Advances in Parallel Computing Algorithms, Tools and Paradigms
D.J. Hemanth et al. (Eds.)

doi:10.3233/APC220005

Prediction on Type-2 Diabetes Mellitus Using Machine Learning Methods

Soumya K N[a,1] and Vigneshwaran P[b]
[a]*Department of Information Science and Engineering, JAIN (Deemed-to-be-University), Karnataka. India.*
[b]*Department of Networking and Communications, SRM Institute of Science and Technology, Chennai, India.*

Abstract. Diabetes has become one of the most fatal diseases as a result of lifestyle changes, food habits, and decreased physical exercise. Diabetes is believed to afflict 422 million people globally, according to the latest WHO estimates. Having said that, the Type II category of diabetes is more fatal because it is determined by the body's insulin resistance. Furthermore, Type II diabetes has been linked to complications with the kidneys, eyes, and heart. A big number of scientists are also looking into the possibility of a link between diabetes and cancer. We present an overview of such discoveries as well as our cancer research efforts in this report. Dimensionality reduction, Classification, and Clustering are applied in the proposed work to compare with the existing classifiers. PIMA Indian diabetes datasets and Stanford AIM-94 dataset is considered as the benchmark dataset for performing experimentation.

Keywords. Dimensionality Reduction, Machine Learning, Diabetes, Cancer, Cancer risk, PIMA

1. Introduction

They claim that the only constant in the Universe is changing. The socioeconomic growth of emerging nations such as India has resulted in a paradigm change in the lifestyle of its population. This impacts not only the metropolitan population, but also the semi-urban and rural communities. Changes in lifestyle have led in changes in food habits, levels of physical activity and so on. It is easy to remark that this has resulted in what may be referred to as lifestyle diseases, even though many of them are technically disorders, with Diabetes being the most chronic and lethal. Diabetes is an issue, not a disease, characterized by hyperglycemia, which causes damage to the eyes, kidneys, heart, and nervous system. Diabetes is grouped into two categories: type 1 and type 2. In the former situation, the pancreas produces little or no insulin on its own, but in the latter case, the body develops insulin resistance. Governments must ensure that patients have access to low-cost care in order for them to survive. The World Health Organization (WHO) claims that there are around 422 million inhabitants on the globe, primarily in nations with a low and moderate-income, suffering from diabetes, with 1.6 million fatalities directly attributable to cancer. People with Type II diabetes are more

[1]K. N. Soumya, Department of Information Science and Engineering, JAIN University, Karnataka, India; E-mail: kn.soumya@jainuniversity.ac.in.

likely to develop cancer symptoms later in life, according to research. Many scholars from all over the world have previously undertaken extensive research on cancer and diabetes. However, many programmers were developed in silos, making it difficult to assess if diabetes causes or contributes to cancer. With the introduction of High-Performance Computing (HPC), it is now feasible to scale up calculations at a fraction of the cost. In this paper, we explain our efforts to find a convincing explanation for type II diabetes patients getting cancer symptoms, as well as an overview of the efforts made by colleagues throughout the world to find the common cause.

2. Literature Survey

2.1 Diabetes

Diabetes Mellitus is a condition in which the quantity of sugar in the blood cannot be controlled. This metabolic disorder is quite widespread nowadays, either because the body does not produce enough insulin or does not respond to the insulin that is produced. Diabetes affects 37 million people globally, as per the World Health Organization (WHO), and the number is anticipated to more than double in 2030. Diabetes claimed the lives of 50 lakh individuals in 2012. Eighty percent of those killed came from poor and middle-class families. Diabetes affects 5 crores and over in India, and this figure will rise to 7 crores in a few years. India is ranked second in the world. “Insulin-Dependent Diabetes Mellitus” was the initial name for type 1 diabetes. Diabetes can occur at any age, but it must be identified before the age of 20. Insulin-producing cells or beta cells in the pancreas are damaged in this kind of diabetes.

Type 2 diabetes was originally classified as non–insulin-dependent diabetes since it was diagnosed in adults above the age of 20. Gestational diabetes during pregnancy canaries when the pancreas fails to produce the proper quantity of insulin in the body; these three types of diabetes necessitate treatment, and if detected and treated early, early complications can be avoided heading.

2.2 Cancer

Malignancy is an infection described by uncontrolled cell increase and division. While particular sorts of cells create and partition at a remarkable rate, others might develop and isolate all the more leisurely. Each cell in the body has a characterized reason and has a restricted life expectancy. At the point when a cell's life expectancy lapses, it is advised to pass on, just to be supplanted by one more by the body. A carcinoma is described as the obliteration of the solid encompassing tissues which will ultimately spread to different parts.

This kind of cancer by and large starts in the tissues that are nearer to internal or external surfaces of the body. For the most part, this kind of malignant growth happens when the DNA of a cell is either changed or harmed. A few models are prostate, breast cancer, lung, and colorectal malignant growth. A sarcoma for the most part starts from the harmed cells of connective tissue. Extensively talking, this incorporates fat, ligament, bone and so forth. This can be found in any piece of the body. A sarcoma has many related subtypes which are portrayed by the tissue and kind of cell. Sarcoma can be extensively sorted into 2 kinds’ viz. bone and delicate issue.

Leukemia is likewise called a blood malignant growth wherein the capacity of the body to battle contamination is compromised. Leukemia is for the most part connected with issues in delivering blood and typically influences the White Blood Cells (WBC). It is by and large described by a big number of strange platelets. Lymphoma is one more sort of malignancy that begins in the lymphatic framework. It is characterized by lymphocyte alterations and an exceptionally large number of lymphocytes. Lymphoma can be extensively ordered into 2 kinds for example Hodgkin's and non- Hodgkin's. Bo Zhu et al. [1] have dealt with finding the relationship [2-5] among diabetes and colorectal malignancy. Their reliance is on meta-analysis, which reports the influence of diabetes on colorectal guess based on some companion studies and guarantees to have given predicted findings. The additional case to have completed a thorough look on numerous data sets and performed a meta-examination on the obtained data of almost 2 million people; It was investigated the endurance rate and endurance risk that was malignant growth explicit.

