An Intelligent Diagnostic Framework For Healthcare Services Using Optimal Deep Learning Model

K. Veerasekaran, P. Sudhakar

Abstract: Presently, people affected by diabetes mellitus (DM) get increased in every single day. Many of the diabetes patients are completely unaware about the health issues or risks factors related to it. This paper presents a new classification model for predicting type II DM (T2DM). The proposed model intends to enhance the classifier rate and build the model highly adaptable to numerous dataset. Using a sequence of pre-processing steps, the presented model comprises of two processes namely clustering and classification. Initially, K-means clustering technique is employed to cluster the dataset. Then, the logistic regression (LR) classifier is employed for data classification. Here, the parameter tuning of LR with deep learning models such as Adam and Root Means Square (RMS) Prob model takes place. The experimental validation using PIMA Indians diabetes dataset is employed and detailed investigation takes place under diverse domains. The simulation outcome indicated that the presented LR-Adam with after K means clustering model shows outstanding results on the applied dataset under several validation parameters.

Index Terms: Diabetes, Clustering, Classification, LR, K-means.

1 INTRODUCTION
Diabetes mellitus (DM) is one of the commonly available chronic diseases which undergo characterization of high blood glucose. Almost 50% of every diabetic patient has genetic problem which is an important cause of DM. The improper functioning of the pancreas generating adequate insulin and the body's incompetent usage of insulin are equally pathologic reasons for the occurrence of DM. A set of two varieties of DM exists namely type 1 and type 2 DM referred as T1DM and T2DM [1]. The pathogenesis of T1DM represents the secretion of injured β-cells in the pancreas eliminating it from reaching the minimum blood glucose level in time. The secretion deficiency and resistivity of insulin causes the T2DM, also termed as non-insulin dependent DM. From the previous decades, the number of diabetes patients is exponential increased globally as shown in Fig. 1. Besides, it is noted that the ratio of male diabetic people is comparative higher than the female diabetic persons. Based on a study made by authorized statistics, people affected by diabetes in China reaches around 110 million in 2017. It indicates that the China is the highest diabetic affected people worldwide. The International Diabetes Federation (IDF) presented a recent report on DM in Diabetes Atlas revealing that around 415 million gets affected by diabetes in 2015.

With respect to the rise in diabetic patient count, it is predicted that the number reaches to 642 million, or 1:10 ratio. For lowering the morbidity and decrease the impact of DM, it is very important to concentrate on maximum risk group of people with DM. Based on a recent study by World Health Organization (WHO), the higher risk of DM exists under the following factors: age less than 45, irregular physical activities, heredity, Hypertension or heart disease or gestation female. For analyzing the high-risk groups of DM, it is needed to make use of recent information technology. So, data mining techniques has found to be a proper tool to analyze the study. Data mining is a computation procedure used to discover the patterns exist in massive dataset by including approaches in the field of machine learning, database and statistics [2]. A major aim of these approaches is clustering, classification, association and so on. It holds a sequence of procedure includes in an automatic or partially automatic way for extracting and exploring fascinating, concealed characteristics from massive data quantity. The high data quality and effectively employed models forms the two basic views of data mining. It is employed to distinct domains like weather prediction, market examination, engineering analysis, and customer link management. But, the applicability in healthcare is yet to be improved. For instance, each hospital includes an overabundance of fundamental and healthcare details related to the patient. It is needed to reform, and filter useful data for performing clinical examination and diagnosis [3,4]. It is
practical to believe that there are a mixture of expensive patterns and coming up for researcher people to discover it. Among the huge number of massive number of diabetes affected people, most of them do not aware of their health quality. So, it is believed to develop a model which classifies the patients to suspected or confirmed patients in the last five years from the initial investigation period for high-risk in DM group. Presently, data mining techniques are employed in an increased rate for predicting the probability of the disease [5]. Various models and techniques have been developed by research people and marked the significant area of the research. By analysing various research works, it is observed that the Pima Indians Diabetes Dataset is commonly employed. [6] introduced a hybrid prediction model (HPM) which makes use of K-means clustering technique intended to validate the selected class labels of provided data and employed C4.5 model intended to build a final classification technique. [7] underwent a comparison of the classifier results of the multilayer perception (MLP) over ID3 and J48 models. The outcome indicated that the pruned J48 tree model attains superior classification. [8] introduced artificial metaplasticity on MLP (AMMLP) as a classification model for diabetes which attains the optimal accuracy of around 90%. For obtaining helpful and proper data, a pre-processing method is introduced and the variable selection takes place in a proper way. [9] underwent an analysis of the advantages of various pre-processing models to predict DM. It is observed that the pre-processing models enhanced the classifier accuracy. [10] investigated the risk factors of T2DM depending upon the FP-growth and Apriori models. [11] presented a receiver operating characteristic (ROC) area, sensitivity, and specificity measures are applied for validation. Based on the effective prediction model, a proper model is needed to make it comfortable in various situations. [12] introduced an android application for resolving the limited awareness of DM. It utilized DT model for predicting the level of DM. It also offers suggestions to the patients related to DM. This paper presented a new classification model for predicting type II DM (T2DM). The proposed model intends to enhance the classifier rate and build the model highly adaptable to numerous dataset. Using a sequence of pre-processing steps, the presented model comprises of two processes namely clustering and classification. Initially, K-means clustering technique is employed to cluster the dataset. Then, the logistic regression (LR) classifier is employed for data classification. Here, the parameter tuning of LR with deep learning models such as Adam and Root Means Square (RMS) Prob model takes place. The experimental validation using PIMA Indians diabetes dataset is employed and detailed investigation takes place under diverse domains. The simulation outcome indicated that the presented LR-Adam with after K means clustering model shows outstanding results on the applied dataset under several validation parameters.

