Selection and Assessment of Diabetes Risk Factors with Data Analysis and Machine Learning Techniques Using Physiological Data

Sodavy Tong
Department of Computer Science /
Kasetsart University
Bangkok / Thailand
sodavy.t@ku.th

Dangjai Souvannakitti
Department of Physiology /
Phramongkutklao College of Medicine
Bangkok / Thailand
drdangjai@pcm.ac.th

Nopadon Juneam*
Department of Computer Science /
Kasetsart University
Bangkok / Thailand
fscindj@ku.ac.th

ABSTRACT — This research presents a method for selecting and assessing diabetes risk factors with data analysis and machine learning techniques using physiological data, where the real data set used is from the Royal Dunyapabbumbad Foundation. The data set consists of the detailed physical examination data of 453 samples. The results of the data analysis show the statistical correlation of factors that significantly affects the risk levels given by the traditional diabetes risk score modeling, leading to the selection of proper risk factors for diabetes prediction towards the samples in the data set. The properness of the selected risk factors is then assessed via the predictive models generated by machine learning algorithms including Random Forest, k-Nearest Neighbors, Logistic Regression, Support Vector Machine, and Multilayer Perceptron. The experimental results show that using risk factors including genders, family history of diabetes, waist circumference levels, blood pressure levels, hypertension, family history of hypertension, family history of cardiovascular disease, smoking habits, and BMI levels, results in the highest overall performing models.

Keywords *— diabetes risk factors, predictive analytics, feature selection, machine learning, data science*

1. Introduction

According to the 2019–2020 Thai National Health Survey of the Thai population through physical examinations, the prevalence of diabetes among Thai individuals aged 15 years and older was found to be 9.5% [1]. Most of the cases were Type 2 Diabetes. The progression of Type 2 Diabetes in most patients is typically gradual and does not present with acute symptoms. The disease is caused by the pancreas's inability to produce enough insulin to regulate blood sugar levels effectively, along with the body's development of insulin resistance. Without early screening, patients are often unaware that they have abnormal blood sugar levels indicative of diabetes. And if left untreated, chronic and serious complications can arise, such as complications of the eyes, kidneys, and nerve problems, arterial blockages, coronary artery disease, and ischemic stroke, etc.

The survey also found that up to 30.6% of diabetic patients had never been diagnosed and were unaware that they had the disease [1]. Therefore, diabetes screening to detect asymptomatic individuals with diabetes is important, as early diagnosis and treatment can prevent long-term complications. However, universal diabetes screening is not feasible for everyone due to the high cost of the process. Diabetes risk assessment prior to screening is therefore an important step to select only those patients with a high probability of having diabetes for further screening. The benefit of risk assessment is that it makes the screening process more economical and cost-effective.

Generally, diabetes risk assessment commonly uses findings from studies on risk factors. Nevertheless, there are numerous risk factors for diabetes, and their impact on disease development varies depending on ethnicity and geographical conditions. Consequently, risk assessment must consider all relevant factors in combination [2].

Table 1. Conversion of data into diabetes risk scores.

Risk factors	Risk score					
	0	1	2	3	4	5
Age (year)	34-44	45-49	≥ 50	-	-	-
Gender	Male	-	Female	-	-	-
BMI levels (kg/m^2)	< 23	-	—	≥ 23 and < 27.5	-	≥27.5
waist circumference (cm)	< 90 (Male)	-	≥ 90 (Male)	-	-	-
	< 80 (Female)	-	≥ 80 (Female)	-	-	-
hypertension	No	-	Yes	-	-	-
Family history of diabetes	No	-	-	-	Yes	-

* Corresponding author

Table 2. Interpretation of the diabetes risk score sum.

Total score	Risk of diabetes in the next 12 years	Chance of developing diabetes in the future	Risk Levels
≤ 2	< 5%	1 in 20 person (5%)	Low
3-5	5%-10%	1 in 12 person (8.33%)	Medium
6-8	11%-20%	1 in 7 person (14.29%)	High
> 8	≥ 20	1 in 3-4 person (25%-33.33%)	Very High

For the assessment of Type 2 Diabetes risk in Thailand, the findings from a long-term (Cohort Study) [2] have been utilized. This study examined several risk factors that can be easily assessed through questionnaires and basic physical examinations, such as age, gender, body mass index (BMI), waist circumference, hypertension, and family history of diabetes. The collected data is then calculated into Diabetes Risk Scores according to Table 1, and the total score is interpreted into Risk Levels according to Table 2. The study results indicate that these scores can predict the risk of developing diabetes over the next 12 years with reasonable accuracy in the Thai population. However, assessing diabetes risk for specific subgroups requires further analysis to select and evaluate risk factors appropriate for those particular groups.

