Prospects for the use of artificial intelligence technologies in predicting the risk of cardiovascular disease (CVD) outcomes in patients with type 2 diabetes mellitus

[1] Alimova D.A., [2] Ikramov A.A., [1]Trigulova R.Kh., [3] Mukhtarova Sh.Sh., [3] Ismailov S.I., Alikhanova N.M., Takhirova F.A.

raisa_trigulova@mail.ru

[1] Republican Specialized Scientific-Practical Medicine Center of Cardiology, Tashkent, Uzbekistan

[2] Institute of Mathematics, Tashkent, Tashkent, Uzbekistan

[3] Tashkent Pediatric Medical Institute, Tashkent, Uzbekistan

Republican Specialized Scientific and Practical Medical Center of Endocrinology, Tashkent, Uzbekistan

Abstract: The study investigates the role of COVID-19 in anamnesis and other features on mortality in 114 patients with type 2 diabetes and cardiovascular diseases during their two-year observation. We performed statistical comparison of two groups of patients and found differences in distributions of some features such as Creatinine, Glomerular Filtration Rate, Total cholesterol, Uric acid, Natriuretic peptide, Ejection fraction, Atrial diastolic flow rate, Isovolumic relaxation time, and others. We also used Logistic Regression and 4-fold crossvalidation to reduce feature space. We built the optimal model using Logistic Regression to predict mortality in two years after medical examination with Concordance Index = 0.955 on test set and Concordance Index = 0.987 on training set.

Keywords: machine learning, cardiology, diabetes, covid-19.

1. Introduction

Diabetes type II and cardiovascular disease are considered risk factors for increased severity of COVID-19 and worse outcomes, including higher mortality. The overall percentage of patients with diabetes affected by COVID-19 ranges from 5.3% to 20% [1]. Recent study [2] now suggests that therapy with dipeptidyl peptidase-4 inhibitors (i-DPP-4, sitagliptin) was neutral for serious adverse cardiovascular events, including stroke, in studies of cardiovascular outcomes in patients with type 2 diabetes mellitus. In a retrospective case-control study [3] conducted in northern Italy during the COVID-19 pandemic, treatment with sitagliptin during hospitalization was associated with reduced mortality and improved clinical outcomes in these patients.

Another monocentric study [4] assessing the impact of antidiabetic drugs on survival in patients with COVID-19 described the relationship between the use of i-DPP-4 and a statistically significant reduction in mortality (11 patients, one of whom died).

However, treatment with sitagliptin was associated with worse outcomes. In a cohort of diabetic patients admitted to the National Center for Infectious Diseases (NCID) in Singapore, patients treated with i-DPP-4 were more likely to require admission to an intensive care unit [5].

Thus, to evaluate the potential improvement in survival associated with i-DPP-4 in post-COVID- 19 patients we started this study.

Treatment with most glucagon-like peptide 1 (GLP1) analogues in patients with type 2 diabetes reduced the incidence of serious adverse cardiac events in studies of cardiovascular outcomes. At the same time, we did not find any studies on the tolerability of this class of drugs after COVID-19.

In this article we evaluate mortality risk among patients with type 2 diabetes and cardiovascular diseases, taking into account the past COVID-19 in anamnesis.

2. Materials and Methods

The follow-up of patients started even before the beginning of the COVID-19 epidemic. The study included 114 patients diagnosed with CHD (coronary heart disease) and with type 2 diabetes (DM2) (49 men and 65 women) have being observed in the CHD Unit of the REPUBLICAN SPECIALIZED CARDIOLOGY SCIENTIFIC-PRACTICAL MEDICINE CENTER since January 2020. The mean age was 63.2 ± 8.8.

