Prediction of Glycated Haemoglobin Based on Routine Blood Count Tests to Support the Diagnosis of Diabetes Mellitus

Glauco Cardozo (glaucocardozo@gmail.com)
Instituto Federal de Educacao Ciencia e Tecnologia de Santa Catarina https://orcid.org/0000-0002-3694-6047

Guilherme Rettore Andreis
Institute of Biomedical Engineering, Federal University of Santa Catarina

Sandra Cossul
Institute of Biomedical Engineering, Federal University of Santa Catarina

Annelise Correa Wengerkievicz Lopes
Santa Luzia Laboratory

Silvia Modesto Nassar
Department of Informatics and Statistics, Federal University of Santa Catarina

Jefferson Luis Brum Marques
Institute of Biomedical Engineering, Federal University of Santa Catarina

Research

Keywords: machine learning, laboratory tests, diabetes mellitus, diagnosis

Posted Date: July 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-38916/v1

Abstract

Background

Currently 8.8% of the World's population aged from 20 to 79 have diabetes mellitus (DM); of this total is estimated that 50% have not been diagnosed and do not know they have the disease. The most common laboratory tests used for diagnosis include blood glucose (FPG) and glycated haemoglobin (HbA_1c). The HbA_1c test has advantages over FPG, therefore being recommended in diagnoses of DM. Early diagnoses are essential to prevent complications caused by DM; however, the symptoms of the initial stage are present in only 40% of the carriers, symptomless carriers oddly pursue the DM test. In a lifetime patients performs a series of laboratory exams for health analysis which is stored as laboratory data, the computational approach offers enormous potential in health data analysis discovering relevant results overlooked by physicians. Use machine learn approach on data stored of routine blood count laboratory tests to predict HbA_1c diagnosis.

Method:

Using laboratory results from data stored of HbA_1c, was formed six data groups composed of individuals: healthy and pre-diabetic (HP); healthy and diabetic (HD); pre-diabetic and diabetic (PD); healthy and non-healthy (HN); non-diabetic and diabetic (ND); and healthy, pre-diabetic and diabetic (HPD) patients. For each data group, was tested the K nearest neighbours (KNN), support vector machines (SVM), random forests (RF), naive Bayes (NB) and artificial neural network (ANN) models. Assessment of model performance was carried out using sensitivity, specificity, precision and negative prediction.

Results

The KNN model applied to the ND group had the best performance in the diagnosis of diabetes, resulting in a sensitivity value of 53.6% and an accuracy of 90.1%. The classification after regression with the neural network model (ANNr) and the ND group had a more general result, with a sensitivity of 74.3% and an accuracy of 77.2%. Analysing only the values for the regression, the neural network model presented a mean square error of 0.36 for the final test base with a correlation of 0.85.

Conclusions

We conclude that machine learning-based computational models can predict HbA_1c values from other routine laboratory tests. Thus, they can assist in the detection of diabetes and act as a warning for undiagnosed cases.

Background

The carbohydrates we eat break down into glucose, which is one of the primary sources of energy used by cells. Produced by the pancreas, insulin is a hormone that acts as a kind of key, allowing blood glucose to be carried into cells and produce energy. Diabetes mellitus (DM) is a chronic metabolic disorder caused by a deficiency in insulin production or an inability of the body's cells to make use of it properly. Over time, this causes an increase in blood glucose levels known as hyperglycemia, which can cause numerous health complications.[1].

There are three main classifications of DM: type 1, type 2, and gestational. Type 1 diabetes can occur at any age, being more frequent in children and adolescents and corresponding to 10–20% of cases. It is characterised by little or no insulin production due to the destruction of pancreatic β cells and requires daily insulin injections to keep glucose levels under control [1].

Type 2 diabetes, on the other hand, accounts for more than 90% of cases, usually occurring in older individuals (over 40 years of age), although it can also occur in young people and children [2]. Its main characteristic is tissue resistance to insulin action, the causes of which have yet to be fully clarified, but it is strongly related to behavioural factors, such as eating habits, physical inactivity, and obesity [1]. The diagnosis relies on laboratory tests or the appearance of chronic complications when the disease is already advanced [1].

According to the 8th edition of *IDF Diabetes Atlas* [1], approximately 8.8% of the World's population aged 20 to 79, or 425 million people, have DM. By 2045 it is predicted to reach a total of 693 million people aged between 18 and 99 years. Of this total, about 50% have not been diagnosed and do not know they have the disease, causing delayed treatment and increasing health costs [1]. Worldwide, DM costs account for about 1/8 of total health spending; it is also among the diseases with the highest death tolls, accounting for more than 10% of deaths worldwide [1].

Thus, early diagnosis is essential to avoid further complications and reduce treatment costs. However, early diagnosis does not frequently occur, as almost half of those affected by the disease are unaware of the disease [1].

Diabetes is diagnosed via analysis of laboratory tests, such as fasting plasma glucose (FPG) or plasma glucose (2 h-PG) 2 h after ingestion of 75 g of glucose (OGTT), in addition to glycated haemoglobin (HbA_1c). The accuracy of all these tests can be used to diagnose diabetes mellitus but may differ. [3]. Studies show that, compared to the cut-off points for FPG and HbA_1c, the two hour PG value diagnoses more people with diabetes [3]. However, HbA_1c has advantages, such as international standardisation of assays, lower biological variability, being unaffected by acute stress, and no need for fasting, among others [4]. Thus, given the characteristics of the methods presented, HbA_1c has been increasingly indicated for screening and diagnosis of diabetes [5].

For the fasting blood glucose test (FPG), patients with glucose levels below 100 mg/dL are considered healthy. Patients with glucose levels between 101 and 125 mg/dL are pre-diabetic, and patients with glucose equal to or above 126 mg/dL are considered diabetic. However, this test requires the patient to fast for at least eight hours [3]. If glucose presents values above 200 mg/dL (even without glucose intake or fasting) and the patient presents symptoms, that patient is considered diabetic [3].

For the 2 h glucose test, after 75 g of glucose intake, patients results are considered healthy if their glucose level is below 140 mg/dL, considered pre-diabetes if their glucose is between 140 and 199 mg/dL and considered diabetes if the glucose is greater than or equal to 200 mg/dL [3]. When using HbA_1c, patients are classified as healthy if HbA_1c is below 5.7%, pre-diabetic if HbA_1c is between 5.7% and 6.4%, and diabetic if it is equal or greater than 6.5% [3].

Considering that approximately 60% of patients present no symptoms in the initial phase of the disease [6], patients must perform some of these tests to detect DM, but most people without symptoms do not pursue these tests. In observing the frequency of tests performed by a laboratory in Florianópolis, Santa Catarina, Brazil in 2017, the blood count with analytes was found to be the most performed exam. This year, FPG was the third most performed test, with HbA_1c occupying the 52nd position. Measuring glucose 2 h after ingestion of 75 g of glucose was in the 409th position.

The characteristics presented hinder early diagnoses of diseases such as type 2 diabetes. Patients often have several exams throughout their lives that may be useful in analysing their health; however, physicians may overlook relevant results or fail to notice patterns in the laboratory dataset, because valuable information related to a diagnosis may be too subtle and more difficult to be identified by a human without adequate computational support [7].

To interpret these results correctly, clinicians must evaluate many tests and interpret them, along with other clinical data, while considering patient history. Although this manual approach to exam interpretation is standard in most cases, computational approaches to laboratory data integration and analysis offer great potential in the search for diagnoses [8].

Clinical laboratories present most test results as individual numerical values. However, the results of these tests viewed in isolation usually have limited usefulness in obtaining a diagnosis. Luo [8], in his ferritin study, found that laboratory tests often include redundant information. Thus, through machine learning-based models, it was possible to predict the results of ferritin laboratory tests from the result sets of other laboratory tests from each patient, providing additional information to refine the diagnosis.

In the same study, Luo [8] found that when measuring ferritin in laboratory tests, they found a high false-negative rate when compared to the computational model. This illustrates that with access to large databases, intelligent systems can improve the interpretation of laboratory test results.

Similarly, Gunčar [9] found that machine learning models can be used to predict hematological diseases using blood tests only. In the study, Gunčar says that laboratory tests have more information than that

commonly considered by health professionals.

With significant evolution in recent years [10], machine learning methods are powerful tools in supporting medical diagnoses. Studies [11, 9, 12] have shown that these methods are capable of predicting and identifying diseases based on laboratory tests and clinical data with similar accuracy to a human specialist. Other studies [13, 14, 15] have also been able to assist in the diagnosis of diabetes by making use of machine learning techniques.

