

¹ Natalya V. Frolova² Nadezhda V.
Yuferova³ Dmitry A. Zheltov

Preventing Stroke Risk using an Algorithm-Based Medical Network



Abstract: - According to the World Health Organization, 15 million people worldwide suffer from a stroke every year. Of these, five million die, and another five million experience prolonged periods of immobility. In four out of 10 fatal cases, patients could have survived if their blood pressure had been controlled in a timely manner. There are many developments in the early prevention of stroke. The authors reviewed the latest achievements in this area of diagnostics and got the idea that it is necessary to create an Internet platform that works on the principle of an algorithm-based social network. Users could enter their basic medical data and receive an estimate of the risk of different types of stroke, as well as recommendations for preventing them. The more people using this platform, the more accurate the risk analysis would be. This project could make a contribution to machine learning in medicine, as well as to improving the quality of life of millions of people.

Keywords: Algorithm-based, medical network, preventing risk, stroke.

I. INTRODUCTION

Many people are not aware of the dangers and risks associated with diseases such as heart attack and stroke. Experts emphasize several reasons for this condition: outdated methods of informing, complexity in understanding of specialists, insufficient budgets for social events.

To date, the governments of all countries and relevant authorities take various measures to inform citizens about the risks and dangers of diseases of the cardiovascular system, as evidenced by the presence of many articles and activities carried out for this purpose [1].

Since medicine is a very specialized field that requires a high level of competence to understand, there is a risk that the average person may not understand all the risks of a phenomenon, even when explained by a competent person. The presence of professional terminology can only confuse the average person. It is to avoid such situations that various health organizations have resorted to simplifying all the background information and reducing it to a list of rules to be followed in the prevention or occurrence of stroke.

Despite the fact that stroke is a serious problem that gets more serious every year, the methods of its treatment and rehabilitation have not changed for a long time, except for the improvement of medications. Today, the recognized fact of effective reduction of stroke morbidity and mortality is primary prevention and informing people about the risks of stroke.

The article analyzes the software tools that exist today, approved by recognized health experts, for primary stroke risk identification, shows their limitations and proposes an approach to the development of software (software) that meets modern standards.

Artificial intelligence techniques were used in the study to identify the best models for obtaining information about the influencing factors on the indicator under investigation. For the development were used such technologies as: HTML, CSS, Bootstrap, JavaScript.

Data from the World Health Organization (WHO) were used to estimate the probability of stroke using econometric modeling.

¹ Perm State University, Perm, Russia. nvf_psu@mail.ru

² Perm State University, Perm, Russia. Yuferova@econ.psu.ru

³ Saint Petersburg Electrotechnical University "LETI", Saint Petersburg, Russia. jeltov.dima@yandex.ru

II. SOFTWARE PRODUCTS FOR ASSESSING THE LIKELIHOOD OF STROKE

Looking on the Internet, one can find many proposals for determining the risk of stroke. Such software products use artificial intelligence methods, machine learning algorithms. Let us classify them.

Software products aimed at informing the population about the risk of stroke differ in the methods of distribution (Internet or mobile application) and methods of information (reference material or risk calculator).

The method of distribution of a software product is very important in order to reach the target audience, and it also determines which software development technologies will be used. For example, when distributing via the Internet, only Internet access is required to use the service, as the client server part does not require installation. Such accessibility contributes to the recruitment of a larger audience as opposed to a mobile application. The web application is also updated automatically and does not require the user to download client files. Web application is also easier and cheaper to promote due to advanced settings of targeted advertising on various search sites and free placement on the Internet, while promotion of mobile application requires paying a fee to place it in specialized application stores such as "GooglePlay", "Apple store" and "AppGalery". Mobile application development also requires more time than web application, this is due to the complexity of the technology stack used for development. To ensure the availability of mobile application requires development for different platforms, such as "IOS" or "Android", which also have their own technological features and rules for data usage. That's why most of the popular services start with a web application and only in the process of scaling (adding new functionality) a mobile application is created.

When you search for "Stroke Risk" you can find many different services in the app store and on the internet, but most of them are either reference material or stroke risk calculators. Among such apps, the services that provide reference material are the most numerous, this is because such software is much easier and faster to develop than statistical calculators, however, if one looks at the number of downloads of each type of app in the app store, one will realize that the user is more interested in stroke calculators. This is due to the fact that it is difficult for a person far from the medical profession to perceive reference material even in the presence of the desire to get answers to a worrying question. That is why stroke calculators are more popular in the search for simple and easy to understand quick answers.