2 PROPOSED MODEL

2.4 Pre-processing
In predictive models, data quality plays a predominant role. The WEKA tool compresses many filter types to pre-process the data. Here, few approximation models are used for the optimization of the actual dataset. Initially, an analysis takes place on every feature medical implications and its relationship with DM. In addition, the pregnancy count related to the DM is also determined. So, a transformation takes place from numerical to nominal attributes. The value of '0' represented absence of pregnancy and '1' denotes the presence of pregnancy. Next, the missing as well as wrong values present in the dataset because of error and irregulation. Many of the imprecise simulation outcomes take place by useless values. For instance, in the actual dataset, the blood pressure and BMI values should be 0 indicating that the real values are missed. For reducing the impact of missing values, average value from the training data is used for replacing every missing value.

2.2 K-means clustering technique
K –means [13] is a famous clustering model that gathers \( n \) data points to \( K \) clusters through minimizing the data point distance from \( K \) cluster heads in a repeated way. It is a type of unsupervised learning technique employed to solve the clustering issue. It is a simple and easy process which classifies the provided dataset to a specified cluster count. The fundamental concept lies in the definition of \( k \) clusters, one for every individual cluster. The position of centres takes place in a crucial way because of the nature of diverse locations which results in distinct outcome. So, the optimal model lies in the placement of keeping it farther from one another. The next step considers every point belonging to a provided dataset and links to its nearest cluster centre. Upon the absence of completed points, the beginning process is carried out and earlier group age is also carried out.

\[ J = \sum_{i=1}^{c} \sum_{j=1}^{c} (\|x_i - v_j\|)^2 \]

In this point, it is needed to carry out a determination of \( k \) novel centroids as barycentre of the clusters attained from the earlier step. Once a set of \( k \) new centroids are computed, an effective binding procedure is carried out among the identical dataset points and nearest new centre. A creation of loop takes place. As a result, it is observed that the \( k \) centres alter the locations in a sequential way until no modification takes place. Finally, this model aims to the minimization of the objective function known as squared error function and is represented as follows:
where \( \|x_i - v_j\| \) represents the Euclidean distance among \( x_i \) and \( v_j \), \( c_i \) denotes the data point count in \( i^{th} \) cluster and \( c \) represents of cluster centre count. The determination of distance takes place by Euclidean distance or cosine measure. In K-means model, the procedure of choosing the value of \( K \) for a dataset with undetermined class count is a crucial process. At this point, elbow model, information criterion method and so on are applied for the selection of \( K \) value. The advantages of this method are faster, simple to implement, robust and offer optimal performance in case of distinct dataset. Fig. 2 shows the clustering outcome of k-means technique under \( N=60 \) and \( c=3 \).