Given the facts presented, this study is intended to make use of a database of laboratory tests to predict possible diseases in individual patients. The main goal is to try to predict or assist in the diagnosis of diabetes mellitus through routine examinations and machine learning techniques.

Results

The results for each classification model were compiled for each data group to compare the performace of models. In Fig. 1A, are plotted sensitivity and specificity values for the HP group. This graph represents the main diagonal of the confusion matrix. In Fig. 1B, precision (positive predictive values) and negative predictive values for the same group, whose values come from the second diagonal of the confusion matrix.

To compare the performance of the models over the HP group, the metrics used were F1 Score and Accuracy, as shown in the plot in Fig. 1C.

In general, models for this group presented low sensitivity, indicating the models' reduced ability to distinguish healthy from pre-diabetic individuals.

In Fig. 2A, are plot sensitivity and specificity values for the HD group. In Fig. 2B, precision (positive predictive values) and negative predictive values.

The results obtained with the HD group showed better performance, especially regarding sensitivity, which was expected because the correct classification of pre-diabetic individuals represents the most significant difficulty.

Models trained with the HD group are a good option for a previous classification of healthy and diabetic individuals. However, pre-diabetic individuals will be misclassified into one of the two classes.

Among the trained models, SVM had the best performance, with F1-Score equals 86.6%.

In Fig. 3A, are plotted sensitivity and specificity values for the PD group, and in Fig. 3B precision and NPV.

In contrast to the results obtained in the HP group (Fig. 1A, Fig. 1B and Fig. 1C), the results from the PD group showed higher sensitivity and lower specificity, except for the KNN model. Precision yielded similar results. However, the models had proportionally better results than in the HP group. Figure 3C shows that both F1-score, as well as accuracy, presented similar results between the different models.

In Fig. 4A, are plotted sensitivity and specificity values for the HN group. In Fig. 4B, precision values and negative predictive values for the same group. In Fig. 4A and Fig. 4B, in addition to the same models plotted in the previous groups, is also added the classification performed after regression with the neural network model, identified here as ANNr.

In this group, the KNN model had the highest specificity. However, the sensitivity was worst in comparison to the sensitivity of the other models. The ANNr model (classification after regression) presented the best cost-benefit.

Analysing the F1-Score and Accuracy values between the different models (Fig. 4C), ANN had the best result, as opposed to the KNN that had the worst, although it had the highest precision.

In Fig. 5A, are plotted sensitivity and specificity values for the ND group. In Fig. 5B, precision values and negative predictive values for the same group, also adding the classification performed after regression with the ANNr model.

Analysing the results, we observed that the KNN model presented higher specificity and precision, although the sensitivity was the lowest among the models.

In Fig. 5C are plotted the F1-Score and Accuracy values for the different models on the ND group. Similar to the HN group, here, the KNN model had the worst result, although it also had the highest precision.

When we analyze the values in the confusion matrix (Fig. 6), we can see that, although this model identifies only 53.0% of the total of real diabetics, it has a specificity of 98.9% and consequently an accuracy of 89.7%. This result gives us more confidence in the expected positive results, since only 11.3% of those classified by the model as diabetic will be false positives.

In Fig. 7, is plotted the hit rate for each model for the HPD group.

The analysis of Fig. 7 reveals that the main problem of the models is to classify the individuals with pre-diabetic correctly. However, this was expected, given the diffuse feature in the classification this categories.

For this group of data, the Random Forest had a more general result, with the best performance in the classification of pre-diabetic individuals.

In Fig. 8, is plotted the confusion matrix of the classification model using random forest (RF), which achieved an hit rate of approximately 80% for HD patients. The problem relates to the classification of pre-diabetic patients, which as with other models, was in most cases around 50%.

Given the different data groups and models tested, the classification after regression with the ANNr for the ND group was particularly useful. In Fig. 9, is presented the confusion matrix with the results obtained with this model. The model achieved a sensitivity of 70.7%, a specificity of 96.4%, an precision of 80.3%, and an NPV of 94.1%.

Analysing only the values for the regression, the ANNr model presented a mean square error of 0.36 for the final test base.

Discussion

Among different tests to diagnose diabetes, there are advantages and disadvantages to every method.

Analyzing the results for the various classification models and the different data groups (HP, HD, PD, HN, ND and HPD), the performance of the models stands out according to the specific characteristics of each group. Thus, each model and group will be better indicated depending on the search. Thus, each model and group will be better indicated depending on the search.

The ideal is to have models with high specificity and high precision, which means fewer false positives. Similarly, as the sensitivity increases, the model achieves more correct ratings. However, this is not as important as precision, because, even if the model does not rank many results correctly, the ones identified as positive will be mostly correct.

In this case, both the KNN model and the ANNr model, both trained with the ND group, could be used to identify false negatives for the FPG tests. In the case of the KNN model, we have a sensitivity of 53% but an accuracy of 89.7%. The ANNr model was more general, with a sensitivity of 70.7%, but with an accuracy of 80.3%

Analysing the models with the HPD group all models (except KNN), presented a similar performance in the classification of healthy and diabetic patients, with the ANNr model presenting the best overall result. In the case of pre-diabetic patients, all models had difficulty in the classification (Fig. 7).

In the case of regression, the neural network model (ANNr) was also very satisfactory in predicting HbA1c values with an average square error [16] of 0.36 and a correlation of 0.85 on the test data set.

Conclusion

Observing the fact that the models have more difficulty with the classification of the HP group in relation to the DP group, we are induced to think that pre-diabetic individuals are more similar to healthy individuals than to diabetics. The same is confirmed when we observe the better performance of the models with the ND group compared to the HN group.

In general, we conclude that machine learning-based computational models can predict HbA1c values as from other laboratory tests. These findings for diagnoses of DM without the use of HbA1c exam implies a series of advantages for the health care system and the patient, as the main advantage being the early detection of the disease, which can be overlooked by the lack of symptoms. Likewise, such models can help detect false negatives on the FPG test and identify diabetic individuals, being used as an alert for undiagnosed cases.

Methodology

In this study, were used a database of laboratory tests performed by the Santa Luzia laboratory in Florianópolis, SC, Brazil throughout 2017. The study was approved by the ethics committee of the Federal University of Santa Catarina under registration number 02203918.0.0000.0121. All simulations were performed in Python, using the Jupyter [17] environment and the scikit-learn libraries [18].

Initially, from the database was selected the tests most frequently performed in comparison to HbA1c, we removed tests with non-quantitative values, tests with uneven distribution, and samples with missing data.

Following the methodology of the 8th edition of the IDF Diabetes Atlas [1], was selected patients aged between 19 and 99 years old, outliers referring to the other entries were kept because they are directly related to the pathologies.

The factor analysis technique was applied to select the most relevant parameters concerning the prediction of HbA1c. In this process, input variables are tested in order to obtain the best result and evaluate the impact that each one has on the output variable. Input variables are grouped according to contribution to the model, with an influencing factor assigned to each group and resource within groups [19, 16].

Pre-processing resulted in a base with 14 main parameters (Table 1) and 57,710 samples. According to Hb1Ac classification, the database was unbalanced, with 60% of the samples classified as healthy individuals, 25% with pre-diabetes and 15% with diabetes.

Table 1
Input parameters selected through factor analysis

Factor	Factor influence (%)	Most influential parameter	Parameter influence (%)
MR2	0.12	Absolute haematocrit count (Hct)	0.98
MR1	0.11	White blood cells - leukocytes (WBC)	0.96
MR3	0.10	Percent lymphocytes (Lymph%)	0.97
MR7	0.10	Creatinine	0.95
MR5	0.08	Mean corpuscular volume (MCV)	0.98
MR6	0.07	Fasting plasma glucose (FPG)	0.98
MR15	0.07	Percent basophils (Baso%)	0.92
MR9	0.06	Percent eosinophils (Eos%)	0.99
MR8	0.06	Percent monocytes (Mono%)	0.95
MR10	0.06	Mean corpuscular haemoglobin concentration (MCHC)	0.98
MR12	0.05	Triglyceride	0.97
MR11	0.04	Absolute platelet count (Plt)	0.94
MR13	0.04	HDL cholesterol	0.71
MR14	0.03	Ages	0.85

The pre-processed data set was normalised with a mean of 0 and a standard deviation of 1. The normalised data set was randomly divided into three parts: training, validation, and testing. First, 20% of the total data set was separated for the final test. The remaining 80% were divided by 70% for training and 30% for validation.

