УДК 004.62

APPLICATION MACHINE LEARNING TO CONTROL STUDENTS TRAJECTORY

F.M.Alimova
senior Lecturer

Tashkent University of Information Technologies named after Muhammad al Khorazmiy. E-mail: fotima.pm@mail.ru

F.M.Alimova

She graduated from the Tashkent University of Information Technologies named after Muhammad al Khorazmiy. TUIT teacher on the estate of Muhammad al Khorazmiy

Abstract. Successful and productive development of the country's digital economy is a key factor in sustainable development, production growth in all areas of socio-economic activity, which increases the country's competitiveness, the quality of life of citizens, ensures economic growth and national sovereignty. Currently, modern vocational education is moving to a qualitatively new level in connection with the introduction of a competency-based approach, which aims to provide students with tools for both understanding and action, allowing them to perceive new socio-economic realities, as well as navigate in changing conditions learning and work.

The authors of the article are offered a multi-parameter model that analyzes all the parameters of a graduate based on big data and provides estimates for the qualifications of graduates.

Keywords: education, teaching methods, intellectual analysis, assessment, qualification, competence, learning process, innovation.

Today, the labor market requires highly qualified, highly competent personnel. The curricula of universities are adjusted to the requirements of the labor market. E-education around the world is rapidly developing and the main problem is the timely provision of students with high-quality educational information. This task cannot be solved without analyzing the large flow of information that enters the information environment of e-education from the participants in the educational process - students, teachers, administration, etc. There are many different types of data, both structured and unstructured, that are difficult to process with traditional statistical methods.

E-education reveals new, sometimes hidden, relationships in big data, new knowledge (data mining), which can be used to improve the educational process and increase the efficiency of its management. To classify electronic educational resources, identify patterns (templates) of students with similar psychological, behavioral and intellectual characteristics, develop individualized curricula, the article proposes to use big data analysis methods.

To date, many software applications for big data mining have been developed. These software products can be used for classification, clustering, regression and network analysis of educational information. The use of these methods in e-education will allow teachers to receive information about students in a timely manner, quickly respond to any changes in the learning process, and make timely changes to educational content. As well as the data obtained make it possible to manage the student's educational trajectory.

The trend that takes into account the individual characteristics and personal qualities of students gives a transition to student-centered learning standards. Today, it is especially important in the context of the introduction of such "future technologies" as expert systems, robotics and additive manufacturing methods. Educators are making efforts to find a model of an individual approach in mass education, which will allow the use of an adaptive approach in education that takes into account the individual characteristics of each student. In the 19th century, this was achieved through the individual work of the teacher with each of the students, identifying his preferences, inclinations, determining the material that the student did not learn. This is not possible in modern streaming teaching, as teachers are overloaded due to the increase in the number of classes that the teacher works with. In addition to this, the number of students, subjects taught and the amount of “paper work” are growing. In this situation, many children have reduced motivation to study, which remains with them even at the stage of obtaining higher education at the university. One of the results is that the vast majority of university graduates work outside their specialty. Of course, there are exceptions to these rules, but, in general, the situation does not change. As a solution to the problem, it can be proposed to build an individual learning trajectory. In the case of mass streaming education, the task of constructing an individual learning path can be solved using machine learning algorithms and statistical methods.

Elective modules allow students to independently form up to 30% of the educational program. The task of this part of the educational program is to teach the student to make decisions independently, make informed choices, find their own deficits and ways to fill them, focusing on the goals and objectives set for themselves, understanding the image of the profession, analyzing the external situation, their own experience, changes in metropolis environment [2]. Such a system of choice allows each student to complete an individualized training program and receive a unique competency map as a result. Models for building educational trajectories within the modules make it possible to maximally systematize and logically connect the elements of the modules and the technologies for their implementation through the practical application of the acquired skills in everyday and professional activities, as well as through the acquisition of experience in social, project and research work. The choice of these modules is carried out by students every semester in the information environment. After that, study groups are formed from among the students who signed up for a particular module. The ranking of the most popular elective modules among students in the 2019-2020 academic year is as follows: psychology of interpersonal relationships, psychology of emotions, psychology of family and family education, self-development and personal growth, psychology of conflict, life hacks for the future professional, psychological technologies for discovering and developing oneself, emotional well-being and personal achievements, history of cinema, psychology of extreme situations, japanese is easy for everyone.