The following apps were identified by searching the internet and the GooglePlay app store:

- "Stroke Riskometer" Mobile application app: Release date 12 August 2021, over 10 thousand downloads;
- "CHA2DS2-VASc HAS-BLED" Mobile application: Release date 19 February 2018, more than 10 thousand downloads;
- "Calculator CHA2DS2-VASc"[2] Web App;
- "SCORE Calculator"[3] Web App;
- "ACC/AHA 2013 Calculator"[4] Web App;
- "CHA2DS2-VASc Calculator"[5] Web app.

Familiarization with stroke risk calculators was started with mobile applications because of the assumption of their more advanced functionality. This will allow an objective assessment of the comprehensiveness of the assessment systems themselves and the functionality of the applications, which forced developers to move to mobile platforms.

The Stroke Riskometer has been endorsed by major international neurological organizations and WHO specialists. It is a mobile application with a stroke risk calculator and minimal background information. The test includes 21 questions assessing various indicators such as: age, gender, blood pressure, height, weight, presence of heart disease, race.

After the test, it is suggested to evaluate the results on a diagram of relative risk or on a diagram of absolute risk, as well as to estimate the risk (probability of stroke) in 10 years. There are also reference materials in the "Prevention" and "D.D.A.R." sections. Despite the fact that the application was released in 2021, it clearly does not meet modern standards of mobile application development. This fact is given by the standard elements of controllers from the markup language XAML, standard font size and type, as well as the lack of animation. Also, the application does not allow to estimate the risk of stroke for a given period of time and does not tell in too much detail what evaluation system it uses, unlike other calculators.

The "Calculator CHA2DS2-VASc HAS-BLED" application is an even simpler variation of the stroke risk calculator. It contains 2 tabs with a test for 2 different rating scales. The test consists of questions that assess indicators such as: age, gender, presence of hypertension, whether you have had a stroke, and whether you have diabetes. The assessment is based on two assessment scales CHA2DS2 and the more modern CHADS, but these assessment methods do not include some stroke risk factors. The objectivity of such an assessment method is still debated not in favor of this scale. This application, unlike the other calculators, has no reference information and also uses a less objective method of assessment. The interface also contains standard controllers from the XAML markup language, which indicates that this application was developed in a hurry.

Unfortunately, many popular mobile applications for stroke risk assessment do not meet modern standards of mobile application development. All reviewed instances had a basic interface and a standard set of controllers and diagrams. Some of the representatives did not have localization into popular languages, which negatively affects public awareness of the risks assessed in this application.

When evaluating web applications aimed at stroke risk assessment, it was found that most of such calculators are located on the websites of official medical institutions, where it is also possible to find background information and interviews with specialists. For example, when evaluating the SCORE calculator, which uses the assessment method of the same name, it can be seen that the web site belongs to the Yamal Center for Public Health and Medical Prevention. The calculator consists of a series of questions assessing: country of residence, age, gender, blood pressure, smoking, and cholesterol level. After entering the relevant information, a coefficient is displayed, the interpretation of which can be viewed on the same web page.

The SCORE (Systematic Coronary Risk Evaluation) scale was developed to assess the risk of fatal cardiovascular disease within 10 years. The scale was based on data from cohort studies conducted in 12 European countries (including Russia) with a total of 205178 people. Although the SCORE risk assessment is recognized by the medical community as fairly objective, it is still not detailed, reducing the answer to whether the chance of stroke is high (if > 5%) or low. The web application, although it has a basic interface, assesses risks objectively enough and also belongs to the portal of a medical institution with reference materials and feedback from specialists, which positively affects its informative value.

The ACC/AHA 2013 calculator as well as SCORE calculates only the ten-year risk of stroke and has the same list of control questions, but has no reference material. Detailed methods of calculating probability can be found in the "basement" of the website where the calculator is located. There are also links to various studies and articles. ACC/AHA is an American College of Cardiology (ACC) and American Heart Association (AHA) report [6] that specified the model used by this calculator.

The CHA2DS2-VASc calculator, as well as the " Calculator CHA2DS2-VASc HAS-BLED" application, uses a scale slightly supplemented by CHADS2 for risk assessment, unlike its classic version CHA2DS2-VASc takes into account various vascular diseases. The calculator itself is located on a medical forum that allows a person to read the interpretation of the results in more detail if desired. Although the objectivity of even a modified version of this scale is still debatable, the calculator itself, as well as the forum has an eye-pleasing interface.

