International Journal of Research in Engineering, Science and Management
Volume-3, Issue-2, February-2020
www.ijresm.com | ISSN (Online): 2581-5792

Learning Techniques - Classification of Insulin-Dependent Diabetes Mellitus Among Adults

I. Carol[1], D. Richard[2]
[1,2]Assistant Professor, Department of Information Technology, St. Joseph's College, Trichy, India

***Abstract*: Insulin-dependent diabetes mellitus (IDDM) is a set of correlated diseases in which the body cannot regulate the amount of sugar in the blood. It mainly affects the adult and characterized by chronic hyperglycemia associated with disturbances of carbohydrate, fat, and protein metabolism due to absolute or relative deficiency in insulin secretion and/or action. It causes long term damage, dysfunction and failure of various organs such as eyes, kidneys, nerves, heart and blood vessels. A new methodology is used to find the stages of Insulin-dependent diabetes mellitus using Convolutional Neural Network (CNN). The symptoms and stages of Insulin-dependent diabetes mellitus are classified by using CNN technique. The hidden layers of a CNN typically consist of convolutional layers, pooling layers, fully connected layers and normalization layers. If you have fully-connected layers at the end of your convolutional network, implementing dropout is easy. It helps us to know the various stages of Insulin-dependent diabetes mellitus and to predict the recommend preclusion to patients those who are affected by Insulin-dependent diabetes mellitus and provide implication to that patient.**

***Keywords*: Convolutional neural network (CNN), Insulin-dependent diabetes mellitus (IDDM).**

1. Introduction

Insulin-dependent diabetes mellitus is a disorder caused when the body doesn't make enough insulin. This high blood sugar produces the classical symptoms of polyuria (frequent urination), polydipsia (increased thirst) and polyphagia (increased hunger). If left untreated, diabetes can lead to blindness, kidney disease, nerve disease, heart disease, and stroke. Insulin-dependent diabetes mellitus often simply referred to as diabetes is a condition in which a person has high blood sugar, so that the body needs a special concern and provides extensive solutions to remain staying healthy, effervescent and lucid. This learning finds the knowledge by using supervised and unsupervised learning algorithms such as, Convolutional neural network. To find the efficiency of neural network technique by using the Root Mean square error and Mean absolute error [4], [5]. Beyond that the low level of attributes in Patient's records falls in border phase of IDDM. Convolutional neural network has been employed to envisage the knowledge about the disease. Symptoms and stages are a class variables used for classification [19], [20]. A network composed of more than one layer of neurons, with some or all of the outputs of each layer connected to one or more of the inputs of another layer. The first layer is called the input layer, the last one is the output layer, and in between there may be one or more hidden layers.

A. IDDM - approach

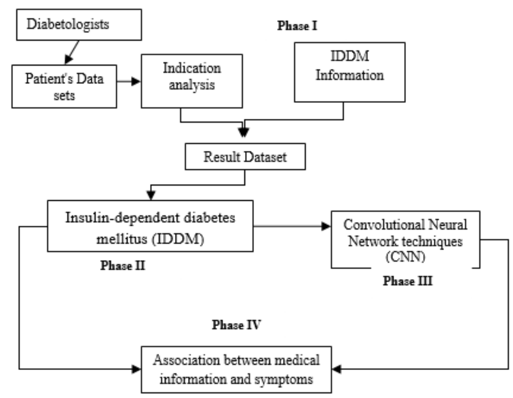

Fig. 1. Overall architecture diagram

In Phase I, the patient's dataset has been collected from diabetologists. The patient's dataset contains the symptoms of IDDM such as Excessive thirst, Fatigue, hay fever, Dry Skin, blurred vision, sudden weight loss, tingling hands and feet etc. The information has been examined and used in this experiment. In this dataset, demographic information is gathered from the patients. The result dataset contains the demographic and patient's symptoms information is used for analysis [17], [18].

In Phase II and III, Predictive and Descriptive learning methods are used to find the impact of disease in the urban and rural area. The methods are one of the Convolutional neural network respectively and it is allocated to IDDM stages of the patients [15]. The six sigmoid nodes are used as inputs, the weights are assigned to each node, and their six output layers are classified based on IDDM stages. According to the stages based classification, the ten sigmoid nodes are used as input and three output layers are classified based on the stages of IDDM [12]. In Phase IV, the association between symptoms and IDDM stage is analyzed using motivating measures.