Smoking, from bad health habits, in 19 patents (16.6%), normal BMI up to 25 was in 8 patients (7.01%). The frequency of previous myocardial infarction in 46 patients (40.3%), percutaneous interventions in 29 patients (25.4%), coronary by-pass surgery in 7 patients (6.14%). The ratio of FC according to NYHA was as follows: FC III in 88 patients (77.1%), FC II in 15 patients (13.1%), FC I in 11 patients (9.6%). All patients regularly took background therapy, i.e. antiplatelet drug (96.4%), statins (87.7%), beta-blockers (99.2%), RAAS and ACE blockers took 51 patients (55.2%), sartans - 71 patients (62.2%). In addition, patients took antihyperglycemic drugs, i.e. metformin (n-30; 26.3%), sitagliptin/metformin (n-30; 48.2%), and liraglutide (n-30; 25.4%).

Inclusion criteria: presence of antiplatelet drug confirmed coronary heart disease (CHD) in combination with type 2 diabetes; stable forms of CHD; consent of the patient to participate in the medical study. Exclusion criteria: acute MI, stroke; age over 70; type 1 diabetes; permanent forms of AF; CKD stage 4; refusal of the patient to participate in the research. All patients underwent general clinical studies (examination, anamnesis, blood pressure, heart rate, auscultation, palpation, anthropometric data).

During the observation period, 98 criteria were selected for each patient. Among them are age, gender, height, weight, history of COVID-19, background therapy and antihyperglycemic agents taken, blood test results (lipid and carbohydrate spectrum, creatinine, brain natriuretic peptide), EchoCG and ECG parameters.

Among the patients included in the observation, there were 7 fatal cases, out of 26 who recovered from COVID-19. In this regard, the available results of a two-year study were divided into groups: with and without a history of COVID-19, and fatal cases on the individual basis.

Logistic regression [7] was used to produce a mortality prediction model. This Machine Learning method allows not only to classify patients and assign them their mortality risk level, but also constructs easily explainable models. It can also be used to determine important features for the prediction. All methods are provided by libraries scikit-learn 1.1.1 and scipy 1.5.3 for python 3.7.

To measure model's efficiency, we use commonly applied in medical research C-Index [8] (Concordance Index, also known as ROC AUC).

2.1 Preparation of datasets

The whole dataset was divided on negative and positive (lethal outcomes) cases. 80 % of negative cases were randomly selected for training set. 4 positive cases were randomly selected and added to the training set. The remaining negative and positive cases formed the testing set. As Logistic Regression is sensitive to class imbalance during its training, we duplicated positive cases in the training set 20 times to match the number of negative cases in it.

2.2 Statistical analysis of patients' data

We use different statistical methods to analyze dataset including The Brunner-Munzel test [6]. This test was chosen to examine similarities two subsets – with COVID-19 and without it in anamnesis – as it does not require both subsets to have same standard deviation or same distribution. We also compare two datasets using Chi-squared test. We measured standard statistical parameters of the dataset to describe the population in study.

The results are given in Table 1.

Tab. 1 Statistical parameters of some features of patients in the dataset

Feature	Mean and standard deviation	Minimum and maximum values	First and third quartile values	Median
Age, years	62.5 ± 9.2	35 - 75	58- 68.8	65
Weight, kg	87.3 ± 15.6	60 - 168	76.02 - 97.2	85
Height, cm	165.5 ±	150 -	160 -	16

