International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

Early Stage Disease Prediction by using BPNN-NB Ensemble Technique

Issac. P.J
Research Scholar,
Department of Computer Science, Rayalaseema University, Kurnool, India.

Allam Appa Rao
Former VC, JNTU, Kakinada, India,
Chairman, Institute of Bioinformatics and Computational Biology, Visakapatanam, India

***Abstract*— Many diseases turn fatal when left untreated and most of the patients are not aware that they have a certain disorder. Hence, detecting disease at an early stage must be important for improving the life span of the affected people. Diabetes and Cancer are two of the most commonly occurring diseases that can be predicted and treated at early stages. In this paper, a novel classification technique is implemented that combines the Back Propagation Neural Network (BPNN) and Naïve Bayes (NB) classifiers for predicting the cancer and diabetes. The proposed technique is analysed for different evaluation metrics like accuracy, precision, recall and false positive rate. However, this algorithm can also be used to predict lots of other diseases by using certain datasets, which will be a future scope of our work.**

Keywords*— *Diseases, Diabetes, Cancer, prediction, BPNN, NB, Ensemble

I. INTRODUCTION

The number of chronic diseases is rising on a daily basis. The increase in the living standards of humans is also increasing the prevalence of chronic diseases [1]. Chronic diseases are a major source of death in countries like the United states, China and India [2]. Therefore, it is necessary to conduct risk assessments for chronic diseases. Collection of Electronic Medical records (EMR) is getting very convenient [3].

Diabetes is a disease that is prevalent among many people among the present generation. Its prevalence is increasing worldwide at a rapid pace. It is a disorder that is linked to the metabolism and takes place when the pancreas is not able to create sufficient insulin. In some cases, a high level of insulin is created and hence it is difficult for the body to control this insulin Huang et al. [4]. This makes it difficult for the people due to its long term complications which is the tissue damage and organ failure placing an additional burden on healthcare system [5]. The prevalence of heart disease among the patients of DM are more prevalent at up to 40 times than those without DM [6]. T2DM causes lots of deaths around the world and the number of diabetic patients have increased significantly [7], [8]. The disease occurs mostly among the older age group; however it is also rising among the younger age group. It not only causes heart diseases, but also leads to strokes, blindness, renal failures, etc. and most of these take place at the initial stage. This shortens the lifespan of affected morbid people. There are different factors that are responsible for the disorder and this includes diet, genes and lifestyle. Even though there is advancement in the healthcare, the diabetes is still on the rise. It is estimated that around 150 million people are affected by this disorder [9]. Hence, it is necessary to control and prevent diabetes at an early stage in order to improve the quality of the life of the patients.

There are many complications that may arise due to diabetes like damage of the heart, eyes, blood vessels and nerves. Diabetic Nephropathy (DN) is one of the major complications of diabetes. This complication has affected around 30% of the patients with type 1 diabetes (T1DM) and around 25 to 40% of the patients who have type 2 diabetes (T2DM) [8]. When a patient has DN, then there is a high level of albumin in the urine continuously and sudden reductions in filtration rates of glomules. There is also an increase in the blood pressure. This threatens the patient's quality of life [10]. Initially, it cannot be observed easily and forms between 10 and 15 years for the T1DM patients. This is not clearly known among the T2DM patients and may vary from person to person [11]. However, when it is detected at a later stage, it is difficult to cure it since the kidneys have already been damaged [12]. Hence, it is necessary to create a prediction model that uses different clinical parameters like gender, BMI and diabetic history. Additionally, gender and genetic data may a play a major role in predicting the DN which may improve the efficiency.

