Data Mining Methods to Improve Clinical Trials in Diabetic Patients

Chandeep Kaur* and Olufemi Muibi Omisakin

Department of Applied Business, Nelson Marlborough Institute of Technology, Auckland Campus, New Zealand

*Corresponding author: Chandeep Kaur, Department of Applied Business, Nelson Marlborough Institute of Technology, Auckland Campus, New Zealand, Tel: +919560583665; E-mail: chandeep.virdi@gmail.com

Abstract

Diabetes is on the rise in New Zealand with over 250,000 New Zealanders enrolled in primary health organisations for diabetes (predominantly Type 2) and an estimated 100,000 more remain undiagnosed. Records indicate that 68.5% of Pacific women are prone to gestational diabetes (GDM) and Type 2 diabetes compared with Maori with 47.3% while the diabetes rate of Indian women is only 5.5%.

Clinical trials, regardless of the therapeutic area, collect a huge amount of data which can be structured or unstructured in nature. Therefore, it becomes difficult to analyse such huge amount of data. As a result this research evaluates different data mining methods capable of analysing big clinical data. Using primary and secondary research, this study proposes one of the best techniques that can be used in data mining of such huge data. To accomplish this, the study briefly discusses different data mining methods. The approach aims to extract information from clinical trial data mining methods to improve on them analytically, and above all to make a recommendation for the best possible data mining and analytical technique.

Keywords: Diabetic research; Machine learning; Bayes classification; Framework; Data mining; Type 2 diabetes; Big Data

Introduction

Clinical trials generate data on safety and efficacy. The drugs that are present in the market have already gone through a clinical trial in which patients are enrolled and the drug is checked for its efficacy and safety through the analysis of data collected. The final data report is then submitted to Food and Drug Administration (FDA) for approval. Clinical trials are being conducting all over world taking a similar approach and using similar systems to analyse data. With so many trials on-going with increasing numbers of diabetes cases, it becomes difficult to analyse such huge data. This research focuses on Type 2 diabetes which accounts for nearly 90% of all cases of diabetes. There are trials still running in the same manner as they were years ago. Therefore, this study aims to assess data mining methods such as Big data, Naïve Bayes Classification, and Machine Learning within clinical trials and their effectiveness in the analysis of data.

Conditions such as pre-diabetes cannot be called "diabetes" as the abnormal glucose level is not high enough for a diagnosis of diabetes. This is a situation which occurs when the glucose in the blood is higher than normal. People who are overweight or people with high blood pressure or people who have a family history of diabetes are at risk of the disease. According to New Zealand Herald [1], in the past 10 years, the number of New Zealanders with diabetes has doubled from 125,000 to 250,000. In the past 10 years there has been more than 40 new diabetes diagnoses every day [2].

Diabetes is a very common disease, and it can be prevented by maintaining a healthy lifestyle. Type 2 diabetes usually occurs after the age of 30. However, there are now cases of Type 2 diabetes seen in the teenagers and children. Type 2 diabetes occurs when insulin is not produced in the body to retain normal levels of blood sugar (glucose). Body needs glucose, but if the blood glucose is too high, it can damage the body within a particular time frame. According to bpac better medicine [3], there has been an increase in diabetes cases amongst the younger generation requiring effective data mining and analysis. This study discusses different data mining methods that can be beneficial for clinical trials of diabetic patients. Therefore, effective procedures or processes which can be thoroughly used in the clinical trials for effective diabetes management need to be recognised. This study will be beneficial by identifying the right model for the prediction of diabetes in future and thereby reducing the incidence of this disease. The study will assess people's responses through a survey questionnaire to ascertain the trends of data mining methods and how they are being used.

This study starts with an introduction followed by the literature review, the method and methodology chapter, data analysis and interpretation after which findings and contributions were discussed.

Research question

How often do clinical researchers or healthcare companies use Big Data or Naïve Bayes Classification data mining methods?

Research objectives

- To determine and compare traditional research methods and the Big Data method used in clinical trials for diabetes.
- To assess the Naïve Bayes Classification and other data mining methods that can be used in clinical trials.

Literature Review

Background

Increases in Type 2 diabetes have been observed in children recently. In New Zealand, type 2 diabetes in children was first observed in the 1990s. A study carried out between 1995 and 2015 found that 104 young people under the age of 15 were treated for type 2 diabetes [4]. There is an overall increase in type 2 diabetes cases amongst Pacific Island youth by jumping to 3.6 per 100,000 and 3.3 per 100,000 for young Maori [5]. Type 1 diabetes is an auto-immune disease which cannot be prevented, but type 2 diabetes is something that can be prevented by leaving a healthy lifestyle [2].

