An Enhanced Approach for the Prioritisation in Patent Ductus Arteriosus (PDA) Services Using Data Mining Techniques

Noura Alzeer, Hessah A. Alsalamah, Alaaeldin Hafez
College of Computer and Information Sciences, King Saud University
Riyadh, Saudi Arabia
e-mail: noura.alzeer@gmail.com, {halsalamah, hafez}@ksu.edu.sa

Abstract— **The past few decades have witnessed a huge increase in the number of heart disease patients globally. An estimated 17.7 million people died from Cardio-Vascular Diseases (CVDs) in 2015, representing 31% of all global deaths. Saudi Arabia is no exception when it comes to the growing number of CVDs patients. The most common CVDs in Saudi Arabia is Patent Ductus Arteriosus (PDA), which is a persistent opening between two major blood vessels leading from the heart. Unlike many other fields of medicine, PDA is time-related and cannot be delayed. Thus, there is a need to have a Decision Support System (DSS) that can analyze the crucial statistics in deciding the level of urgency for each child that requires PDA surgery. This research proposes an approach to classify the PDA patients according to their urgency level based on their current conditions and the results of scans and tests they have undergone. An enhanced approach, which uses Data Mining (DM) techniques, is proposed to manage waiting lists and determine the priority among patients, as well as diagnosis of certain diseases. This approach aims to support early prevention and intervention to increase life expectancy and well-being of PDA patients. Our proposed approach achieved an accuracy of 99.15%.**

Keywords- Data Mining (DM); Healthcare; Patent Ductus Arteriosus (PDA); Waiting Lists Management.

I. INTRODUCTION

The alarming growth in pediatric and adult CVDs is now a major threat to world health [2]. The causes of CVDs include, but are not limited to diet through bad eating habits, obesity, and modern day lifestyle included stress. Other reasons may include general heart disease which can be hereditary, age and routine life pressure which can all be contributory to CVDs. There are at least eighteen distinct types of congenital heart defects that are recognized, with many additional anatomic variations [3].

This study focuses on one type, called PDA. In pre-term babies, the closure of the Ductus Arteriosus (DA) may be delayed after birth and is influenced by factors, such as gestation and development at birth, Respiratory Distress Syndrome (RDS), artificial mechanical ventilation, infection and lack of antenatal steroid treatment. These risk factors are more common amongst babies who require intensive care [2]. Persistent patency of the DA is therefore more common in pre-term infants; in over 50-60% of cases where babies are born at less than 29 weeks' gestation, the DA does not naturally close. This is known as a PDA [3].

In Saudi Arabia, physicians undertake a detailed medical examination of all pre-term infants, including examining the infant specifically for any PDA symptoms [4]. If any such symptoms are discovered, the newborn infant will receive immediate treatment. However, in full-term healthy babies, conditions, such as PDA are rarely checked. It is only if physicians or parents notice any PDA symptoms or if the child's non-diagnosed condition becomes critical to the point when it require prompt medical intervention that any treatment for PDA is given. Because of the increasing numbers of PDA patients in Saudi Arabia, it is important to raise the awareness of this disease and to educate parents, explaining the symptoms of this disease to ensure that patients are diagnosed as early as possible and thus receive the proper treatment as early as possible. It is also important for hospitals, once a patient is diagnosed with PDA, to ensure the patients have access to treatment in a timely manner. DM techniques, especially classification, have been used in many different medical fields, such as managing waiting lists and determining priorities among patients, as well as the diagnosis of certain diseases. In this research, we propose a hybrid approach of classification, clustering and association rules to determine not only life and death cases, but also cases which are both urgent and not urgent, and the urgency level within each category using the association rules. A DSS is used when scheduling patients or pre-planning a date for their scheduled treatment, procedure or surgery, and it is believed to have a positive impact on the current process.

The paper is organized as follows. Section II explains the importance of this work. Section III describes the theoretical background. Section IV presents related works to this research. Section V outlines the proposed research methodology, followed by Section VI, which shows how DM is used to build a model to diagnose PDA patients according to their urgency level. Section VII discusses the results obtained from the proposed model. Finally, the conclusion and future work of this research are found in Section VIII.