	8.5	196	170	5
BMI	31.8 ± 5.8	21 - 58.1	27.6 - 34.6	30.4
Fasting blood glucose, mmol/l	9.4 ± 3.1	4.5 – 17.0	7.2 - 11.3	8.6
Glycemia postprandial, mmol/l	13.4 ± 4.5	8.0 - 27.2	10 - 16.3	13.1
Glomerular Filtration Rate, ml/min/1,73m²	66.4 ± 21.9	28.0 – 100.0	53.5 - 81.2	64.6
Creatinine,mk/mol/l	94.7 ± 28.6	49 – 172.0	77.5 - 104	91
Total cholesterol, mg/dl	186.3 ± 54.9	88 – 238.0	152.3 - 216.5	180.5
Low density lipoprotein Cholesterol, mg/dl	104.3 ± 49.5	35.0 – 189.0	70 - 133	96
High density lipoprotein Cholesterol, mg/dl	37.9 ± 9.8	20.0 – 58.0	30 - 44	36
Uric acid, mkmol/l,	6.5 ± 2.02	3.1 – 12.0	5.2 - 8.2	6.3
Natriuretic peptide, pg/ml	1885.8 ± 1069	365 - 4568	1003.5 - 2588	1467
Ejection fraction, %	56.8 ± 9.8	42.0 – 72.0	50.3 - 64	60.1
E/A, ratio of early and atrial diastolic flow rates	0.87 ± 0.37	0,52 – 2.64	0.64 – 0,97	0.75
End diastolic volume EDV, ml	53.6 ± 5.7	42.0 – 74.0	49 - 58	53
End systolic volume ESV, ml	37.1 ± 5.8	26.0 – 58.0	33 - 41	36
Interventricular septal, mm	11.9 ± 1.3	9.0 – 15.0	11 - 13	12
Posterior wall of the left Ventricle, mm	11.2 ± 1.5	7.0 - 14.5	10 - 12	11
Isovolumic relaxation Time IVRT, ms	108 ± 21	55.0 – 143.0	98 - 122.8	110
Deceleration time DT, ms	202.2 ± 33.7	90.0 – 254.0	188.3 - 225	208
e'lateral, cm/s	8.8 ± 1.6	5.2 - 14.5	8 - 9.3	8.6
e'septal, cm/s	6.8 ± 1	4.4 –	6.2 - 7.3	6.8

		10.0		

2.3 Search for optimal feature subspace

The number of features for each patient is too large to use in practical terms. We performed a special search for the optimal feature subspace using cross-validation. Initially we built Logistic Regression using all features, trained and tested it using 4-fold cross-validation only on training set. Each fold had one positive case. From 4 generated models tested on validation we get median C. We then tried to eliminate a single feature with increase in C Index each time performing 4- fold cross-validation. We continued this process moving towards increase in C Index and decrease of feature space dimensions.

Similar approach we used previously in modeling COVID-19 severity in patients [9, 10].

3. Results

Firstly, we divided the dataset onto two groups: with and without COVID-19 in anamnesis and analyzed their likeliness. The Brunner-Munzel test compared two groups to determine whether any feature is distributed differently among those patients. As a result, only two features were distributed differently (with $p < 0.05$). They are Coronary artery bypass grafting, Glycemia postprandial. We also have three features with $p < 0.076$: Sodium, Uric acid and Natriuretic peptide levels in blood.

Next, we applied chi-squared test. As there are 75 patients without COVID-19 and only 39 with COVID-19 in anamnesis, we needed to match the data. For each feature we split the interval of its values on up to 12 bins. Then we calculated number of patients in each bin for each group. We replaced values of the second group with respectful bin's median. We generated a new list of values based on a frequency of each bin for the first group and placed each bin's median as many times as this frequency multiplied by 39 (total number of patients in the second group). Now, both lists contain the same total number of values and can be compared using chi-squared.

Here, we have difference with $p < 0.05$ in distribution of the following features: Creatinine, Glomerular Filtration Rate, Total cholesterol, Low density lipoprotein cholesterol, Triglycerides, Uric acid, Natriuretic peptide, Ejection fraction, Atrial diastolic flow rate, Isovolumic relaxation time, and DTE. This chi-squared test helps to compare not just means and standard deviations of two groups, but the actual distributions and their tails.

3.1 Logistic Regression and Feature Selection

We found optimal subspace of 24 features. Next, we trained Logistic Regression model on whole training set. Coefficients of the Model are presented in Table 2. C Index on test set is 0.955 while C Index on training set is 0.987. The ROC curve of the Model on the test set is presented in Fig. 1.