As the cases of Type 2 diabetes are increasing lot of patient's data that is collected by clinical trials. Therefore, using a more realistic data collection and analytical tools are required. There is no fixed technique to use, and they all require a lot of manpower (especially manual review) to analyse the data. Therefore, a lot of research is being done to come up with more efficient technique to analyse clinical data. For instance, Naïve Bayes Classification and Big Data analytical tools are believed to be better than the traditional one. This study compares traditional research methods with Bayes Classification and Big Data and summarise trends in current trials for diabetic patients and find out how Naïve Bayes classification and Big Data can be useful in clinical trials for diabetic patients.

Application of data mining techniques in medical sciences is becoming crucial in converting available data into valued knowledge. This research engaged in methodical reviews of uses of machine learning and data mining tools in diabetic research regarding the prognosis and diagnosis of Type 2 diabetes and complications from the disease. In diabetes research, data mining is an important method used for processing large volumes of data on diabetes; and it helps to generate knowledge. Therefore, Big Data, traditional data mining methods and Bayes Classification will be studied to ascertain their effectiveness in diabetes research.

In diabetes mellitus research, data mining is an important method used for large volumes of data available for diabetes; it helps in getting more knowledge about it. Therefore, big data, traditional data mining methods and Bayes classification will be studied to understand their effectiveness over the diabetes mellitus.

Big Data is the analysis of huge datasets computationally to understand, identify and disclose arrangements, trends and links, especially relating to human behaviour and interactions. Using Big Data can help forecast or analyse diabetes amongst undiagnosed patients and manage extensive data [6]. The authors argue that health tracking devices like FitBit help in recording data of the patient and reporting the same.

Rise in diabetes cases in younger generation

Many clinical trials are conducting to understand the precursors that are interlinked with the diabetes and its causes. Recently, machine learning, data mining and Big Data methods have been used for clinical trials for predicting diabetes and other diseases in patients. It is also being used to understand the precursors that are interlinked with diabetes and its causes

Recently new cases of Type 2 diabetes came up in younger generation in New Zealand. Healthcare providers are looking for new methods to overcome this issue and help discover diabetes early in patients to help them prevent other related diseases as diabetes is linked with higher health risks **(Figure 1)**. The below chart indicates the rate of increase in the number of Type 2 diabetes cases in New Zealand among younger people that there is a high increase in the number of Type 2 diabetes cases in New Zealand amongst younger people [3].
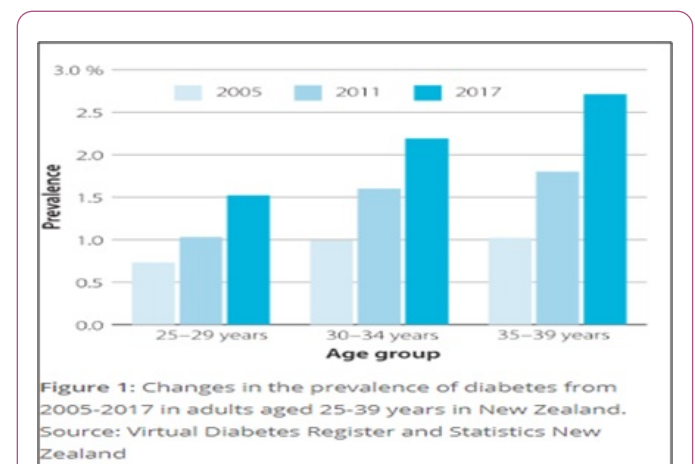

Figure 1: Changes in the prevalence of diabetes from 2005-2017 in adults aged 25-39 years in New Zealand. Source: Virtual Diabetes Register and Statistics New Zealand

Figure 1 A rising tide of type 2 diabetes in younger people: What can primary care do?

From the above it is quite clear that the diabetes cases are increasing in New Zealand; which means that the diabetes self-management education system and analysis methods are having some loopholes which impact on the overall analysis of clinical data recorded for such trials. Diabetes can be controlled by managing obesity, maintaining quality of life and diabetes self-management education (DSME).

