Hybrid Multi-Cloud based Disease Prediction Model for Type II Diabetes

M. Durgadevi, R. Kalpana

Abstract: Advancements in health informatics pave the way to explore new medical decision making systems which are characterized by an exponential evolution of knowledge. In the medical domain, disease prediction has become the centre of research with the increasing trend of healthcare applications. The predictive knowledge for the diagnosis of disease highly depends on the subjective knowledge of the experts. So the development of a disease prediction model in time is essential for patients and physicians to overcome the problem of medical distress. This paper explores a hybrid approach (Cooperative Ant Miner Genetic Algorithm) for classifying the medical data. Three benchmarked Type II diabetic datasets (US, PIMA, German) from the UCI machine learning repository were used to analyze the effectiveness of the disease prediction model. The devised classification algorithm with a Soft-Set approach was deployed in a Multi-Cloud environment for enhancing the storage and retrieval of data with reduced response and computation time. The cooperative classification algorithm in the cloud database distinguishes the diseased cases from the normal ones .The soft set theory analyzes the severity of the diseased cases by calculating the percentage of diabetic risk using soft intelligent rules and stores them in a separate knowledge base. Thus the proposed model serves as a suitable tool for eliciting and representing the expert's decision which aids in prediction of Type II diabetic risk percentage leading to the timely treatment of patients.

Keyword: Disease Prediction, Ant Miner Algorithm, Genetic Algorithm, Soft Sets, Multi cloud storage

I. INTRODUCTION

Recently many advanced storage techniques were developed to store the excess data files on the cloud in such a way that they can be accessed remotely. But the reliability of these storage techniques is not guaranteed. Every cloud has its independent element in such a way that when a foreign cloud wants to make use of the storage of another cloud, it has to trust its respective element. This utilization of storage as service of another cloud is made possible through the use of a multi-cloud storage database. Therefore, the complete cloud-based disease prediction model of Type II Diabetes was optimized by the support of a Multi cloud storage system. Three cloud storage providers were used namely Dropbox, Google Drive and Copy.

M. Durgadevi, pursuing her Ph.D. degree in the Department of Computer Science and Engineering at Pondicherry Engineering College, Puducherry, India.
  Dr.R. Kalpana, Professor in the Department of Computer Science and Engineering at Pondicherry Engineering College, Puducherry India.

Three cloud storage providers were used namely Dropbox, Google Drive and Copy. The input data comprising the Type II Diabetic dataset obtained from the UCI repository and the patient's real time data collected from hospitals along with their personal details are accumulated in the cloud database. The prediction and diagnosing system predicts the Type II diabetes using the cooperative classification algorithm and the soft set estimates the percentage of diabetic risk. Therefore the health records of the patient are categorized as normal and diseased. The set of antecedent consequent rules generated by the classification algorithm aids in making strong decisions on the health records.

In this paper, a new cloud-based healthcare application developed for monitoring, predicting and diagnosing type II diabetes has been discussed. This cloud-based application enables the people to easily access and avail the online healthcare services around the globe. The devised classification algorithm (Cooperative Ant Miner Genetic algorithm) was incorporated into the cloud environment for predicting the Type II Diabetes. The developed soft set theory was also integrated with the cloud storage for analyzing the severity of the Type II Diabetes. This severity is estimated in terms of risk percentage.

II. RELATED STUDIES

The achievability and sustainability of a country's economic and social growth is partly dependent on the medical sector of the country [1]. Preserving current updated patients' data is a herculean task where the advanced data mining technologies ease the data entry process [2]. Different disease prediction models have been presented to assist doctors for dealing with the diagnostic problems. The diagnosis model takes care of all the decision-making process when the doctor comes across an abnormal or an unprecedented case of clinical disorder. In recent years, the advanced growths of the prediction models have set aside the physician's decision by the maintenance of patient's medical history [3]. Presently, the Artificial Intelligence (AI) approaches are widely employed in healthcare applications [4].Several researches have been done on disease prediction models as a complementary solution to the healthcare [5]. These typical models are commonly applied for identifying the solution to the medical issues. The statistical tools and Machine Learning (ML) models are recently developed to assist the process of disease diagnosis [6]. The disease prediction model can investigate the massive and complex medical data in medical with the use of effective methodologies, namely, data mining tools, meta heuristic algorithms and soft-computing methods [7, 8].

Hybrid Multi-Cloud based Disease Prediction Model for Type II Diabetes

Personalized treatments offers better treatment by integrating patient health with the predictive analytics and also help the future researches to improve the process of disease diagnosis [9]. At present, doctors have a constraint of selecting a particular collection of diagnoses for determining the risk of the patient using the symptomatic history and existing genetic information [10]. In future, the healthcare domain may involve more advanced devices and biosensors with refined health measurement abilities which possibly will hit the market, enabling massive amount of data available for such cutting-edge ML-based healthcare technologies [11].The ML offers several interesting applications in the domain of clinical trials and research [12].Since the clinical trials lead to waste of time and money, the application of ML techniques can help to find the significant clinical trial candidates [13]. The ML techniques are also used for verifying the real-time monitoring and the data access of the trial participants. This process is done by identifying the best sample size to be validated which leverages the power of electronic records to minimize the data-based errors [14]. Several data classification models are available in the literature such as Linear Classifiers (LR), Decision Trees (DT), Support Vector Machines (SVM) and Neural Networks (NN) [15].