2. Review of related modeling impetus

A. Hidden layer activation function

To use a logistic (sigmoid) activation function for the hidden layers. A logistic function is recommended. Here Figure: 2

show the plot of a logistic activation function:

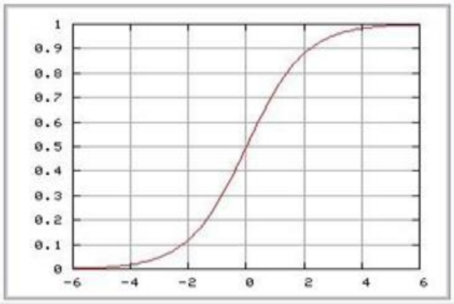

Fig. 2. Logistic Activation Function

B. The Output Report Generated Using IDDM dataset

1) Project Parameters

============ Project Parameters ============
Target variable: STAGES OF IDDM
Number of predictor variables: 9
Type of model: convolutional Neural Network (CNN)
Number of layers: 3 (1 hidden)
Hidden layer activation function: Logistic
Output layer activation function: Logistic
Type of analysis: Classification
Category weights (priors): Data file distribution
Misclassification costs: Specified cost matrix
Validation method: Random sampling (90%)
Input data file: J:\IDDM documents\DATASET.csv
Number of variables (data columns): 10
Data sub setting: Use all data rows
Number of data rows: 300

The Project Parameters section of this IDDM report displays a summary of the options and parameters user can selected on the various property pages for the model. It reveals that the classification techniques are employed for finding the predicted knowledge using CNN. In this model, nine predictor variable are used and one target for variable for prediction. Random sampling method is used for validating the data. Only three layers are used for finding the knowledge from this modeling using logistic activation.

2) Summary of variables

The Table 1 displays information about each variable in the IDDM dataset. The first column shows the name of the variable, the second column shows how the variable was used; the possibilities are Target, Predictor, Weight and Unused. The third column shows whether the variable is categorical or continuous, the forth column shows how many data rows had missing values on the variable, and the fifth column shows how many categories (discrete values) the variable has. In the case of continuous variables, the number of categories (such as Patients and Stages of IDDM) will be limited by the value specified for "Max. Categories for predictor variables" on the model design property page.

Table 2
Classification Table for Training Data

Category	Actual category		Misclassified category		Percentage of cost	
	Count	weight	Count	weight	percent	cost
Primary	99	99	0	0	0	0
Nonsevere	63	63	0	0	0	0
Severe	108	108	0	0	0	0
total	**270**	**270**	**0**	**0**	**0**	**0**

Table 3
Classification Table for Validating Data

Category	Actual category		Misclassified category		Percentage of cost	
	Count	weight	Count	weight	percent	cost
Primary	11	11	0	0	0	0
Nonsevere	7	7	0	0	0	0
Severe	12	12	0	0	0	0
Total	30	30	0	0	0	0

3) Classification Summary

It reveals that misclassification does not exist in the validation. The cost and weight information are tabulated based on classification in Table 2 and 3.

In this dataset, 30 objects are used for training and 270 objects are used for testing. Each category is classified and finds the weights are calculated using convolutional neural network.

4) Confusion Matrix Table

A "Confusion Matrix" provides detailed information about how data rows are classified by the model. The matrix has a row and column for each category of the target variable. The categories shown in the first column are the actual categories of the target variable. The categories shown across the top of the table are the predicted categories. The numbers in the cells are the weights of the data rows with the actual category of the row and the predicted category of the column. Here table 4 shows the IDDM datasets- confusion matrix. The numbers in the diagonal cells are the weights for the correctly classified cases where the actual category matches the predicted category. The

Table 1
Summary of Variables in IDDM datasets Using DTREG

No.	Variables	Class	Type	Missing Rows	Categories
1	Patients	Predictor	Categorical	0	30
2	Excessive thirst	Predictor	Continuous	0	12
3	Urinary Infection	Predictor	Continuous	0	2
4	Wrinkles	Predictor	Continuous	0	2
5	Skin problems	Predictor	Continuous	0	2
6	Fatigue	Predictor	Continuous	0	2
7	Hunger	Predictor	Continuous	0	2
8	Hayfever	Predictor	Continuous	0	2
9	Yeast infections	Predictor	Continuous	0	2
10	Stages of IDDM	Target	Categorical	0	3

off-diagonal cells have misclassified row weights. For IDDM dataset, the Non severe category was slightly misclassified as Primary and severe category.