We tried different grouping strategies to explore the data fully creating different groups of data. The objective was to compare the classification performance of the models before these different groups. Based on the HbA1c exams patients are classified as healthy if their HbA_{1c} is below 5.7%, pre-diabetic if their HbA_{1c} is between 5.7% and 6.4%, and diabetic if it is equal to or greater than 6.5%, for different data evaluation six groups of data were created, consisting of:

🟢🟡 **Healthy and pre-diabetic (HP) patients**. Removing diabetic individuals

🟢🔴 **Healthy and diabetic (HD) patients**. Removing pre-diabetic individuals

🟡🔴 **Prediabetic and diabetic (PD) patients**. Removing healthy individuals

🟢🟡🔴 **Healthy and non-healthy (HN) patients**. In this group, the 'non-healthy' category consisted of pre-diabetic and diabetic patients.

🟢🔴 **Non-diabetic and diabetic (ND) patients**. In this group, the 'non-diabetic' category consisted of healthy and pre-diabetic patients.

🟢🟡🔴 **Healthy, pre-diabetic, and diabetic patients (HPD).**

The groups were trained with the following classification models:

- **K nearest neighbours (KNN)**
- **Support vector machines (SVM)**
- **Random forests (RF)**
- **Naive Bayes (NB)**
- **Artificial neural network (ANN)**

And a regressor model:

- **Artificial neural network (ANNr).**

As metrics for classification model evaluation was measured accuracy, sensitivity (Recall), specificity, precision (positive predictive value - PPV), and negative predictive value (NPV), to perform exploratory analysis. To evaluate the results of the final test dataset was used the F1 score, which gives a real representation of the results in unbalanced data.

All models were trained with the six different groups of data created. For each model, several values of hyperparameters were tested and adjusted, always to improve the results and reduce overfitting between the training and validation base. Finally, the test basis was used to assess the model's performance.

The neural network used in the two models was of the multilayer perceptron type, using the Keras library [20]. In the regression model, after data prediction, the outputs also were classified according to groups:

HN, ND, and HPD. The metrics used were the same as those used for the other classifiers, besides the mean square error and the scatter plot for visualisation.

Abbreviations

DM
Diabetes mellitus
FPG
Fasting plasma glucose
PG
Plasma glucose
HbA_1c
Glycated haemoglobin
HP
Healthy and pre-diabetic
HD
Healthy and diabetic
PD
Pre-diabetic and diabetic
HN
Healthy and non-healthy
ND
Non-diabetic and diabetic
HPD
Healthy, pre-diabetic and diabetic
KNN
K nearest neighbours
SVM
support vector machines
RF
Random forests
NB
Naive Bayes
ANN
Artificial neural network
ANNr
Artificial neural network with classification after regression
IDF
International Diabetes Federation
2 h-PG

Plasma glucose 2 h after ingestion of 75 g of glucose

OGTT

Glucose tolerance test

PPV

Positive predictive values

NPV

Negative predictive values

Declarations

Acknowledgements

Not applicable.

Author information

Affiliattions

Federal Institute of Santa Catarina, Florianópolis, SC, Brazil.

Glauco Cardozo

Institute of Biomedical Engineering, Federal University of Santa Catarina, Florianópolis, SC, Brazil.

Guilherme Rettore Andreis , Sandra Cossul and Jefferson L. B. Marques

Santa Luzia laboratory, Florianópolis, SC, Brazil.

Annelise Correa Wengerkievicz

Department of Informatics and Statistics, Federal University of Santa Catarina, Florianópolis, SC, Brazil.

Silvia Modesto Nassar

Contributions

ACW assisted in the interpretation and use of the database; JLBM and SN guided and supervised the study; SC reviewed and wrote the paper; GRA revised and drafted the work in addition to assisting in the methodology; GC carried out the study, developing the methodology, being the principal author of the work.

Corresponding author

Correspondence to Glauco Cardozo: glauco.cardozo@ifsc.edu.br

Availability of data and materials

The data used are the property of the laboratory and cannot be made available.

Consent for publication

Not applicable.

Competing interests

We declare that we have no competing interests.

Ethics approval and consent to participate

The study used a database of laboratory tests performed by the Santa Luzia laboratory, in Florianópolis-SC, throughout 2017. The study was approved by the ethics committee of the Federal University of Santa Catarina under registration number 02203918.0.0000.012.

Funding

No funding was received.

References

1. Federation ID, “IDF Atlas 9th edition and other resources,” *International Diabetes Federation. (*2019*). IDF Atlas 9th edition and other resources. Retrieved January 13, 2020, from* https://www.diabetesatlas.org/en/resources/, 2019. [Online]. Available: https://www.diabetesatlas.org/en/resources/?gclid=CjwKCAjwkPX0BRBKEiwA7THxiGGu4-2c0TG2qc5kNjW4IN3fZxM4g-QbZUX8CeHB_z8ccB_cet-KeBoC-6wQAvD_BwE. [Accessed: 20-Apr-2020].
2. Rao PV. Type 2 diabetes in children: Clinical aspects and risk factors. Indian J Endocrinol Metab. Apr. 2015;19(7):S47–50.
3. “AMERICAN DIABETES ASSOCIATION. STANDARDS OF MEDICAL CARE IN DIABETES-2017.”.
4. Nathan DM, et al., “International expert committee report on the role of the A1C assay in the diagnosis of diabetes,” *Diabetes Care*, vol. 32, no. 7. American Diabetes Association, pp. 1327–1334, 01-Jul-2009.
5. Sacks DB. “A1C versus glucose testing: A comparison,” *Diabetes Care*, vol. 34, no. 2. American Diabetes Association, pp. 518–523, Feb-2011.
6. Diabetes UK, “Prediabetes | Diabetes UK,” *Diabetes UK*, 2020. [Online]. Available: https://www.diabetes.org.uk/preventing-type-2-diabetes/prediabetes. [Accessed: 20-Apr-2020].
7. Baron JM, Dighe AS. “The role of informatics and decision support in utilization management,” *Clinica Chimica Acta*, vol. 427. pp. 196–201, 01-Jan-2014.
8. Luo Y, Szolovits P, Dighe AS, Baron JM. Using Machine Learning to Predict Laboratory Test Results. Am J Clin Pathol. Jun. 2016;145(6):778–88.

9. Gunčar G, et al. An application of machine learning to haematological diagnosis. ” Sci Rep. 2018;8(1):411.
10. Peek N, Combi C, Marin R, Bellazzi R. “Thirty years of artificial intelligence in medicine (AIME) conferences: A review of research themes,” *Artificial Intelligence in Medicine*, vol. 65, no. 1. Elsevier B.V., pp. 61–73, 01-Sep-2015.
11. Roy SK, Hom J, Mackey L, Shah N, Chen JH, “Predicting Low Information Laboratory Diagnostic Tests.,” *AMIA Jt. Summits Transl. Sci. proceedings. AMIA Jt. Summits Transl. Sci.*, vol. 2017, pp. 217–226, 2018.
12. Wong J, Murray Horwitz M, Zhou L, Toh S, “Using Machine Learning to Identify Health Outcomes from Electronic Health Record Data,” *Curr. Epidemiol. Reports*, vol. 5, no. 4, pp. 331–342, Dec. 2018.
13. Zheng T, et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inform. Jan. 2017;97:120–7.
14. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine Learning and Data Mining Methods in Diabetes Research. Comput Struct Biotechnol J. Jan. 2017;15:104–16.
15. Olivera AR, et al., “Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes - ELSA-Brasil: accuracy study,” *Sao Paulo Med. J.*, vol. 135, no. 3, pp. 234–246, Jun. 2017.
16. “factor-analyzer · PyPI.” [Online]. Available: https://pypi.org/project/factor-analyzer/. [Accessed: 18-Apr-2020].
17. “Project Jupyter. | Home.” [Online]. Available: https://jupyter.org/. [Accessed: 19-Jun-2020].
18. “1.6. Nearest Neighbors — scikit-learn 0.22.2 documentation.” [Online]. Available: https://scikit-learn.org/stable/modules/neighbors.html. [Accessed: 25-Apr-2020].
19. Goodfellow I, Bengio Y, Courville A, *Deep Learning*. MIT Press, 2016.
20. “Home - Keras Documentation.” [Online]. Available: https://keras.io/. [Accessed: 20-Oct-2019].

Figures

Figure 1A - Sensitivity and specificity of models for the HP group.

