. In addition to the background, this chapter also includes the research questions, the research objectives, the scope, and the limitations of the study.

II. Chapter 2 provides a detailed overview of the background and related work on the problem of predicting suicide attempts. An overview of suicide attempt predictive modeling in the context of suicide risk assessment tools, data-driven approaches, and knowledge-driven approaches are discussed.

III. Chapter 3 provides an overview of the methodology used to achieve the research objectives of this study and describes the procedures performed.

IV. Chapter 4 demonstrates the explainable learning algorithms for predicting suicide attempts.

V. Chapter 5 presents an ontology model to semantically represent the classification risk of suicide attempts for individuals at risk of suicide attempts.

VI. Chapter 6 focuses on the information fusion-based explanation generation method by combining predictions from explainable machine learning and ontology models to support decision-making for the fusion-based framework of explainable suicide attempt prediction.

VII. Finally, Chapter 7 summarises and concludes the main contributions of the proposed work and presents future directions.

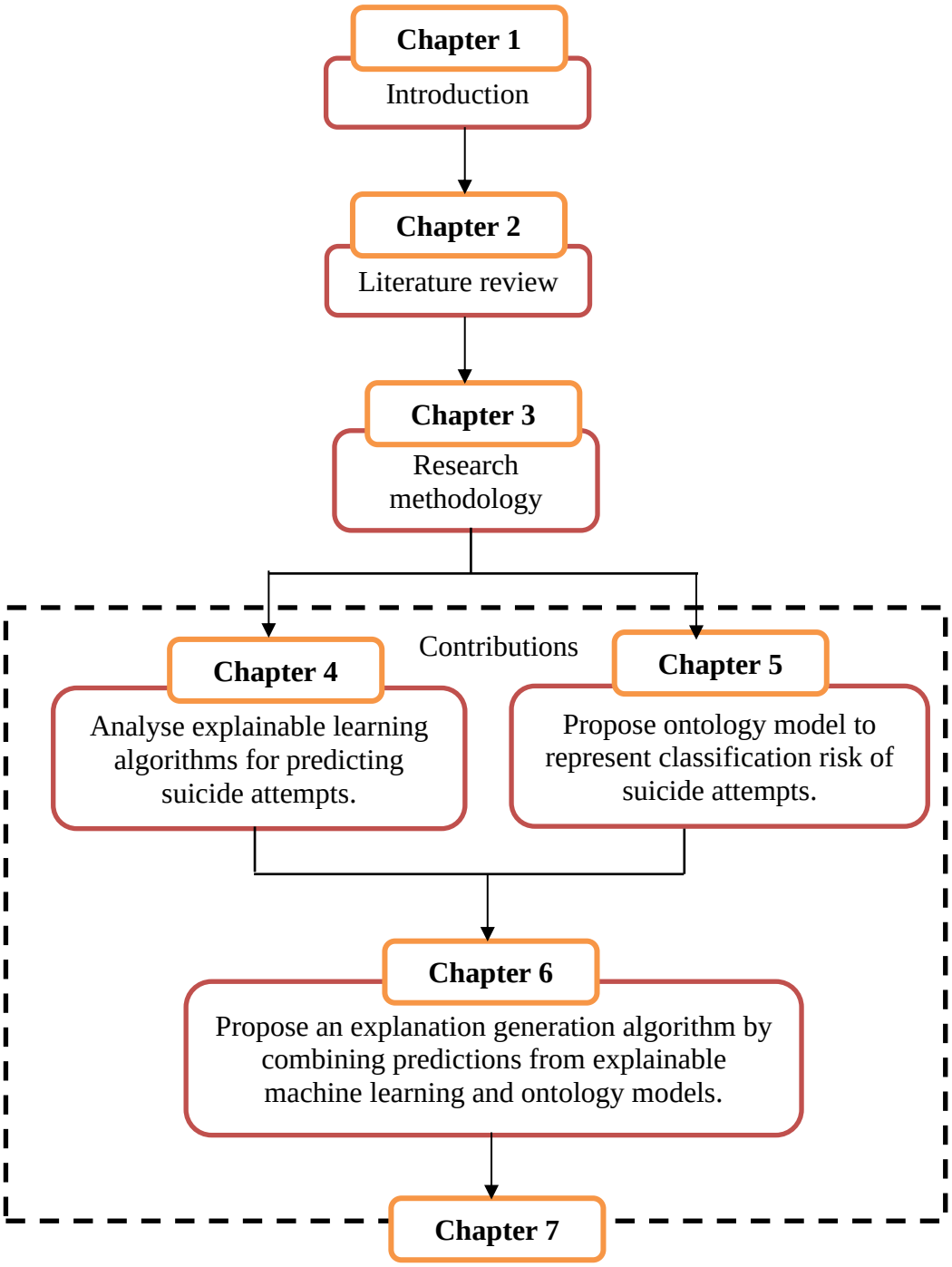

Figure 1.1 Structure of the thesis

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Given the increasing number of suicide cases worldwide, approaches to support, complement, and improve decision-making in the early screening and detection of suicidal behaviour are needed. In this chapter, related works in the area of suicidal behaviour prediction to assist clinicians and psychiatrists are discussed. First, an overview of suicidal behaviour prevention is provided. Next, predictive modeling is discussed in terms of current approaches using suicide risk assessment tools, data-driven approaches, and knowledge-driven approaches. Also, the importance of risk factors influencing the predictive modeling is presented in this chapter and explainable learning approaches are presented to support the explainability of predictive models for suicide attempts. A framework for explainable models in medical applications is discussed and the text explanation generation method is presented in the context of information fusion in the medical domain. Finally, potential trends and research directions emerging from the review are summarized in the context of this research study.