From the literature review, it was found that most related research has applied machine learning techniques with the goal of creating models to predict diabetes risk from the datasets used. Examples include predicting diabetes risk in the middle-aged population in Sweden [3], predicting diabetes risk in rural populations in China [4], and long-term diabetes risk prediction in [5], etc. Our research aims to propose a method for selecting and assessment diabetes risk factors for the study's sample using data analysis and machine learning techniques with Physiological Data.

In this research, we use a real dataset from the Royal Dunyapabbumbad Foundation [6]. This dataset comprises detailed physical examination data from 453 individuals, with a total of 32 physiological attributes collected. Considering the incompleteness of the dataset for predictive use, we applied statistical data analysis techniques to select predictive features. The results from the Analysis of Variance (ANOVA) demonstrated a statistically significant relationship between factors influencing traditional risk assessment levels. This led to the selection of appropriate risk factors for predicting diabetes for a sample group in the dataset. The suitability of the selected risk factors was evaluated through predictive models (Classifiers) built using machine learning algorithms, including Random Forest, k-Nearest Neighbors, Logistic Regression, Support Vector Machine, and Multilayer Perceptron. Experimental results from this research indicate that using risk factors such as gender, family history of diabetes, waist circumference, blood pressure level, hypertension, family history of hypertension, family history of cardiovascular disease, smoking habits, and BMI resulted in the best overall predictive performance on the test dataset. This confirms the appropriateness of these risk factors for predicting diabetes in the sample group from the dataset.

The remainder of this research paper is organized as follows: Section 2 discusses the dataset used in this research and its characteristics; Section 3 presents the relevant background, including data analysis techniques and machine learning methods employed in the research; Section 4 describes the methodology for selecting diabetes risk factors using data analysis techniques; Section 5 explains the methodology for assessing diabetes risk factors using machine learning techniques; Section 6 discusses the results obtained from the risk assessment; Section 7, the final section, provides a summary of the research.

2. Dataset Description

In this research, we used physiological data from the Royal Dunyapabbumbad Foundation [6]. The use of this dataset for research is part of the project to develop a holistic, personalized healthcare platform. The dataset consists of detailed physical examination records of 453 anonymous volunteers and contains a total of 32 physiological attributes collected from the dataset. These attributes can be divided into two types, including 18 categorical attributes and 14 numerical attributes. Furthermore, the attributes can be grouped into three main groups:

- Group 1: General physical attributes, including gender, age, height, weight, BMI, waist circumference, smoking habits, smoking history, alcohol drinking, and alcohol drinking history. This group contains a total of 10 attributes.
- Group 2: Attributes related to existing medical conditions, such as those associated with diabetes, hypertension, coronary artery disease, arrhythmia, gout, hyperlipidemia, and family history of various chronic diseases. There are 17 attributes in this group.
- Group 3: Attributes related to the risk level of chronic diseases, including diabetes risk level (based on traditional assessment), BMI level, waist circumference level, blood pressure level, and risk level for cardiovascular disease. This group consists of 5 attributes.

From the exploration of the dataset characteristics for predictive use in diabetes risk, it was found that the number of samples classified by Class or Risk Level is shown in Table 3. In addition, it was also found that the dataset has issues with data incompleteness, as follows:

2.1. Missing Values in the Dataset

The dataset contains samples with missing values due to incomplete data collection. When considering all attributes, it was found that there are 7 attributes with more than 30% missing values,

which include: blood lipid level, cholesterol level, LDL (bad fat), HDL (good fat), triglyceride level, uric acid level, and plasma glucose level.

2.2. Issues with Predictive Feature Selection

The dataset contains an excessive number of features due to extensive data collection. Most of these features may not be relevant for diabetes prediction. Additionally, redundancy was observed among several features collected in the dataset.

Table 3. Number of samples in the dataset, training set, and test set.

