

Программные системы: теория и приложения

Двуязычный электронный научный журнал

№ 1  2025

Bilingual Online Scientific Journal

Program Systems: Theory and Applications

Том 16 Выпуск 1(64) 2025 г.

СОДЕРЖАНИЕ

Научная статья

ПРИКЛАДНЫЕ ПРОГРАММНЫЕ СИСТЕМЫ

Винокуров И. В.[✉]. *Использование модели Mask R-CNN для сегментации объектов недвижимости на аэрофотоснимках (Англ., Рус.)* .. 3–23, **24–44**

Научная статья

ПРОГРАММНОЕ И АППАРАТНОЕ ОБЕСПЕЧЕНИЕ РАСПРЕДЕЛЕННЫХ И СУПЕРКОМПЬЮТЕРНЫХ СИСТЕМ

Бавенко В. Н.[✉], Невечеря А. П.. *О реализации QR-разложения на трехмерном систолическом массиве* **45–56**, 57–59

Научная статья

МЕДИЦИНСКАЯ ИНФОРМАТИКА

Грибова В. В., Шуматов В. Б., Лебедев С. В., Шалфеева Е. А., Шестопапов Е. Ю., Окунь Д. Б.[✉], Ковалев Р. И., Шепета Е. И., Федорищев Л. А., Лифшиц А. Я.. *Комплекс онтологий как модель системы интеллектуальной поддержки в реабилитации пациентов, перенесших инсульт* **61–79**, 80–82

Научная статья

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ

Хубиев К. Ю.[✉], Семенов М. Е.. *Мультимодальное предсказание цен акций на примере российского рынка ценных бумаг (Англ., Рус.)* 83–106, **107–130**

Авторский указатель **131**

Click the flag at a top corner of any page to switch the language, please!

Author index 133

Contents 135



Using the Mask R-CNN model for segmentation of real estate objects in aerial photographs

Igor Victorovich **Vinokurov**

Financial University under the Government of the Russian Federation, Moscow, Russia

Abstract. The mass appearance of illegal and unregistered in the Unified State Register of Real Estate (USRRE) real estate objects complicates cadastral registration for many entities at the territorial and administrative levels. Traditional methods of identifying objects of this type, based on manual analysis of geospatial data, are labor-intensive and time-consuming.

To improve the efficiency of this process, it is proposed to automate the detection of objects in aerial photographs by solving the instance segmentation problem using the Mask R-CNN deep learning model. The article describes the preparation of a dataset for this model, examines the main quality metrics, and analyzes the results obtained. The efficiency of the Mask R-CNN model in practice is shown for solving the problem of detecting construction projects that are not registered in the USRRE. (*Linked article texts in English and in Russian*).

Key words and phrases: Cadastral registration, aerial photography analysis, instance segmentation, Mask R-CNN, PyTorch

2020 *Mathematics Subject Classification:* 68T20; 68T07, 68T45

For citation: Igor V. Vinokurov. *Using the Mask R-CNN model for segmentation of real estate objects in aerial photographs*. Program Systems: Theory and Applications, 2025, **16**:1(64), pp. 3–44. (*In English, in Russian*). https://psta.psiras.ru/read/psta2025_1_3-44.pdf

Introduction

Cottage settlements, dacha cooperatives and gardening associations have a high density of development, which complicates the automatic detection of illegally built real estate objects and objects not registered in the Unified State Register of Real Estate on aerial photographs (USRRE).

For the implementation of aerial photography, the Roskadastr PPC uses quadcopters. The resulting photographs are used to search for real estate objects subject to mandatory registration. Such objects include:

- country houses intended for permanent residence,
- objects standing on a foundation (baths, terraces, summer kitchens),
- buildings on plots for individual housing construction.

Manual search limits the efficiency and accuracy of this process. With the development of deep learning models such as RetinaNet [1], SOLO [2], YOLACT [3], Mask R-CNN [4] and others, it becomes possible to significantly reduce the time and improve the quality of image analysis due to automatic detection of objects. The models listed above are a powerful tool for implementing instance segmentation, which allows not only to identify objects in an image, but also to accurately highlight their boundaries. This is especially important for tasks related to the analysis of dense buildings, where objects can be partially overlapped or located in close proximity to each other. The use of instance segmentation implemented by these models is best suited for analyzing geospatial data, since the presence of masks and bounding boxes around the found object instances allows not only to anticipate the presence of search objects, but also to estimate their sizes. The latter can be useful when assessing changes in the size of objects associated with their reconstruction.

This paper presents the results of a study of the applicability of the Mask R-CNN (Mask Region-based Convolutional Neural Network) model for detecting real estate in cottage settlements, summer cottage cooperative and gardening associations. The peculiarity of the proposed approach is the processing of aerial photographs in order to identify the bounding frames of the detected objects, compared with certain geographic coordinates, and the comparison of their presence or significant change in size with the results of similar processing of aerial photographs for the previous period, usually a year. This approach allows us to identify changes in development and record buildings that may not have official registration in USRRE.

1. Overview of works on object recognition in aerial photographs using Mask R-CNN

Mask R-CNN is an extension of the Faster R-CNN model, designed to solve the problems of object detection and image segmentation [5]. A distinctive feature of Mask R-CNN from Faster R-CNN is the creation of accurate masks for each of the detected objects in the image. The model consists of two main components – object detection and object segmentation [4].

To extract features from the input image, Mask R-CNN uses the ResNet architecture [6] or another convolutional neural network. These features are then passed to a Region Proposal Network (RPN) [4, 6], which generates Regions of Interest (RoI) – predicted object locations. All obtained RoIs are passed to the main classifier, which determines the class of the object and specifies its boundaries.

Unlike Faster R-CNN, Mask R-CNN has an additional branch responsible for creating segmentation masks. This branch uses a small convolutional network to generate binary masks for all objects in the RoI. The mask size corresponds to the size of the RoI.

The model is trained using a loss function that combines the losses from classification, coordinate regression, and segmentation, which allows the model to simultaneously optimize all three tasks [4].

Mask R-CNN copes well with occluded objects and complex backgrounds due to its ability to generate accurate masks. However, its performance may depend on the quality of the training dataset and the chosen model architecture. Overall, Mask R-CNN is a strong tool for object detection and segmentation tasks in images, providing high accuracy and flexibility in application.

Currently, there are quite a lot of works on recognizing objects of various types in photographs and aerial photographs using this model, which indicates the effectiveness of its application for solving practical problems.

The paper [4] presents the basic architecture of Mask R-CNN and its application to various segmentation problems, including instance segmentation of objects.

In [7], the authors apply Mask R-CNN to detect buildings in high-resolution satellite images. High segmentation accuracy and advantages of using the Mask R-CNN model for solving such problems are demonstrated.

Detection of buildings in photographs from natural disaster zones using the Mask R-CNN model and pre-processing of the dataset, in order to improve the efficiency of its use, is given in [8]. The paper demonstrates the advantages of this model over other models that implement object detection in photographs.

The paper [9] solves the problem of localizing building polygons in high-resolution satellite images. The efficiency of the Mask R-CNN model for obtaining real building contour boundaries in photographs with high object density is demonstrated.

A method for detecting and segmenting ships at the pixel level using the Mask R-CNN model is proposed in [10]. The advantages of using this model are noted and the efficiency of the proposed method is evaluated.

In [11], it is shown that extracting building contours from satellite images is a complex task due to differences in scale, structure and types of buildings. To solve this problem, it is proposed to use the Mask R-CNN model. The efficiency of its use is demonstrated. The results of extracting individual buildings from satellite images are proposed to be used for applications that automate population assessment, implement urban planning and others.

Detection of modern urban architecture using the Mask R-CNN model trained on a dataset consisting of elements of modern architectural styles is given in [12]. As a result of comparing the obtained results, the efficiency of using Mask R-CNN is shown in comparison with other models.

The work [13] shows the difficulty of detecting buildings and various types of structures on satellite images due to illumination, building density, different types of terrain and other factors. To effectively solve this problem, it is proposed to use the Mask R-CNN model and our own dataset with improved image augmentation.

The use of the Mask R-CNN model for recognizing various objects on satellite images and aerial photographs for the purpose of automating mapping and maintaining local maps up to date is described in the work [14].

In [15], the Mask R-CNN model is proposed to be used to maintain the accuracy of terrain maps for effective response to natural disasters and catastrophes. It is proposed to update terrain maps based on aerial and satellite images. The work obtained more than acceptable results for photographs of different quality, resolutions and color channels.

In [16, 17], a hybrid approach to extracting building contours from low-resolution satellite images using the Mask R-CNN model is proposed. This approach opens up prospects for the development of automated tools for processing satellite images, and their effective use in land use monitoring and disaster response.

2. Justification for the choice of the model

Based on the above review, the Mask R-CNN model can be reasonably chosen for implementing instance segmentation of real estate objects in aerial photographs, since it has the following advantages.

(1) Accuracy

- The Mask R-CNN model demonstrates higher accuracy in detecting objects compared to other models [9, 10].
- The two-stage approach of the Mask R-CNN model (region proposal and segmentation [4–6]) allows for more accurate localization and segmentation of objects while simultaneously generating their masks.

(2) Versatility

- Mask R-CNN can be easily adapted to solve a variety of problems, including object detection and contouring, instance segmentation and semantic segmentation [4–6].
- The modular architecture of this model allows you to easily add or remove components as needed.

(3) Robustness

- Mask R-CNN is more robust to noise and distortion in images [15].
- The two-stage approach of this model helps reduce the number of false positives and improves the overall robustness [4–6].

(4) Support for objects of different sizes

- Mask R-CNN can effectively segment objects of different sizes, from small to large [4, 8].

- The region proposal mechanism [4] allows this model to detect objects regardless of their type and size (e.g. COCO [5]).
- (5) Less overfitting issues
- Mask R-CNN is less prone to overfitting [11].
 - Using multiple loss functions and a two-stage approach helps the model to generalize better [4–6, 18]

Mask R-CNN models also have some drawbacks that are not critical when solving the problem of automating image processing, these are relatively low speed of operation and complexity of implementation [19].

