Journal name here manuscript No.
(will be inserted by the editor)

An AI-based Intelligent System for Healthcare Analysis Using Ridge–Adaline Stochastic Gradient Descent Classifier

N Deepa[1] · **Prabadevi B**[1] · **Praveen Kumar Reddy M**[1] · **G Thippa Reddy***[1] · **Thar Baker***[2] · **Ajmal Khan M**[3] · **Usman Tariq**[4]

Received: date / Accepted: date

Abstract Recent technological advancements in information and communication technologies introduced smart ways of handling various aspects of life. Smart devices and applications are now an integral part of our daily life; however, the use of smart devices also introduced various physical and psychological health issues in modern societies. One of the most common health care issues prevalent among almost all age groups is diabetes mellitus. This work aims to propose an Artificial Intelligence (AI) – based intelligent system for earlier prediction of the disease using Ridge Adaline Stochastic Gradient Descent Classifier (RASGD). The proposed scheme RASGD improves the regularization of the classification model by using weight decay methods, namely Least Absolute Shrinkage and Selection Operator(LASSO) and Ridge Regression methods. To minimize the cost function of the classifier, the RASGD adopts an unconstrained optimization model. Further, to increase the convergence speed of the classifier, the Adaline Stochastic Gradient Descent classifier is integrated with Ridge Regression. Finally, to validate the effectiveness of the intelligent system, the results of the proposed scheme have been compared with state-of-art machine learning algorithms such as Support Vector Machine and Logistic Regression methods. The RASGD intelligent system attains an accuracy of 92%, which is better than the other selected classifiers.

[1] Vellore Institute of Technology, Vellore , Tamil Nadu, India.
[2] Liverpool John Moores University, Liverpool, Merseyside, United Kingdom .
[3] Department of Electrical and Computer Engineering, COMSATS University Islamabad-Attock Campus, Attock 43600, Pakistan
[4] College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Khraj, 11942, Saudi Arabia.
E–mail:deepa.rajesh@vit.ac.in, prabadevi.b@vit.ac.in, praveenkumarreddy@vit.ac.in, drajmal@cuiatk.edu.pk, u.tariq@psau.edu.sa
* Corresponding Author: G. Thippa Reddy, Thar Baker,
E–mail: thippareddy.g@vit.ac.in, t.baker@ljmu.ac.uk*

Keywords Artificial Intelligence · Health care · Machine Learning · Regression · Adaline Stochastic Gradient Descent

1 Introduction

Table 1: List of Acronyms

Sl No.	Acronym	Definition
1	AI	Artificial Intelligence
2	RASGD	Ridge Adaline Stochastic Gradient Descent Classifier
3	LASSO	Least Absolute Shrinkage and Selection Operator
4	WHO	World Health Organization
5	SGD	Stochastic Gradient Descent
6	SVM	Support Vector Machine
7	NHANE	National Health and Nutrition Examination Survey
8	PCA	Principal Component Analysis
9	MNIST	Modified National Institute of Standards and Technology
10	ADALINE	Adaptive Linear Neuron
11	TP	True Positive
12	TN	True Negative
13	FP	False Positive
14	FN	False Negative

The main source of energy in living beings is the blood sugar (or) blood glucose, that is generated from the intake of food. When the blood sugar is too high in one's body, it leads to a disease known as diabetes. Pancreas generates a hormone known as "insulin," which helps the glucose generated from food consumption passed on to the cells that lead to energy. At times, enough amount of insulin is not generated in our bodies, or some of the insulin doesn't get used, which leads to the accumulation of glucose in the cells. This situation will lead to severe health problems. High sugar levels can lead to severe and fatal diseases like stroke, heart diseases, kidney failures, nerve damage, etc. Different categories of diabetes are type-1 diabetes, type-2 diabetes, and gestational diabetes.

Type-1 diabetes is a condition in which enough insulin is not generated in the body. In type-2 diabetes, the body does not make enough insulin or use enough insulin for the generation of energy. Gestational diabetes is mostly observed in pregnant women, which may, in turn, lead to type-2 diabetes. Tye-2 diabetes is the most common of these categories [11].

Diabetes is one of the most common diseases found in the world population as per the World Health Organization (WHO); there were 422 million adults that were affected by diabetes. The changing food habits, lack of exercise to the body lead the rise of the number of adults being affected by diabetes to 8.5 % in 2014 from 4.7% of the adult population in 1980. Also, it is estimated that 1.1 million adolescents and children are affected by diabetes. Approximately 4

million peoples' death is caused by diabetes every year. The effect of diabetes doesn't stop with the individuals being affected by it, and it passes on to the next generation too as a hereditary disease. According to WHO in 2017, almost 850 billion US Dollars has been spent for the treatment of diabetes in the entire world[23].

From the above discussion, it is clear that early detection and treatment of diabetes help people save a lot of money and also prevent humans from being affected by other severe diseases. In recent years many researchers have used machine learning algorithms[6,8,10,9] to understand the patterns based on the risk factors of diabetes and hence classify the diabetes diseases [12].

The dataset used in this work is originally from the National Institute of Diabetes and Digestive and Kidney Diseases [17]. The objective of the paper is to develop an intelligent system for health care analysis using machine learning classifiers which could handle the data overfitting, select relevant features even in the presence of highly correlated parameters, minimize the cost function of the classifier applied and to get better convergence speed of the classifier. In this work, the Stochastic Gradient Descent (SGD) classifier [18] is used for classifying the diabetes dataset. Adaline is integrated with SGD for predicting continuous values in order to learn the coefficients of the model and also to attain a fast convergence rate. To handle overfitting, ridge regression and LASSO regression methods are integrated with Adaline SGD.

The main contributions of this paper are:

1. To handle the overfitting of data, Ridge regression is integrated with Adaline stochastic gradient descent classifier.
2. The feature subset is obtained in the presence of highly correlated variables and missing values by using the Ridge regression method.
3. Unconstrained optimization step is included to minimize the cost function in gradient descent classifier.
4. Instead of using predicted class labels, the RASGD employs "Adaline" with stochastic gradient descent for predicting continuous values to learn the coefficients of the model. Also, it is more powerful in determining the precision of predicted values.
5. The robustness of RASGD is obtained by integrating with Stochastic Gradient Descent (SGD) with Adaline to attain better convergence speed.