Risk Level (Class)	All samples (100%)	Training set (70%)	Test set (30%)
Low (0)	209 (46.14%)	146 (46.06%)	63 (46.32%)
Medium (1)	106 (23.40%)	79 (24.92%)	27 (19.85%)
High (2)	131 (28.92%)	86 (27.13%)	45 (33.09%)
Very High (3)	7 (1.54%)	6 (1.89%)	1 (0.74%)
Total	**453** (100%)	**317** (100%)	**136** (100%)

3. Data Analysis and Machine Learning Fundamentals

This section discusses the fundamentals related to data analysis techniques using statistical methods and machine learning applied in the research.

3.1. Data Analysis Techniques Using Statistical Methods

3.1.1. Analysis of Variance (ANOVA)

ANOVA is a technique used to test the differences between two or more sample groups. It involves calculating the variance, which is a basic statistical measure used to assess deviation or the distribution of variable values within a sample group compared to the mean. ANOVA is based on the assumption that sample groups with non-significant differences will exhibit overlapping or similar value distributions, and conversely, groups with significant differences will not.

In feature selection, ANOVA can be applied to test the relationship between features and classes. This is done by considering the ratio of the variance of the feature values within each class or the Within-Group Mean Variance to the variance of the feature values between classes or the Between-Group Mean Variance. This ratio is called the F-ratio or F-value. If the F-value is close to 1, it indicates a tendency for the feature to be independent of the class. Conversely, if the F-value is greater than 1, it suggests the feature is not independent of the class.

In this research, we used One-Way ANOVA because it was appropriate for the characteristics of our dataset. The mathematical formula is as follows:

Consider a feature x of interest from l different sample groups. Let n_j be the number of samples in group j. Let x_{ij} represent the feature value of sample i in group j, and let $\bar{x}_j$ be the Individual Mean of feature x for samples in group j, as shown in Equation 1, which displays the calculation for $\bar{x}_j$.

$$\bar{x}_j = \frac{\sum_{\mathrm{i}=1}^{n_j} x_{ij}}{n_j} \tag{1}$$

The Within-Group Variance, denoted as $\mathrm{SS_w}$, can be measured using the sum of squared deviations from the individual means, as shown in Equation 2.

$$SS_W = \sum_{j=1}^{l} \sum_{i=1}^{n_j} (\mathrm{x}_{ij} - \bar{x}_j)^2 \tag{2}$$

Let $\bar{x}$ represent the Grand Mean of the feature x, as shown in Equation 3, which illustrates the calculation of $\bar{x}$.

$$\bar{x} = \sum_{1}^{l} \bar{x}_j \tag{3}$$

The Between-Group Variance, denoted as SS_B, can be measured using the sum of squared deviations from the grand mean, as shown in Equation 4.

$$SS_B = \sum_{j=1}^{l} (\bar{x}_j - \bar{x})^2 \tag{4}$$

Equations 5 and 6 show the calculations for the within-group mean variance and the between-group mean variance, denoted as MS_W and MS_B, respectively.

$$MS_B = \frac{SS_B}{k-1} \tag{5}$$

$$MS_W = \frac{SS_W}{N-k} \tag{6}$$

Equation 7 shows the calculation of the F-value, denoted by the symbol F.

$$F = \frac{MS_B}{MS_W} \tag{7}$$

3.1.3. P-Value

The P-value, or probability value, is a measure used to assess the overall probability of an observed outcome under the assumption of no difference between sample groups, also known as the Null Hypothesis. Typically, the result from a hypothesis test, such as the F-value, is converted into a P-value. This value indicates the probability of obtaining the observed F-value under the F-distribution. If the resulting P-value is very low, it means that the observed outcome is highly unlikely to occur under the null hypothesis. In testing the null

hypothesis, the P-value that can either confirm or reject the hypothesis depends on the predefined statistical significance level, or Alpha Level (or α), which is used as a criterion. Generally, and in this research, the alpha level is set at 0.05 or α = 5%.

3.2. Related Machine Learning Techniques

3.2.1. Supervised Learning

Machine learning involves applying Learning Algorithms to create a model that can explain the relationship between data in a dataset. In this research, we applied Supervised Learning for Multi-Class Classification. The main process of this technique is Model Training, which uses a training set consisting of both explanatory variables and explained variables. The explained variables must have class labels, also known as labeled data. If we let the set of explanatory variables be x and the class be the explained variable y, then model training is the application of a learning algorithm to the training set. This results in a model h that can explain the relationship from x to y, or in other words, $h(x) = y$.

In this research, we applied several learning algorithms to build models, including Random Forest (RF), k-Nearest Neighbors (k-NN), Logistic Regression (LR), Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP). Readers can refer to [4] for more detailed information about all of these learning algorithms.