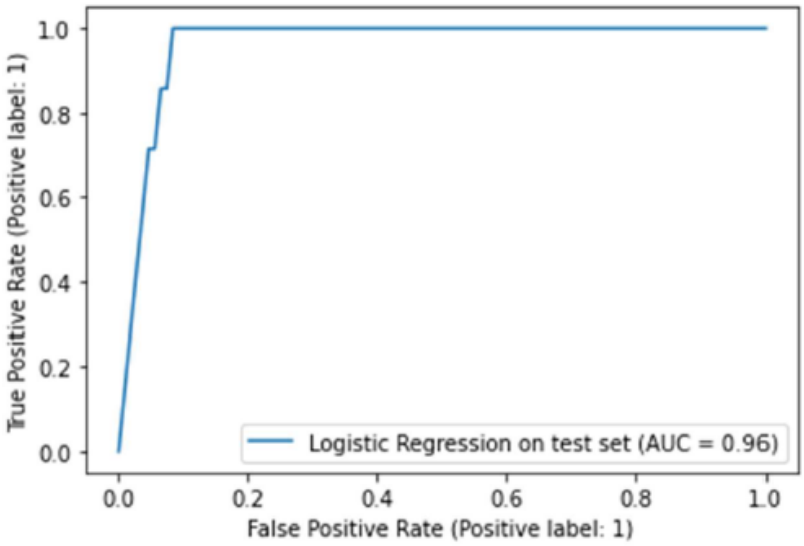

Fig. 1 ROC curve of the optimal model of Logistic Regression on test set

Tab. 2 Coefficients of features in the optimal Logistic Regression Model

Feature	Coefficient
COVID in anamnesis	11.280
Glomerular Filtration Rate, ml/min/1,73m²	-147.539
Creatinine, mk/mol/l	-89.924
Total cholesterol, mg/dl	6.723
Low density lipoprotein cholesterol, mg/dl	-23.516
High density lipoprotein cholesterol, mg/dl	-19.538
Natriuretic peptide, pg/ml	-0.447
Ejection fraction,%	-48.367
A, atrial diastolic flow rate, cm/s	-82.275
E/A, ratio of early and atrial diastolic flow rates	6.093
EDD End diastolic, mm	53.484
ESD End systolic, mm	111.256
Interventricular septum, mm	-10.761
Posterior wall of the left ventricle, mm	66.397
Left atrium,	69.473
E/e`	7.547
Isovolumic relaxation time, ms	26.587
DTE, early mitral flow deceleration time, ms	74.172
e'lateral, cm/s	-26.811
e'septal, cm/s	-6.432
Dosage of Liraglutid	-0.756
Dosage of sitagliptin	-21.288
Dosage of meftormin	-62.650
Angiotensin II receptor antagonists	-4.395

4. Discussion

The signs in Table 2 demonstrate the input of each feature into prediction. An increase in features with negative coefficients lowers the total risk of mortality predicted by the model.

The study followed patients for only two years and can only be used to model mortality within a 2-year timeframe. As it continues, we will be able to build survival models in the future.

We reviewed related work to support the selected subspace of the features and to find scientific explanation to the model.

Large-scale clinical studies have shown that incretin-based therapy does not have a positive effect on eGFR. Most reports of renal outcomes for incretins only analyze the statistical data on the combined renal outcome, which combines albuminuria (protein in the urine), a decrease in the estimated glomerular filtration rate (eGFR) and end-stage renal disease (ESRD). The LEADER study [11] investigated the effect of the GLP-1 agonist liraglutide on CVD outcomes. In liraglutide-treated patients with fewer CV events than placebo (13.0 vs 14.9%, hazard ratio CI 0.87–9.97, $p < 0.001$ for non-inferiority, $p < 0.01$ for superiority). The occurrence or exacerbation of nephropathy was less with liraglutide (5.7 vs 7.2%, hazard ratio CI 0.67–0.92, $p = 0.003$). Again, the improvement in renal outcomes was due to a lower incidence of albuminuria rather than a decrease in renal function.