II. IMPORTANCE OF THE STUDY

Many guidelines have been published to determine the importance of diagnosing the urgency of PDA. This urgency can be summarized into two categories; life and death. Any delay in diagnosing the urgency could have a

 ISBN: 978-1-61208-618-7

tragic fatal outcome, but early diagnosis and intervention can result in the increased life expectancy and well-being of the patient.

This study uses DM techniques to study and analyze PDA patients' records collected from King Fahd Medical City (KFMC) in Riyadh, Saudi Arabia, to build a new standard model to diagnose PDA and determine the urgency levels of PDA patients, depending on their existing medical condition, via results of scans and other tests carried out. It also presents recommendations for testing newborn babies for any PDA symptoms. This will help parents to receive the appropriate timely treatment for PDA for their infant before the condition worsens and requires surgery. The model is tested on new patients and its results are compared to those results provided by the physicians at KFMC to determine its accuracy. The model is also presented to further specialist heart physicians from different hospitals in Saudi Arabia to be tested. Their comments and recommendations are used to modify the original model to reach its final result. Having a DSS, which uses a standard model to provide an accurate urgency level for each PDA case, will save patients and their family time and money instead of trying to obtain various consultations from different physicians in different hospitals. Ultimately, there is also the cost-saving factor to the physician and hospital to be considered in reducing multiple appointments, as well as reducing multiple tests. The mere reduction in appointments alone opens up the availability of allocating an appointment to another patient. Overall, the hospital will also save large sums of money from the reduction of multiple examinations, tests and procedures, some of which are unnecessary as they may well be repetitive and thus not conducive to the well-being of the patient.

III. Related Work

Extensive research has been conducted studying the long waiting list problems, especially within medical fields. Some research focused on how to arrange priorities of these waiting lists and others dealt with how to reduce the general waiting times. Either way, what is definitive is that, generally, there is far greater demand over supply that results in waiting times and more extensive waiting lists.

A recent study [5] used statistical techniques to compare waiting lists for hip and knee arthroplasty at Groote Schuur Hospital in Cape Town in South Africa with the waiting lists kept by the surgeons at that hospital. The results showed that the hospital's waiting list is inaccurate because of the poor data management; for example, in many cases, the list did not have the current contact details of the patients. It also identified unfairness of waiting times among the patients. The research recommended the use of a scoring-based prioritization system which uses clinical diagnosis and tests results, radiographic and societal parameters to manage the waiting lists fairly.

A significant amount of research focused on using DM techniques in general in the healthcare sector [6]-[9] showed the benefits of applying different types of DM techniques in disease diagnosis and finding side effects of certain medicines, for example. However, they did not focus on one type of disease or on the type of DM techniques. Other studies used one or more DM techniques to predict or diagnose one type of disease; for example, recent research [10] used Naïve Bayes and WAC (Weighted Associated Classifier) to predict whether a patient has a risk of heart diseases or not. The study used factors, such as age, sex, whether diabetic, height, weight, blood pressure, cholesterol, fasting blood sugar and hypertension to predict the possibility of having heart diseases. In another study, the focus was on using DM techniques in oral cancer patients [11] and the authors showed that all the models built for predicting the survivability of oral cancer patients show similar results and performance. However, they did prove that the TreeBoost model is slightly better than the others as all 18 predictors they used in their research are considered for each spit. They showed the experimental results of probability adjustment, threshold analysis and lift-gain are also slightly better in the TreeBoost model. Therefore, they concluded that the TreeBoost classification model is an effective system for determining the survival rate of oral cancer patients. Rogers and Joyner [12] shows a special type of DM technique, called SAS software (Statistical Analysis Software), that can be utilized as a solution for the increasing costs in the healthcare sector and to improve the quality of service and care provided by hospitals and healthcare centers. The authors of the study discussed the DM methodology used in SAS software and applied it when studying the factors affecting the healthcare industry. It was shown that it can be used to solve serious and critical problems in the healthcare sector. This research is an example of applying types of DM (Statistical Analysis in this case) on medical data sets. Finally, in a closely-related study to this research topic [13], the authors submitted a survey of current techniques of knowledge discovery in databases, using DM techniques that are used in today's medical research, particularly in Heart Disease Prediction. When predicting a heart attack, 15 characteristics are listed and with basic DM technique, other approaches, such as Artificial Neural Network (ANN), Time Series, Clustering and Association Rules and soft competing approaches can also be combined.