2.3 Logistic regression algorithm

A classifier intends to construct a model which maps the data items to a provided class depending upon the present data. It is applied in the extraction of important data items from the model or predicting the nature of the data. Under various cases, the dependent variable of the LR model is binary-classification. It indicates that the LR model is generally applied for solving the two-category difficulty. A major aim of this simulation lies in the prediction of the presence of diabetes or not, that is generally a binary-classification problem. In addition, the LR model is generally employed in data mining, diagnosis of diseases, particularly classification of healthcare data. As a summary, LR is applied for the prediction of existence or non-existence of diabetes disease. LR model depends upon the linear regression model as shown in Eq. (1):

\[
P = \alpha + \beta_1x_1 + \beta_2x_2 + \ldots + \beta_mx_m
\]

A classification issue is almost comparable to the linear regression issue which predicts only the continuous values. It manages uniform sensitivity all through the real number field. The prediction values of any classifier could be \( 0/1 \). The outcome will be \( 1 \) when the value is higher than the threshold; else the outcome will be \( 0 \). The output parameter range of LR lies in the range of \([0,1]\). Depending upon LR, it appends a sigmoid function layer (non-linearity). The characteristics are initially totalised in a linear way and prediction takes place utilizing sigmoid function. Some important formulas are provided here:

\[
Pr(Y = +1|X) = \beta X \text{ and } Pr(Y = -1|X) = 1 - Pr(Y = +1|X)
\]

\[
\downarrow \sigma(x) := \frac{1}{1 + e^{-x}} \in [0,1]
\]

\[
Pr(Y = +1|X) = \sigma(\beta X) \text{and } Pr(Y = -1|X) = 1 - Pr(Y = +1|X)
\]

Here, a classification takes place under positive and negative classes. Here, \( Y \) represents the presence of diabetes. \( X \) independent variables indicates the set of 8 attributes in the actual dataset. Each dependent variable \( X \) is allocated to a coefficient value known as \( \beta \) indicating the weight. Once it is investigated using LR model, the dataset offered that every value has a weight. The diverse weights indicate distinct correlation among \( X \) and \( Y \). The parameters in LR can be tuning to achieve better classifier outcome. In this study, a set of two DL models namely Adam and RMSProp is applied for tuning the parameters present in LR which are discussed in the following sections.

2.4 Adam model

The Adam algorithm is explained here. Let \( f(\theta) \) indicates a noisy objective function: a stochastic scalar function which can be differentiable with respect to the parameter \( \theta \). It is interested to minimize the expected values of this function, \( \mathbb{E}[f(\theta)] \) with respect to \( \theta \). Under \( f_1(\theta), \ldots, f_T(\theta) \), it is denoted the realization of the stochastic function at subsequent time steps \( t = 1, \ldots, T \). The stochasticity comes from validation at arbitrary sub-samples (mini batches from (minibatches) of datapoints, or occur from inherent function noise. With \( g_t = \nabla_\theta f_t(\theta) \), a gradient is denoted, i.e. vector of partial derivatives of \( f_t \), based on \( \theta \) validated at timestep \( t \). This technique will update an exponential moving average of the gradient \( (\bar{m}_t) \) and squared gradient \( (\bar{v}_t) \) where the hyper parameters \( \beta_1, \beta_2 \in [0,1] \) manages an exponential decay rate of the motion average. The movement of average itself are estimation of the first moment and second moment of the gradient. At the same time, the movement of average undergo initialization as \( \bar{m}_t \) results to an estimation of the moment which is biased to zero, particularly in the initial time steps, and particularly when the decay rate is low. An essential characteristic of Adam’s update rule is its cautious alternative of step size. It is assumed that \( \beta_1 = 0 \), the efficient step considered in parameter space at time step \( t \) is \( \Delta = \frac{\tilde{m}_t}{\sqrt{\tilde{v}_t}} \). It holds a set of 2 upper limits: \( |\Delta_1| \leq a \cdot (1 - \beta_1)/\sqrt{1 - \beta_2} \) in the case \( 1 - \beta_1 > \sqrt{1 - \beta_2} \), \( \|\Delta_1\| \leq \alpha \) otherwise. The initial case takes place in many sparsity cases, if a gradient has been zero at every time step except that at present time step. For low sparse case, the effectual step size is low. When \( 1 - \beta_1 = \sqrt{1 - \beta_2} \), it is defined that \( \|\bar{m}_t\| < 1 \). so \( \|\Delta_1\| < \alpha \).