The IoTs offers a way for developing Internet-linked things using the computer networks. The IoT states that rather than the usage of significant computing devices viz., laptops, tablets and mobile phones, it is efficient to have many low powerful gadgets namely, watches, umbrella, refrigerator and air-conditioner [16].The frequently used daily services like room freshener and motor vehicles linked together undergo intelligent programming using computer processors. These processors are supported through highly reliable sensors which offer real-time output [17].Therefore, the linked devices or the things have computation and communication abilities. Apart from the usage of uncomplicated devices like umbrella and lamp, infrastructures can also be linked using network communication [18]. These delighted objects in IoT holds various technical reasoning abilities to carry out the assigned task in the medical sector without the need of any personal details.

Cloud Computing and IoT are advantageous in various ways when they are connected to one another. The monitoring system developed by the integration of these two technologies will screen the data of a patient effectively even at the remote sites which will be helpful for the doctors [19]. The IoT is usually supported by the cloud framework for enhancing the results with respect to high resource utilization, memory, energy and processing capabilities. Cloud computing also gains favor from IoT [20] technologies by improving the possibility to manage the present world and to offer various novel services in a distributed and dynamic way. The Cloud- based IoT techniques are very helpful in offering effective services to the healthcare applications to monitor and access records from any remote area [21]. The cloud-based healthcare applications were employed to gather essential details like adequate modifications in the medical attributes at a certain interval and transmit those serious health attributes in a fixed period of time. The IoT devices and the sensor reading stored in the cloud can be used to diagnose the disease at the appropriate time, prior to severity of the disease. A comprehensive diagnosis process, disease prevention, manual and psychological impairments in humans come under the healthcare monitoring [22]. The field of healthcare evolves rapidly in most of the countries. The healthcare domain might be considered as a rich data source as it produces high amounts of underutilized data involving administrative reports, standard findings and electronic clinical records [23]. A detailed analysis on the different data mining techniques employed in the healthcare field [24-26] was done with respect to the type of disease predicted, preprocessing technique used, datasets used and the performance analysis factors.

III. SYSTEM ARCHITECTURE OF MULTI CLOUD BASED DISEASE PREDICTION MODEL

The architecture of the multi-cloud-based disease prediction model is shown in Fig.1.The disease prediction model comprises six elements: Patient's personal data, dataset from a standard repository (UCI/Kaggle), patient's medical data and a multi-cloud database, a cooperative algorithm prediction system and a severity investigator with a separate knowledge base.
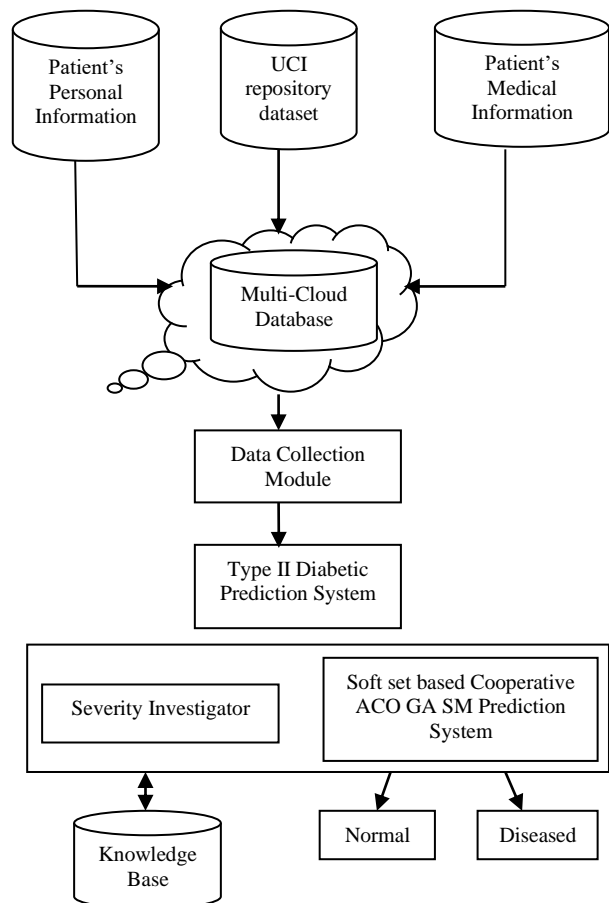

Fig. 1.Multi-cloud based disease prediction model

The dataset collected from the hospitals, clinical labs comprising the complete history of the patients is stored in the multi cloud database.

The data highly needed for the prediction of Type II diabetes is filtered from the multi cloud by the process of data collection. The filtered data is again stored in the multi cloud for easier access. The Type II diabetic prediction model has the capability to predict the disease by the cooperative Ant-Miner Genetic Algorithm. The classification algorithm classifies the patients records into normal and diseased. The prediction system has a sub-component called the severity investigator which helps to determine the percentage of diabetic risk using the intelligent soft rules which are stored in the knowledge base.

IV. SOFT SET BASED COOPERATIVE ANT MINER GENETIC ALGORITHM

A new disease prediction model has been presented by the use of Ant-miner algorithm [27]. This paper illustrated that the proposed ant-miner model can be employed to explore decision rules from the expert's data in both qualitative as well as quantitative way. A set of three feature selection approaches was employed to reduce the unwanted and repeated features in the dataset and thereby to enhance the results of the presented classification model. The classifier results attained by the use of three feature selection methods were highly sensitive to the type of dataset being utilized. The feature selection approaches enhances the rate of accuracy compared to the performance attained without the use of feature selection methods. Hence, the proposed Ant-Miner model has achieved better classification results than the other top classification models.

