QUANTUM-LIKE MEDICAL DIAGNOSTIC SYSTEMS FOR DECISION-MAKING
FINAL THESIS

Candidate:
Alexander Iftene
Supervisor:
Dr. Catarina Moreira
School Name:
Science and Engineering Faculty
Queensland University of Technology - QUT

Contents

List of Figures

1 PROJECT TITLE

Quantum-Like Medical Diagnostic Systems For Decision-Making

2 NAMES

Candidate: Alexander Iftene

Supervisor: Dr. Catarina Moreira.

Dr. Catarina Moreira is a lecturer in the School of Information Systems and an associate editor for the *Springer Nature* journal *BMC Bioinformatics*, section *Machine learning and artificial intelligence*. She is a pioneer in the development of quantum-like probabilistic graphical models for decision making to empower human decision-making. She is leading the emerging *QuLBIT: Quantum-Like Bayesian Inference Technologies* project which focuses on the development of non-classical probabilistic graphical models to empower human decision-making. Her work also comprises machine learning / deep learning models, multi-modal image data fusion techniques for medical decision-making and innovative human interactive probabilistic models for explainable AI.

3 EXECUTIVE SUMMARY

It is the purpose of this thesis to provide a set of contributions, both theoretical and empirical, of quantum-like models for medical diagnostic systems as an alternative mathematical approach that has the potential of enhancing our current understandings about the structure of human behaviour and the way medical decisions are made. This will be modelled using deep and machine learning techniques in an effort to create an applicable framework unifying classical and quantum probability theory for medical diagnostics.

4 BACKGROUND AND LITERATURE REVIEW

4.1 Introductory Statement

Cognitive biases and personality traits (such as risk aversion and ambiguity) may lead to diagnostic inaccuracies Literature suggests that in medical diagnosis, at least one cognitive bias or personality trait were found to affect physicians (like overconfidence, lower tolerance to risk, anchoring effect and availability biases) (Saposnik et al., 2016). In the field of cognitive psychology, literature also suggests that under uncertainty, humans tend to violate the laws of probability and logic (Tversky and Kahneman, 1974; Kahneman and Tversky, 1982). It seems that under uncertainty, humans do not follow the normative axioms of expected utility theory. These findings call for new methods and framework to assist human decision-making, namely in the medical diagnostic field.

The basis for reliable and efficient systems to determine potential terminal health problems is not accounting for the fundamental concept of classical probability theory and logic in the scope of human decision-making support systems. This is a paramount field of medical technological advancements as the need for an unbiased and deterministic system is more reliable than a professional medical inference from a human of which violates classical probability theory and can therefore potentially compromise diagnosis's. This is the case because through many cognitive science experiments, resulting conclusions determine that humans violate the laws of probability theory and logic and create decisions based on a marginal amount of uncertainty. Ironically, this has opened up a gap in the science and engineering community to bridge the ideas of classical probability theory and logic and quantum probability theory to form a unified computational framework. The purpose of this thesis will be to focus on the medical decision making for diagnostics and treatments for any illness or disease through the unification of classical and quantum probability theory and translating them into a decision-making framework using probabilistic inference and machine learning techniques.

4.2 Literature Review

4.2.1 Decision-making systems in the Literature

For the specific case of medical decision-making, it is well known that the medical diagnosis and treatments are biased towards the medical professional's belief and experience. Work from George Bergus found that when provided with medical evidence about a certain patient, physicians would classify the probability of this patient as having a urinary tract disease differently according to the order the medical evidence was provided (Bergus R et al., 2002). There have been countless other studies on this field of science and engineering that have indicated much of the same, which is one of the main drives into this thesis concept and the developments henceforth.

It is renown throughout the literature of cognitive science and economics that when it comes to decision-making under uncertainty, humans by default will opt for preferences that don't maximise the expected utility theory or that are inconsistent with the axioms of expected utility theory, leading to decisions that are either sub optimal, paradoxical or even irrational (Tversky and Kahneman, 1974). The decisions made by humans are based on beliefs concerning likelihood of uncertain events such as the diagnosis of a future potentially terminal illness, disease or cancer. Despite the medical knowledge and a life-time of experience in the field, the unwavering truth is that the diagnosis will be based off a series of patterns, numerical odds or subjective probabilities which is subjectively based on heuristic principles. These paradoxical decisions can result in cognitive biases (Kahneman and Tversky, 1982), contradictions of major economic principles such as the Sure Thing Principle or to the laws of probability theory and logic (Trueblood et al., 2014).

Through many iterations of research throughout the years, it has given insights into the vast range of human cognition which contradicts the widely accepted research on logic and probability theory. Unfortunately for current decision support systems, they are cannot accommodate sub-optimal or irrational human decisions, because they are built using recursive learning algorithms. Ironically, these algorithms are based on probability theory. The decision systems based in neural networks are designed to minimize cost functions and to pursue optimisation (Mazurowski et al., 2008). This demonstrates the axioms in probability theory when they are involved in normative decision-making frameworks which are classified as ”correct”. When accounting for preferences under uncertainty, the research shows that models based on classical theories of rational choice determine decisions for individuals, as opposed to guiding them. However, when choices from an individual do not align with axioms of classical probability theory, it does not mean that the choices are ”wrong” or ”incorrect”, but instead is effective at advancing the individual's desired intentions.

Certain models that are not only optimal for rational decisions, but also accounting for sub-optimal, irrational and paradoxical systems would greatly benefit the advancement of decision support systems in the medical field since there would be a more optimised agreement between the system and the decision maker. This has the potential to provide the ultimate unification of the rational and irrational worlds of decision making science.

If one was to revisit the fundamental core concepts that underline generative decision models of humans, we will make decisions that violate the laws of probability theory while under uncertainty. If this is the case then the question begs as to if these laws are in themselves limited to fully express the nature of cognitive biases and decision-making in humans. A more malleable and flexible probability theory could provide improved insights and accommodate many paradoxical findings reported in the literature. Fortunately for the cognitive science community, one such framework that has been extensively studied in the domain of human cognition is the quantum probability theory concept (Trueblood et al., 2014).

4.2.2 Quantum-Like Bayesian Technologies for Decision Making

A reputable probabilistic framework called the Quantum-Like Bayesian network (Moreira., 2017) is an example model that is well established in the literature for predicting and accommodating paradoxical human decisions across different decision scenarios ranging from psychological experiments (Moreira and Wichert, 2016) to real-world credit application scenarios (Moreira et al., 2020a). The main difference between a Bayesian Network (BN) and a Quantum-Like Bayesian Network (QLBN) is the way one specifies the values of conditional probability tables. In traditional BNs, one uses real numbers to express probabilities and numerical paradoxes, however, in quantum mechanics, these probabilities are expressed as probability amplitudes which are represented as complex numbers. As quantum mechanics explains the component of the wave functions of particles, it also explains that value of the wave function of a given particle relative to space and time is relative to the position of the particle. When making probabilistic inferences using QLBNs, one must convert these quantum amplitudes into classical probability values by measuring the squared magnitude of these amplitudes (this is exercised using Born's Rule (Wang et al., 2013)).

