ISSN 2278-3091
Volume 8, No.6, November – December 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse142862019.pdf
https://doi.org/10.30534/ijatcse/2019/142862019

Designing and Developing Predictive Rehabilitation Management System for Patient Registry in northern Borneo

Soo Fun Tan[1], Fatimah Ahmedy[2], Suun Yie Teh[1]
[1]Faculty of Computing & Informatics, Universiti Malaysia Sabah, Malaysia
[2]Faculty of Medicine & Health Sciences, Universiti Malaysia Sabah, Malaysia
fatimahmedy@ums.edu.my

ABSTRACT

Patient registry is an important entity in healthcare information system. It serves as the platform for accurate and fast documentations of patients' demographics, clinical information, functional improvements, treatment outcomes and managerial details. With influx of massive data gathered in registry database, predictive data analytics has increasingly been applied as mean to predict and prognosticate outcomes from diseases and treatments, based on specific patients' characteristics. This article presents the design and development of predictive system intended for rehabilitation patient registry for information visualisation. It highlights the technical aspects of designing and developing the predictive system.

Key words: Predictive system, predictive data analytics, rehabilitation registry, rehabilitation management system

1. INTRODUCTION*

In the era of digital health, implementation of healthcare information system is rising globally leading to massive data and robotising patients' information into hundreds, if not more of patient registries worldwide [1]. Patient registry is an organised system for collecting, storing, retrieving and analysing information on individuals with diseases or conditions that predispose to the occurrence of a health-related event [2]. Layers of databases are required in the developed registry system due to the heterogeneity of the stored data and information, ranging from simple demographics to complex clinical details and multifaceted treatment outcomes. Such complexities present challenges for data queries and big data analysis especially for predictive analytics. Predictive analytics highlight the importance of having a system in the current state of healthcare to delivery preventive and effective disease management.

Rehabilitation registry in Sabah, a state located in northern Borneo, is still at its infancy with lack of data sharing and integration. Rehabilitation is a subspecialty of medicine that focused on prevention, prognostication and treatment of disabling diseases such as stroke, traumatic brain injury and spinal injury. A registry permits precise documentation of demographics, clinical information, functional improvement outcome of rehabilitation care services and research [3]. Before patient registries become available, clinical documentations and management were done manually requiring period of time and labour intensive. When data retrieval was needed, either for patients' management, audit purpose, investigative cause or medicolegal issue, the process was time consuming, inefficient and insecure in terms of protecting patients' confidentiality.

The conventional ways of clinics and hospitals for handling patients' health information and medical records include handwritten manual method, manual retrieval of medical record and manual extraction of data from the clinical notes, resulting in delays with medical treatment for patients due to heavy administrative task handled by the staffs. These lead to higher margin of human error would affecting the healthcare facility operation indirectly. Nevertheless, data sharing although convenient, gives rise to issues on patients' confidentially and privacy. In a survey among doctors and patients on information privacy in electronic medical records, patients do have some reservations regarding the usage of confidential information by third parties, but they still they appreciated the potential advantages from sharing and aggregation of electronic information when used specifically for management of their own health [4]. Doctors generally articulated positive views about electronic medical records systems with fewer apprehensions regarding potential risks on patients' medical information privacy [5].

On that note, a common patient registry merely acts as a database with real-time feature allowing access from various healthcare facilities. As such, data retrieval, although faster with the real-time and cloud-based system, is still warranted before transferred onto a data analytical platform for descriptive and predictive statistical analysis. In the field of medicine, data analysis is imperative in order to investigate causal and effect, risk factors and predictors of disease, and stipulate the outcomes and trends from treatments. Based on these motivations, an ideal patient registry would have following features: 1) primary purpose of digitizing patient records for an efficient data retrieval, 2) real-time feature for

* This work was supported by Universiti Malaysia Sabah's research grant.

registering and accessing patients' data effectively at different locations by different users, and 3) aptitude for descriptive and predictive data analytics for overall monitoring and benchmarking of rehabilitation care services.

From clinical standpoint, registries in developed countries are moving towards predictive analytical purposes. For example, in the UK, data registry from Global Burden of Diseases Injuries and Risk Factors Study (GBD, 2010) was further utilised to scrutinise the patterns of health loss and the leading preventable risks [6]. The China National Stroke Registry is a nationwide prospective registry initiated in 2007 aiming to evaluate the delivery and quality of stroke care [7]. National Cancer Registry Programme in India was developed for epidemiological studies and developing human resources for planning and monitoring cancer control activities [8]. Patient registries in the field of rehabilitation medicine have evolved for the past decade since the introduction and commendation of using World Health Organisation International Classification of Functioning, Disability and Health (ICF) framework that recognised the needs to identify, quantify and determine the disabilities, prognostications and predictors [9]. Model Systems Knowledge Translation Center at American Institute for Research developed comprehensive physical medicine and rehabilitation databases that not only store and register desired related information, but able to generate statistical outcomes for traumatic brain injury, spinal cord injury and burn.