3.2.2. Model Tuning Techniques

To apply learning algorithms and build models that yield the best results, it is essential to properly configure the Hyperparameters of the algorithm to suit the training dataset. Inappropriate settings may lead to a decrease in the model's accuracy when applied to the test dataset. This issue can arise from either an Overfitting Model, where the model is too accurate with the training data but fails to explain the relationship in the test data, or an Underfitting Model, where the model is not accurate with the training data and also fails to explain the relationship in the test data.

In this research, we applied model tuning using the Grid Search method in combination with 6-Fold Cross Validation. 6-Fold Cross Validation is a method used to evaluate the average performance of a model across six iterations. The training dataset is randomly divided into six parts. In each iteration, one part is used as the validation set to test the model's performance, while the remaining five parts are used to train the model. Model tuning using Grid Search in conjunction with 6-Fold Cross Validation, the process searches for the set of hyperparameter values that result in the best average performance of the model on the validation sets.

4. Risk Factor Selection using Data Analysis Techniques

The selection of diabetes risk factors using physiological data in this research consists of three main steps: Data Pre-processing, ANOVA of the Sample Groups in the Dataset, and Selection of Risk Factors Using the Results of ANOVA, respectively. The details of each step are described as follows:

4.1. Data Pre-Processing

The goal of data pre-processing is to prepare the dataset for predictive analysis. This involves several sub-steps:

4.1.1 Data Cleaning and Data Transformation

In this step, the dataset is converted into a suitable numerical format. As a result, all data are transformed into numeric values. The features in the dataset can be divided into two types: Categorical features, consisting of 18 features, and Numerical features, consisting of 14 features.

4.1.2. Handling of Missing Values

Due to the presence of missing values in the dataset, this step involved removing certain features with more than 30% missing data. As a result, 25 usable features remained in the dataset, consisting of 18 categorical features and 7 numerical features.

4.2 ANOVA of the Sample Groups in the Dataset

In this step, ANOVA was applied to test the independence between each feature and the sample groups categorized by risk levels. The F-values and P-values were calculated using the dataset after missing values had been handled. Table 4 presents the features that passed the ANOVA filtering with a significance level of α=0.05, resulting in a total of 17 selected features.

Table 4. Results from ANOVA (α=0.05)

No.	Attribute	Attribute Type	F-Value	P-Value
1	family history of diabetes	categorical	187.37	0
2	waist circumference	numerical	33.55	0
3	weight	numerical	25.92	0
4	BMI	numerical	18.81	0
5	BMI level	categorical	16.03	$1.00*10^{-9}$
6	waist size level	categorical	15.09	$3.40*10^{-9}$
7	diastolic blood pressure	numerical	12.41	$1.08*10^{-7}$
8	blood pressure level	categorical	11.92	$2.06*10^{-7}$
9	hypertension	categorical	11.80	$2.41*10^{-7}$
10	systolic blood pressure	numerical	11.34	$4.42*10^{-7}$
11	Cardiovascular risk level	categorical	9.33	$6.35*10^{-6}$
12	gender	categorical	8.09	$3.31*10^{-5}$
13	family history of hypertension	categorical	7.05	0.000133
14	family history of cardiovascular disease	categorical	5.52	0.001049
15	height	numerical	5.12	0.001799
16	age	numerical	5.10	0.001856
17	Smoking habits	categorical	5.03	0.002039

4.3 Selection of Risk Factors Using the Results of ANOVA

Upon reviewing the features in Table 4, we found that many were

still redundant as a result of the data collection process. This data redundancy appeared in both numerical and categorical features, such as BMI/BMI level/weight/height; waist circumference/waist circumference level; and blood pressure level/systolic blood pressure/diastolic blood pressure. However, these redundant features could be used interchangeably for risk factor selection. Based on this, we selected features to form three sets of risk factors as follows:

- Risk Factor Set **A**: A set consisting of all 17 features that passed the ANOVA filtering.
- Risk Factor Set **B**: A set of risk factors with reduced redundancy, created by selecting only the numerical features. This set consists of 12 features: *gender, age, height, waist circumference, weight, diastolic blood pressure, systolic blood pressure, family history of diabetes, hypertension, family history of hypertension, family history of cardiovascular disease,* and *smoking habits.*
- Risk Factor Set **C**: A set of risk factors with reduced redundancy, created by selecting only the categorical features. This set consists of 9 features: *gender, family history of diabetes, waist circumference level, blood pressure level, hypertension, family history of hypertension, family history of cardiovascular disease, smoking habits,* and *BMI level.*