This research focuses on using the basic information about the patient, such as age and results of scans, to provide an initial diagnosis of PDA. This helps to predict PDA cases as early as possible, especially in small towns which are far from large hospitals in major cities. It helps hospitals schedule urgent cases of PDA in waiting lists according to their urgency level.

IV. Theoretical Background

In this study, DM techniques are used to diagnose PDA and to arrange patients according to their urgency level. To gain a better understanding of this study, it is important to

 ISBN: 978-1-61208-618-7

first understand what PDA is, what DM techniques are and how they work.

A. Patent Ductus Arteriosus (PDA)

PDA is "a persistent opening between two major blood vessels leading from the heart. The opening, called the ductus arteriosus, is a normal part of a baby's circulatory system before birth that usually closes shortly after birth. If it remains open, however, it is called a patent ductus arteriosus." [14]. This opening allows a great deal of blood to enter the lungs and heart, causing extra blood pressure on the lungs and enlarging and weakening the heart. There are no clear causes for this disease, but genetics may play a role. A defect in one or more gene could prevent the DA from closing normally after birth [14]. Symptoms of PDA depend on the size of the defect. While small PDA can have no signs sometimes, large PDA can be detected as a result of different causes. These include poor eating habits and consequently very slow growth, tiring and sweating during crying or eating, a rapid heart-beat and fast breathing [14]. Treatment for PDA can vary from medication or catheter-based procedures to surgery, depending on the age of the child and the size of the PDA [14] [15]. As PDA patient numbers increase, the pressure on hospitals to provide urgent care grows, whether this be interventional care or damage repair. Unlike other fields of medicine, in PDA, time is of essence and intervention cannot be delayed. Thus, the need to create a system that can speedily analyze the statistics is crucial in determining the level of urgency for each child that requires PDA-related surgery.

B. Data Mining (DM)

DM is a process, which finds useful patterns from large quantities of data [16]. It is the core step of the knowledge discovery process [16]. These steps are:

- Data selection: This is the selection of applicable data and records from the one or more databases applicable to the relevant study.
- Data preprocessing: This step involves the removal of irrelevant items and data entries that are not connected to the work environment. Data preprocessing allows the transforming of the original data into a suitable shape to be used by a mining algorithm, before starting the process of DM algorithm analysis [17].
- Data transformation and enrichment process: This step contains the calculating of the new characteristics from the existing characteristics.
- Data Mining: The application of DM techniques to the data to obtain new patterns and models.

DM can be carried out using numerous methods, depending on the type of data used and the result required from DM. In this study, the focus will be on three DM techniques: Classification, Clustering and Association Rule.

1) Classification:

As the main process [18] that is included in DM systems, this maps the various types of data into well-defined subgroups and classes. Another name for the classification process is Supervised Learning. Supervised Learning consists of two main stages; the first phase is a model construction and the second is model usage. Model construction: This phase contains the operations of building the used model in the structures of the attributes.
Model usage: All the operations that are used in utilizing the model.

2) Clustering:

Clustering is a process of grouping physical or abstract objects into classes of similar objects. Clustering and Classification are both grouping methods. Clustering is an unsupervised classification, whereas Classification is a supervised grouping. Classification and Prediction are also related techniques. Classification predicts class labels [19] whereas Prediction predicts continuous-valued functions. The major clustering algorithms are:

- Partitioning algorithms: Construct various partitions and then evaluate them according to certain criteria.
- Hierarchy algorithms: Create a hierarchical decomposition of the set of data (or objects) using certain criteria.
- Density-based: Based on connectivity and density functions.
- Grid-based: Based on a multiple-level granularity structure.
- Model-based: A model is hypothesized for each of the clusters and the idea is to find the best fit of that model for each other.