Ability for predictive data analytics from a database system permits a decision-making based on available data in a predictive or adaptive pattern after describing and analyzing the situation from sensing, actuation and control, in short term called as predictive system. The models from this system allow healthcare providers to have better insight on the wider scenario as they connect patients' symptoms and characteristics or treatments to predict future outcomes [10]. From clinical management point of view, such predictive system enhances decision-making at a clinic or hospital setting according to information on hand to perform actions after summarising the situation from the gathered information. The use of predictive systems technology in the healthcare sector has been shown to produce more accurate diagnostic tools for provision of better treatment and quality of life for patients [11].

Predictive analytics is being said as one of the smart elements due to its ability to understand the future based on data. In reality, there is no statistical algorithm that can predict imminent events with full certainty but predictive analytics has the ability to estimate the outcomes based on existing data trends, statistics and calculations. This feature is very important in healthcare for preparing and arranging adequate healthcare facilities and equipment to face the anticipated circumstances or conditions based on the calculation of probability. It supports an integrated health system for a coordinated and comprehensive person-centered care to those who would benefit the most, i.e. the patients. Therefore, predictive analytics in healthcare has beneficial prospects for meaningful improvements in effectiveness, productivity, expenses and population health with targeted interventions toward individuals at risk [12]. Implementation of this feature would not only enhance the data visualisation but display relevant, meaningful facts rather than a mere display of descriptive information.

The purpose of this article is to present on the design and development of predictive system intended for rehabilitation patients registry for information visualisation. It highlights the technical and clinical aspects of designing and developing the predictive system.

2. REVIEWS ON EXISTING SYSTEMS AND CHALLENGES

Three existing electronic clinical management systems are discussed and contrasted based on their strengths and weaknesses: OpenClinic GA , Clinical Management System and Ispirithalaya Hospital Management [12] **-** [15] (Table 1). OpenClinic GA is an open source, multilingual, integrated hospital information management system focusing on both clinical databases and administrative domains [13]. This system is capable to generate descriptive statistical reports allowing end users to have an overview of the data that can be exported for secondary purposes. There is insufficient data for detailed descriptive data analytics as most of the descriptive analytics displayed were basic demographics that does not assist much for providing clinician or administrative staff a deeper insight such as issue on readmission after discharge from certain conditions.

Table 1: Comparison among OpenClinic GA, Clinical Management System and Ispirithalaya Hospital Management System

Dimension/ System	**OpenClinic GA System**	**Clinical Managemen t System**	**Ispirithalaya Hospital Managemen t System**
Accessibility	High	Moderate	Low
Quality of content	Good	Average	Average
Responsiveness	Moderate	Moderate	Moderate
Security and privacy	Good	Good	Good
Reliability	Good	Average	Average
Reporting tool	Good	Average	Average
Descriptive Analytic	Does not exist	Does not exist	Does not exist
Predictive Analytic	Does not exist	Does not exist	Does not exist
User Access Control	Exist	Exist	Exist

Clinical Management System System by Baratali Ghadamalizadeh is a healthcare computer system created to computerize clinical operation and automate medical workflows [14]. This particular system provides assistance and convenience for hospital staff roles including administrator, physicians, laboratory, imaging and pharmacy assistances by providing administrative control over management with added value of enhanced reporting and real-time information. It shares similar disadvantage like OpenClinic GA pertaining to inadequacy of data for predictive analytics. Ispirithalaya Hospital Management is an open source project built to assist in the management of a healthcare system allowing administration role for doctor and patient [15]. It does not have either descriptive or predictive analytics features to aid hospital administrative staff for decision making in managing and determining clinical–related problems.

The proposed system, Rehabilitation Management System (RMS) is an integrated information system designed to manage clinical information and rehabilitation activities of patients while the management system serves as a framework designed for policies, procedures and management processes to smoothen the operation of the rehabilitation department as the end-user of this system. RMS is a paperless record system with coordination between various units within the department leading to improved efficiency in care, speedy services delivery and higher accuracy of analytical results. In a way, RMS does acts as an e-health platform for healthcare professionals to gain access for patient’s record with clinical assessments and functional outcomes tools onboard. However, the proposed system is has predictive technology system, a software system with aptitude for managing, coordinating, and integrating related rehabilitation management activities as well as care operation based on different access control level in order to aid for a better decision-making at a clinic or hospital setting provided with the descriptive and predictive information gathered and analysed.