Rest of the paper is organized as follows: Section 2 summarizes the recent state of the artworks on diabetes disease classification. The proposed model is presented in Section 3. Experimentation results are discussed in Section 4, followed by conclusions and future work in Section 5.

2 Literature Survey

Recently, a large amount of interest has been observed in the research community to apply various Machine Learning and Artificial Intelligence methods for predicting diabetes at the earliest. In [21], the authors utilized the Pima Indian

diabetic dataset from the UCI repository to conduct an experimental analysis for accurate diabetic prediction. For this process, the authors extracted the best features and classification models using six feature selection techniques and ten different types of classifiers. These training parameters have been used to predict the accuracy rate. Furthermore, the proposed approach can be extended by taking more number of instances and apply deep learning approaches.

In [36], the authors proposed a framework for the detection of type 2 diabetes using machine learning and feature engineering approaches. The authors analyzed data from the electronic health records of 123,241 patients. The experiment was conducted on 300 randomly selected patients. It involves several steps, such as choosing the best features to eliminate sparse and noisy data. Out of 36 features, the best eight features are selected during this process. In the several popular Machine Learning algorithms have been used for classification. Finally, the trained attributes are used to estimate the performance of the system. The future scope of the work is to improve the system by taking into account the multilingual datasets in cross-domain.

Similarly, in [5], the authors proposed a framework for the detection of type 2 diabetes by combining toenails and machine learning approaches. The proposed technique works in different steps; in the first step, feature reduction is performed to recognize the related attributes that will reduce the number of features and remove unnecessary or noisy information. In the second step, ensembling techniques are used for classification. The authors have performed experiments to evaluate their proposed technique by using the publicly available UCI datasets. The results have shown that the proposed model outperformed the existing approaches. However, to achieve more robust results, there is a need to perform further experimentation with extended datasets.

In [20], the authors adopted the Gaussian process for the classification of diabetes dataset and then carried out a comparative analysis with other existing classifiers, such as Quadratic Discriminant Analysis, Naive Bayes, Linear Discriminant Analysis. The authors used the Pima diabetic dataset from the UCI repository to conduct an experimental analysis for accurate diabetic prediction. This dataset consists of 768 patients, of which 268 are diabetic and 500 are controlled patients. The experimental results suggest that the authors have achieved better accuracy compared to other classifiers. Furthermore, the proposed approach can be extended by taking more number of instances and apply deep learning approaches and the future work can be expanded further by using regression models to predict the seriousness of the disease.

In [14], the authors have proposed a technique for predicting diabetes called the Cuckoo Search Optimized Reduction and Fuzzy Logic Classifier. The proposed technique works in two steps; in the first step, feature reduction is performed using a cuckoo search algorithm to recognize the related attributes that will reduce the number of features and remove unnecessary or noisy information. In the second step, the fuzzy rules are produced and then these rules are applied to classify the diabetic disease. The authors have performed experiments to evaluate their proposed technique by using the dataset collected

from an India hospital. The results have shown that the proposed prediction algorithm outperformed the existing approaches by attaining a better accuracy rate. This work can be expanded further by considering the complexity of space and time.

In [19], the authors proposed linear Support Vector Machine (SVM) and bagged trees to predict diabetic disease. The authors conducted an experimental analysis of the National Health and Nutrition Examination Survey (NHANE) dataset. In this work, firstly, some statistical approaches have been applied to treat imbalanced and missing data. Later, Principal Component Analysis (PCA) is used for feature extraction. The experiment was conducted using MATLAB with 140 data samples. The proposed system outperformed other methods considered. Besides, the system can be extended by taking larger data samples, and regression models can be applied.

Similarly, in [31], the authors have adopted several classification algorithms for predicting the person with diabetes at an early stage. During this process, the authors have used three different classification algorithms like Support Vector Machine, Decision Tree, and Naive Bayes. The efficiency of all these three algorithms is assessed using different measures such as accuracy, precision, Recall, F-measure. The results show that Naive Bayes achieves 76.30% accuracy and outperforms other algorithms with the highest accuracy rate. This work can be further extended by taking advanced classifiers to achieve a high accuracy rate.

The Stochastic Gradient Descent(SGD) method is a gradient descent approach used for larger datasets to minimize the objective function. In [3], the authors investigated the performance of different classifiers to predict the disease at an earlier stage. The authors used two different datasets during this process. For both datasets, the SGD classifier has outperformed other classifiers in terms of accuracy and achieved an absolute mean error of 0.1916.

In [30], the authors used Guided Stochastic Gradient Descent Algorithm to reduce inconsistency and error rates in large datasets. The authors have adopted a guided search approach for improving the effectiveness of the SGD. Experimental results indicate that the guided search approach outperforms conventional approaches with a 3% improvement in accuracy rate. This work can be expanded by adding several real-time disease prediction algorithms. This may significantly improve performance as compared to classical machine learning techniques.

In [34], the authors proposed a new training model to improve SGD's performance in deep neural networking. During this process, the authors used inconsistent training that dynamically adjusts the rate of loss error and lowers mean batch error. The experimentation was conducted on the Modified National Institute of Standards and Technology database (MNIST) real-time datasets and improved performance in terms of time complexity and reduced loss rate. This work can be further expanded by using variance reduction techniques to increase the convergence rate.

In [32], the authors integrated the Barzilai–Borwein model into Stochastic Gradient Descent to increase system performance in terms of convergence

rate, execution time, sensitivity. Experimental results were compared with the conventional SVM model. The proposed approach produces a higher convergence rate, decreases execution time and lowers the sensitivity rate. In this paper, a Stochastic Gradient descent classifier is used for the development of the AI-based Intelligent system.