In this research, a portioning technique called K-Means is used [19]. This works with numerical values only, which suit the numerical measurements of the PDA scans.

3) Association Rule Mining (ARM):

The concept of ARM is considered to be one of the main studies in DM methods. Such rules associate one or more attributes of a dataset with another attribute, producing an if–then statement concerning attribute values. The original problem is the market basket analysis which tries to find all the interesting relations between the bought products. Sequential pattern mining attempts to find inter-session patterns, such as the presence of a set of items followed by another item in a time-ordered set of sessions or episodes. The association rule is an implication expression of the form $X \rightarrow Y$, where X and Y are disjoint item sets. The strength of the association rule can be measured by its support and confidence. Support determines how often a rule is applicable to a given dataset, while confidence determines how frequently items in Y appear in transactions that contain X [20].

 ISBN: 978-1-61208-618-7

V. RESEARCH METHODOLOGY

In this research, DM techniques are used to solve the problem of diagnosing and determining the urgency level of PDA patients. DM can assist in classifying patients according to the urgency of their cases in order to recommend the correct course of treatment, or pre-plan a date for their scheduled treatment, procedure or surgery. Within this study, data relating to PDA patients will be gathered from patients' records at KFMC, including their test results, image scans, medical history and physicians' diagnosis. This data will be stored in tables after being checked and cleansed to eliminate any incomplete records and to transform the data into numeric values or classify it into categories to prepare it for the application of different DM techniques.

Meetings with cardiology physicians at KFMC were initially conducted to discuss their approach to determine the urgency level for each PDA case. Physicians explained the main tests they carry out and scanned the images PDA patients have undertaken. Physicians provided a detailed explanation for each test and scan result, whilst showing the circumstances under which these results are considered normal, abnormal or critical. This helped gain an initial understanding of exactly how the urgency of PDA patient's urgency level is determined. Afterwards, DM techniques were applied to this data to find one standard model to classify patients according to their urgency level. In this research, Waikato Environment for Knowledge Analysis (WEKA) application was used to apply DM techniques. WEKA contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization [21], which facilitates the application of different DM techniques on the data and compares the results in order to understand the relationships between the patients' attributes (details such as age, weight, height and test results) in order to obtain the most accurate model to classify patients according to the urgency level. Following this, testing was carried out on new patients using the proposed model. The results were compared to the results carried out by cardiology consultant physicians to ascertain its effectiveness and success. In case the proposed model yields inaccurate results, it was automated into a useful and easy to use tool. Otherwise, DM techniques were applied again to obtain an improved model, which was then tested again.

The proposed DSS can help in classifying the urgency level of PDA patients as speedily and efficiently as possible based on their test results, sometimes without the need to be examined, at this stage, by a physician. The measures to test the effectiveness of the system are based on two aspects:

- Time needed to determine the urgency level for a specific PDA case. Currently, physicians at KFMC meet for 30-60 minutes to consult on a single case and make a decision. Therefore, the new suggested tool should allow for considerably less time.
- Accuracy of the urgency level suggested by the tool. The tool will be used to determine the urgency level for new PDA patients. The consultant physicians at the hospital determine its accuracy and check the results given by the tool.

VI. USING DM TO BUILD MODEL TO DIAGNOSE PDA PATIENTS

In this research, DM techniques are applied on PDA patients' records.

A. Applying Classification on PDA Patient Records

After collecting PDA patients' records from KFMC, the data was stored in an Excel file in Comma Separated Values (CSV) format, which can be used in WEKA. The records included a copy of ultrasound scans photos. The details about the PDA dimensions are kept in a separate database, which is accessed, only by the physicians. The patient's record stores the PDA size as small, moderate or large. Attributes of the records are explained in Table 1.

After applying classification using decision tree classification, the tree in Fig. 1 was generated. The correctly classified instances are: 234, whilst the incorrectly classified are only 2 records, giving this model a very high accuracy percentage of 99.15%.