A new operator called suppressor mutation was employed to develop an efficient rule miner known as Genetic Algorithm-Suppressor Mutation (GA-SM) to predict diabetes [28]. This operator enabled the GA to generate a logical, optimum and precise rule set for the applied datasets by the elimination of redundant rules. The GA-SM model attained the best ranking with respect to all the classification measures compared to the other top classification methods. In addition, the GA-SM model has excellently performed by producing less number of rules for predicting the presence of diabetes efficiently.

A mathematical model based on soft set is employed to determine the percentage of diabetic risk [29]. Since the rule set is unsuitable for the application to soft sets straightly, the fuzzy membership values of the dataset is attained by the mathematical formula suggested by the experts. The dataset used for analysis has been provided by a comprehensive diabetic care and multispecialty centre. The real data set obtained from the clinical laboratory holds the details related to HbA1 c, PG-F, PG-PP and age data of patients. The HbA1 c, PG-F, PG-PP and age data are provided as input and the output is treated as the percentage of diabetic risk. The results prove the appropriate prediction of the diabetic risk by the soft set model over the other existing mathematical models.

V. MULTI-CLOUD BASED TYPE II DIABETIC PREDICTION MODEL

The multi-cloud-based disease prediction model consists of three stages. The Type II diabetic data from the UCI repository and hospital medical records are gathered together. In the second stage, the gathered data is stored on the multi-cloud database. The third stage analyzes the data in the multi-cloud thoroughly and classifies the data as normal and diseased. Apart from these stages, there is an investigation module which analyzes the severity of the diseased cases. These stages are explained in detail as follows:

A. Collection of data

This stage is responsible for the collection of data which exceeds the normal values. The highlight of this stage is that it selects only the needed features for the prediction. The data from the UCI repository and the real time data collected from hospitals are mapped together. This mapped information is processed in the cloud database and distributed to the doctors/physicians.

B. Data storage on multi-cloud database

The storage, handling and retrieval of large volumes of data from the UCI repository and real time medical data are a tedious task. The multi-cloud environment provides adequate space for handling this massive amount of health data. A file in a multi cloud is split into a chunk which has the capability to replicate in two or more cloud storage providers. The replication of data is done through the Redundant Residue Number System (RRNS). The RRNS resolves the following issues:

- The data is continuously made available in at least one cloud storage even when the other cloud storages are permanently or temporarily out of range. This resolves the problem of data redundancy.
- The RRNS in a multi-cloud storage system also resolves the issue of confidentiality by making the data in the cloud unintelligible so that the cloud provider does not split the data.

C. Disease prediction with severity investigator

In this phase, a set of soft rules is generated by the soft set approach for analyzing the severity of the disease. The severity investigator selects the significant features responsible for the occurrence of the disease and calculates the percentage of diabetic risk based on the specified time constraints.

VI. RESULT AND DISCUSSION

The multi-cloud storage system has been validated in three cloud storage providers, namely Dropbox, Google Drive and Copy. The data classification is the foremost task which segregates the data as normal and disease affected with severity.

Hybrid Multi-Cloud based Disease Prediction Model for Type II Diabetes

Table- I: Comparative analysis in terms of classification measures

Metric	Dataset	k-NN	NB	SVM	DT	Multi- Cloud Based Model
Sensitivity	**US**	88.234	78.234	79.125	77.123	**92.156**
	PIMA	**91.125**	81.345	82.145	80.125	90.278
	German	87.145	83.763	84.663	76.145	**88.792**
Specificity	**US**	83.278	82.391	83.191	82.268	**87.548**
	PIMA	86.256	81.256	82.156	85.245	**93.267**
	German	83.178	82.236	82.136	82.167	**92.266**
Accuracy	**US**	82.167	76.523	79.423	81.156	**89.178**
	PIMA	85.349	83.124	**88.134**	84.238	87.279
	German	86.234	77.267	79.167	85.123	**90.376**
F-Score	**US**	82.345	82.159	85.169	81.234	**89.297**
	PIMA	81.267	85.634	**88.734**	82.256	87.725
	German	83.678	85.239	89.241	82.567	**91.267**
AUC	**US**	84.356	87.239	89.239	81.247	**93.287**
	PIMA	85.378	84.189	87.178	83.276	**92.189**
	German	83.169	82.356	84.245	82.159	**91.389**
Kappa	**US**	0.825	0.834	0.884	0.839	**0.928**
	PIMA	0.817	0.827	0.877	0.828	**0.956**
	German	0.799	0.815	0.845	0.811	**0.974**

The datasets from the UCI repository and hospital data have been evaluated in terms of sensitivity, specificity, F- measure and accuracy. The performance of the multi cloud storage has been compared against the other top data classification approaches like SVM, KNN, DT and NB for analyzing the efficiency of the multi cloud storage. The obtained results were tested with ten-fold cross validation for confirming the performance of the disease prediction model in the multi-cloud storage system.