Artificial Intelligence techniques utilize various algorithms for identifying common patterns in large datasets and include machine learning. It can be used for predicting the results of new data based on trained data [13]. These algorithms are a potential tool for predicting and making decisions in lots of applications [14]. Increasing the accuracy of the prediction techniques is the main issue for the prediction models and hence it should also work efficiently for multiple datasets. However, machine learning algorithms require lots of data and hence the size of the dataset must be large for better accuracy. The number of available datasets obtained from different EMR are extensive and they are available for various types of diseases [15]. They have been used to predict various diseases from the literature. There are lots of machine learning techniques that have been used for predicting the diseases which are Support Vector Machines (SVM), Association rule mining, Naïve Bayes, Neural Networks, Random forest tree, Logistical Regression, etc. These methods may be used as a hybrid technique with other techniques for better accuracy

A bio-inspired heterogeneous technique has been presented in [16], where the data is collected from EMR in real time. An efficient algorithm has been studied in Qiu & Sha. [17] for minimizing the cost of predicting the prediction. The data from EMR is obtained and potential solutions have been given. An efficient flow estimating algorithm has been

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

proposed by Wang et al. [18] for storing and use the data of EMR for disease prediction.

There have been multiple attempts at constructing the prediction models for Diabetes Mellitus (DM), which have been done previously using SVM and other statistical methods [19], [20]. This was found to be efficient when single features are used for the evaluation. However, since DM is caused by a combination of different factors, a designed model for one feature may not be suitable for other features. It would also be difficult for the patients to explain their symptoms and take the action to stop. Hence, we need a user friendly technique that used EMR of the patients to read the entire patients' history and obtain the features in order to predict DM at an early stage. There are many ways of detecting these diseases. Since early detection is necessary, novel methods must be developed for detecting the different diseases at early stage.

Another method of obtaining the data from the EMR is through data mining. The data from EMR can be obtained with certain conditions depending on the diseases. It can predict, cluster, associate and recognize the patterns. They have been used extensively in predicting lots of diseases like diabetes related diseases and cancer related diseases.

Dagliati et al [21] have used ML techniques for embedding data mining and combining the different strategies for obtaining knowledge from data. T2DM has been predicted by using data from EMR of EU. Data is taken from approximately 1000 patients. Feature processing techniques like missing values have been performed by using logical regression and random forest tree. An accuracy of 83.8% has been obtained. But classification has not been performed in this study.

Other studies have been reported for different strategies which enable the prior detection of cancer [22]–[25]. They have used the circulation of miRNAs for detecting and identifying cancer. These techniques have lower sensitivity based on the detection stages and different types of cancer like malignant and benign. Different aspects for predicting the results of cancer based on the trails in gene expression have been discussed in [26], [27]. Even though these studies have lots of advantages, they also have limitations in predicting the cancer. Also, large data is required for the validation in these methods.

Makno et al. [28] has used another prediction technique for diabetes and cancer prediction by using time series and logistic regression with data collected from an EMR. Features have been extracted and 22 factors are considered for predicting early stage cancer and DM. An accuracy of 74% has been obtained. Also, Hemodialysis was detected in this paper which is one of the most common causes of diabetes. Zhu et al. [29] has used data mining for predicting the diabetes at an early stage. PCA and logistic regression have been used along with k-means clustering since the latter cannot attain a high accuracy individually. Even though diabetes has been predicted accurately, less features were selected and hence it wouldn't work well for cancer prediction. The datasets have also been changed and tested in other studies for predicting the disease. A prediction algorithm that has utilized k-means clustering in validated data has been used in Patil et al. [30] with an accuracy of 92.38%. Multiple Layer Perception (MLP) has been studied in Ahmad et al. [31] by utilizing neural networks and they have been compared with ID3 and J48. It has been seen from the results that the J48 tree algorithm, worked better with a greater accuracy of 89.3%.

A predictive model has been developed in Park et al. [32] for evaluating the number of women surviving breast cancer. Different classification techniques like ANN, SVM, and others have been compared by using the data from SEER cancer database. The dataset contains more than 150,000 records with 16 main factors [33]. Also, a class variable survivability has been considered which identifies the number of patients that have survived. The most common factors considered are the number of nodes, size of the tumors, the age at which it is diagnosed. The accuracy of the different classifiers has been compared and it was 65% and 61% for ANN and SVM respectively. This is seen to have a low value of accuracy which should be increased in our study. The survival prediction was assessed for lung cancer patients by using ANN in [34]. The different studies show that predicting the diseases is a tedious task, especially when different datasets are used. This study aims to create a prediction model that aims to increase the existing metrics especially accuracy. Both cancer and diabetes will be predicted by using two different datasets.