The consequences of the review recommend that the diabetic patients will have a decreased life expectancy by no less than 5 years particularly in subjects experiencing colorectal, colon, and rectal malignancy by a huge factor of 18, 19, and 16 percent individually; when compared with that of non-diabetic subjects. They express that a portion of the examinations [6-11] show that the general endurance of colorectal malignancy patients with diabetes mellitus had shown a diminished risk in endurance. Nonetheless, there as on that their meta-investigation [12-14] proposed that diabetes adversely affected the colorectal malignant growth is by and large an endurance perspective.

A review on Diabetes Mellitus and Breast Cancer by Kimberly et al. [15] to play out a fundamental audit and perform meta-investigation to concentrate on the effect of diabetes on breast malignant growth. They have utilized EMBASE and MEDLINE information bases. They have utilized relative terms to diabetes mellitus and malignant growth for information mining. Their hunt incorporated the terms like "glucose prejudice", "hyperglycemia", "cancer", "dangerous neoplasm" and so forth They further express that the 2 examinations performed on the malignancy explicit mortality gave blended out comes and further expressed that Srokowski et al [16] noticed an expanding pattern in breast cancer explicit mortality with prior diabetes mellitus.

These individuals appear to have undergone chemotherapy. He also demonstrated that a greater proportion of women with diabetes mellitus displayed symptoms of advanced breast cancer than non-diabetic patients. The findings also suggested that people with existing diabetic mellitus were at a higher risk of repeated chemotherapy.

A portion of the studies [17-19] in the comparative lines that were led suggested the presence of a positive relationship between breast malignant growth and previous diabetes mellitus, for example, women who had previously been diagnosed with diabetes mellitus were more likely to have breast cancer. While other researchers were conducting research to establish the possible link between cancer and preceding diabetes mellitus disease. Konstantinos et al [20-22] did a review to assess the legitimacy of such a conceivable affiliation and evaluation of hazard of casualty. He utilized the information sources from PubMed, EMBASE, Cochrane Database and afterward played out a meta-investigation across the information to find if the subjects with type II diabetes are at risk of creating malignancy or have a low endurance score. During their review, they tracked down that numerous meta-investigation reports that type II diabetes has a positive connection with a raised risk of different sorts of cancer viz. live, breast, pancreas, and so forth [23-25]. Nonetheless, their assessment has some

misclassification because diabetes evaluation was performed at pattern far before the malignancy analysis, and as a result, they are non-differential. Nonetheless, at the end of their review, they express an admonition that, while they do not deny type II diabetes being related to various types of malignant growth based on different other writing they would have referred to, the results of the studies that show the solid tendency of type II diabetes being related with the arena of creating cancer or, surprisingly, more dreadful passing from cancer, exhibit a light hint of predisposition. They emphasis the importance of forming alliances and adopting improved assessment procedures for type II diabetes in order to achieve more feasible outcomes. Dániel Végh et al. [26] took up an intriguing review on the relationship of type II diabetes with Oral Tumors as Hungary has the biggest number of oral malignant growth and countless sort II diabetes rates. They chose patients who had previously dangerous forms of cancer [27-29]. According to the findings of their review, type II diabetes was found in 25.9 percent of the participants and IFG in 20.6 percent of the subjects who were diagnosed with oral cancer. The findings also suggest that 46.5 percent of the participants diagnosed with oral cancer had ongoing metabolic difficulties. The investigations of the greater part of the analysts demonstrate that type II diabetes is related to a raised risk of malignancy. Additionally, some epidemiological information shows that diabetes affects tumors and all the more particularly so in some hyper nearby malignancies like colorectal, pancreas, and liver.

Hyperinsulinemia, portrayed by a serious level of insulin obstruction is suspected to track down the missing connection. This is because of the speculation that insulin may have a potential mitogenic impact [30-33]. Similarly, some studies suggest that the hyperglycemic condition itself may be cancer-causing since it raises oxidative pressure [34-35]. Hiroshi Noto et al. [36] have been focusing on these notions in order to find a possible link between malignant growth and type II diabetes. Profound Neural Network (DNN) is a design of ANNs that utilizes different layers among input and the yield layers. The advantage of DNNs is that they normally tie down the proper numerical perspective to shift input over to yield regardless of whether the information is linear or non-linear. With the HPC and GPUs turning out to be more open and conservative it has become a lot simpler to use such innovation without any problem. Meng-Hsuen Hsieh et al., have utilized DNN to concentrate on the relationship of type II diabetes with colorectal cancer [37-40].

They have utilized k-overlap cross-approval with the worth of k being 10 as a metric for this situation to decide the KPIs like misfortune capacity of indicator, high effect hyper-boundaries, and so on. Their DNN model comprised of a 37- dimensional information layer and 3 secret layers of 30 measurements every which brought about 1 scalar yield layer. They used stochastic inclination drop [38-40], which they claim to have improved by using sped up angle drop [41-43]. The high dimensionality of the information indicates that the information layers will be densely connected, making it necessary for the enactment capacity to be extremely sensitive and hence Rectified Linear Unit (ReLU) [44] was picked as the actuation work for input and secret layers. They utilize soft max enactment [45] for the yield layer because they are not interested in a twofold characterization. Because exactness was not a reliable measure of the indicator task due to unequal information appropriation, [46-48] they instead used accuracy, affectability, and F1 scores. They stated in the final comments of their review that depending on the DNN model that they had planned, they could see the association of type II diabetes and colorectal malignant development in people who had previously had type II diabetes. Given that the cancer has a wide range of classes and effects; it

appears to be more feasible to conduct the investigation to discover a possible link between type II diabetes and a confined/site-specific cancer. Among many such studies, one that has piqued our interest is the investigation of type II diabetes and breast cancer. P Boyle et al. [49] have conducted tests to investigate if there is a link between type 2 diabetes and breast cancer. They initially focused on the overall danger, and their certainty stretches were deduced empirically, followed by calculating corresponding fluctuation.