In addition to the models discussed above, the YOLO model [20], which has become quite popular recently, can be used to solve the instance segmentation problem. Despite the high speed of obtaining results, this model has significant drawbacks – low quality of recognition of groups of small objects due to the limited number of candidates for bounding boxes (two) and the possibility of duplicating bounding boxes for the same object [21]. A comparison of the results of real estate recognition with similar results obtained using the latest version of the YOLO model is given in p. 6

3. Purpose of the work

As noted above, photographing the territories under study by the «Roscadastr» PLC control center is carried out using quadcopters. The quadcopter (at the time of writing, it is Phantom 4 RTK) implements movement along a predetermined trajectory. The camera automatically takes pictures upon reaching certain points along the route, forming either its full image or images of route fragments (in most cases, with overlap).

The purpose of this work is to study the applicability of the Mask R-CNN model for the implementation of automatic or automated processing of the obtained images and the identification of illegally built or unregistered real estate objects on them. To achieve the stated goal, the following tasks were solved in the work:

- (1) creation of our own dataset for training the model,
- (2) analysis of the quality of the model,
- (3) analysis of the results of using the model to detect illegal construction sites or real estate objects not registered in the USRRE.

TABLE 1. Object types and their labels

Object type (class)	Label name and class segmentation color
Cottage or summer house	<i>building</i>
Greenhouse (hothouse)	<i>greenhouse</i>
Outbuilding	<i>outbuilding</i>
Vehicle	<i>vehicle</i>
Swimming pool	<i>swimming</i>

4. Dataset creation

4.1. Image annotation

The main task in generating a dataset used for training and testing the model is labeling or annotating the original images, which is a polygonal contouring of objects recognized in photographs. The accuracy of the contouring determines the accuracy of object detection.

Currently, there is a large amount of software that implements image annotation. To generate the dataset, open source software LabelMe was selected.

The annotation process involves assigning a label to each recognizable object of a particular type. When annotating aerial photographs, labels were assigned to the following 5 types of objects, Table 1.

The main type of objects is a cottage or a summer house. Other types of objects are used to prevent false recognition of the latter and for the purpose of possible continuation of work in this direction.

The annotation results were saved in the JSON format supported by LabelMe. When forming the dataset, 435 photographs obtained from a quadcopter were annotated. An example of a photograph of a fragment of a summer house cooperative and the polygonal contouring of 2 classes of objects – summer house and greenhouse in LabelMe implemented during the annotation process are shown in Figure 1, 2a and 2b, respectively.

The dataset formed for training and testing the model is a collection of the above number of aerial photographs and the same number of JSON files in the LabelMe software format, containing arrays of coordinates of polygonal contouring points and names for objects of 5 types.

For training and testing the model, 80% and 20% of the elements of this dataset were randomly selected, respectively. The distribution



FIGURE 1. Photograph of a fragment of a summer cottage cooperative taken by a Phantom 4 RTK quadcopter



(a) A summer cottage and a greenhouse (b) Contouring of objects of 2 types

FIGURE 2. Dataset element – image of objects and their polygonal contouring on a fragment of an image

diagram of objects belonging to these classes on aerial photographs is shown in Figure 3.

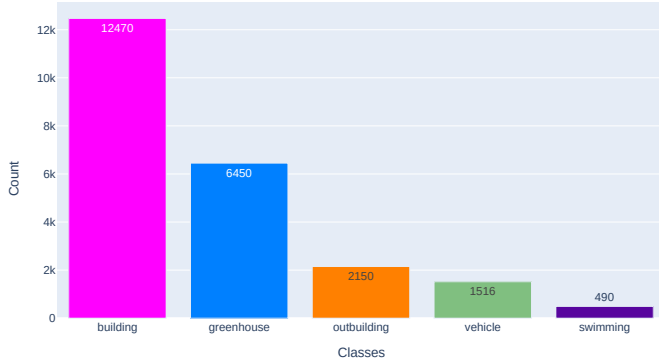


FIGURE 3. Distribution of classes in the dataset photographs

4.2. Dataset augmentation

To improve the generalization ability of the model, a pipeline of transformations implemented in the `PyTorch` package was formed and applied to the images from the training dataset. At the very beginning, based on the IoU (Intersection over Union) metric, images were cropped while preserving the ROI. Then, stochastic color transformations were applied, including variations in brightness, contrast, saturation, and hue, as well as random grayscale transformation (`RandomGrayscale`) and histogram equalization (`RandomEqualize`).

Further increase in diversity consisted in reducing the number of color levels and horizontal reflection of images using the methods `RandomPosterize` and `RandomHorizontalFlip` respectively. In addition, scaling of images to the maximum size while maintaining the aspect ratio, padding to a square shape, and resizing to the required size using antialiasing were implemented. The final step was to convert the data type to `torch.float32` with scaling of pixel values and validation of bounding box coordinates (`SanitizeBoundingBoxes`).

5. Model creation and exploration

5.1. Model creation

Mask R-CNN is characterized by a relatively high complexity of implementation. As a result, successful application of this model requires careful tuning of the parameters and network architecture. A fairly fast and

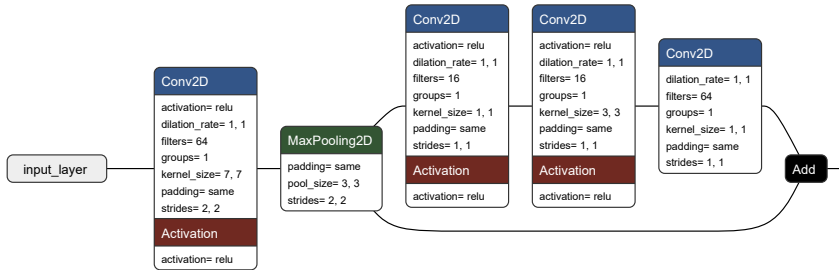


FIGURE 4. The initial convolutional layer in ResNet-101 and the first residual block [22]

accurate object detection model `maskrcnn_resnet50_fpn_v2` from the `torchvision` package was chosen in this work. To improve the efficiency of feature extraction (backbone), the ResNet-101 architecture with Feature Pyramid Network (FPN) [4] was chosen in this model. The general principles of its formation and training are described quite well in [22] and [23].

The model is pre-trained on the COCO dataset [5]. For this model, the optimal number of epochs for additional training was experimentally found, the optimizer `Adam` and the scheduler `OneCycleLR` (the mechanisms that determine the decrease in weights during training and the speed of this process) were selected, the number of output channels was changed, and the partitioning for anchor boxes – pre-defined rectangular frames that are used to propose potential RoIs when detecting objects in an image was configured. The code for generating and exploring the model was written using the PyTorch machine learning library [24].

The initial convolutional layer of ResNet-101 was modified to efficiently implement instance segmentation of objects of different scales. Several parallel convolutional layers with kernels of different sizes were added before the layer with a sufficiently large kernel size (7×7), Figure 4 and Figure 5.

This potentially allows the model to simultaneously take into account both contextual features (kernel 5×5) and detailed characteristics of objects (kernels 1×1 and 3×3). In addition, parallel convolutional layers allow accelerating the convergence of the model.

Attempts to vary the residual blocks of this model did not lead to any significant improvement in object detection.

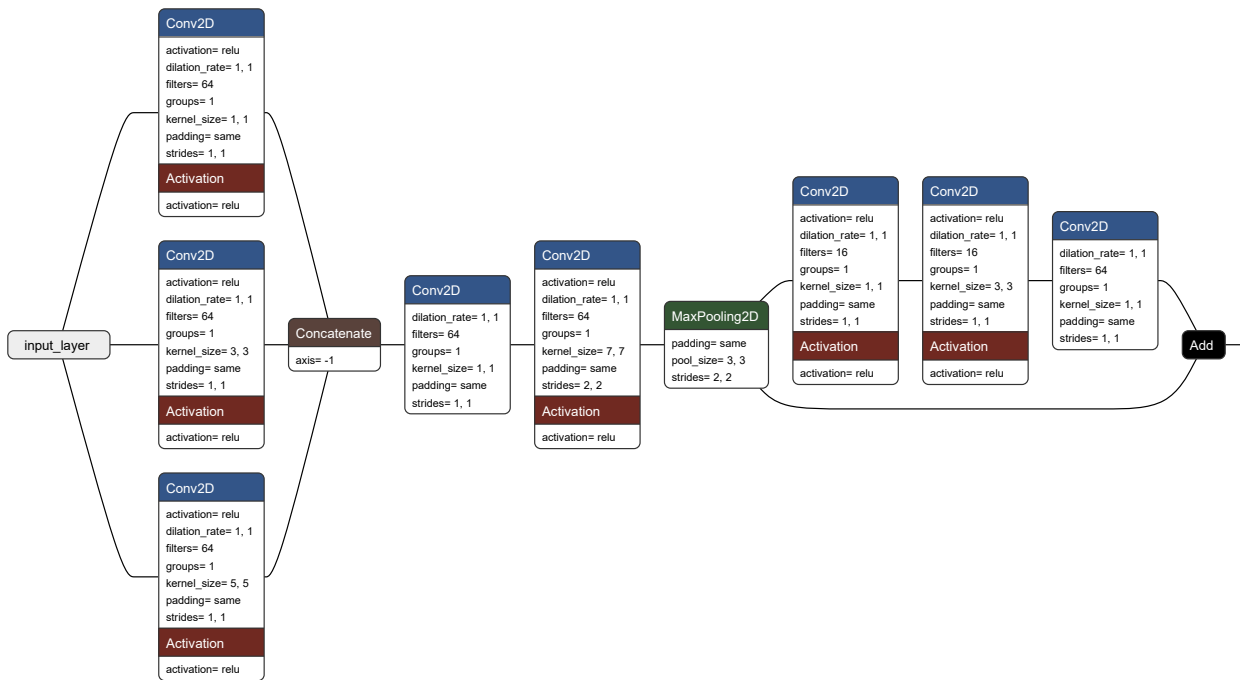


FIGURE 5. New parallel convolutional layers in ResNet-101

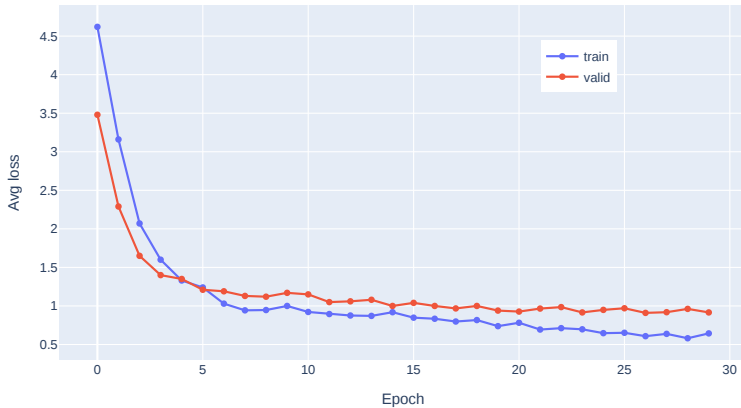


FIGURE 6. Change in the mean value of the model loss function

5.2. Model exploration

To analyze the accuracy of the model during the experimental studies, 3 key metrics were calculated – **Loss**, **mAP** and **CS**.