The main attribute for classifying the patients in the tree is the current age.

TABLE I. ATTRIBUTES IN PDA PATIENTS' RECORDS FROM KFMC

Attribute Name	Type	Values
Sex	Nominal	F or M
Current age	Numerical	Integers from 0 for new born
Height	Numerical	Integers
Weight	Numerical	Decimal numbers
Heart rate	Numerical	Integers
PDA last size	Nominal	**Non**: the patient does not have previous scan **Small, Mod, Large**
PDA current size	Nominal	**Small, Mod, Large**
Change	Nominal	**Yes** **No** **New**: if the patient does not have previous scan
Period to monitor	Numerical	Integers
Decision	Nominal	**Wait**: the patient gets medical treatment and waits for another scan after 3 or more months **Urgent**: if the patient needs Surgery or Catheter Procedure

 ISBN: 978-1-61208-618-7

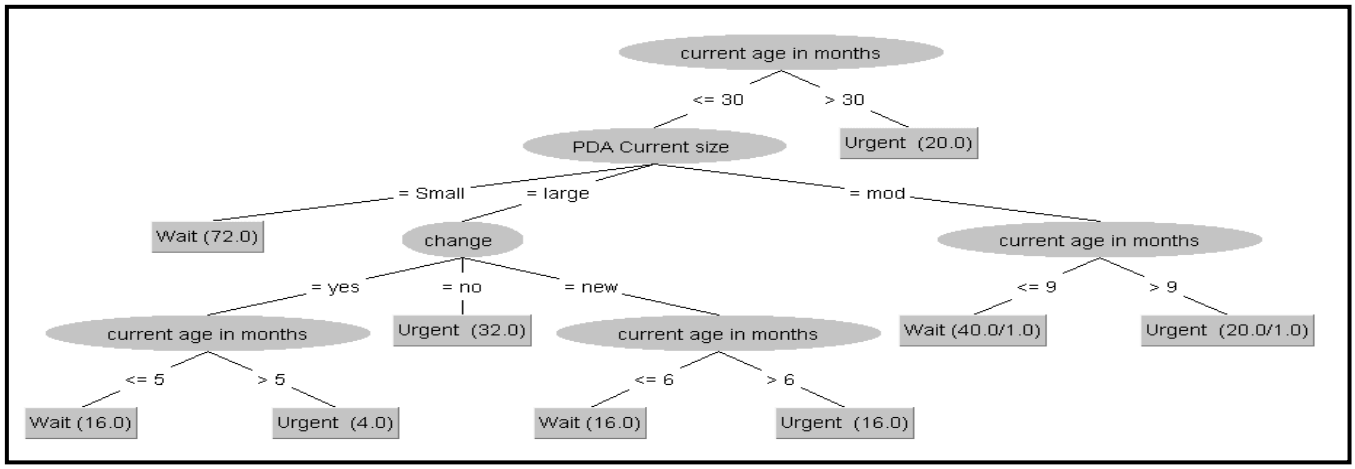

Figure 1. Decision Tree resulting from applying classification on PDA patients' records.

1. If the age of the patient is older than 30 months, then it is an Urgent case. This is applied in 20 cases.
2. If the child's age is less than, or equal to, 30 months and the current size of PDA is small then the patient receives medical treatment and waits for 3 months for a further scan. This is applied in 72 cases.
3. If the age of the patient is less than, or equal to, 30 months of age and the current size of PDA is moderate then there are two cases depending on the current age:
    - If the age is less than, or equal to, 9 months then the patient receives medical treatment and waits 3 months for another scan. This is applied in 40 cases, with the exception of one case where the decision was revealed as Urgent.
    - If the age is more than 9 months, then the patient case is urgent. This is applied in 20 cases, except for one case whereby the decision was to receive medical treatment and wait for 3 months for a further scan.
4. If the age is less than, or equal to, 30 months and the current size of PDA is large then there are 3 cases depending on whether there was a change on the PDA size.
    - No change: the case is classified as Urgent. This is applied in 32 cases
    - There was change: there are 2 cases based on the age:
        i. Age less than, or equal to, 5 months: receive medical treatment and wait 3 months for new scan. This is applied in 16 records
        ii. Age is greater than 5 months: it is an Urgent case. This is applied in 4 cases.
    - The change = New, which denotes a new patient with no previous record. There are 2 cases based on the current age:
        i. Age less than, or equal to, 6 months: receive medical treatment and wait 3 months for new scan. This is applied in 16 cases.
        ii. Age is greater than 6 months: it is an Urgent case. This is applied in 16 cases.