In the realm of probabilistic uncertainties where random variables are not observed, this measurement made from Born's Rule can be used to attenuate the classical probabilities through quantum interference effects and therefore allow QLBNs to accommodate for all these paradoxical findings concerned with human data. In order to develop a QLBN we must make use of BNs in order sustain the link between probability theory and graph theory. A property of graph theory is the assumption that one can build complex systems by combining smaller components, which in reality is easier to simulate as opposed to calculating all possible events. BNs provide the decision problem in smaller packets that can be unified to perform required inferences and only the relevant inferences that are needed will be computed. A typical representation of a classical BN can be seen below in Fig.1.

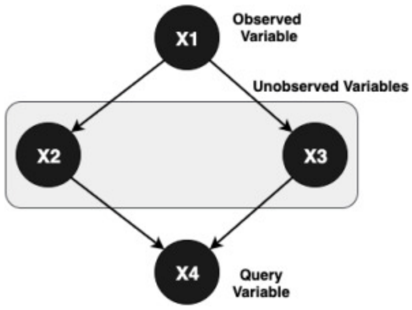

Figure 1: A simple Bayesian Network representation (Moreira et al., 2020b).

These quantum-like models alone are probabilistic models and therefore cannot be applied in decision-making settings because this implies the notion of utility. For instance, it is not enough that the doctor knows the probability of a patient having a disease or illness. The doctor should rather act on the information provided by the system and decide what is the best procedure or treatment for the patient, this is where the importance of utility comes in. The medical professional must maximise the utility of the decision which will equate to the improvement of well-being of the patient.

The decision-making systems of today are based off classical probability theory and logic which cannot predict paradoxical decisions made by humans. Instead of accommodating for this important aspect of human decision-making science, they are usually ignored and classified as irrational with the ironic conclusion that they are ”wrong”. In this sense, the quantum-like Bayesian networks have been recently extended to Quantum-Like Utility Diagrams (QLUD), which use the expected utility theory as the main mathematical framework to compute the maximum expected utility of some action. The difference is that the probabilities that are computed in the Expected utility are derived from a quantum-like Bayesian network, which again can produce interference effects in probabilistic inferences under uncertainty and influence the final expected utility. An example representation of a functional QLBN is depicted in Fig.2 below:

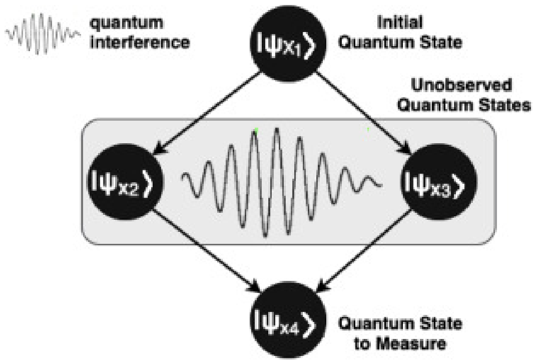

Figure 2: A simple Quantum-Like Bayesian Network representation (Moreira et al., 2020b).

4.2.3 Other Quantum-Like Models in the Literature

There are many examples of quantum-like models that exist in the literature, a few that will be mentioned due to their relevance to the subject matter are the Quantum-Like Approach, the Quantum Dynamic Model and the Quantum-Like Prospect Theory. These models delve into how we could apply them to predict the results associated with decision-making problems specifically related to computer-aided diagnostic systems (CAD).

4.2.4 The Quantum-Like Approach

The Quantum-Like Approach was proposed by A.Khrennikov and this model corresponds to a general contextual probability space from which the classical and and quantum probability models can be derived. A contextual probability example would be the Vaxjo Model where the context relates to the circumstances that form the setting for an event of which it can be fully understood. An exmaple of this would be best explained using cognitive science with which mental contexts are taken into account, same with social sciences where social contexts are considered and so on and so forth when regarding other domains (Moreira and Wichert, 2016).

A set of variables is inherently associated with the context where in quantum mechanics an observable corresponds to a self ad jointed operator on a complex Hilbert Space. The Vaxjo Model models these possible events mathematically with their respective values and dissonance.

$$Pr_{context} = (C, \Theta, \pi) \tag{1}$$

An example using this equation can be explained by using the context C as an element of various contexts for an observable value of the elements of Θ. The probability of the value of one observable is expressed in terms of the conditional probability which is in reference to the probability distribution given by π.

$$\pi(\Theta, C) = Pr(a = \alpha | C) \tag{2}$$

If the quantum realm is considered, the above equation can be interpreted as the selection with respect to the result of $a = \alpha$ of a measurement performed in a. For the contextual probability model, the Vaxjo framework corresponds to a model M described by the context and the set of observable O. All of this corresponds to a probability distribution of some observables belonging to a specific context. We can assume that for a context C there are dichotomous observables and that each of these take some values of a and β respectively. The Vaxjo Model can be built from the generalised structure of the quantum law of total probability. This means that the formula is a combination of the classical probability theory with a supplementary term called the interference term which does not exist in classical probability and enables the representation of interference between quantum states (Moreira and Wichert, 2016).

$$Pr(b = \beta) = Classical_Probability(b = \beta) + Interference_Term \tag{3}$$

Using the Vaxjo Model probabilities for trigonometric spaces will not provide a complete description of a quantum system because it can violate the axiom of probability theory. An algorithm was proposed in the literature that actually extends the Vaxjo Model and is able to accommodate the positivity axiom. This algorithm was proposed by A.Khrennikov and it is called the Quantum-Like Representation Algorithm (QLRA). The quantum complex amplitudes can be obtained from classical probability by using Born's Rule as stated before, however, one can simplify Born's Rule for two variables using the following equation (Moreira and Wichert, 2016).

$$\begin{aligned} Pr(\beta|C) = Pr(\alpha_1|C) * Pr(\beta|\alpha_1, C) + \\ Pr(\alpha_2|C) * Pr(\beta|\alpha_2, C) + 2 * \sqrt{Pr(\alpha_1|C) * Pr(\beta|\alpha_1, C)} * \\ \sqrt{Pr(\alpha_2|C) * Pr(\beta|\alpha_2, C)} \end{aligned} \tag{4}$$

This equation corresponds to the representation of quantum law and probability through the Vaxjo Model. The Quantum-Like Approach model can be extended to more complex decision scenarios that can include two or more random variables. The Quantum Dynamic Model on the other hand offers the ability to perform quantum time evolution which involves the creation of doubly stochastic matrix which represents the rotation of the participant's beliefs(Moreira and Wichert, 2016).

4.2.5 The Quantum Dynamic Model

The Quantum Dynamic Model is the work of Busemeyer and uses the notion of doubly stochastic matrices to preserve unit length operations and to obtain a probability value that doesn't require normalisation to occur(Trueblood et al., 2014). Using the Prisoner's Dilemma model for the principle of application, the participant's actions are represented by superposition vectors with all possible actions which are modelled by the use of wave functions which are quite popular in quantum mechanics and probability theory. This stochastic matrix can only be modelled by the use of an auxiliary Hamiltonian matrix where all values correspond to a certain value of parameters representing payoffs of the defect and cooperate actions of the individual (Moreira and Wichert, 2016).

Using this model, one can estimate the quantum probability of a certain situation given the participant's choices. It is however a very general framework that can lead to many different probabilities which solely depend on how the parameters are chosen to fit. An example of this is shown in a previous study by Moreira and Wichert (2016) where to illustrate this concept they decided to fix one of the parameters used in the modelling of the Prisoner's Dilemma game in the Quantum Dynamic Model and vary the intervals. The below figure demonstrates a few possible probabilities that can be obtained by the Dynamic Quantum Model for the Prisoner's Dilemma game.

