International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 08, 2020
ISSN: 1475-7192

Sparse Balanced SVM Based Detection of Type2 Diabetes Using FIIMG Dataset

[1]Dr Naveen Kumar S, [2]Matampalli Satheesh

ABSTRACT

An adequate Type 2 Diabetes unified administration system and regular timely checkup has key role in treatment of Type-2 Diabetes at initial stages. In Recent years there is rapid increase of evolution of Machine learning technique and FIIMG Dataset which is category of Electronic Health Record. Over fitting, Model interpretability and computational cost are the challenges while managing these much of information. Based on these challenges, we proposed a Machine Learning technique called Sparse Balanced Support Vector Machine (SB-SVM) Based Type 2 Diabetes discovering by using Electronic Health record dataset named FIMMG dataset. We have collected data for Type 2 diagnosis from uniform age group that related to Electronic Health records such as exemptions, examination and drug prescription. Machine Learning and Deep neural networks are mainly used in solving task. Results proved that Sparse Based SVM provide better predictive performance and computation time when compared to techniques that are present in existing system. To increase model interpretability, we introduced induced sparsity which manages data which have high dimension.

Keyword: *SB-SVM, FIMMG dataset, Type 2 Diabetes*

I. INTRODUCTION

As per survey of World Health Organization(WHO),worldwide death registered for diabetes will prominently increase around 9%(More than 400 million worldwide deaths) by 2030[3]. In developing nations nearly 50% of diabetic cases goes undiagnosed. The fact is that Type 2 Diabetes Symptoms marked less than of other type of diabetes (Type-1). To prevent, International Diabetes Foundation (IDF) concluded that initial treatment and regular medication will save lives from Type2 Diabetes [2]. Generally, According to National Health System Survey in 2015, People spending more on treatment of diabetes as income spending for the treatment of diabetes is approximately 15% of the income spending for the treatment of other diseases[1] .

[1]*Associate Professor, Department of Computer Science and Engineering, Audisankara College of Engineering & Technology, Gudur, Andhra Pradesh*
[2]*M.Tech Scholor, Department of Computer Science and Engineering, Audisankara College of Engineering & Technology, Gudur, Andhra Pradesh*

Received: 12 Mar 2020 | Revised: 22 Apr 2020 | Accepted: 09 May 2020

A more effective integrated management solution is proposed which have doctors and specialists who are specialists or efficient in their field provides efficient solution to control the health care costs related to diabetes. Mostly all General practitioners clinics are now provided with Electronic Health records which can store health history of patients that contains several heterogeneous information such as Demographic, Monitoring, lifestyle and clinical records. According to Health Information Technology (HIT),The strong investment is due to Electronic Health Records will reduce medical errors ,healthcare costs and morality[4], [5]. The national healthcare and efficiency is increased due to Electronic Health records will decrease use of laboratory tests and outpatient visits [3][6].

Modeling a Decision Support System (DSS) exploit biomedical information to convert Electronic Health record into knowledge as large amount of data is recorded in Electronic Health Records which has large number of patients' data. In this case, Physicians are able to predict the health outcomes by generating patterns that are generalized from Machine Learning models[7]. The main purpose of proposed work is to present sparse balanced Support Vector Machine (SB-SVM), which essential features are extracted from FIMMG dataset which is part of electronic health records that can be used in detection of Type 2 Diabetes. We have collected data related to Electronic Health Records such as exemptions, examination and drug prescriptions before T2D diagnosis which excludes features that revealed in regular checkup for Type 2 Diabetes patients.

Moreover, we considered age group of 40 to 60 years range where patient age is not sporadically admissible in ender that can define Type2Diabetes patients situation .

To measure the reliability of our approach we create three research questions:

- Case I: All set of Electronic Health Records features will be able to predict Type 2 Diabetes by using Sparse Based-SVM Approach.
- Case II: A Subset of Electronic Health Records features collected before Type 2 Diabetes clinical diagnosis will able to predict Type 2 Diabetes by using Sparse Based -SVM. ?
- Case III: A subset of Electronic Health Record features collected before Type 2 Diabetes from a specific age group which is able to predict Type 2 Diabetes by using Sparse Based-SVM.

The paper is structured as follows: Section II explain about brief description of the Literature of Electronic Health records for Type 2 Diagnosis; Section III demonstrate about proposed Sparse Based-Support Vector Machine approach; Section IV explains about Discussion and results; Section V demonstrate about conclusions and future work; Section VI shows references of that paper.