The efficient magnitude of the steps occupied in parameter space at every time step is roughly limited by the step size setting \( a, i.e., |\Delta_1| \leq a \). Adam makes use of initialization bias correction terms.

2.5 RMSProp model

RMSProp model is based on the exponentially weighted average of the gradients such as gradient descent with momentum with the variation in the updating parameters. Assuming an example where it is tried to optimize a cost function which holds contours such as below and the red dot indicates the location of the local optimum. The start gradient descent starts from point ‘A’ and after an iteration of the gradient descent is ended at point ‘B’, the other side of the ellipse as shown in Fig. 3. Next, other step of gradient descent ended with the point ‘C’. Under every round of gradient descent, it is moved towards the local optimum with up as well as down oscillations. When high learning rate is utilized, then the vertical oscillation holds high magnitude, then, the vertical oscillation reduces the gradient descent and eliminates it from using high learning rate. The bias has the responsibility for the vertical oscillations whereas ‘weight’ represents the motion in the horizontal direction. Once the updates the bias, it reduces the vertical oscillation and when ‘weights’ are updated with maximum values. In the backward propagation, \( dW \) and \( db \) parameters are used for updating \( W \) and \( b \) as indicated below.

\[
W = W - \text{learning rate} \times dW
\]
In RMSprop, instead of utilizing dW and db not dependent on one another for every epoch, exponentially weighted averages of the square of dW as well as db is applied.

\[ S_{dW} = \beta * S_{dW} + (1 - \beta) * dW^2 \]  
(8)

\[ S_{db} = \beta * S_{db} + (1 - \beta) * db^2 \]  
(9)

Where \( \beta \) is another hyperparameter which takes the value of 0 and 1. It considers the weight among the average of preceding values and the square of present values for calculating the new weighted average. Once the exponentially weighted averages are calculated, it will update the variables.

\[ W = W - learning\ rate * \frac{dW}{\sqrt{S_{dW}}} \]  
(10)

\[ b = b - learning\ rate * \frac{db}{\sqrt{S_{db}}} \]  
(11)

SdW is comparatively small and Sdb is comparatively high.

### 3. EXPERIMENTAL VALIDATION

With an intention of analyzing the outcome of the applied classifier models, a benchmark PIMA diabetes dataset is utilized [5]. It comprises a number of 768 instances under the set of eight features. The attributes involved are Preg, Plas, Pres, Skin, Insu, BMI, Pedi and Age. Fig. 4 shows the distribution of the features present in the dataset. All the 768 instances fall into two class labels namely positive and negative. A total of around 34.9% of instances comes under (+)ve class and the rest of the 65.10% comes under (-)ve class. This class distribution of the dataset is demonstrated in Fig. 5. These details are given in Table 1.