Challenges in data outcome in diabetes

Quick prognosis of pre-diabetes and diabetes could help check diabetes mellitus. Both diseases are common and inflict a significant public health drain. The big prevalence of early detection through the motion of big early checking out of asymptomatic patients has yet to be actually confirmed, and rigorous trials to provide evidence are pretty unlikely to arise. A sturdy prediction for damaging consequences is the duration

of glycemic burden, yet efficient interventions do exist to lower the chance of DM development from pre-diabetes and reduce the prevalence of complications springing up from the disease. An established risk factor for Type 2 diabetes is overweight, and sadly many sufferers are too regularly prescribed anti-diabetic medications that increase their appetite and are a major reason for weight gain. The most usual pills related to weight gain in treating diabetes are sulfonylureas, thiazolidinedione (TZDs), and insulin. Current research has also indicated that hypothyroidism is a risk factor for new onset of Type 2 diabetes, which is also related to the usage of statins, which are extensively-used cholesterol-decreasing drugs. The subsequent generation of research for drug efficacy and remedy improvement endeavours will be advantageous for applying modern technologies which use Artificial Intelligence and Big Data analytics on the molecules, genetics, mechanism of movement, and digital clinical records. The art of forecasting clinical effects, outcomes and risk elements can be employed by evolving targeted treatment plans and creating medical trials, as well as helping produce better-differentiated results with higher therapeutic and business options. Therefore, pre-medical experimental data can be immediately linked with potential scientific results. Some of the challenges faced doing these are provided below.

Obesity a main precursor

Obesity is considered and proven to be main precursor of Type 2 diabetes. With obesity, the body's blood glucose and haemoglobin A1C levels also increases. This is the main cause of diabetes in patients. The treatment for the hyperglycaemia of Type 2 diabetes is closer control of blood glucose [7]. Some of the other risk factors which also need to be controlled along with controlling blood glucose and Haemoglobin A1C levels are smoking, controlling blood pressure, and controlling cholesterol [8]. There are patients who prefer to undergo weight loss surgery, which helps in lowering the weight of the person, thereby reducing the blood glucose and Haemoglobin A1C levels as well.

Quality of Life (QoL)

Patients with diabetes are assessed for Quality of Life (QoL) which is normally confused with its usual meaning. The term Quality of Life is composed of four components: Physical, mental, cognitive, psychological and social. Diabetes worsens a person's Quality of Life, and if it exists with another illness, it is exacerbated.

Diabetes self -management education system

Self-management of diabetes includes agreement to pharmacotherapy, proper diet, routine exercise, timely blood glucose monitoring, foot care, eye care and dental care. The primary diabetes self- management system also includes support given by the healthcare professional.

The maximum examined model of diabetes care is the Chronic Care Model (CCM), which is an organised model of care for patients with long-lasting illnesses. These types of processes and methods give an opportunity to be able to see further and review challenges faced during diabetes management.

Treating older diabetes

Health care for older people poses lot of challenges to advanced practice nurses or nurse practitioners and nurses with doctoral level qualifications in the science of learning. Basically, diabetes specialist nurse (DSN) face lot of challenges because as people grow older they develop more health-related problems including diabetes. Many older people show lot of problems in managing their diabetes because of living alone or having lost their partners. However, some older people manage their diabetes effectively and live liberated lives. Maintaining the right blood glucose level in the body requires several complex actions such as having nutritious food, regular exercise, taking medicines on time and checking the blood glucose level regularly to manage diabetes effectively [9]. Some other problems linked with the maintenance of diabetes are reduced sleep quality (PSQI) or dementia and it was noted that out of 25, one person develops dementia between the ages of 70-79.

Mindful Attention Awareness Scale (MAAS)

The Mindful Attention Awareness Scale (MAAS) scores patients based on age, sex, race/ethnicity, family history of diabetes, and youth financial status and reflect glycaemic control and self-mind. Care is regularly characterised by the patient's capacity to take care of ordinary typical errands nonjudgmentally while taking care of physical and mental procedures. An expanded MAAS implies a patient is impressively more prone to have great glycaemic control with typical blood glucose (BG) and not in great danger of Type 2 diabetes. Better BG control is associated with dispositional care, generally because of the lower body weight and feeling of better control in patients with more prominent levels of care. "Dispositional care" indicates a natural, yet factor, quality, where most people have different abilities to centre around and know about the "occasion" or occasions that are occurring [10].

Big data analytics in diabetes management

Big Data comprises the collection and assimilation of huge and mixed datasets, whereas education analytics include inspection of the patterns in single or collected datasets. Big Data uses several data resources and proposes that the maximum Big Data which could be employed by healthcare provider, education research most probably was not learnt for education purposes. To understand the importance of Big Data from this research perspective, a brief comparison between

traditional research methods and Big Data research method is tabulated in **Table 1**.

Table 1 A comparison of traditional research methods and the big data method.