Their findings indicate that their meta-analysis includes as many as 40 risk indicators from their writing study. The findings also revealed that in participants with preexisting type II diabetes, the Summary Relative Risk (SRR) was viewed as 1.27 over a given time period and the affectability investigation produced quantitatively insignificant results. When the meta-examination was limited to the imminent gathering of subjects the distinction in the SRR was immaterial and the SRR was 1.23 with a distinction of 0.04. Unexpectedly, when the equivalent meta-examination was done on the review subjects the SRR was altogether higher at 1.36. According to their review, the SRR was non-digressive to the extent that the subjects' menopausal status was concerned about an SRR worth of 0.86; nevertheless, the SRR in post-menopausal subjects was regarded as 1.15, and the difference was actually significant.

To summarize their findings, we can say that they discovered a much stronger link between diabetes and breast cancer [50], but only in postmenopausal [51] women. In one of the most recent studies, a few scientists attempted to use group learning approaches to discover a possible association between diabetes and breast cancer. In this review by Shaboni et al. [52], scientists used criteria such as gravida count, glucose [53-56] concentration, diastolic pulse, and the results of an insulin test observed for 2 hours. They guarantee that their suggested model outperforms Adaptive KNN, the decision tree, and group perception. The increase in risk is comparable to many of these studies on Caucasian populations and additionally suggests that an expanded risk of malignant development in Diabetes Mellites subjects may happen globally with a negligible magnitude for most types of malignant growth. The increase in risk is comparable to many of these studies on Caucasian populations and moreover suggests that an expanded hazard of malignant growth in Diabetes Mellites subjects may happen globally with a negligible magnitude for most types of malignant growth. As he concludes, DM individuals from Asian Population-based studies had a small increase in the risk of most forms of cancer, with middle-aged men having DM having a higher cancer risk. Zidian Xie [57] conducted an exhaustive audit on forecasting models for type 2 diabetes using a variety of AI computations including, SVM, Decision tree, logistic regression, neural network, random forest, and Gaussian Naive Bayes. By examining their forecast, effectiveness on the informative index of the test, In terms of AUC, sensitivity, specificity, and accuracy, our prediction models performed similarly in anticipating DM type 2.

In any case, the neural network prediction model outperformed the others in terms of accuracy, specificity, and AUC. The decision tree prediction model, on the other hand, had the most affectability. The Neural Network model provided the best model execution among the eight predictive models, with the highest AUC esteem; nonetheless, the Decision tree model is preferred for beginning evaluating for Diabetes mellites type 2 since it had the highest sensitivity and, thus, detection rate. He affirmed recently detailed risk factors and distinguished sleeping time and recurrence two additional possible risk factors for type 2 diabetes have been found as a result of studies. In this assessment, pancreatic cancer is possibly the deadliest malignancy because its

initial diagnosis is difficult, and most patients have effectively proceeded to unrespectable circumstances with determination. [58] Menghsuenhsieh two models were created in Taiwan to predict the probability of getting pancreatic cancer in T2DM patients. The data used in this review revealed that the LR model outperformed the ANN model in the prediction of pancreatic cancer.

The findings could help predict pancreatic cancer by observation, early detection, and therapy in people who have certain risk markers. More research from different countries is required to see whether our discoveries are applicable elsewhere. The model compiled risk variables for pancreatic cancer. The Chi-square test and the students' test were used to compare differences between direct and indirect factors. The region under the ROC bend of both forecast models was compared to the optimum value of 1.Understanding the relationship between diabetes and cancer is going to be one of the healthcare community's most difficult tasks. This also implies a better understanding of the role of glucose-lowering medicines. The use of glucose-lowering medication appears promising because it impacts the aetiology of cancer in people with type 2 diabetes, but there are some limitations to consider. J. A. Johnson and B. [59] Carstensen discovered a strong link between liver and pancreatic malignant growths, as well as reverse causality, where cancer caused diabetes. There were no notable discoveries between DM type 1 with malignant growths identifying with DM type 2. Individuals with type 2 diabetes were found to have a higher risk of stomach cancer in both Western and Asian countries.

Hyperglycemia was discovered, followed by hyperinsulinemia, with the latter choice promoting cancer cell proliferation. RCT reports have been debatable because they made the occurrence of malignant look higher, and on the other end, two zeitgeister of preliminary insulin analogues with a long half-life indicating no proof of an increased risk of malignant growth but they were limited by a small sample size of patients for a short term. Observational examinations are preferred, with a longer time span between successive meetings and high-quality information gathering to ensure a legitimate outcome in such complex cases with a large amount of conflicting data. Diabetes Mellitus is a chronic pancreas malignancy characterized by an inability to produce enough metformin.

This increases the concentration of galactose in the blood, which has an effect on a range of important designs such as veins and neurons. NIDDM (Non-Insulin Dependent Diabetes Mellitus) is very inescapable, representing more than 90% of all diabetes cases around the world. Ashrita Kannan, P. Vigneshwaran [60] have investigated Type 2 diabetes which is linked to getting older, being overweight, and having a family background of the illness. Grouping of cancer is finished utilizing the Random Forest. This model finds the optimal weight for the information many times and then fits it to the classification. The prediction approach is used to categories the data and delivers an outcome indicating whether a diabetic patient is likely to get cancer. It categories cancer types such as breast, liver, and colon. The Random Forest computation has a preparing precision of 99 percent. Diabetes can cause a variety of problems, including damage to the heart, veins, eyes, kidneys, and nerves. Modern medicine has had to deal with a massive amount of data collection, analysis, and utilization in order to treat complex clinical concerns. The review by Shahabeddin Abhari and Sharareh R [61] looked into AI algorithms and methodologies for T2DM treatment, with a focus on AI strategies. The assessment assumes that AI-assisted diabetes consideration is a useful technology. According to the findings, many approaches, including FL, ES, NLP, and robots, have not been used to DM type 2

autonomously. Only 6% of distributions used the KB strategy, while 71% used ML and 23% used a combination of AI strategies. Diabetes and cancer have a puzzling association due to their multi factorial nature. Hui Chen [62] presented and described the dataset in this research by dividing it into two collections. A) In this study, DM type 2 patients with linked liver cancer were compared to DM type 2 patients without cancer. B) In a study, T2DM combined liver cancer patients were compared to DM type 2 integrated other cancer patients. Both patients are divided into ten groups, nine of which are used as training sets and one as a test set. A logistic multivariate regression using a step-by-step approach forward conditional method was used for training. The data comes from the National Clinical Medical Science Data Center's Diabetes Dataset (301 Hospital).