Table 4
Confusion Matrix

	Actual category	Predicted Category		
		Primary	Non severe	Severe
Training Data	Primary	11	0	0
	Non Severe	0	7	0
	Severe	0	0	12
Validation Data	Primary	99	0	0
	Non Severe	0	63	0
	Severe	0	0	108

5) Variable importance table

The variable importance table gives a ranking of the overall importance of the predictor variables.

Table 5
Variables of Importance

S. No.	Variables	Importance
1	Excessive thirst	71.734
2	Urinary Infection	34.509
3	Wrinkles	33.083
4	Skin problems	33.083
5	Fatigue	30.584
6	Hunger	10.567
7	Hayfever	8.731
8	yeast infections	0.566

Importance scores are computed by using information about how variables were used as primary splitters and also as surrogate splitters. If a primary splitter is slightly better than a surrogate, then the primary splitter may "mask" the significance of the other variable. By considering surrogate splits, the importance measure calculated by giving more accurate measure of the actual and potential value of a predictor [21]. To get the most accurate measure of importance, the user should select the option "Always compute surrogate predictors" on the Missing Data property page. The importance score for the most important predictor is scaled as 71.734. Other predictors will have lower scores. Only predictors with scores greater than zero.

3. Results and Discussion

Based on the data set, the diagnosis was made by a physician with training and qualifications in diabetologists. For the purpose of this study, children and teenagers with IDDM and atopic dermatitis are excluded. Standard treatment is advised for the children and teenagers which is the same as children and teenagers with IDDM. The physician's interpretation of clinical data and clinical images are stored in the medical databases [22], [23]. An expert medical knowledge and specialized learning techniques to understand the meaning of unstructured data explanation. The dataset contains 300 patients' objects. The technique has been employed to categorize the patients and their symptoms using predictive modeling software and it evaluates based on the errors occurred in the classification, which is shown in table 6.

Table 6
Error Calculation

S. No.	Types of Error	Value
1.	Mean Absolute Error	0.0125
2.	Root Mean Squared Error	0.0133
3.	Relative Absolute Error	0.5362
4.	Root Relative Squared Error	0.7089

Here the main notified errors are root mean squared and mean absolute error, which are minimum in this model. This reveals that the model classifies the dataset perfectly. The finding divulges that densely populated group of patients with close similarities based on the stages of IDDM are classified.

4. Conclusion

In this Paper, popular learning methods are used to predict the patient information. In Convolutional neural network technique, the three stages are identified as primary, non-severe and severe [8]. It divulges that there is a perfect classification. Standard treatment is advised for the IDDM affected patients [3] especially for adults (>20). It comprises general recommendation on diet changes, regular exercise, possibly insulin shots and hormone tablet for acute flare-ups of the patients [14]. The correlation study unveils that the symptoms and stages have strapping association. The association between each entity is identified by using classification [16].

References

[1] Cardona, F., Morcillo, S., Gonzalo-Martin, M. and Tinahones, F. (2005). The apolipo protein E genotype predicts postprandial hypertrigly ceridemia in patients with metabolic syndrome. Journal of Endocrinology and Metabolism, 90(5), 2972-2975.

[2] Dandona, P., Aljada, A., Chaudhuri, A., Mohanty, P. and Garg, R. (2005). Metabolic syndrome: A comprehensive perspective based on interactions between obesity, diabetes and inflammation. Circulation, 111(11), 1448-1454.

[3] Homer J, Jones A, Seville D, Essien J, Milstein B, Murphy D. 2004. The CDC diabetes system modeling project: Developing a new tool for chronic disease prevention and control. 22nd International Conference of the System Dynamics Society, Oxford, England.

[4] Ida J. Hatoum, Frank B. Hu, Jeanenne J. Nelson, and Eric B. Rimm, "Lipoprotein-Associated Phospholipase A2 Activity and Incident Coronary Heart Disease Among Men and Women with NIDDIDD Miabetes", Diabetes, vol. 59, May 2010.