Although today there are many software products designed to inform the population, their study leads to the conclusion that they do not meet modern standards of application development, so, for example, none of the reviewed web applications has adaptive design for mobile devices, while most of the world's population uses mobile Internet. Even the Stroke Riskometer app, which is endorsed by international health leaders, does not have a decent design and functionality.

In order to develop the best analog, it will be necessary to create a web application and also to provide it with adaptive design for mobile devices to inform a potentially larger number of people, and also to introduce on the website with the web application reference materials on the interpretation of the obtained coefficient. To make the application more transparent, it is worth writing a model specification, establishing feedback for consultation with experts, and adding the possibility to consent to data collection for subsequent studies.

III. STATISTICAL MODEL FOR DETERMINING THE RISK OF STROKE

Let us consider the specification of the proposed logit and probit models embedded in the proposed tool that determine the probability of stroke occurrence depending on the selected factors.

Based on the analysis of the articles, we will select the following variables for the econometric model: subject patient identification number, sex, age, presence of hypertension, presence of heart disease, marital status, type of employment, place of residence, mean glucose level, body mass index, smoking status, and presence of stroke. Cross-sectional data were selected for the study. There were 5110 observations in the initial sample. The information was obtained from official WHO (World Health Organization) sources. Due to some missing data in the provided sample, two variables (Type of activity-work_type and Place of residence- Residence_type) had to be dropped, and observations with missing data and obvious outliers had to be removed.

The initial analysis of the data is quite interesting: the sample has a mean patient age of 44 years with a variance of 18 years, the coefficient of variation indicates heterogeneity in the data, there is a slight rightward bias in the sample, the minimum age is 10 years and the maximum is 82 years, which shows how diverse and representative the values in the sample are.

The constructed correlogram showed that the correlations between the dependent variable (had a stroke - sick or no stroke - healthy) and the significant variables do not exceed 0.25, which can be called an average correlation. This result is due to the fact that the number of sick people relative to healthy people is much smaller, and in medicine such a relationship can be called tangible.

Let us consider several models and select the best one for further inclusion in the proposed tool. We use the logit model, which is able to account for both quantitative and qualitative factors [7] and compare it with the result from the probit model (Table I):

Table I. Models

Model (logit)	$df\$stroke = df\$hypertension + df\$heart_disease + df\$age + df\$avg_glucose_levels$
Model (probit)	$df\$stroke = df\$hypertension + df\$heart_disease + df\$age + df\$avg_glucose_levels$

Checking for multicollinearity showed that there was no multicollinearity in the logit and probit models. Marginal effects were calculated and their mean values were found, which we will interpret as coefficients (see Table II).

Table II. Estimation of average marginal effects for some variables

Logit			
$df\$hypertension$	$df\$heart_disease$	$df\$age$	$df\$avg_glucose_level$
0.0232318031	0.0203477845	0.0028905515	0.0002004317
Probit			
$df\$hypertension$	$df\$heart_disease$	$df\$age$	$df\$avg_glucose_level$
0.0248571587	0.0224458280	0.0027009395	0.0001942943

Based on the obtained results, we can draw conclusions. The obtained coefficients can be interpreted as the amount by which the probability of stroke increases when the variable is increased by one. Thus, for example, the presence of heart disease in a patient according to the calculated coefficients increases on average the probability of stroke by 2%.

Neural networks have also been considered as competitive models. At this stage, we have stopped on logit, probit models.

IV. CONCLUSION

In the process of creating the web application, there was a problem of not having enough data in the sample for some indicators, which had a bad effect on the estimation of their marginal effects. The search for other datasets was also unsuccessful due to the fact that most of the datasets related to medicine are paid due to the large number of private organizations in this industry and the legislation of some countries regarding the disclosure of medical information.

On this basis, it was decided to add the possibility of voluntary and at the same time anonymous transfer of test data for further study, which contributes to the development of the platform and also enables other researchers in the field to use these data (Fig.1).