II. METHODOLOGY

In this paper, disease prediction is performed by using a novel machine learning algorithm. Different diseases can be predicted by using the corresponding dataset. In this work, cancer and diabetes are predicted by using the proposed algorithm. Data from EMR is obtained and used as a dataset. The dataset will have the necessary features for disease prediction. However, not all the features will be required for the prediction. Hence, the data must be cleaned and processed before using it. After pre-processing, classification is done, which looks into the features and predicts them accordingly. This paper has used a hybrid classification technique that combines Naïve Bayes and BPNN algorithm. The classification is performed and evaluation metrics are obtained for comparison.

A. Dataset

The dataset is initially selected. The data is collected from a large hospital or multiple small hospitals. This data is compiled into a dataset, since uniformity must be maintained only the common features are taken into account. Our dataset contained 5 years data obtained from EMR. The features considered for diabetes are level of glucose, blood pressure, history of pregnancy, thickness of the skin, insulin percentage, BMI, age, etc. There are some common features in the cancer dataset. However, some other features that are available in the cancer dataset are other carcinogenic parameters. The data set will be split into a train set and a test set. Around 70% of the data will be sent to the former and 30% to the latter.

The list of features in both the datasets have been shown in the tables. These features change depending upon the type of diseases. The breast cancer dataset contains most the average values of ten components related to the dimensions of the breast, whereas the diabetes dataset uses different features like the vital features for the patients

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

TABLE 1: LIST OF FEATURES AVAILABLE IN THE CANCER DATASET.

Mean SE Worst	Radius
	Texture
	Perimeter
	Area
	Smoothness
	Compactness
	Concavity
	Concave points
	Symmetry
	Fractal Dimension

TABLE 2: LIST OF FEATURES AVAILABLE IN THE DIABETES DATASET.

Pregnancies
Glucose
Blood Pressure
Skin Thickness
Insulin
BMI
Diabetes Pedigree Function
Age

B. Pre-Processing

The data is now pre-processed with a combination of different techniques. The text in the dataset must be cleaned and processed before getting classified. Noises are first removed in the document for correcting the undesirable data. To remove the noises, linear transform is performed which is a type of statistical transform. This will make the data distribution more significant to the classifier. Since there will be many unnecessary features, only the necessary features have to be extracted in order to predict more efficiently. Principle Component Analysis (PCA) is used in this work for identifying the significant features and extracting them. The data is initially scaled since some of the features may have more data than the others. Hence, the features with less data will be scaled so that equal representation will be given to all the features. After extracting the necessary data, the unnecessary data is removed and replaced with significant data by using normalizing.

C. Classification

The extracted data has to be classified for more efficient prediction. Once the pre-processing technique is completed, classification must be performed. Hence, a hybrid technique which combines BPNN and Naive Bayes is utilized. Also, a gradient boosting method is used for improving the quality of each iteration. The training data is used for training the dataset. The two optimization techniques are combined and this combination is trained by the dataset. After the training process, the pre-processed and feature extracted data is fed into the classifiers for efficient prediction.

Python is used in this work for the implementation process. A combination of BPNN and Naïve Bayes classifiers along with PCA as a pre-processing technique has been used in this work. The following metrics will be calculated in this work. The accuracy has to be calculated where it is the percentage of the true and false works detected correctly. The accuracy must be as high as possible. The other metrics that will be calculated are the true positives, true negatives, false positives, and false negatives. With these, precision and recall are also calculated.