II. LITERATURE SURVEY

In [8], To increase the efficiency of conventional Machine Learning Models such as Decision Tree, Bayesian Algorithm, K-Nearest Neighbor, Zheng proposed a framework that based on defining high level features that can be used in feature extraction implemented by further computational effort.

Received: 12 Mar 2020 | Revised: 22 Apr 2020 | Accepted: 09 May 2020

In [11], A feature selection or extraction has implemented with normal baseline statistical methods such as ReliefF, SVM and recursive elimination feature which will predict Type2Diabetes condition from imbalanced dataset which cannot support the high dimensional data.

In [12], To govern enormous features in finding more occurrences without considering any enhancing feature selection stage their proposed method interacts with classification technique. For the training & validation stage wrapper feature selection approaches such as chi-squared, ANOVA and kendall's algorithm that can increase in computational complexity which is not always linked with learning method.

In [13], Age Marker is the main selective feature in evidencing for diabetic condition to rank the important attributes by ReliefF which is a filter based feature selection strategy, They employed supervised algorithms demonstrated machine learning algorithms such as boosting and bagging to classify the dataset which are able to introduced induced sparsity which manages data that can have high dimension which increases model interpretability.

In [14], the main theme of the paper is to develop an algorithm which can manage unbalanced data and high dimensional data without executing any selection strategies which contains of supplementary features. Their approach perfectly predicts the Type 2 Diabetes patients by considering conventional machine learning models.

In [15], To discover Type 2 Diabetes from 40-60 years age group ,DT model was proposed and Adaboost technique is used to reduce the generalization error while dealing with data that have high dimension.

In [16], SMOTE algorithm is proposed which is used to solve unbalanced class problem that can deal with minority classes. Inducing of sparsity apart from their approach which increase understandably of ML techniques. Moreover, without drawing synthetic samples they can deal with high imbalanced technique which is not consistent and data that are not real class which is not uniform.

In [17],Yu et al.To solve Two classification problems related to condition of type 2 diabetes ,Support Vector Machine model is proposed which is able to discriminate between control subjects and Type 2 Diabetes while dealing with high dimensional data who are not suffering from diabetes is taken into consideration.

In summary, the main contributions of the proposed work are:

- Novel FIMMG Dataset is collected and employed.
- For Treating chronic Type 2 Diabetes core algorithm is designed.
- Sparse Based-Support Vector Machine is proposed which produce better result when compared to existing system.
- The experimental tests evaluated in hospital scenario.

III. PROPOSED SYSTEM

The bound of forecasted Sparse Based-Support Vector Machine was tailored in 0,1 interval by considering sigmoid function without any change of Support Vector Machine error function. The parameters were analysed with

bionomical likelihood by adding post-processing step such it makes mapping realized. The measured probabilistic outputs of Support Based-Support Vector Machine shows the predicted response (yp) based on threshold th = 0.5:

$$P(y_p = T2D|f) = \frac{1}{1+exp(Af+B)} > th \quad y_p = T2D$$
$$else \quad y_p = Control$$

Many Solutions regarding class-imbalance problem was proposed at both algorithmic and data levels. Analysis of different forms of resampling such as directed resample, random resample, oversample with report genesis suggest that we have to use techniques in algorithm level. In order to maximize the macro-recall metric which affects high unbalanced data we have to regulate decision threshold in validation set. The Result of the Sparse Based Support Vector Machine generates an uncalculated measure which is non probability value. The final step grant change the amount of Sparse Based-Support Vector Machine into posterior probability which represents important matter that can combined with Decision Support which helps us to detect early stage of Type 2 Diabetes by figure out the prediction of Type2 Diabetes patients.. The main idea for our proposed approach is to usage of sigma model that can fit posterior directly so that parameters A and B will be able to give best probability outputs. Rather than one parameter, the proposed sigma model has mainly two parameters which analysed that can give complete idea about prediction[18].

The main theme for application of the proposed system may focus on two different challenges:

(i) Majority degree of redundancy and data with high dimension will have several noisy data .

(ii) The introduction of First Norm LASSO regularised that can allow inducing the sparsity in increasing the intractability of linear regression technique which can predict result but also it tells about why and how the prediction was made. Here we have to reduce variance of class from some efficient to zero which can improve the prediction accuracy by LASSO regularizer.