The regression techniques, also called regularization models such as Ridge and Lasso, are popularly used for feature selection in the machine learning domain. An approach is proposed by integrating Ridge and Lasso methods with the Bayesian method [1]. A model was developed to forecast the price of rice in Thailand using Lasso and Ridge regression methods. In this model, Ridge and Lasso regression methods are applied to diminish the values of variables [4]. Lasso and Robust regression techniques were used to select the efficiency of a collector in the Solar dryer. The results showed that the Lasso provided better accuracy compared with a robust regression technique [16]. A model was developed for the selection of genome and prediction of its breeding value in animal and plant breeding using six regression techniques, namely Adaptive Lasso, Lasso, Ridge BLUP, Ridge, adaptive elastic net, and elastic net [22]. A model was developed for the prediction of failures in corporate sectors using regression methods such as Ridge regression and Logistic Lasso. The results showed that Ridge performed well compared to the Logistic Lasso method [24].

From the above survey, it is evident that the existing SGD classifiers could not handle the correlation between the data, missing values, and overfitting of data. Also, the existing works on SGD classifiers suffered from lower convergence rates. The proposed RASGD overcome these drawbacks using weight decay methods, namely Ridge and LASSO, and integrating Adaline with SGD classifier, which performs classification of diabetes dataset with higher accuracy.

3 Proposed System

The steps involved in the proposed work are:

1. Load the diabetes dataset.
2. Pre-processing of the dataset.
   (a) Data transformation is performed by min-max standardization method.
   (b) Feature selection and regularization are carried out by weight decay methods.
3. Split the dataset into training and testing data.
4. Train the Adaline Stochastic Gradient Descent classifier using the training data.
5. Evaluate the performance of the proposed model by using the testing dataset with several metrics.
6. The proposed Ridge-Adaline stochastic gradient descent classifier model is compared with Adaline stochastic gradient descent classifier, Lasso-adaline

stochastic gradient descent classifier, Support vector machine algorithm and Logistic regression method.

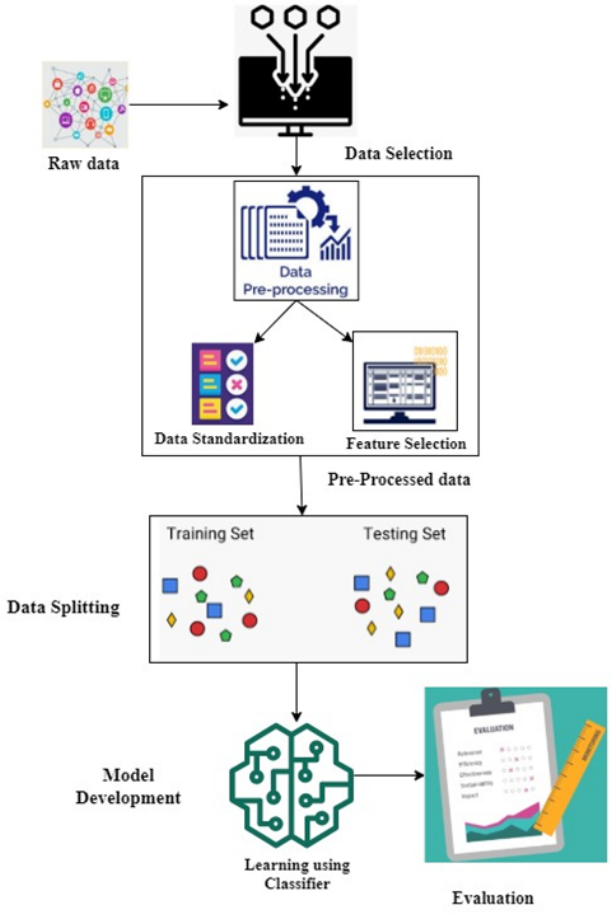

Fig. 1: Proposed Framework for AI-based Intelligent System

The proposed framework is depicted in Figure 1. The RASGD took the raw data and pre-processed for generalization, which involves two processes, namely Data standardization using the min-max algorithm followed by feature selection using Weight decay methods viz., LASSO and Ridge Regression algorithms. The pre-processed data is divided into a training set and test set, which are used to train the model and to evaluate the results, respectively. Adaline Stochastic gradient descent classifier is used for the development of an AI-based intelligent system. The trained model is tested with test data using several metrics such as precision, F1-Score, and accuracy.

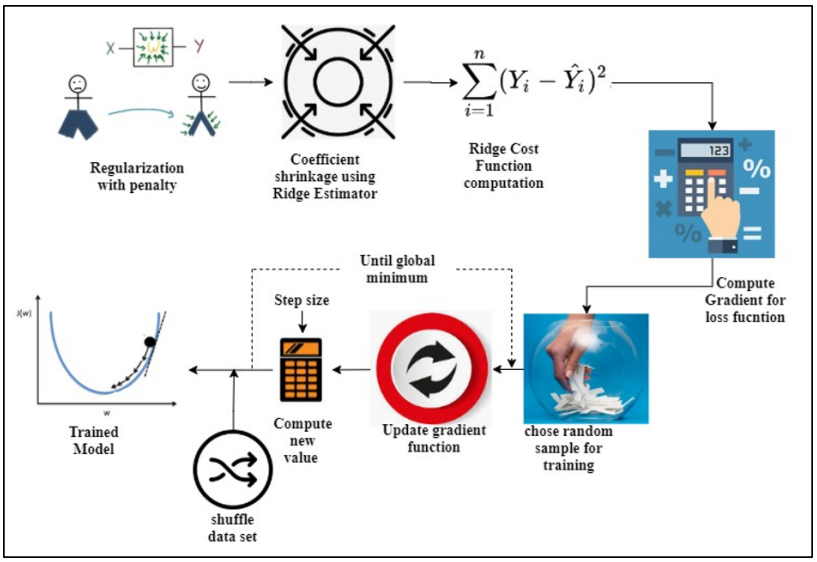

Fig. 2: Framework for Ridge Adaline Stochastic Gradient Descent Classifier

The working model of RASGD is illustrated in Figure 2. The proposed RASGD model is an intelligent system developed for health care analysis which performs the following tasks such as handling data overfitting, selection of relevant features, handling high correlation among the features, reduction of the cost function of the classifier and obtain better convergence speed. To treat noisy data and missing values, feature selection is made before classification. Ridge regression is used to identify the feature subset relevant to the scenario even when the features are highly correlated. Ridge regression method identifies the feature subset by regularising the coefficients with penalty value, uses ridge estimator to shrink the coefficients, and compute the cost function. This regression belongs to L2 type of regularization in which it adds a penalty called L2 penalty. L2 penalty is calculated as the square of the magnitude of the coefficients. The cost function in Ridge regression method is updated by simply summing the penalty values.