D. Results

The algorithm is implemented using python programming and the following results are obtained. The dataset is split into training and testing dataset with training dataset being larger than the testing dataset at a 7:3 ratio. Here two datasets are utilized, one each for diabetes and cancer. While the cancer dataset has totally 788 values, the cancer dataset contains 570 values. 70% has been taken for training the dataset while 30% has been taken for testing. The number of values allotted for training and testing individually is shown in table 3.

TABLE 3: TRAINING AND TESTING DATA SIZE

	Cancer	Diabetes
Training	399	537
Testing	171	231

The obtained results are tabulated in the work in form of confusion matrix which contains the four values of TP, FP, TN, FN. In the diabetic data using the proposed technique, 203 values out of 231 have been classified correctly. In this, 63 diabetic patients and 140 non diabetic has been classified correctly. On the other hand, 17 diabetic patients have been classified as non-diabetic and 11 non diabetic patients have been classified as diabetic. It has an accuracy of 88%, precision of 78% and recall of 31%. Meanwhile, for the cancer dataset, 156 values out of 171 have been classified correctly. In this, 100 cancer patients and 56 normal has been classified correctly. On the other hand, 8 cancer patients have been classified as being normal and 11 normal people have been classified as having cancer. It has an accuracy of 91.23%, precision of 93.46% and recall of 92.99%. The confusion matrix for the cancer and diabetes dataset is shown in table 4and 5.

TABLE 4: CONFUSION MATRIX – CANCER DATA

Index	**Yes**	**No**
Yes	100	8
No	7	56

TABLE 5: CONFUSION MATRIX – DIABETES DATA

Index	Yes	No
Yes	63	17
No	11	140

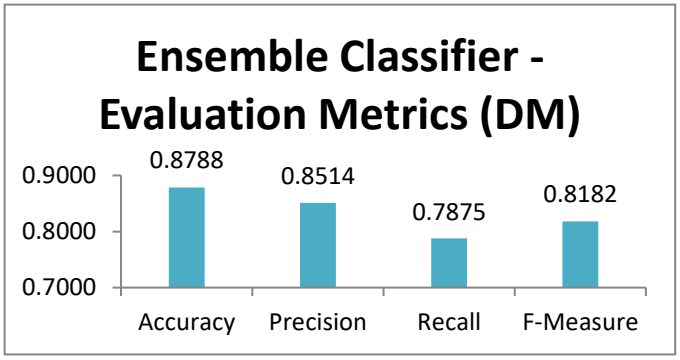

FIG. 1: EVALUATION METRICS FOR DIABETES

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

The accuracy, Precision, Recall and F-measure for diabetes is given in figure 1. Similarly, the same parameters for cancer prediction is shown in figure 2.

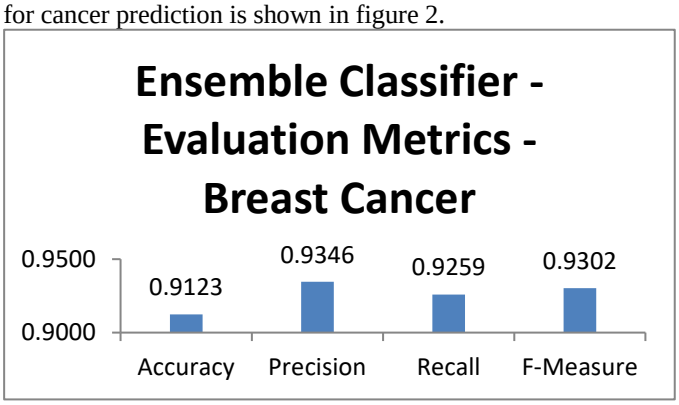

FIG. 2: EVALUATION METRICS OF BREAST CANCER

Accuracy is percentage of values that are classified correctly. The accuracy of the performed system is also measured separately for each technique and also for the proposed combined technique for both the diseases. For the diabetes dataset, the accuracy of the individual BPNN is 86%, whereas it is 76% for Naïve Bayes. When the technique is combined into an ensemble classifier, the accuracy is increased to 88%. Similarly, it is 94% for individual BPNN, and 0.96 for Naïve Bayes, the combined technique has an accuracy of 95%. In spite of having a smaller training set for the cancer dataset, the accuracy is higher, since the number of features in the dataset is high. The values are tabulated in table 4 and plotted in fig 1