The previous model uses the PDA size and patient age to provide an accurate initial diagnosis of PDA.

B. Applying Clustering on PDA Patient Records

One of the clustering techniques provided by WEKA is called Simple K-Means [21]. WEKA uses the mean of the attributes to group the records into K groups or clusters, such that the records in each group are closer to each other (similar to each other) than they are to records in other groups. Clustering works better with numerical attributes than with nominal attributes. The records of PDA patients collected from KFMC have both numerical attributes, such as, age, height and weight, and nominal attributes, such as, PDA size, and gender. In order to apply a Simple K-Means algorithm, nominal attributes should be changed into numerical attributes.

The specialists from KFMC who were consulted for this study explained that gender is not a major attribute in deciding the urgency level of PDA patients. Research and publications of study articles related to PDA treatments also suggest that although PDA is more common in girls than boys, gender does not affect the diagnosis of the disease. For these reasons, sex attribute is not used in the clustering

 ISBN: 978-1-61208-618-7

technique. The other nominal attribute is PDA size, and it is replaced with the average size as follows:

- Small less than 1.5mm. Therefore, it is replaced by 1.25.
- Moderate between 1.5mm and 2mm, so it is replaced by 1.75.
- Large greater than 2mm, so it is replaced by 2.25.

Applying simple K-Means making K = 2, which means the results will give 2 clusters, one for patients who receive treatment and wait for 3 months to have another scan and a class for patients who need urgent treatment. The two clusters are shown in Table 2.

TABLE II. RESULTS OF APPLYING CLUSTERING ON PDA PATIENTS' RECORDS

Attribute	**Cluster #1**	**Cluster #2**
Current age in months	15.5109	7.0625
Height	71.9565	62.7451
Weight	10.6685	6.9181
Heart rate	115	108.3333
PDA last size	1.2391	1.125
PDA current size	1.9457	1.6111
Decision	**Urgent**	**Wait**
Number of records	92	144

The table shows that:

- Older patients (average 15.5 months) are more likely to be classified as urgent, while younger patients (average 7.1 months) receive treatment and remain under observation.
- As expected, older patients have higher height and higher weight, so height and weight averages are related to the same clusters as age.
- Higher heart rate was related to Urgent cases while lower heart rate is related to patients who receive treatment to remain under observation.
- The important attribute here is the current size of PDA: average 1.9 (which means moderate or large PDA) means Urgent case, while smaller PDA with average 1.6 is not urgent.

These results comply with the results obtained from the decision tree used in the classification.

C. Applying Association Rule on PDA patient records

Association rule technique is applied to those patients whose cases are classified as "Urgent" to determine the urgency level. Doctors determine the urgency level of the patient based on many factors, such as age, size of PDA, tests and ultrasound scans. There are three levels of urgency:

- Level1: needs immediate surgery.
- Level2: needs surgery within a month or two.
- Level3: needs to undergo more tests and scans to determine a time for surgery.

To apply the association rule, all the attributes should be nominal, so the patient's age is transformed into age group. Also, the heart rate was transformed from numbers into nominal: low, normal, and high. The height and weight were also transformed into nominal values.