The University is constantly improving models and methods for choosing elective modules. In order to form in the information environment a personalized set of recommendations from elective modules for building individual educational trajectories of students, a solution was developed and tested based on data mining (artificial intelligence). Recently, intellectual analysis of educational data (Educational Data Mining, EDM) has been increasingly used in the information environment of the university and is being introduced as new services to improve the educational process [3]. One such example is the development of a recommender system at the University of California at Berkeley, which aims to help students make decisions about choosing elective courses [4]. Neural networks, in particular RNNs, are used as the main methods for generating recommendations.

In, a hybrid multicriteria recommender system with genetic optimization is used to help students with the choice of elective courses. It uses two multi-criteria systems: the first is based on a collaborative filtering model, and the second is based on content filtering. As input parameters with information about the student, the student's grades for previous courses, the level of satisfaction and the chosen direction of education are used. The following parameters are used to describe training courses: information about teachers, competencies obtained from the training course, area of knowledge of the training course, description of the course in the form of keywords

[6]. Machine learning algorithms are used as the main ones for generating recommendations [7]. In particular, in [8] a comparative analysis is performed for popular algorithms kNN (k-nearest neighbors), singular value decomposition (P-SVD), sparse linear method (SLIM). Approbation of these methods was carried out on the basis of a Chinese university, where among the three listed methods, kNN and SLIM showed the best values.

In connection with the transition of the education system to a competency-based approach actual is the problem of evaluating learning outcomes, as well as building an individual trajectory of student learning, the solution of which requires the use of modern information technologies. In accordance with the federal state standards of higher professional education (FGOS VPO) of the third generation, which determine the requirements for the results of mastering basic educational programs, up to 50% of disciplines have a variable character, those. depends on the choice of the student. It's significant is reflected in the results of the formation of various competencies.

The article discusses models, methods and algorithms for finding the optimal individual educational trajectory student

References

[1] Data mining in customer relationship management / F.M. Alimova / TATU xabarlari 2(29)/2014, Toshkent, 8-11 p..

[2] Control of knowledge in the test. / F.M. Alimova, U.Giyosov, A.Abdullayev // “XXI аср ва технология соҳасидаги устувор йўналишлар”, VII Халқаро илмий конференцияси бўйича мақолалар тўплами, том 2, Тошкент, 2014 г. 473-477 с.

[3] Каримова В.А, Алимова Ф.М. / Оценивание знаний студентов при преподавании специальных дисциплин на опыте университета ИНХА (Южная Корея). Труды Северо-кавказского филиала Московского технического университета связи и информатики. Часть 2./ Подготовлены по результатам международной молодёжной научно-практической конференции СКФ МТУСИ «ИНФОКОМ-2015», Ростов-на-Дону, 20-25 апреля 2015 года, стр. 90-92

[4] Хусу, А.П. Шероховатость поверхностей (теоретико-вероятностный подход) / А.П. Хусу, Ю.Р. Виттенберг, В.А. Пальмов. – М.: Наука, 1975. – С. 344.

[5] Линник, Ю.В. Математически-статистическкое описания профиля поверхности при щлифовании / Ю.В. Линник, А.П. Хусу // Инженер. Сборник АН СССР. – М. Академиздат, 1954. – С. 432.

[6] Найак, П.Р. Применение модели случайного поля для исследования случайных поверхностей / П.Р. Найак // Проблемы трения и смазки. – 1971. – №3. – С. 85-89.

[7] Харин Ю. С. Теория вероятностей, математическая и прикладная статистика: учебник / Ю. С. Харин, Н. М. Зуев, Е. Е. Жук. - Минск: БГУ, 2011. – 463 с.

[8] Julius O. Smith III. Mathematics of Discrete Fourier Transformation (DFT) with audio applications. – W3K Publishing, 2007. - 322 p.

ПРИМЕНЕНИЯ МАШИННОГО ОБУЧЕНИЯ ПРИ УПРАВЛЕНИИ ТРАЕКТОРИЕЙ СТУДЕНТОВ

Ф.М. АЛИМОВА

Старший преподаватель ТУИТ

Ташкентский университет информационных технологий имени Мухаммада ал Хоразмий.

Аннотация. Успешное и продуктивное развитие цифровой экономики страны является ключевым фактором устойчивого развития, роста производства во всех сферах социально-экономической деятельности, что повышает конкурентоспособность страны, качество жизни граждан, обеспечивает экономический рост и национальный суверенитет. В настоящее время современное профессиональное образование переходит на качественно новый уровень в связи с внедрением компетентностного подхода, целью которого является предоставление обучающимся инструментов как понимания, так и действия, позволяющих воспринимать новые социально-экономические реалии, а также ориентироваться в изменяющихся условиях обучения и работы.