Loss in Mask R-CNN allows you to evaluate how accurately the model predicts classification, coordinate regression and object segmentation. The closer the values of this metric to 0 (0%), the more accurate the model's result. The change in the average value of this metric for all 5 classes from the training and testing (valid) datasets depending on the training epoch is shown in Figure 6.

The mAP (mean Average Precision) metric is the average value of the precision metric over all classes of objects. A similar metric mAP50-95 is calculated for different thresholds of overlapping bounding boxes IoU (Intersection over Union) – from 50% to 95%. Both metrics take into account both precision (**precision**) and recall (**recall**) at different IoU levels. The closer the values of each of these metrics are to 1 (100%), the more accurate the model's output is. The dependence of the mAP and mAP50-95 metrics for all classes from the training and valid datasets on the training epoch is shown in Figure 7.

The Confidence Score (CS) metric represents the probability that an object belongs to a certain class. CS is a measure of the model's confidence that the predicted object is actually present in the bounding

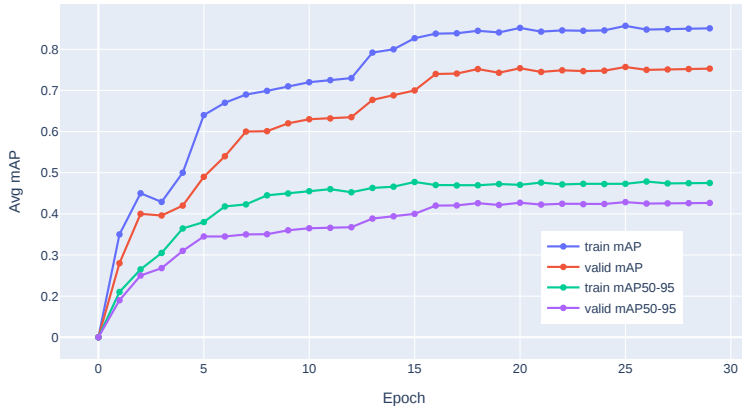


FIGURE 7. Change in the average value of mAP and mAP50-95 metrics in bounding box detection

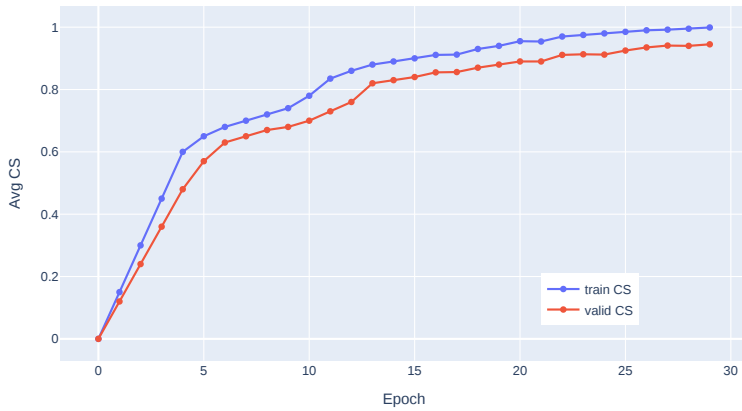


FIGURE 8. Change in the average value of the CS metric during object detection

box. This metric is one of the most important for models implementing object detection. CS values range from 0 to 1 (100%); higher values indicate greater model confidence in the correctness of its prediction. The dependence of this metric for all classes from the training and valid datasets on the training epoch is shown in Figure 8.

TABLE 2. Accuracy metric values at the last epoch of model training

Dataset	Loss	mAP	mAP50-95	CS
train	0.741	0.851	0.475	0.999
valid	0.915	0.753	0.416	0.991



FIGURE 9. Detecting real estate objects in images from the test dataset

Table 2 shows the values of the accuracy metrics for the training and test datasets at the last epoch of model training, see Figure 6–8.

Calculating accuracy metric values in [Google Colab Pro](#) in the T4 GPU runtime took about 80-100 minutes. From the graphs and table above, it follows that the Mask R-CNN model trained on its own dataset has more than acceptable accuracy in detecting real estate objects of interest to us in aerial photographs. Several examples of detecting summer houses from the datasets for training the model and for testing it are shown in Figure 9, 10, respectively.

When using this model, minor errors were detected in 3-5% of cases, manifested in the form of incorrect identification objects or incorrect segmentation of these boundaries. The occurrence of these errors is associated with the density of the segmented objects, with an insufficiently representative dataset, the presence of noise in the images and insufficient



FIGURE 10. Detection of real estate objects in images from the test dataset

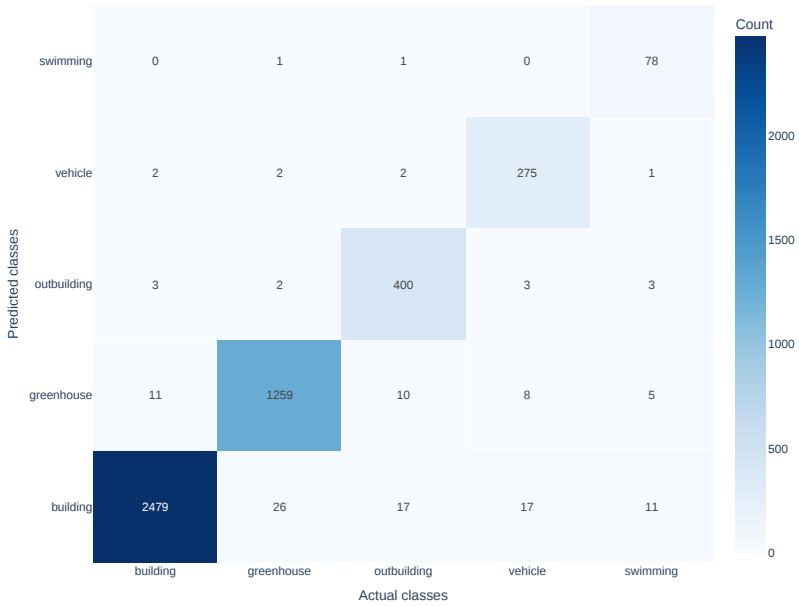


FIGURE 11. Model confusion matrix for the test dataset

sharpness of the image area. The matrix of errors in detecting real estate objects in images from the test dataset is shown in Figure 11.

Example incorrect identification for several objects of the *greenhouse*, *outbuilding* and *vehicle* classes is shown in Figure 12. It is easy to see that all objects of the *building* class, due to their large number in the training dataset, are identified correctly, which is quite sufficient for the above stated purpose of the work.

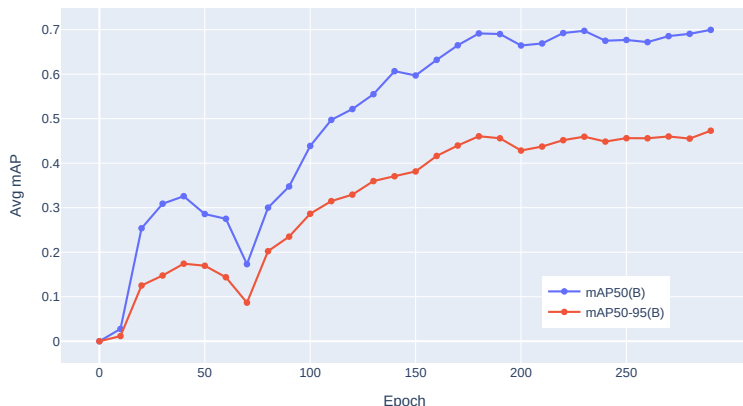


FIGURE 13. Change in the average value of the mAP and mAP50-95 metrics of the YOLO model when detecting bounding boxes

Figure 12,14) definition of object masks by the Mask R-CNN model was noted compared to YOLO. Incorrect (inaccurate) definition of an object mask can be critical for the «Roscadastr» PLC information system.

Overall, the YOLO model from Ultralytics left a favorable impression with its ease of use and the availability of ready-made functionality for conducting experiments and analyzing the results obtained. In some cases, depending on the specifics of the problem being solved and the dataset, using YOLO models allows you to get a slightly better result compared to Mask R-CNN [25].

7. Practical implementation of the results

The results obtained in the course of work to improve the efficiency of detecting cottages and summer houses on aerial photographs using the Mask R-CNN model were implemented in the beta version of one of the subsystems of the of the «Roscadastr» PLC information system. The main goal of this subsystem is to determine the presence of registration of construction objects in the USRRE. This subsystem implements a process that includes orthotransformation, creating a digital terrain model (DTM), obtaining reference points and calculating the coordinates of real estate objects [26].

A photograph taken from a quadcopter is divided into a number of fragments (tiles), which are analyzed for the presence or absence

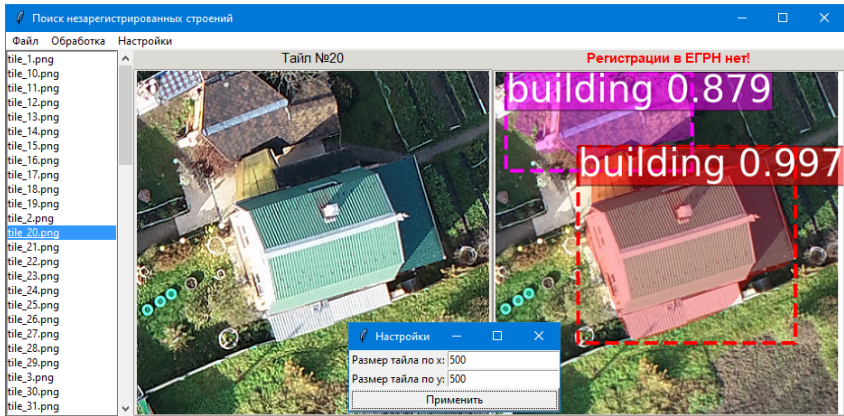


FIGURE 15. An example of detecting an unregistered summer house on an aerial photograph fragment

PLC information system based on Mask R-CNN uses this model to recognize objects in images obtained from a quadcopter, which significantly reduces the time and resources spent on identifying illegal and unregistered construction, and increases the efficiency of land use control.