- Group a: less than 6 months
- Group b: between 6-12 months
- Group c: between 12-36 months
- Group d: older than 36 months

These ranges are based on the physicians' suggestion at KFMC. After applying association rule on the patients' records, the following rules were obtained:

- PDA Current size =large 103
  ==> urgency level=level1 103
- Age group=a 95
  ==> urgency level=level1 95
- Age group=a PDA Current size =large 39
  ==> urgency level=level1 39
- Age group=b PDA Current size =large 30
  ==> urgency level=level1 30
- Age group=b PDA Current size =mod 12
  ==> urgency level=level2 12
- Age group=c PDA Current size =mod 51
  ==> urgency level=level1 51
- Age group=c PDA Current size =large 45
  ==> urgency level=level1 45
- Age group=c PDA Current size =small 64
  ==> urgency level=level3 64
- Age group=d 72
  ==> urgency level=level1 72
- Age group=a change =no 39
  ==> PDA Current size =large 39

All the above rules have a confidence = 1, which means they are 100% true based on the PDA patients' records from KFMC. Table 3 shows a summary of these rules.

TABLE III. RESULTS OF APPLYING ASSOCIATION RULE DM ON THE URGENT CASES OF PDA PATIENTS

Urgency level	**Cases**
Level1	PDA Size = Large or Age is less than 6 months or Age between 12-36 months and PDA size = Mod or Age older than 36 months
Level2	Age between 6-12 months and PDA size = Mod
Level3	Age between 12-36 months and PDA = small

VII. RESULTS

DM classification and clustering techniques were used to diagnose PDA and classify patients into two categories: urgent cases and non-urgent, which are referred to here by Wait cases where the patient receives medical treatment and waits for another check-up after 3 or more months. Urgent cases were then arranged into three levels of urgency using association rule DM techniques. The models presented by this research were tested by cardiology specialists at KFMC who subsequently approved them. They suggested using age

 ISBN: 978-1-61208-618-7

range instead of actual age in months to make the diagnosis more general. The suggested age groups are:
a: 0-6 months / b: 6-12 months / c: 12-36 month / d: older than 36 months
After applying the specialists' recommendations, the resulting model is shown in Table 4.

TABLE IV. THE RESULT OF APPLYING DM TECHNIQUES ON PDA PATIENTS' RECORDS

Case Description	Diagnosis	Urgency level
Age group = a, b & PDA current size = Small	Wait	
Age group = a & PDA current size = Large & Change = yes	Wait	
Age group = b, c & PDA current size = Large & Change = yes	Urgent	Level1
Age group = a, b, c & PDA current size = Large & Change = no	Urgent	Level1
Age group = a & PDA current size = Large & Change = new	Wait	
Age group = b, c & PDA current size = Large & Change = new	Urgent	Level1
Age group = a & PDA current size = Mod	Wait	
Age group = b & PDA current size = Mod	Urgent	Level2
Age group = c & PDA current size = Mod	Urgent	Level1
Age group = c & PDA current size = Small	Urgent	Level3
Age group = d	Urgent	Level1

VIII. CONCLUSION AND FUTURE WORK

This study presented a hybrid approach to diagnose PDA as early as possible using simple tools. The introduction of an ultrasound scan that is readily available in almost all medical centers in Saudi Arabia is one method of detection and diagnosis. This study also recommends increasing the awareness of PDA amongst first time parents, as well as parents with more than one child, as early as possible, to ensure that they are educated to understand what PDA is and to seek medical help as soon as they notice any symptoms in their baby, such as shortness of breath, sweating while feeding, struggling to breathe and nurse at the same time or quite simply being too attached to nursing. To a parent, some of these symptoms may not appear to be any obvious cause for alarm, but with prior knowledge of PDA, early diagnosis and intervention could quite well positively impact on health and life expectancy. Ultimately, the goal is to produce a longer living and healthier Saudi population.

It is expected that when increasing the awareness of PDA, this will, in turn, increase the number of PDA patients in Saudi Arabia. The result will create a strain on already struggling hospitals and also substantially increase the number of patients needing to go onto a waiting list to see a consultant as well as those patients who have already received a diagnosis but require ongoing medical treatment to prevent surgery, as well as to maintain their health. Overall, the burgeoning hospital outpatient clinics for those patients with PDA will need to be overhauled to ensure that all patients are attended to in timely and appropriate manner.