Ключевые слова: образование, методы обучения, интеллектуальный анализ, оценивание, квалификация, компетентность, процесс обучения, инновации.

UDC 004.032.26

STATIC SIGNATURE VERIFICATION BASED ON MACHINE LEARNING

U.Yu. Akhundjanov
phd student at the United Institute of Informatics Problems, National Academy of Sciences of Belarus

V.V. Starovoitov
doctor of engineering sciences, professor, chief researcher UIIP NAS of the Republic of Belarus

United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Republic of Belarus E-mail: umidjan_90 @ mail.ru

U.Yu. Akhundjanov

Graduated from Tashkent University of Information Technologies named after Muhammad al-Kharazmiy Fergana branch. PhD student at the United Institute of Informatics Problems of the National Academy of Sciences of Belarus. Conducts research on off-line handwritten signature verification.

V. V. Starovoitov

Doctor of Sciences and professor of computer science. He is a Principal research fellow at the United Institute of Informatics Problems, National Academy of Sciences of Belarus (UIIP NAN Belarus). Award: The State Prize of the Republic of Belarus in science. Research interests of professor Starovoitov are processing and analysis of digital images obtained in different parts of the electromagnetic spectrum. He has published over 150 papers.

Abstract. This paper describes the results of handwritten signature recognition. A handwritten signature database of 40 people made on paper and a publicly available Bengali handwritten signature database of 100 people were used for the experiments. A handwritten signature database of 40 people was collected with 10 authentic and 10 fake signatures for each person made by other people. A Bengali handwritten signature database of 100 people was collected 24 authentic and 30 forged signatures for each person. For this experiment, 20 people were randomly selected from the Bengal Handwritten Signature Database. Four options were used to reduce the signatures to sizes: 200×120, 250×150, 300×150, and 400×200 pixels for classification. These images served as input data for the proposed network architecture.

As a result of testing the proposed approach, the average accuracy of correct classification for the first base of handwritten signatures reached 90.04%. For the base of Bengal handwritten signatures 97.50%.

Keywords: Recognition, verification, handwritten signature, classification, FRR, FAR.

Introduction.

Handwritten signatures are an undeniable and unique way of confirming a person's identity. Because of its simplicity and uniqueness, it occupies an important place in the field of behavioral biometrics. Signatures are the most widely used biometric attribute, they are widely used in many banks, business transactions and documents that are approved with signatures and therefore secure authentication becomes an imperative.

Biometrics by the type of biometric parameters used are divided into two types into physiological and behavioral, where physiological features include facial shape, fingerprint, iris, retina, DNA. [1, 2, 7, 8], behavioral biometrics include handwritten signature, gait, voice. [6, 9].

With the development of technology today, there are a large number of financial transactions that need to be verified for authenticity. Today, most institutions actively use traditional signature verification methods. For the most part, traditional methods are manual and require experienced professionals for this purpose. Manual verification is time consuming and is a completely subjective process which depends greatly on the experience of the specialist verifying the signature in question. Biometrics plays an important role in development of a modern automatic identification and verification method [10].

Handwritten signature identification can be done statically in online mode and dynamically in off-line mode. Static or off-line signature recognition is performed after its image on paper has been digitized. The digital images are then transformed and analyzed [3]. In dynamic or online recognition systems the analysis begins during its creation. Additionally, information about the sequence of x- and y-coordinates of the signature points, information about the pressing force, writing speed etc. is collected. The static mode of signature verification has fewer informative features, which makes its process more complicated [11].

Many different approaches have been proposed to solve this problem. The accuracy of their recognition was tested on publicly available datasets, such as GPDS960, GPDS-4000, MCYT, BHSig260 and CEDAR, etc. All of these datasets contain three groups of signatures, genuine, random and qualified fakes.

The use of neural network technology helps to verify signatures more accurately. This is because neural networks effectively build non-linear dependencies, which describe the data more accurately, they are more robust to noise in the input data, and adapt to changes in the data. Reviews of these works are given in [3-6].

The authors of [12] proposed a method for static signature verification based on a convolutional neural network. They have investigated, that in the process of signature verification the manually created features have no or very little resemblance to the signature. The authors reported that convolutional neural networks produce more relevant features than manually created features. This paper used publicly available GPDS, PUC-PR datasets to evaluate the effectiveness of the method. They stated that their approach achieved the lowest EER (ratio of falsely accepted fakes to total fakes), but there was an imbalance between the false positive rate (FPR) and false negative rate (FNR). The authors later extended their work [11] and analyzed the deeply studied features that were extracted in [12]. They investigated different architectures and reported the lowest EER in the literature on the GPDS dataset.