Adaline stochastic gradient descent classifier is used for classification. It computes the gradient (i.e. the slope) for the loss function. Gradient descent means moving down the slope to reach the lowest point on the curve. This is done iteratively until a minimum point is reached. Stochastic gradient descent is a discriminative learning method that models from the observed data and classifies them. Stochastic gradient descent classifier, when optimized with hyperparameters, classifies the instances in lesser time than linear regression. The loss function is the error between the predicted value and actual value. Adaline stochastic gradient descent classifier considers only one sample for each iteration and it helps in fast convergence. Each sample is chosen randomly and after each iteration, the dataset is shuffled for more randomness. The new

value is computed based on the old value and the step size. The learning rate determines the step size. This process is repeated until a global minimum is reached. The trained model obtained is evaluated with the test dataset.

3.1 Data Visualization using Scatterplot matrix:

The diabetes dataset collected from Kaggle repository is used in this paper for the development of AI-based Intelligent system. It contains eight attributes and records of around 800 patients. To visualize the data in the dataset, a scatterplot matrix is used. This plot is a matrix comprising a group of scatterplots that shows the relationship between every pair of variables. From the scatterplot matrix, it is easier to infer the most substantial relation and the weakest one. This, in turn, helps to interpret the correlation between each pair of variables. The scatterplot matrix of the diabetes dataset is shown in Figure.3.

Methodologies used:

3.2 Feature Selection methods:

Feature selection is one of the challenging tasks in the field of statistics. Many research works have been carried out in this area in order to standardize and optimize this task to fit into any kind of data. Feature selection helps in selecting the relevant variables suitable for the given problem and also it reduces the number of variables by removing the variables which contain redundant information. Based on the selection of estimation metrics, the feature selection algorithms are divided into three main methods, namely filter methods, wrapper methods and embedded methods [27, 25, 15, 2, 7].

Filter methods work independently and eliminate the irrelevant features before applying the appropriate classification algorithm [26]. These methods consider the intrinsic properties of the dataset to calculate the evaluation metric. The relevant features are selected using the calculated values. The features with low calculated value are removed and the features with high values against the evaluation metric are retained in these kinds of feature selection methods [28]. The main disadvantage of the filter methods is their unidimensional nature. The evaluation metrics are computed for each and every variable individually, but the associations between the variables are neglected. Due to this nature of filter methods, the classification accuracy of the developed model is less. The advantage of using these filter methods is that these methods are applied to the classification model in the initial stage. Therefore the methods give a rapid performance and they are extensible.

In Wrapper methods, the feature selection is determined by the prediction accuracy of the developed model. It is iterative in nature and feature subsets are selected by these wrapper methods in each iteration. Their performance

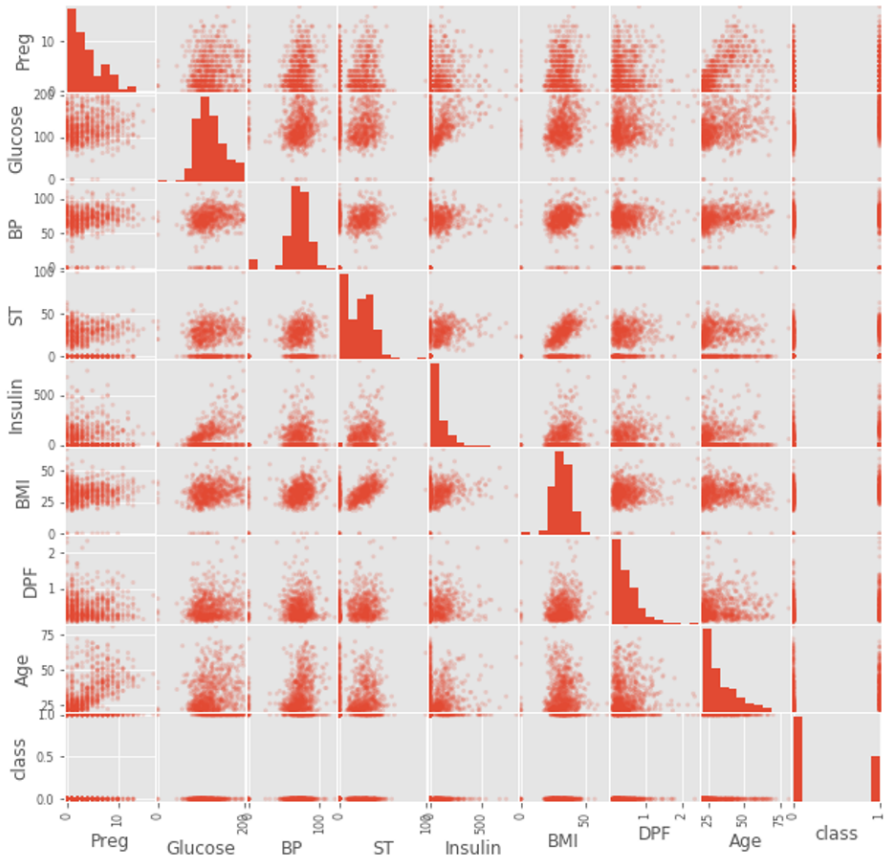

Fig. 3: Scatterplot Matrix of Attributes Related to Diabetes Dataset

is assessed by the algorithm applied for classification purposes. When the dataset consists of high dimensional data, the number of feature subset selection process iterations is increased. Therefore the feature selection algorithm is wrapped in the selected classification algorithm to find the optimal variable subset. As the wrapper methods are included in the classification algorithm, it is proved that the selected features produce excellent prediction accuracy. Therefore the selection of appropriate wrapper feature selection methods depends upon the selection of suitable classification algorithms. These wrapper methods are multidimensional in nature and they consider the associations between the features for subset selection. The disadvantages of these methods are they are prone to overfitting and need huge computation costs depending upon the chosen classifier.