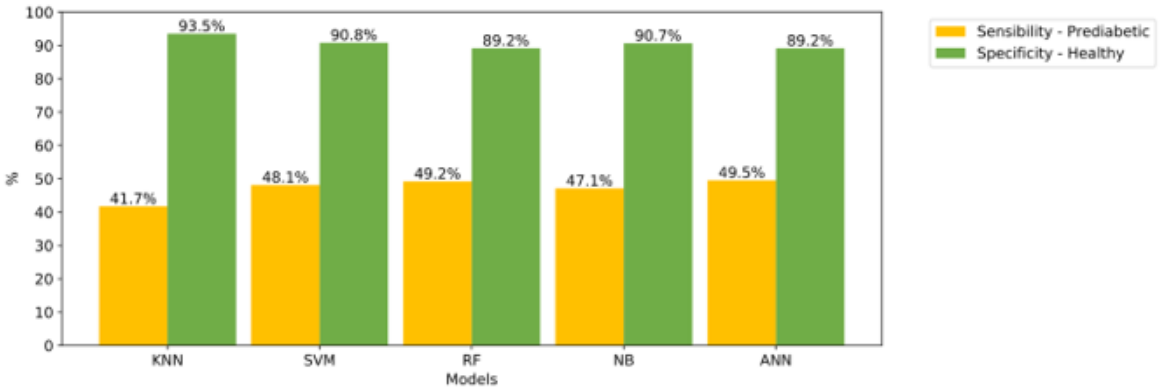

Figure 1B – Precision (PPV) and NPV of models for the HP group.

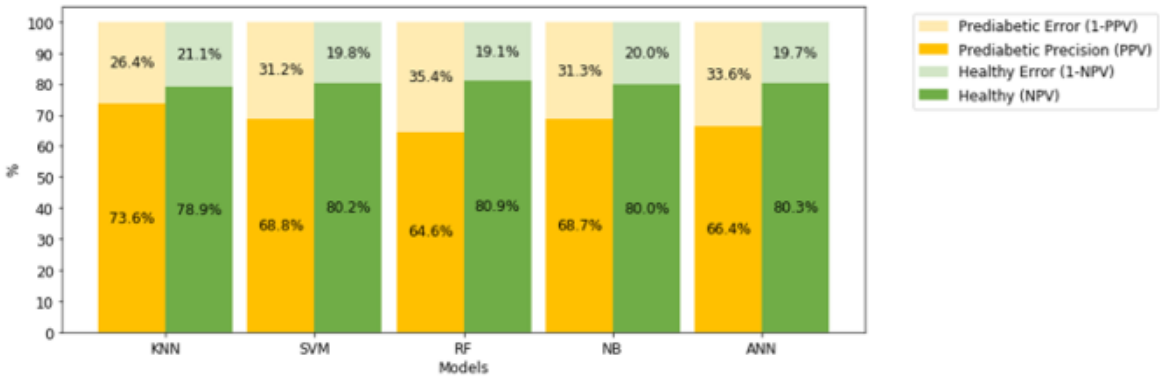

Figure 1C –F1-Score and Accuracy metrics of models for the HP group.

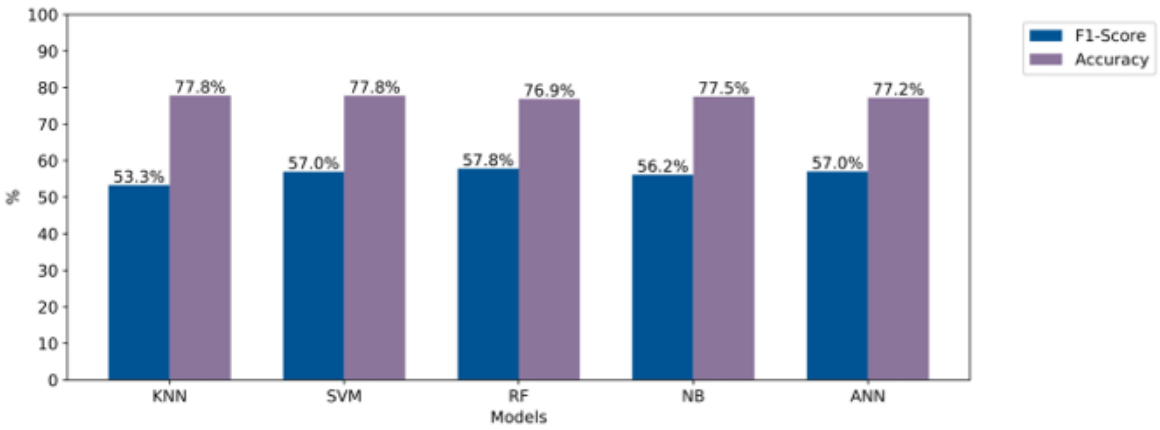

Figure 1

A - Sensitivity and specificity of models for the HP group. B – Precision (PPV) and NPV of models for the HP group. C –F1-Score and Accuracy metrics of models for the HP group.

Figure 2A - Sensitivity and specificity of models for the HD group.

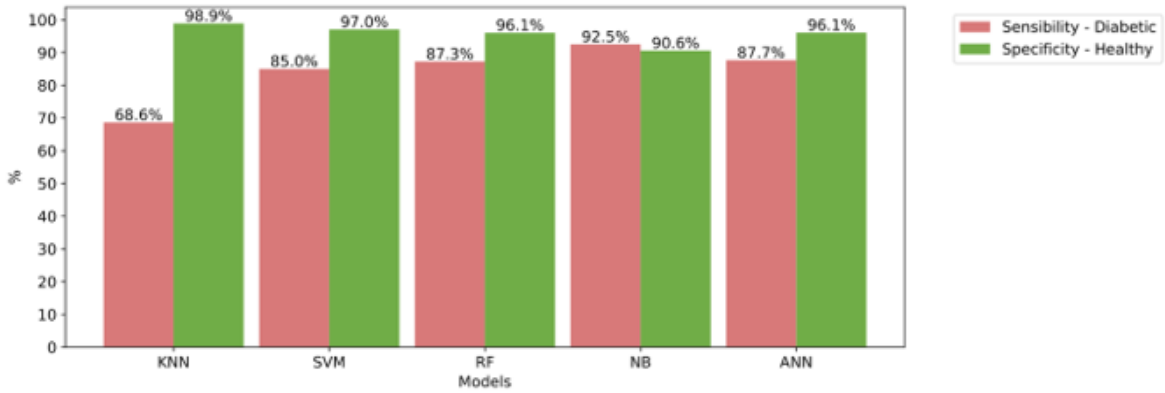

Figure 2B - Precision (PPV) and NPV of models for the HD group.

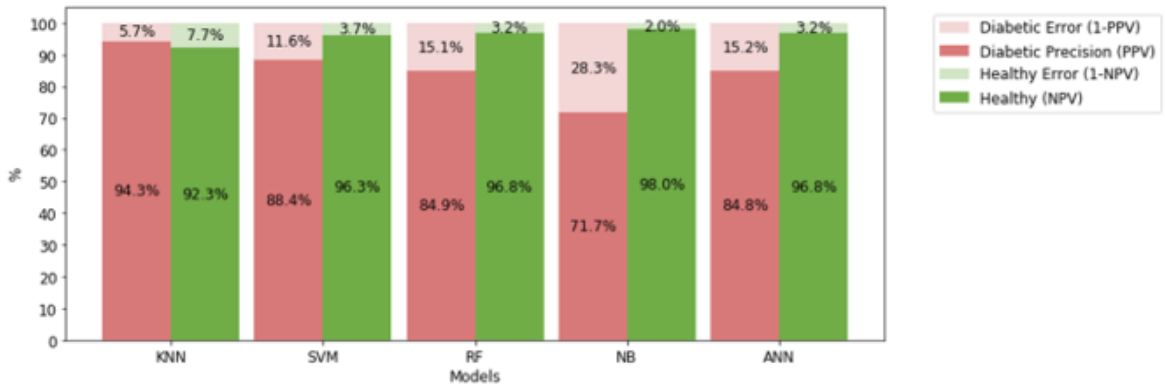

Figure 2C – F1-Score and Accuracy metrics of models for the HD group.