Aspects	Traditional methods	Big Data method
Data collection instruments	Purpose-built data gathering instruments	Opportunistic mining of pre-existing and dynamically accumulating data
Data cleaning	Scrupulous confirmation of data quality	Allowing uncertainty in data quality
Data collection intrusiveness	High	Low
Data acquisition cost per subject	High	Low
Methodology	Governs what data is essential	Reacts to what data is available
Analysis	Prevailing desktop approaches and tools (SPSS, NVivo, Atlas.ti)	Potentially new tools required to parse and report on very large datasets
Sample properties	Tightly defined	Loosely defined (depending on markers in the data to specify population features)
Contexts	One to a few	Many (hinges on how a context is defined)
Replicability	Potentially problematic to sustain or recreate study context, resources	Comparatively easy to access and reanalyse data; experiments may be open-ended
Sample size	Variable, tending to small	Variable, tending to very large
Temporality	Static final reports	Dynamic dashboards
Comparison of traditional research methods and the big data method		

All kinds of data can be examined using Big Data. This data can include data such as diagnoses, vital signs, prescribed medicines and several types of reported symptoms and diseases can be examined using Big Data. Clinical trial data is in two forms i.e., digital and paper. Data can be in any form like Electronic Health records (EHR) of a patient, clinical reports, doctor's prescription, diagnostic reports and images etc. Data that is raw and can be evaluated further includes blood pressure readings, height, weight and glucose measurements etc. Data mining and big data has been used by a lot of researchers to predict models for diabetes. Currently, the researchers have revealed that a patient's record can be studied using data analytics to foresee the future risk of metabolic syndrome as well as diabetes in patients who currently do not have the disease. There are some different data mining methods which are used in clinical trials, discussed in next section.

Data mining techniques

Data mining is not about fetching new data. Instead, it is about drawing conclusions from the patterns and new knowledge from data which is already collected or recorded. There are seven different techniques of data mining used in general which are highly effective; and which are set out below. Different data mining methods use different types of techniques from those listed below [11].

Tracking patterns: Tracking patterns is one of the simple techniques used to understand and find out patterns in the concerned dataset. This is usually recognition of the unusualness of data occurring at regular intervals.

Classification: Classification is a comparatively difficult data mining technique that requires collection of various aspects together into obvious categories, which can be further used to draw conclusions.

Association: Association is related to following patterns but is more specific to dependently related variables. In this type of technique, specific events or attributes which are extremely connected with another event are looked into.

Outlier detection: In this type of technique, we don't only look into patterns, but we also look into the outliers in the data.

Clustering: Clustering is like classification. However, it involves grouping portions of data organised collectively on the basis of their similarities.

Regression: It is a form of arranging and demonstrating the possibility of certain variables in the presence of other variables.

Prediction: This is one of the most treasured data mining techniques. It plans for the type of data which will be seen in future, just by recognising the patterns of the past, which are enough to chart a precise forecast.

Having discussed the techniques, this study moves onto how vast data is handled efficiently through Big Data.

Handling enormous data efficiently through Big Data

Big Data is a system which provides the opportunity to handle vast unstructured data and convert it into structured data to analyse it. Implementing Big Data to handle machine learning or artificial intelligence models which require vast amounts of data is the most efficient way of handling and analysing data.

However, professionals involved with clinical data such as healthcare professionals will need to be trained on such platforms as Big Data and Artificial Intelligence. Big data requires programming language, and for that some professional needs to be trained on that first in order to proceed with analysing the data via the Big Data analytics method. Once a person knows how to use big data, data analytics gets much easier and quicker. Therefore, the clinical industry must have this incorporated as a mandatory process to be followed. Amongst the techniques for data mining mentioned above, a few which are used for Big Data, are tracking patterns, classification, association, clustering and prediction [12]. There has been a significant development and popularity observed in the Big Data datasets in many countries. This is attributable to the following below:

- Healthcare is collecting increased amounts of data from electronic health records, home monitoring and biomedical research.
- Developments in technology and processes for data administration and the fresher analytics that are being developed in private divisions allow us to collect, recover, examine and remove knowledge from Big Data groups
- Greater datasets can be brought into use through Big Data after attaining a higher understanding of the diseases.

Apart from Big Data, Naïve Bayes Classification is also used to manage diabetes clinical data as discussed below.

Naïve Bayes Classification Model in Diabetes research

Studies have already been performed on different data analysis and data mining methods in clinical trials related to Type 2 diabetes. These researches were done to see which analysis model is the most effective and reliable. The Naïve Bayes model utilises the classic Naïve Bayes algorithm that assumes that all of the factors are independent of one another [13].

Bayesian algorithm is a graphical model that encodes probabilistic relationships with the variables of interest. Many papers have been published and many researches have been done on the prediction of diabetes using the Bayesian algorithm, for example Kumari, Vohra and Arora [14] in their "Prediction of diabetes using Bayesian network" on data mining methods that help in predicting diabetes. The study was based on experimental results. The accuracy of the Bayesian algorithm was calculated, and it showed that accuracy was 99.51% which is quite high.