The authors of [13] in their paper applied a Siamese convolutional network architecture for signature verification. A Siamese network has two identical networks with common weights, the same parameters and configuration, which accept different pairs of images as input. A Siamese network has two identical networks with common weights, identical parameters and configuration that take different pairs of images as input. These two networks are connected using a contrast loss function. According to the loss function, the similarity score between the two images is computed using the Euclidean distance, during back propagation the parameters are updated in the same way in both networks. The network was trained to reduce the distance between the "genuine - genuine" pair and increase the distance between the "genuine - fake" pair. The authors evaluated their method on completely different datasets, e.g., BHSig260, GPDS, CEDAR. But this method requires a large amount of time and high computational power, since two networks are trained simultaneously.

For estimation of efficiency of recognition and verification such indexes are used, as an error of the first kind FRR (ratio of the number of incorrectly rejected authentic signatures to the total number of authentic signatures), an error of the second kind FAR (ratio of the number of incorrectly accepted fakes to the total number of fakes) and measure EER - the level of equal probability of errors, at which FAR and FRR are equal [14].

FAR and FRR are determined by the formulas:

$$FAR = FPR = \frac{FP}{FP + TN},$$ FPR = False positive rate;

$$FRR = FNR = \frac{FN}{FN + TP}$$, FNR = False negative rate;

FP (***False positive***) - False positive solution, also called 1st kind error. The model predicted a positive result, but in fact it is negative;

TP (***True positive***) - a true positive solution. The model predicted a positive outcome, the prediction matched reality;

FN (***False negative***) - False negative decision, also called 2nd kind error. The model predicted a negative result and in fact it was positive;

TN (***True negative***) - a true negative solution. The model predicted a negative result, the prediction matched reality;

To evaluate the classification of our model, we used a function (Accuracy). The authors of the article [10] believe that the Accuracy function determines the share of correct answers and can be briefly translated as correctness or accuracy. When the number of objects of both classes is equal, this function can be used to estimate the classification results.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Preparation of data for handwritten signature recognition on images.

Two handwritten signature databases were used as experimental data for training the handwritten signature recognition system, one of which contained 800 handwritten signature images of 40 people. The database contained 10 authentic and 10 fake signatures of each person. Figure 1 shows examples of handwritten signatures for the first database.

This database of handwritten signatures was collected with the help of students at the Fergana branch of the Muhammad al-Khwarizmi Tashkent University. The signature samples were scanned at 800 dpi (dots per inch) and each signature was cut at 850×550 pixels. Figure 2 shows examples of Bengali handwritten signatures for the second base. A Bengali handwritten signature database of 100 people was collected with 24 authentic and 30 fake signatures for each person. For this experiment, 1,080 handwritten signatures of 20 people were randomly selected from the Bengal handwritten signature database.

The images of the handwritten signatures were converted to halftone and then to binary. For this purpose, a method of Otzu was used. This method is used to calculate a threshold t that minimizes the average segmentation error, i.e., the average error from deciding whether image pixels belong to an object or background [15-16].

Applications of a convolutional neural network.

A convolutional neural network is a very broad class of architectures, the main idea of which is to reuse the same parts of the neural network to handle different small local sections of inputs.

To distribute the image classes, directories were created, with two subdirectories created in each directory, according to the names of the classes: genuine and forced.

Experiments were performed with the reduction of captions to 200×120, 250×150, 300×150, and 400×200 pixels.