Further research on the application of the Mask R-CNN model will be aimed at increasing its accuracy by expanding and optimizing the training dataset, as well as integrating this model into existing land use monitoring and control systems. In addition, a promising direction is the development of algorithms for automatic verification of the model's results based on comparison with the USRN data.

References

- [1] T.-Y. Lin, P. Goyal, R. B. Girshick, K. He, P. Dollár. *Focal loss for dense object detection*, Computing Research Repository (CoRR), 2017, 10 pp. [arXiv](#) 1708.02002
- [2] X. Wang, T. Kong, Ch. Shen, Y. Jiang, L. Li. *SOLO: Segmenting objects by locations*, Computing Research Repository (CoRR), 2019, 19 pp. [arXiv](#) 1912.04488
- [3] K. Duda, A. Ivanov. "On decidability of amenability in computable groups", *Archive for Mathematical Logic*, **61** (2022), pp. 891–902.
- [4] K. He, G. Gkioxari, P. Dollár, R. B. Girshick. *Mask R-CNN*, Computing Research Repository (CoRR), 2017, 12 pp. [arXiv](#) 1703.06870 4, 5, 7, 8, 12
- [5] S. Ren, K. He, R. B. Girshick, J. Sun. *Faster R-CNN: Towards real-time object detection with region proposal networks*, Computing Research Repository (CoRR), 2015, 14 pp. [arXiv](#) 1506.01497 5, 7, 8, 12

- [6] K. He, X. Zhang, S. Ren, J. Sun. *Identity mappings in deep residual networks*, Computing Research Repository (CoRR), 2016, 15 pp. [arXiv:1603.05027](#)  ⁵, ^{7, 8}
- [7] T.-Y. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick. *Microsoft COCO: Common objects in context*, Computing Research Repository (CoRR), 2014, 15 pp. [arXiv:1405.0312](#)  ⁶
- [8] Y. Xu, L. Wu, Z. Xie, Z. Chen. “Building extraction in very high resolution remote sensing imagery using deep learning and guided filters”, *Remote. Sens.*, **10**:1 (2018), id. 144, 18 pp.  ^{6, 7}
- [9] Q. Han, Q. Yin, X. Zheng, Z. Chen. “Remote sensing image building detection method based on Mask R-CNN”, *Complex Intell. Syst.*, **8** (2022), pp. 1847–1855.  ^{6, 7}
- [10] K. Zhao, J. Kang, J. Jung, G. Sohn. “Building extraction from satellite images using Mask R-CNN with building boundary regularization”, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (18–22 June 2018, Salt Lake City, UT, USA), IEEE, ISBN 9781538661017, id. 242, 4 pp.  ^{6, 7}
- [11] X. Nie, M. Duan, H. Ding, B. Hu, E. K. Wong. “Attention Mask R-CNN for ship detection and segmentation from remote sensing images”, *IEEE Access*, **8** (2020), pp. 9325–9334.  ^{6, 8}
- [12] M. Jenila Vincent, P. Varalakshmi. “Extraction of building footprint using MASK-RCNN for high resolution aerial imagery”, *Environmental Research Communications*, **6**:7 (2024), id. 075015, 17 pp.  ⁶
- [13] X. Zhu, L. Hu, J. Wang. “Urban modern architecture recognition based on Mask-RCNN and ECA attention mechanism”, Fifth International Conference on Geoscience and Remote Sensing Mapping (ICGRSM 2023) (13–15 October 2023, Lianyungang, China), Proc. SPIE, vol. **12980**, 2024, ISBN 9781510672789, id. 129801D.  ⁶
- [14] R. Raghavan, D. Chander Verma, D. Pandey, R. Anand, B. Kumar Pandey, H. Singh. “Optimized building extraction from high-resolution satellite imagery using deep learning”, *Multimedia Tools and Applications*, **81**:29 (2022), pp. 42309–42323.  ⁶
- [15] D. Ulanov, A. Syrov. “Building footprint extraction based on RGBD satellite imagery”, CS230 Deep Learning (Winter 2020, Stanford University, CA), 2020, 11 pp.  ⁷
- [16] A. Solanki, R. K. Singh, B. Demeneze. “Aerial pictures semantic segmentation applying deep learning”, *International Journal of Trendy Research in Engineering and Technology*, **5**:1 (2021), pp. 42–48.  ⁷
- [17] A. NourEldeen, M. E. Wahed. “Enhanced building footprint extraction from satellite imagery using Mask R-CNN and PointRend”, *Bulletin of Electrical Engineering and Informatics*, **5**:13 (2024), pp. 3601–3608.  ⁷
- [18] K. He, X. Zhang, S. Ren, J. Sun. *Deep residual learning for image recognition*, Computing Research Repository (CoRR), 2015, 12 pp. [arXiv:1512.03385](#)  ⁸

- [19] Ch. J. Mills. *PyTorch Mask R-CNN tutorial*, GitHub repository, 2023. [URL](#) ↑⁸
- [20] J. Redmon, S. Divvala, R. B. Girshick, A. Farhadi. *You Only Look Once: Unified, real-time object detection*, Computing Research Repository (CoRR), 2015, 10 pp. [arXiv](#) [1506.02640](#) [doi](#) ↑⁸
- [21] R. Khanam, M. Hussain. *YOLOv11: An overview of the key architectural enhancements*, 2024, 9 pp. [arXiv](#) [2410.17725](#) [doi](#) ↑⁸
- [22] T.-Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, S. J. Belongie. *Feature pyramid networks for object detection*, Computing Research Repository (CoRR), 2016, 10 pp. [arXiv](#) [1612.03144](#) [doi](#) ↑¹²
- [23] A. Waleed. *Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow*, GitHub repository, 2017. [URL](#) ↑¹²
- [24] E. Stevens, L. Antiga, T. Viehmann. *Deep Learning with PyTorch*, Manning Publications, New York, 2020, ISBN 9781617295263, 520 pp. ↑¹²
- [25] R. Sapkota, A. Dawood, M. Karkee. “Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments”, *Artificial Intelligence in Agriculture*, **13**:1 (2024), pp. 84–99. [doi](#) ↑¹⁹
- [26] V. F. Bulavitsky. “Use of drones to obtain aerial photographs terrain”, *Elektronnoe nauchnoe izdanie «Uchyonye zametki TOGU»*, **4**:4 (2013), pp. 1747–1755 (in Russian). ↑¹⁹

Received	21.10.2024;
approved after reviewing	24.12.2024;
accepted for publication	11.01.2025;
published online	31.01.2025.

Recommended by

dr. V. P. Fralenko

Information about the author:



Igor Victorovich Vinokurov

Candidate of Technical Sciences (PhD), Associate Professor at the Financial University under the Government of the Russian Federation. Research interests: information systems, information technologies, data processing technologies

[ID](#) 0000-0001-8697-1032
e-mail: igvvinokurov@fa.ru

The author declare no conflicts of interests.



Использование модели Mask R-CNN для сегментации объектов недвижимости на аэрофотоснимках

Игорь Викторович **Винокуров** 

Финансовый Университет при Правительстве Российской Федерации, Москва, Россия

Аннотация. Массовое появление незаконных и незарегистрированных в Едином Государственном Реестре Недвижимости (ЕГРН) объектов недвижимости осложняет кадастровый учёт для многих субъектов территориального и административного уровня. Традиционные методы выявления объектов подобных типов, основанные на ручном анализе геопространственных данных, трудоёмки и требуют значительного времени.

Для повышения эффективности этого процесса предлагается автоматизировать обнаружение объектов на аэрофотоснимках путём решения задачи инстанс-сегментации с использованием модели глубокого обучения Mask R-CNN. В статье описана подготовка набора данных для этой модели, исследованы основные метрики качества и проанализированы полученные результаты. Показана эффективность модели Mask R-CNN при обнаружении объектов недвижимости, не имеющих регистрации в ЕГРН. (*Связанные тексты статьи на английском и на русском языках*)

Ключевые слова и фразы: Кадастровый учёт, анализ аэрофотоснимков, инстанс-сегментация, Mask R-CNN, PyTorch

Для цитирования: Винокуров И. В. *Использование модели Mask R-CNN для сегментации объектов недвижимости на аэрофотоснимках* // Программные системы: теория и приложения. 2025. Т. 16. № 1(64). С. 3–44. (Англ.+русс.)
https://psta.psisras.ru/read/psta2025_1_3-44.pdf

Введение

Коттеджные поселения, дачные кооперативы и садовые товарищества обладают высокой плотностью застройки, осложняющей автоматическое обнаружение на аэрофотоснимках незаконно построенных объектов недвижимости и объектов, не зарегистрированных в ЕГРН.

Для реализации аэрофотосъёмки в ППК «Роскадастр» используются квадрокоптеры. На полученных фотографиях ищутся объекты недвижимости, подлежащие **обязательной регистрации**. К таким объектам относятся:

- дачные дома, предназначенные для постоянного проживания,
- объекты, стоящие на фундаменте (бани, террасы, летние кухни),
- строения на участках для индивидуального жилищного строительства.

Ручной поиск ограничивает эффективность и точность этого процесса. С развитием моделей глубокого обучения, таких как RetinaNet [1], SOLO [2], YOLACT [3], Mask R-CNN [4] и других, появляется возможность значительно сократить время и повысить качество анализа изображений за счёт автоматического детектирования объектов. Перечисленные выше модели позволяют не только идентифицировать объекты на изображении, но и точно выделять их границы. Последнее особенно важно для анализа плотных застроек, где объекты могут быть частично перекрыты или находится в близком соседстве друг с другом. Использование инстанс-сегментации, реализуемой этими моделями, наилучшим образом подходит для анализа геопространственных данных, поскольку наличие масок и ограничивающих рамок вокруг найденных экземпляров объектов позволяет не только обнаружить объекты, но и оценить их размеры. Это помогает заметить изменения размеров в результате перестройки объекта.