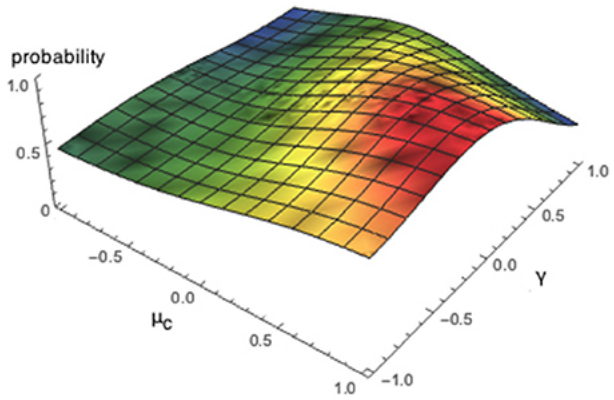

Figure 3: Illustration all possible probabilities that can be obtained by varying the parameters γ and μC (Moreira and Wichert, 2016).

These parameters are designed in a psychological setting for the operation of Quantum Dynamic Models. Parameters are incorporated to model dissonance effects and payoffs of the participants which also provide an approximation for the psychology of the problem that is not observed in other quantum models. The biggest disadvantage to Quantum Dynamic Models is the dependency on Hamiltonian matrices because creating a Hamiltonian creates a dynamic in itself where it is required that all possible interactions of the decision to be known. The Hamiltonian also grows exponentially with the complexity of the decision problem and the computation of the unitary operator from such matrices is complex. This is why approximations are used instead of the former because of the sheer complexity of the calculations involved in using matrices of this magnitude (Moreira and Wichert, 2016).

4.2.6 Computer Aided Medical Diagnostic Systems

A Computer-aided Medical Diagnostic systems (CAD) in the medical field is at the forefront of cutting-edge intelligence systems in the interface of medicine in computer science and engineering. CAD systems are utilised to emulate the way a medical professional makes medical diagnosis. More advanced CAD systems can analyse clinical data and provide deeper and fresh insights of knowledge, which can enhance current diagnostic systems and improve their capabilities over time. The way this can be achieved is to have feedback systems implemented so that they may interpret successful and unsuccessful data from different sources (Yanase and Triantaphyllou, 2019). While medicine and computer science have advanced exponentially in the past few decades, the respective fields have become increasingly complex and diverse. Intelligent CAD use artificial intelligence (AI), data mining and machine learning approaches to dissect medical data that can often be complex and involve massive amounts of empirical data. Systems as such greatly benefit the process of diagnostic decisions in a wide range of illnesses and diseases (Giger, 2018). The contributing factors that have led to the development of CAD systems in medicine are the complexity of the medical diagnosis process, the size of the clinical data relevant to certain illnesses and diseases and the preexisting rules provided by the user to form diagnostic guidelines for the machine.

An example of an early development and application of CAD systems was in the 1960s when a number of radiologists had worked on a CAD system to detect abnormalities in medical images (Winsberg et al., 1967). In this paper, the reading of radio graphs of asymptomatic patients via automated systems by means of optical scanning and computer interpretation was investigated. It was found that through implementing a computational system to calculate the densities of the film used to capture mammograms of patients, the comparison of two breasts with different shapes and sizes could then be done more uniformly. A reliable and effective CAD system is known as a generic expert system. The mechanics of a general expert system is broken down into sections as shown in Fig.3. The user interacts with the inference engine which can use either a forward or backward chaining approach to determine the resulting decision. A computer scientist is usually in charge of the knowledge base which is comprised of a set of diagnostic rules and facts which are set to be true as they have knowledge pertinent to the purpose of the expert system.

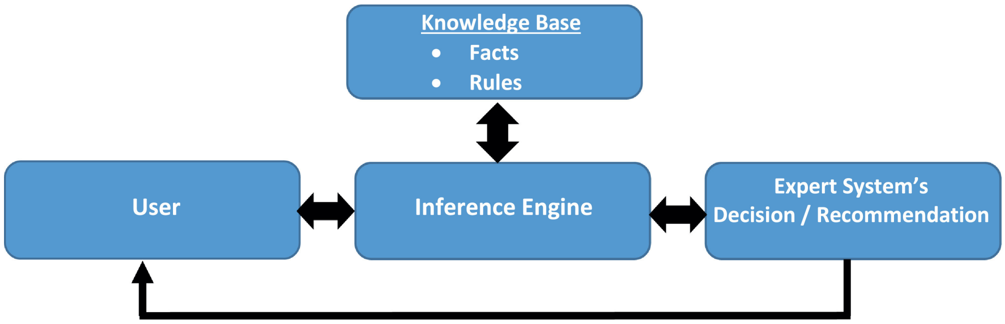

Figure 4: Simplified architecture of a general purpose expert system for medical diagnosis (Yanase and Triantaphyllou, 2019).

Generally the logic comprises of various IF-THEN statements which is depicted in Fig.4 below, where an ideal implementation would consist of an antecedent part (a logical expression that can neither be true or false) and the consequent part which is expected to be true if the antecedent part is true. The antecedent part is the combination of n premises symbolised as P_1 and P_2 and so on which are either true or false. These expert systems were created in an effort to emulate the classical diagnostic decision-making processes that humans are expected to obey. This is an excellent demonstration of how the simple IF-THEN logic can be represented as a method of obtaining a knowledge base for the medical professional.

4.2.7 Implementations of Machine Learning Techniques in CAD

The machine learning approaches applied to medical diagnostics are divided into two categories; supervised and unsupervised. Both approaches utilise data sets for training their algorithms which is useful for developing a model to infer data points on a newer set of data. Supervised learning data sets are trained by parsing observations along with a class value, for example, certain mammograms may show benign signs of cancer while others show malignant cases. On the other hand, in unsupervised learning the data sets are not associated with a class value where the idea is to investigate whether the data can be grouped into clusters which will aid in the understanding of the diagnostic process. Unsupervised learning is more challenging

```
IF      [(Premise P1) AND (Premise P2) AND ... (Premise Pn)]
THEN    (Premise Pk)

An example from MYCIN:
IF:     [ (The infection which requires therapy is meningitis)      AND
          (Organisms were not seen on the stain of the culture)     AND
          (The type of the infection is bacterial)                  AND
          (The patient has been seriously burned) ]
THEN:   [ (There is suggestive evidence (0.5)) (pseudomonas-aeruginosa
          is one of the organisms which might be causing the infection) ]
```

Figure 5: The pseudo-code of the IF-THEN methodology of logic rules as obtained from the MYCIN expert system (Buchanan, 1984).

to cultivate than supervised learning as there is less readily available information from the system to work with.

The main benefit with using machine learning approaches on a medical diagnostic system is that the processing power of a machine vastly supersedes that of a human brain, meaning that immense amounts of clinical data can be parsed and analysed and produce potentially important inferences. Machine learning allows for the identification of data consistencies and inconsistencies, manage more complex dimensions of data and derive probabilistic assessments more efficiently and effectively. This is what is meant by an ”intelligent” system as it can improve the performance over time by utilising the success and failures of the system they derive simultaneously. An excellent example of such theoretical applications is the well-known supervised approach based on the inferences of decision trees. A decision tree (as seen in Fig.5) is essentially the methodology which partitions the training data set into groups of a hierarchical manner where there is only one class which is classified as predominant. The information along the root node (start of the decision tree) and a leaf node (extension of the decision tree) corresponds to a pattern based on the inputs of the data sets which can be used to identify new data points of unknown class values, as seen below.