2.2 Suicidal Behaviour Prevention

Suicide is a leading cause of death worldwide. Priority for suicide research and prevention is needed for world health services. According to the World Health Organization (2014), nearly 800,000 people are estimated to die by suicide, and suicide rates are increasing significantly each year (World Health Organization, 2021). Reports on the Health for World's Adolescence indicate that suicide is the second leading cause of death among young people aged 10 to 24, and 76% of global suicides are committed

in low-and middle-income countries (Ritchie et al., 2015). Suicide rates for the global population are shown in Figure 2.1 (Global Burden of Disease Collaborative Network, 2020). Globally, 9 people per 100,000 died by suicide, and the number is expected to increase each year (Roth et al., 2018). Thus, it is clear that suicide is a growing public health problem, and that the phenomena of suicide are highly complex and dynamic, encompassing multiple risk factors such as psychological, biological, environmental, and clinical factors, as well as marked differences among age groups, genders, and geographic regions (Turecki et al., 2019).

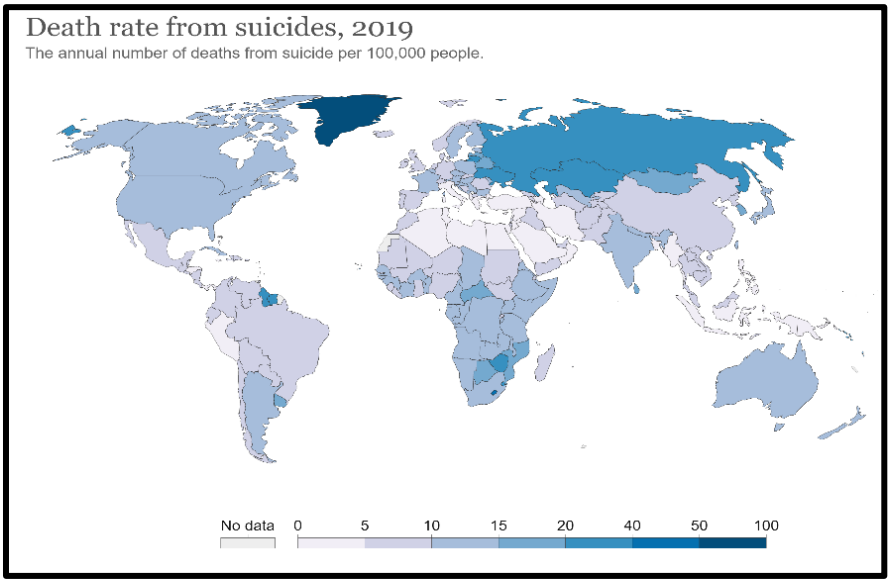

Figure 2.1 The annual number of deaths from suicide across the world *(Source: Global Burden of Disease, 2020)*

Although suicide is one of the leading causes of death among young people, global suicide rates show that people aged 70 years and older are most likely to die by suicide, as shown in Figure 2.2 (Global Burden of Disease Collaborative Network, 2020). The suicides worldwide follow a standard pattern: the older the age group, the higher the death rate. Therefore, the number of suicides in the older population is high compared to other populations (Ritchie et al., 2015; Roth et al., 2018).