3. Dataset used for Experimentation

According to the survey, there are extremely few dataset records available for experimentation. PIMA Indian Dataset, UCI dataset, and Stanford AIM 94 datasets containing seventy patient records are the benchmark datasets used for diabetes prediction. The sections that follow provide an overview of the PIMA Indian dataset and the UCI dataset with 21 attributes.

3.1 Pima Dataset:

Pima Indians diabetes dataset is collected from the National Institute of Diabetes and Digestive and Kidney Diseases. The dataset's primary purpose is to identify patients' health records in order to determine whether a patient is diabetic or not. Table 1 shows the parameters of the PIMA Indian diabetes dataset, which contains 769 records.

Table 1. PIMA Dataset Parameters

S. No	Parameter Name	S.No	Parameter Name
1	Num_Preg	5	Insulin
2	Glucose_Conc	6	BMI
3	Diastolic_bp	7	Diab_Pred
4	Thickness	8	Age

Table 2. Stanford AIM-94 Diabetes Dataset

S.No	Parameter Name	S.No	Parameter Name
1	Regular insulin dose	11	Post-supper blood glucose measurement
2	NPH insulin dose	12	Pre-snack blood glucose measurement
3	Ultra Lente insulin dose	13	Hypoglycemic symptoms
4	Unspecified blood glucose measurement	14	Typical meal ingestion
5	Unspecified blood glucose measurement	15	More-than-usual meal ingestion
6	Pre-breakfast blood glucose measurement	16	Less-than-usual meal ingestion
7	Post-breakfast blood glucose measurement	17	Typical exercise activity
8	Pre-lunch blood glucose measurement	18	More-than-usual exercise activity
9	Post-lunch blood glucose measurement	19	Less-than-usual exercise activity
10	Pre-supper blood glucose measurement	20	Unspecified special event

The original matrix included 21 distinct diabetes prediction criteria. 29330 records are utilized for testing reasons. While preprocessing the dataset, the majority of the

parameters produce negative outcomes. Table 2 provides the parameters utilized in this experiment to classify the data for the patient's records as diabetes or not.

3.2 Proposed Work

Diabetes detection from current datasets is always a difficult task. The Pima Indian dataset and the UCI dataset are utilized for testing reasons. The proposed method analyses several diabetes markers and determines whether they are diabetic prone or not. Data is preprocessed and classified into many classifications. Dimensionality reduction is used to remove undesirable or noisy data from preprocessed datasets. Principal component analysis (PCA) is used to reduce dimensionality by taking into account the identity matrix. Equation (1) describes the PCA that was done on the dataset.

$$D=\mathrm{O}-\mathrm{II}T\ \mathrm{O}() \tag{1}$$

Equation (1) is derived from the original matrix of records and features. Furthermore, an identity matrix with the number of rows of an original matrix with the identity value is created. To obtain the output matrix, dimensionality reduction is conducted on the original pre-processed matrix. The final matrix is created by using the similarity measure to transform the dimensionally reduced matrix. To acquire the optimal similarity, the similarity measure is computed using the transformation matrix. Classification of the matrix is performed using Naïve Bayes similarity measure. Naïve Bayes is used as a benchmark classification method used to classify the given dataset whether the patient is normal or abnormal. Advanced naïve Bayes algorithm is proposed to perform similarity identification for the given dataset. Though many approaches are used in the literature, the scope of this work was carried out using feature clustering and feature similarity that was specified in [63-65]. The motivation to carry out this work is using the feature similarity measures shown in [66].

$$\psi min=\psi i+\tau i.\mu i.ai\{i\epsilon(1,2,3,\ldots\ldots\ldots n)\} \tag{2}$$

Equation (2) represents the variance of the transformed matrix to compute the similarity between the features. Where τi belongs to the random variable with index ranging from -1 to 1; μi is the binary mutation range with having the precision between 0and1.The probability distribution for identifying the precision ways ranges between 0 and 1. A conventional cross-over is applied on ‘n’ variables to compare with the new individual variables. After evaluating equation (2), selection of the random variables is performed on the non-binary variables to identify the new elements. Binary metaheuristics algorithm is applied which leads to better accuracy.

After performing the product of the identity matrix and the transpose of it represented in equation, the transformation of the original matrix is produced (1). Where 'n' is the number of records available in the supplied dataset. The flowchart of the technique used to obtain the dataset is shown in Figure 1. Researchers frequently consider the PIMA diabetes dataset for experimenting in the classification of diabetes patients as normal or abnormal. In this research, the Stanford AIM-94 dataset is considered for experimentation.

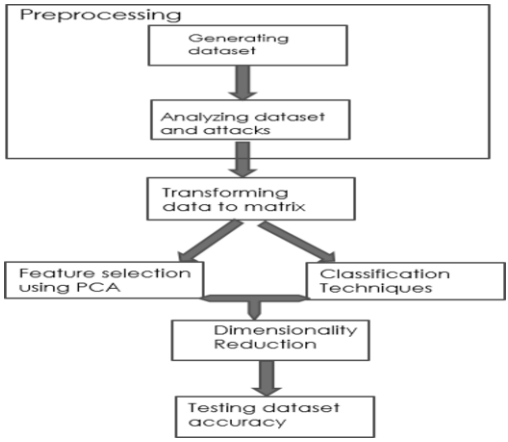

Figure 1. Flowchart of Proposed Approach

4. Discussions and Results

Statistics are gathered by conducting multiple experiments with the PIMA dataset, the UCI dataset, and the Stanford AIM-94 dataset, each of which has 29360 records preprocessed to 3478 records. The dataset is dimensioned and classified using the J48, Nave Bayes, Random Tree, and Random Forest models, in addition to the suggested model. Below mentioned in Table 3 to Table 7 gives the confusion matrix of the different classifiers with class labels normal or abnormal; the confusion matrix along with the accuracy, precision, and recall are calculated to identify the accuracy of the existing model with the previous models.