<table>
<thead>
<tr>
<th>Description</th>
<th>Pima Indian Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Instances</td>
<td>768</td>
</tr>
<tr>
<td>Number of Attributes</td>
<td>8</td>
</tr>
<tr>
<td>Number of Class</td>
<td>2</td>
</tr>
<tr>
<td>Percentage of Positive Samples</td>
<td>34.90%</td>
</tr>
<tr>
<td>Percentage of Negative Samples</td>
<td>65.10%</td>
</tr>
</tbody>
</table>
As explained in previous section, K-means clustering technique is applied in the earlier stage and the results attained are clearly illustrated in Fig. 6. The figure evidently stated the total number of instances, list of available features and number of clustered instances. Here, a possible of 4 clustered instances can takes place. It is provided that a set of 79 instances are correctly cluster under the class label '0' and 182 instances are accurately clustered under the class label '0'. However, a group of 86 instances under class label '1' is misclustered as 0. In line with, a group of 421 instances under class label '0' is misclustered as 1. The bar chart inside the figure depicts the number of correctly clustered instances as 79 and 182 under the class labels of 0 and 1 respectively.

A sample visualization of the clustered output is shown in Fig. 7. Here, a set of four kinds of clustered instances are shown with four different indicators. Fig. 8 provided an extensive accuracy analysis of the applied classifier models before and after clustering. As seen in Fig. 8a, it is noticed that the accuracy rate of Adam is gradually increased and reaches to a maximum of around 74%. Additionally, Fig. 8b shows that the accuracy rate is gradually increased and reaches to a maximum of around 75%. As seen in Fig. 8c, it is noticed that the accuracy rate of Adam after clustering is gradually increased and reaches to a maximum of around 98%. Additionally, Fig. 8d shows that the accuracy rate of RMSProp is gradually increased and reaches to a maximum of around 98%.
98.45%. These values pointed out that the clustering process considerably raised the accuracy values of the RMSProb model. Fig. 9 provided an extensive loss rate analysis of the applied classifier models before and after clustering. As seen in Fig. 9a, it is noticed that the loss rate of Adam is gradually decreased and reaches to a minimum of around 0.55%. Additionally, Fig. 9b shows that the loss rate is gradually increased and reaches to a maximum of around 0.52%. As seen in Fig. 9c, it is noticed that the loss rate of Adam after clustering is gradually increased and reaches to a maximum of around 0.05%. Additionally, Fig. 9d shows that the loss rate of RMSProp is gradually increased and reaches to a maximum of around 0.035%. These values pointed out that the clustering process considerably minimized the loss values of the RMSProb model.

Fig. 9. Loss Rate analysis: (a) Adam before clustering, (b) RMSProp before clustering, (c) Adam after clustering, (b) RMSProp after clustering

Fig. 10 shows the generated confusion matrix of the applied classifiers. As shown in Fig. 10(a), the Adam model before clustering properly classifies a set of 447 instances as positive and 128 instances as negative. In addition, the Fig. 10(b) represents that the RMSProp before clustering correctly classifies a set of 466 instances as positive and 108 instances as negative. As shown in Fig. 10(c), the Adam model after clustering properly classifies a set of 78 instances as positive and 180 instances as negative. In addition, the Fig. 10(d) represents that the RMSProp after clustering correctly classifies a set of 75 instances as positive and 180 instances as negative.

Fig. 10. Confusion Matrix(a) Adam before clustering, (b) RMSProp before clustering, (c) Adam after clustering, (b) RMSProp after clustering

Table 2 made a comparative study of the applied classifiers under distinct validation parameters. Figs. 11 and 12 shows the results attained by distinct classifiers under various measures.

Table 2 Performance Evaluation of Different Classifiers with Proposed Method on Diabetes Dataset

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Precision</th>
<th>Recall</th>
<th>Accuracy</th>
<th>F-score</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam+LR</td>
<td>89.40</td>
<td>76.15</td>
<td>74.86</td>
<td>82.24</td>
<td>40.19</td>
</tr>
<tr>
<td>RMSProp+LR</td>
<td>93.20</td>
<td>74.44</td>
<td>74.74</td>
<td>82.77</td>
<td>37.59</td>
</tr>
<tr>
<td>K-means+ Adam+LR</td>
<td>96.20</td>
<td>97.44</td>
<td>98.08</td>
<td>96.82</td>
<td>95.44</td>
</tr>
<tr>
<td>K-means+ RMSProp+LR</td>
<td>94.93</td>
<td>97.40</td>
<td>97.70</td>
<td>96.15</td>
<td>94.51</td>
</tr>
</tbody>
</table>