In the experiments, data on diabetic patients was collected from different hospitals. Some pre-processing was utilised to produce good quality data. In pre-processing, attributes are identified and selected from the data collected, and then normalisation of the data and finally numerical discretisation occur. However, there were many challenging experiences when experiments were conducted to determine the best method to predict diabetes and how to manage diabetes cases effectively. The researcher assumes that using smart technology could be helpful to minimise this type of occurrence.

Smart technology used for data collection for diabetes

There are devices available in the market that can help in the self-management of diabetes. Such devices are fitted with sensors that help in collecting data from the patient and sending it to the healthcare providers. These existing devices are being worked upon by researchers for mass manufacturing like special shoes or socks that can detect blood supply. Other similar devices are smart contact lenses or smart applications within mobile phone devices.

However, this technology is not used properly by users and that is why it becomes difficult to collect proper data for diabetes management. Incomplete data cannot provide proper analysis results. Machine learning has also gained a lot of attention when it was used as part of a system as summarised in the next section.

Machine learning in diabetes management

There are the instances where machine learning has been brought into use especially in diabetic studies and has proved helpful to diabetic patients. Advanced Bolus Calculator has been developed to capture the variations in blood glucose levels linked to lifestyle habits i.e., exercise or stress. The Advanced Bolus Calculator supported by case-based reasoning will do the personalised bolus recommendations once it adapts to a person's physiology with repeated use. This uses an "ABC4D" device in order to record everyday data to analyse lifestyle habits. Advanced Bolus Calculator is one of the methods of machine learning used for diabetes management **(Figure 2)**.
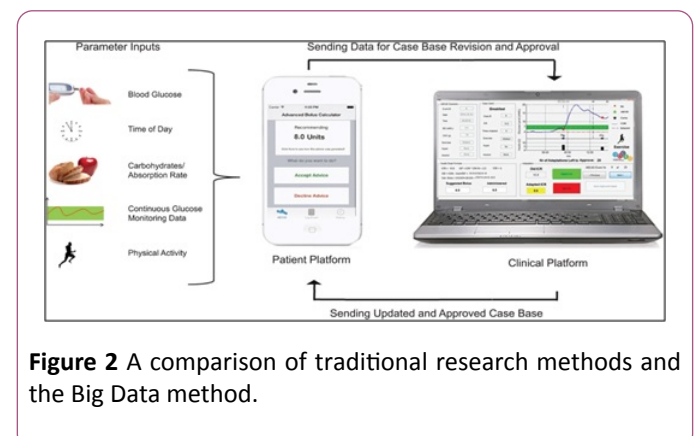

Figure 2 A comparison of traditional research methods and the Big Data method.

An Advanced Bolus Calculator for Type 1 diabetes: System Archtecture and Usability Results [2]. Machine learning signifies a shift in the clinical pattern from the same old fixed management approaches to a data driven precision medicine. For instance, the "Look Ahead" trial determined that decreasing weight does not reduce heart related risks in overweight people with Type 2 diabetes. However, re-analysis on machine learning recognised a subgroup of participants whose weight loss was indeed proven beneficial, and also found that this effect was being obscured by another subgroup of harmful effects [15-19].

Research Methodology

This research uses both secondary and primary research for data collection. Secondary research such as books or e-books, webpages, journals, and newspaper articles regarding the clinical trials and clinical data management for diabetic patients was used. In primary research, the study used a survey questionnaire and personal interview for data collection while opinions of some experts in the clinical trial industry were noted.

A survey questionnaire was designed to achieve the objective of the study:

- To determine and compare traditional research methods and the Big Data method used in clinical trials for diabetes.
- To assess the naïve Bayes classification and other data mining methods that can be used in clinical trials.

Interviews were conducted with employees from the clinical research industry to get more details on the trends followed and how data analysis is done in companies nowadays. Interviews were carried out to support the data collected through the survey questionnaire. During the interview process, the researcher discussed with the experts on the Naïve Bayes classification and other data mining methods and whether their companies were using this method in their clinical trials. The reason for choosing this question for the interview was that this way we will get an understanding of data analysis methods used in clinical research, and we will get to know if people are aware of the Big Data and Naïve Bayes classifications.

Ethical Considerations

Ethical approval for the research was issued by Nelson Marlborough Institute of Technology. All ethical considerations relating to participants were taken care of properly. The research considered the following basic ethical considerations regarding participants who were interviewed: Participants were requested to sign a participant information sheet voluntarily and a consent form before the interviews was conducted. Participants were advised of their rights to privacy and confidentiality. No major risk was associated with the interviews. Interviews were conducted with minimal or no risk. The researcher adjusted her style to suit each participant's personality and emotional state. Data was collected from

participants with integrity, and all secondary research work utilised in this report has been properly cited and referenced.