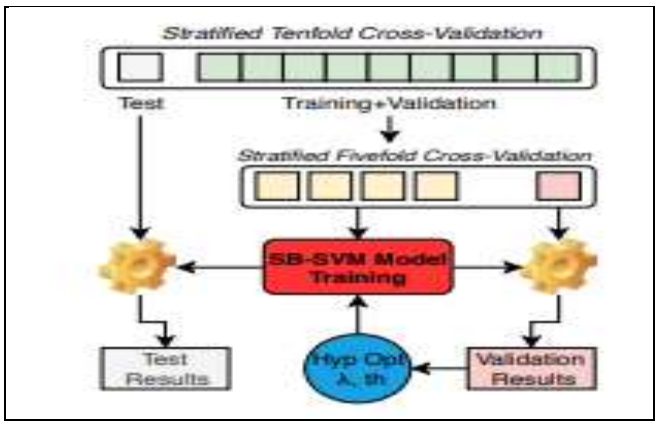

Fig.1: Proposed SB-SVM Approach

The main purpose of proposed work is to present sparse balanced Support Vector Machine (SB-SVM), which essential features are extracted from FIMMG dataset which is part of electronic health records that can be used in detection of Type 2 Diabetes. We have collected data related to Electronic Health Records such as exemptions, examination and drug prescriptions before T2D diagnosis which excludes features that revealed in regular checkup for Type 2 Diabetes patients.

Moreover, we considered age group of 40 to 60 years range where patient age is not sporadically admissible in ender that can define Type2Diabetes patient's situation[9][10].

IV. RESULTS AND DISCUSSION

We experimented the proposed Sparse Based-Support Vector Machine in order to explain answers that can be explored in existing system and we correlated our work with ML algorithms such as Bagging, K-NN, Linear discriminative model, Gauss, Linear regression model and Deep Learning projects such as Single Layer Perceptron (SLP) and Deep Neural Network already proposed in the finding result of osteoporosis respectively.

Work	Model	Recall%		AUC%	
	Baseline	mean	std	mean	std
[16], [18], [21]	SVM Lin	74.12	(4.02)	81.68	(5.60)
[16], [18], [21]	SVM Gauss	71.96	(4.22)	81.98	(4.84)
[18]	KNN	69.23	(4.97)	70.97	(5.06)
[17], [18], [20]	DT	80.99	(3.34)	87.79	(4.17)
[18], [20], [22]	RF	77.81	(5.66)	86.30	(4.24)

Table-I:Proposed System Compared with existing approaches

The Evaluation of our Sparse Based Support Vector Machine was evaluated by using following conditions:

1) Quality: Quality is computed by considering non-weighted mean for every class.

2) Reminiscence: Reminiscence is computed by considering non-weighted mean for every class.

3) ROC: It is drawn by considering FPR on x-axis and TPR on y-axis by considering various threshold values.

4) AUC: AUC denotes Classifier in which rank can be choosen as positive number is greater than that of negative number

5) L_0 Measure :It computes based on the zero coefficients that presented in the model which can denotes as sparsity.

Received: 12 Mar 2020 | Revised: 22 Apr 2020 | Accepted: 09 May 2020

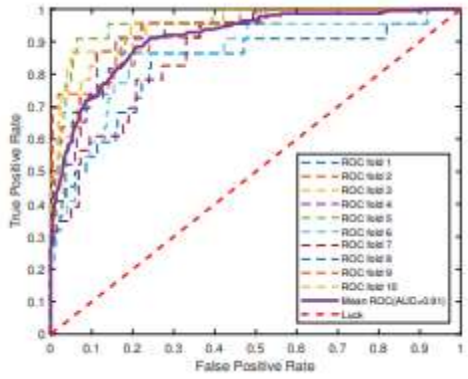

Fig 2. ROC Curves For Different System Approaches

When Compared to precision, Recall achieved best performance in the optimization of validation set. The ROC curves for different system approaches are shown in above figure and it shows the averaged curve of each and every fold in which all considerations of ROC curve will lies greater than the chance level which is denotes as red point line in the above figure. The AUC and reminiscence of Sparse Based-Support Vector Machine are compared with KNN, DT.LR, SVM Lin and SVM Gauss which can perfectly analyses about Type2 Diabetes patient's condition. The AUC and reminiscence for the Sparse Based-Support Vector Machine are slightly greater than, RF (recall: t18 = 0.514, p = .09),DT (recall: t18 = 0.514, p = .61; AUC: t18 = 1.652, p = .12) and The Sparse Based Support Vector Machine is superior than other baseline methods by analysis of ROC Curves(Figure 2). The Sparse Based Support Vector Machine surpasses ($p < .05$) conventional machine learning models .The Roc Area and recall of Sparse Based-Support Vector Machine compared with ReliefF, Support Vector Machine, once comparison is done the proposed Sparse Based-Support Vector Machine identify features that contribute to the decision that clearly demonstrate between normal people and Type 2 Diabetes patients. The Sparse Based-Support Vector Machine offers a particular technique of finding our main aim, as SB-SVM is linear and easily interpretable when compared with existing approaches. The computed l_0 measure was 0.52[19].