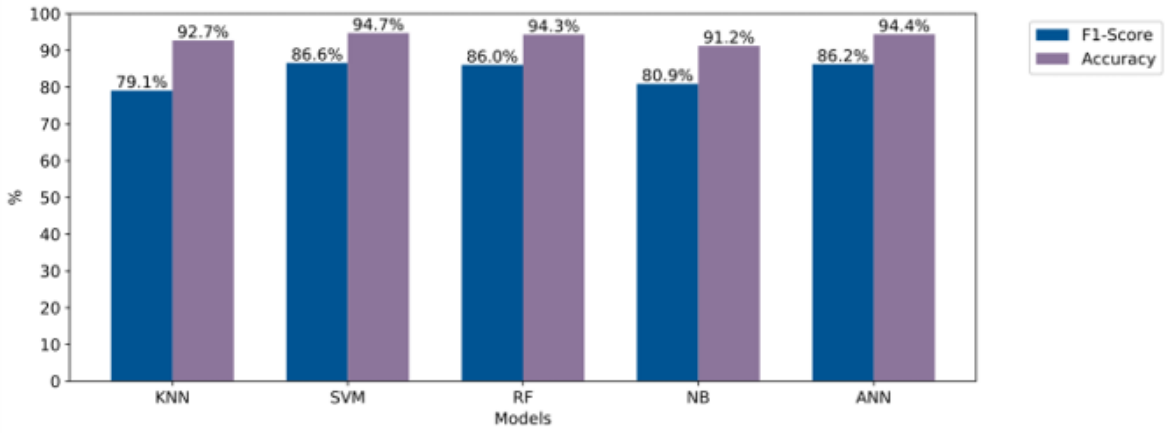

Figure 2

A - Sensitivity and specificity of models for the HD group. B - Precision (PPV) and NPV of models for the HD group. C – F1-Score and Accuracy metrics of models for the HD group.

Figure 3A - Sensitivity and specificity of models for the PD group.

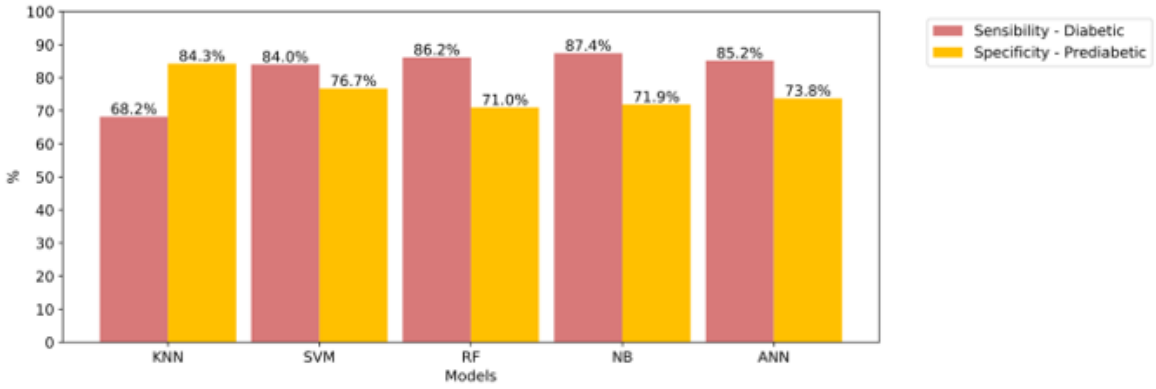

Figure 3B - Precision (PPV) and NPV of models for the PD group.

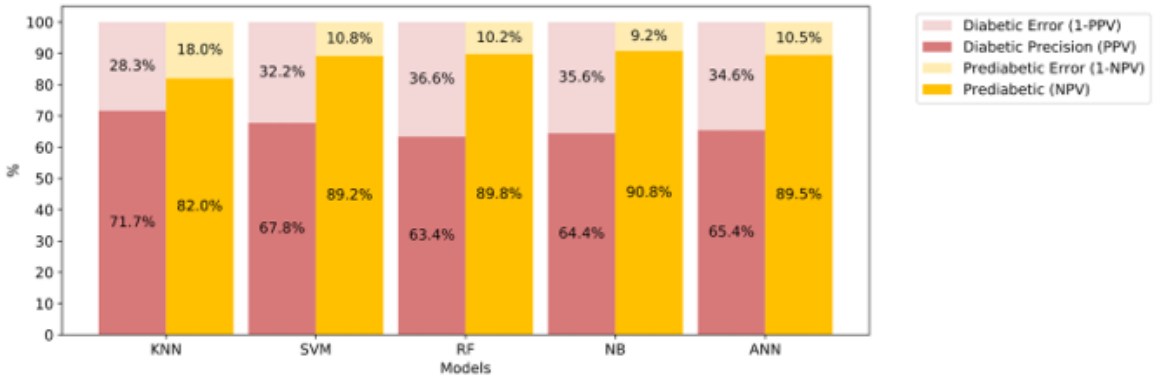

Figure 3C - F1-Score and Accuracy metrics of models for the PD group.

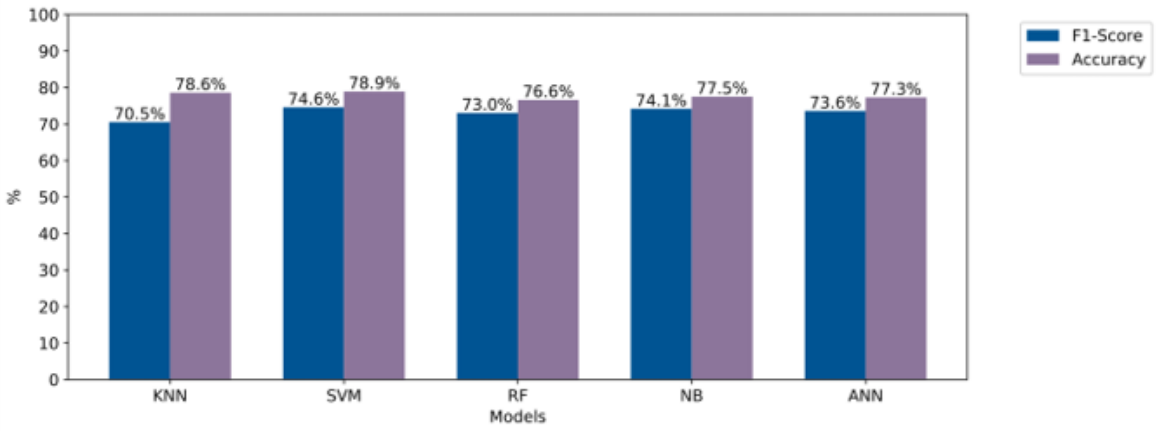

Figure 3

A - Sensitivity and specificity of models for the PD group. B - Precision (PPV) and NPV of models for the PD group. C - F1-Score and Accuracy metrics of models for the PD group.

Figure 4A - Sensitivity and specificity of models for the HN group.

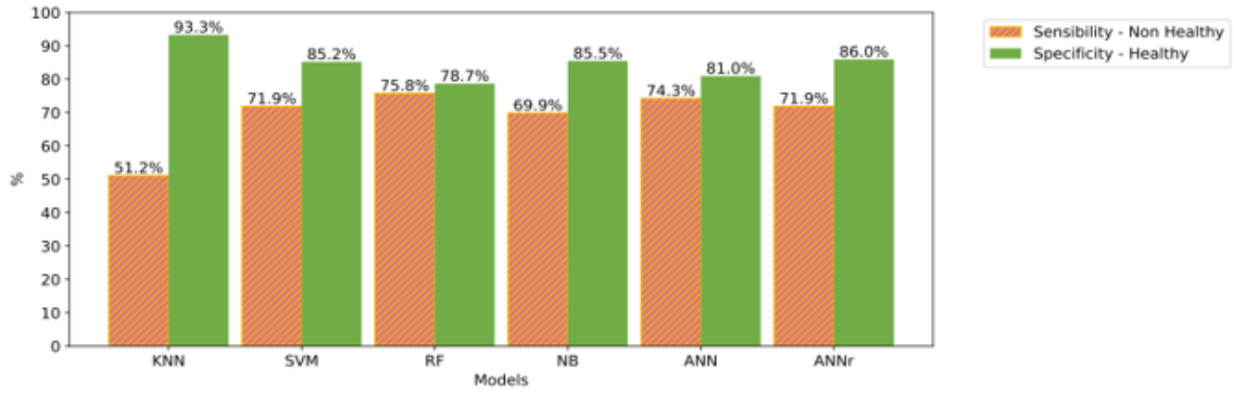

Figure 4B - Precision (PPV) and NPV of models for the HN group.