The TECOS study [12] (DPP-4 inhibitor sitagliptin) showed that sitagliptin was non-inferior to conventional treatment in CV outcomes in patients with established CV disease. The effects of DPP- 4 inhibitors are mainly pleiotropic (anti-fibrotic, antioxidant, anti-apoptotic, and anti-inflammatory).

Many studies of phase 2 and phase 3 aimed at evaluating the hypoglycemic effects of incretin therapy have found favorable changes in cardiovascular risk factors in patients with type 2

diabetes. A GLP-beneficial effect on body weight [13], arterial pressure [14], and lipid profile [15]. A LEADER study assessed factors influencing time to first MACE and identified several factors including HbA1c, body weight, systolic blood pressure (BP) and LDL cholesterol [16].

Brain natriuretic peptide (BNP) concentrations are reduced in people with obesity, insulin resistance, and diabetes, and this deficiency may contribute to their cardiovascular risk. An increase in NT-pro-BNP levels shows [17] a strong correlation with poor prognosis in elderly patients with chronic heart failure. NT-pro-BNP is a better short-term independent predictor of CV mortality than C-reactive protein and albumin excretion in elderly patients with type 2 diabetes according to a population-based study [18].

5. Conclusion

The study demonstrated the difference in patients with type 2 diabetes and cardiovascular diseases based on COVID-19 in anamnesis. Although means and standard deviations were comparable for most of the features, the actual distributions were different according to chi-squared test.

We used a simple Machine Learning algorithm to predict mortality in two years after admission. Logistic Regression is a popular tool in binary classification as it provides explanation of the outcome based on each input value. It also helps to prioritize the future studies as focus would be on features with the largest absolute value.

The optimal model constructed in this study will play a significant role during admission of patients that satisfy criteria of the study. It has C Index of 0.955 on test set. The model also demonstrates a specific role of prescribed diabetic medications on outcome.

We plan to study groups of patients based on medication, to include follow medical examinations to predict changes in physiological indicators and build more robust models.

References

[1] Deng S.Q., Peng H.J. Characteristics of and public health responses to the coronavirus disease 2019 outbreak in China. J Clin Med. 2020. DOI: 10.3390/jcm9020575

[2] Lim S., Oh T. J., Dawson J., Sattar N. Diabetes drugs and stroke risk: intensive versus conventional glucose-lowering strategies, and implications of recent cardiovascular outcome trials. Diabetes Obes. Metab. 22, 2020 – pp. 6–15. DOI: 10.1111/dom.13850

[3] Solerte S., et al. Sitagliptin treatment at the time of hospitalization was associated with reduced mortality in patients with type 2 diabetes and COVID-19: a multicenter case-control retrospective observational study. 2020 Dec; 43(12) – 2999-3006. doi: 10.2337/dc20-1521

[4] Mirani M. et al. Impact of comorbidities, glycemia at admission, and DPP-4 inhibitors in type 2 diabetic patients with COVID-19: a case series from an academic hospital in Lombardy, Italy. Diabetes Care 2020; 43(12): 3042–3049. https://doi.org/10.2337/dc20-1340

[5] Dalan R., et al. The association of hypertension and diabetes pharmacotherapy with COVID-19 severity and immune signatures: an observational study. European Heart Journal - Cardiovascular Pharmacotherapy, Volume 7, Issue 3, May 2021 – pp. e48–e51. https://doi.org/10.1093/ehjcvp/pvaa098

[6] Karch J. Psychologists Should Use Brunner-Munzel's Instead of Mann-Whitney's U test as the Default Nonparametric Procedure. 2021. DOI: 10.31234/osf.io/kgdwn.

[7] Ambrish G., Ganesh B., Ganesh A., et al. Logistic regression technique for prediction of cardiovascular disease. Global Transitions Proceedings, Volume 3, Issue 1, 2022 – pp. 127-130, https://doi.org/10.1016/j.gltp.2022.04.008.