Interpretation/Analysis

Seven questions shown in the table below are related basically to ascertaining participant's awareness of Naïve Bayes or Big Data, if any improvements are foreseen in the current data mining methods used in the companies or if any company is currently using Big Data or the Naïve Bayes methods of analysis etc. **(Table 2)**.

Table 2 Ascertaining which data mining method is used in companies.

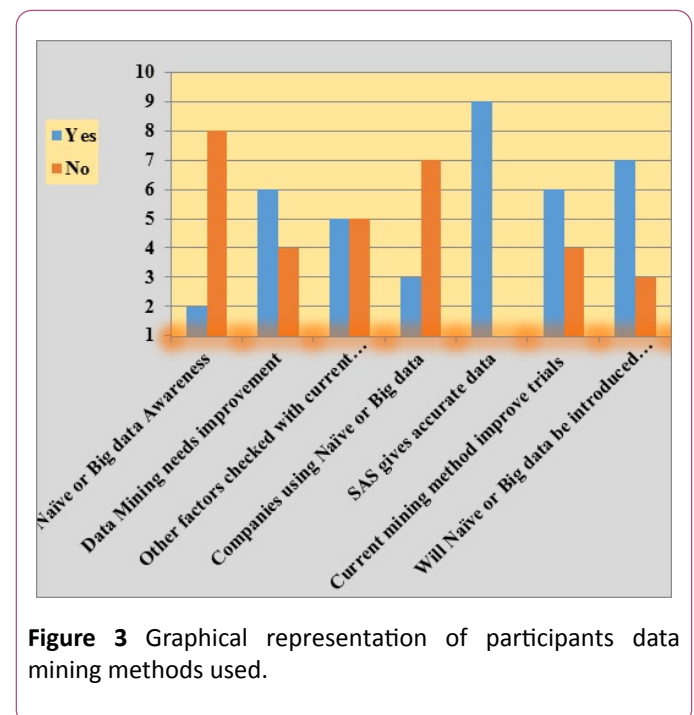| Questions	Yes	No
Naïve or Big Data awareness	2	8
Data mining needs improvement	6	4
Other factors checked with current data mining method	5	5
Companies using Naïve or Big Data	3	7
SAS gives accurate data	9	1
Current data mining method to improve trials	6	4
Will Naïve or Big Data be introduced in company	7	3

Figure 3 Graphical representation of participants data mining methods used.

As can be seen from **Figure 3**, out of 10 questions from the questionnaire, there were seven questions having response type as Yes or No. This makes it easier for a participant to answer and it is easier to analyse such data in graphical format. The questionnaire consists of questions that are basically about getting information on which data mining methods are used in companies.

From **Figure 2**, very few companies are aware of how the Naïve Bayes classification or Big Data can be used in clinical trials. Only 20% of people out of all participants were aware of these data mining methods. However, 60% of people agreed that they need an improvement in the current analysis or data mining method used. On the other hand, 90% of people using the Statistical Analysis System (SAS) tool believed that SAS was giving them accurate data with minimal time consumption. Around 40% believed that the current data mining method used within their company was not contributing to improving clinical trials. Rather, they felt that the analytical method used in their company contributed only towards meeting the protocol requirements as per the project for the sample size considered and collected data accordingly. However, 70% of people felt that Naïve Bayes or Big Data must be introduced into their current company to learn something new and to contribute to making clinical trials more effective and efficient.

Current trend of data mining

The doughnut style graph **(Figure 4)** is designed for one of the questions in the questionnaire, which required the participant to answer which data mining method they were using or had used. Therefore, based on the responses from participants, it was found that 70% participants from the industry were using the Statistical Analysis System (SAS) tool whereas Naïve and Big Data each were used by 10% of participants. It was also noted that the Extraction-Transformation Loading (ETL) method was also being used by 20% of participants **(Table 3)**.
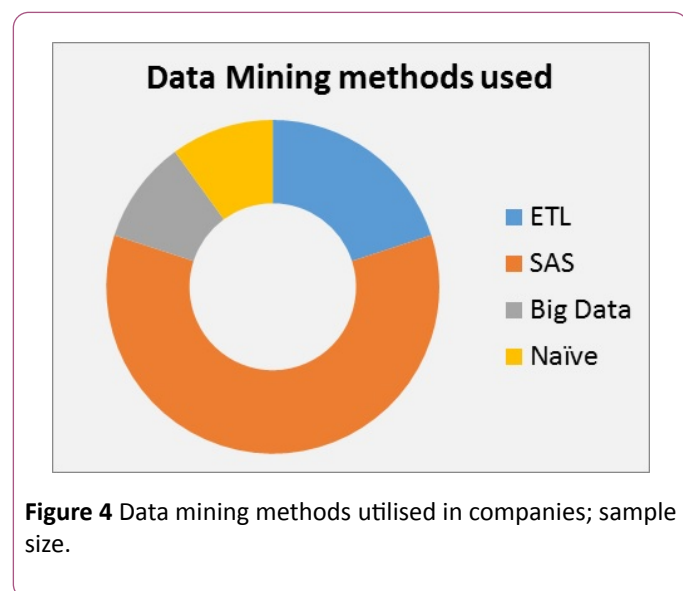

Figure 4 Data mining methods utilised in companies; sample size.