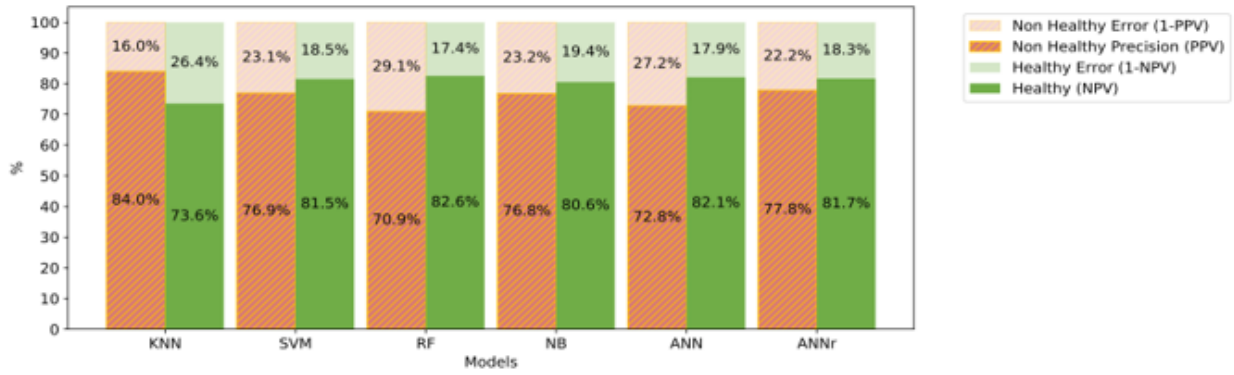

Figure 4C - F1-Score and Accuracy metrics of models for the HN group.

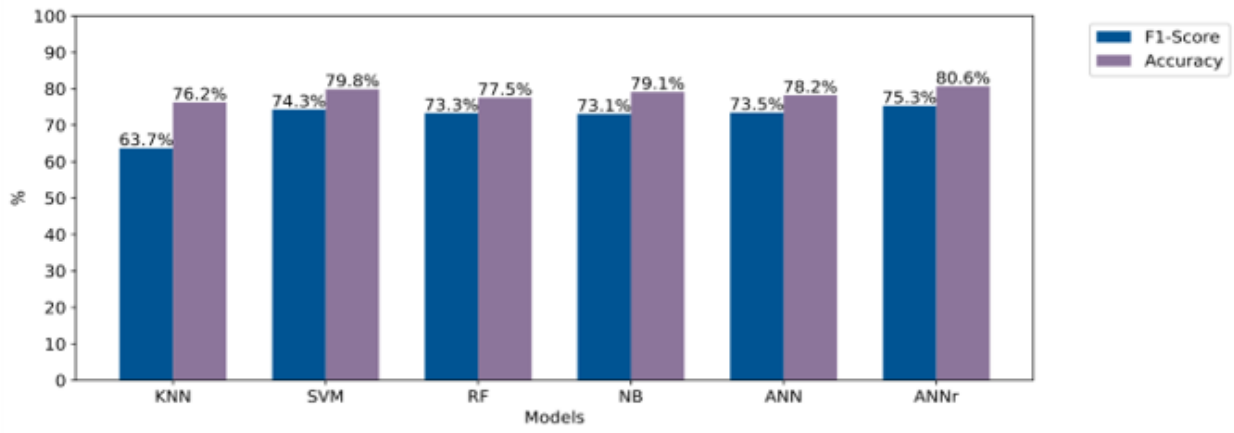

Figure 4

A - Sensitivity and specificity of models for the HN group. B - Precision (PPV) and NPV of models for the HN group. C - F1-Score and Accuracy metrics of models for the HN group.

Figure 5A - Sensitivity and specificity of models for the ND Group.

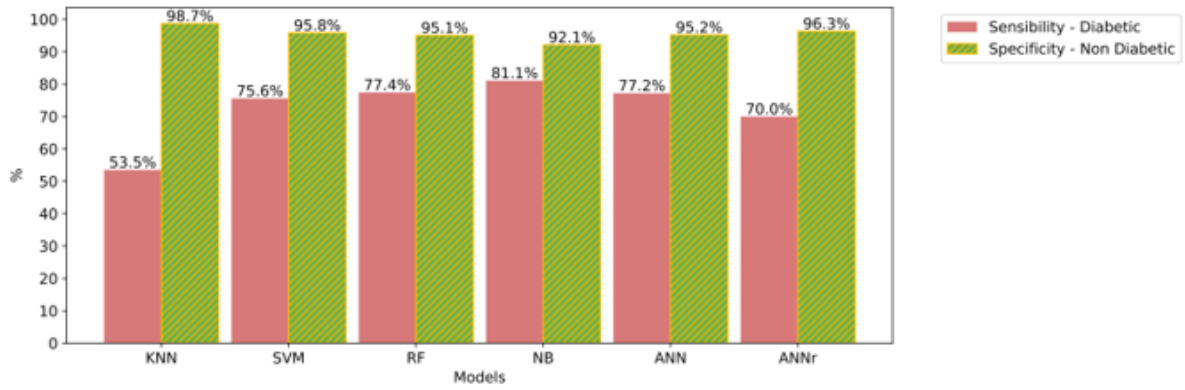

Figure 5B - Precision (PPV) and NPV of models for the ND group.

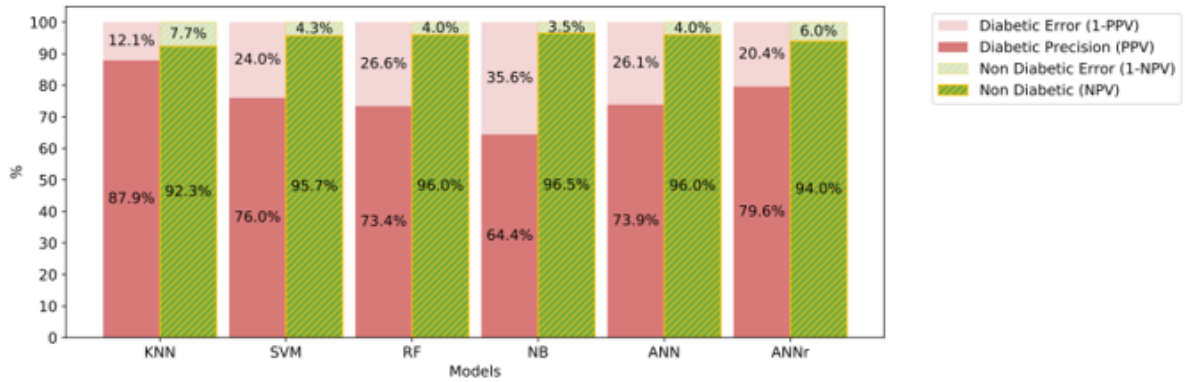

Figure 5C - F1-Score and Accuracy metrics of models for the ND group.

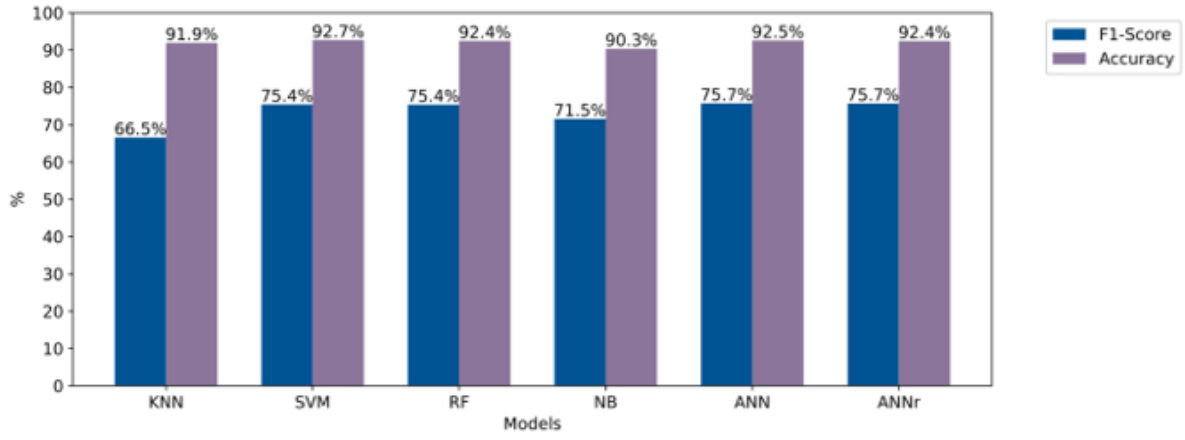

Figure 5

A - Sensitivity and specificity of models for the ND Group. B - Precision (PPV) and NPV of models for the ND group. C - F1-Score and Accuracy metrics of models for the ND group.