A typical alternative to decision trees is the K-nearest neighbour approach (see Fig.6) which is defined by when an arbitrary new data point is presented for classification, all the remaining data points in the training set are examined and the K points which are closest to the arbitrarily chosen data point is selected (Kumar et al., 2006). The struggle of figuring out how to select the ideal K component is easily resolved by considering many candidates for K and then correlating the performance of the inferred model on a testing data set. Typically, the model with the best performance is used to derive the optimal K value for the data set, however, there is not a widely accepted approach on how to achieve this so it is entirely application dependent at this stage.

```
Input:    D = training dataset defined on attributes and with class values;
Output:   Decision tree <Tree_T0> capable of classifying the training dataset D;

Algorithm Build_Decision_Tree (D)
Begin/* Algorithm <Build_Decision_Tree> */
Tree_T0 = { } ; /* initialization */

/* If we have only one class (i.e., D is "pure") or a stopping criterion is invoked, then stop.
   We have reached a "leaf" node of the decision tree.
   Otherwise, determine the best attribute to split the remaining of the data.
   Proceed recursively until only "leaf" nodes are created and no more splits are possible */
IF (D is "pure" or a stopping criterion is invoked) THEN
        (Create a leaf node that corresponds to D and STOP); /* Tree_T0 has only this node */
ELSE
Begin /* else #1 */
        FOR each attribute in a ∈ D DO
        begin /* do-loop #1 */
                Compute the information-theoretic evaluative value if we split on a
        end /* for-loop #1 */
        a_BEST =        Most promising attribute for a split;
        Tree_T0 =       Create a decision node based on attribute a_BEST;
        D_i =           Induced sub-datasets from D based on a_BEST;
        FOR each sub-dataset D_i DO
        begin /* do-loop #2 */
                /* Call recursively Algorithm <Build_Decision_Tree> with argument D_i */
                Tree_T_i = Build_Decision_Tree (D_i);
                Attach Tree_T_i to the corresponding branch of Tree_T0;
        end /* for-loop #2 */
End /* else #1 */
RETURN Tree_T0;
End /* Algorithm <Build_Decision_Tree> */
```

Figure 6: The pseudo-code of a decision tree inference from a data set (Yanase and Triantaphyllou, 2019).

```
Input:  D = the training dataset /* each data point is associated with a class value */
        K = number neighbors to consider (size of dataset D >= K >=1)
        Y = a new data point of unknown class value that we need to classify.
Output:The inferred class value of the new data point Y.

Algorithm K- Nearest_Neighbors (D, K, Y, Class value of Y);
Begin /* Algorithm K-Nearest_Neighbors */
        FOR (each data point X_i in dataset D) DO
        begin /* do-loop #1 */
                Compute the distance between the new point Y to X_i;
        end /* do-loop #1 */

        /* Find the closest K neighbors to the new point Y */
        Determine K_Closest_Neighbors = set of K data points from D that are closest to Y;
        /* Determine the majority class. Ties can be broken randomly */
        Majority_Class = the majority class of the members in K_Closest_Neighbors;
        RETURN Majority_Class as the class of the new point Y;
End /* Algorithm K-Nearest_Neighbors */
```

Figure 7: The pseudo-code representation of a K-nearest neighbour approach to a data set (Yanase and Triantaphyllou, 2019).

The oldest known supervised approach is called the "perceptron" system (see Fig.5) which is an effort to emulate the human brain function. It uses a simple neural network approach where each input field corresponds to an input node and a weighted approach denoted as vector W is used to determine the predicted class. If the weighted value in vector W is greater than or equal to a control parameter t then it will be sorted into the first class of a two class dynamic, and otherwise if it isn't. This system is also based on some control parameters which are calibrated by using a test data set much like the previous approaches.

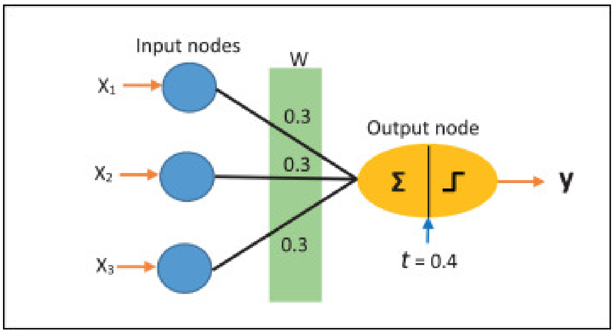

Figure 8: Neural network representation of a "perceptron" based approach (Yanase and Triantaphyllou, 2019).

Artificial Neural Networks (ANN) are a generalised approach of the perceptron system which incorporates the idea of implementing multiple perceptrons where the main difference is a hidden layer which can be seen in Fig.6. This hidden layer acts as a way to infer more generally from a data set as opposed to its predecessor system but the results are improved and optimised. The advantage of ANN systems is that they comprise of three different types based on the number of hidden layers within the system. These neural networks are called Feed-forward, Recurrent and Convolutional, where each type has its own complexity and use for certain cases. Feed-forward NNs are the standard issue ANN where the information travels in an unidirectional manner from an input node to a processing node. The catch is that the hidden layers may or may not be present in the process which makes it more explicable. Recurrent NNs are more widely used and are more complex in that the data travels in multiple directions and the data is stored in processing nodes where the algorithm learns to improve the functionality. Convolutional NNs are the most popular approach to ANN systems today due to their ability to be applied in facial recognition cases. This NN allows for encoded attributes into the input to assume that the processed data is in fact an image.

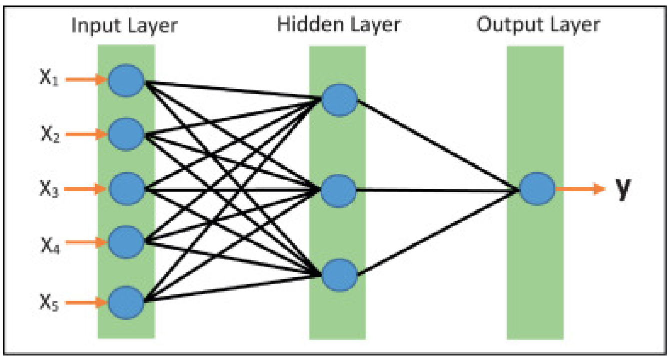

Figure 9: Artificial Neural Network (ANN) representation (Yanase and Triantaphyllou, 2019).