While working on this research, we faced many difficulties. For example, the details of the scans, such as the dimensions of the PDA, were not saved in the patients' records. This information was saved in special files which can be accessed by the physicians only. This made collecting the information needed for the patients extremely difficult. It was not easy to meet the specialists at KFMC for any length of time as they were tremendously busy all the time.

Regarding future work, further study and analysis should be carried out on more patients from different hospitals to obtain a more accurate model to diagnose PDA and determine its urgency level. Additional data should be added to the patients' records to show all the detailed measurements taken from the ultrasound scan.

ACKNOWLEDGMENT

The authors extend their deep appreciation to the Deanship of Scientific Research at King Saud University for funding this work.

REFERENCES

[1] World Health Organization. [Online]: http://www.who.int, [February 2018].
[2] World Heart Federation. [Online]: http://www.world-heart-federation.org [February 2018].
[3] National Heart, Lung, and Blood Institute, "What is the heart?" [Online]: http://www.nhlbi.nih.gov/health/health-topics/topics/hhw/, [February 2018].
[4] Saudi Ministry Of Health website, "Cardio-Vascular Diseases". [Online]: http://www.moh.gov.sa/HealthAwareness/EducationalContent/Diseases/Heartcirculatory/, [February 2018].
[5] T. Kavalieratos, M. Nortje, and R. N. Dunn. "Hip and knee arthroplasty waiting list - how accurate and fair?" *South African Medical Journal*, Apr 2017.
[6] R. Zhang and Y. Katta. "Medical Data Mining, Data Mining and Knowledge Discovery," in *SPIE*, 2002, pp. 305-308.
[7] H. Wang and S. Wang. "Medical Knowledge Acquisition through Data Mining," in *IEEE*, 2008, pp. 1-4.
[8] L. Passmore, J. Goodside, L. Hamel, L. Gonzales, T. Silberstein, and J. Trimarchi, Assessing decision tree models for clinical in-vitro fertilization data. Technical Report TR03-296, Dept. of Computer Science and Statistics, University of Rhode Island, 2003.
[9] S. Patel and H. Patel, "Survey Of Data Mining Techniques Used In Healthcare Domain." *International Journal of Information Sciences and Techniques*, vol. 6, 2016.
[10] A. Patel, S. Gandhi, S. Shetty, and B. Tekwani. "Heart Disease Prediction Using Data Mining." *International Research Journal of Engineering and Technology*, vol. 4, 2017.
[11] N. Sharma and H. Om, "Data mining models for predicting oral cancer survivability," *Netw. Model. Anal. Heal. Informatics Bioinforma.*, vol. 2, no. 4, pp. 285-295, 2013.
[12] G. Rogers and E. Joyner, "Mining your Data for Health Care Quality Improvement," *SAS Institute,* Inc., January 2009.

 ISBN: 978-1-61208-618-7

[13] J. Soni, U. Ansari, and D. Sharma, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction." *International Journal of Computer Applications,* March 2011.
[14] Mayo Clinic, [Online]: http://www.mayoclinic.org/diseases-conditions/patent-ductus-arteriosus/basics/definition. [February 2018].
[15] University of California, Department Of Surgery. [Online]: http://pediatric.surgery.ucsf.edu/conditions--procedures/patent-ductus-arteriosus.aspx, [February 2018].
[16] Mrs. Bharati M. Ramageri*, "*Data Mining Techniques and Applications*", Indian Journal of Computer Science and Engineering,* vol. 1, no. 4, pp. 301-305.
[17] W. Klosgen and J. Zytkow, *Handbook of data mining a knowledge discovery.* Oxford University Press: New York, 2002.
[18] B.K. Bharadwaj and S. Pal, "Data Mining: A prediction for performance Improvement using classification", *International Journal of Computer Science and Information Security*, vol. 9, pp. 136-140, 2011.
[19] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, San Francisco, 3rd edition, 2013.
[20] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," In Proceedings of the 1993 ACM SIGMOD international conference on management of data, pp. 207–216, 1993.
[21] WEKA. [Online]: www.weka.com, [February 2018].

 ISBN: 978-1-61208-618-7