A. Experimental analysis w.r.t sensitivity

Table 1 provides the attained experimental outcome of different classification model with respect to many validation parameters on the applied diabetes datasets. Fig.2 illustrates the comparative performance of diverse techniques in terms of several measures. A lower sensitivity value of 77.123 is achieved by the DT classifier when tested on the US dataset. The NB and SVM models have achieved slightly higher sensitivity values of 78.234 and 79.125respectively. The k-NN model exhibits moderate results with a sensitivity value of 88.234 which exceed the results attained by other methods. However, it shows its incapability to achieve maximum performance over the presented multi-cloud storage method which has obtained the highest sensitivity value of 92.156.

Similarly, when the developed model was applied to the PIMA dataset, inefficient result is attained by the DT classifier with a minimum sensitivity value of 80.125. The NB and SVM classifiers exhibit slightly better sensitivity rates of 81.345 and 82.145. The k-NN technique reached the highest sensitivity value of 91.125 which was somewhat higher than the multi-cloud method. When evaluating the outcome on the German dataset, the worst performance was exhibited by the DT with a sensitivity value of 76.145. At the same time, it is noted that the NB and SVM classifiers also do not achieve any considerable performances. Even though the k-NN method achieved a maximum sensitivity rate of 87.145, it failed to outperform the proposed multi-cloud-based model which gains the sensitivity value of 88.792.

B. Experimental analysis w.r.t specificity

The obtained results of simulation are given in table 1 in terms of specificity over the various diabetes datasets. For US dataset, the moderate performance is demonstrated through the DT which has resulted in specificity rate of 82.268. The NB technique has also attained the subsequent moderate sensitivity rate of 82.391. The SVM and k-NN have attained the same specificity rates of 83.191 and 83.278. The multi-cloud method supersedes all the other classifiers by attaining high specificity rate of 87.548.
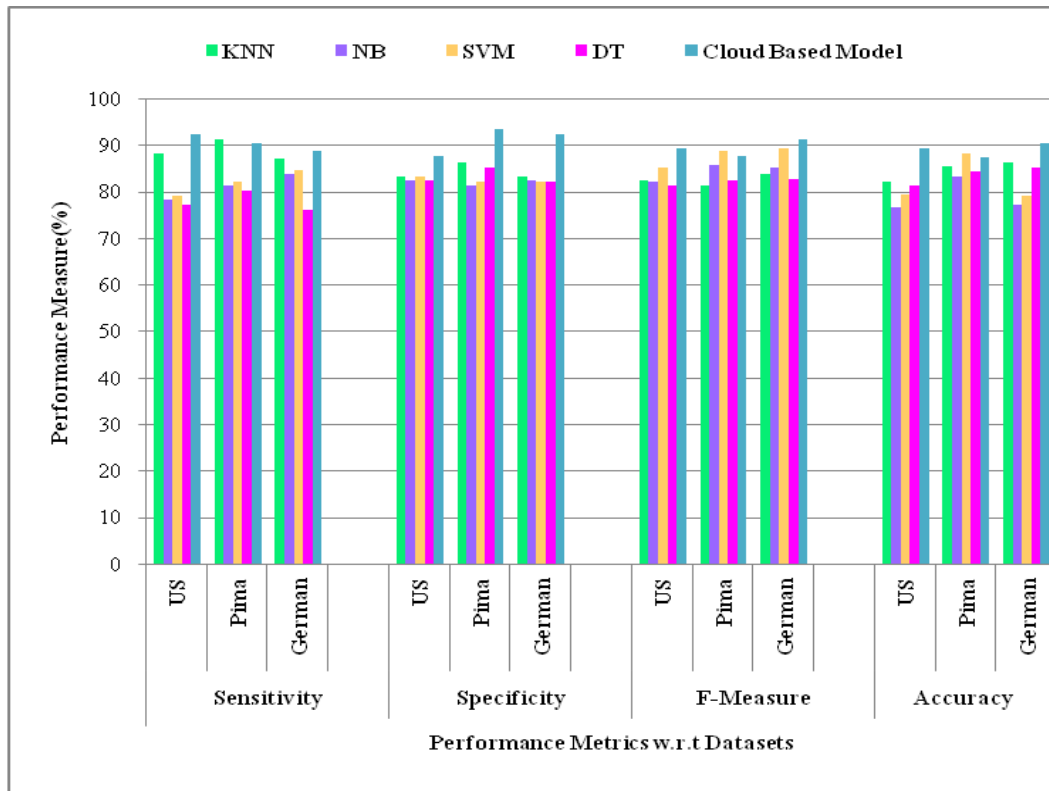
Fig 2. Comparative analysis of classification methods w.r.t several performance measures

For PIMA dataset, the NB attains the lowest specificity rate of 81.256. The SVM is the next worst performer which attains 82.156 as specificity rate. The methods k-NN and DT attains the moderate specificity rates of 86.256 and 85.245 respectively. The proposed method exceeds all the compared techniques by attaining the highest specificity rate of 93.267. For German dataset, the techniques like NB, SVM and DT has attained more or less the same performances of 82.236, 82.136 and 82.167. Whereas, the presented multi-cloud-based model outperforms the other classification approaches by attaining 92.266 as the highest specificity rate.