4.3 Research Problem

In the last section, we reviewed the main concepts and approaches that are used in the literature for interpret-ability and explain-ability of the role human decision-making has in the application of medical diagnostic systems. Although each model provides specific and unique advantages towards the challenges of applicability, they also suffer from different limitations. For example, the fundamentals of contemporary CAD systems fail to account for the obvious flaw in human decision-making science when approaching it from a classical probabilistic point of view. Human beings defy and violate classical probability theory and therefore leave an ambiguous amount of uncertainty for CAD systems to process and identify, which in most cases is done with a large amount of bias. The missing link lies in the QLBN system where uncertainty is modelled more explicitly and accounted for without bias, which if implemented into a CAD system framework which accounts for classical and quantum probability theory, will produce models that are more realistic and certain. The idea is not to decide for the user, but instead guide the course of diagnosis to a conclusion which determines a case that yields the highest utility for the patient all the while providing the medical professional an unbiased and generalised outlook. This dynamic interpretation on human decision-making models for CAD system architectures outlines the necessity for an unbiased and a more generalised approach for medical professionals. QLBNs, classical probabilistic BNs, machine learning applications and utility theory for CAD systems are the key components for this thesis proposal and beg the following research questions:

- **Research Question 1:** *How to combine the elements of Quantum-Like Bayesian Networks, classical probabilistic Bayesian Networks, utility theory and machine learning applications to form into an unbiased and generalised decision-making framework for medical diagnostic CAD systems?*

  In order to provide the user with causal understandings of why certain features contributed to a certain prediction, a theoretical causal abstract model should be defined. This decision-making framework should be able to compute cause and effect relationships between the permuted features based on a strong mathematical theory and based on probabilistic graphical models and machine learning techniques. Therefore, we hypothesise that the system should provide better insights and consequently better decisions due to the improved probabilistic nature of the model. A medical professional can then offer more precise and reliable explanations for the predicted result, while accounting for the utility of the patient.

- **Research Question 2:** *What metrics and/or criteria would be suitable to access the quality of explanations generated from the proposed decision-making framework in the context of a prediction?*

  One of the biggest open research questions in the literature is on how to create metrics that are able to assess how one explanation can be understood to be better than another. By proposing novel and standard metrics for explain-ability, one can promote the reliability of models, leading to a higher trust of human decision-makers. It is necessary to develop new standards and norms in terms of evaluation metrics which are able to assess the predicted model. Furthermore, the time and memory efficiency evaluation are common metrics to evaluate the quality of an algorithm.

- **Research Question 3:** *What is the effectiveness of the proposed framework when applied to generalised predictive domains such as medical decision-making?*

  After defining the theory and the evaluation metrics, an important aspect is to determine the effectiveness of the proposed causal abstract model in a certain domain. For instance, medical decision-making is an important domain that highly benefits with explainable models. Holzinger theorised what could be a suitable model to interpret medical images of human pathologies and the urgent need for such systems (Holzinger et al., 2019). Finance is another field which should require frameworks for customer credit evaluation. In this research, the domain will be constricted to the medical medical-making area where we will investigate the effectiveness of the decision-making framework.

5 PROGRAM AND DESIGN OF THE PROPOSED RESEARCH INVESTIGATION

5.1 Objectives, Methodology and Research Plan

The primary focus of this research work rests in the development of theoretical models and associated algorithms for interpretative models grounded on the notion of certainty and generalisation. This research work will have three major components: the theoretical contribution of a human decision-making framework for interpretation; the algorithmic contribution that promotes understandable, unbiased and generalised results for the user; and the implementation of the framework which will consist in the utilisation of the proposed model and algorithms in a specific domain.

Given that the aim of this research work is to incorporate the notion of an interpretative framework with applicable understandings to the user, one can extend the Bayesian network to a Quantum-Like Bayesian network that will enable the computation of causally related features and the effects they have in the prediction. Note that the computational framework is an approximation of a predictive model represented by the theories and mathematical interpretations of QLBNs and classical and quantum probability theory. To generate an interpretative model as such, four points should be addressed:

- **Aim 1 (Theory):** Leverage on existing Quantum-Like models and extend their notions to CAD.
- **Aim 2 (Algorithms):** Enhance CAD using quantum-like probabilistic inferences to deal with medical decisions with high levels of uncertainty.
- **Aim 3 (Evaluation):** : Evaluate the quality of causality base explanations in a variety of empirical settings.
- **Aim 4 (Application):** Develop an open-source framework, which will configure the proposed approaches to yield explainable predictions and render the results to users in an understandable, visual, and interactive manner.

5.2 Timeline for Completion

The following figure below demonstrates a projected timeline of the implemented methodology that will be carried out throughout the project. Below is the phases that will be implemented in order to achieve the desired project outcome of a functioning decision-making framework that can interpret real-world medical data and model predictions. The ultimate aim of this thesis is to produce a published paper.

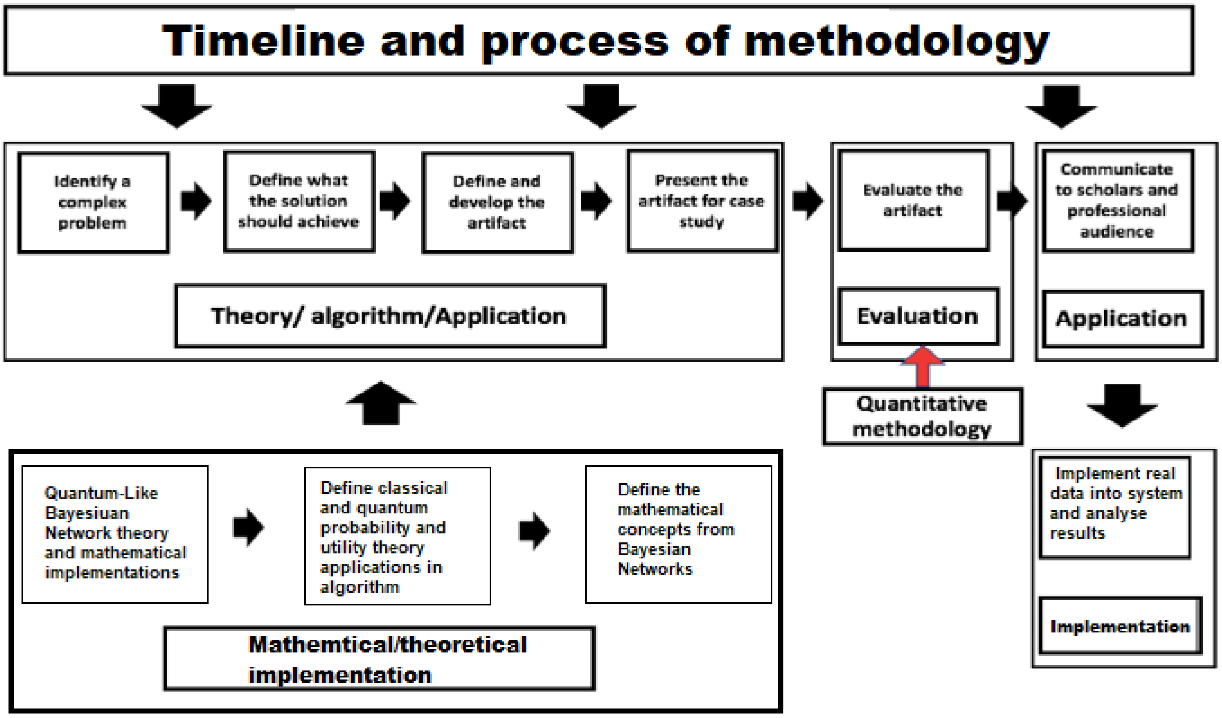

Figure 10: Generalised process of the design process and expected timeline of the research objective.

Phase 1: Literature review, defined problem and background research [Theory]

- Identify the current problem of human decision-making models and use machine learning to focus on the research direction.
- Develop the conceptual model foundations of a Quantum-Like decision-making framework for CAD systems.

Phase 2: Mathematical and theoretical implementations in machine learning algorithm development. [Interpretation]

- Identify the mathematical models and theories that are required to develop the foundations of the framework.