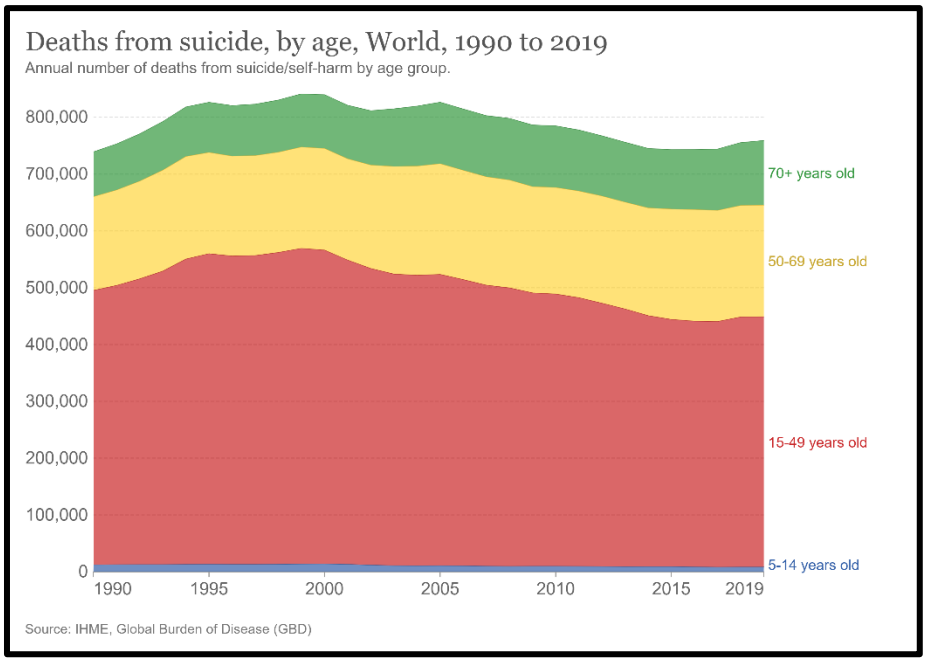

Figure 2.2 The suicide rates by age group *(Source: Global Burden of Disease, 2020)*

In the context of Malaysia, suicide attempts are a growing problem as the suicide rate increases every year. Malaysia has a moderately high suicide rate, at around 12 deaths per 100,000 population compared to other Asian countries (Sinniah et al., 2014). In recent years, the average suicide rate was 5.8 per 100,000 population with an estimated 1841 suicide deaths and 5 deaths per day (Lew et al., 2022). According to the National Institute of Health Malaysia, three out of ten adults over the age of 16 suffer from mental health problems, leading to an increase in the suicide rate. In addition, the National Health Morbidity Survey 2017 showed that suicidal behaviour among Malaysian adolescents has increased recently (Institute for Public Health, 2018). Furthermore, there is still a high incidence of unreported cases, as suicide is illegal in Malaysia and is not accepted by most religions. The incidence poses challenges to mental health providers, clinicians, and psychiatrists, and raises questions about how to improve early detection of suicide and prevention strategies (Abu Bakar et al., 2023; Lew et al., 2022).

Suicide prevention is a global health priority that has critical socio-economic implications, as highlighted by the United Nations' Third Sustainable Development Goals (UN SDGs), which include suicide mortality as a mental health indicator to promote mental health and well-being worldwide. Nearly one million people die by suicide each year, and suicide rates have increased significantly worldwide in recent years (World Health Organization, 2014). In 2021, Health for World's Adolescents study showed that suicide is the third leading cause of death among young people aged 10 to 19. In the context of Malaysia, the suicide rate among adolescents has increased (Sinniah et al., 2014). Therefore, this needs to be addressed as an important public health concern, as suicidal behaviour contributes to an increase in national expenditures and a decrease in productivity (Fonseka et al., 2019).

According to De Leo et al. (2021), suicidal behaviour is defined as thoughts and behaviours related to an individual intentionally taking their own life including, suicide attempt, suicidal ideation, suicide plan, and self-harm. Suicide refers to an act of death that is initiated and carried out by an individual to the end of the action (fatal), while a suicide attempt refers to an act in which an individual harms herself or himself with the intention to die and survives (non-fatal). Suicidal ideation refers to an individual having thoughts of suicide, with or without taking their own life, and suicide plan is defined as a formulation (thoughts) of how and when to perform a suicidal act without active preparation. Self-harm refers to a non-fatal act in which an individual harms herself or himself intentionally or without intentionally with a variety of motives, including the wish to die (O'Connor & Nock, 2014). Thus, the most critical and important feature of suicidal behaviour is the suicide attempt, as suicide attempts involve a physical act that can lead to suicide and have a strong predictive power for suicide deaths (Belsher et al., 2019; Klonsky et al., 2016).

The phenomenon of suicide attempts is highly complex and dynamic, encompassing multiple factors such as psychological, biological, environmental, and clinical factors, as well as significant differences between age groups, genders, and geographic regions (Turecki et al., 2019). Therefore, the first step in suicide prevention can be viewed as a classification and prediction task that allows for accurate identification of individuals at risk for suicide, and thus preventive intervention. A better understanding of the risk factors that distinguish those who attempt suicide from those who seriously consider it and those who repeatedly attempt suicide is important for understanding and predicting suicide attempts (Franklin et al., 2017; Kessler et al., 2017).

Medical decision-making in psychiatry is important for the prevention of suicidal behaviour, especially for the prediction of suicide attempts. Medical decision-making is defined as the process of using clinical knowledge and expertise to diagnose and manage a patient's health condition based on patient preferences, available clinical information, and available resources (Kattan & Cowen, 2009). Decision-making to predict suicide attempts is a complex and critical component of patient care and requires a high level of clinical expertise, judgement, and experience. It is important that clinicians and psychiatrists explain their thought process and involve patients as much as possible in the decision-making process, to ensure that the final decision is consistent with the patient’s values and preferences. However, understanding decision-making in predicting suicide attempts is a difficult process as it involves complex patterns of patient’s risk factors and integrating them with the medical knowledge to make a final decision (Boudreaux et al., 2021; Burke et al., 2019). Therefore, there is a need for a predictive model of suicide attempts that can support clinicians’ decision-making.

2.3 Approaches for Suicide Attempt Predictive Modeling

Predictive modeling of suicide attempts is gaining attention in the healthcare setting because of its importance for effective decision-making, especially for clinicians and psychiatrists in today's world. Predictive modeling has been developed to improve the healthcare system, including medical expert systems and clinical decision support systems (Boudreaux et al., 2021; Fonseka et al., 2019). According to Middleton et al., (2016), a clinical decision support system is used to support health-related decisions and actions with clinical knowledge and patient-specific information to improve the healthcare delivery process and enable effective and easy use by clinicians. The system assists clinicians in making decisions and selecting appropriate treatments to improve the classification of an individual suicide attempt and reduce preventable medical errors. However, little research has been conducted on these systems in the field of mental health, especially suicide prevention (Bernert et al., 2020; Burke et al., 2019).

Therefore, predictive modeling of suicide attempts requires an understanding of the approaches used to develop the models and the risk factors (features) used to predict suicide attempts. Developing models to predict suicide attempts is challenging because the uniqueness of individuals and the interaction of risk factors that influence behaviour have increased complexity. Each individual with a suicide attempt is complex, but the occurrence of risk factors for suicide attempts is an essential foundation for suicide prevention (O'Connor & Portzky, 2018). This is supported by Franklin et al. (2017) meta-analysis, in which identifying risk factors is a key component of a national suicide prevention strategy in the context of individuals, communities, and populations that are most vulnerable to suicide. Therefore, several approaches have been explored and are currently being used to develop a predictive model for suicide attempts that will support decision-making, namely suicide risk assessment tools, data-driven approaches, and

knowledge-driven approaches (Belsher et al., 2019; Cheung et al., 2013; Velupillai et al., 2019).

2.3.1 Suicide Risk Assessment Tools

Suicide risk assessment tools are also known as conventional methods for identifying and classifying individuals at risk for suicide (Baek et al., 2021; Runeson et al., 2017). The risk assessment tools commonly used in psychiatry to assess suicide risk include the Columbia Suicide Severity Rating Scale (CSSRS), the Beck Scale for Suicide Ideation (BSSI), and the Beck Hopelessness Scale (BHS) (Baek et al., 2021). The Columbia Suicide Severity Rating Scale (CSSRS) was developed by Posner et al. (2011) and considers the clinical symptoms and risk factors associated with suicide through semi-semi structured interviews. The subscales of CSSRS consist of 6 items assessing the severity and intensity of suicide attempts. Although CSSRS is a valuable tool for assessing suicide risk, the practice of CSSRS is challenging because clinicians must be trained in the administration and scoring of the CSSRS to ensure reliable results and its scope is limited due to its focus on suicidal ideation and attempts, and not all aspects of suicide risk are captured, including social factors and mental health problems.

Beck Scale for Suicide Ideation (BSSI) is a self-assessment questionnaire developed by Beck et al. (1988), which consists of a total of 21 questions and attempts to measure suicidality and its severity. The content of the questions includes topics such as the desire for life and death, the frequency of suicide incidents, and the perceived sense of control to commit suicide. Based on each individual's experience over the past few weeks, a Likert scale is used to measure the total score of the questionnaire. This scale is useful to assess suicide risk, but it is limited to suicidal ideation and does not capture other important risk factors, such as anxiety, depression, and substance abuse.

The Beck Hopelessness Scale (BHS) was also developed by Beck et al. (1974) and consists of a 20-item self-report scale measuring perceived negative attitudes toward the future. The total score is calculated by summing the scores for each question, and the higher the score, the greater the sense of hopelessness. The BHS focuses on hopelessness and does not consider other risk factors for suicide. All assessment tools are self-report questionnaires designed to identify symptoms related to suicide risk and then classify individuals into a suicide risk group. Psychiatrists then interview individuals to determine the severity of suicidal behaviour. In this case, suicide risk is predicted using a structured interview instrument such as the MINI International Neuropsychiatric Interview Suicide Module to classify individuals with suicide attempts. Managing all risk assessment tools is time-consuming and can be challenging in busy clinical settings. The scale is based on self-reported responses, which can be influenced by current mood and personal bias (Runeson et al., 2017).

Currently, conducting a suicide risk assessment has become a compulsory part of the clinical procedure in psychiatry. Suicide risk management guidelines have been developed to assist clinicians is often challenging, and suicide risk assessments are a supplement to this clinical evaluation (Large et al., 2016). According to the American Psychiatric Association (APA) Practice Guideline for the Assessment and Treatment of Patients with Suicidal Behaviours (American Psychiatric Association, 2006), the guideline is a very important step in improving the management of patients at risk for suicide in the healthcare setting. In Malaysia, there is a guideline that focuses on suicide prevention and management (Ministry of Health Malaysia, 2013), which was used manually to assess patients at risk of suicide. Figure 2.3 shows the guidelines and the framework for suicide risk assessment and management in Malaysia.

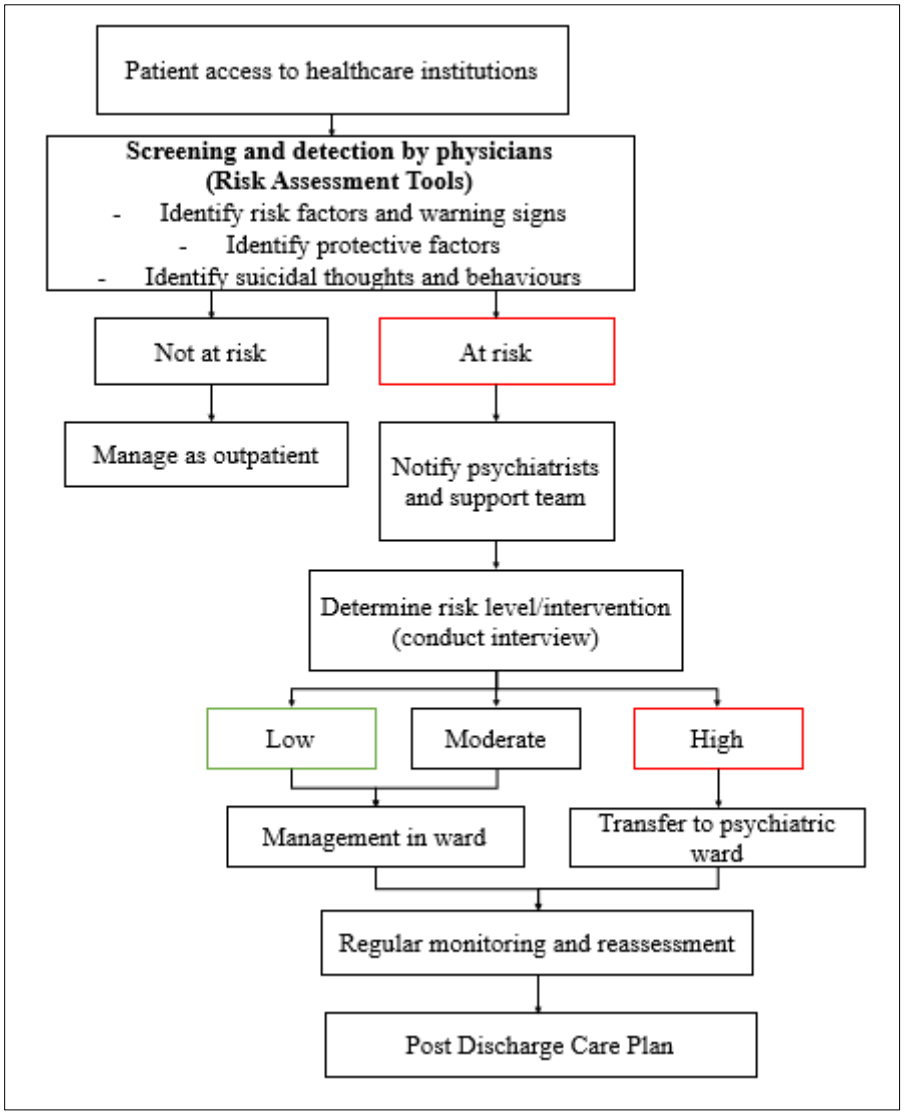

Figure 2.3 Guideline for suicide risk assessment management in hospital *(Source: Ministry of Health Malaysia, 2013)*

The psychiatric evaluation serves as the basis for suicide risk assessment manually. Assessing the degrees of the patient's suicide risk for accurate decisions about immediate safety measures and appropriate treatment settings are standardized and supported by the use of suicide risk assessment tools (Jacobs and Brewer, 2010). However, these risk assessment tools for prediction have been slow to develop and have not improved (Franklin et al., 2017; Large et al., 2016; Ribeiro et al., 2016) due to the costly implementation of new risk assessment tools and capturing risk factors that contextually dependent is complex. A meta-analysis study by Large et al. (2016) found

insufficient reliability and unclear scientific support that risk assessment tools improve the prediction of suicide attempts. Current approaches to generating suicide risk assessment rely on transforming clinical observations into fixed tools for common statistical models. However, it is possible that these approaches will reach their limits, as the development of new risk assessments for suicidal behaviour especially suicide attempts is costly and time-consuming (Velupillai et al., 2019).

In addition to the manual use of risk assessment tools based on risk scores, research studies on suicide attempts have used conventional statistical techniques to identify, classify, and predict an individual's suicide risk, such as chi-square tests, multivariate analyses, and Mann-Whitney U analyses (Belsher et al., 2019; Franklin et al., 2017). Typically, these approaches produce a simple algorithm that requires researchers to use a limited number of risk factors to examine the correlation between risk factors for a simple classification problem and low predictive power for predicting suicide attempts (Cheah et al., 2018). The study by Cheah et al. (2018) shows that the predictive power using the correlation for the number of risk factors (age, income, gender, ethnicity, education, marital status, health status), which does not include employment status, family medical history, and mental health condition is limited. The study only shows that there is a positive correlation between poor health conditions and suicide, without providing the probability of an individual will suicide attempts. Thus, given the dynamics and complexity of suicide attempts, conventional statistical techniques have significantly hampered the effectiveness of informing clinical decision-making.