Table 3. Confusion Matrix of J48 with Aim-94 dataset

Classificaon	Confusion Matrix				Metric						
		Class 1	Class 0	Tota l	TP	FN	FP	TN	Accuracy	Precision	Recall
J48	Class 1	1747	455	2202	1747	455	453	823	73.893	0.794	0.793
	Class 0	453	823	1276	823	453	455	1747	73.893	0.644	0.645
				3478	2570	908	908	2570	73.893	0.739	

Table 3 gives the overall information about the confusion matrix for 3478 preprocessed records. The overall accuracy of the J48 classifier is noted as73.89% and from this, various other parameters like precision, recall, sensitivity, and specificity can be calculated. In similar ways, the accuracy of Naïve Bayes is obtained as 65.095%, RandomTreewith70.529%, RandomForestwith75.043%, and proposed approach with 84.215%. According to the results of the trial, the precise approach of the suggested solution outperforms the existing alternatives. When using the suggested measure on the Stanford Aim-94 dataset, the accuracy and precision outperform all other techniques. When compared to existing techniques, the number of dimensions lowered is relatively good. The performance of reading the data improves significantly when the number of features is reduced. One such important contribution shown is identifying True positive, False Positive values shown in Fig 2 and Fig3.

Table 4. ConfusionMatrixofNaïveBayeswithAim-94dataset

Classification	Confusion Matrix				Metric						
Naïve Bayes		Class 1	Class 0	Total	TP	FN	FP	TN	Accuracy	Precision	Recall
	Class 1	1019	1183	2202	1019	1183	31	1245	65.095	0.970	0.463
	Class 0	31	1245	1276	1245	31	1183	1019	65.095	0.513	0.976
				3478	2264	1214	1214	2264	65.095	65.095	

Table 5. Confusion Matrix of Random Tree with Aim-94dataset

Classification	Confusion Matrix				Metric						
Random Tree		Class1	Class 0	Total	TP	FN	FP	TN	Accuracy	Precision	Recall
	Class 1	1692	510	2202	1692	510	515	761	70.529	0.767	0.768
	Class 0	515	761	1276	761	515	510	1692	70.529	0.599	0.596
				3478	2453	1025	1025	2453	70.529	70.529	

Table 6. Confusion Matrix of Random Forest with Aim-94 dataset

Classification	Confusion Matrix				Metric						
Random Forest		Class 1	Class 0	Total	TP	FN	FP	TN	Accuracy	**Precision**	**Recall**
	Class 1	1762	428	2202	1762	440	428	848	75.043	0.805	0.800
	Class 0	440	848	1276	848	428	440	1762	75.043	0.658	0.665
				3478	2610	868	868	2610	75.043	75.043	

Table 7. Confusion Matrix of Proposed Approach with Aim-94 dataset

Classification	Confusion Matrix				Metric						
Proposed Approach		Class 1	Class 0	Total	TP	FN	FP	TN	Accuracy	Precision	Recall
	Class 1	2369	130	2499	2369	130	419	560	84.215	0.850	0.948
	Class 0	419	560	979	560	419	130	2369	84.215	0.812	0.572
				3478	2929	549	549	2929	84.215	84.215	

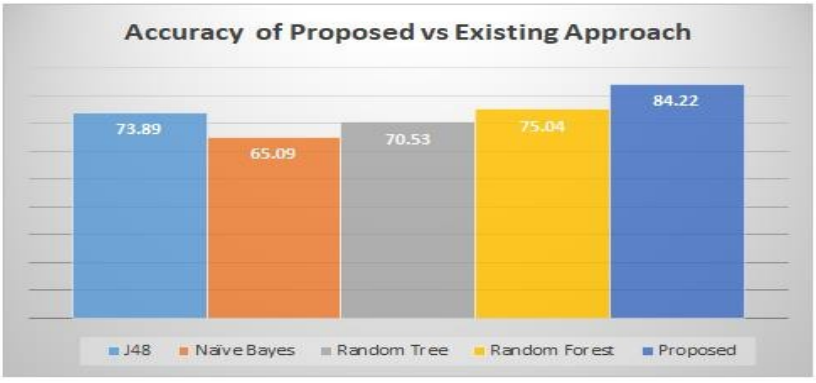

Figure 2. Accuracy of Proposed vs Existing Approach

The results carried out are performed for any conventional dataset with various existing classifiers. The proposed approach's how better accuracy and with other methods, the accuracy and precision are calculated but the results are not promising. The records collected have very few samples, the improvement in identifying the data that is classified as normal or abnormal is a tedious task. Fig 2 and Fig 3 show the various accuracies with the proposed approach compared with existing methods. In addition, the results are shown promising compared to existing methods.