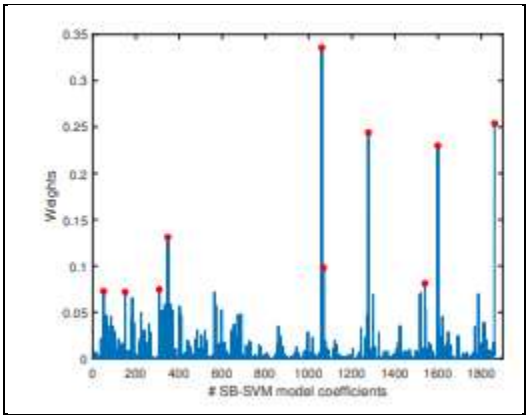

Fig.3 SB-SVM coefficients magnitude and l_0 measure values.

The proposed system and the results shown earlier bring to the following main discussion points:

Received: 12 Mar 2020 | Revised: 22 Apr 2020 | Accepted: 09 May 2020

A. Hospital Use case: The standard diagnosis of Type 2 Diabetes remains challenging and uncertain as the management of diabetic patient is complex and complicated. At this stage, Decision Support System solution provide perfect prediction of Type 2 Diabetes that becomes essential in order to get an idea about prevention strategies that can be applied to sort out T2D patients. According to the survey, Machine Learning algorithms will be able to identify the disease based on symptoms which can prevent spreading disease. From these main points we proposed Sparse Based-Support Vector Model to predict which can predict Type 2 Diabetes patients by using FIIMG Dataset.

B. Sparse Balanced Support Vector Machine efficiencies and results : The experimental results disclose that how Electronic Health Record features are related to early stage of Type 2 diabetes prediction. Here in this case particular point of each technique is sparsely distributed but not unchanging inform. we compared our proposed approach Sparse Based Support Vector Machine with Machine Learning algorithms especially Linear discriminate model, Bagging, Gauss, Linear regression model and Deep Learning projects such as Single Layer Perceptron (SLP) and Deep Neural Network already proposed in the guessing of result which generate initial prediction of T2D. The performed correlations deals about imbalanced setting of the assignment. LASSO regularizes an efficient strategy that can significantly decreases method complexity. FIIMG dataset consists of different unrelated components: i) mislaid values ii) anomalies iii) Doctor transaction types and v) sporadic nature in pathologies.

C. Hospital usecase impact: Decision Support System will provide prediction in order to support the early-stage diagnosis while some hidden patterns unsighted by physicians. An experimental result shows that robustness of the Sparse Based-Support Vector Machine methodology in the clinical scenario. On selecting FIIMG Dataset the model selection criteria produces great reusability and more impact on generated result. Moreover, the method coefficients of sparsity are likely to be change across different Electronic Health Record dataset and different task [20].

V. CONCLUSION AND FUTURE WORK

An adequate Type 2 Diabetes unified administration system and regular timely checkup has key role in treatment of Type-2 Diabetes at initial stages. In Recent years there is rapid increase of evolution of Machine learning technique and FIIMG Dataset which is category of Electronic Health Record. Over fitting, Model interpretability and computational cost are the challenges while managing these much of information. Based on these challenges, we proposed a Machine Learning technique called Sparse Balanced Support Vector Machine (SB-SVM) Based Type 2 Diabetes discovering by using Electronic Health record dataset named FIMMG dataset. We have collected data for Type 2 diagnosis from uniform age group that related to Electronic Health records such as exemptions, examination and drug prescription. Machine Learning and Deep neural networks are mainly used in solving task. Results proved that Sparse Based SVM provide better predictive performance and computation time when compared to techniques that are present in existing system. The SB-SVM approach easily predicts output which is used in linear regression technique.

Received: 12 Mar 2020 | Revised: 22 Apr 2020 | Accepted: 09 May 2020

Future work of this paper is application of nonlinear kernel as Gaussian or polynomial such that it matches main features that is present in non linear space. Imposing LASSO is the solution which can extract main features from non linear space.