5. Diabetes Risk Factor Assessment Using Machine Learning Techniques

To assess diabetes risk factors, we applied machine learning techniques to perform model training on the dataset. The process involved feeding in different sets of risk factors as input, followed by a predictive performance evaluation and a class separability evaluation. The details of these steps are as follows:

5.1. Model Training

We used the pre-processed dataset to train our predictive models with various machine learning algorithms. The input for each model included one of the selected risk factors sets (explanatory variables) and the risk levels determined by the traditional assessment (explained variables). The dataset was randomly split into a training set and a test set using a 70:30 ratio, as shown in Table 3. For model tuning, we applied Grid Search combined with 6-fold cross-validation to find the most optimal hyperparameter values for each learning algorithm, ensuring we created the best-performing models possible.

5.2. Model Prediction Performance Testing

We conducted performance testing of the prediction model using a test set. The model's performance was measured using the following performance metrics:

- Accuracy: The ratio of correctly predicted samples to the total number of samples.
- Precision: The ratio of correctly predicted positive samples (True Positive Samples) to all samples predicted as positive (True and False Positive Samples).
- Recall: The ratio of correctly predicted positive samples (True Positive Samples) to all actual positive samples (Positive Samples).
- F1-score: The Harmonic mean of Precision and Recall.

The model testing results are shown in the graph in Figure 1. From the graph, it can be observed that inputting data using risk factor set A and risk factor set C tends to result in models with the highest overall performance, except in the case of the model built using the Random Forest algorithm, which yields the highest overall performance when data is input using risk factor set B. We also calculated the average values of the performance metrics and ranked the risk factor sets based on their averages, with the results presented in Table 5.

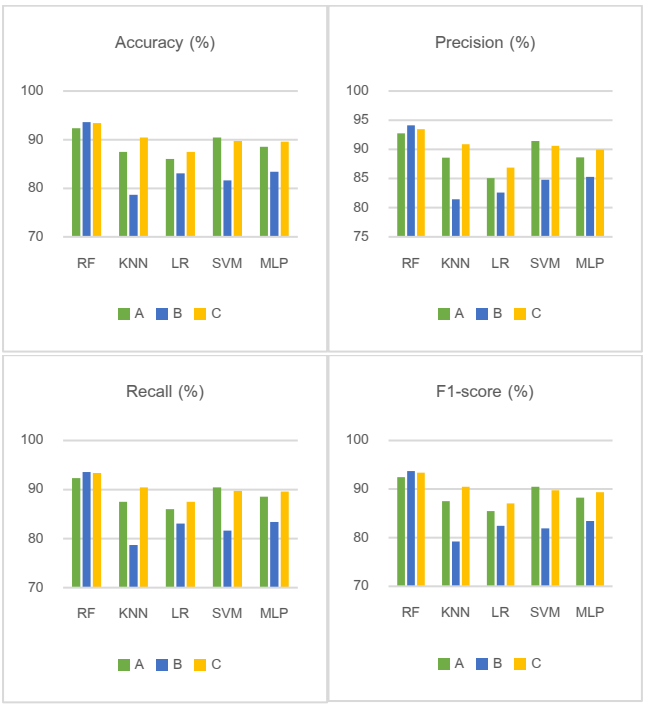


Figure 1. Performance values of the models in testing.

Table 5. Average performance values of the models in testing.

No.	Risk factor set	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1	C (compare with C)	90.12 (0.00)	90.36 (0.00)	90.12 (0.00)	90.01 (0.00)
2	A (compare with C)	88.97 (-1.15)	89.28 (-1.07)	88.97 (-1.15)	88.84 (-1.17)
3	B (compare with C)	84.07 (-6.04)	85.64 (-4.72)	70.06 (-20.06)	84.14 (-5.87)

From Table 5, it can be seen that importing data using risk factor set C resulted in the highest average overall efficiency, while importing data using risk factor set A resulted in the second-highest average

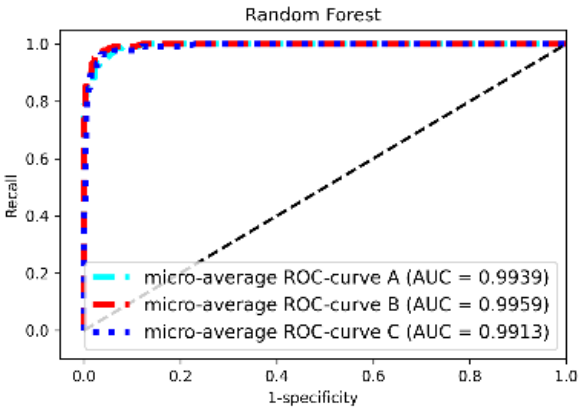
overall efficiency. However, it was found that the average performance values obtained from using risk factor sets A and C are very close to each other.