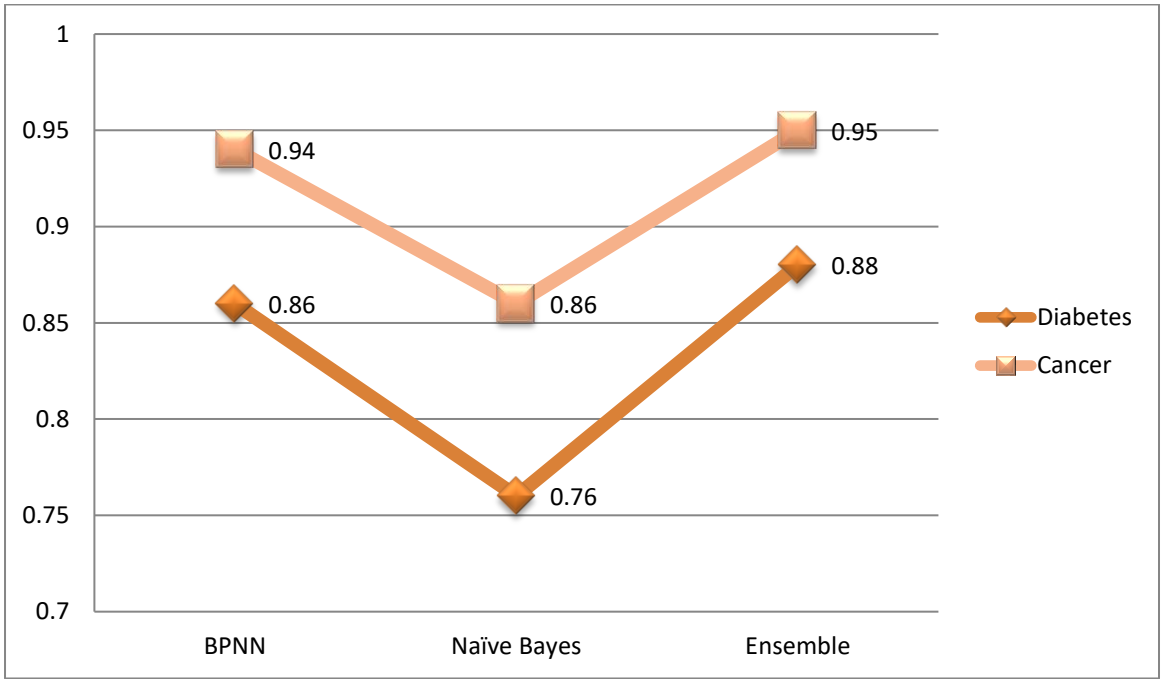

FIG. 3: ACCURACY OF THE IMPLEMENTED TECHNIQUE

TABLE 6: ACCURACY OF THE IMPLEMENTED TECHNIQUE

Technique	Diabetes	Cancer
BPNN	0.86	0.94
Naïve Bayes	0.76	0.86
Ensemble	0.88	0.95

It can be seen from the table that the accuracy of the BPNN algorithm is high when it is used individually. Comparatively, the Naïve Bayes classifier has a lesser accuracy. The combined ensemble technique has an overall higher accuracy in both the datasets.