C. Experimental Analysis w.r.t Accuracy

As shown in the table 1, the worst performance is demonstrated through NB which has resulted in a specificity rate of 76.52 and the subsequent poor performances are demonstrated through the DT and SVM with the accuracy rates of 81.156 and 76.523 respectively on the applied US dataset. Above all, the k-NN method attains a superior accuracy rate of 82.167 compared to all the other methods but it fails to attain superior performance over the multi-cloud method which has attained the highest accuracy rate of 89.178. For the PIMA dataset, the NB attains the worst performance of 83.124, the DT and k-NN attains the moderate accuracy rates of 84.238 and 85.349 whereas the multi-cloud method attains the highest accuracy rate of 87.279. On the applied German dataset, the NB exhibits its inefficiency with the lowest accuracy of 77.267. On the other hand, the DT and SVM classifier also do not achieve any considerable performance. Though the k-NN attains a high accuracy rate of 86.234, it cannot outperform the multi-cloud-based model which gains an accuracy rate of 90.376.

D. Experimental Analysis w.r.t F-Score

The derived results of simulation in terms of F-Score are given in table 1. For US dataset, the moderate performance is demonstrated through the DT which has resulted in F-Score value of 81.234. The k-NN and NB techniques attain more or less the same F-Scores of 82.345 and 82.159. Even though the SVM classifier attains a superior performance of 85.169, it fails to outperform the projected multi-cloud-based model which attains the maximum F-Score rate of 89.297. Thus, the multi-cloud-based method exceeds all the other classifiers by attaining the highest F-Score rate. For PIMA dataset, the k-NN attains the lowest F-Score value of 81.267. The DT classifier is the next poor performer which attains 82.256 as F-Score. The NB and SVM methods attain the F-Score values of 85.634 and 88.734 respectively. The proposed multi-cloud-based method attains the F-Score rate of 87.725 which indicate its efficient classification. For German dataset, the DT attains the poor F-Score rate of 82.567 and the k-NN is the next worst performer which attains 83.678 as F-Score rate. The NB and SVM models obtain 85.239 and 89.241 as moderate F-Score rates. The multi cloud based method outperforms all the other methods by attaining 91.267 as the highest F-Score rate.
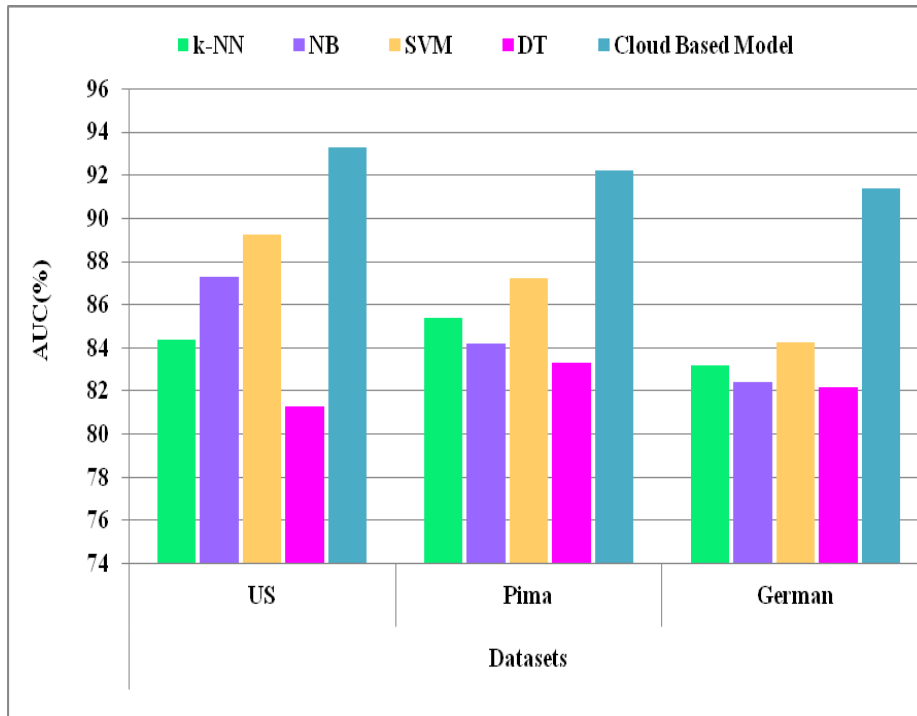
Hybrid Multi-Cloud based Disease Prediction Model for Type II Diabetes

Fig.3.Comparative analysis of various methods in terms of AUC

E. Experimental Analysis w.r.t AUC

The experimental results of various methods in terms of AUC are shown in table 1 and Fig.3. On the whole, the multi cloud method attains the highest AUC over all the applied datasets whereas the DT attains the worst performance in all cases. For the US dataset, DT attains the worst performance of 81.247 which is poor AUC compared to all the other techniques. The k-NN is the next poor performer that attains 84.356 as AUC. The SVM and NB obtained the high AUC rates of 89.239 and 87.239. The multi-cloud-method exceeds all the other methods by obtaining 93.287 as the highest AUC rate. For PIMA dataset, DT attains the lowest AUCof83.276. The NB and k-NN are the next worst performers that attain 84.189 and 85.378 as AUC. Even though, the SVM attains a superior F-score rate of 87.178, it fails in outperforming the multi cloud method that attains 92.189 as the highest AUC rate. For German dataset, the DT and NB attain the poor and more or less similar AUC rates of 82.159 and 82.356. The k-NN attains 83.169 as a moderate AUC rate. Though the SVM exceeds all the other methods by obtaining 84.245 as AUC rate, it fails to establish better results with respect to the multi-cloud method that attained the highest AUC of 91.389. Thus the multi cloud method outperforms all other methods with a maximum AUC.