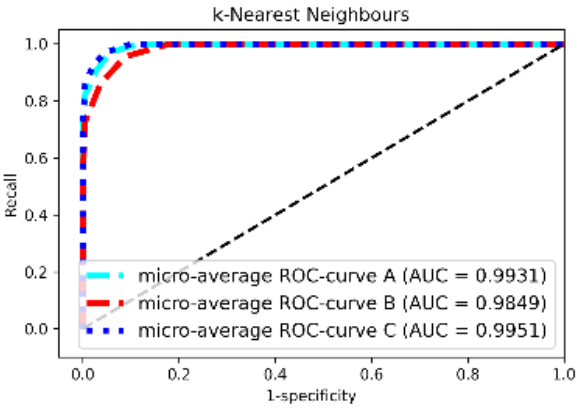
5.3. Risk Level Discrimination Capability Test

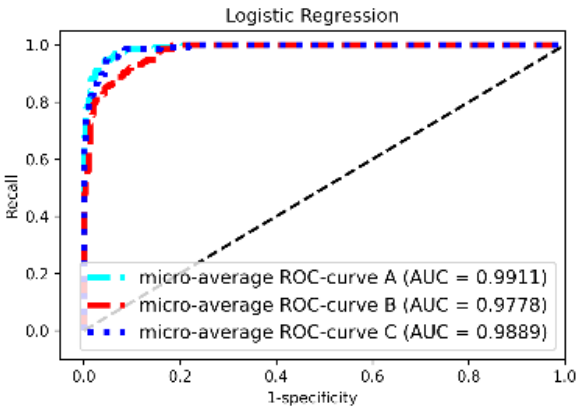
In testing the model's ability to discriminate between risk levels, we conducted experiments using the test set and measured additional performance metrics as follows:

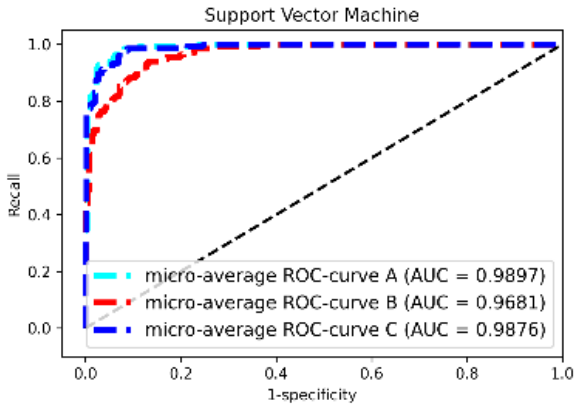
- Specificity: This refers to the proportion of correctly predicted negative samples (True Negative Samples) to the total number of actual negative samples.
- Receiver Operating Characteristic (ROC) Curve: A graph that plots the Recall (on the x-axis) against 1 - Specificity (on the y-axis), obtained by applying different classification thresholds to the model.
- Area Under the ROC Curve (AUC): This is the area calculated under the ROC curve. An AUC value closer to 1.0 indicates that the model has good discriminative ability, while a value of 0.5 indicates the worst performance in discriminating between classes.

Figure 2 shows the ROC curve plotted using the Micro Average ROC values and the AUC values obtained from all models. From the graph, it can be observed that using risk factor set B tends to result in the model having the lowest ability to discriminate between risk levels. In contrast, using risk factor sets A or C tends to result in models with relatively similar and higher discrimination capabilities.