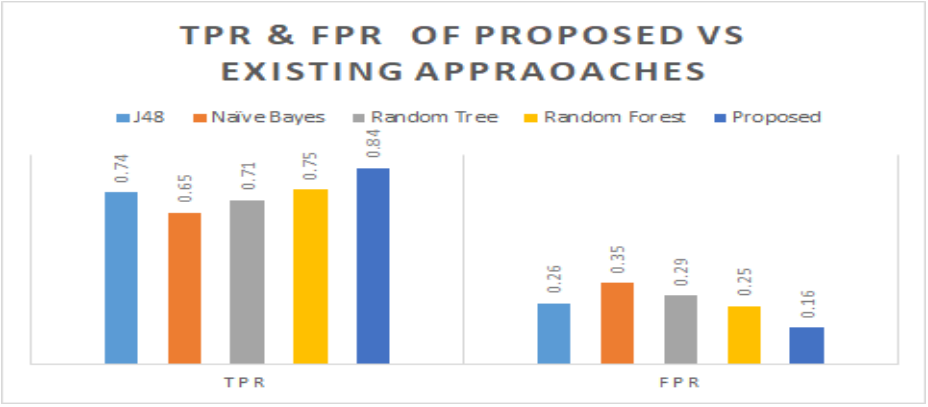

Figure 3. TPR & FPR of Proposed vs Existing Approach

5. Conclusion and Future Work.

Many scopes are available from the front referenced survey and examination for a great deal of exceptional effort that has been put in by particular scientists all around the world. The literature provides immense motivation from their works. A huge fraction of the tests have not had the opportunity to demonstrate unequivocally that a preexisting Form II diabetic disease will result in any type of malignant growth. We firmly agree that expecting the occurrence of any sort of cancer in persons with preceding type II diabetes is critical in knowing the factors that contribute to malignant growth and assuming that can distinguish between harmful and innocuous tumours. In our current work, we are attempting to develop an original methodology for establishing any conceivable link between type II diabetes and any type of case by utilizing information-driven information design based calculations while utilizing Machine learning methods for classifying the data into normal and abnormal with greater accuracy.

References

[1] Zhu B, Wu X, Wu B, Pei D, Zhang L, Wei L (2017) The relationship between diabetes and colorectal cancer prognosis: A meta-analysis based on the cohort studies. April 19, 2017.

[2] Chen L, Magliano D J, Zimmet P Z. The worldwide epidemiology of type 2 diabetes mellitus present and future perspectives. Nature reviews Endocrinology. 2011; 8(4):228±36. Epub 2011/11/09. https://doi.org/10.1038/nrendo.2011.183 PMID:22064493

[3] Klil-Drori A J, AzoulayL, Pollak M N.Cancer, obesity, diabetes, and antidiabetic drugs: is the fog clearing? Nature reviews Clinical oncology. 2016. Epub 2016/10/26.

[4] Klil-Drori A J, AzoulayL, Pollak M N. Cancer, obesity, diabetes, and antidiabetic drugs: is the fog clearing? Nature reviews Clinical oncology. 2016. Epub 2016/10/26.
[5] Mills KT, Bellows CF, Hoffman AE, Kelly TN, Gagliardi G. Diabetes mellitus and colorectal cancer prognosis: a meta-analysis. Diseases of the colon and rectum. 2013; 56(11):1304±19. Epub 2013/10/10. PubMed Central PMCID: PMCPMC3800045.https://doi.org/10.1097/DCR.0b013e3182a479f9 PMID:24105007
[6] Bella F, Minicozzi P, Giacomini A, Crocetti E, Federico M, Ponz de Leon M, et al. Impact of diabetes on overall and cancer-specific mortality in colorectal cancer patients. Journal of cancer research and clinical oncology. 2013; 139(8):1303±10. Epub 2013/05/02. https://doi.org/10.1007/s00432-013-1439-8 PMID: 23633003
[7] Cossor FI, Adams-Campbell LL, Chlebowski R T, Gunter M J, Johnson K, Martell R E, et al. Diabetes, metformin use, and colorectal cancer survival in postmenopausal women. Cancer epidemiology. 2013; 37(5):742±9. Epub 2013/06/19. PubMed Central PMCID: PMCPMC3769471. https://doi.org/10.1016/j.canep.2013.04.015 PMID: 23773299
[8] Chen KH, Shao YY, Lin ZZ, Yeh YC, Shau WY, Kuo RN, et al. Type 2 diabetes mellitus is associated with increased mortality in Chinese patients receiving curative surgery for colon cancer. The oncologist. 2014; 19(9):951±8. Epub 2014/07/26. PubMed Central PMCID: PMCPMC4153450. https://doi.org/10.1634/theoncologist. 2013-0423 PMID: 25061090.
[9] Luo J, Lin HC, He K, Hendryx M. Diabetes and prognosis in older persons with colorectal cancer. British journal of cancer. 2014; 110(7):1847±54.
[10] Tong L, Ahn C, Symanski E, Lai D, Du XL. Temporal trends in the leading causes of death among a large national cohort of patients with colorectal cancer from 1975 to 2009 in the United States. Annals of Epidemiology. 2014; 24(6):411±7. Epub 2014/02/18. https://doi.org/10.1016/j.annepidem. 2014.01.005 PMID: 24529646.
[11] Fransgaard T, Thygesen LC, Gogenur I. Increased 30-day mortality in patients with diabetes undergoing surgery for colorectal cancer. Colorectal disease: the official journal of the Association of Coloproctology of Great Britain and Ireland. 2016; 18(1): O22±9. Epub 2015/10/16.
[12] HuangY, CaiX, MaiW, LiM, HuY. Association between prediabetes and risk of cardio vascular disease and all-cause mortality: systematic review and meta-analysis. BMJ (Clinical research ed). 2016; 355: i5953. Epub 2016/11/25. PubMed Central PMCID: PMCPMC5121106.
[13] Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. Trials. 2007; 8:16. Epub 2007/06/09. PubMed Central PMCID: PMCPMC1920534. https://doi.org/10.1186/1745-6215-8-16 PMID: 17555582
[14] Zhu B, Wu X, Bi Y, Yang Y. Effect of bilirubin concentration on the risk of diabetic complications: A meta-analysis of epidemiologic studies. Scientific reports. 2017; 7:41681. Epub 2017/01/31. PubMed Central PMCID: PMCPMC5278382. https://doi.org/10.1038/srep41681 PMID: 28134328The relationship between diabetes and colorectal cancer prognosis PLOS
[15] Kimberly S .Peairs, Bethany B. Barone, Claire F. Snyder, Hsin-ChiehYeh, Kelly B. Stein, Rachel L. Derr, Frederick L. Brancati, and Antonio C. Wolff Diabetes Mellitus and Breast Cancer Outcomes: A Systematic Review and Meta-Analysis January 1, 2011.
[16] Yancik R, Wesley MN, Ries LA, et al: Effect of age and comorbidity in postmenopausal breast cancer patients aged 55 years and older. JAMA 285: 885-892, 2001.
[17] van de Poll-Franse LV, Houterman S, Janssen- Heijnen ML, et al: Less aggressive treatment and worse overall survival in cancer patients with diabetes: A large population-based analysis. Int J Cancer 120:1986-1992, 2007.
[18] Srokowski TP, Fang S, Hortobagyi GN, et al: Impact of diabetes mellitus on complications and outcomes of adjuvant chemotherapy in older patients with breast cancer. J Clin Oncol 27:2170-2176, 2009.
[19] Fleming S T, Pursley H G, Newman B, et al: Comorbidity as a predictor of the stage of illness for patients with breast cancer. Med Care 43:132-140, 2005.
[20] Konstantinos K Tsilidis assistant professor1 2, John C Kasimis Ph.D. student1, David S Lopez, Evangelia E Ntzani, John PA Ioannidi" Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies" BMJ 2014;350: g7607: https://doi.org/10.1136/bmj.g7607
[21] Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses.BMJ2011;342:d549.
[22] Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. StatMed. 2002;21:1539-58.
[23] Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, et al. Diabetes and cancer: a consensus report. Diabetes Care 2010; 33:1674-85.
[24] Monami M, Lamanna C, Balzi D, et al: Sulphonylureas and cancer: A case-control study. Acta Diabetol 46:279-284, 2009.
[25] Bowker SL, Yasui Y, Veugelers P, et al: Glucose-lowering agents and cancer mortality rates in type 2 diabetes: Assessing effects of time-varying exposure. Diabetologia 53:1631-1637,201