Table 3 Provides information about how many people used a data mining method.

Current data mining method used	Number of people using
ETL	2
SAS	6
Big Data	1
Naïve	1

Out of the 10 questions, there were two questions which had response options such as Agree or Strongly Agree or Neutral or Disagree or Strongly Disagree. Therefore, a different graph depicts the responses received from participants. In this, the two questions relate to whether people would like to consider other data mining methods or not and whether their current data mining methods minimise the time required to process data. Therefore, as per the responses received and shown in **Figure 5** it emerged that four out of 10 people "Agreed" to the question that they would like to consider other data mining methods as well and two out of 10 people "Strongly agreed" with this. On the other hand, four out of 10 disagreed with the same. For the other question for checking time consumption of current data mining methods, five out of 10 people agreed that their current data mining method took minimal time and two out of 10 strongly agreed with it, whereas, three out of 10 disagreed that their current data mining method took minimal time in analysis.
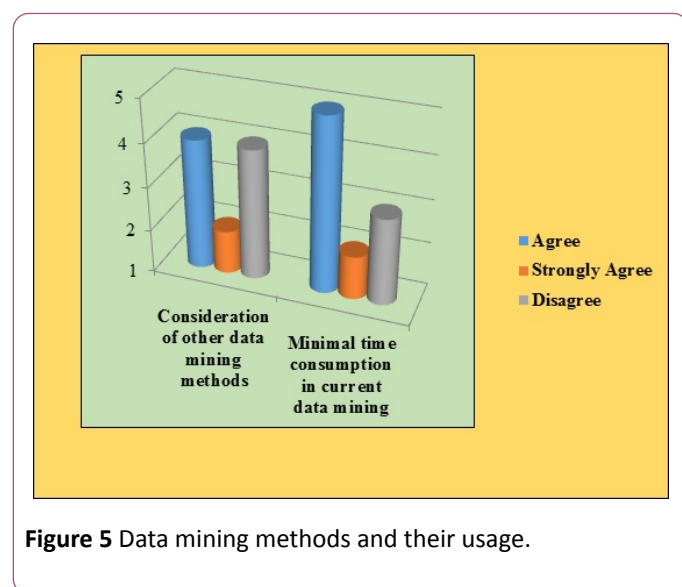

Figure 5 Data mining methods and their usage.

Findings and Observations

This study reveals that bringing Big Data and other data mining methods into use will change the clinical data manager's role. Based on the literature as discussed, it can be said that clinical data managers will have their roles changed to clinical data scientists or analysts soon. In place of just managing clinical data, data scientists will have to observe patterns in the clinical data that could lead to new assumptions and findings and lead to better analysis of the data using such data mining methods. Additionally, using different data mining methods discussed in this study can enhance the capability of getting new and verifiable indications of hidden links between different causes and complications of a disease. However, apart from clinical data scientists and data managers, it is also one of main of responsibilities of other researchers involved to take care of the complications that patients face in day to day life as patients tend to develop other complications apart from

diabetes, especially when they are getting older. Therefore, the researchers involved must investigate such complications and try to include processes and procedures which could eliminate or at least minimise them.

Some of the complications which have been observed in older people are that they develop other health related problems along with having diabetes such as weight management, improper diet and irregularity in medications consumption etc. Normally, it is people above 65 years of age who start developing signs of other health related problems. They are not able to eat right or proper nutrition, or not doing any workout or exercise, and not checking their blood glucose levels by themselves or taking medications on time as prescribed. These complications happen with older people when they are living alone and do not have any helping hand for such things. Some other noteworthy complications concern the sleep cycle with people with low glycaemic control. They develop impaired sleep quality (PSQI). Also, a poor attitude towards diet management, food choices and a decrease in the positive approach to managing diabetes and self-reporting are linked to the bad sleep cycles of these patients.