Figure 1. Examples of handwritten signatures for experiments

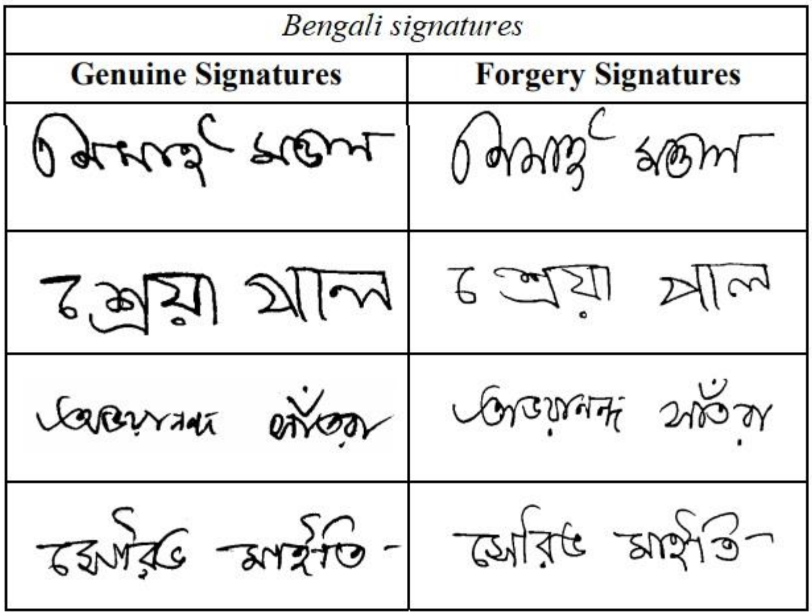

Figure 2. Examples of Bengali handwritten signatures for experiments

The architecture of the convolutional neural network.

The deep learning model used to produce the results is described below:

1. Convolution layer, kernel size 3x3, number of feature maps - 32 pieces, ReLU activation function.
2. Sub-sample layer, maximum value selection from 2x2 square.
3. The convolution layer, kernel size 3x3, number of feature cards - 32 pieces, ReLU activation function.
4. Layer of subsample, maximum value selection from 2x2 square.

5. The convolution layer, kernel size 3x3, number of feature cards - 64 pieces, ReLU activation function.
6. Layer of subsample, maximum value selection from 2x2 square.
7. Layer of conversion from two-dimensional to one-dimensional representation.
8. Full-link layer, 64 neurons, ReLU activation function.
9. Dropout layer. This is a thinning method which is used to average the training results.
10. Output layer, 1 neuron, sigmoid activation function.

Layers 1 to 6 are used to select important features in the image, and layers 7 to 10 are used to evaluate the classification result.

Results.

To train, validate and test the model, 800 handwritten signature images were used for the first base in an 8:1:1 proportion, respectively. Half of them were images of genuine signatures and the other half were images of fake signatures. For the second base, 1080 images of Bengali handwritten signatures in the proportion of 21:4:2, respectively. The computational experiment was performed on the https://colab.research.google.com/ platform.

Figure 3. Training and validation graph with 250x150 image resolution for first base

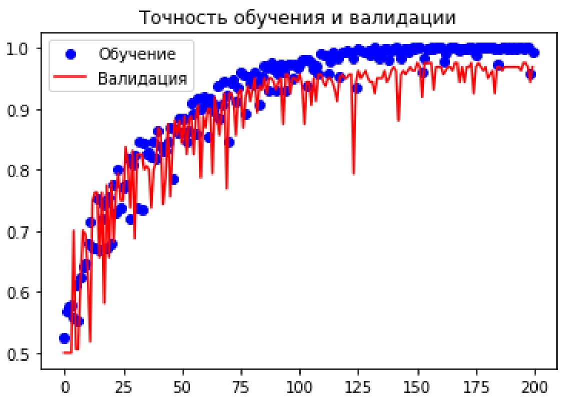

Figure 4. Training and validation graph with image resolution for the Bengali 250x150 base

Table 1. Results of signature recognition

Handwritten Signature Bases	The correctness of recognition with a 200x120 extension	The correctness of recognition with a 250x150 extension	The correctness of recognition with a 300x150 extension	The correctness of recognition with a 400x200 extension
Base 1	88,31	**90,04**	89,12	88,74
Base 2 (Bengali)	94,48	**97,50**	96,40	95,65

Table 1 shows the results of the experiments. The trained neural network model showed the best result in both bases at handwritten signature resolution of 250x150 pixels.

In order to create a handwritten signature recognition system, several programs were developed in Python using deep learning models. The work of this software can be divided into several stages: preparation of the dataset, image acquisition with simultaneous preprocessing, training on the collected data through the prepared learning model. The results of this experiment can be found on GitHub.com [17].

Conclusion.

Off-line signature verification is inferior to on-line technology in accuracy. The results of the experiments described in the article have shown that the approach to handwritten signature verification is promising.

The average accuracy of correct classification of signatures was achieved for the first base on images of size 250x150, and is equal to 90.04%, for the second base on images of size 250x150, and is equal to 97.50%. In the future, it is planned to improve the algorithm and increase the recognition accuracy, as well as to form a larger sample size. The main direction of further research will be the allocation of informative features that allow high recognition accuracy.

References

[1] Старовойтов В.В., Ю. Голуб. Обработка изображений радужной оболочки глаза для систем распознавания. Минск: LAP LAMBERT Academic Publishing, 2018. – 188c.