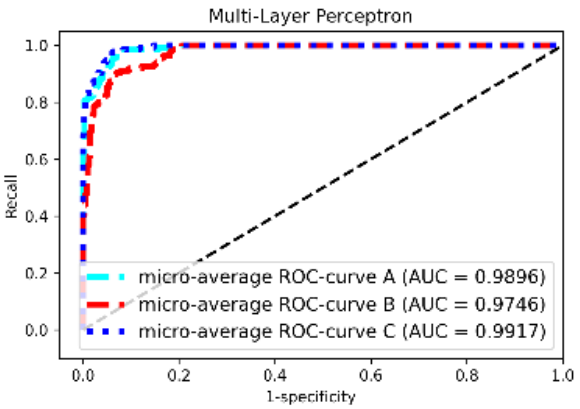


Figure 2. ROC Curve and AUC value obtained from the Model during Testing.

6. Discussion on the Evaluation of Diabetes Risk Factors

From the results of the model's predictive performance and its ability to discriminate between different risk levels, we found that using risk factor sets A and C yielded similarly high performance in the experiments. However, when comparing the two sets, it was observed that risk factor set C is a subset of risk factor set A. Moreover, the comparable performance measured in the experiments provides empirical evidence that risk factor set C can effectively a good substitute for risk factor set A. In other words, this suggests that some of the risk factors in set A may not be necessary for predicting diabetes risk in the sample population of the dataset used.

Furthermore, when comparing risk factor set C with the traditional set of diabetes risk factors used in conventional assessments [2], we found that set C includes four factors not present in the traditional set: *blood pressure level, family history of hypertension, family history of cardiovascular disease,* and *smoking habits*. Conversely, the traditional set includes *age* as a factor, which is not present in risk factor set C. Considering the predictive performance obtained from risk factor set C against the expected performance from the traditional assessment set (which theoretically should yield the highest performance), we found that using risk factor set C resulted in an error of no more than 10%. Additionally, based on the results of the model's ability to discriminate risk levels, it can be concluded that using risk factor set C allows the model to distinguish risk levels at a level comparable to the traditional assessment (compared to a maximum AUC value of 1.0).

7. Conclusion

This research proposes a method for selecting and evaluating risk factors for predicting diabetes in a sample group using data analysis techniques and machine learning with physiological data. The results from ANOVA revealed statistically significant relationships between the factors and the levels of risk assessed using traditional methods. This led to the selection of appropriate risk factors for predicting diabetes in the sample group. The results of diabetes risk factor evaluation using machine learning techniques indicated that risk factor set **C** is highly suitable for accurately predicting diabetes in the dataset's sample group.

Acknowledgements

This research is a collaborative project under the Center of Artificial Intelligence Technology and Innovation for Health, Department of Computer Science, Faculty of Science, Kasetsart University. For this research, we used a physiological dataset from the Royal Dunyapabbumbad Foundation. The use of this dataset is governed by [6].

References

[1] Wichai Aekplakorn, Hataichanok Puckcharern, and Waraporne Satheannopphakao, "The 6th Thailand National Health Examination Survey 2019–2020," Department of Community Medicine, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand, 2021.

[2] Vichai Aekplakorn, Pongamorn Bunnag, Piyamitr Sritara, Sayan Cheepudomvit, Sukit Yamwong, and Rajata Rajatanavin," Diabetes Risk Score for Thai Population," Health Systems Research Journal, vol. 1, no. 3-4 pp. 262-267, Oct.-Dec. 2007.

[3] Lara Lama, Oskar Wilhelmsson, Erik Norlander, Lars Gustafsson, Anton Lager, Per Tynelius, Lars Wärvik, and Clase-Göran Östenson. "Machine learning for prediction of diabetes risk in middle-aged Swedish people," Heliyon, vol 7, no. 7, Jul. 2021.

[4] Liying Zhang, Yikang Wang, Miaomiao Chongjian Wang, and Zhenfei Wang. "Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: the Henan Rural Cohort Study." Scientific Reports, vol 10, Mar. 2020.

[5] Nikos Fazakis, Otilia Kocsis, Elias Dritsas, Sotiris Alexiou, Nikos Fakotakis, and Konstantinos Moustakas, "Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction," IEEE Access, vol. 9, pp. 103737-103757, 2021.

[6] Dangjai Souvannakitti, "Project to Develop a Holistic Individualized Healthcare Platform," Dunyapabbumbad Foundation for Aging and Health under the Royal Patronage of Her Royal Highness Princess Bejaratana Rajasuda Sirisobhabannavadi, Thailand, 2020.

[7] Nuanwan Soonthornphisaj, "Machine Learning," Department of Computer Science, Faculty of Science, Kasetsart University, Bangkok, Thailand, pp. 1-228, Aug. 2020.