Fig. 11. Precision and recall analysis of diverse models

The table values indicated that the Adam-LR model shows worse classification with the precision of 89.40, recall of 76.15, accuracy of 74.86, F-score of 82.24 and kappa value of 40.19. Besides, table values indicated that the RMSProp-LR model shows poor classification with the precision of 93.20, recall of 74.44, accuracy of 74.74, F-score of 82.77 and kappa value of...
37.59. Next, RMSProp-LR model after clustering produces near optimal classifier results with the precision of 94.93, recall of 97.40, accuracy of 97.70, F-score of 96.15 and kappa value of 94.51. However, it is interesting that the Adam-LR model after clustering produces optimal classifier results with the precision of 96.20, recall of 97.44, accuracy of 98.08, F-score of 96.82 and kappa value of 95.44. In summary, the Adam-LR with k-means clustering is considered as the better classification model.

**Fig. 12. Classifier results analysis of diverse models**

For proper validation, a detailed analysis with the recently state of art methods has been made and the results are available in Table 3 and Fig. 13. As mentioned earlier, Adam-LR with K-means model shows maximum classification with the accuracy of 98.08. At the same time, the LR with K-means and FNC models also tries to show competitive results with the accuracy value of 95.42 and 94.50 respectively. Next, the HPM model show moderate classification with the accuracy of 92.36.

**Table 3 Comparison with Recent Methods with Proposed for Applied Dataset in terms of Accuracy**

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Method</td>
<td>98.08</td>
</tr>
<tr>
<td>K-means-LR</td>
<td>95.42</td>
</tr>
<tr>
<td>Fuzzy Neural Classifier (FNC)</td>
<td>94.50</td>
</tr>
<tr>
<td>HPM</td>
<td>92.38</td>
</tr>
<tr>
<td>AMMLP</td>
<td>89.93</td>
</tr>
<tr>
<td>J48 (pruned)</td>
<td>89.30</td>
</tr>
<tr>
<td>J48 (unpruned)</td>
<td>86.60</td>
</tr>
<tr>
<td>Hybrid Model</td>
<td>84.50</td>
</tr>
<tr>
<td>MLP</td>
<td>81.90</td>
</tr>
<tr>
<td>Logistic</td>
<td>78.20</td>
</tr>
<tr>
<td>J48</td>
<td>76.70</td>
</tr>
<tr>
<td>SGD</td>
<td>76.60</td>
</tr>
<tr>
<td>ELM</td>
<td>75.72</td>
</tr>
<tr>
<td>NaiveBay</td>
<td>74.90</td>
</tr>
</tbody>
</table>

**Fig. 13. Accuracy analysis with recently presented models**

Simultaneously, the AMMLP and J48 pruned models attained near identical accuracy values of 89.93 and 89.30 respectively. In line with, the J48 unpruned model shows poor classification with the minimum accuracy of 86.60. Similarly, all the other compared models such as LR, J48, SGD, ELM, NB, Bayesian Network, CART and KNN show ineffective classification over the presented models.

### 4 CONCLUSION

This paper has presented a new classification model for predicting T2DM. The parameters in LR can be tuning to achieve better classifier outcome. Using a sequence of pre-processing steps, the presented model comprises of two processes namely clustering and classification. Initially, K-means clustering technique is employed to cluster the dataset. Then, LR classifier is employed for data classification. In this study, a set of two DL models namely Adam and RMSProp is applied for tuning the parameters present in LR which are discussed in the following sections. With an intention of analyzing the outcome of the applied classifier models, a benchmark PIMA diabetes dataset is utilized. The experimental outcome stated that the Adam-LR with K-means model shows maximum classification with the accuracy of 98.08.

### References


Ms. K Sowjanya, MobDBTest: A machine learning based system for predicting diabetes risk using mobile devices. 2015 IEEE International Advance Computing Conference (IACC)