[2] Chaudhry, S. A. An enhanced lightweight anonymous biometric based authentication scheme for TMIS / S. A.Chaudhry, H. Naqvi, M. K. Khan // Multimedia Tools and Applications - 2017, 22 p. DOI:10.1007/s11042-017-4464-9.

[3] Hafemann, L.G. Offline handwritten signature verification — Literature review / L.G. Hafemann, R. Sabourin, L.S. Oliveira // Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA) – 2017. 8p. DOI:10.1109/ipta.2017.8310112.

[4] Hadeel J.Jriash. Offline handwritten signature verification system using neural network / J.Jriash Hadeel, A. Z. Abdullah Nada // International Journal of Computer Science and Mobile Computing. – 2015. Vol.4, Issue.10.– P. 403-412.

[5] Impedovo S. Verification of Handwritten Signatures: an Overview / S. Impedovo, G. Pirlo // 14th International Conference on Image Analysis and Processing. – 2007. – P.191-196. DOI:10.1109/iciap.2007.4362778.

[6] Foroozandeh, A. Offline Handwritten Signature Verification and Recognition Based on Deep Transfer Learning / A. Foroozandeh, A.H. Ataollah, H. Rabbani // International Conference on Machine Vision and Image Processing. – 2020, 7p. DOI:10.1109/mvip49855.2020.918748.

[7] De Marsico, M. Iris recognition through machine learning techniques: A survey / M. De Marsico, A. Petrosino, S. Ricciardi // Pattern Recognition Letters – 2016, 14 p. DOI:org/10.1016/j.patrec.2016.02.001.

[8] Sharma S. Identity verification using shape and geometry of human hands / S. Sharma, S. R. Dubey, S. K. Singh, R. Saxena, R. K. Singh // Expert Systems with Applications – 2015. –P. 821–832. DOI: 10.1016/j.eswa.2014.08.052.

[9] Wan C. A Survey on Gait Recognition / C. Wan, L Wang, V. V. Phoha // ACM Computing Surveys. - 2018, 35p. DOI:10.1145/3230633.

[10] Ferrer, M. A. Robustness of Offline Signature Verification Based on Gray Level Features / M.A. Ferrer, J. F. Vargas, A. Morales, A.Ordonez // IEEE Transactions on Information Forensics and Security – 2012. – Vol.7, Issue.3.– P. 966–977. DOI:10.1109/tifs.2012.2190281.

[11] Hafemann L.G.Analyzing features learned for offline signature verification using deep cnns / L.G. Hafemann , R. Sabourin, L.S. Oliveira // 23rd international conference on Pattern recognition (ICPR). IEEE – 2016. – P. 2989–2994. DOI:10.1109/icpr.2016.7900092.

[12] Hafemann L. G. Writer-independent feature learning for Offline Signature Verification using Deep Convolutional Neural Networks / L.G. Hafemann, R. Sabourin, L.S. Oliveira // International Joint Conference on Neural Networks (IJCNN) – 2016. –P. 2576–2994. DOI:10.1109/ijcnn.2016.7727521.

[13] Jagtap, A. B. Siamese Network for Learning Genuine and Forged Offline Signature Verification / A. B. Jagtap, D. D. Sawat, R. S. Hegadi // Recent Trends in Image Processing and Pattern Recognition – 2019. – P. 131–139. DOI:10.1007/978-981-13-9187-3_12.

[14] Starovoitov V. V., Golub Y. I. Comparative study of quality estimation of binary classification. Informatics. – 2020. – Vol. 17, no. 1, P. 87−101 (in Russian).

[15] Исрафилов, Х.С. Исследование методов бинаризации изображений / Х.С. Исрафилов // Вестник науки и образования. – 2017. –Т.2.- № 6(30). − С. 43–50.

[16] Янковский, А.А. Критерии выбора метода бинаризации при обработке изображений лабораторных анализов. АСУ и приборы автоматики [Электронный ресурс]. – Режим доступа: https://cyberleninka.ru/article/n/kriterii-vybora-metoda-binarizatsii-pri-obrabotke-izobrazheniy-laboratornyh-analizov/viewer. – Дата доступа: 25.12.2021.

[17] Akhundjanov U.Yu. My_signature_verifiction / U.Yu. Akhundjanov // https://github.com [Electronic resource]. – 2022. Mode of access: https://github.com/MrUmidjan90/My-signature verification/blob/main/Bingali.ipynb– Date of access: 27 February 2022.