In the embedded type of feature selection method, a threshold score is calculated by the model developed using the classification algorithm. The developed model provides the evaluation score for each feature. From the scores,

features with a high score will be retained and features with a low score will be removed. The measures for evaluating the performance of these selected features are computed by the developed model. In this kind of feature selection method, the model is developed once for the computation of feature scores. Therefore the computation cost of these embedded methods is less when compared to other methods such as wrapper methods, which require high computation cost because of its iterative nature [29]. Also, the embedded type of feature selection method needs a threshold value to be calculated for the selection of features [33].

3.3 LASSO Regression method:

LASSO and RIDGE are two embedded type of weight decay methods used for feature selection. LASSO technique is a robust method which performs regularization task and feature selection. Regularization is the process of reducing errors and avoiding overfitting. In the LASSO method, the sum of the absolute scores of the parameters is calculated and a constraint is applied on those scores such that the calculated sum should be less than an upper bound value. Therefore the method applies a regularization step in which it updates the coefficient values of the regression variables by reducing few to zero. In the feature selection step, the features which have a non-zero coefficient value in the regularization step are included in the feature subset for the development of the classification model. LASSO method helps to get good prediction accuracy and reduce overfitting. The disadvantages of LASSO methods are they do not give measures of uncertainty and they are not consistent in some applications. It is very difficult to determine feature selection when the variables are extremely correlated in nature. The cost function of Lasso regression is as follows [13]:

$$\sum_{i=1}^{m} (y_i - \check{y_i})^2 = \sum_{i=1}^{m} \left(y_i - \sum_{j=0}^{n} w_j * x_{ij} \right)^2 + \lambda \sum_{j=0}^{n} |w_j| \tag{1}$$

Where m is the number of instances, n is the number of features in the dataset, x denotes the predictor variable, y denotes response variable, λ is the penalty value(regularization parameter), and w represents coefficients. The regularization parameter can be controlled to update the coefficient values.

3.4 Ridge Regression Method:

Ridge Regression method also performs regularization and feature selection tasks. This regression method determines the feature subset, which is relevant to the classification problem even when the variables are highly correlated. It helps to deal with missing values in the variables as well. The Ridge regression method applies a special type of estimator for shrinkage of coefficients, namely

ridge estimator. This regression belongs to L2 type of regularization in which it adds a penalty called L2 penalty. L2 penalty is calculated as the square of the magnitude of the coefficients. The cost function in Ridge regression method is updated by simply summing the penalty values. The cost function used in Ridge regression is as follows:

$$\sum_{i=1}^{m} (y_i - \check{y_i})^2 = \sum_{i=1}^{m} \left(y_i - \sum_{j=0}^{n} w_j * x_{ij} \right)^2 + \lambda \sum_{j=0}^{n} w_j^2 \tag{2}$$

Where m is the number of instances, n is the number of features in the dataset, and λ is the penalty value. The Ridge regression applies constraints on w coefficients. The penalty term is used for the regularization of the coefficients. This type of regression is used to shrink the coefficients and which leads to the reduction of complexity of the model. As the Ridge regression helps to obtain feature subset even when there are correlations between the variables and can deal with missing values of the variables, this regression is used in this work.

3.5 Adaline Stochastic Gradient Descent Classifier:

Gradient descent is an efficient approach and provides the best results for sparse data. Here the gradient refers to the slope. In literal terms, gradient descent means moving down the slope (descending down) to reach the lowest point on the curve. This is done iteratively until a minimum point is reached. The three variants of gradient descent include batch, stochastic and mini-batch gradient descent approaches. These variants are derived based on the number of samples taken for each iteration. As per the semantics of the term stochastic, the stochastic gradient descent works based on random probability. Instead of training the whole data, it selects a random set of data (a few samples) from the given dataset for each iteration. This few sets of samples chosen for an iteration is called a batch. The conventional gradient descent considers the whole dataset as a batch to avoid noise for accurate classification. Nevertheless, it will be difficult when the data set is enormous. Thus evolved the Stochastic gradient descent classifier, which solves this problem by considering the subset of the sample in random for each iteration. Stochastic gradient descent is a discriminative learning model that models from the observed data and classifies them. Stochastic gradient descent considers the batch size as one i.e., one sample in an iteration. So in each iteration, the gradient of the cost function for a single sample is computed. Also, stochastic gradient descent is faster than conventional gradient descent as the updates are done immediately after training each sample. Stochastic gradient descent classifier, when optimized with hyperparameters, classifies the instances in lesser time than linear regression. ADALINE is ADAptive Linear Neuron, which is a binary classifier similar to Percepton [35]. Widrow-rule of Adaline updates the weights

using the linear activation function. The Adaline stochastic gradient descent classifier Algorithm is as follows:

1. Take the derivatives of the loss function for each feature. The loss function is given as $J(\theta) = (\check{y} - y)^2(x)$ where $\check{y}$ is the predicted value and y is the actual value with respect to x.
2. Compute the gradient ∇ of the loss function in step 1.
3. Choose a random initial value of the features to start θ_0
4. Update the gradient function by inducing the feature values
5. Compute the step size as $step\_size = Gradient * Learning\_Rate$
6. Compute the new feature value as: New_value=old_value-step_size the new_value is updated in opposite direction of the gradient.

$$\theta_1 = \theta_0 - (step\_size * \nabla j(\theta))$$

7. Shuffle the data points after each iteration Repeat steps 3 to 5 until the gradient is closer to zero

The data points are selected randomly at each step. The Learning_Rate has a more significant influence on the algorithm. It is better if Learning_Rate is higher at the beginning as it makes the algorithm to take huge steps and decrease while it reaches the minimum value, to avoid missing the minimum point. Learning_Rate is initialized to 0.01.

4 Results and Discussions

To remove the irrelevant, missing, and redundant values from the dataset, the RSAGD preprocess the data using the min-max standardization technique. After preprocessing the data, the feature selection methods are used to obtain the reduced feature subset to describe the decision class and eliminates redundant and irrelevant variables. The feature selection step helps to handle missing data and considers interactions among the variables. It helps to make the classification algorithms run faster even when high dimension dataset are involved. And also, these feature selection algorithms are used to reduce overfitting.

Adaline Stochastic Gradient Descent classifier is applied to the dataset initially for classification purposes without performing the feature selection step. The decision boundary region diagram of the Adaline Stochastic Gradient Descent classifier is shown in Figure 4. The classification results are shown in Table 2, and the results are not satisfactory. Then Lasso and Ridge Regression algorithms are applied for the feature selection process, and the Adaline Stochastic Gradient Descent algorithm is used for the classification of diabetes dataset. Figure 5 and Figure 6 show the decision boundary region diagrams of the Adaline Stochastic Gradient Descent algorithm with Lasso regression and Ridge regression. It is clear from the results that the Adaline Stochastic Gradient Descent classifier does not perform well without Lasso and Ridge Regression algorithms. Further, for validation, the diabetes dataset is applied

to other classifiers such as Support Vector Machine algorithm and Logistic Regression method, and the decision boundary region diagrams of these classifiers are shown in Figure 7 and Figure 8. From these figures, it is proved that the Ridge Adaline Stochastic Gradient Descent algorithm performs better than the Adaline Stochastic Gradient Descent classifier, Lasso Adaline Stochastic Gradient Descent algorithm, Support Vector Machine and Logistic Regression method.

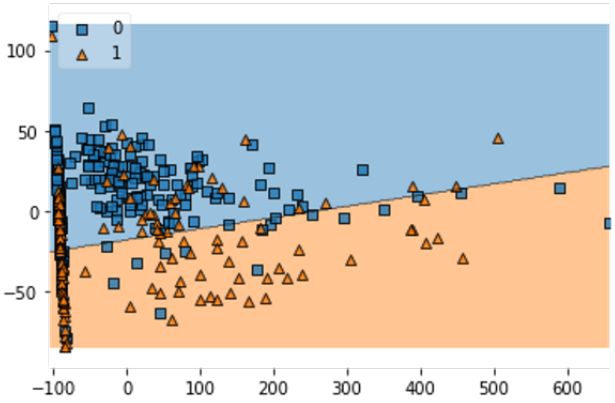

Fig. 4: Decision Region Boundary for Adaline Stochastic Gradient Descent Classifier

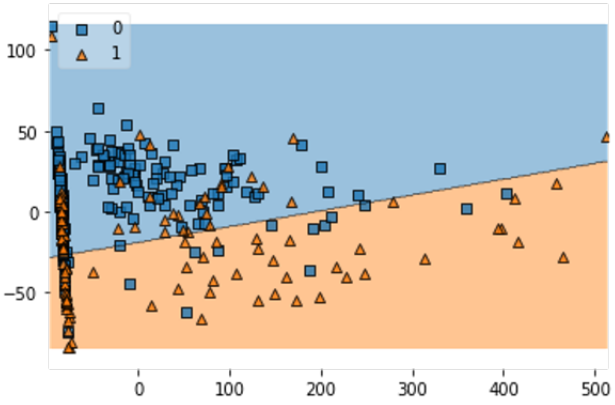

Fig. 5: Decision Region Boundary for Lasso Adaline Stochastic Gradient Descent Classifier

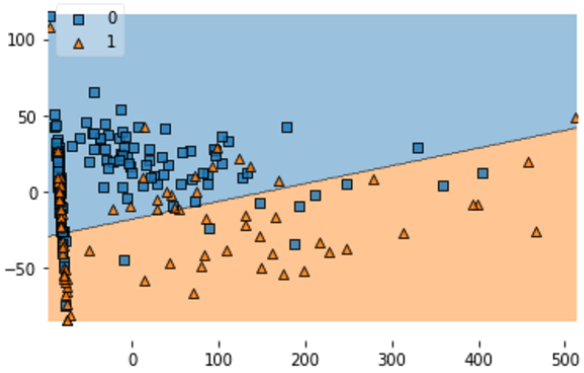

Fig. 6: Decision Region Boundary for Ridge Adaline Stochastic Gradient Descent classifier

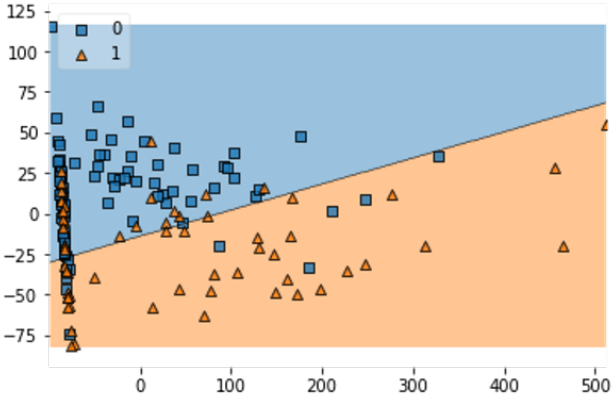

Fig. 7: Decision Region Boundary for Support Vector Machine Classifier

The classification results of the Adaline Stochastic Gradient Descent classifier, Lasso Adaline Stochastic Gradient Descent, Ridge-Adaline Stochastic Gradient Descent, Support Vector Machine, and Logistic regression classifiers are shown in Table 2. The basic metrics used for the evaluation of classification and prediction models are Precision, Recall, F1-score. In order to calculate these metrics, True Positive (TP), True Negative(TN), False Positive(FP) and False Negative(FN) parameters should be identified from the classification results.

True positive is the number of instances that are correctly predicted as positive values. If the patient is diagnosed with diabetes, the value of the class label is 1, which is positive. True Negative is the instances that are correctly predicted as negative values. If the patient is not diagnosed with diabetes, the value of class label is 0, which is negative. False Positive is the instance which has to be predicted with negative value but actually predicted with a positive

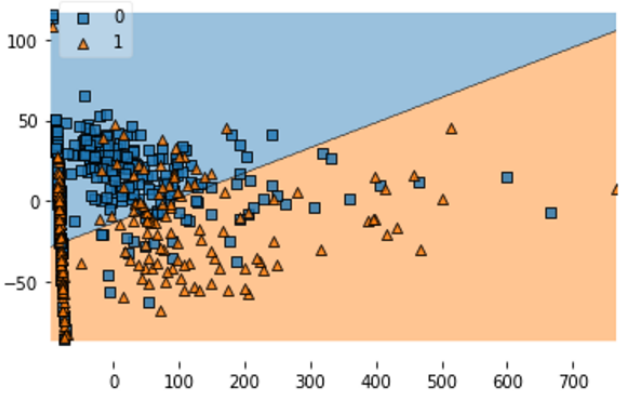

Fig. 8: Decision Region Boundary for Logistic Regression Classifier

value. Here the patients who are not suffering from diabetes are wrongly predicted as suffering from diabetes. False Negative is the instance which has to be predicted with negative value but predicted with a positive value. Here the patients who are not suffering from diabetes are wrongly predicted as suffering from diabetes.

Precision is defined as the number of instances with true positive values divided by the number of instances with true positive plus false positive values. Recall or Sensitivity is defined as the number of instances correctly predicted as positive, divided by the total number of positive instances. F- Score is defined as the harmonic mean of the two measures precision and recall. The values of these metrics range from 0 to 1. The value 1 is said to be a good score and the value 0 is the worst score for any classification problem. The formulae for different evaluation metrics are given below:

$$Precision = TP/(TP + FP) \tag{3}$$

$$Recall = TP/(TP + FN) \tag{4}$$

$$F1\ Score = 2 * (Recall * Precision)/(Recall + Precision) \tag{5}$$

In order to validate the results obtained from the AI-based intelligent system, the results from other classifiers such as Support Vector Machine algorithm and Logistic Regression method are considered and shown in Table 2

From Table 2, it has been proved that the Ridge-Adaline Stochastic Gradient Descent method outperforms Lasso-Adaline Stochastic Gradient Descent method, Support Vector Machine and Logistic Regression methods. The Ridge Adaline Stochastic method has obtained an accuracy of 92%, Precision of 91% for Class 0 and 85% for Class 1, Recall value of 91% for Class 0 and 81% of Class 1, F1-Score of 87% for Class 0 and 82% for Class 1. The overall accuracy scores of Adaline Stochastic Gradient Descent is 75%, Lasso-Adaline Stochastic Gradient Descent is 81%, Ridge-Adaline Stochastic Gradient Descent is

Table 2: Classification results of different classifiers

Adaline Stochastic Gradient Descent	**Precision**	**Recall**	**F1-score**
Class 0	0.78	0.88	0.82
Class 1	0.67	0.50	0.57
Macro average	0.72	0.69	0.70
Weighted average	0.74	0.75	0.74
Accuracy	0.75		
Lasso-Adaline Stochastic Gradient Descent	**Precision**	**Recall**	**F1-score**
Class 0	0.80	0.89	0.84
Class 1	0.72	0.56	0.63
Macro average	0.76	0.72	0.74
Weighted average	0.77	0.78	0.77
Accuracy	0.81		
Ridge-Adaline Stochastic Gradient Descent	**Precision**	**Recall**	**F1-score**
Class 0	0.91	0.91	0.87
Class 1	0.85	0.81	0.82
Macro average	0.88	0.86	0.84
Weighted average	0.89	0.88	0.85
Accuracy	0.92		
Support Vector Machine	**Precision**	**Recall**	**F1-score**
Class 0	0.81	0.84	0.82
Class 1	0.66	0.60	0.63
Macro average	0.73	0.72	0.73
Weighted average	0.76	0.76	0.76
Accuracy	0.76		
Logistic Regression	**Precision**	**Recall**	**F1-score**
Class 0	0.73	0.96	0.83
Class 1	0.67	0.17	0.27
Macro average	0.70	0.57	0.55
Weighted average	0.71	0.72	0.66
Accuracy	0.72		

92%, Support Vector Machine is 76% and Logistic Regression is 72%. From the evaluation metrics, it has been validated that when the Ridge Regression method is integrated with a Stochastic Gradient classifier can produce better results for classification problems.

5 Conclusion and Future Work

In this paper, an AI-based Intelligent system is proposed for health care analysis. Pima Diabetes dataset has been applied to the developed system for analyzing the performance of the model. The weight decay methods have been applied to the dataset for feature selection and regularization namely LASSO and Ridge regression. Adaline Stochastic Gradient Descent classifier has been used for the development of the classification model. Ridge regression per-

forms well with Adaline stochastic Gradient Descent classifier than LASSO. Further, the classification results of the developed AI-based Intelligent system are compared with the results obtained from other classifiers such as Support Vector Machine and Logistic Regression methods. Among these methods, the Ridge-Adaline Stochastic Gradient Descent classifier outperforms the other methods with a classification accuracy of 92%. The developed model can be applied to any healthcare dataset for the prediction of disease. In the future, the scalability and robustness of the proposed model can be tested with a high dimensional dataset.