В работе приведены результаты исследования применимости модели Mask R-CNN (Mask Region-based Convolutional Neural Network) для обнаружения объектов недвижимости в коттеджных поселениях, дачных кооперативах и садовых товариществах. Особенностью предлагаемого подхода является обработка аэрофотоснимка с целью выявления ограничивающих рамок обнаруженных объектов, сопоставленных с определёнными географическими координатами, и сравнение их наличия или значимого изменения размеров с результатами аналогичной обработки аэрофотоснимков за прошлый период, как правило год. Такой подход позволяет выявлять изменения в застройке и постройки, не имеющие официальной регистрации.

1. Обзор работ по распознаванию объектов на аэрофотоснимках с использованием Mask R-CNN

Mask R-CNN является расширением модели Faster R-CNN, предназначенной для решения задач детектирования объектов и сегментации изображений [5]. Отличительной особенностью Mask R-CNN от Faster R-CNN является создание точных масок для каждого из обнаруженных на изображении объектов. Модель состоит из двух основных компонентов – обнаружение объектов и их сегментация [4].

Для извлечения признаков из входного изображения Mask R-CNN использует архитектуру ResNet [6] или другую свёрточную нейронную сеть. Эти признаки затем передаются в сеть предположения региона (RPN, Region Proposal Network) [4, 6], которая генерирует области интереса (RoI, Region of Interest) – предполагаемые местоположения объектов. Все полученные RoI передаются в основной классификатор, который определяет класс объекта и уточняет его границы.

В отличие от Faster R-CNN, в модели Mask R-CNN реализована ещё одна ветвь, отвечающая за создание масок сегментации. Эта ветвь использует небольшую свёрточную сеть с целью генерации бинарных масок для всех объектов в RoI. Размер маски соответствует размеру RoI.

Для обучения модели используется функция потерь, которая объединяет потери от классификации, регрессии координат и сегментации, что позволяет модели одновременно оптимизировать все три задачи [4].

Mask R-CNN хорошо справляется с перекрывающимися объектами и сложными фонами благодаря способности генерировать точные маски. Однако её производительность может зависеть от качества обучающего набора данных и выбранной архитектуры модели. В целом, Mask R-CNN является сильным инструментом для задач детекции и сегментации объектов в изображениях, обеспечивая высокую точность и гибкость в применении.

В настоящее время существует достаточно большое количество работ по распознаванию объектов различных типов на фотографиях и аэрофотоснимках с использованием этой модели, что свидетельствует об эффективности её применения для решения практических задач.

В работе [4] представлена основная архитектура Mask R-CNN и её применение к различным задачам сегментации, включая инстанс-сегментацию объектов.

В [7] авторы применяют Mask R-CNN для обнаружения зданий на спутниковых снимках высокого разрешения. Показывается высокая

точность сегментации и преимущества использования модели Mask R-CNN для решения подобных задач.

Обнаружение зданий на фотографиях из зон стихийных бедствий с использованием модели Mask R-CNN и предварительной обработки набора данных, с целью повышения эффективности его использования, приведено в [8]. В работе показаны преимущества этой модели перед другими моделями, реализующими детектирование объектов на фотографиях.

В работе [9] решается задача локализации полигонов зданий на спутниковых снимках высокого разрешения. Показывается эффективность применения модели Mask R-CNN для получения реальных границ контуров зданий на фотографиях с высокой плотностью объектов.

Метод обнаружения и сегментации кораблей на уровне пикселей с использованием модели Mask R-CNN предлагается в [10]. Отмечается преимущества использования этой модели и оценивается эффективность предлагаемого метода.

В [11] показывается, что извлечение контуров зданий из спутниковых снимков является сложной задачей из-за различий в масштабах, структурах и типов зданий. Для решения этой задачи предлагается использовать модель Mask R-CNN. Показывается эффективность её использования. Результаты извлечение отдельных зданий из спутниковых снимков предлагается использовать для приложений, автоматизирующих оценку населения, реализующих городское планирование и других.

Детектирование современной городской архитектуры с использованием модели Mask R-CNN, обученной на наборе данных, состоящим из элементов современных архитектурных стилей, приведено в [12]. В результате сравнения полученных результатов показана эффективность использования Mask R-CNN по сравнению с другими моделями.

В работе [13] показывается сложность обнаружения зданий и различного типа построек на спутниковых снимках из-за освещённости, плотности застроек, различных типов рельефов местности и других факторов. Для эффективного решения этой задачи предлагается использовать модель Mask R-CNN и собственный набор данных с усовершенствованной аугментацией изображений.

Применение модели Mask R-CNN для распознавания различных объектов на спутниковых снимках и аэрофотоснимках с целью автоматизации картографирования и поддержания карт местностей в актуальном состоянии описано в работе [14].

В [15] модель Mask R-CNN предлагается использовать для поддержания точности карт местностей с целью эффективного реагирования

на стихийных бедствия и катастрофы. Актуализировать карты местностей предлагается на основе аэрофотоснимков и спутниковых снимков. В работе получены более чем приемлемые результаты для фотографий с различным качеством, разрешениями и цветовыми каналами.

В [16, 17] предлагается гибридный подход к извлечению контуров зданий из спутниковых снимков низкого разрешения с использованием модели Mask R-CNN. Такой подход открывает перспективы для разработки автоматизированных инструментов обработки спутниковых снимков, и их эффективному использованию при мониторинге землепользования и реагирования на стихийные бедствия.

2. Обоснование выбора модели

Для реализации инстанс-сегментации объектов недвижимости на аэрофотоснимках, исходя из проведённого выше обзора, может быть обосновано выбрана модель Mask R-CNN, поскольку она обладает следующими преимуществами.

(1) Точность

- Модель Mask R-CNN демонстрирует более высокую точность детектирования объектов по сравнению с другими моделями [9, 10].
- Двухэтапный подход модели Mask R-CNN (предложение региона и сегментация [4–6]) позволяет более точно локализовать и сегментировать объекты с одновременным формированием их масок.

(2) Универсальность

- Mask R-CNN можно легко адаптировать для решения различных задач, включая обнаружение и контуризацию объектов, их инстанс- и семантическую сегментацию [4–6].
- Модульная архитектура этой модели позволяет легко добавлять или удалять её компоненты по мере необходимости.

(3) Надежность

- Mask R-CNN более устойчива к шумам и искажениям в изображениях [15].
- Двухэтапный подход этой модели помогает уменьшить количество ложных срабатываний и улучшить общую надежность [4–6].

(4) Поддержка объектов разных размеров

- Mask R-CNN может эффективно сегментировать объекты разных размеров, от маленьких до больших [4, 8].

- Механизм предложения региона [4] позволяет этой модели обнаруживать объекты независимо от их типа и размера (например, СОСО [5]).

(5) Меньше проблем с переобучением.

- Mask R-CNN менее склонна к переобучению [11].
- Использование нескольких функций потерь и двухэтапный подход помогают модели лучше обобщать результаты [4–6, 18]

Модели Mask R-CNN присущи и некоторыми недостатками, которые не являются критичными при решении задачи автоматизации обработки изображений, – это относительно невысокая скорость работы и сложность реализации [19].

Помимо рассмотренных выше моделей, для решения задачи инстанс-сегментации может быть использована и достаточно популярная в последнее время модель YOLO [20]. Несмотря на высокую скорость получения результатов, эта модель обладает существенными недостатками – невысоким качеством распознавания групп небольших объектов из-за ограниченного числа кандидатов для ограничивающих рамок (две) и возможностью дублирования ограничивающих рамок для одного и того же объекта [21]. Сравнение результатов распознавания объектов недвижимости, с аналогичными результатами, полученных с использованием последней версии модели YOLO, приведено в п. 6

3. Постановка задачи

Как уже было отмечено выше, фотографирование исследуемых ППК «Роскадстр» территорий осуществляется с использованием квадрокоптеров. Квадрокоптер (на момент написания статьи – это Phantom 4 RTK) реализует движение по заранее заданной траектории. Камера автоматически делает снимки при достижении определенных точек маршрута, формируя либо его полное изображение, либо изображения фрагментов маршрута (в большинстве случаев с перекрытием).

Целью данной работы является исследование применимости модели Mask R-CNN для реализации автоматической или автоматизированной обработки полученных снимков и выявления на них незаконно построенных или незарегистрированных объектов недвижимости. Для достижения поставленной цели в работе решались следующих задач:

- (1) формирование собственного набора данных для обучения модели,
- (2) анализа качества работы модели,
- (3) анализ результатов применения модели для обнаружения объектов незаконного строительства или объектов недвижимости, не зарегистрированных в ЕГРН.

4. Формирование набора данных

4.1. Аннотирование изображений

Основной задачей при формировании набора данных, используемого для обучения и тестирования модели, является отметка или аннотирование исходных изображений, представляющее собой полигональную контуризацию распознаваемых на фотографиях объектов. Точность контуризации определяет точность обнаружения объектов.

В настоящее время существует большое количество ПО, реализующего аннотирование изображений. Для формирования набора данных было выбрано ПО с открытым исходным кодом LabelMe.

Процесс аннотирования предполагает присваивание метки каждому распознаваемому объекту того или иного типа. При аннотировании аэрофотографий метки присваивались следующим 5-ти типам объектов, таблица 1.

Таблица 1. Типы объектов и их метки

Тип (класс) объекта	Имя метки и цвет сегментации класса
Коттеджный или дачный домик	<i>building</i>
Теплица (парник)	<i>greenhouse</i>
Хозяйственная постройка	<i>outbuilding</i>
Транспортное средство	<i>vehicle</i>
Бассейн	<i>swimming</i>

Основным типом объектов является коттеджный или дачный домик. Остальные типы объектов используются с целью предотвращения ложного распознавания последних и с целью возможного продолжения работы в данном направлении.

Результаты аннотирования сохранялись в формате JSON, поддерживаемом LabelMe. При формировании набора данных было осуществлено аннотирование 435 полученных с квадрокоптера фотографий. Пример фотографии фрагмента дачного кооператива и реализованная в процессе аннотирования полигональная контуризация 2-х классов объектов – дачного домика и теплицы в LabelMe приведены на рисунке 1, 2а и 2б соответственно.

Сформированный для обучения и тестирования модели набор данных представляет собой совокупность из указанного выше количества аэрофотоснимков и такого же количества JSON-файлов в формате ПО LabelMe, содержащих массивы координат точек полигональной контуризации и имена для объектов 5-ти типов.