F. Experimental Analysis w.r.t Kappa

The obtained results of simulation are given in table 1 in terms of kappa over the various diabetes datasets and are shown in Fig.4. For US dataset, the poor performance is demonstrated through the k-NN of 0.825. The technique NB and DT attain more or less the same kappa rates of 0.834 and 0.839. Even though the SVM attains a superior kappa rate of 0.884, it failed in exceeding the multi-cloud method that attains highest kappa rate of 0.928. For PIMA dataset, the k-NN attains the lowest kappa rate of 0.817. The NB is the

next worst performer which attains 0.827 as kappa rate. The methods SVM and DT attain the kappa rates of 0.877 and 0.828. The multi-cloud method exceeds all the compared techniques by attaining highest kappa value of 0.956. For German dataset, the techniques like NB and DT attains more or less the same poor performances of 0.815 and 0.811.Though the SVM attains the enhanced performance by achieving 0.845 as the kappa rate, the multi-cloud method outperforms all the methods by attaining 0.974 as the highest kappa rate. Thus the multi- cloud method shows efficient results in terms of kappa irrespective of the applied datasets.

G. Experimental Analysis w.r.t Response Time

Table 2 and Fig.5 shows the experimental outcomes of the compared methods in terms of response time w.r.t the applied datasets. For US dataset, the DT attains the worst performance of 72.156ms which was the highest time taken to respond to the user. The response time of k-NN and SVM was 50.234ms and 48.256ms.Though the NB has attained a low response time of 37.251, it has failed to outperform the multi-cloud method which has obtained 25.463 as the lowest response time. For the PIMA dataset, the DT requires the highest response time of 68.274ms.The k-NN is the next worst performer which needs 53.189ms as response time. The NB and SVM need 32.368ms and 44.367ms as response time. The multi-cloud model requires only 23.267 as response time which is comparatively very low. For the German dataset, the DT requires a maximum response time of 65.368ms and k-NN is the next method which requires high response time of 52.176ms. The NB and SVM need 30.254ms and 41.299ms as response time. As minimum response time is required for effective performance, the multi cloud model takes only 20.236ms. As a result, in all the cases, the multi cloud method is superior to all the other methods.
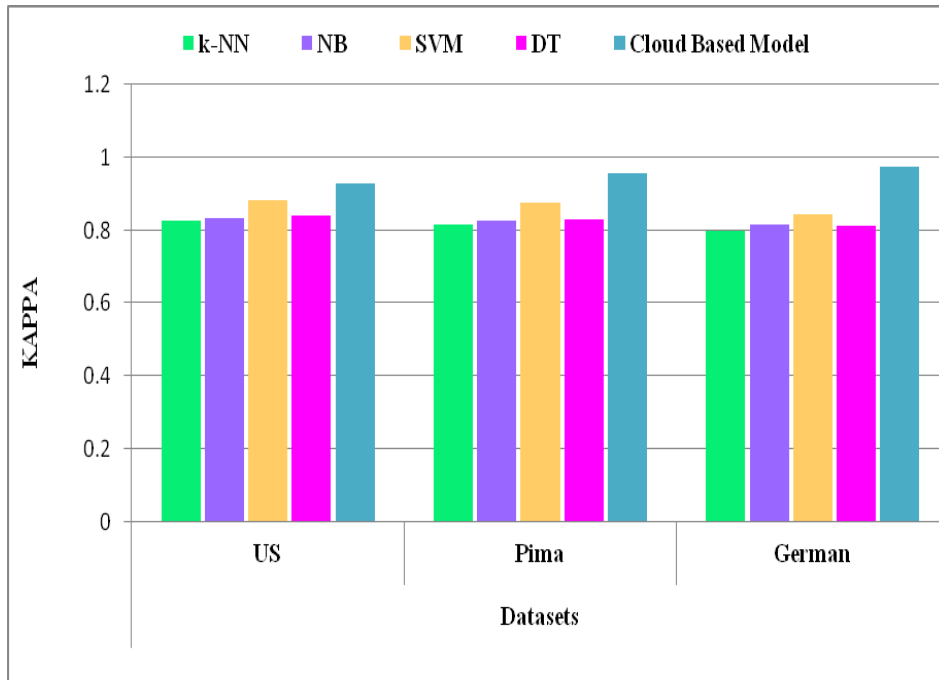
Fig.4.Comparative analysis of various methods in terms of kappa

Table 2. Comparative analysis in terms of response time

Performance Metric	Dataset	k-NN	NB	SVM	DT	Multi Cloud Based Model
Response Time(ms)	**US**	50.234	37.251	48.256	72.156	**25.463**
	PIMA	53.189	32.368	44.367	68.274	**23.267**
	German	52.176	30.254	41.299	65.368	**20.236**

H. Experimental Analysis w.r.t Computation Time

Table 3 and Fig.6 show the experimental results of the compared methods in terms of computation time on different datasets. For US dataset, the DT requires a maximum computation time of 89.283ms. At the same time, the NB also results in high computation time of 83.473ms but is slightly lower than the time needed by DT. Moreover, the SVM and k-NN require computation time of 77.236ms and 68.723ms. Interestingly, the multi-cloud model takes a minimum of 40.345ms which is significantly lower than those of the other methods. For the PIMA dataset, the DT requires more computation time of 81.342ms. Then, NB is the next worst performer which needs 75.389ms as the computation time.