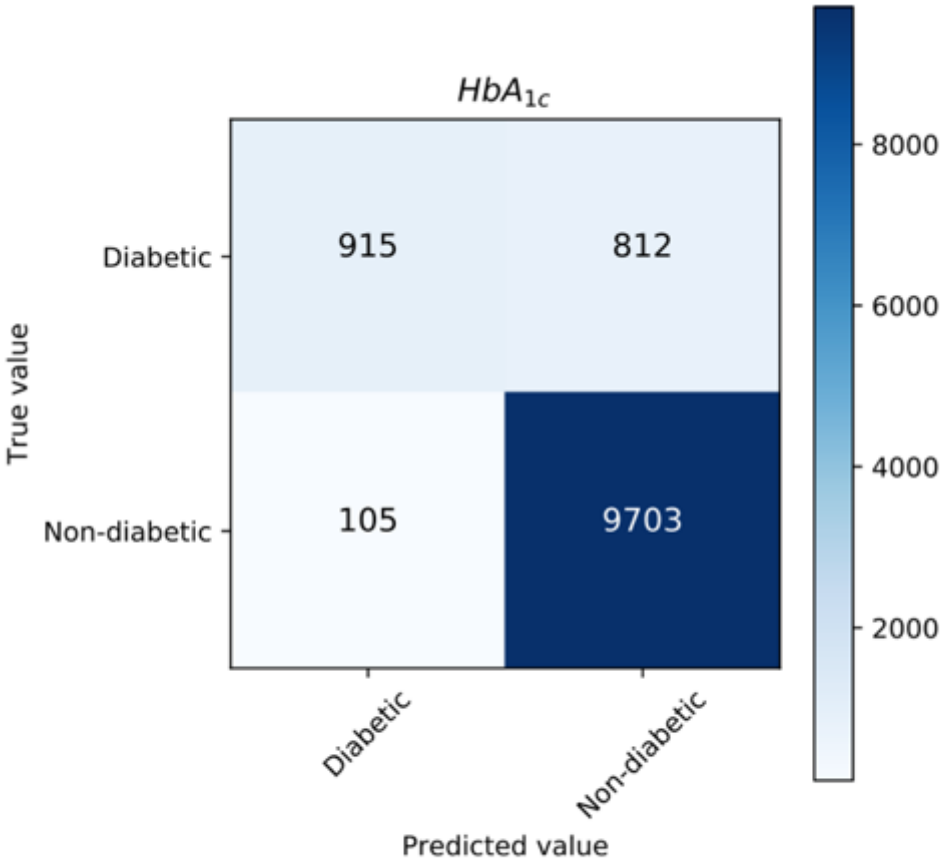

Figure 6

KNN model confusion matrix for the ND group.

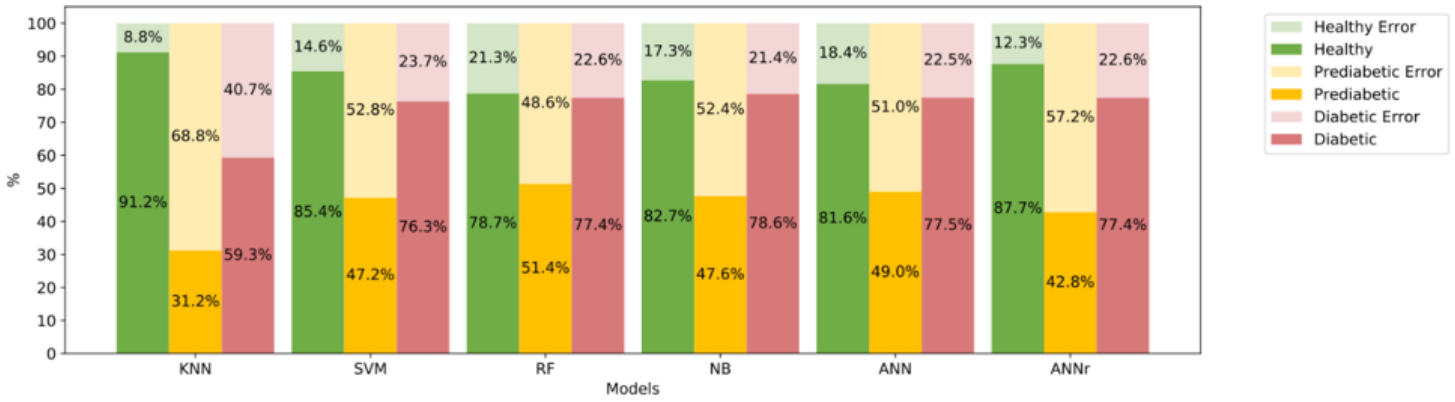

Figure 7

Hit ratio for the three classes, according to each model.

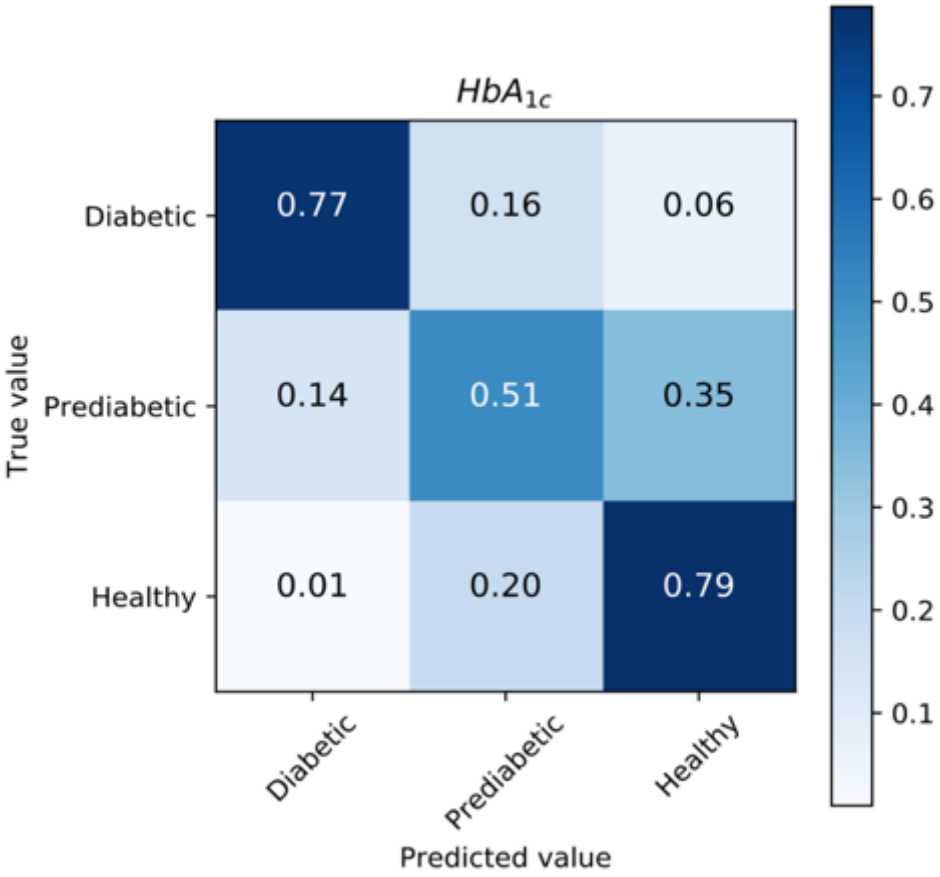

Figure 8

Confusion matrix with the percentage of the random forest (RF) model for the HPD group.