Phase 3: Testing and evaluate the system [Explanation]

- Map results from framework into understandable and interpretative explanations for the user.
- Build a standard for evaluating the theory-based algorithm.
- Apply the model using real-world medical diagnostic data.

Phase 4: Open-Source framework, complete thesis write up, and publication

- Development of an open-source explainable framework for human decision-making CAD systems incorporating quantum and classical theories and applications.
- Thesis write-up.
- Publish paper.

5.3 Future Work and Conclusion

Throughout the process of developing the foundations to create this system, I have learnt a great deal in regards to theoretical concepts to do with classical and quantum probability theory and the integration needed to achieve a QLBN that can be used in real-world CAD systems. The majority of my time was used to research and learn about the mathematical theories of the system I am to build as there is a lot of background research that needs to be done in order to start implementing models used for machine learning. The approach to this project has not changed and the supervisor and I have decided that the course is in line with the trajectory the project is needed to head in order to create this system. I have not to this date been able to start constructing my system as I still have not achieved the level of understanding to implement these complex mathematical systems into a machine learning framework. The fact of the matter is this will also be the first one-of-a-kind system to be created using my supervisor's work and my developments from it, there exists no such technology to date of the system I am to create in this thesis project. In order to make sure that this system is a success, it is imperative that I familiarise myself completely with the necessary foundations.

The initial design phase will consist of porting the mathematical theories derived from classical and quantum realms to a Python framework where the modelling of such theories will take place. This will be the most difficult part as this is where the foundations of the proposed system resides. After this is completed, then the construction of the prototype of the machine learned model will be executed. During this phase, it will be analysed if the model is engineered efficiently enough to accommodate for mock-up data to be imported and trained in the proposed machine learning model. This phase will be another difficult process as there are a multitude of things that can go wrong during the learning and training phase of the model. There will be other methods and avenues explored during this process to ensure that the maximum amount of resources is used to get a fully functioning QLBN system that can operate using real world medical diagnostic data. Obviously, the final stage is to have the system fully functioning and conditioned to take any data given and create results that satisfy the quantum and classical probabilistic theories and accounting for heuristic biases all the while accommodating the maximum utility of the patient. Overall, it is expected that the project outcomes will be achieved as per the project timeline and a never before seen technology will be created for use in the medical field. Real-time data will be collected, verified and tested. Thankfully for this project, there is no cost for the production and the only difficulties that can be mentioned are to do with the process of development: background knowledge and research, availability of real-world data, and time.

6 RESULTS

6.1 Diabetes Discrete Dataset

	Diabetes	times_pregnant	glucose_conc	Diastolic_BP	Triceps_thk	2_hr_insulin	BMI	Pedigree	Age
0	1	(4.0, 7.0]	(131.0, 155.0]	(70.0, 76.0]	(33.0, 40.0]	(-0.01, 63.0]	(33.2, 35.5]	(0.59, 0.78]	(41.5, 81.0]
1	-1	(-0.01, 1.0]	(-0.01, 93.0]	(60.0, 66.0]	(18.0, 24.0]	(63.0, 94.0]	(25.9, 30.0]	(0.07500000000000001, 0.24]	(20.99, 22.0]
2	1	(-0.01, 1.0]	(131.0, 155.0]	(23.99, 60.0]	(33.0, 40.0]	(167.0, 230.0]	(39.1, 67.1]	(0.78, 2.42]	(32.0, 41.5]
3	1	(2.0, 4.0]	(-0.01, 93.0]	(23.99, 60.0]	(29.0, 33.0]	(63.0, 94.0]	(30.0, 33.2]	(0.24, 0.32]	(24.0, 27.0]
4	1	(1.0, 2.0]	(155.0, 198.0]	(66.0, 70.0]	(40.0, 63.0]	(230.0, 846.0]	(30.0, 33.2]	(0.07500000000000001, 0.24]	(41.5, 81.0]
...	...	...	...	...	...	...	...	...	...
389	1	(-0.01, 1.0]	(155.0, 198.0]	(82.0, 110.0]	(40.0, 63.0]	(230.0, 846.0]	(39.1, 67.1]	(0.07500000000000001, 0.24]	(24.0, 27.0]
390	1	(-0.01, 1.0]	(119.0, 131.0]	(82.0, 110.0]	(33.0, 40.0]	(94.0, 125.0]	(35.5, 39.1]	(0.78, 2.42]	(32.0, 41.5]
391	-1	(1.0, 2.0]	(-0.01, 93.0]	(23.99, 60.0]	(24.0, 29.0]	(-0.01, 63.0]	(25.9, 30.0]	(0.59, 0.78]	(20.99, 22.0]
392	-1	(7.0, 17.0]	(93.0, 103.0]	(70.0, 76.0]	(40.0, 63.0]	(167.0, 230.0]	(30.0, 33.2]	(0.07500000000000001, 0.24]	(41.5, 81.0]
393	-1	(4.0, 7.0]	(119.0, 131.0]	(70.0, 76.0]	(18.0, 24.0]	(94.0, 125.0]	(25.9, 30.0]	(0.24, 0.32]	(27.0, 32.0]

394 rows × 9 columns

Figure 11: Discrete data of the Diabetes dataset used for modeling the BN. (Moreira et al., 2020b).

6.2 Diabetes Bayesian Network Model

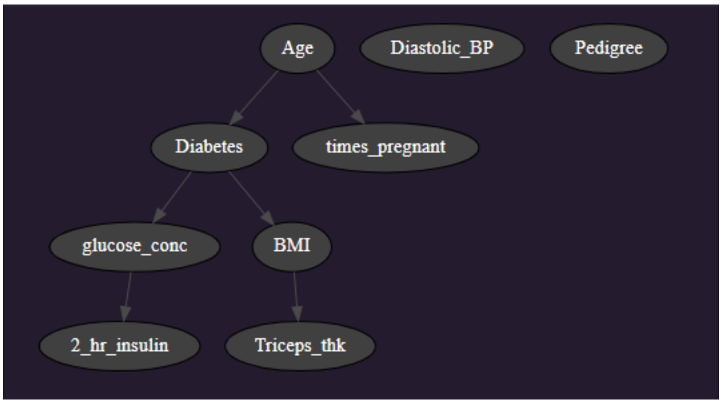

Figure 12: Bayesian Newtork of the Diabetes dataset. (Moreira et al., 2020b).

	Diabetes	
Age	-1	1
"(20.99, 22.0]"	0.9082	0.0918
"(22.0, 24.0]"	0.8465	0.1535
"(24.0, 27.0]"	0.7494	0.2506
"(27.0, 32.0]"	0.5554	0.4446
"(32.0, 41.5]"	0.5409	0.4591
"(41.5, 81.0]"	0.3640	0.6360

	times_pregnant				
Age	"(-0.01, 1.0]"	"(1.0, 2.0]"	"(2.0, 4.0]"	"(4.0, 7.0]"	"(7.0, 17.0]"
"(20.99, 22.0]"	0.6613	0.2467	0.0911	0.0004	0.0004
"(22.0, 24.0]"	0.5752	0.2203	0.1696	0.0344	0.0006
"(24.0, 27.0]"	0.4112	0.2645	0.2645	0.0592	0.0005
"(27.0, 32.0]"	0.2697	0.1747	0.3330	0.2063	0.0164
"(32.0, 41.5]"	0.1967	0.0332	0.2131	0.3275	0.2294
"(41.5, 81.0]"	0.1214	0.0156	0.0458	0.3330	0.4841

Figure 13: Visual representation of the nodal relaltionship between Diabetes and Age. (Moreira et al., 2020b).