The NB and k-NN do not show considerable performances. The multi-cloud model needs only a very low computation time of 30.276ms. For the German dataset, the DT needs a high computation time of 77.325ms. The NB is the next method that takes a higher computation time of 72.478ms. The k-NN and SVM need 60.725ms and 65.341ms as the computation time respectively. The multi-cloud method requires only 28.324ms as the computation time. Thus in all the cases, the multi-cloud method is superior to all the other techniques.

Table 3. Comparative analysis in terms of computation time

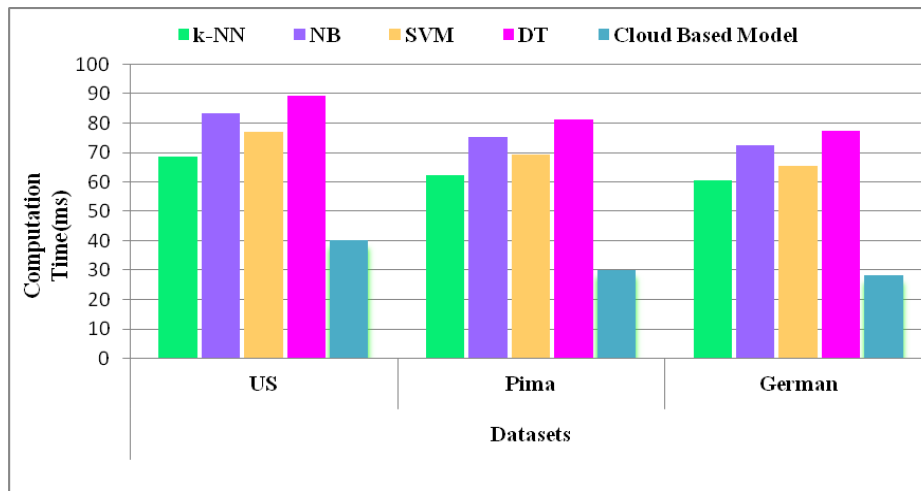| Performance Metric	Dataset	No of Attributes	No of Instances	k-NN	NB	SVM	DT	Multi cloud based Model
Computation Time(ms)	US	55	100000	68.723	83.473	77.236	89.283	**40.345**
	PIMA	8	768	62.342	75.389	69.238	81.342	**30.276**
	German	8	768	60.725	72.478	65.341	77.325	**28.324**

Fig.6.Comparative analysis of various methods in terms of computation time

VII. CONCLUSION

The healthcare sector seemed to be an important field for generating a huge quantity of data in various forms which can be useful to predict different diseases and also assist physicians to take clinical decisions. The data mining techniques have the ability to look for new and valuable information from the available massive quantity of data. In this paper, a hybrid approach for disease prediction has been built in a multi- cloud storage environment using three cloud storage providers which diagnose and monitor the severity level of the disease. A new framework has been constructed using the cooperative Ant-Miner-Genetic-Algorithm for diagnosing the disease and the severity is identified by the soft set theory. A multi-cloud database has been used to improve the performance of the prediction system since single-cloud storage cannot maintain the upload and download of data simultaneously. A hybrid multi-cloud based disease prediction model for Type II diabetes has been developed for better organization and retrieval of the stored data which has achieved **sensitivity (92.156), specificity (93.267), accuracy (90.376), AUC (93.287), F-Score (91.267), Kappa (0.974) with response time (20.236ms) and computation time (28.324ms).** The obtained results clearly show that the classification rules derived from the Cooperative Ant Miner-Genetic Algorithm has significantly performed better than the other top data mining classification algorithms. In future, the Cooperative Ant Miner Genetic Algorithm may be implemented on real time medical applications that require optimization.

REFERENCES

1. M.A. Uddin and M. Masih, "Finance, growth and human development: An Islamic economic development perspective",2015
2. M. Chen, S. Mao, and Y. Liu, "Big data: A survey", Mobile networks and applications, vol.19, no.2, pp.171-209, 2014.
3. P. Nielsen, "Information and Communication Technologies for Development. Strengthening Southern-Driven Cooperation as a Catalyst for ICT4D: 15th IFIP WG 9.4 International Conference on Social Implications of Computers in Developing Countries, ICT4D 2019, Dar es Salaam, Tanzania, May 1–3, 2019, Proceedings, Part I. Springer", 2019.
4. R. Bellazzi and B.Zupan, "Predictive data mining in clinical medicine: current issues and guidelines", International journal of medical informatics, vol.77, no. 2, pp.81-97, 2008
5. H. Chen, R.H. Chiang, and V.C. Storey, "Business intelligence and analytics: From big data to big impact", MIS quarterly, vol. 36, no. 4,2012.
6. M.I.Jordan and T.M.Mitchell, "Machine learning: Trends, perspectives, and prospects", Science, vol. 349, no. 6245, pp.255-260, 2015.
7. D.W.Bates, S.Saria, L. Ohno-Machado, A. Shah, and G.Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients", Health Affairs, vol. 33, no. 7, pp.1123-1131, 2014.
8. J.Lee, F.Wu, W.Zhao, M. Ghaffari, L. Liao, and D.Siegel, "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications", Mechanical systems and signal processing, vol. 42, no.1-2, pp.314-334,2014.
9. R.Miotto, L.Li, B.A. Kidd, and J.T Dudley, "Deep patient: an unsupervised representation to predict the future of patients from the electronic health records", Scientific reports, vol. 6, pp.26094,2016 (23)
10. D.F. Polit, and C.T. Beck, "Nursing research: Generating and assessing evidence for nursing practice", Lippincott Williams & Wilkins,2008.
11. S.E. Dilsizian and E.L. Siegel, "Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment", Current cardiology reports, vol. 16, no. 1, p.441, 2014.