TABLE 7: ACCURACY OF PREVIOUSLY OBTAINED TECHNIQUES [35]

Classifiers	Accuracy	Source
Logistical Regression	77%	Detrano et al. [7]
Naïve Bayes	81.48%	Cheung [12]
Fuzzy, KNN, AIRS	87%	Polat et al. [8]
GA – AWIS	87%	Ozsen and Gunes [9]
Decision Tree Techniques	84.1	Shouman et al. [15]
AIS	84.5%	Polat et al. [13]

The accuracy of the conventional systems are shown in table 7. Each researcher used various techniques and have obtained different sets of accuracy. The conventional algorithms have been compared with the proposed technique and shown in figure 4. It can be seen that the proposed technique is superior to the existing technique in terms of accuracy.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

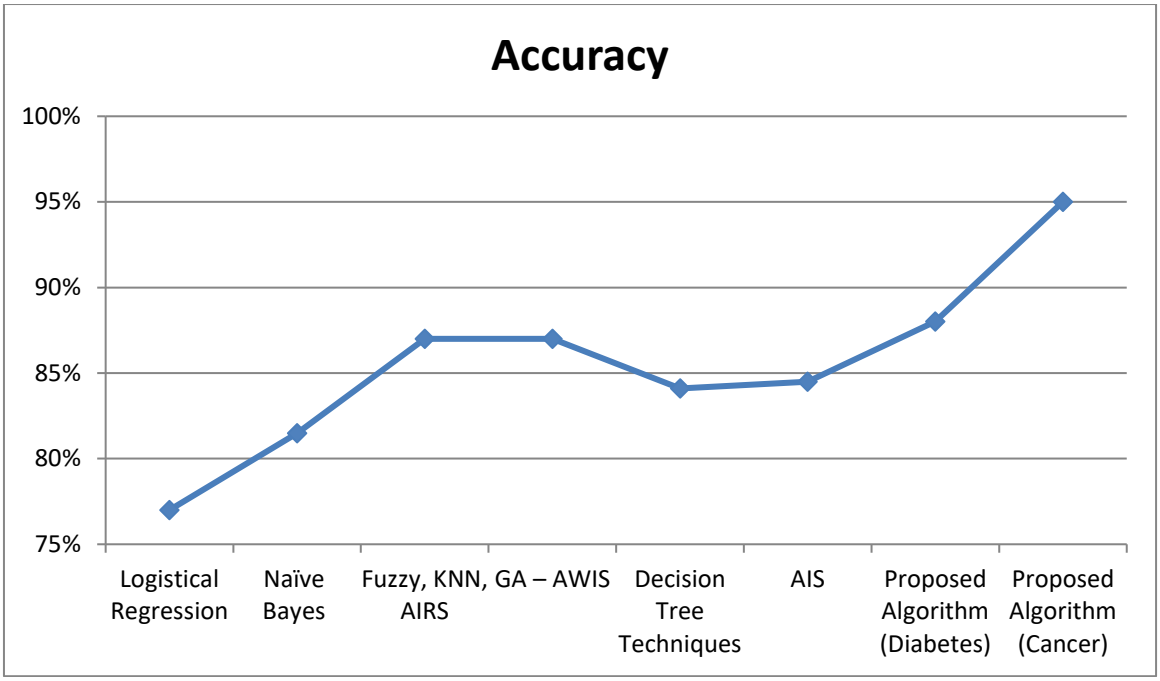

FIG. 4. COMPARISON OF ACCURACY WITH EXISTING TECHNIQUES.

CONCLUSION

The classifier combines the two individual classifiers namely, Naïve Bayes and BPNN for calculating the maximum accuracy, prediction, recall, F-measure, error rate, specificity, and false rate. A comparative analysis with each classifier was carried out for cancer and diabetes. The number of correct classification made by Naïve Bayes' prediction model was 192 and misclassifications were 39 for DM prediction for "n=462". By comparing the false predictions and the various performance indices across all the three classifiers, it was proved that the proposed ensemble classifier provided an accurate and better prediction compared to NB and BPNN classifiers with and without scaling the data. The range of improvement made by the proposed ensemble classifier for diabetes was between 4.09% and 10.53%. Similarly, the prediction was performed for breast cancer. The correct classifications made by Naïve Bayes' model were 144 and the false predictions were found to be 27. The total predictions made were 171 and had an accuracy of 91.23%, 93.46% precision, and 92.59% recall. Comparatively, cancer had a better prediction than diabetes. This might be due to the difference in the size of the datasets; however, a high accuracy has been obtained in this work.