[26] Dániel Végh, Dorottya Bányai, Péter Herman1, Zsolt Németh and Márta U jpál "Type-2 Diabetes Mellitus and Oral Tumors in Hungary: A Long-Term Comparative Epidemiological Study" ANTICANCER RESEARCH 37: 1853-1857 (2017).
[27] Wu CH, Wu TY, Li CC, Lui MT, Chang KW, and Kao SY: Impact of diabetes mellitus on the prognosis of patients with oral squamous cell carcinoma: A retrospective cohort study. Ann Surg Oncol 17: 2175-2183, 2010.
[28] Faulds MH and Dahlman-Wright K: Metabolic diseases and cancer risk.CurrOpinOncol24:58-61, 2012.
[29] Goutzanis L, Vairaktaris E, Yapijakis C, Kavantzas N, Nkenke E, Derka S, Vassiliou S, Acil Y, Kessler P, Stavrianeas N, PerreaD, Donta I, Skandalakis P and Patsouris E: Diabetes may increase the risk for oral cancer through the insulin receptors ubstrate-1and focal adhesion kinase pathway. Oral Oncol 43:165-173, 2007.
[30] Werner H and Katz J: The emerging role of the insulin-like growth factors in oral biology. JDentRes83: 832-836, 2004.
[31] Carboni JM, Lee AV, Hadsell DL, Rowley B R, Lee FY, Bol DK, Camuso AE, Gottardis M, Greer AF, Ho CP, Hurlburt W, Li A, Saulnier M, Velaparthi U, Wang C, Wen ML, Westhouse RA, Wittman M, Zimmermann K, Rupnow B A, and Wong TW: Tumor development by transgenic expression of a constitutively active insulin-like growth factor I receptor. Cancer Res 65:3781-3787, 2005.
[32] Suba Z and Ujpal M: Correlations of insulin resistance and neoplasms. MagyOnkol 50: 127-135, 2006.
[33] White MF: The insulin signaling system and the IRS proteins. Diabetologia. 1997;40 Suppl 2: S2–S17
[34] Stocks T, Rapp K, Bjorge T, et al: Blood glucose and risk of incident and fatal cancer in the metabolic syndrome and cancer project (Me-Can): Analysis of six prospective cohorts. 2009. PLoSMed. 6(12):e1000201.
[35] Salahudeen A K, Kanji V, Reckelhoff J F and Schmidt AM Pathogenesis of diabetic nephropathy: Aradical approach. Nephrol Dial Transplant 12: 664-668, 1997.
[36] Meng-HsuenHsieh, Li-MinSun, Cheng-LiLin, Meng-JuHsieh, KyleSun, Chung-Y.Hsu, An-KuoChou and Chia-Hung Kao "Development of a Prediction Model for Colorectal Cancer among Patients with Type 2 Diabetes Mellitus Using a Deep Neural Network" J.Clin.Med. 2018, 7,277; doi:10.3390/jcm7090277
[37] Deng, L.; Gui, Z.; Zhao, L.; Wang, J.; Shen, L. Diabetes mellitus and the incidence of colorectal cancer: An updated systematic review and meta-analysis. Dig. Dis. Sci. 2012, 57, 1576–1585
[38] [38] Yuhara, H.; Steinmaus, C.; Cohen, S.E.; Corley, D.A.; Tei, Y.; Buffler, P.A. Is diabetes mellitus an independent risk factor for colon cancer and rectal cancer? Am. J. Gastroenterol. 2011, 106, 1911–1921.
[39] Sinagra, E.; Guarnotta, V.; Raimondo, D.; Mocciaro, F.; Dolcimascolo, S.; Rizzolo, C.A.; Puccia, F.; Maltese, N.; Citarrella, R.; Messina, M.; et al. Colorectal cancer risk in patients with type 2 diabetes mellitus: A single-center experience. J. Biol. Regul. Homeost. Agents 2017, 31, 1101–1107.[PubMed].
[40] Jiang, Y.; Ben, Q.; Shen, H.; Lu, W.; Zhang, Y.; Zhu, J. Diabetes mellitus and incidence and mortality of colorectal cancer: A systematic review and meta-analysis of cohort studies. Eur. J. Epidemiol. 2011, 26, 863– 876.
[41] Sutskever, I.; Martens, J.; Dahl,G.; Hinton, G. On the importance of initialization and momentum in deep learning. PMLR 2013, 28, 1139–1147.19.
[42] Martens, J. Deep learning via Hessian-free optimization. In Proceedings of the 27th International Conference on Machine Learning (ICML), 2010.
[43] Martens, J. and Sutskever, I. Learning recurrent neural networks with hessian-free optimization. In Proceedings of the 28th International Conference on Machine Learning (ICML),pp.1033–1040, 2011.
[44] Martens, J. and Sutskever, I. Training deep and recurrent networks with hessian-free optimization. Neural Networks: Tricks of the Trade, pp. 479–535, 2012.
[45] Meng-Hsuen Hsieh et al. Development of a Prediction Model for Colorectal Cancer among Patients with Type2 Diabetes Mellitus Using a Deep Neural Network, J. Clin. Med. 2018, 7, 277; doi:10.3390/jcm7090277
[46] Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. Available online: https://arxiv.org/pdf/ 1412.6980.pdf (accessed on 25 July2018).
[47] Dozat, T. Incorporating Nesterov Momentum into Adam. In Proceedings of the International Conference on Learning Representations Workshop, San Juan, Puerto Rico, 2–4 May 2016.
[48] Shraboni Rudra a, Minhaz Uddin a, Mohammed Minhajul Alam. Forecasting of Breast Cancer and Diabetes Using Ensemble Learning, the Intl J Comp. Comm Inf Vol. 1 Iss. 1 the Year 2019
[49] P Boyle1 , M Boniol*,1, A Koechlin1, C Robertson1 Diabetes, and breast cancer risk: a meta-analysis British Journal of Cancer (2012) 107(9), 1608 –1617
[50] Coughlin SS, Calle EE, Teras LR, Petrelli J, Thun MJ (2004) Diabetes mellitus as a predictor of cancer mortality in a large cohort of US adults. Am J Epidemiol 159:1160–1167