Для обучения и тестирования модели было случайным образом выбрано 80% и 20% элементов этого набора данных соответственно.



Рисунок 1. Фотография фрагмента дачного кооператива с квадрокоптера Phantom 4 RTK



(а) Фрагмент изображения с дачным домиком и теплиц

(б) Контуризация объектов 2-х типов

Рисунок 2. Элемент набора данных – изображение объектов и их полигональная контуризация

Диаграмма распределения на аэрофотоснимках объектов, принадлежащих этим классам, приведена на рисунке 3.

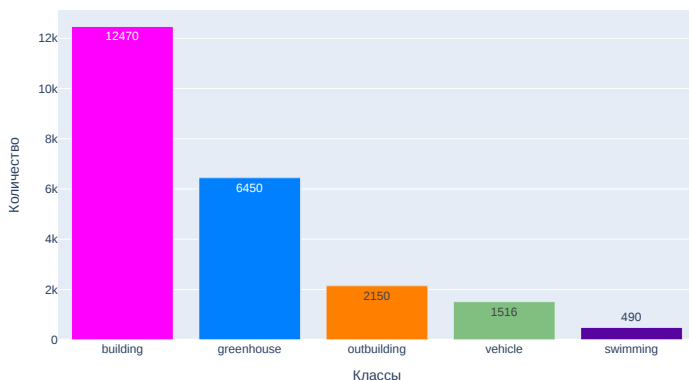


Рисунок 3. Распределение классов на фотографиях набора данных

4.2. Аугментация набора данных

Для повышения обобщающей способности модели сформирован и применён к изображениям из обучающего набора данных конвейер преобразований, реализованных в пакете `PyTorch`. В самом начале на основе метрики `IoU` (Intersection over Union) осуществлялось обрезка изображений с сохранением ROI. Затем применялись стохастические преобразования цвета, включающие вариации яркости, контрастности, насыщенности и цветового тона, а также случайное преобразование в градации серого (`RandomGrayscale`) и выравнивание гистограммы (`RandomEqualize`).

Дальнейшее увеличения разнообразия заключалось в уменьшении количества цветовых уровней и горизонтального отражения изображений с использованием методов `RandomPosterize` и `RandomHorizontalFlip` соответственно. Помимо этого было реализовано масштабирование изображений до максимального размера с сохранением соотношения сторон, дополнения до квадратной формы, и изменении размера до требуемого с использованием антиалиасинга. На заключительном этапе осуществлялось преобразование типа данных в `torch.float32` с масштабированием значений пикселей и валидация координат ограничивающих рамок (`SanitizeBoundingBoxes`).

5. Формирование и исследование модели

5.1. Формирование модели

Mask R-CNN характеризуется относительно высокой сложностью реализации. Как следствие, успешное применение этой модели требует тщательной настройки параметров и архитектуры сети. В работе была

выбрана достаточно быстрая и точная модель обнаружения объектов `maskrcnn_resnet50_fpn_v2` из пакета `torchvision`. Для повышения эффективности извлечения признаков (backbone) в этой модели была выбрана архитектура ResNet-101 с Feature Pyramid Network (FPN) [4]. Общие принципы её формирования и обучения достаточно хорошо описаны в [22] и [23].

Модель предварительно обучена на наборе данных COCO [5]. Для этой модели было экспериментально найдено оптимальное количество эпох дообучения, выбраны оптимизатор **Adam** и шедюлер **OneCycleLR** (механизмы, определяющие уменьшение весов во время обучения и скорость этого процесса), изменено количество выходных каналов и настроено разбиение для якорных боксов – заранее определенных прямоугольных рамок, которые используются для предложения потенциальных RoI при обнаружении объектов на изображении. Код формирования и исследования модели написан с использованием библиотеки машинного обучения PyTorch [24].

Для эффективной реализации инстанс-сегментации объектов различных масштабов, первоначальный свёрточный слой ResNet-101 был модифицирован. Перед слоем с достаточно большим размером ядра (7×7) были добавлены несколько параллельных свёрточных слоёв, имеющих ядра различных размеров, рисунок 4, 5.

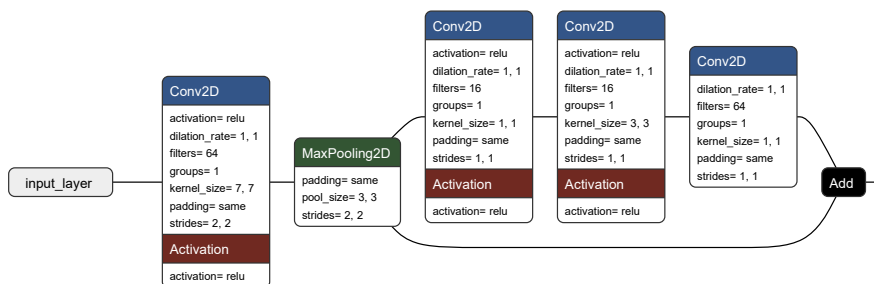


РИСУНОК 4. Начальный свёрточный слой в ResNet-101 и первый residual-блок [22]

Это потенциально позволяет модели одновременно учитывать как контекстуальные особенности (ядро 5×5), так и детальные характеристики объектов (ядра 1×1 и 3×3). Помимо этого, параллельные свёрточные слои позволяют ускорить сходимость модели.

Попытки вариативного изменения residual-блоков этой модели к какому-либо значимому улучшению детектирования объектов не привели.

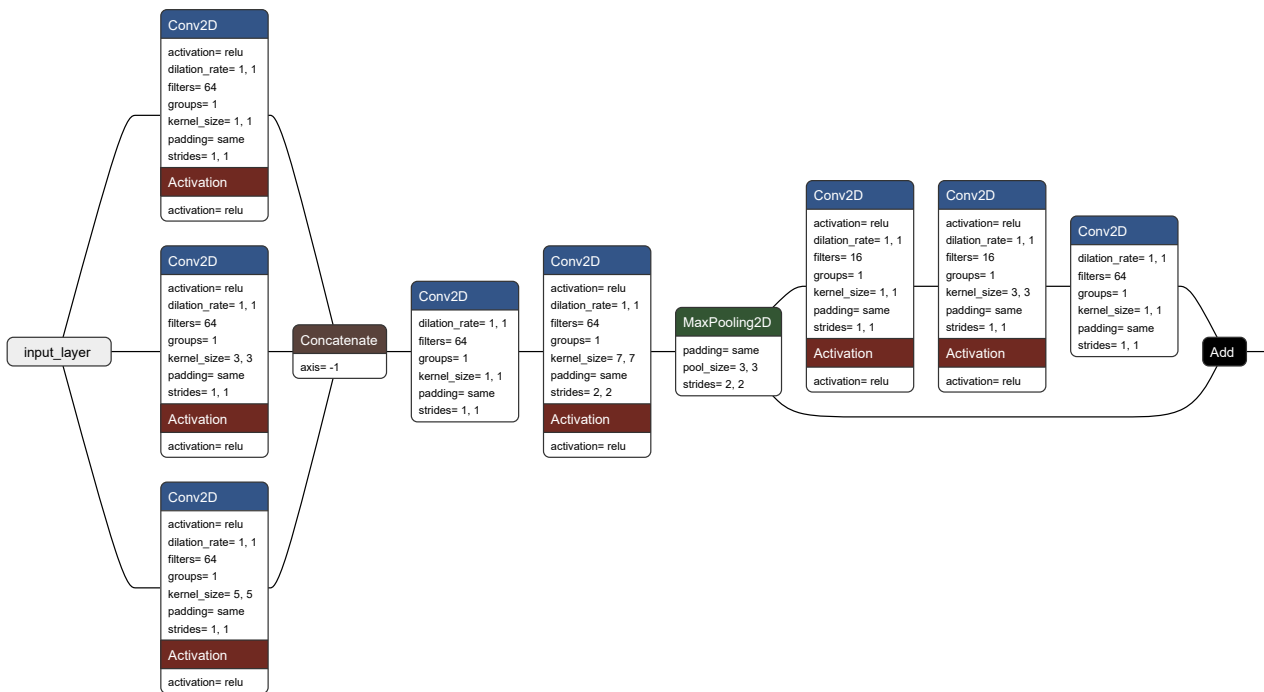


Рисунок 5. Параллельные свёрточные слои в ResNet-101

5.2. Исследование модели

Для анализа точности работы модели в процессе проведения экспериментальных исследований осуществлялось вычисление трёх ключевых метрик – **Loss**, **mAP** и **CS**.

Loss в Mask R-CNN позволяет оценить, насколько точно модель предсказывает классификацию, регрессию координат и сегментацию объектов. Чем ближе значения этой метрик к 0 (0%), тем более точным является результат работы модели. Изменение среднего значения этой метрики для всех 5-ти классов из наборов данных для обучения (train) и тестирования (valid) в зависимости от эпохи обучения приведено на рисунке 6.

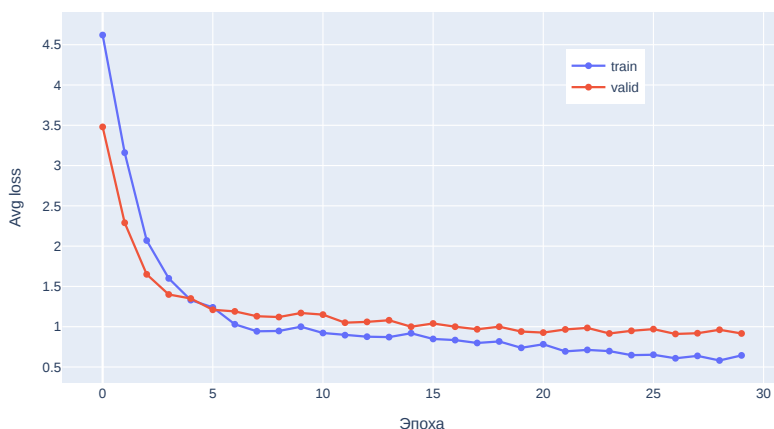


РИСУНОК 6. Изменение среднего значения функции потерь модели

Метрика **mAP** (mean Average Precision) является средним значением метрики точности по всем классам объектов. Аналогичная метрика **mAP50-95** рассчитывается при различных порогах перекрытия ограничивающих рамок **IoU** (Intersection over Union) – от 50% до 95%. Обе метрики учитывают как точность (**precision**), так и полноту (**recall**) на различных уровнях **IoU**. Чем ближе значения каждой из этих метрик к 1 (100%), тем более точным является результат работы модели. Зависимости метрик **mAP** и **mAP50-95** для всех классов из наборов данных для обучения (train) и тестирования (valid) от эпохи обучения приведена на рисунке 7.