[5] C. Deeb, "Diabetes Technology During the Past 30 Years: A Lot of Changes and Mostly for the Better", Diabetes L Spectrum, 2008. 21, pp 78- 83, 2008.

[6] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P.W. Duin, "Is independence good for combining classifiers?," in Proc. Int. Conf. Pattern Recognition (ICPR), vol. 2, Barcelona, Spain, 2001, pp. 168–171.

[7] Miller, A., Blott, B., & Hames, T, "Review of Neural Network Applications in Medical Imaging and Signal Processing. Medical and Biological Engineering and Computing", (1992)30(5), 449- 464.

[8] Moriarty DG, Zack MM, Kobau R. 2003. The Centers for Disease Control and Prevention's Healthy Days measures: Population tracking of perceived physical and mental health over time. Health and Quality of Life Outcomes 1(1): 37.

[9] S. H. Min, I. Han: Optimizing Collaborative Filtering Recommender Systems Lecture Notes in Artificial Intelligence vol. 3528, 2005, pp. 313–319.

[10] Stewart WF, Ricci JA, Chee E, Hirsch AG, Brandenburg NA (June 2007). "Lost productive time and costs due to diabetes and diabetic neuropathic pain in the US workforce". J. Occup. Environ. Med. 49(6):672–9.

[11] Tang, W., Hong, Y., Province, M., Rich, S., Hopkins, P., Arnett, D., Pankow, J., Miller, M. and Eckfeldt, J. (2006). Familial clustering for features of the metabolic syndrome. Diabetes Care, 29(3), 631-636.
[12] Williams. D. Prevost, T., Whichelow, M., cox, B., Day, N. and Wareham, N. (2000). A cross-sectional study of dietary patterns with glucose intolerance and other features of the metabolic syndrome [Abstract]. British Journal of Nutrition, 83(3), 257-266.
[13] Wilson, P., D'Agostino, R., Parise, H., Sullivan, L., and Meigs, J. (2005). Metabolic Syndrome as a precursor of cardiovascular disease and NIDD IDDM diabetes mellitus. Circulation, 112, 3066-3072.
[14] Adarsh, P., Jeyakumari, D, Multiclass SVM-based automated diagnosis of diabetic retinopathy. In: Communications and Signal Processing (ICCSP), 2013 International Conference on. IEEE; 2013, pp. 206–210
[15] Wong WL, Su X, Li X, Cheung CMG, Klein R, Cheng C-Y, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. Lancet Glob Health. 2014; 2, pp. 106–116.
[16] Abràmoff MD, Lou Y, Erginay A, Clarida W, Amelon R, Folk JC, et al. Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning. Invest Ophthalmol Vis Sci. 2016; 57: 5200–5206.
[17] E. Renard, J. Place, M. Cantwell, H. Chevassus, and C. C. Palerm, “Closed-loop insulin delivery using a subcutaneous glucose sensor and intraperitoneal insulin delivery feasibility study testing a new model for the artificial pancreas,” Diabetes Care, vol. 33, no. 1, pp. 121–127, 2010
[18] R. Hovorka, “Closed-loop insulin delivery: From bench to clinical practice,” Nature Rev. Endocrinol., vol. 7, no. 7, pp. 385–395, 2011.
[19] Yingjie Zhang Byron C. Wallace, A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification IJCNLP, 2017.
[20] Y. Kim, “Convolutional neural networks for sentence classification, 2014.
[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105.
[22] Akm Ashiquzzaman, Abdul Kawsar Tushar, Md Rashedul Islam, Dongkoo Shon, Kichang Im, Jeong-Ho Park, Dong-Sun Lim, Jongmyon Kim Reduction of overfitting in diabetes prediction using deep learning neural network. IT Convergence and Security, Springer (2018), pp. 35-43.
[23] Michael A. Pfeifer, Daniel Cook, Joel Brodsky, David Tice, A Reenan, Sally Swedine, Jeffrey B. Halter, Daniel Porte, “Quantitative evaluation of cardiac parasympathetic activity in normal and diabetic man Diabetes, 31 (4), (1982), pp. 339-345.