6.3 Diabetes Bayesian Network Inferences

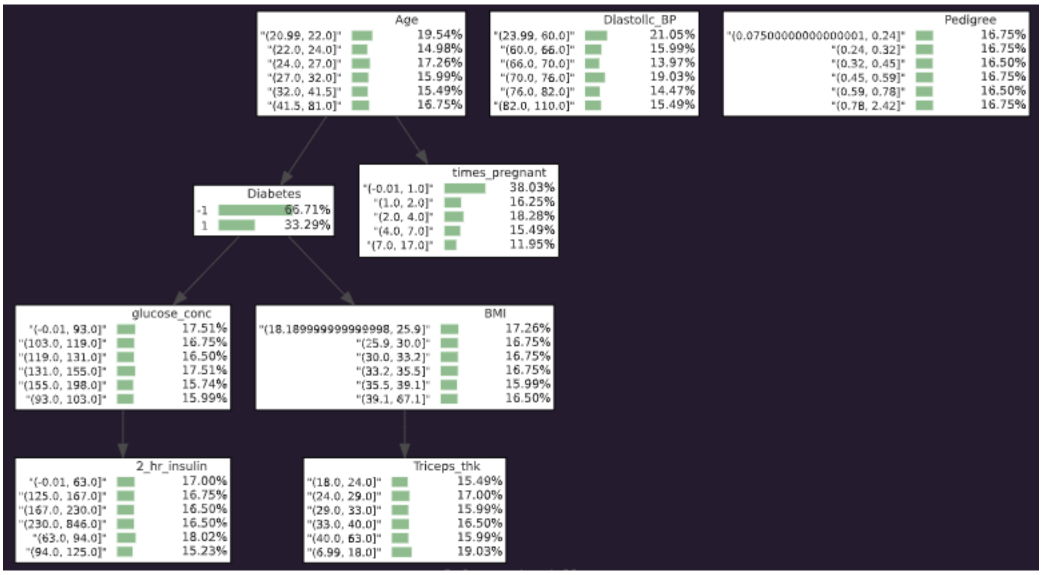

Figure 14: Computed inferences of the Diabetes dataset. (Moreira et al., 2020b).

6.4 Visualising Quantum States of Diabetes and BMI Relationship

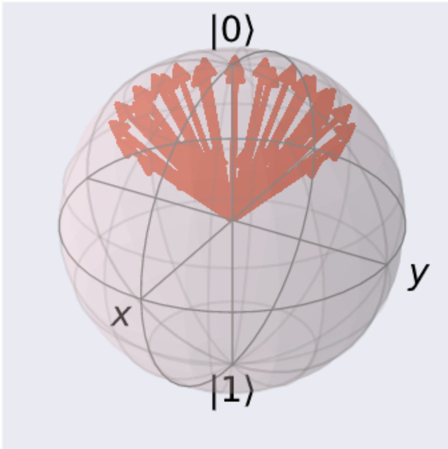

Figure 15: Computed Diabetes Quantum States. (Moreira et al., 2020b).

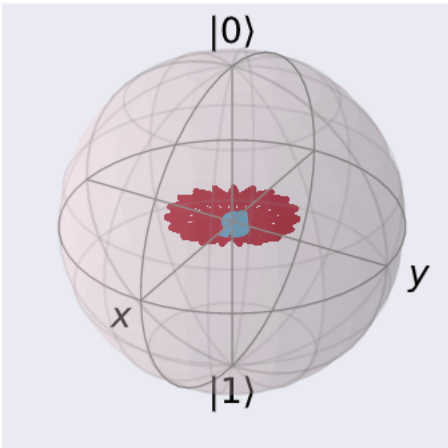

Figure 16: Computed BMI Quantum States. (Moreira et al., 2020b).

6.5 Computing Density Operator

```
#Compute the Full Joint Distribution
classical_joint = computeFullJoint(bn, var="Diabetes")
#Compute the superposition states
quantum_superposition = createSuperpositionState( classical_joint )
gnb.sideBySide(classical_joint, quantum_superposition[:], captions=["Classical Full Joint", "Quantum Superposition"])
```

Figure 17: Computed Classical Joint Distribution. (Moreira et al., 2020b).

```python
def my_density( superposition ):
  density = np.zeros((2,2), dtype=complex)
  for i in range(0, 2):
    for j in range(0, 2):

      real = superposition[:][0][i].real * superposition[:][0][j].real
      imaginary = superposition[:][0][i].imag + conj(superposition[:][0][j]).imag

      density[i][j] = complex(real, imaginary)

  return Qobj(density)
```

Figure 18: Density operator computation. (Moreira et al., 2020b).

```python
rho = my_density(quantum_superposition)
gnb.sideBySide(classical_joint,
               pd.DataFrame(rho[:], columns=["Diabetes", "BMI"]),
               pd.DataFrame(np.expand_dims(np.real(np.diag(rho[:])),0), columns=["Diabetes,BMI"]),
               captions=["Classical Full Joint Distribution",
                         "Quantum Density Operator",
                         "Classical Full Joint Distribution from Density Operator"])
```

Figure 19: Computed Classical Full Joint and Quantum Density Operator. (Moreira et al., 2020b).

6.6 Computing the Marginal Distribution and Quantum Partial Traces

```python
var_name = "Diabetes"

# classical inference
ie=gum.LazyPropagation(bn)
ie.makeInference()

quantum_marginal_distr = computePartialTrace( rho[:], [0,1], axis = bn.idFromName(var_name) )
classical_marginal_distr = ie.posterior( var_name )

gnb.sideBySide(classical_marginal_distr,
               pd.DataFrame(quantum_marginal_distr[:], columns=["Diabetes", "BMI"]),
               pd.DataFrame(np.expand_dims(np.diag(quantum_marginal_distr[:]),0), columns=["Diabetes", "BMI"]),
               captions=["Classical Marginal Distribution",
                         "Quantum Partial Trace over %s" %(var_name),
                         "Classical Marginal Distribution from partial trace"])
```

Figure 20: Classical Marginal Distribution and Quantum Partial Trace of the Diabetes dataset. (Moreira et al., 2020b).

7 DISCUSSION

7.1 Using QLBNs for Diabetes Dataset

The idea of this project was to optimise the QULBIT framework so that it can be applied to a real world dataset from the UCI repository, which in this case, the diabetes dataset was used. The following steps were necessary to achieve the above results for the Diabetes dataset: optimising the functions for computing the density matrix and partial trace, setting all the interference terms such that I can get the minimum possible interference and retrieve the posterior probabilities. Another experiment that was conducted was to optimise the functions for computing the density matrix and partial trace, setting the interference terms such that the maximum interference is possible and then retrieve all the posterior possibilities.

The QLBN framework managed to discretize the data in such a way that it can be used for processing the relevant BN as seen in Figure 11. The diabetes dataset had many data points and we will see later on how this will play an integral role in run-time. The BN model in Figure 12 as depicted demonstrates the associations between age, diabetes and the number of times pregnant. We are most interested in the relationship between diabetes and BMI as this is what is going to be modelled reflectively using the QLBN framework. The visualisation of the data is then furthermore explored with the Figure 13 displaying the relationships of people of certain ages who are more likely to get diabetes and comparatively, the amount of times someone has fallen pregnant with relationship to age.