Метрика **CS** (Confidence Score) представляет собой вероятность того, что объект принадлежит к определенному классу. **CS** является мерой уверенности модели в том, что предсказанный объект действительно

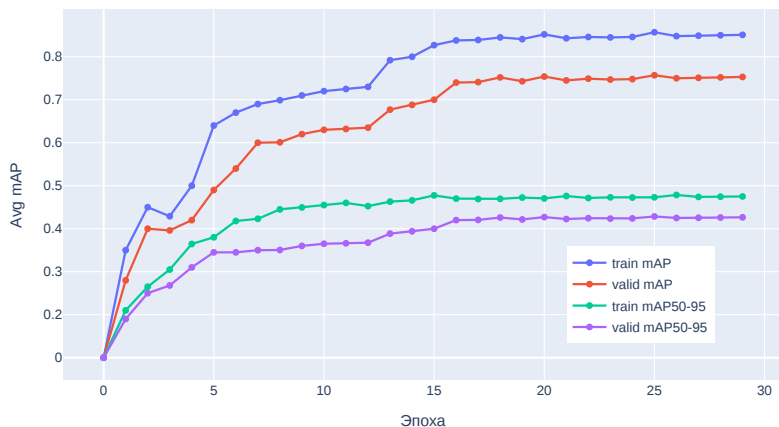


Рисунок 7. Изменение среднего значения метрик mAP и mAP50-95 при детектировании ограничивающих рамок

присутствует в ограничивающей рамке. Эта метрика особенно важна для моделей, реализующих детектирование объектов. Значения CS варьируются от 0 до 1 (100%); более высокие значения указывают на большую уверенность модели в правильности своего предсказания. Зависимость этой метрики для всех классов из наборов данных обучения (train) и тестирования (valid) от эпохи обучения показано на рисунке 8.

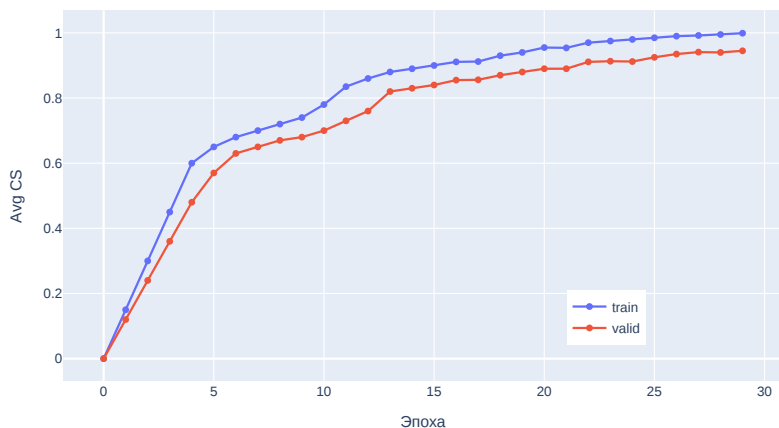


Рисунок 8. Изменение среднего значения метрики CS при детектировании объектов

В таблице 2 приведены значения метрик точности для обучающего и тестового наборов данных на последней эпохе обучения модели, см. рисунки 6–8.

ТАБЛИЦА 2. Значения метрик точности на последней эпохе обучения модели

Набор	Loss	mAP	mAP50-95	CS
train	0.741	0.851	0.475	0.999
valid	0.915	0.753	0.416	0.991

Вычисление значений метрик точности в Google Colab Pro в среде выполнения T4 GPU занимало порядка 80-100 минут. Из приведенных выше графиков и таблицы следует, что обученная на собственном наборе модель Mask R-CNN обладает более чем приемлемой точностью выявления интересующих нас объектов недвижимости на аэрофотоснимках. Несколько примеров детектирования дачных домиков из наборов данных для обучения модели и для её тестирования приведены на рисунках 9 и 10 соответственно.

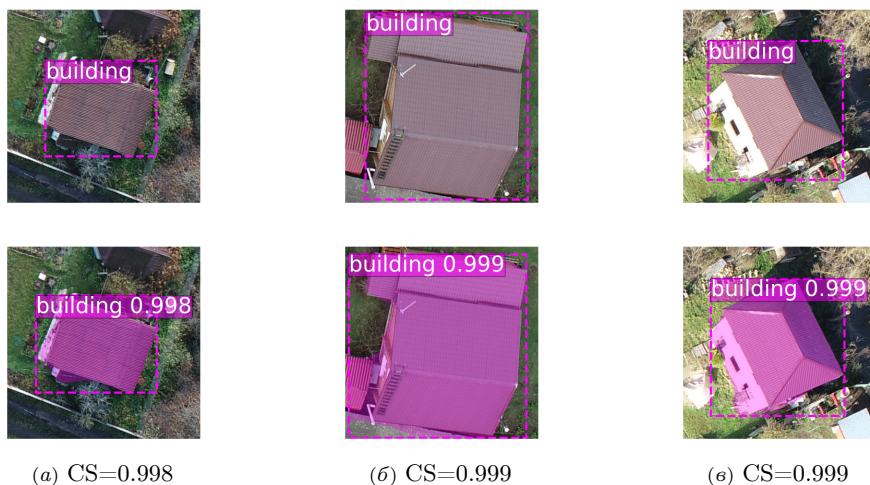


РИСУНОК 9. Детектирование объектов недвижимости на изображениях из тестового набора данных

При использовании этой модели в 3-5% случаев были выявлены незначительные ошибки, проявляющиеся в виде неверной идентификации объектов или неправильной сегментации границы. Появление этих ошибок связано с плотностью расположения сегментируемых объектов, с недостаточно представительным набором данных, наличием шумов на снимках



Рисунок 10. Детектирование объектов недвижимости на изображениях из тестового набора данных и недостаточной резкостью области изображения. Матрица ошибок детектирования объектов недвижимости на изображениях из тестового набора данных показана на рисунке 11.

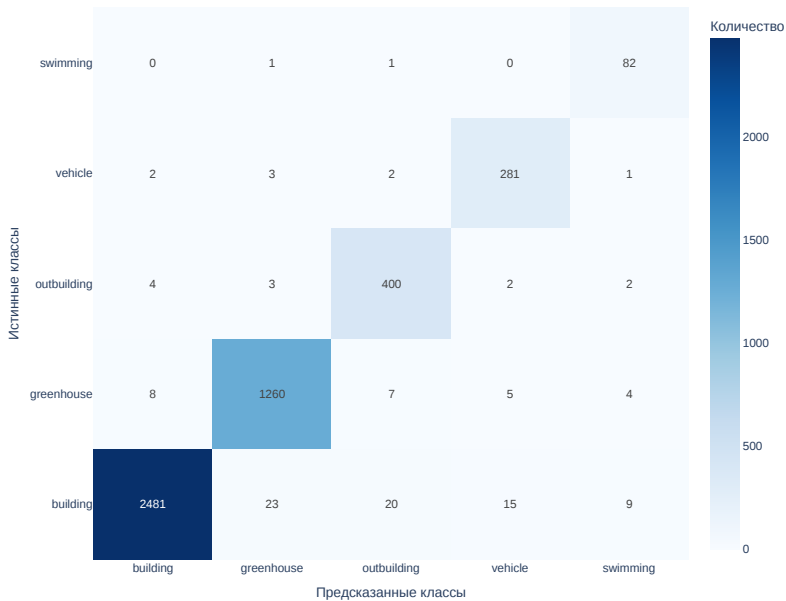


Рисунок 11. Матрица ошибок модели для тестового набора данных

Пример неверной идентификации для нескольких объектов классов *greenhouse*, *outbuilding* и *vehicle* приведён на рисунке 12. Легко заметить, что все объекты класса *building*, из-за достаточно большого их количества в обучающем наборе данных, идентифицируются верно, что является вполне достаточным для поставленной выше цели работы.

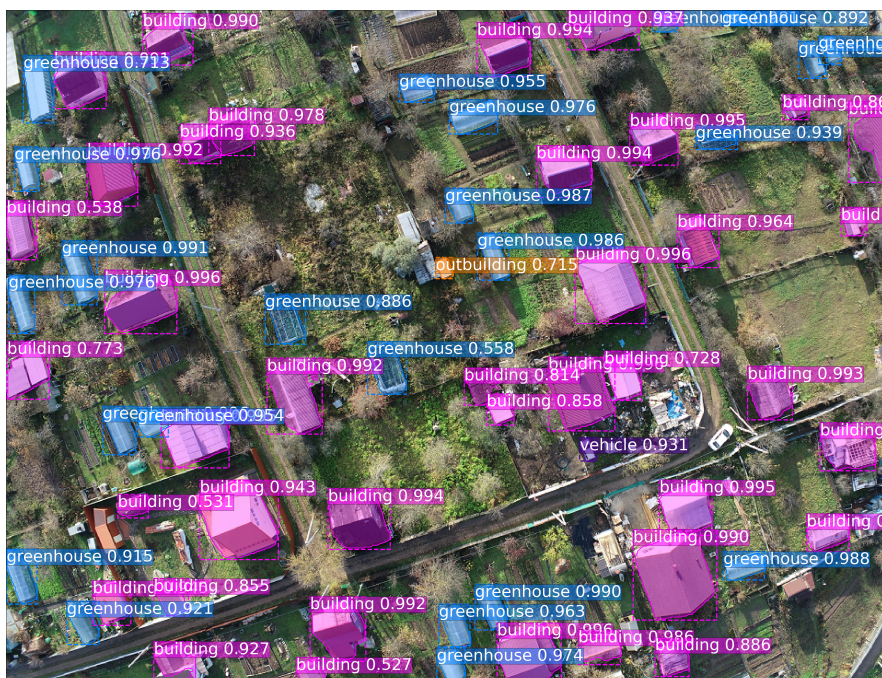


Рисунок 12. Результаты детектирования объектов разных типов на фотографии фрагмента дачного кооператива (см. рисунок 1)

6. Сравнение с моделью YOLO

Модель глубокого обучения YOLO в настоящее время интенсивно развивается и наравне с Mask R-CNN может быть использована для реализации детектирования объектов недвижимости на аэрофотоснимках. Исследования, аналогичные приведённым выше, были проведены и для модели *yolo11n-seg* от Ultralytics. На рисунке 13 приведены зависимости значений метрик точности mAP и mAP50-95 при детектировании ограничивающих рамок объектов недвижимости от эпохи обучения. Из этого рисунка видно, что эти метрики достигают значений, сопоставимых с полученными моделью Mask R-CNN, на большем количестве эпох.