12. M. Follen, R. Castaneda, M.Mikelson, D.Johnson, A. Wilson, and K. Higuchi, "Implementing health information technology to improve the process of health care delivery: a case study", Disease Management, vol. 10, no. 4, pp.208-215,2007

13. S. Timmermans and M. Berg, "The gold standard: The challenge of evidence-based medicine and standardization in health care", Temple University Press,2010.

14. I.Anagnostopoulos, S. Zeadally and E. Exposito, "Handling big data: research challenges and future directions",The Journal of Supercomputing, vol. 72, no. 4, pp.1494-1516,2016.

15. O. Patsadu, C. Nukoolkit and B. Watanapa, "Human gesture recognition using Kinect camera", In 2012 Ninth International Conference on Computer Science and Software Engineering IEEE,pp. 28-32,2012.

16. O. Vermesan and P. Friess, "Internet of things: converging technologies for smart environments and integrated ecosystems. River publishers",2013.

17. P.P.Pereira, J. Eliasson, R. Kyusakov, J. Delsing, A. Raayatinezhad, and M.Johansson, "Enabling cloud connectivity for mobile internet of things applications", In 2013 IEEE seventh international symposium on service-oriented system engineering, IEEE, pp. 518-526,2013.

18. W.Stallings, "Foundations of modern networking: SDN, NFV, QoE, IoT, and Cloud. Addison-Wesley Professional",2015.

19. E. Borgia, "The Internet of Things vision: Key features, applications and open issues", Computer Communications, vol. 54, pp.1-31,2014.

20. A. McEwen and H. Cassimally, "Designing the internet of things, John Wiley & Sons", 2013.

21. D. Minoli, "Building the internet of things with IPv6 and MIPv6: the evolving world of M2m communications, John Wiley & Sons", 2013.

22. R. B. Lukmanto and E.Irwansyah, "The Early Detection of Diabetes Mellitus (DM) Using Fuzzy Hierarchical Model",Procedia Computer Science, vol. 59, no. 1, pp. 312-319, 2015.

23. K. Polat, andS. Güneş, "An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease", Digital Signal Processing, vol. 17, no. 4, pp. 702-710, 2007.

24. S. Gajare and S. Sonawani,"Improved Automatic Feature Selection Approach for Health Risk Prediction."2018 Second International Conference on Computing Methodologies and Communication (ICCMC),vol. 1, no. 1, pp. 23-42, 2018.

25. K. Lakshmi, "Utilization of Data Mining Techniques for Prediction and Diagnosis of Tuberculosis Disease Survivability", International Journal of Modern Education and Computer Science, vol. 5, no. 8, pp. 8-17, 2013.

26. R. SanakalandS. T. Jayakumari, "Prognosis of Diabetes Using Data mining Approach-Fuzzy C Means Clustering and Support Vector Machine",International Journal of Computer Trends and Technology, vol. 11, no. 2, pp. 94-98, 2014.

27. Durgadevi M and Kalpana R, Medical Distress Prediction based on Classification Rule Discovery using Ant-Miner Algorithm, in Proceedings of the IEEE International Conference (11th International Conference on Intelligent Systems and Control, ISCO'17), pp. 88-92, Karpagam College of Engineering and Technology, Coimbatore, India, 2017.

28. Durgadevi M and Kalpana R, A Cooperative GA-SM Prediction Model for healthcare services, e Government Applications for Smart Cities, Inderscience Publishers,2019.

29. Durgadevi M and Kalpana R, Diabetic Risk Prediction based on Association Rule Mining using Soft Sets, International Journal of Pure and Applied Mathematics (IJPAM), Vol. 116, pp. 259-265, 2017.

AUTHORS PROFILE

M. Durgadevi, the author of this paper is currently pursuing her Ph.D. degree in the Department of Computer Science and Engineering at Pondicherry Engineering College, Puducherry, India. She was born in December 1991 at Puducherry. She received her B.Tech. in Computer Science and Engineering from Manakula Vinayagar Institute of Technology, Pondicherry University, Puducherry in 2013, her M.Tech in Network and Internet Engineering from Pondicherry University in 2015.She worked as an Associate in Cognizant Technology Solutions, Coimbatore from July, 2015 till December, 2015. She worked as a Guest Lecturer in the Department of Information Technology at Pondicherry Engineering College during her research period from July,2016 till December,2016. She has published six research papers in reputed International Journals and Conferences. Her research interests include data mining, evolutionary algorithms and cloud computing.

Dr.R. Kalpana, is currently working as a Professor in the Department of Computer Science and Engineering at Pondicherry Engineering College, Puducherry India. She obtained her PhD in Computer Science and Engineering from the Department of Computer Science, Pondicherry University, Puducherry, India, in 2013. She received her M.Tech in the field of Computer Science and Engineering from the Pondicherry University, India, in 1998. She obtained her B.Tech in the field of Computer Science and Engineering from the Pondicherry University, India, in 1996. She is having around 19 years of academic experience and her research interest includes parallel and distributed systems and algorithm.

Retrieval Number: C5680029320 /2020©BEIESP
DOI: 10.35940/ijeat.C5680.029320
Journal Website: www.ijeat.org

2335

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication