

Article

# Deep Neural Network for Predicting Diabetic Retinopathy from Risk Factors

Ganjar Alfian <sup>1</sup>, Muhammad Syafrudin <sup>2,\*</sup>, Norma Latif Fitriyani <sup>2</sup>, Muhammad Anshari <sup>3</sup>, Pavel Stasa <sup>4</sup>, Jiri Svub <sup>4</sup> and Jongtae Rhee <sup>2</sup>

<sup>1</sup> Industrial Artificial Intelligence (AI) Research Center, Nano Information Technology Academy, Dongguk University, Seoul 04626, Korea; ganjar@dongguk.edu

<sup>2</sup> Department of Industrial and Systems Engineering, Dongguk University, Seoul 04620, Korea; norma@dongguk.edu (N.L.F.); jtrhee@dongguk.edu (J.R.)

<sup>3</sup> School of Business & Economics, Universiti Brunei Darussalam, Gadong BE1410, Brunei; anshari.ali@ubd.edu.bn

<sup>4</sup> Department of Economics and Control Systems, Faculty of Mining and Geology, VSB–Technical University of Ostrava, 70800 Ostrava-Poruba, Czech Republic; pavel.stasa@vsb.cz (P.S.); jiri.svub@vsb.cz (J.S.)

\* Correspondence: udin@dongguk.edu

Received: 21 August 2020; Accepted: 17 September 2020; Published: 19 September 2020



**Abstract:** Extracting information from individual risk factors provides an effective way to identify diabetes risk and associated complications, such as retinopathy, at an early stage. Deep learning and machine learning algorithms are being utilized to extract information from individual risk factors to improve early-stage diagnosis. This study proposes a deep neural network (DNN) combined with recursive feature elimination (RFE) to provide early prediction of diabetic retinopathy (DR) based on individual risk factors. The proposed model uses RFE to remove irrelevant features and DNN to classify the diseases. A publicly available dataset was utilized to predict DR during initial stages, for the proposed and several current best-practice models. The proposed model achieved 82.033% prediction accuracy, which was a significantly better performance than the current models. Thus, important risk factors for retinopathy can be successfully extracted using RFE. In addition, to evaluate the proposed prediction model robustness and generalization, we compared it with other machine learning models and datasets (nephropathy and hypertension–diabetes). The proposed prediction model will help improve early-stage retinopathy diagnosis based on individual risk factors.

**Keywords:** retinopathy; risk factor; machine learning; deep neural network; recursive feature elimination; deep learning

## 1. Introduction

Diabetes is a chronic disease associated with abnormal blood glucose (BG) levels. Patients with type 1 diabetes (T1D) cannot control their BG naturally due to lacking insulin secretion, while for type 2 diabetes (T2D), the body cannot utilize its produced insulin [1,2]. Thus, T1D patients must administer insulin via injection or an insulin pump to achieve a near-normal glucose metabolism [3]. For T2D patients, a healthy diet, physical exercise, and drug administration are suggested to control BG levels and prevent many complications. Diabetes patients commonly develop acute complications, such as hypoglycemia (BG < 70 mg/dL) and hyperglycemia (BG > 180 mg/dL) if they fail to carefully self-manage BG levels [4]. Excessive BG (hyperglycemia) can result in long-term complications, e.g., retinopathy [5,6], nephropathy or kidney disease [7,8], and cardiovascular disease [9,10]. Diabetic retinopathy (DR) is the most common ocular complication from diabetes and the leading cause of visual impairment among

patients [11]. Thus, it is important to accurately predict related complications to help prevent their progression [12], and individual significant risk factors can be utilized as input features to improve prediction model performance.

Considerable research effort has recently focused on developing of retinopathy prediction models using machine learning based on individual risk factors, aiming for high accuracy and good generalization. Pre-diagnosis models have been used for different populations and have shown significant performance predicting diabetic retinopathy [13–17]. Deep neural networks (DNNs) are machine learning systems that incorporate many layers to learn more complex patterns, achieving exceptional prediction accuracy for many applications. In particular, DNNs have shown excellent performance by improving classification accuracy compared with conventional models [18–23]. DNNs have also been trained to successfully predict diabetic retinopathy based on retinal fundus images [24–29]. Furthermore, previous studies have established positive impacts from support vector machine–recursive feature elimination (SVM-RFE) as the feature selection algorithm on improving classification accuracy [30–37], especially for DNNs [38,39].

However, none of these previous studies included DNNs with RFE for retinopathy prediction based on individual risk factors. Therefore, the current study integrated DNN and SVM-RFE for retinopathy to improve prediction accuracy. We employed SVM-RFE to extract significant risk factors and DNN to subsequently generate higher accuracy prediction models. Utilizing the identified significant retinopathy predictors (risk factors) provides optimized prediction models. Thus, this study will help diabetic patients to reduce diabetic retinopathy risk, the major cause of blindness in such patients.

The remainder of this paper is organized as follows. Section 2 discusses related works for retinopathy prediction, including related DNN and RFE applications for healthcare. Section 3 details the proposed retinopathy prediction model, and Section 4 discusses experimental results for the proposed model. Section 5 summarizes and concludes the paper, including discussing study limitations and future research directions.

## 2. Literature Review

This section discusses previously proposed machine learning models for retinopathy pre-diagnosis, with particular attention regarding DNN and RFE for health-related datasets.

### 2.1. Diabetic Retinopathy Prediction

Prediction models based on machine learning and employing individuals' risk factors as input features helped improve retinopathy pre-diagnosis for Iran, Korea, US, and Taiwan populations. Hosseini et al. [13] used a logistic regression (LR) model to predict retinopathy for 3734 T2D patients from the Isfahan Endocrinology and Metabolism Research Center, Iran. The model achieved AUC = 0.704, sensitivity = 60%, and specificity = 69%. Oh et al. [14] proposed retinopathy risk prediction based on Lasso. They used a dataset from the Korea National Health and Nutrition Examination Surveys (KNHANES) V-1, with 327 patients randomly selected as training data and 163 as validation. The proposed model achieved AUC = 0.81, accuracy = 73.6%, sensitivity = 77.4%, and specificity = 72.7%.

Retinopathy prediction has also been applied for US patients. Ogunyemi and Kermah [15] investigated machine learning models for predicting retinopathy from six health centers in South Los Angeles, comprising 513 T2D patients. The dataset was split into approximately 80% for training and the remainder for testing. They compared RUSBoost ensemble and AdaBoost model predictions and showed that RUSBoost ensemble was superior to AdaBoost, achieving accuracy = 73.5%, AUC = 0.72, sensitivity = 69.2%, and specificity = 55.9%. LR, support vector machine (SVM), and artificial neural network (ANN) models also helped improve retinopathy prediction performance for the overall US population [16] using an updated dataset comprising 27,116 T1D and T2D patients from the Los Angeles County Department of Health Services. Combining ANN and the synthetic minority oversampling technique (SMOTE) achieved superior results with AUC = 0.754, sensitivity = 58%, and specificity = 80%. Tsao et al. [17] proposed SVM-, decision tree (DT)-, ANN-, and LR-based models

to predict DR based on several risk factors for T2D patients. The dataset was gathered from a private hospital in Northern Taiwan, incorporating 430 normal and 106 DR patients. The SVM based model was superior to the other algorithms considered, achieving accuracy = 79.5% and AUC = 0.839.

## 2.2. Deep Neural Network

Deep neural network models have been used in many previous studies to improve prediction accuracy compared with other models. For diabetes prediction-related issues, DNNs have only been applied to predict T2D. Kim et al. [22] used phenotype and genotype data from the Nurses' Health Study (NHS) with data from the Health Professionals Follow-up Study (HPFS) to evaluate the prediction model performance. The proposed DNN outperformed LR, with AUC = 0.931 and 0.928 for male and female patients. Stacked autoencoders in DNN were also applied for T2D diabetes classification [23]. Their model was applied to a Pima Indian dataset and achieved classification accuracy = 86.26%.

Diabetic retinopathy is a complication of diabetes that causes damage to the blood vessels in the retina and leading to vision impairment. Therefore, an accurate prediction model for pre-diagnosis retinopathy would be very useful to improve patient health outcomes. DNN have shown good performance diagnosing retinopathy from retinal fundus images, including datasets from Otago and Messidor [24], and three clinical departments in Sichuan Provincial People's Hospital [25]. Parmar et al. [26] employed a convolutional neural network to detect DR from retinal images and their model outperformed others considered. Furthermore, the ResNet architecture model was utilized to detect DR from fundus images achieving an excellent classification accuracy [27,28]. Finally, Gadekallu et al. [29] employed a DNN with grey wolf optimization (GWO) and principle components analysis (PCA) to optimize the parameters and reduce dimensionality, respectively, to predict DR based on extracted features from retinal imaging. However, these studies only diagnosed retinopathy from retinal fundus images, and to our best knowledge, no previous study considered DNNs for retinopathy based on risk factors.

## 2.3. Recursive Feature Elimination

Many previous studies employed feature selection to improve the model prediction accuracy. Guyon et al. [30] introduced RFE to select the most significant gene(s) for cancer classification and, hence, improve classification model accuracy. The algorithm calculates a rank score and eliminates the lowest-ranking features. Previous studies showed significant performance improvements by employing RFE, including predicting mental states (brain activity) [31,32], Parkinson [33], skin disease [34], autism [35], Alzheimer [36], and T2D [37]. They showed that SVM-RFE achieved superior performance than several comparison methods. In addition, previous studies demonstrated DNN accuracy improvement by integrating RFE as the feature selection algorithm [38,39]. The experimental results showed that integration of SVM-RFE to DNN algorithms achieved best prediction accuracy as compared to other methods.

To our best knowledge, only Kumar et al. [37] considered RFE for diabetes prediction. Kumar et al. used SVM-RFE to identify the most discriminatory gene target for T2D. These identified significant genes could then be focused on as potential drug targets. However, although SVM-RFE was employed to extract significant features for T2D, this was not applied to the DR dataset. Similarly, previous DR studies used DNNs to classify disease from retinal fundus images only, and risk factors were not utilized as DNN input features. Therefore, the current study proposed SVM-RFE and DNN to improve DR prediction accuracy from individual risk factors. To our best knowledge, this is the first time SVM-RFE and DNN using individual risk factors were employed to improve DR prediction accuracy.

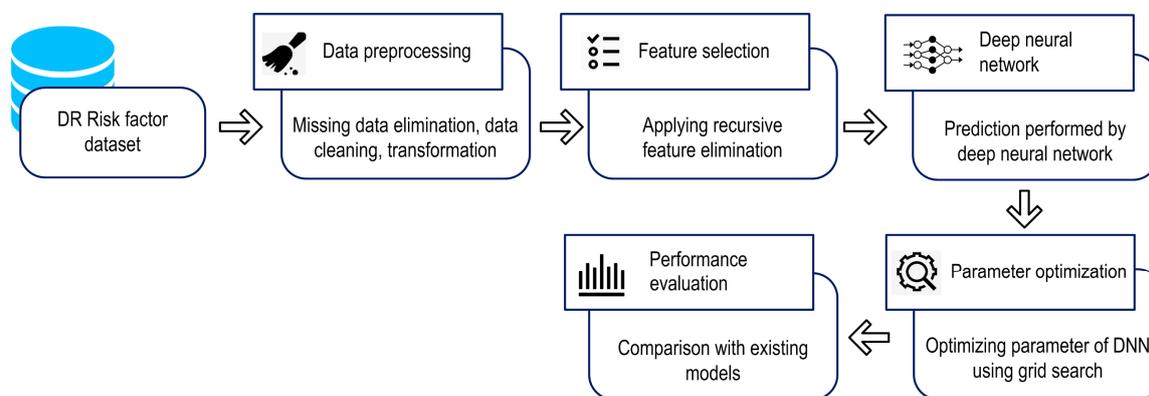
### 3. Methodology

#### 3.1. Datasets

Previous studies established that T1D or T2D patients tend to develop complications, such as retinopathy, nephropathy, cardiovascular disease (CVD), etc. Therefore, we proposed a DNN-based model to predict whether T1D or T2D patients will later develop DR. The dataset was collected by Khodadadi et al. [40] and related with diabetes complications in Lur and Lak populations of Iran. Informed consent was obtained from patients, and the dataset was made publicly available by previous authors (<https://data.mendeley.com/datasets/k62fdsnwkg/1>). The dataset was gathered from 133 diabetic patients covering known risk factors for neuropathy, nephropathy, diabetic retinopathy (DR), peripheral vessel disease (PVD), CVD, food ulcer history, and dawn effect. Originally, the dataset consisted of 24 information gathered from diabetic patients (T1D and T2D). We removed irrelevant features, leaving 14 potentially DR-relevant risk factors, as shown in Table 1. The class label (retinopathy) was assigned when the subject had symptomatic cases with a history of laser or surgical therapy. The objective of our study was to classify whether a diabetic patient will develop diabetic retinopathy (DR) in the future.

#### 3.2. Design of Proposed Model

Figure 1 shows the proposed DNN model to predict DR diagnosis from several risk factors, based on a public DR dataset. Data pre-processing removed inappropriate and inconsistent data. During the pre-processing stage, data normalization was applied by rescaling real valued numeric attributes into  $[0, 1]$ . Missing values in the numeric and nominal attributes were replaced by mean and mode, respectively. Then RFE removed irrelevant features, and a DNN-based prediction was developed using the grid search algorithm to optimize the model hyperparameter and, hence, maximize DNN performance. Performance was evaluated by comparing the proposed and other best-practice machine learning models from previous studies.



**Figure 1.** Proposed deep neural network model for diabetic retinopathy (DR) prediction.

We used stratified 10-fold cross-validation (CV), a variation of k-fold CV for the proposed and comparison machine learning models. In k-fold CV, the dataset is split into k subsets of equal size and the instances for each subset or fold are randomly selected. Each subset, in turn, is used for testing and the remainder for the training set. The model is evaluated k times such that each subset is used once as the test set. However, in stratified k-fold cross-validation, each subset is stratified so that they contain approximately the same proportion of class labels as the original dataset. By this procedure, the variance among the estimates are reduced and the average error estimate is reliable. Furthermore, our dataset is imbalanced, with 32% of subjects classified as DR patients. A previous study demonstrated that stratified k-fold CV is generally considered superior to regular CV, particularly

for unbalanced datasets [41]. Figure 2 shows the overview of the model validation based on CV and stratified CV applied to the two-class dataset.

**Table 1.** Diabetic retinopathy dataset.

No	Attribute	Description	Type	Range
1	BMI	Subject's body mass index	Numeric	18–41
2	DM duration	Subject's diabetes duration (y)	Numeric	0–30
3	A1c	Subject's average blood glucose level over the past 3 months (mg/dL)	Numeric	6.5–13.3
4	Age	Subject's age (y)	Numeric	16–79
5	FBS	Subject's fasting blood sugar level (mg/dL)	Numeric	80–510
6	LDL	Subject's low-density lipoprotein level (mg/dL)	Numeric	36–267
7	HDL	Subject's high-density lipoprotein level (mg/dL)	Numeric	20–62
8	TG	Subject's triglyceride level (mg/dL)	Numeric	74–756
9	Sys BP	Subject's systolic blood pressure (mmHg)	Numeric	105–180
10	Dias BP	Subject's diastolic blood pressure (mmHg)	Numeric	60–120
11	Sex	Subject's sex	Categorical	0 = Female 1 = Male
12	DM type	Subject's diabetes type (T1D or T2D)	Categorical	0 = T1D 1 = T2D
13	DM treat	Subject's diabetes treatment	Categorical	0 = Both (Insulin and oral agent) 1 = Insulin 2 = Oral agent
14	Statin	Subject's statin status (frequently used as part of diabetes care)	Categorical	0 = Ator (atorvastatin) 1 = No statin 2 = ROS (rosuvastatin)
15	Retinopathy (class)	Subject's retinopathy status	Categorical	0 = No (91) 1 = Yes (42)

### 3.3. Recursive Feature Elimination (RFE)

Feature selection removes redundant and irrelevant features to improve machine learning quality and efficiency. This study applied RFE with SVM kernels, i.e., linear function, to evaluate feature significances for DR dataset [30]. SVM-RFE works first by training the dataset with the SVM classifier. Next, the ranking weights for all features are computed. Finally, the feature with smallest weight is deleted. This process is repeated until no features are left, with later eliminated features having higher ranks. The bottom ranked ones are the least informative and removed in the first iteration. Thus, irrelevant features are gradually eliminated and important features retained for classification. The process is summarized in Algorithm 1.

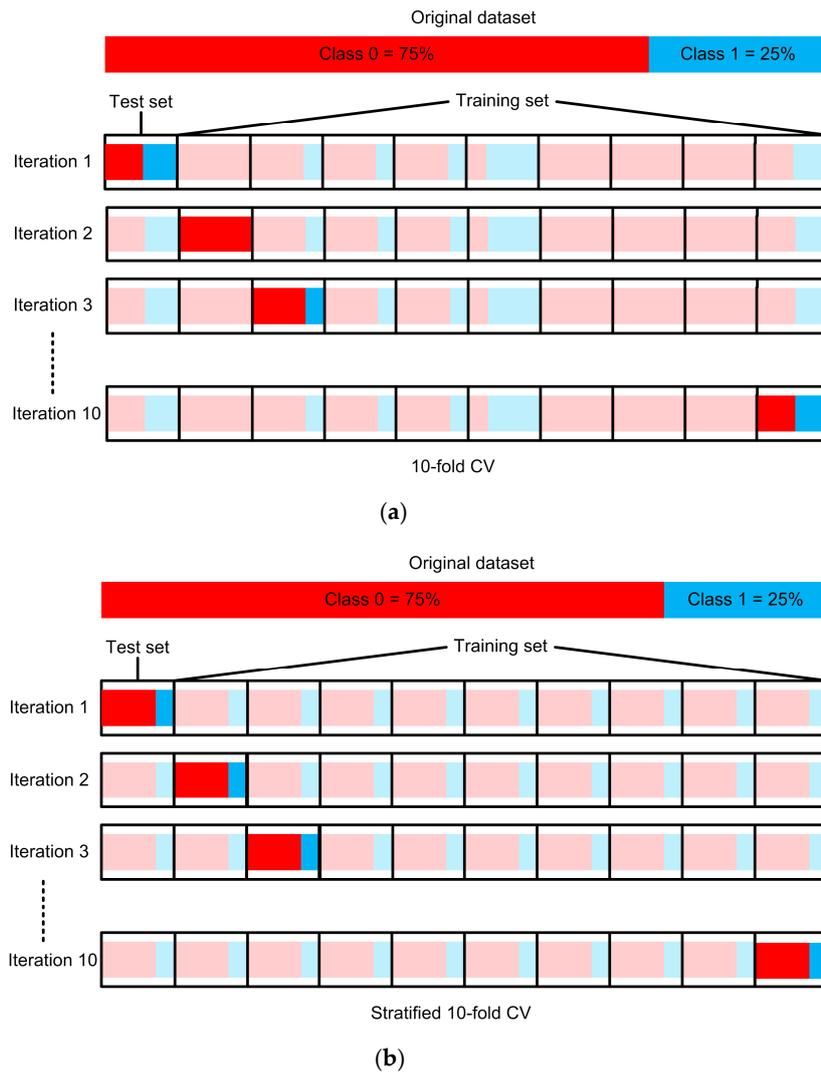


Figure 2. Model validation (a) 10-fold cross-validation (CV) and (b) stratified 10-fold CV.

---

**Algorithm 1.** SVM-RFE pseudocode

---

**Input:**  $X_0 = (x_1, x_2, \dots, x_m)^T$   
 $y = (y_1, y_2, \dots, y_m)^T$   
 $s = [1, 2, \dots, n]$   
 $r = []$

**Output:**  $r$

**while**  $s$  is not empty **do**

$X = X_0(:, s)$   
 $\alpha = SVM\_train(X, y)$   
 $w = \sum_k \alpha_k y_k x_k$   
 $c_i = (w_i)^2$   
 $f = argmin(c)$   
 $r = [s(f), r]$   
 $s = s(1 : f - 1, f + 1 : length(s))$

**end while**  
**return**  $r$

---

The SVM-RFE algorithm can be divided into three steps, which are input, calculation of the weight of each feature, and removing the lowest ranked feature. In Algorithm 1, during the input stage,  $X_0$  is defined as the training sample,  $y$  is class labels,  $s$  is subset of remaining features, and  $r$  is feature sorted list. In the next process, the weight calculation of each feature is conducted where the algorithm repeats the process until the list of  $s$  is empty. The new training sample  $X$  is defined according to the remaining features  $s$ . The set of paired inputs and outputs is used by training the classifier  $\alpha$ . The calculation of the weight vector  $w$  and ranking criteria  $c_i$  is conducted at this stage. When the lowest ranking feature  $f$  was obtained, the feature sorted list  $r$  can be updated. At the last stage, the feature with the smallest ranking criterion is removed and  $s$  is updated. The users can stop the iteration when the list of  $s$  is not empty, so that desired number of features can be obtained.

In Algorithm 1,  $SVM\_train$  (linear SVM) is utilized to learn from the set of paired inputs  $X$  and outputs  $y$ . The linear SVM classifies training data by mapping the original data onto a high dimensional feature space and finding the linear optimal hyperplane to separate instances of each class from the others [42]; for the case of separating training vectors belonging to two linearly separable classes:

$$(\mathbf{x}_i, y_i), \mathbf{x}_i \in R^n, y_i \in \{+1, -1\}, i = 1, \dots, n \tag{1}$$

where  $\mathbf{x}_i$  is a real valued  $n$ -dimensional input vector and  $y_i$  is the class label associated with the training vector. The separating hyperplane is determined by an orthogonal vector  $\mathbf{w}$  and bias  $b$ , which identify points that satisfy

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{2}$$

Thus, the classification mechanism for linear SVM can be expressed as

$$\max_{\alpha} \left[ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \right] \tag{3}$$

with constraints

$$\sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \dots, n \tag{4}$$

where  $\alpha$  is the parameter vector for the classifier hyperplane, and  $C$  is a penalty parameter to control the number of misclassifications.

Figure 3 shows the attribute ranking for the DR dataset. We investigated the impact of the top  $k$  features for DNN accuracy. The SVM-RFE was executed to remove irrelevant features, and important features ( $k$ ) are utilized as the input for the DNN. This process is repeated for all possible  $k$  features. Finally, we found that the first (top) 13 DR features is the optimal number of features that can maximize DNN accuracy. Hence only systolic blood pressure (Sys BP) was removed from the DR dataset. Section 4.2 discusses optimizing the number of features for the dataset.

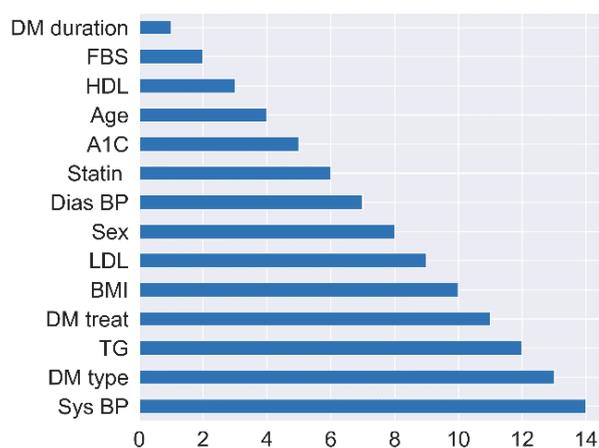


Figure 3. Attribute ranking for diabetes retinopathy (DR) datasets.

We also compared the performance of SVM-RFE with other feature selection methods on improving DNN accuracy, such as chi-squared, ANOVA, and extra trees. The result of feature selection execution for the dataset as shown in Table 2. The feature selection methods showed different results in terms of extracting the most important features. A higher value of score, f-value, and gini importance indicate the importance of the features. The impact of a different feature selection on DNN accuracy is presented in Section 4.2.

**Table 2.** Feature selection results for the diabetes retinopathy (DR) dataset.

No	Attribute	Feature Selection Model			
		RFE (Rank)	Chi-Squared (Score)	ANOVA (F-Value)	Extra Trees (Gini Importance)
1	BMI	10	0.225	2.352	0.056
2	DM duration	1	5.474	49.028	0.161
3	A1c	5	1.054	4.780	0.081
4	Age	4	1.352	24.473	0.098
5	FBS	2	0.970	12.349	0.088
6	LDL	9	0.520	6.207	0.059
7	HDL	3	0.571	8.726	0.077
8	TG	12	0.643	6.419	0.077
9	Sys BP	14	1.992	18.519	0.083
10	Dias BP	7	1.734	18.738	0.070
11	Sex	8	2.669	4.641	0.045
12	DM type	13	0.127	1.870	0.009
13	DM treat	11	0.779	2.889	0.061
14	Statin	6	0.064	0.176	0.031

The proposed SVM-RFE generated the top five features, i.e., DM duration, FBS, HDL, Age, and A1c as significant risk factors for retinopathy. Other feature selection methods, such as chi-squared, ANOVA, and extra trees, generated a different subset of important features. In chi-squared, the attributes DM duration, Sex, Sys BP, Dias BP, and Age were identified as the top five features. Furthermore, the attributes DM duration, Age, Dias BP, Sys BP, and FBS were identified as the top five features in ANOVA, while in extra trees, the result is similar to ANOVA except that A1C was included in the top five features instead of the attribute Dias BP. These top five features generated by the feature selection methods can be utilized as the input features for the deep neural network to improve classification accuracy.

Finally, the SVM-RFE identified the attribute Sys BP as an irrelevant feature, while the attribute Statin was recognized by chi-squared and ANOVA as a less important feature. The attribute DM type was discovered by extra trees as an unimportant feature; therefore, these irrelevant features generated by feature selection must be removed as input for the classifier. Excluding these irrelevant features is expected to improve DNN performance. A more detailed discussion regarding our significant risk factors for retinopathy and a comparison with the results from previous studies are presented in Section 4.3.

### 3.4. Proposed Deep Neural Network

We employed a DNN model to predict DR from the risk factors dataset. A DNN is an ANN class with multiple hidden layers between input and output layers and has recently become highly successful for classification. A DNN is fully connected; hence, each unit receives connections from all units in the previous layer. Thus, each unit has its own bias and a weight for every pair of units in two consecutive layers. The net input was calculated by multiplying each input with its corresponding

weight and then summing. Each unit in the hidden layer took the net input and applied an activation function. Thus, network computation with three hidden layers can be expressed as

$$h_i^{(1)} = \varphi^{(1)}\left(\sum_j w_{ij}^{(1)}x_j + b_i^{(1)}\right) \tag{5}$$

$$h_i^{(2)} = \varphi^{(2)}\left(\sum_j w_{ij}^{(2)}h_j^{(1)} + b_i^{(2)}\right) \tag{6}$$

$$h_i^{(3)} = \varphi^{(3)}\left(\sum_j w_{ij}^{(3)}h_j^{(2)} + b_i^{(3)}\right) \tag{7}$$

and

$$y_i = \varphi^{(4)}\left(\sum_j w_{ij}^{(4)}h_j^{(3)} + b_i^{(4)}\right) \tag{8}$$

where  $x_j$  is the input units,  $w$  is weight,  $b$  is bias,  $y$  is output units,  $h_i^{(l)}$  is units in the  $l$ th hidden layer, and  $\varphi$  is the activation function. In our study, we used and evaluated several activation functions, such as sigmoid, hyperbolic tangent, and rectified linear unit (ReLU), which are presented in detail in Equations (9)–(11), respectively.

$$\varphi(v) = \frac{1}{1 + e^{-v}} \tag{9}$$

$$\varphi(v) = \frac{e^{2v} - 1}{e^{2v} + 1} \tag{10}$$

$$\varphi(v) = \max\{v, 0\} \tag{11}$$

The DNNs were trained using the back propagation (BP) algorithm [43], which compares the prediction result with the target value and modifies each training tuple’s weights to minimize error between prediction and target values. In measuring a good prediction model to predict the expected outcome, a loss function is required. Our study focused on binary classification, where the number of classes is two. The cross-entropy loss function can be calculated as

$$Loss = -\sum_i (y'_i \log(y_i) + (1 - y'_i) \log(1 - y_i)) \tag{12}$$

where  $y'_i$  is true probability and  $y_i$  is predicted probability value. This process was iterated to produce optimal weights, providing optimal predictions for the test data. Figure 4 shows the proposed DNN model to predict DR from individuals’ risk factors.

We utilized the grid search algorithm [44] to automatically select the best parameters for the proposed deep neural network (DNN) model. We then applied a grid search for the training set and measured cross-validation to obtain the best prediction model, as shown in Figure 5. The objective for this method was to select the best hyperparameter for the proposed DNN model to achieve the highest accuracy for DR. We found a five-hidden-layers network with different neurons each, ReLU as activation function, and SGD as weight optimization were the best parameters for the DNN. Table 3 shows optimized hyperparameter details for the proposed DNN for DR. Finally, we applied these parameters to the proposed DNN model.

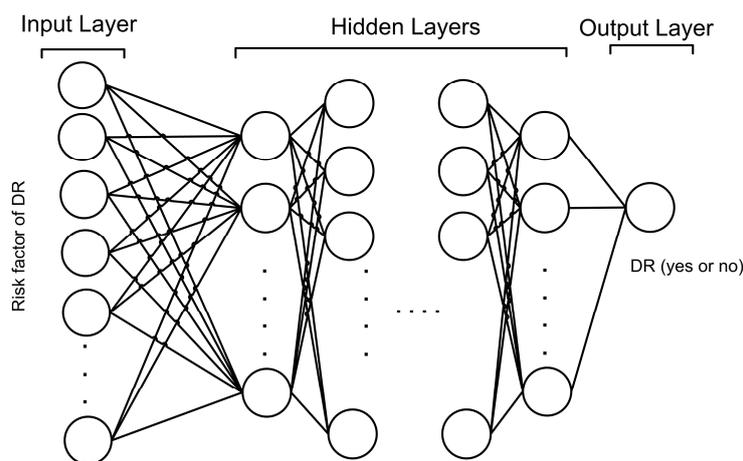


Figure 4. Proposed deep neural network (DNN) for diabetes retinopathy (DR) prediction.

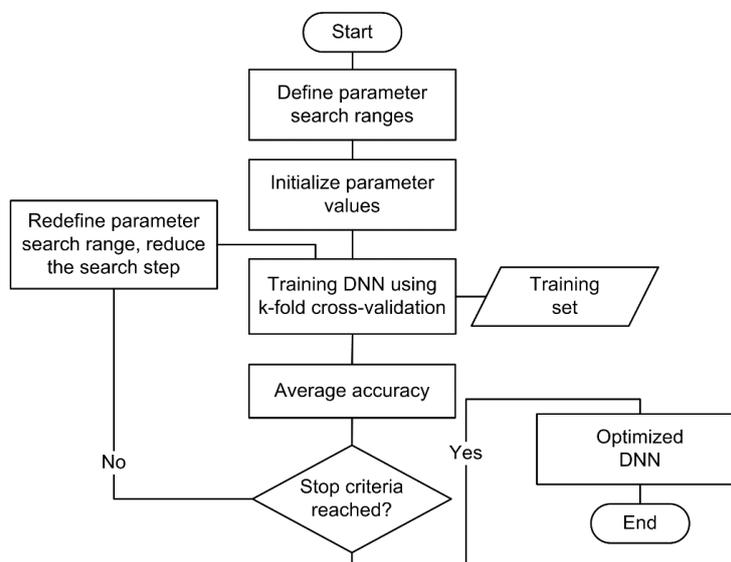


Figure 5. Grid search algorithm for deep neural network (DNN) hyper-parameter optimization.

Table 3. Optimized hyperparameter using grid search.

Hyperparameter	Optimized Value
Hidden layer size	100, 64, 128, 64, 32
Activation function	ReLU
Alpha	0.0001
Initial learning rate	0.01
Maximum iteration	500
Optimization algorithm	SGD

Notes: ReLU = rectified linear unit, SGD = stochastic gradient descent.

### 3.5. Experimental Setup

Machine-learning models were applied to distinguish DR from the public dataset. Classification and feature selection methods were implemented in Python V3.7.3, utilizing the Scikit-learn V0.22.2 library [45]. We used Scikit-learn default parameters to simplify implementing other models. We performed the experiments on an Intel Core i5-4590 computer with 8 GB RAM running Windows

7 64 bit. Average performance metrics, such as accuracy (%), precision (%), sensitivity or recall (%), specificity (%), F1 score (%), and AUC were obtained by conducting 10 runs of stratified 10-fold CV.

#### 4. Results and Discussion

This section considers the proposed DNN model performance and feature selection model impacts and discusses DR risk factors. We also verified a good generalization capability by applying our results to other public datasets (nephropathy and hypertension–diabetes).

##### 4.1. Prediction Model Performances

We applied the proposed DNN with RFE model to predict DR using known risk factors and compared the outcomes with several current best-practice data-driven models that have wide acceptance and a proven track record for accuracy and efficiency: k-nearest neighbor (KNN), C4.5 decision tree (DT), support vector machine (SVM), naïve Bayes (NB), and random forest (RF). Table 4 compares model performance metrics, averaged from 10 runs of stratified 10-fold CV. The proposed DNN with RFE model was superior to the traditional models with respect to accuracy, sensitivity, specificity, F1 score, and AUC, achieving 82.033%, 76.000%, 80.389%, 71.820%, and 0.804, respectively. In terms of detecting the positive cases, the KNN model achieved the highest precision (79.714%); however, our proposed model generated the highest recall (76.000%). The precision is the ratio of correctly predicted positive cases to the total predicted positive cases. On other hand, the recall indicates the accuracy of a model in predicting the positive cases for the cases where the actual condition is positive. Therefore, to identify prediction of positive cases (retinopathy) accurately, the model should focus more on recall rather than precision. In the medical area, it is common to focus more on sensitivity (recall) and specificity to evaluate medical tests [46]. Furthermore, it is also important to detect positive cases accurately, since when the model fails to detect the retinopathy, it will lead to blindness in such patients. Finally, the proposed model achieved 5.308% accuracy improvement compared with the current best-practice DR models.

**Table 4.** Performance metrics for diabetes retinopathy prediction.

Method	Accuracy	Precision	Sensitivity (Recall)	Specificity	F1	AUC
KNN	77.418	79.714	49.500	69.806	56.492	0.698
DT	75.989	63.095	65.000	73.222	59.558	0.732
SVM	78.846	75.214	51.500	71.361	56.333	0.714
NB	73.022	54.970	66.000	70.889	56.939	0.709
RF	78.352	62.500	39.000	67.889	45.944	0.679
Proposed model (DNN + RFE)	82.033	72.937	76.000	80.389	71.820	0.804

Accuracy rate is the most common metric for classifier performance. However, class distribution must be considered for unbalanced datasets using specific classifier metrics. ROC is a useful tool to provide evaluation criteria for unbalanced datasets [47]. The ROC curve contrasts false positive and false negative outcomes, as shown in Figure 6, where AUC indicates overall classification performance [48], with the best model having AUC ≈ 1. Figure 6 shows ROC curves analysis for the proposed and other considered classification models for DR dataset. The proposed model achieved the highest AUC = 0.80.

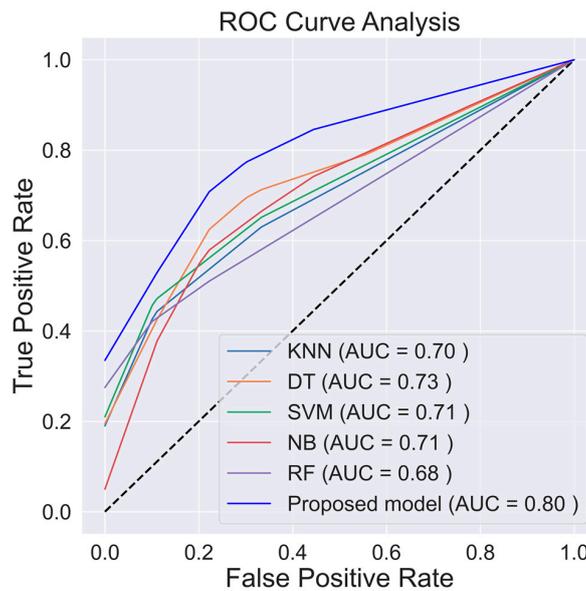


Figure 6. ROC curves for the diabetes retinopathy (DR) dataset.

Thus, the proposed model achieved significantly improved metrics compared with current best practice classification methods. Specific impacts for RFE and other feature selection methods on performance accuracy are presented in Section 4.2.

4.2. Feature Selection Impacts

The optimal number of features is required to implement RFE. Therefore, we investigated the impact of top  $k$  features on DNN accuracy. The full dataset included 14 features, and RFE expected to remove irrelevant features. Figure 7a shows the impact of the best  $k$  features as defined by RFE on DNN model accuracy. Optimal,  $k = 13$  for DR dataset, and maximum DNN model accuracy are achieved when including only these defined optimal features. The result showed that removing high number of features leads to the reduction of accuracy; therefore, the highest accuracy can be achieved by removing a small number of features.

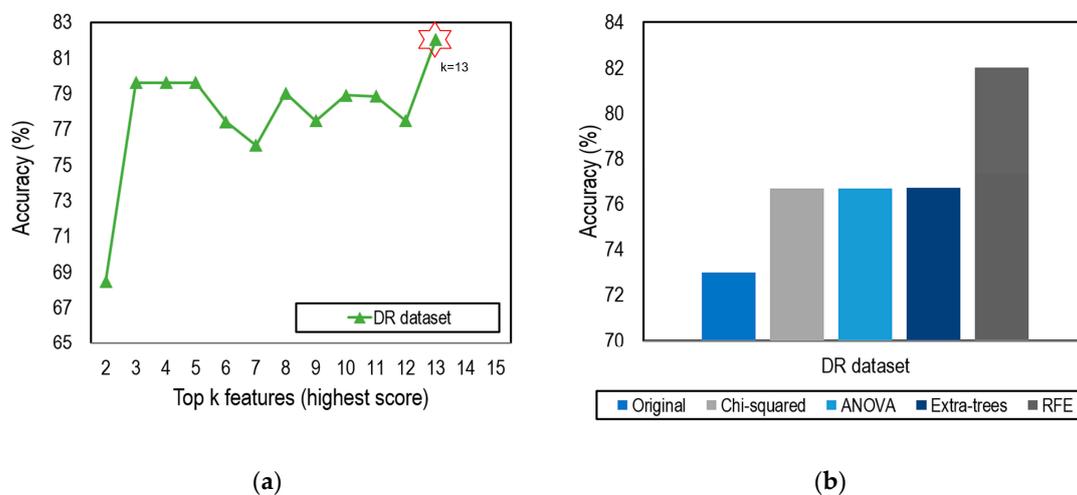


Figure 7. Proposed DNN model accuracy (a) with respect to number of top  $k$  features defined from recursive feature elimination (RFE) and (b) other feature selection models for the diabetes retinopathy (DR) dataset.

Based on the strategy of searching, the feature selection can be categorized into three methods, such as wrapper, filter, and embedded methods [49,50]. Filter methods measure the relevance of features by their correlation with the target variable, while wrapper methods utilize a learning machine to measure the usefulness of a subset of features according to their predictive power. Embedded methods perform feature selection in the process of training based on specific learning machines. In our study, we used SVM-RFE as an application of wrapper methods, while for filter methods, ANOVA and chi-squared were utilized. For embedded methods, we utilized the extra trees algorithm to extract relevant features [51]. Figure 7b shows the impact of the feature selection method on DNN model accuracy. The feature selection based on extra trees automatically selected the optimal number of features = 7. We also investigated and found that the optimal number of features for ANOVA and chi-squared are same (13 features). RFE generated superior accuracy, up to 9.07% for DR dataset compared to outcomes without feature selection. However, the other considered feature selection methods performed poorly, with only slight accuracy improvements for the DR dataset using chi-squared, ANOVA, and extra trees. Thus, RFE was the best choice for the DNN, providing the maximum DR prediction accuracy.

#### 4.3. Risk Factors and Previous Studies

Table 5 compares the proposed and previous models for DR. Previous prediction models with different input variables were applied to various populations, including Iran, South Korea, US, and Taiwan. To our best knowledge, no previous study considered DNN with RFE for DR prediction. The proposed DNN and RFE model achieved superior performance compared with all previous models in terms of model accuracy. However, SVM [17] and Lasso [14] achieved superior AUC for DR. Most of the previous studies used holdout for model validation, whereas we used stratified 10-fold CV to avoid overfitting.

**Table 5.** Comparison with previous studies for predicting diabetes retinopathy (DR).

Dataset	Population	Study	Method	Number of Features	Model Validation	Accuracy	AUC
DR	Iran	[13]	LR	9	-	-	0.704
	South Korea	[14]	Lasso	19	Holdout (67:33)	0.736	0.810
	United States	[15]	Ensemble RUSBoost	11	Holdout (80:20)	0.735	0.720
	Taiwan	[17]	SVM	10	Holdout (80:20)	0.795	0.839
	United States	[16]	ANN + SMOTE	8	Holdout (66:34)	-	0.754
	Iran	Current	DNN + RFE	13	Stratified 10-fold CV	0.820	0.804

Notes: DNN = deep neural network; RFE = recursive feature elimination; LR = logistic regression; SVM = support vector machine; ANN = artificial neural network; SMOTE = synthetic minority oversampling technique.

Risk factors and their relative importance vary across the world and the previous studies showed that important predictors could be retrieved for DR. Hosseini et al. [13] used age, BMI, sex, diabetes duration, and HbA1c; whereas Oh et al. [14] identified fasting BG, triglyceride, low BMI, and insulin therapy as strong predictors. Other studies also identified insulin use [17] and A1C [16] as important risk factors. Our proposed study used RFE to select the best variables and identified the top five risk as diabetes duration, fasting BG, HDL, Age, and A1c; which are largely consistent with selected risk factors from previous studies.

Directly comparing these results is inappropriate, since they were derived from different datasets, pre-processing methods, and validation methods. Therefore, Table 5 should not be considered as strong evidence regarding model performance, but it provides a general comparison and allows discussions regarding the proposed model and previous approaches. We used a public dataset for the current study, which was limited to small populations in Iran. Benchmarking machine learning models will become somewhat fairer as other datasets become publicly available.

#### 4.4. Another Diabetes Dataset

To evaluate the proposed prediction model robustness and generalization, we compared it with other machine learning models and datasets (diabetic nephropathy and hypertension–diabetes). Diabetic nephropathy (DN) is a serious kidney-related complication (kidney disease) relatively commonly developed by T1D and T2D patients. The DN dataset we employed was related to DR (Table 1), provided by Khodadadi et al. [40]. The dataset was gathered from 133 diabetic patients with 73 among them developing DN. Thus, we used the same 14 input features as in Table 1 but with a different output class, i.e., DN. The description of DN dataset can be seen in Appendix A (see Table A1). The objective for this dataset was to classify whether the diabetic patient would develop DN.

We also gathered a dataset from the National Health Insurance Sharing Service (NHISS) Korea comprising applicant's general health data 2013–2014 [52]. The original input variables were age group (BTH\_G), systolic blood pressure (BP), diastolic BP, fasting BG, BMI, and sex. The dataset included four classes, where the subject was diagnosed to have hypertension, diabetes (T1D or T2D), hypertension and diabetes, or healthy (no diabetes or hypertension history). We converted this multiclass into a binary classification problem, transforming it into healthy or disease (hypertension, diabetes), and removed fasting BG, since this variable is closely related with diabetes diagnosis. We randomly selected 1000 individuals from approximately 1 million, and hence, the final dataset comprised 761 healthy and 239 disease (hypertension, diabetes) patients. The description of NHISS dataset can be seen in Appendix A (see Table A2). The objective for this dataset was to classify whether the subject would develop disease (i.e., hypertension or diabetes).

Table 6 compares classification performance for the proposed DNN + RFE model with other machine learning models (KNN, DT, SVM, NB, and RF). Average accuracy and AUC were calculated from 10 runs of stratified 10-fold CV. The proposed model achieved superior accuracy and AUC for both datasets: 84.121% and 0.839 for the DN and 81.600% and 0.702 for the NHISS datasets, respectively.

**Table 6.** Proposed prediction model performance compared with other models for public datasets.

Method	DN		NHISS	
	Accuracy	AUC	Accuracy	AUC
KNN	81.813	0.814	80.600	0.690
DT	81.978	0.817	80.900	0.679
SVM	83.297	0.834	80.700	0.665
NB	67.527	0.649	79.600	0.701
RF	82.747	0.825	80.775	0.664
Proposed DNN + RFE	84.121	0.839	81.600	0.702

Notes: DN = diabetic nephropathy dataset, NHISS = hypertension–diabetes dataset from NHISS, Korea. RF = random forest; DNN = deep neural network; RFE = recursive feature elimination; KNN = k-nearest neighbor; SVM = support vector machine; DT = decision tree; NB = naïve Bayes.

Applying RFE, we selected 13 features for the DN dataset and 3 for the NHISS dataset (BTH\_G, Systolic BP, and BMI), and we used grid search to optimize the DNN hyperparameter. Thus, we found the optimal DNN design for the DN dataset to be five hidden layers (100, 64, 128, 64, 32), ReLU activation function, and a maximum of 100 iterations. For NHISS dataset, DNN with five hidden layers (100, 64, 128, 64, 32) and tanh activation function has achieved highest accuracy. The best optimization algorithms were identified as LBFGS (limited-memory Broyden–Fletcher–Goldfarb–Shanno) and Adam for the DN and NHISS datasets, respectively. These additional experiments confirmed the proposed model robustness toward different healthcare domains.

## 5. Conclusions and Future Work

The current study combined DNN with RFE to predict DR. The model is expected to help individuals foresee DR danger based on risk factors during initial disease phases. A public dataset

incorporating DR risk factors was utilized, and the proposed model performance was compared with previous best-practice KNN, DT, SVM, NB, and RF models. The proposed model outperformed conventional classification models and most other previous models, achieving accuracy = 82.033%.

We applied RFE to identify significant features from all datasets, i.e., extracting the highest DR risk factors. Combining RFE and DNN improved the prediction accuracy as compared with all other considered feature selection methods (chi-squared, ANOVA, and extra trees). The proposed DNN + RFE model improved the accuracy (9.07%) compared with DNN without feature selection. Thus, machine learning combined with feature selection can effectively detect DR. This offers increased cost-effectiveness for health care systems, where decision support based on the proposed prediction model could provide decision opinions. We hope this study will help reduce the DR risk for diabetic patients, which is the major cause of blindness.

The dataset used here was from a relatively small and quite specific population; hence, the prediction model outcomes cannot not be simply generalized for broader application. Similarly, the identified important risk factors might not be appropriate for other populations. Thus, as it stands, the proposed model would be unsuitable for clinical trials due to dataset limitations. Therefore, the proposed approach should be extended to other clinical datasets and compared widely with other prediction and feature selection models. Once model validation is extended to broader datasets, other risk factors affecting DR could be identified.

**Author Contributions:** Conceptualization, G.A. and P.S.; data curation, G.A. and J.S.; formal analysis, N.L.F.; methodology, G.A. and M.S.; project administration, J.R.; resources, P.S. and J.S.; software, G.A. and M.S.; validation, M.A.; visualization, N.L.F.; writing—original draft, G.A.; writing—review & editing, M.S. and M.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper receives no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript.

ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AUC	Area under the ROC Curve
BG	Blood Glucose
BMI	Body Mass Index
CV	Cross Validation
CVD	Cardiovascular Disease
DBP	Diastolic Blood Pressure
Dias BP	Diastolic Blood Pressure
DM	Diabetes Mellitus
DN	Diabetic Nephropathy
DNN	Deep Neural Network
DR	Diabetic Retinopathy
DT	Decision Tree
FBS	Fasting Blood Sugar
GWO	Grey Wolf Optimization
HbA1c	Hemoglobin A1c
HDL	High-density Lipoproteins
KNN	K-Nearest Neighbor
LBFGS	Limited-memory Broyden–Fletcher–Goldfarb–Shanno
LDL	Low-density Lipoprotein
LR	Logistic Regression
NB	Naïve Bayes
NHISS	National Health Insurance Sharing Service
PCA	Principal Components Analysis

PVD	Peripheral Vessel Disease
ReLU	Rectified Linear Units
RF	Random Forest
RFE	Recursive Feature Elimination
ROC	Receiver Operating Characteristic
SBP	Systolic Blood Pressure
SGD	Stochastic Gradient Descent
SMOTE	Synthetic Minority Over-sampling Technique
SVM	Support Vector Machine
SVM-RFE	Support Vector Machine–Recursive Feature Elimination
Sys BP	Systolic Blood Pressure
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
TG	Triglyceride

### Appendix A

Table A1 describes the diabetes nephropathy (DN) dataset, comprises residents’ risk factors from 133 diabetic patients with 73 among them developing DN (Khodadadi et al. [40]). The dataset is made publicly available by previous authors (<https://data.mendeley.com/datasets/k62fdsnwk/1>).

**Table A1.** Diabetic nephropathy (DN) dataset.

No	Attribute	Description	Type	Range
1	BMI	Subject’s body mass index	Numeric	18–41
2	DM duration	Subject’s diabetes duration (y)	Numeric	0–30
3	A1c	Subject’s average blood glucose level over the past 3 months (mg/dL)	Numeric	6.5–13.3
4	Age	Subject’s age (y)	Numeric	16–79
5	FBS	Subject’s fasting blood sugar level (mg/dL)	Numeric	80–510
6	LDL	Subject’s low-density lipoprotein level (mg/dL)	Numeric	36–267
7	HDL	Subject’s high-density lipoprotein level (mg/dL)	Numeric	20–62
8	TG	Subject’s triglyceride level (mg/dL)	Numeric	74–756
9	Sys BP	Subject’s systolic blood pressure (mmHg)	Numeric	105–180
10	Dias BP	Subject’s diastolic blood pressure (mmHg)	Numeric	60–120
11	Sex	Subject’s sex	Categorical	0 = Female 1 = Male
12	DM type	Subject’s diabetes type	Categorical	0 = T1D 1 = T2D
13	DM treat	Subject’s diabetes treatment	Categorical	0 = Both (Insulin and oral agent) 1 = Insulin 2 = Oral agent
14	Statin	Subject’s statin status (frequently used as part of diabetes care)	Categorical	0 = Ator (atorvastatin) 1 = No statin 2 = ROS (rosuvastatin)
15	Nephropathy (class)	Subject’s nephropathy status	Categorical	0 = No (60) 1 = Yes (73)

Table A2 describes the NHISS Korea dataset, comprises residents' risk factors from 1000 individuals with 239 among them developing disease (hypertension, diabetes) [50]. The NHISS Korea dataset is available online (<https://nhiss.nhiss.or.kr/bd/ab/bdabf003cv.do>).

**Table A2.** National Health Insurance Sharing Service (NHISS) Korea dataset.

No	Attribute	Description	Type	Range
1	BTH_G	Age group of a subject	Categorical	0 = 20–24 1 = 25–26 2 = 27–28 ... 26 = greater than 75
2	SBP	Subject's systolic blood pressure (mmHg)	Numeric	84–190
3	DBP	Subject's diastolic blood pressure (mmHg)	Numeric	50–120
4	BMI	Subject's body mass index	Numeric	15.6–39.9
5	SEX	Subject's sex	Categorical	0 = Male 1 = Female
6	DIS (class)	Subject's disease (hypertension, diabetes) status	Categorical	0 = No (761) 1 = Yes (239)

## References

1. American Diabetes Association. Standards of medical care in diabetes. *Diabetes Care* **2006**, *29* (Suppl. 1), S4–S42.
2. American Diabetes Association. Introduction: Standards of Medical Care in Diabetes. *Diabetes Care* **2018**, *41*, S1–S2. [[CrossRef](#)] [[PubMed](#)]
3. Silverstein, J.; Klingensmith, G.; Copeland, K.; Plotnick, L.; Kaufman, F.; Laffel, L.; Deeb, L.; Grey, M.; Anderson, B.; Holzmeister, L.A.; et al. Care of Children and Adolescents with Type 1 Diabetes: A statement of the American Diabetes Association. *Diabetes Care* **2005**, *28*, 186–212. [[CrossRef](#)] [[PubMed](#)]
4. Goldstein, D.E.; Little, R.R.; Lorenz, R.A.; Malone, J.I.; Nathan, D.; Peterson, C.M.; Sacks, D.B. Tests of Glycemia in Diabetes. *Diabetes Care* **2004**, *27*, 1761–1773. [[CrossRef](#)] [[PubMed](#)]
5. Klein, R. Relationship of Hyperglycemia to the Long-term Incidence and Progression of Diabetic Retinopathy. *Arch. Intern. Med.* **1994**, *154*, 2169. [[CrossRef](#)]
6. Mohamed, S.; Murray, J.C.; Dagle, J.M.; Colaizy, T. Hyperglycemia as a risk factor for the development of retinopathy of prematurity. *BMC Pediatr.* **2013**, *13*, 78. [[CrossRef](#)]
7. Schrijvers, B.F.; De Vriese, A.S.; Flyvbjerg, A. From Hyperglycemia to Diabetic Kidney Disease: The Role of Metabolic, Hemodynamic, Intracellular Factors and Growth Factors/Cytokines. *Endocr. Rev.* **2004**, *25*, 971–1010. [[CrossRef](#)]
8. Alicic, R.Z.; Rooney, M.T.; Tuttle, K.R. Diabetic Kidney Disease: Challenges, Progress, and Possibilities. *Clin. J. Am. Soc. Nephrol.* **2017**, *12*, 2032–2045. [[CrossRef](#)]
9. Selvin, E.; Marinopoulos, S.; Berkenblit, G.; Rami, T.; Brancati, F.L.; Powe, N.R.; Golden, S.H. Meta-Analysis: Glycosylated Hemoglobin and Cardiovascular Disease in Diabetes Mellitus. *Ann. Intern. Med.* **2004**, *141*, 421. [[CrossRef](#)]
10. Ormazabal, V.; Nair, S.; Elfeky, O.; Aguayo, C.; Salomon, C.; Zuñiga, F.A. Association between insulin resistance and the development of cardiovascular disease. *Cardiovasc. Diabetol.* **2018**, *17*, 122. [[CrossRef](#)]
11. Cheung, N.; Mitchell, P.; Wong, T.Y. Diabetic retinopathy. *Lancet* **2010**, *376*, 124–136. [[CrossRef](#)]
12. Golubnitschaja, O. Advanced Diabetes Care: Three Levels of Prediction, Prevention & Personalized Treatment. *Curr. Diabetes Rev.* **2010**, *6*, 42–51. [[CrossRef](#)] [[PubMed](#)]
13. Hosseini, S.M.; Maracy, M.R.; Amini, M.; Baradaran, H.R. A risk score development for diabetic retinopathy screening in Isfahan-Iran. *J. Res. Med. Sci.* **2009**, *14*, 105–110. [[PubMed](#)]
14. Oh, E.; Yoo, T.K.; Park, E.-C. Diabetic retinopathy risk prediction for fundus examination using sparse learning: A cross-sectional study. *BMC Med. Inform. Decis. Mak.* **2013**, *13*, 106. [[CrossRef](#)]

15. Ogunyemi, O.; Kermah, D. Machine Learning Approaches for Detecting Diabetic Retinopathy from Clinical and Public Health Records. *AMIA Annu. Symp. Proc.* **2015**, *2015*, 983–990.
16. Ogunyemi, O.I.; Gandhi, M.; Tayek, C. Predictive Models for Diabetic Retinopathy from Non-Image Teleretinal Screening Data. *AMIA Jt. Summits Transl. Sci. Proc.* **2019**, *2019*, 472–477.
17. Tsao, H.-Y.; Chan, P.-Y.; Su, E.C.-Y. Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms. *BMC Bioinform.* **2018**, *19*, 283. [[CrossRef](#)]
18. Mallick, P.K.; Ryu, S.H.; Satapathy, S.K.; Mishra, S.; Nguyen, G.N.; Tiwari, P. Brain MRI Image Classification for Cancer Detection Using Deep Wavelet Autoencoder-Based Deep Neural Network. *IEEE Access* **2019**, *7*, 46278–46287. [[CrossRef](#)]
19. Akyol, K. Comparing of deep neural networks and extreme learning machines based on growing and pruning approach. *Expert Syst. Appl.* **2020**, *140*, 112875. [[CrossRef](#)]
20. De Falco, I.; De Pietro, G.; Della Cioppa, A.; Sannino, G.; Scafuri, U.; Tarantino, E. Evolution-based configuration optimization of a Deep Neural Network for the classification of Obstructive Sleep Apnea episodes. *Future Gener. Comput. Syst.* **2019**, *98*, 377–391. [[CrossRef](#)]
21. Koshimizu, H.; Kojima, R.; Kario, K.; Okuno, Y. Prediction of blood pressure variability using deep neural networks. *Int. J. Med. Inform.* **2020**, *136*, 104067. [[CrossRef](#)] [[PubMed](#)]
22. Kim, J.; Kim, J.; Kwak, M.J.; Bajaj, M. Genetic prediction of type 2 diabetes using deep neural network. *Clin. Genet.* **2018**, *93*, 822–829. [[CrossRef](#)] [[PubMed](#)]
23. Kannadasan, K.; Edla, D.R.; Kuppli, V. Type 2 diabetes data classification using stacked autoencoders in deep neural networks. *Clin. Epidemiol. Glob. Health* **2019**, *7*, 530–535. [[CrossRef](#)]
24. Ramachandran, N.; Hong, S.C.; Sime, M.J.; Wilson, G.A. Diabetic retinopathy screening using deep neural network: Diabetic retinopathy screening. *Clin. Exp. Ophthalmol.* **2018**, *46*, 412–416. [[CrossRef](#)]
25. Gao, Z.; Li, J.; Guo, J.; Chen, Y.; Yi, Z.; Zhong, J. Diagnosis of Diabetic Retinopathy Using Deep Neural Networks. *IEEE Access* **2019**, *7*, 3360–3370. [[CrossRef](#)]
26. Parmar, R.; Lakshmanan, R.; Purushotham, S.; Soundrapandiyan, R. Detecting Diabetic Retinopathy from Retinal Images Using CUDA Deep Neural Network. In *Intelligent Pervasive Computing Systems for Smarter Healthcare*; Sangaiah, A.K., Shantharajah, S., Theagarajan, P., Eds.; Wiley: Hoboken, NJ, USA, 2019; pp. 379–396. ISBN 978-1-119-43896-0.
27. Shankar, K.; Perumal, E.; Vidhyavathi, R.M. Deep neural network with moth search optimization algorithm based detection and classification of diabetic retinopathy images. *SN Appl. Sci.* **2020**, *2*, 748. [[CrossRef](#)]
28. Ayhan, M.S.; Kühlewein, L.; Aliyeva, G.; Inhoffen, W.; Ziemssen, F.; Berens, P. Expert-validated estimation of diagnostic uncertainty for deep neural networks in diabetic retinopathy detection. *Med. Image Anal.* **2020**, *64*, 101724. [[CrossRef](#)]
29. Gadekallu, T.R.; Khare, N.; Bhattacharya, S.; Singh, S.; Maddikunta, P.K.R.; Srivastava, G. Deep neural networks to predict diabetic retinopathy. *J. Ambient Intell. Human Comput.* **2020**. [[CrossRef](#)]
30. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **2002**, *46*, 389–422. [[CrossRef](#)]
31. Gysels, E.; Renevey, P.; Celka, P. SVM-based recursive feature elimination to compare phase synchronization computed from broadband and narrowband EEG signals in Brain–Computer Interfaces. *Signal Process.* **2005**, *85*, 2178–2189. [[CrossRef](#)]
32. Yin, Z.; Zhang, J. Operator functional state classification using least-square support vector machine based recursive feature elimination technique. *Comput. Methods Programs Biomed.* **2014**, *113*, 101–115. [[CrossRef](#)] [[PubMed](#)]
33. Senturk, Z.K. Early diagnosis of Parkinson’s disease using machine learning algorithms. *Med. Hypotheses* **2020**, *138*, 109603. [[CrossRef](#)] [[PubMed](#)]
34. Chatterjee, S.; Dey, D.; Munshi, S. Integration of morphological preprocessing and fractal based feature extraction with recursive feature elimination for skin lesion types classification. *Comput. Methods Programs Biomed.* **2019**, *178*, 201–218. [[CrossRef](#)] [[PubMed](#)]
35. Wang, C.; Xiao, Z.; Wang, B.; Wu, J. Identification of Autism Based on SVM-RFE and Stacked Sparse Auto-Encoder. *IEEE Access* **2019**, *7*, 118030–118036. [[CrossRef](#)]
36. Richhariya, B.; Tanveer, M.; Rashid, A.H. Diagnosis of Alzheimer’s disease using universum support vector machine based recursive feature elimination (USVM-RFE). *Biomed. Signal Process. Control* **2020**, *59*, 101903. [[CrossRef](#)]

37. Kumar, A.; Sharmila, D.J.S.; Singh, S. SVMRFE based approach for prediction of most discriminatory gene target for type II diabetes. *Genom. Data* **2017**, *12*, 28–37. [[CrossRef](#)]
38. Chen, Z.; Pang, M.; Zhao, Z.; Li, S.; Miao, R.; Zhang, Y.; Feng, X.; Feng, X.; Zhang, Y.; Duan, M.; et al. Feature selection may improve deep neural networks for the bioinformatics problems. *Bioinformatics* **2020**, *36*, 1542–1552. [[CrossRef](#)] [[PubMed](#)]
39. Karthik, S.; Srinivasa Perumal, R.; Chandra Mouli, P.V.S.S.R. Breast Cancer Classification Using Deep Neural Networks. In *Knowledge Computing and Its Applications*; Margret Anouncia, S., Wiil, U.K., Eds.; Springer: Singapore, 2018; pp. 227–241. ISBN 978-981-10-6679-5.
40. Khodadadi, B.; Mousavi, N.; Mousavi, M.; Baharvand, P.; Ahmadi, S.A.Y. Diagnosis and predictive clinical and para-clinical cutoffs for diabetes complications in Lur and Lak populations of Iran; a ROC curve analysis to design a regional guideline. *J. Nephropharmacol.* **2018**, *7*, 83–89. [[CrossRef](#)]
41. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence—Volume 2, Montreal, QC, Canada, 20–25 August 1995; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1995; pp. 1137–1143.
42. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
43. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
44. Jiménez, Á.B.; Lázaro, J.L.; Dorransoro, J.R. Finding Optimal Model Parameters by Discrete Grid Search. In *Innovations in Hybrid Intelligent Systems. Advances in Soft Computing*; Corchado, E., Corchado, J.M., Abraham, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; Volume 44, pp. 120–127. ISBN 978-3-540-74971-4.
45. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
46. Kim, H.-E.; Kim, H.H.; Han, B.-K.; Kim, K.H.; Han, K.; Nam, H.; Lee, E.H.; Kim, E.-K. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: A retrospective, multireader study. *Lancet Digit. Health* **2020**, *2*, e138–e148. [[CrossRef](#)]
47. Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [[CrossRef](#)]
48. Huang, J.; Ling, C.X. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 299–310. [[CrossRef](#)]
49. Guyon, I.; Elisseeff, A. An Introduction to Feature Extraction. In *Feature Extraction. Studies in Fuzziness and Soft Computing*; Guyon, I., Nikravesh, M., Gunn, S., Zadeh, L.A., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; Volume 207, pp. 1–25. ISBN 978-3-540-35487-1.
50. Miao, J.; Niu, L. A Survey on Feature Selection. *Procedia Comput. Sci.* **2016**, *91*, 919–926. [[CrossRef](#)]
51. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [[CrossRef](#)]
52. National Health Insurance Sharing Service (NHIS) Korea. Available online: <https://nhiss.nhis.or.kr/bd/ab/bdabf003cv.do> (accessed on 20 August 2020).