Calculating the inferences of the dataset was the next step to further the dissection of the data. In Figure 14, the representation of all the nodes with the corresponding inferences is shown and it is obvious that there is no strong correlation between diabetes and the other attributing factors for this dataset. The classical joint distribution is the result of calculating the inference terms as they determine the probability distribution of one or more of the variables given other variables. From using this technique it is obvious that there is a low chance of having diabetes given the relationships of the nodes that consist of BMI and glucose concentration. This can be attributed to conditional independence of the nodes given the Markov blanket that it represents.

Quantum states are depicted by the random variables that make up the network which are represented. This means that these variables are represented as a probability distribution of complex probability amplitudes over the different outcomes that the random variable can have, as opposed to being specified as a real number. This can be seen in traditional BNs, however, in QLBNs, one needs to distinguish the two types of quantum states which demonstrate the root nodes (1) and the child nodes (2). Child nodes are represented as statistical distributions of quantum states, which means that the child is represented as an ensemble where the condition probabilities can be found in different quantum states in the ensemble. When analysing Figure 15, the quantum states depict the pure state of the configurations for a given ket function according to the relevant phases. Comparatively, Figure 16 demonstrates the ensemble states from the quantum states representing configurations of a ket function phase according to its parent ket function phase.

Calculating the superposition and the density operators was done by computing the densities relevant to the real and imaginary numbers associated with the quantum bayesian network model. As it is known, the relevant interference terms of the calculated density matrix will exhibit the quantum states which are interferring with the classical distribution. These are the terms that the QLBN framework are after as this will help calculate the uncertainties that are not accounted for using the classical BN calculations. In quantum thery, all individual quantum states contained in a Hilbert Space are defined by a superposition state which is represented by a quantum state vector, comprising of the occurrence of all events of the system. The density operator uses the superposition terms in a way to describe the system where one can compute the probabilities of finding each state in a network. The density operator in this example uses the Hermitian matrix where we account for the number of quantum states in the network that contains the classical full joint distribution. The sum of the diagonal of the elements in this matrix are the classical terms that are used.

The density operator also contains quantum interference terms in the off-diagonal elements, which are the core of this model. It is precisely through these off-diagonal elements that, during the inference process, one is able to obtain quantum interference effects, and consequently, deviations from normative probabilistic inferences. This is what is calculated and computed in Figure 19.

Realistically speaking, computing the probabilities is not achievable without computing the probabilities using the partial trace methodology. This is possible during the inference process which isolates the subgroup of quantum systems from larger systems. It is necessary sometimes to trace out a subgroup of quantum states from a larger group which in many cases is provided by two different quantum states. This is done by using a partial trace function which can be seen in Figure 20 where we convert the complex probability amplitudes computed in the partial trace operation into real probability values. Independent on the query of how it is done, the operation is reliant on the query made by the decision maker, so we use an operator in the code to select all entries of the density operator that satisfy the assignment of the specified query. This is done in the same way as a full joint distribution has queries selected which allows for the operator to be defined as a column vector with zeros in all positions. This allowed for the acquiring of maximum and minimums for the expected datasets.

Since this is a fixed basis in our system, the partial trace over the density operator is simplified by using the computePartialTrace function, where the summation is executed over all possible combinations of values of the unobserved values. Since BN use urinary tables where the probability is conditional attributed to double stochasticity, the final scores must be normalised in order for the data to be presentable. Double stochasticity implies that of a square matrix which requires each row and each column of non-negative real numbers to add up to one. Expanding this will result in the marginalisation of the dataset which has been already calculated using the classicalmarginaldistr function. After this has been computed, finally it can be seen that the interference terms cancel and the QLBN collapses into its classical counterpart. This means that the QLBN framework is an abstracted model of the classical BN model since it represents both classical and quantum behaviours.

To summarise this project, the diabetes dataset has been successfully used to demonstrate the differences between the QLBN and BN models utilising specific computational techniques and optimisation in order to achieve the relevant interference terms which ideally create the utility in the decision-making space for this specific dataset. Although most of the operations between the QLBN and classical BN is concerned with similar use cases for mapping BNs, the underlying factor is that joint distributions and marginalisation are computed quite differently using the two different models. By representing the possible states of the diabetes dataset as complex quantum probability amplitudes in a complex Hilbert Space, the full joint distribution of the network using a density matrix operator was achieved. The main diagonal elements of this matrix correspond to the classical elements and the off diagonal elements correspond to the quantum elements, where it will enable deviations from normative probabilistic inferences which are obtained in traditional bayesian network models. The QLBN framework and its quantum interference effects are integral and are the result of the interference caused by the quantum wave function. As it is known in other relevant literature, the Newtonian world that most activities take place have growing evidence to show for quantum effects that can manifest in the macroscopic realms.

8 APPENDIX

GitHub link: https://github.com/player1011/QULBIT-Thesis-Project

References

Bergus R, G., Levin P, I., and Elstein S, A. (2002). Presenting risks and benefits to patients. *Journal of General Internal Medicine*, 17:612–617.

Buchanan, B. (1984). Rule based expert systems. *The MYCIN Experiments of the Stanford Heuristic Programming Project*.

Giger, M. L. (2018). Machine learning in medical imaging. *Journal of the American College of Radiology*, 15(3, Part B):512 – 520. Data Science: Big Data Machine Learning and Artificial Intelligence.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., and Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9:e1312.

Kahneman, D. and Tversky, A. (1982”). Variants of uncertainty. *Cognition*, 11(2):143 – 157.

Kumar, V., Tan, P.-N., and Steinbach, M. (2006). Introduction to data mining. *Instructor's Solution Manual*, pages 160–169.

Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A., and Tourassi, G. D. (2008). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural Networks*, 21(2-3):427–436.

Moreira, C., Hammes, M., Kurdoglu, R. S., and Bruza, P. (2020a). Qulbit: Quantum-like bayesian inference technologies for cognition and decision. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Moreira, C., Tiwari, P., Pandey, H. M., Bruza, P., and Wichert, A. (2020b). Quantum-like influence diagrams for decision-making. *Neural Networks*.

Moreira, C. and Wichert, A. (2016). Quantum-like bayesian networks for modeling decision making. *Frontiers in Psychology*, 7:11.

Moreira., C. A. P. (2017). Quantum probabilistic graphical models for cognition and decision. *Science*, 1:236.

Saposnik, G., Redelmeier, D., Ruff, C. C., and Tobler, P. N. (2016). Cognitive biases associated with medical decisions: a systematic review. *BMC Medical Informatics and Decision-Making*, 16.

Trueblood, J. S., Pothos, E. M., and Busemeyer, J. R. (2014). Quantum probability theory as a common framework for reasoning and similarity. *Frontiers in Psychology*, 5:322.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–1131.

Wang, Z., Busemeyer, J. R., Atmanspacher, H., and Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science*, 5(4):672–688.

Winsberg, F., Elkin, M., Macy, J., Bordaz, V., and Weymouth, W. (1967). Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. *Radiology*, 89(2):211–215.

Yanase, J. and Triantaphyllou, E. (2019). A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Systems with Applications*, 138:112821.