References

1. Bedoui, A., Lazar, N.A.: Bayesian empirical likelihood for ridge and lasso regressions. Computational Statistics & Data Analysis **145**, 106917 (2020)
2. Bhattacharya, S., Kaluri, R., Singh, S., Alazab, M., Tariq, U., et al.: A novel pca-firefly based xgboost classification model for intrusion detection in networks using gpu. Electronics **9**(2), 219 (2020)
3. Bilge, L., Dumitraş, T.: Before we knew it: an empirical study of zero-day attacks in the real world. In: Proceedings of the 2012 ACM conference on Computer and communications security, pp. 833–844 (2012)
4. Boonyakunakorn, P., Nunti, C., Yamaka, W.: Forecasting of thailand's rice exports price: Based on ridge and lasso regression. In: Proceedings of the 2nd International Conference on Big Data Technologies, pp. 354–357 (2019)
5. Carter, J.A., Long, C.S., Smith, B.P., Smith, T.L., Donati, G.L.: Combining elemental analysis of toenails and machine learning techniques as a non-invasive diagnostic tool for the robust classification of type-2 diabetes. Expert Systems with Applications **115**, 245–255 (2019)
6. Deepa, N., Ganesan, K.: Aqua site classification using neural network models. AGRIS on-line Papers in Economics and Informatics **8**(665-2016-45133), 51–58 (2016)
7. Deepa, N., Ganesan, K.: Mahalanobis taguchi system based criteria selection tool for agriculture crops. Sādhanā **41**(12), 1407–1414 (2016)
8. Deepa, N., Ganesan, K.: Multi-class classification using hybrid soft decision model for agriculture crop selection. Neural Computing and Applications **30**(4), 1025–1038 (2018)
9. Deepa, N., Ganesan, K.: Decision-making tool for crop selection for agriculture development. Neural Computing and Applications **31**(4), 1215–1225 (2019)
10. Deepa, N., Ganesan, K.: Hybrid rough fuzzy soft classifier based multi-class classification model for agriculture crop selection. Soft Computing **23**(21), 10793–10809 (2019)
11. for Disease Control, C., Prevention, et al.: National diabetes statistics report, 2017. Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services **20** (2017)
12. Fitriyani, N.L., Syafrudin, M., Alfian, G., Rhee, J.: Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. IEEE Access **7**, 144777–144789 (2019)
13. Fonti, V., Belitser, E.: Feature selection using lasso. VU Amsterdam Research Paper in Business Analytics pp. 1–25 (2017)
14. Gadekallu, T.R., Khare, N.: Cuckoo search optimized reduction and fuzzy logic classifier for heart disease and diabetes prediction. International Journal of Fuzzy System Applications (IJFSA) **6**(2), 25–42 (2017)
15. Gadekallu, T.R., Khare, N., Bhattacharya, S., Singh, S., Reddy Maddikunta, P.K., Ra, I.H., Alazab, M.: Early detection of diabetic retinopathy using pca-firefly based deep learning model. Electronics **9**(2), 274 (2020)
16. Javaid, A., Ismail, M., Ali, M.K.M., et al.: Efficient model selection of collector efficiency in solar dryer using hybrid of lasso and robust regression. Pertanika Journal of Science & Technology **28**(1) (2020)

17. Kaggle: Diabetes detection (2019(Accessed on March, 2019)). https://www.kaggle.com/uciml/pima-indians-diabetes-database
18. Kalimeris, D., Kaplun, G., Nakkiran, P., Edelman, B., Yang, T., Barak, B., Zhang, H.: Sgd on neural networks learns functions of increasing complexity. In: Advances in Neural Information Processing Systems, pp. 3491–3501 (2019)
19. Kaur, G., Chhabra, A.: Improved j48 classification algorithm for the prediction of diabetes. International Journal of Computer Applications **98**(22) (2014)
20. Maniruzzaman, M., Kumar, N., Abedin, M.M., Islam, M.S., Suri, H.S., El-Baz, A.S., Suri, J.S.: Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm. Computer methods and programs in biomedicine **152**, 23–34 (2017)
21. Maniruzzaman, M., Rahman, M.J., Al-MehediHasan, M., Suri, H.S., Abedin, M.M., El-Baz, A., Suri, J.S.: Accurate diabetes risk stratification using machine learning: role of missing value and outliers. Journal of medical systems **42**(5), 92 (2018)
22. Ogutu, J.O., Schulz-Streeck, T., Piepho, H.P.: Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. In: BMC proceedings, vol. 6, p. S10. Springer (2012)
23. Organization, W.H., et al.: Classification of diabetes mellitus (2019)
24. Pereira, J.M., Basto, M., da Silva, A.F.: The logistic lasso and ridge regression in predicting corporate failure. Procedia Economics and Finance **39**, 634–641 (2016)
25. Reddy, G.T., Khare, N.: Heart disease classification system using optimised fuzzy rule based algorithm. International Journal of Biomedical Engineering and Technology **27**(3), 183–202 (2018)
26. Reddy, G.T., Reddy, M.P.K., Lakshmanna, K., Kaluri, R., Rajput, D.S., Srivastava, G., Baker, T.: Analysis of dimensionality reduction techniques on big data. IEEE Access **8**, 54776–54788 (2020)
27. Reddy, G.T., Reddy, M.P.K., Lakshmanna, K., Rajput, D.S., Kaluri, R., Srivastava, G.: Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis. Evolutionary Intelligence pp. 1–12 (2019)
28. Reddy, T., RM, S.P., Parimala, M., Chowdhary, C.L., Hakak, S., Khan, W.Z., et al.: A deep neural networks based model for uninterrupted marine environment monitoring. Computer Communications (2020)
29. RM, S.P., Bhattacharya, S., Maddikunta, P.K.R., Somayaji, S.R.K., Lakshmanna, K., Kaluri, R., Hussien, A., Gadekallu, T.R.: Load balancing of energy cloud using wind driven and firefly algorithms in internet of everything. Journal of Parallel and Distributed Computing (2020)
30. Sharma, A.: Guided stochastic gradient descent algorithm for inconsistent datasets. Applied Soft Computing **73**, 1068–1080 (2018)
31. Sisodia, D., Sisodia, D.S.: Prediction of diabetes using classification algorithms. Procedia computer science **132**, 1578–1585 (2018)
32. Sopyła, K., Drozda, P.: Stochastic gradient descent with barzilai–borwein update step for svm. Information Sciences **316**, 218–233 (2015)
33. Suppers, A., Gool, A.J.v., Wessels, H.J.: Integrated chemometrics and statistics to drive successful proteomics biomarker discovery. Proteomes **6**(2), 20 (2018)
34. Wang, L., Yang, Y., Min, R., Chakradhar, S.: Accelerating deep neural network training with inconsistent stochastic gradient descent. Neural Networks **93**, 219–229 (2017)
35. Widrow, B., et al.: Adaptive" adaline" Neuron Using Chemical" memistors.". (1960)
36. Zheng, T., Xie, W., Xu, L., He, X., Zhang, Y., You, M., Yang, G., Chen, Y.: A machine learning-based framework to identify type 2 diabetes through electronic health records. International journal of medical informatics **97**, 120–127 (2017)