[8] Shiffman D., Louie J.Z., Devlin J.J., Rowland C.M., Mora S. Concordance of Cardiovascular Risk Factors and Behaviors in a Multiethnic US Nationwide Cohort of Married Couples and Domestic Partners. JAMA Netw Open. 2020; 3(10):e2022119. doi:10.1001/jamanetworkopen.2020.22119

[9] Ikramov A., Adilova F., Anvarov K., Khadjibaev A. COVID-19 Severity Prediction in Patients Based on Anomaly Detection Approach. In: Yang, XS., Sherratt, S., Dey, N., Joshi, A. (eds) Proceedings of Sixth International Congress on Information and Communication Technology.

Lecture Notes in Networks and Systems, vol 235. Springer, Singapore. 2021 – p. 611-618. https://doi.org/10.1007/978-981-16-2377-6_56

[10] Ikramov A., Anvarov K., Sharipova V., Iskhakov N., Abdurakhmonov A., Alimov A. Machine Learning Algorithms in Application to COVID-19 Severity Prediction in Patients. In: Long, G., Yu, X., Wang, S. (eds) AI 2021: Advances in Artificial Intelligence. AI 2022. Lecture Notes in Computer Science(), vol 13151. Springer, Cham. 2022 – p. 344-355. https://doi.org/10.1007/978-3-030-97546-3_28

[11] Marso S.P., Daniels G.H., Brown-Frandsen K., et al. Liraglutide and cardiovascular outcomes in type 2 diabetes. N Engl J Med. 2016 Jul 28; 375(4):311-22. DOI: 10.1056/NEJMoa1603827

[12] Green J.B., Bethel M.A., Armstrong P.W., et al. Effect of sitagliptin on cardiovascular outcomes in type 2 diabetes. N Engl J Med. 2015 Jul 16; 373(3):232-42. DOI: 10.1056/NEJMoa1501352

[13] Monami M., Dicembrini I., Nardini C., et al. Effects of glucagon- like peptide- 1 receptor agonists on cardiovascular risk: a meta- analysis of randomized clinical trials. Diabetes Obes Metab. 2014 Jan; 16(1):38-47. DOI: 10.1111/dom.12175

[14] Fonseca V.A., DeVries J.H., Henry R.R., et al. Reductions in systolic blood pressure with liraglutide in patients with type 2 diabetes: Insights from a patient-level pooled analysis of six randomized clinical trials. Journal of Diabetes and its Complications Volume 28, Issue 3, May–June 2014 – pp. 399-405; DOI: 10.1016/j.jdiacomp.2014.01.009

[15] Monami M., Lamanna C., Desideri C.M., et al. DPP-4 inhibitors and lipids: Systematic review and meta-analysis. Adv Ther. 2012 Jan; 29(1):14-25. DOI: 10.1007/s12325-011-0088-z

[16] Buse J.B., Bain S.C., Mann J.F.E., et al. LEADER Trial Investigators. Cardiovascular risk reduction with liraglutide: an exploratory mediation analysis of the LEADER trial. Diabetes Care. 2020 Jul;43(7):1546-1552. DOI: 10.2337/dc19-2251

[17] Shi L., Zhang Y., Zhang J., Gao Y., et al. Application of blood pre-albumin and NT-pro BNP levels in evaluating prognosis of elderly chronic heart failure patients. Exp Ther Med. 2020;20:1337–1342. DOI: 10.3892/etm.2020.8865

[18] Casale Monferrato Bruno G., Landi A., Barutta F., et al. N-terminal probrain natriuretic peptide is a stronger predictor of cardiovascular mortality than C-reactive protein and albumin excretion rate in elderly patients with type 2 diabetes: the Casale Monferrato population-based study. Diabetes Care. 2013; 36:2677–2682. DOI: 10.2337/dc13-0353