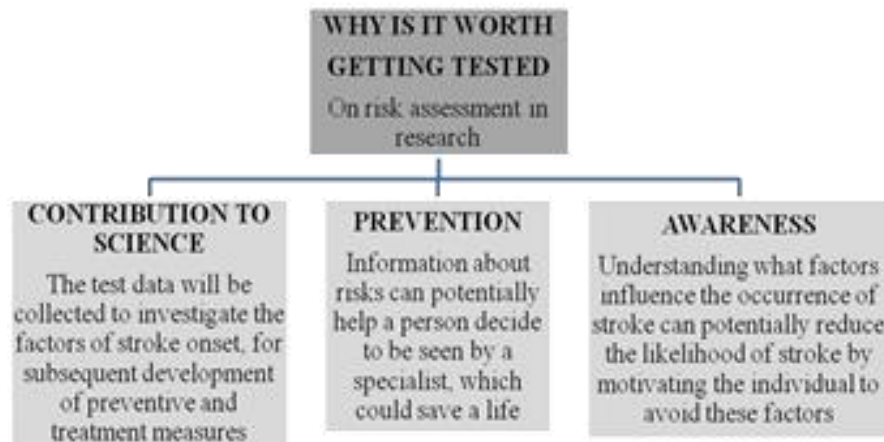


Figure 1. Benefit section: mobile

However, despite the problems associated with parameter estimation, both the calculator and the model itself are keeping pace with competitors in terms of the number of parameters, and potential development from the data left by users will help to surpass all current software solutions in free access.

The website also included blocks of background information to inform the public about the risks and how to detect stroke, which were simplified into columns with sequential instructions and pictures for better understanding of how to proceed. These blocks are also adaptable, making them easy to access regardless of the device used. To select the reference information, various stroke detection methodologies were studied and then a method called UDAR [8] was chosen, which reduces the stroke detection procedure to 4 simple steps.

The size of this article did not allow for a more detailed description of the program interface. At this point, the Stop Stroke platform is a minimal viable product, that is, sufficient for presentation to the public, but has the potential for development because of the built-in functionality for data collection.

Targeted advertising is expected to be emphasized and will help to customize the delivery of information about the platform to groups at risk of stroke. Targeted advertising will also help to encourage large numbers of people to leave their data, which will then be used for stroke research and platform development. The number of potential users using the platform is critical, as the number of users leaving their data, as well as their diversity, directly affects the representativeness of the sample.

REFERENCES

- [1] V. L. Feigin, M. O. Owolabi on behalf of the World Stroke Organization–Lancet Neurology Commission Stroke Collaboration Group, "Pragmatic solutions to reduce the global burden of stroke: a World Stroke Organization–Lancet Neurology Commission", *Lancet Neurology*, 2023, vol. 22, No. 12, [https://www.thelancet.com/pdfs/journals/lanneur/PIIS1474-4422\(23\)00277-6.pdf](https://www.thelancet.com/pdfs/journals/lanneur/PIIS1474-4422(23)00277-6.pdf).
- [2] "Calculator CHA2DS2-VASc", PJC, <https://www.rlsnet.ru/med-calculators/kardiologiya/skala-cha2ds2-vasc-2> (accessed 8 April 2024).
- [3] "Calculator SCORE: risk of death from cardiovascular disease", The portal of the YNAO, http://www.yamalcmp.ru/poleznye_kal_kulyatory/kal_kulyator_score/ (accessed 8 April 2024).
- [4] "Cardiovascular risk assessment (10-year, ACC/AHA 2013)", MSD Manuals, <https://www.msdmanuals.com/medical-calculators/ACCAHA2013-ru.htm> (accessed 8 April 2024).
- [5] "CHA2DS2-VASc: Chronic Heart Failure Scale", WpCalc, <https://wpcalc.com/shkala-cha2ds2-vasc/> (accessed 8 April 2024).
- [6] N. Stone, J. Robinson, A. Lichtenstein, et al., "2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: A Report of the American College of Cardiology/American Heart Association

- Task Force on Practice Guidelines”, J Am Coll Cardiol, 2014, Jul, 63 (25_Part_B), p.2889–2934, <https://doi.org/10.1016/j.jacc.2013.11.002>.
- [7] M.V.Kirpischikova, M.V.Radionova, N.V. Frolova, “Methods of Machine Learning in Assessing the Financial State of an Enterprise”, in: Rocha, A., Isaeva, E. (eds) Science and Global Challenges of the 21st Century - Science and Technology, Perm Forum 2021, Lecture Notes in Networks and Systems, Springer, Cham, 2022, vol. 342, p.1006-1021, https://doi.org/10.1007/978-3-030-89477-1_91.
- [8] “A simple test to save a life: UDAR”, Center for Maternity and Childhood Protection of Magnitogorsk, <https://cmd74.ru/news/zapomnite-prostoj-test-dlya-spaseniya-zhizni-udar/> (accessed 8 April 2024).