REFERENCES

[1] P. Groves, B. Kayyali, D. Knott, and S. Van Kuiken, "The 'big data' revolution in healthcare. Accelerating value and innovation," Jan. 2013.

[2] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: towards better research applications and clinical care," *Nat. Rev. Genet.*, vol. 13, no. 6, pp. 395–405, Jun. 2012.

[3] I. Segura-Bedmar, C. Colón-Ruíz, M. Á. Tejedor-Alonso, and M. Moro-Moro, "Predicting of anaphylaxis in big data EMR by exploring machine learning approaches," *J. Biomed. Inform.*, vol. 87, pp. 50–59, Nov. 2018.

[4] G. M. Huang, K. Y. Huang, T. Y. Lee, and J. T. Y. Weng, "An interpretable rule-based diagnostic classification of diabetic nephropathy among type 2 diabetes patients," *BMC Bioinformatics*, 2015.

[5] D. A. IDF, "Idf Diabetes Atlas - 8th Edition," 2019. [Online]. Available: http://www.diabetesatlas.org/. [Accessed: 31-Jan-2019].

[6] S. Thomas and J. Karalliedde, "Diabetic nephropathy," *Medicine (Baltimore).*, vol. 47, no. 2, pp. 86–91, Feb. 2019.

[7] The Statistics Portal, "Diabetes - Statistics and Facts," 2019. [Online]. Available: https://www.statista.com/topics/1723/diabetes/. [Accessed: 29-Apr-2019].

[8] World Health Organization, "Global Status Report On Noncommunicable Diseases," 2014. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/148114/9789241564854_eng.pdf?sequence=1. [Accessed: 31-Jan-2019].

[9] G. Danaei *et al.*, "National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2·7 million participants," *Lancet*, vol. 378, no. 9785, pp. 31–40, Jul. 2011.

[10] C. Y. Hong and K. S. Chia, "Markers of diabetic nephropathy.," *J. Diabetes Complications*, vol. 12, no. 1, pp. 43–60, 1998.

[11] A. Ntemka, F. Iliadis, N. Papanikolaou, and D. Grekas, "Network-centric Analysis of Genetic Predisposition in Diabetic Nephropathy.," *Hippokratia*, vol. 15, no. 3, pp. 232–7, Jul. 2011.

[12] F. Sharifiaghdas, A. H. Kashi, and R. Eshratkhah, "Evaluating percutaneous nephrolithotomy-induced kidney damage by measuring urinary concentrations of β2-microglobulin.," *Urol. J.*, vol. 8, no. 4, pp. 277–82, 2011.

[13] M. Motwani *et al.*, "Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis.," *Eur. Heart J.*, vol. 38, no. 7, pp. 500–507, Feb. 2017.

[14] A. Chandiok and D. K. Chaturvedi, "Machine learning techniques for cognitive decision making," in *2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI)*, 2015, pp. 1–6.

[15] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Heal. Inf. Sci. Syst.*, vol. 2, no. 1, p. 3, Dec. 2014.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

[16] D. Tian, J. Zhou, Y. Wang, Y. Lu, H. Xia, and Z. Yi, "A Dynamic and Self-Adaptive Network Selection Method for Multimode Communications in Heterogeneous Vehicular Telematics," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3033–3049, Dec. 2015.

[17] M. Qiu and E. H.-M. Sha, "Cost minimization while satisfying hard/soft timing constraints for heterogeneous embedded systems," *ACM Trans. Des. Autom. Electron. Syst.*, vol. 14, no. 2, pp. 1–30, Mar. 2009.

[18] J. Wang, M. Qiu, and B. Guo, "Enabling real-time information service on telehealth system over cloud-based big data platform," *J. Syst. Archit.*, vol. 72, pp. 69–79, Jan. 2017.

[19] B. H. Cho, H. Yu, K.-W. Kim, T. H. Kim, I. Y. Kim, and S. I. Kim, "Application of irregular and unbalanced data to predict diabetic nephropathy using visualization and feature selection methods.," *Artif. Intell. Med.*, vol. 42, no. 1, pp. 37–53, Jan. 2008.

[20] R. K. K. Leung *et al.*, "Using a multi-staged strategy based on machine learning and mathematical modeling to predict genotype-phenotype risk patterns in diabetic kidney disease: a prospective case-control cohort analysis.," *BMC Nephrol.*, vol. 14, p. 162, Jul. 2013.

[21] A. Dagliati *et al.*, "Machine Learning Methods to Predict Diabetes Complications," *J. Diabetes Sci. Technol.*, 2018.

[22] O. Fortunato *et al.*, "Assessment of Circulating microRNAs in Plasma of Lung Cancer Patients," *Molecules*, vol. 19, no. 3, pp. 3038–3054, Mar. 2014.

[23] H. M. Heneghan, N. Miller, and M. J. Kerin, "MiRNAs as biomarkers and therapeutic targets in cancer☆," *Curr. Opin. Pharmacol.*, vol. 10, no. 5, pp. 543–550, Oct. 2010.

[24] D. Madhavan, K. Cuk, B. Burwinkel, and R. Yang, "Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures," *Front. Genet.*, vol. 4, no. 116, 2013.

[25] K. Zen and C.-Y. Zhang, "Circulating MicroRNAs: a novel class of biomarkers to diagnose and monitor human cancers," *Med. Res. Rev.*, vol. 32, no. 2, pp. 326–348, Mar. 2012.

[26] S. Koscielny, "Why Most Gene Expression Signatures of Tumors Have Not Been Useful in the Clinic," *Sci. Transl. Med.*, vol. 2, no. 14, p. 14ps2-14ps2, Jan. 2010.

[27] S. Michiels, S. Koscielny, and C. Hill, "Prediction of cancer outcome with microarrays: a multiple random validation strategy.," *Lancet (London, England)*, vol. 365, no. 9458, pp. 488–492, 2005.

[28] M. Makino *et al.*, "Artificial Intelligence Predicts Progress of Diabetic Kidney Disease-Novel Prediction Model Construction with Big Data Machine Learning," *Diabetes*, vol. 67, no. Supplement 1, p. 539–P, May 2018.

[29] C. Zhu, C. U. Idemudia, and W. Feng, "Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques," *Informatics Med. Unlocked*, p. 100179, Apr. 2019.

[30] B. M. Patil, R. C. Joshi, and D. Toshniwal, "Hybrid prediction model for Type-2 diabetic patients," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8102–8108, Dec. 2010.

[31] A. Ahmad, A. Mustapha, E. D. Zahadi, N. Masah, and N. Y. Yahaya, "Comparison between Neural Networks against Decision Tree in Improving Prediction Accuracy for Diabetes Mellitus," in *Digital Information Processing and Communications*, 2011, pp. 537–545.

[32] K. Park, A. Ali, D. Kim, Y. An, M. Kim, and H. Shin, "Robust predictive model for evaluating breast cancer survivability," *Eng. Appl. Artif. Intell.*, vol. 26, no. 9, pp. 2194–2205, Oct. 2013.

[33] N. Howlader *et al.*, *SEER Cancer Statistics Review*. National Cancer Institute. Bethesda, 2012.

[34] Y.-C. Chen, W.-C. Ke, and H.-W. Chiu, "Risk classification of cancer survival using ANN with gene expression data from multiple laboratories," *Comput. Biol. Med.*, vol. 48, pp. 1–7, May 2014.

[35] D. M. Mythili T., N. Padalia, and A. Naidu, "A Heart Disease Prediction Model using SVM-Decision Trees-Logistic Regression (SDL)," *Int. J. Comput. Appl.*, vol. 68, no. 16, pp. 1–5, 2013.