[51] de Waard F, Baanders-van Halewijn EA (1974) A prospective study in general practice on breast-cancer risk in postmenopausal women. Int J Cancer 14:153–160
[52] Dr. Prof. Neeraj, Sakshi Sharma, Renuka Purohit, and Pramod Singh Rathore, Prediction of Recurrence Cancer using J48 Algorithm, 2nd Inter. Conf. Comm.Elect. Syst., (2017) 386-390.[6]
[53] Deepika Verma and Dr. Nidhi Mishra, Analysis and Prediction of Breast Cancer and Diabetes disease datasets using Data mining classification Techniques, Inter. Conf. Intell. Sust. Sys. (2017)533-538.
[54] Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 2014, 15,1929–1958
[55] He,H.; Garcia, E.A. Learning from imbalanced data. IEEE Trans. Knowl.Data Eng. 2009,21,1263–128.
[56] Sui-Foon Lo, Shih-Ni Chang, Chih-Hsin Muo, Shih-Yin Chen, Fang-Yin Liao, Shu-Wei Dee, Pei-Chun Chen3andFung-ChangSung (2013).A modest increase in the risk of specific types of cancer types in type2 diabetes mellitus patients. DOI: Int. J. Cancer: 132, 182–188 (2013).
[57] ZidianXie, Olga Nikolayeva, MS; Jiebo Luo, Dongmei Li. Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques. Centers for Disease Control and Prevention. VOLUME 16, E130 PUBLIC HEALTH RESEARCH, PRACTICE, AND POLICY SEPTEMBER 2019.
[58] Meng hsuenhsieh, li-Min sun, Cheng-li, lin Meng-Ju hsieh, Chung-Y hsu Chia-hung Kao. Development of a prediction model for pancreatic cancer in patients with type 2 diabetes using logistic regression and artificial neural network models. Cancer Management and Research 2018:106317–6324.
[59] J. A. Johnson & B. Carstensen & D. Witte & S. L. Bowker & L. Lipscombe & A. G. Renehan. Diabetes and cancer (1): Evaluating the temporal relationship between type 2 diabetes and cancer incidence. Diabetologia (2012) 55:1607–1618
[60] Ashrita Kannan, P. Vigneshwaran, R. Sindhuja, and D. Gopikanjali. "Classification of Cancer for Type 2 Diabetes Using Machine Learning Algorithm." A. Kannan et al M. Tuba et al. (eds.), ICT Systems and Sustainability, Advances in Intelligent Systems and Computing 1077.
[61] Shahabeddin Abhari, Sharareh R. Niakan Kalhori, Mehdi Ebrahimi, Hajar Hasannejadasl, and Ali Garavand. "2 Diabetes Artificial Intelligence Applications in Type Mellitus Care: Focus on Machine Learning Methods ". 2019 The Korean Society of Medical Informatics. ISSN2093-3681
[62] Hui Chen, Yi Xin, Yuting Yang, Fei Li, Guoliang Cheng and Xinxin Zhang. "Related Factors and Risk Prediction of Type 2 Diabetes Complicated with Liver Cancer," Proceedings of 2019 IEEE International Conference on Mechatronics and Automation.
[63] A. Nagaraja, U. Boregowda, K. Khatatneh, R. Vangipuram, R. Nuvvusetty and V. Sravan Kiran, "Similarity-Based Feature Transformation for Network Anomaly Detection," in IEEE Access, vol. 8, pp. 39184-39196, 2020, DOI:10.1109/ACCESS.2020.2975716.
[64] Nagaraja, A., Uma, B. and Gunupudi, R. Found Sci (2019). https://doi.org/10.1007/s10699-019-09589-5.
[65] Arun Nagaraja, Shadi Aljawarneh, and Prabhakara H. S. 2018. PAREEKSHA: a machine learning approach for intrusion and anomaly detection. In Proceedings of the First International Conference on Data Science, E-learning and Information Systems (DATA '18). ACM, New York, NY, USA, Article 36, 6pages.DOI: https://doi.org/10.1145/3279996.3280032
[66] Nagaraja, A., Sravan Kiran, V., Prabhakara H. S, and Rajasekhar, N. A membership function for intrusion and anomaly detection of low-frequency attacks. In Proceedings of the first international conference on data science, e-learning and information systems (DATA '18). New York: ACM. https ://doi.org/10.1145/32799 96.3280031