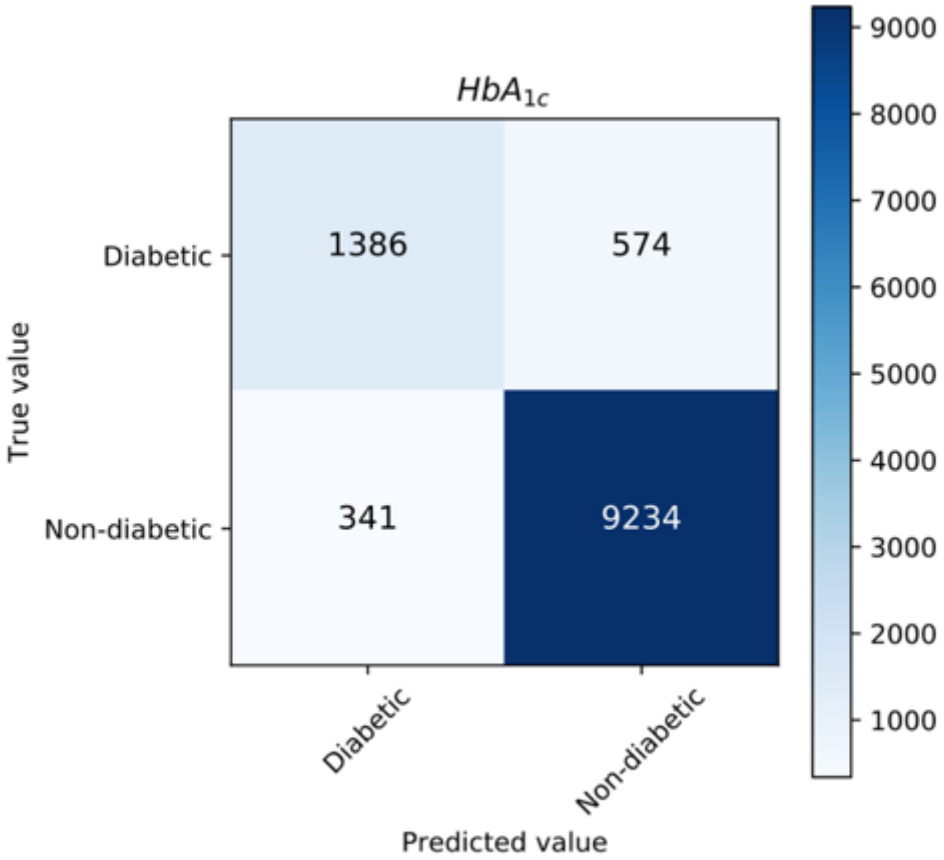

Figure 9

Confusion matrix for the ANNr model and later classification for the ND group.

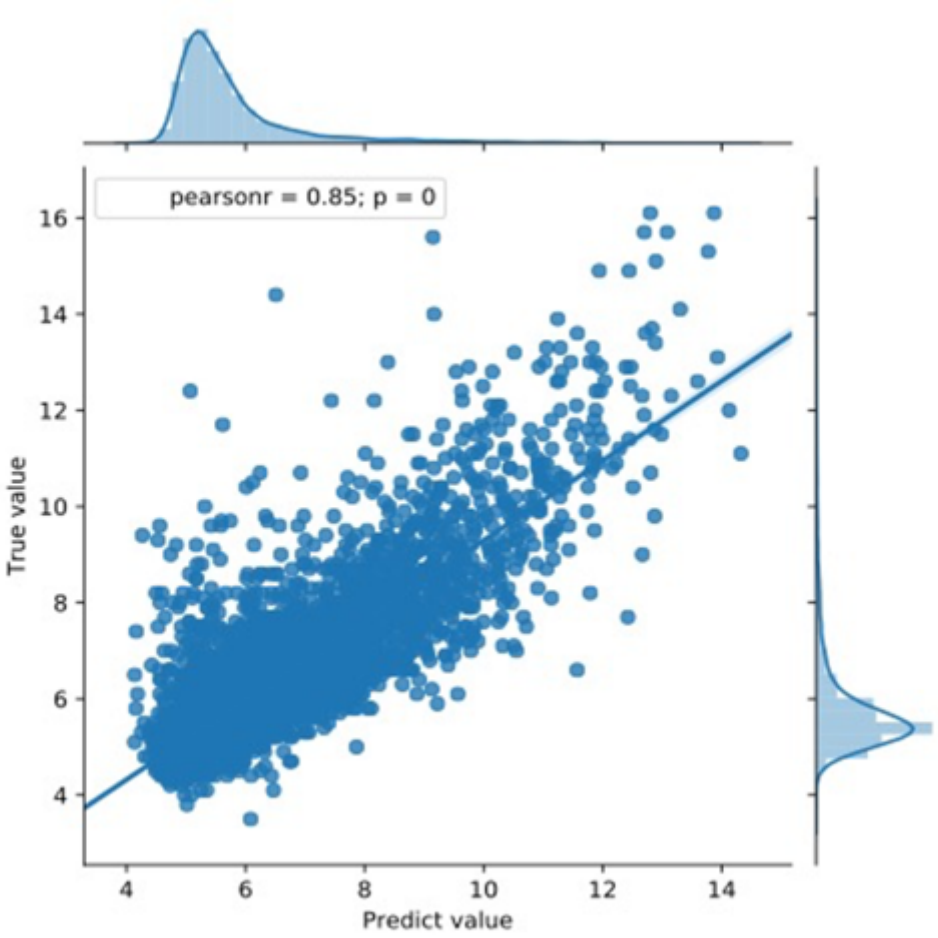

Figure 10

Distribution of predicted HbA1c values to actual values.

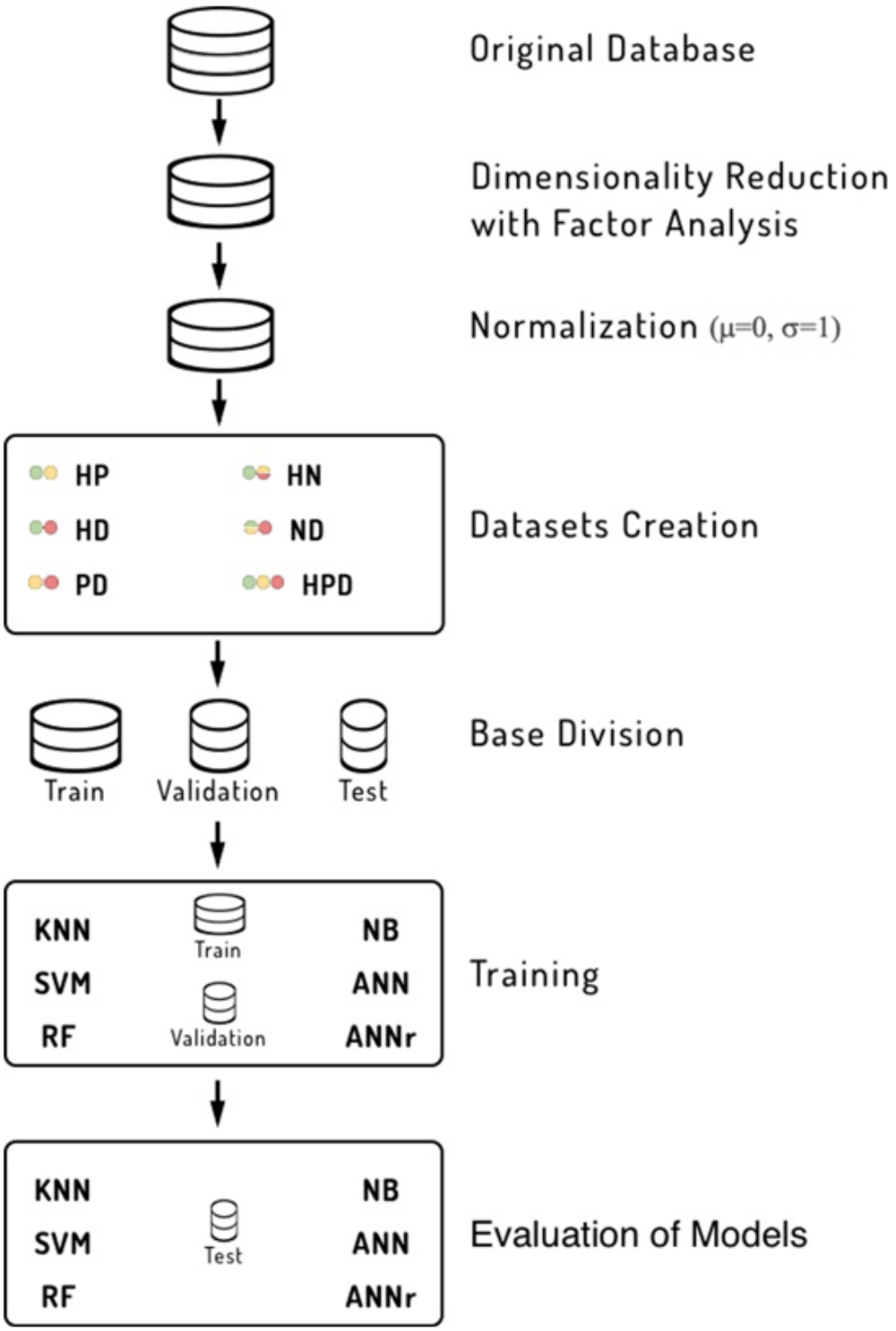

Figure 11

Methodological process diagram.