The above issues and observations can be dealt with through proper diabetes management which can be done with collaborative support from researchers involved like physicians, nurses, researchers, and data scientists etc. Many techniques or methods have been developed to fetch data easily about the patient without their having to submit the same to the doctors or nurses or data scientists. For example, machine learning is one of the techniques which utilises smart devices to capture data such as heart beats, oxygen level, vital signs etc. of the patient, and these devices are integrated in the database system where all the records are maintained and are available to the physicians or nurses to analyse. Therefore, such enormous data collected can be evaluated, analysed and estimated for any signs of improvement or decline in the health of the patient. Such data is not manually reviewed, analysed or evaluated. Rather, data mining methods should be further used to handle such vast data. Therefore, Big Data comes into play at this point and supports the analysis and prediction model of this data. As discussed previously in this report, there are around five types of data mining techniques which are utilised in the Big Data analysis which covers up if any improvement or decline in the health of patient is observed or if a healthy person is developing any signs or symptoms of getting any disease including diabetes. However, this study has helped us to understand what existing data mining methods available and which ones are used amongst the clinical research companies to analyse the data.

Conclusion

This report was completed by doing a survey questionnaire in order to ascertain current trends of data mining methods being used by different companies, and whether some effective procedures or processes are required for use in clinical trials for effective diabetes management, as it is difficult at times to assess the reasons for a particular outcome

which prevents complete eradication of the disease due to lack of effective tools to assess the data recorded in such trials.

This research was done as the number of cases of diabetes, especially Type 2 diabetes has increased recently as Type 2 diabetes accounts for nearly 90% of all cases of diabetes. Currently there are trials which are still running in the same manner as they used to years back, and a gap was predicted in the data mining methods or analytical methods being used in clinical research. Therefore, this study was done to evaluate all data mining methods currently being used. To meet the research requirement, different data mining methods were studied to assess the effectiveness of each one of them.

This research evaluated a lot of data mining methods, but it is concluded that Big Data proves to be very efficient and effective as it can handle structured and unstructured data like medical records, diagnostic images etc. and interpret its data during the analysis step. Therefore, Big Data can be very useful in diagnosis, or prediction of early signs of diabetes. The researcher believes if Big Data is made a standard data mining method across all trials, it will be much easier to collect, evaluate and analyse data and improve the quality of clinical trials all over the world.

References

1. https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=11539392

2. Pesl P, Herrero P, Reddy M, Xenou M, Oliver N, et al. (2016) An advanced bolus calculator for type 1 diabetes: A system architecture and usability results. IEEE J Biomed Health Inform 20(1): 11-17.

3. https://bpac.org.nz/2018/diabetes.aspx

4. http://www.liggins.auckland.ac.nz/en/about/news/news-2018/type-2-diabetes-slowly-rising-in-auckland-kids-pacific-maori-highest-rates.html

5. Sjardin N, Reed P, Albert B, Mouat F, Carter PJ, et al. (2018) Increasing incidence of type 2 diabetes in New Zealand children <15 years of age in a regional-based diabetes service, Auckland, New Zealand. J Paediatr Child Health. 2018 Apr 24.

6. Wang L, Alexander CA (2016) Big data analytics for medication management in diabetes mellitus IJSN 1(1): 42.

7. Chan RS, Woo J (2010) Prevention of overweight and obesity: how effective is the current public health approach. Int J Environ Res Public Health 7(3): 765-783.

8. Poon AK, Juraschek SP, Ballantyne CM, Steffes MW, Selvin E (2014) Comparative associations of diabetes risk factors with five measures of hyperglycemia. Int J Environ Res Public Health 2(1): 000002.

9. Yakaryılmaz FD, Öztürk ZA (2017) Treatment of type 2 diabetes mellitus in the elderly. World J Diabetes 8(6): 278.

10. Lybbert TJ, Wydick B (2017) Hope as aspirations, agency, and pathways: poverty dynamics and microfinance in Oaxaca, Mexico. NBER Work pp: 22661

11. https://www.datasciencecentral.com/profiles/blogs/the-7-most-important-data-mining-techniques

12. https://datafloq.com/read/5-major-data-mining-techniques-being-used-big-data/3352

13. http://cs229.stanford.edu/proj2017/final-reports/5244347.pdf

14. Kumari M, Vohra R, Arora A (2014) Prediction of diabetes using Bayesian Network. IJCSIT 5(4): 5174-5178.

15. Buch V, Varughese G, Maruthappu M (2018) Artificial intelligence in diabetes care. Diabet Med 35(4): 495-497.

16. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I (2017) Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 15: 104-116.

17. Klonoff DC (2009) The increasing incidence of diabetes in the 21st century. J Diabetes Sci Technol 3(1): 1-2.

18. Jowitt LM (2016) Gestational diabetes in New Zealand ethnic groups. Integr Mol Med 3(2): 583-589

19. Huang Y, McCullagh P, Black N, Harper R (2007) Feature selection and classification model construction on type 2 diabetic patients' data. Artif Intell Med 41(3): 251-262.