REFERENCES

[1]Alejandra, "Arterial evaluation in type 2 diabetes mellitus ", Annual International Conference of the IEEE Engineering in Medicine and Biology ,vol.978,no. 10,pp.3584-3592,2010.

[2] Faezeh marzbanrad ,slade Mathews,chandan karmakar, "Methodological comparsions of Heart Rate Variablity Analysis in patients with Type 2 Diabetesand Angiotensin converting Enzyme Polymorphism",IEEE Journal of Biomedical and Health Informatics,Vol.10,no.10,2168-2194,2015.

[3] Kei sato,kazua yana,"Long and short term QT-RR Interval covariablity in Type 2 Diabetes",IEEE Journal on Health Informatics,Vol.02,no.4,38-41,2014.

[4]Yang Guo,Yan Hu,"Using Bayes Network for prediction of Type 2 Diabetes",International Conference For Internet Technology and secured transcations, vol. 145, no. 11, pp. 162–172, 2012.

[5] Huiye,Li, "An Analysis and Research of Type 2 Diabetes TCM Records Based on Text Mining",BIBM vol. 165, no. 15, pp. 1417–1426, 2018.

[6] Gao Juan,Han Yi-wen "Type 2 Diabetes data processing with EM and C4.5 Algorithm, vol. 172, no. 3, pp. 114–120, 2007.

[7] Mohammed Alotaibi, "A Mobile Diabetes educational system for fasting Type 2 diabetics in Saudi arabia," Health Affairs, vol. 30, no. 4, pp. 360–363, 2015.

[8] Mehri,Hammani, "ACE gene I/D polymorphism and obesity in patients with type 2 diabetes mellitus" ICEMIS, vol. 30, no. 3, pp. 353–363, 2017.

[9] Moreno,Miquel "Type 2 Diabetes screening test by means of pulse oximeter" IEEE Transactions on Biomedical engineering, vol. 370, no. 18, pp. 1498–1499, 2017.

[10] Bae,Park,"Risk Prediction Using common rare genetic variants",BIBM, vol. 370, no. 18, pp. 1412–1417, 2017

[11] Hani Hagras,Heba Mohammed "Towards Developing Type 2 Fuzzy Logic Diet Recommendation system for Diabetics" 10th Computer Science and Electrical Engineering, vol. 378, no. 16, p. 1219, 2018.

[12] Fellaji,Azmani,Akharif "Bayesian approach for minimizing nepropathy risk of patients with type 2 diabetes" SITA-14, vol. 25, no. 8, pp. 1595–1610, 2014.

[13] Lawi,Tharamani, "A Premilinary study on identifying probable bio marker of Type 2 Diabetes using Recursive Feature Extraction" EIconCIT, vol. 25, no. 8, pp. 1595–1610, 2018.

[14] S.D.Patek,L.Magni, "Modular Closed Loop Control of Diabetes," IEEE Transactions on Biomedical Engineering, pp. 830–834,2012.

[15] Bipan Tudu,Chakraborthy, "Comparsion of ANN Models to predict LDL level in Diabetes Mellitus type 2," IEEE Journal of Biomedical Informatics and Health, vol. 54, pp. 290–303, 2010.

[16] Hannah Al Ali ,Wiam Boutayeb , Abdesslam Boutayeb , Nora Merabet, "A mathematical model for type 1 diabetes, on the effect of growth hormone**",** 8th International Conference on Modeling Simulation and Applied Optimization, vol. 43, no. 1, pp. 54–59, 2019**.**

[17] Oana Geman ; Roxana Toderean ; Maria Magdalena Lungu , "Challenges in nutrition education using smart sensors and personalized tools for prevention and control of type 2 diabetes", IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME), vol. 44, no. 4, pp. 262–272, 2017**.**

[18] Frank Nani,Jin"Mathematical modeling and simulations of the pathophysiology of Type 2 Diabetes Mellitus"8th International conference on Bio Medical Engineering and informatics, vol. 123, pp. 124–131, 2015.

[19] Raid M Khalil,Adel Al Jumaly, "Machine Learning Based Prediction of depression among type 2 diabetics," ISKE, vol. 57, no. 4, pp. 69–110, 2017.

[20] Tushar Ghorpade,Vimla Jethani, "Diabetic retinopathy detection through image mining for type 2 diabetes" International Conference on Computer Communication and Informatics (ICCCI), vol. 85, pp. 121–127, 2017.

Received: 12 Mar 2020 | Revised: 22 Apr 2020 | Accepted: 09 May 2020