The reporting tool is one of the elements provided by the RMS to smoothen the clinical and managerial operations. It allows the presentation of data in better visualizations, provided with the data extraction as useful information for an enhanced demonstration of the message in a much convenient method. Reporting in RMS is exclusively important to benefit the management team with an improved understanding of patients’ condition and progress, as well as to perform better management in the future events, for example anticipating likelihood for admissions among those with traumatic spinal cord injury at interval periods after the injury based on the patients’ clinical characteristics and functional outcomes. Four main concerns of reporting tool are functionality, performance, usability, and flexibility. With these, data is able to be translated into useful information that is actionable to directly or indirectly aid the decision making of the rehabilitation department to meet patients’ treatments objectives. Besides, with interactive reports, specific filtering and navigation of data can be made. Researchers are concentrating on comprehending the user practices and preponderance towards the use of data visualization. Dashboard customization is one of the most utilized approaches that provide a perceptive atmosphere for examining, re-arranging and exploring multiple visual data representations from various data sources. These feature aids users to better retrieve, experiment and familiarize themselves with their data [16].

Nevertheless, the existing RMS has two functional features that may present as a problem with its reporting tool: 1) limitation of report presentation with non-personalized dashboard, and 2) descriptive and predictive data analytics have to be performed manually, i.e. data have to be extracted manually and statistically analysed using other relevant software that can be time consuming and labour intensive. Massive unorganized information in a healthcare facility has created unnecessary data flooding in the system. The report served as an important tool to condense gathered information and display them in the form of a simplified list of data collection or chart. However, filtration of data is not effective enough for the complex and complicated report, making sorting and searching process of the information inefficient while retrieval of data analysis result is inconvenient.

Reporting tools in existing RMS are lack of customization specific user roles while the presentations are mostly unresponsive, displayed in the textual and tabular method. Data can be re-arranged from lowest to highest or in stem-and-leaf plot format for textual method while tabular method can be presented in a frequency distribution table (FDT), relative FDT, cumulative FDT or contingency table. Both of methods of presentation is often ineffective to quantity such large data leading to inaccurate analysis. However, tables and graphs preparation are still crucial in the analysis and generation of results in order to organize the gathered information in a transparent and concise manner motivating researchers for further studies on developing tools for data visualization techniques [17]. Information can be presented efficiently with a significant visual appeal to attract users, making the results easily comprehensible [18].

Implementing descriptive and predictive data analytics into the RMS would not only imply the ability for summarizing and describing the massive data but also objectively illustrate the past behaviors for a better insight on the occurring trends. The evaluation metrics over time would likely explain affects or influences on the future outcome, clinically and functionally. Innovating and manipulating data from e-health platform is made possible from massive information, inferring the importance of tremendous data input. In principle, predictive analytics are made possible by sorting out the affecting factor, and this is performed using advanced tools such as Tableau, QlikView, Google Analytics etc. It uses statistical models and forecast techniques to identify the future probabilities and trends about what could possibly happen in the future. By implementing descriptive and predictive analytics, a system can work better in providing information related to both historical insight and future trends based on the data collected.

3. DESIGN PRINCIPLE

Descriptive data analytics is the most basic form of analytics where examination of data is performed to answer the question of "What has happened?" and characterized data with visualizations such as graphs, bar chart, pie charts, tables or generated narratives. It uses data aggregation and data mining to yield useful information by providing insights based on the historical summary, which is useful for the preparation of further analysis [19]. It analyses both historical and current data to find out reasons for success or failure event. It describes data in an easy and understandable form by summarizing raw data for simpler interpretation of people. Descriptive and predictive analytics are often related.

Descriptive data analytics prepare the data for predictive data analytics and frequently implemented to help improve the decision-making. There are several types of measures for descriptive data analytics used in Predictive RMS. Arithmetic mean is equal to the sum of all the values in the data set divided by the number of values in the data set. For median, it is the middle value of the variable that divides the series of data in two equal parts, upper and lower while the mode is the value that shows the highest frequency in the series or the most repeated value in a distribution. Mode is the only measure of central tendency that can be used for data measured in a nominal scale. Measure of dispersion is based on the standard deviation and variance. In the measure of position, data is sorted from lowest to highest using quartiles, percentiles and standard score, based on the type of data in the database.

Predictive data analytics is well known for its capability to make predictions about unknown future events based on historical data. It uses various forms of statistical techniques from data mining, statistics, modeling, and machine learning to analyze current and historical facts to make predictions based on the risks and opportunities identified. Generally, past data is being used as the extracted information to build a mathematical model and predictive data analytics is then implemented to forecast the trends and pattern of the market by making predictions. Before predictive data analytics technologies were widely available, healthcare organizations normally devised strategies by manually reviewing forecasts to project clinical outcomes, estimation of admissions and cost for treatments in certain period of time. Predictive data analytics make this process easier and more automated with the help of various tools and analytics software. However, there are several issues to be considered in order to implement predictive analytic methods effectively. First is to understand the approach that should be used to predict the outcome. Second, is to find new measurement sources that be incorporated for improving the predictions. Third, is to plan how to make the resultant predictions as actionable by identifying the benefits and improvements. Fourth, is to consider how to account for the many potential causes of outcomes in predictive models to suggest more accurate modeling, particularly for population management [20].

Predictive data analytics often play a key role in further works and these data will be most effective for informing healthcare decisions particularly to forecast the likelihood of certain disease complications or risk factors affecting specific demographic populations [21]. This allows preventive healthcare measures to take place and predict the likelihood of treatment failures so that preventive care can be administered [22]. Healthcare providers, sales and marketing professionals, systems admins, etc. in both the public and private industries use predictive data analytics technology. It gives an organization a competitive advantage where knowledge gained from predictive model can aid decision-making to bring benefits to the organization. It is being adopted more frequently to ease the whole operation as massive amounts of information in health and patients databases have been searched and analyzed using technology and statistical methods to predict outcomes for different types of patients.

3.1 Statistical Modeling and Machine Learning Techniques

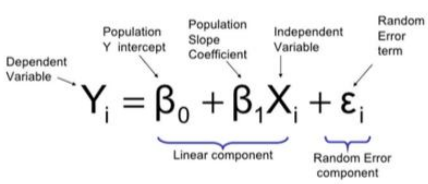
There are several models that can be applied for predictive data analytics depending on the situation. Generally, two types of techniques are considered for predictive RMS; Statistical Modeling and Machine Learning. Statistical modeling (SM) focuses on the formalization of the relationship between variables (independent and dependent) in the form of mathematical equations. It is much related to the math fields that prediction of outcomes is conducted based on finding the relationship or association with preference for simpler models over complex ones.

Both approaches aimed to learn about the underlying phenomena by using massive data generated in the process. Engagement in ML requires effort for selecting the right algorithms, formulating the code, stipulating the parameters, testing the algorithm, and scrutinizing the accuracy of the resultant predictions. For SM, the engagement warrants choosing and specifying a statistical model family, checking the goodness of fit, analyzing the accuracy of predictions, and deciphering projected effects [23]. However, with massive information and heteregeneiy of the clinical data, one or two sets of algorithms may not achieve the aims. ML is better suited for emphasizing prediction while SL focused more on estimation and inference. However, overfitting and dimensionality have been the two biggest problems faced by ML, both historical and ongoing. For overfitting, a model of ML exhibits biases towards the training data and does not generalize to new data, and/or variance. For dimensionality, data understanding in ML is difficult for algorithms to work within multiple dimensions. For SM, there are numerous types of regression models available to use depending on the dependent variable and the type of model that provides the best fit. Normally the suitable type of model to be implemented in a system can be decided by focusing and determining the type of a dependent variable of data that measure continuous, categorical and count data.

3.2 Regression Analysis

In SM, regression analysis is a set of statistical processes to estimate the relationships among variables. SM implies independent and dependent variables in almost all cases where the dependent variable is often represented on the Y-axis while the independent or explanatory variable is represented on the X-axis in modeling charts. Both dependent and independent variables may be single or multiple, quantitative or qualitative. Regression analysis allows the comparison of the effects of variables measured on different scales. These techniques are mostly driven by three metrics; number of independent variables, type of dependent variables, and shape of the regression line. There are various techniques for modeling to be adapted in different situations for predictions data analytics. The most commonly known techniques, linear regression and logistic regression are used and will be further discussed.

Linear regression is widely utilised modeling technique entailing a continuous dependent variable, continuous or discrete independent variable(s) and linear regression line [24]. Linear regression ascertains an association between the dependent variable (Y) and one or more independent variables (X) using the best fit regression line. It is denoted by an equation that predicts the value of a dependent variable based on a given independent variable. The formula for Linear Regression is displayed in Figure 5.



(1)

Y is a continuous dependent variable and X is an independent variable (can be in the form of continuous or binary) or other discrete domains. ε is a term for the variance that is not explained by the model and is usually named "error". Individual dependent values denoted by Yj can be elucidated by adapting the equation to Yj=b0+∑(biXij)+εj.

Multiple linear regression is similar to linear regression except it contains more than one independent variables and Least Square Method is the commonest method used to obtain the best fit line. It calculates the best-fit line for the observed data by minimizing the sum of the squares of the vertical deviations from each data point to the line. Because the deviations are first squared, upon added, there is no canceling out between positive and negative values. Following this, the evaluation of model performance can be done using the metric R-square.

However, major concerns of using linear regression are that there must be a linear relationship between independent and dependent variables as it is very sensitive to outliers that can affect the regression line and eventually the forecasted values. On the other hand, multiple linear regression suffers from multicollinearity, autocorrelation, heteroscedasticity where multicollinearity could increase the variance of the coefficient estimates making the estimates very sensitive to minor changes in the model resulting in unstable coefficient estimates. In the case of multiple independent variables, forward selection, backward elimination and stepwise approach are better for selecting the most significant independent variables. Logistic regression is the technique used for any data with an event outcome exist as either one in two forms, for example, the probability of presence or absence of a complicated urinary tract infection after spinal cord injury. The response provided would be in a binary form (e.g. Yes/No). It is widely used for classification problems in which linear relationship between dependent and independent variables is not required. The formula for Logistic Regression is displayed in Figure 2.

$$P = \frac{e^{a+bX}}{1+e^{a+bX}} \quad \text{or} \quad P = \frac{1}{1+e^{-(a+bX)}} \tag{2}$$

P is the probability of a 1 (the proportion of 1s, the mean of Y), e is the base of the natural logarithm (about 2.718) with a and b as the model's parameters. The value of a yields P when X is zero and b reflects the changes in the probability upon changing X as a single unit. As the association between X and P is nonlinear, b does not have a forthright interpretation as it does in ordinary linear regression. Hence, a link and logit function is more appropriate upon dealing with a dependent variable that has binomial distribution.

Logistic functions are used to identify how the probability of an event is affected by one or more dependent variables. The independent variables should not be correlated with each other. Functions used for ordinal and multiclass values dependent variablea are Ordinal logistic regression and Multinomial logistic regression respectively. All significant variables are included utilizing stepwise method for estimating the logistic regression as ways to avoid overfitting and underfitting. Estimating parameters of the statistical model in given observations is based on Maximum Likelihood Estimation method but it requires large sample sizes.

There are multiple types of regression models that can be selected; hence it is important to choose the best-suited technique based on the type of data variables (independent versus dependent), dimensionality and fitting. The simplest way to adopt is to apply linear regression for an outcome in a continuous form but implement logistic regression for binary and discrete outcomes. It is advisable to use linear regression initially for fitting a model and subsequently determine if such model provides an adequate fit by checking the residual plots. Several factors have to be considered prior selecting the right regression model that best suit for data. Before selecting a model, the relationship and impact of variables have to be identified during data exploration. Metrics such as R-square, adjusted R-square, AIC, BIC and error term are some available examples to be used for fit of models comparison. Dividing the data set into two groups (train and validate) with a mean squared difference between the observed and predicted values would allow some prediction of the accuracy.

There are several advantages of linear regression, for example, interpretability and computing speed. Computational time for logistic regression fitting is slower due to the iterative process of Maximum Likelihood involved but this is not noticeable if small or moderate-sized dataset are being used for fitting a simple model. Linear regression is faster as it uses ordinary least squares (OLS) estimation. OLS does not take into account that the linear probability model is heteroskedastic with residual variance of p(1-p), nevertheless the heteroscedasticity is considered minor if the value is between 0.20 and 0.80. OLS estimates can be improved by using heteroscedasticity-consistent standard errors or weighted least squares. For this reason, linear regression is recommended at initial stage of analysis.

4. SYSTEM ANALYSIS AND DESIGN

4.1 System Analysis

Small group discussion using interview guides was conducted to analyse end user requirement for Predictive Rehabilitation Management System (Predictive RMS). Following the agreement with the studied hospital, interviews include discussion on data requirements, business process and query practices, as well as the condition of the required visualisation report. A brief description of the purpose and objectives of the Predictive RMS was relayed to interviewees consisted of rehabilitation medicine clinical specialists. The system is collecting information for Sabah Rehabilitation Registry and gathering data from three common rehabilitation diseases admitted to a rehabilitation ward: stroke, spinal cord injury (SCI), and traumatic brain injury (TBI). Data collected are primarily divided onto 5 domains: patients' demographics, disease-related clinical information, rehabilitation assessment, functional outcomes and admission details.

The interview gathered information on the anticipated functions for clinical purposes, how the registry would be utilised at the clinical setting, preferred user interface design, customisation of sorting and filtering features for ease of searching process and main concept of dashboard content. The discussion also inquired on the expected total patients for registry with numbers and levels of end users for the system.

In reference to descriptive and predictive data analytics, it is suggested and preferred for the data analytics to be based on these 5 domains with implementation of visualization report feature as well.

4.2 Data Flow Diagram (DFD)

Figure 1 illustrates the data flow diagram (DFD) of the Predictive RMS with three different user roles (system user, system administrators, and clerk) consisting of eight modules: 1) Registering Module, 2) Login Module, 3) Storing of Patient Details, 4) Descriptive Data Analytics, 5) Predictive Data Analytics, 6) Visualization, 7) Report Management, 8) Dashboard Management. Both the system user and hospital administrator are able to conduct descriptive data analytic, predictive data analytic, visualization and report function where the system will get the function and category option, compute and display them to the users. Computed data will be stored in a database named Report for detail purposes.

4.3 System Architectural Design

A total of seven entities exist in the Entity Relationship Diagram (ERD) with description of attributes described in the data dictionary for the proposed Predictive RMS: Users, Patients, Patient_TBI, Patient_Stroke, Patient_SCI, Reports, and Settings. Denormalized ERD is used to increase the performance by decreasing running time process (Figure 2).

5. PRELIMINARY IMPLEMENTATION

5.1 Software Tools for Development

Throughout the whole process of developing the system, Sublime Text is the cross-platform text editor used for both frontend and backend programming due to its capability in supporting many programming languages. PhpMyAdmin is utilized to handle and administrate MySQL, the relational database management system of the project with Xampp as the cross-platform control panel to host the server.

5.2 Preliminary System Implementation

Multiple tables are created for the database based on three different diseases; stroke, SCI and TBI.

5.3 Preliminary Testing and Evaluation

Data entry of patients' information onto Predictive RMS is still ongoing during the narration of this article. The data is yet to achieve a certain threshold in order to test the modeling used for predictive data analytics. The system is not yet running on real-time basis until the analytical parts are fully tested and evaluated for their accuracies.

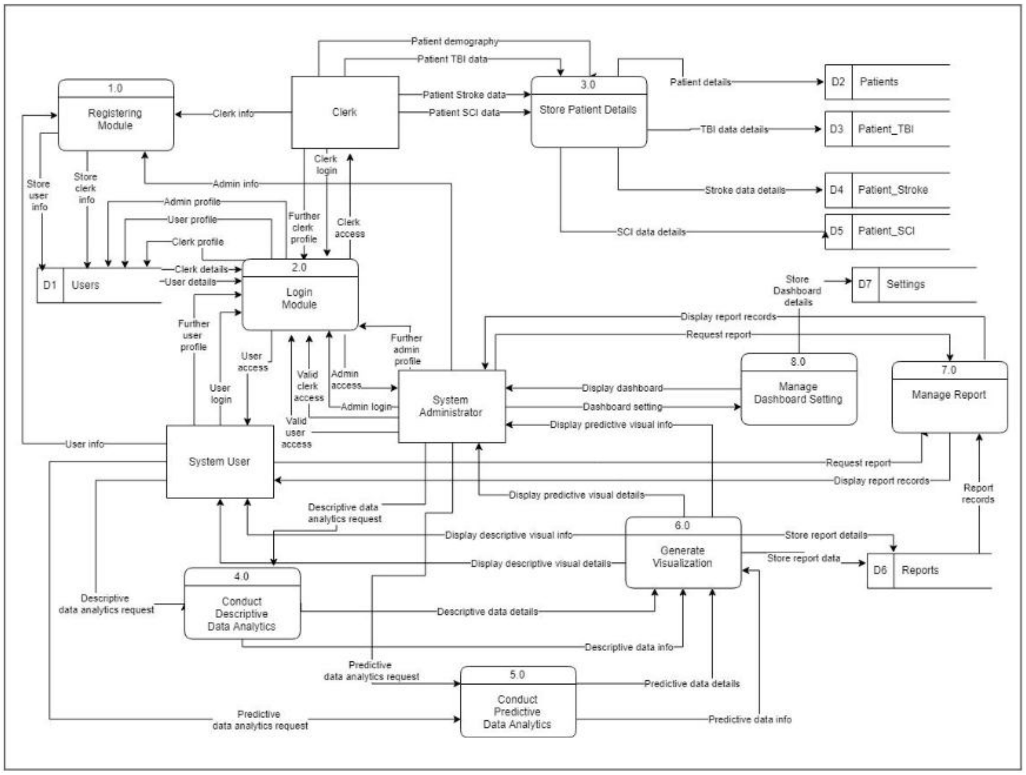


Figure 1: Data Flow Diagram for Predictive RMS

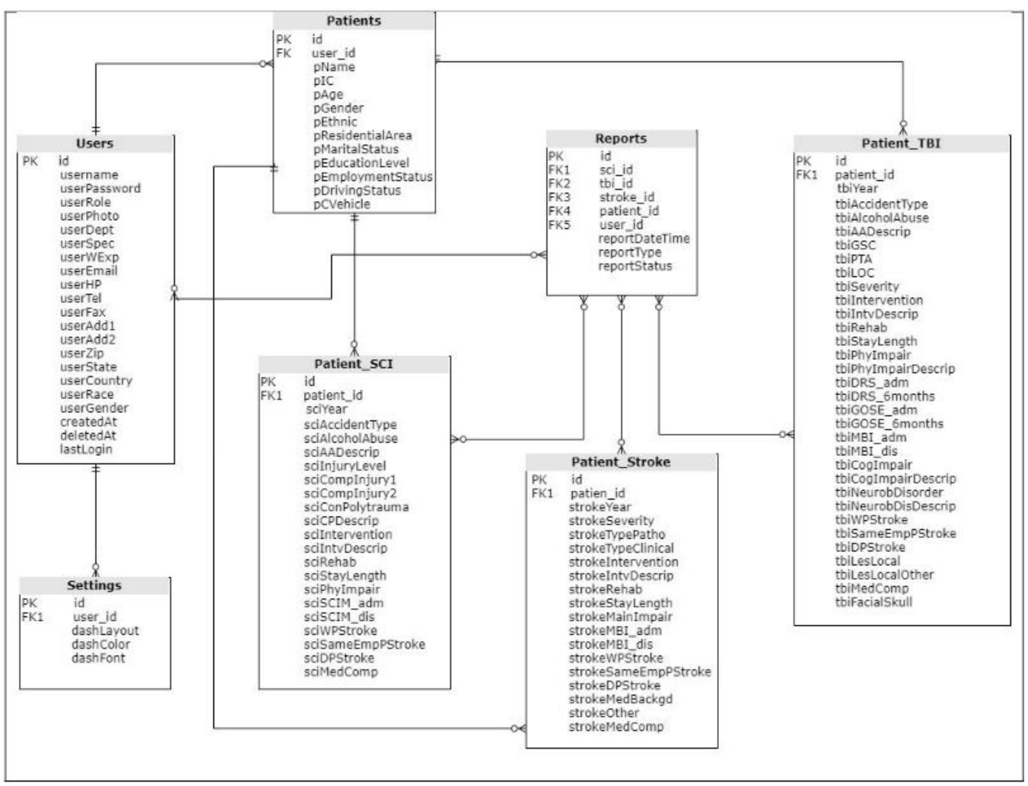


Figure 2: Denormalisation of ERD in Predictive RMS

6. DISCUSSION

The preliminary implementation has shown that Predictive RMS can be advantageous in supporting the patient registry. Further manipulation of patients' data would be the next step to evaluate the accuracy of statistical modelling implemented for predictive data analytics. Expected output are aimed to explore rehabilitation treatment outcome and predict the outcome at interval period after a certain disease is acquired. The latter is crucial for healthcare professionals as these would guide the subsequent therapeutic interventions without exhausting extra cost or resources.

REFERENCES

1. A. Maqsood. Study of Big Data: **An industrial evolution review of applications and challenges,** *Int J Adv Trends Comput Sci Eng,* vol. 6, no. 3, pp. 31-33, May 2017.
2. R.E. Gliklich, N.A. Dreyer and M.B. Leavy. ***Patient Registries,*** 3rd ed. Rockville, U.S.:Agency for Healthcare Research and Quality. 2014, ch 1.
3. P.W. New, F. Simmonds and T. Stevermuer. **A population-based study comparing traumatic spinal cord injury and non-traumatic spinal cord injury using a national rehabilitation database,** *Spinal cord*, vol. 49, no. 3, pp. 397, Mar. 2011. https://doi.org/10.1038/sc.2010.77
4. F. Riordan, C. Papoutsi, J.E. Reed, C. Marston, D. Bell and A. Majeed. **Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK,** *Int J Med Inform*, vol. 84, no. 4, pp. 237–247, Apr. 2015.
5. M.M. Bouamrane and F.S. Mair. **A study of general practitioners' perspectives on electronic medical records systems in NHSScotland,** *BMC Med Inform Deci Mak,* vol. 13, pp. 58, Dec. 2013.
6. C.J. Murray, M.A. Richards, J.N. Newton, K.A. Fenton, H.R. Anderson, C. Atkinson and T. Braithwaite. **UK health performance: findings of the Global Burden of Disease Study 2010,** *The Lancet*, vol. 381, no. 9871, pp. 997-1020, Mar. 2013.
7. Y. Wang et al. **The China National Stroke Registry for patients with acute cerebrovascular events: design, rationale, and baseline patient characteristics,** *Int J Stroke*, vol. 6, no. 4, pp. 355-361, Aug. 2011.
8. G.K. Rath and A.K. Gandhi. **National cancer control and registration program in India,** *Indian J Paediatr,* vol. 35, no. 4, pp. 288, Oct. 2014.
9. G. Stucki, J.D. Reinhardt, G. Grimby and J. Melvin. **Developing research capacity in human functioning and rehabilitation research from the comprehensive perspective based on the ICF-model,** *Eur J Phys Rehabil Med*, vol. 44, no. 3, pp. 343-351, Sept. 2008.
10. S. Kumar, P. Sharma and madhusudan. **Predictive analytics: Cloud computing,** *Int J Adv Trends Comput Sci Eng*, vol. 5, no. 3, pp. 6-12, May 2016.
11. R. Vullers, J. Penders, S. Brongersma, J.L.P. Wijsman, M. Altini and C. Van Hoof. **Wearable Smart Systems for Personal Healthcare and Lifestyle,** in *Smart Systems Integration*, Amsterdam, 2013, pp -.
12. A. Giga. **How health leaders can benefit from predictive analytics,** in *Healthcare Management Forum,* Los Angeles, 2017, pp. 274-277.
13. G. Karara, F. Verbeke, E. Ndabaniwe, E. Mugisho and M. Nyssen. **OpenClinic GA Open Source Hospital Information System Enabled Universal Health Coverage Monitoring and Evaluation in Burundian Hospitals,** *Med Info,* 2017, pp. 738-742.
14. B. Ghadamalizadeh. **Clinical Management System,** 2018, Retrieved from Source Forge: https://sourceforge.net/projects/clinicmanagementsystem/
15. D.T. Kanishka Panamaldeniya. **Ispirithalaya Hospital Management**, 2015. Retrieved from Source Forge. https://sourceforge.net/projects/ispirithalaya/
16. M. Elias and A. Bezerianos. **Exploration Views: Understanding Dashboard Creation and Customization for Visualization Novices,** in IFIP Conference on Human-Computer Interaction, Berlin, 2011, pp. 274-291.
17. A. Erraissi and A. Belangour. **Meta-modeling of Big Data visualization layer using On-line Analytical Processing (OLAP),** *Int J Adv Trends Comput Sci Eng*, vol.8, no. 4, pp. 990-998, Jul. 2019. https://doi.org/10.30534/ijatcse/2019/02842019
18. R.P. Duquia, J.L. Bastos, R.R. Bonamigo, D.A. González-Chica and J. Martínez-Mesa. **Presenting data in tables and charts,** *An Bras Dermatol*, vol. 89, no. 2, oo. 280–285, Apr. 2014.
19. W. Raghupathi and V. Raghupathi. **An Overview of Health Analytics,** *J Health Med Informat* vol. 4, pp. 132, Oct. 2013.
20. D.W. Bates, S. Saria, L. Ohno-Machado, A. Shah and G. Escobar. **Big data in health care: using analytics to identify and manage high-risk and high-cost patients,** *Health Affairs,* vol. 33, no. 7, pp. 1123-31, Jul. 2014.
21. H. Alharthi. **Healthcare predictive analytics: An overview with a focus on Saudi Arabia,** *J Infec Public Health,* vol. 11, no. 6, pp. 749-56, Nov. 2018.
22. I.N. Soyiri and D.D. Reidpath. **An overview of health forecasting,** *Environ Health Prev Med,* vol 18, no. 1, pp. 1–9, Jan. 2013.
23. S. Makridakis, E. Spiliotis and V. Assimakopoulos. **Statistical and Machine Learning forecasting methods: Concerns and ways forward.** *PLoS One,* vol. 13, no. 3, pp. e0194889, Mar. 2018. https://doi.org/10.1371/journal.pone.0194889
24. A. Schneider, G. Hommel and M. Blettner. **Linear Regression Analysis: Part 14 of a Series on Evaluation of Scientific Publications,** *Dtsch Arztebl Int*, vol. 107, no. 44, pp. 776–782, Nov. 2010.