Достаточно часто при анализе полученных результатов были выявлены случаи соотнесения обнаруженного объекта с разными классами. На рисунке 14, отображающем результаты детектирования, аналогичные приведённым на рисунке 12, видно несколько таких случаев (например, с объектами классов *building* и *outbuilding*). При исследовании результа-

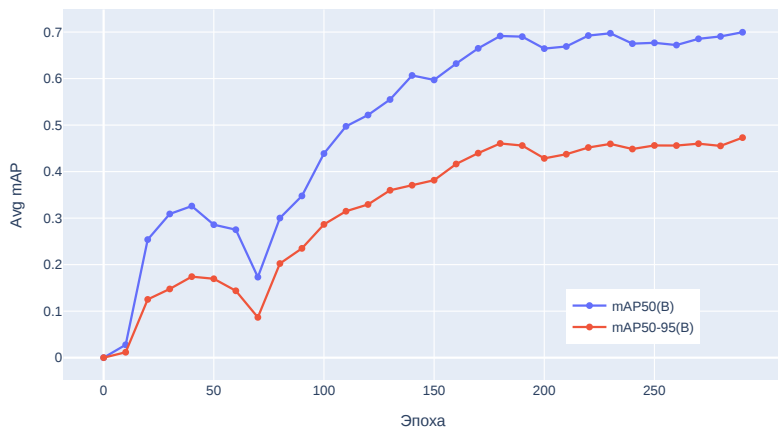


РИСУНОК 13. Изменение среднего значения метрик mAP и mAP50-95 модели YOLO при детектировании ограничивающих рамок

тов, полученных с использованием Mask R-CNN ни одного аналогичного случая выявлено не было. Кроме этого, при анализе достаточно большого количества результатов детектирования, было замечено более точное (до 10%-15%, см. рисунки 12 и 14) определение масок объектов моделью Mask R-CNN по сравнению с YOLO. Неправильное (неточное) определение маски объекта может быть критичным для ИС ППК «Роскадастр».

В целом, модель YOLO от Ultralytics оставила благоприятное впечатление простотой использования и наличием готового функционала для проведения экспериментов и анализ полученных результатов. В ряде случаев, в зависимости от специфики решаемой задачи и набора данных, использование моделей YOLO позволяет получить немного лучший результат, по сравнению с Mask R-CNN [25].

7. Практическая реализация результатов

Результаты, полученные в ходе работ по повышению эффективности обнаружения коттеджных и дачных домиков на аэрофотоснимках с использованием модели Mask R-CNN, были реализованы в бета-версии одной из подсистем информационной системы (ИС) ППК «Роскадастр». Основная цель этой подсистемы заключается в определении наличия регистрации объектов строительства в ЕГРН. В этой подсистеме реализуется процесс, включающий ортотрансформирование, создание цифровой модели местности (ЦММ), получение точек привязки и расчёт координат объектов недвижимости [26].

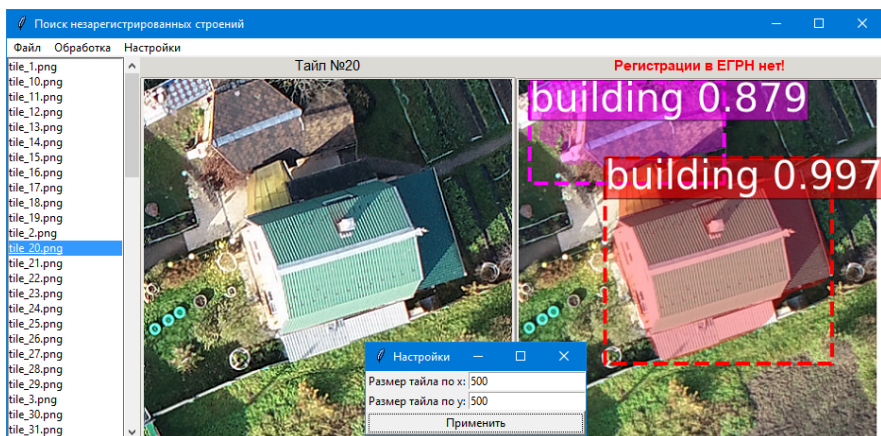


Рисунок 15. Пример выявления незарегистрированного дачного домика на фрагменте аэрофотоснимка

инстанс-сегментации, демонстрируя приемлемую точность по метрикам Loss, mAP, mAP50-95 и CS.

Разработанная бета-версия подсистемы ИС ППК «Роскадастр» на базе Mask R-CNN использует эту модель для распознавания объектов на изображениях, полученных с квадрокоптера, что существенно сокращает время и ресурсы, затрачиваемые на выявление нелегальной и незарегистрированной застройки, и повышает эффективность работы по контролю за использованием земельных ресурсов.

Дальнейшие исследования по применению модели Mask R-CNN будут направлены на повышение её точности в результате расширения и оптимизации обучающего набора данных, а также на интеграцию этой модели в существующие системы мониторинга и контроля за использованием земельных ресурсов. Кроме того, перспективным направлением является разработка алгоритмов автоматической верификации результатов работы модели, основанных на сравнении с данными ЕГРН.

Список использованных источников

- [1] Lin T.-Y., Goyal P., Girshick R. B., He K., Dollár P. *Focal loss for dense object detection.* – Computing Research Repository (CoRR). – 2017. – 10 pp. arXiv:1708.02002 ↑25
- [2] Wang X., Kong T., Shen Ch., Jiang Y., Li L. *SOLO: Segmenting objects by locations.* – Computing Research Repository (CoRR). – 2019. – 19 pp. arXiv:1912.04488 ↑25
- [3] Duda K., Ivanov A. *On decidability of amenability in computable groups* // Archive for Mathematical Logic. – 2022. – Vol. 61. – Pp. 891–902. ↑25

- [4] He K., Gkioxari G., P. Dollár, Girshick R. B. *Mask R-CNN*.– Computing Research Repository (CoRR).– 2017.– 12 pp. arXiv 1703.06870 25, 26, 28, 29, 33
- [5] Ren S., He K., Girshick R. B., Sun J. *Faster R-CNN: Towards real-time object detection with region proposal networks*.– Computing Research Repository (CoRR).– 2015.– 14 pp. arXiv 1506.01497 26, 28, 29, 33
- [6] He K., Zhang X., Ren S., Sun J. *Identity mappings in deep residual networks*.– Computing Research Repository (CoRR).– 2016.– 15 pp. arXiv 1603.05027 26, 28, 29
- [7] Lin T.-Y., Maire M., Belongie S. J., Bourdev L. D., Girshick R. B., Hays J., Perona P., Ramanan D., P. Dollár, Zitnick C. L. *Microsoft COCO: Common objects in context*.– Computing Research Repository (CoRR).– 2014.– 15 pp. arXiv 1405.0312 26
- [8] Xu Y., Wu L., Xie Z., Chen Z. *Building extraction in very high resolution remote sensing imagery using deep learning and guided filters* // Remote. Sens.– 2018.– Vol. 10.– No. 1.– id. 144.– 18 pp. 27, 28
- [9] Han Q., Yin Q., Zheng X., Chen Z. *Remote sensing image building detection method based on Mask R-CNN* // Complex Intell. Syst.– 2022.– Vol. 8.– Pp. 1847–1855. 27, 28
- [10] Zhao K., Kang J., Jung J., Sohn G. *Building extraction from satellite images using Mask R-CNN with building boundary regularization* // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (18–22 June 2018, Salt Lake City, UT, USA).– IEEE.– ISBN 9781538661017.– id. 242.– 4 pp. 27, 28
- [11] Nie X., Duan M., Ding H., Hu B., Wong E. K. *Attention Mask R-CNN for ship detection and segmentation from remote sensing images* // IEEE Access.– 2020.– Vol. 8.– Pp. 9325–9334. 27, 29
- [12] Jenila Vincent M., Varalakshmi P. *Extraction of building footprint using MASK-RCNN for high resolution aerial imagery* // Environmental Research Communications.– 2024.– Vol. 6.– No. 7.– id. 075015.– 17 pp. 27
- [13] Zhu X., Hu L., Wang J. *Urban modern architecture recognition based on Mask-RCNN and ECA attention mechanism*, Fifth International Conference on Geoscience and Remote Sensing Mapping (ICGRSM 2023) (13–15 October 2023, Lianyungang, China), Proc. SPIE.– vol. 12980.– 2024.– ISBN 9781510672789.– id. 129801D. 27
- [14] Raghavan R., Chander Verma D., Pandey D., Anand R., Kumar Pandey B., Singh H. *Optimized building extraction from high-resolution satellite imagery using deep learning* // Multimedia Tools and Applications.– 2022.– Vol. 81.– No. 29.– Pp. 42309–42323. 27
- [15] Ulanov D., Syrov A. *Building footprint extraction based on RGBD satellite imagery*, CS230 Deep Learning (Winter 2020, Stanford University, CA).– 2020.– 11 pp. 27, 28
- [16] Solanki A., Singh R. K., Demeneze B. *Aerial pictures semantic segmentation applying deep learning* // International Journal of Trendy Research in Engineering and Technology.– 2021.– Vol. 5.– No. 1.– Pp. 42–48. 28
- [17] NourEldeen A., Wahed M. E. *Enhanced building footprint extraction from satellite imagery using Mask R-CNN and PointRend* // Bulletin of Electrical Engineering and Informatics.– 2024.– Vol. 5.– No. 13.– Pp. 3601–3608. 28

- [18] He K., Zhang X., Ren S., Sun J. *Deep residual learning for image recognition*.– Computing Research Repository (CoRR).– 2015.– 12 pp. arXiv:1512.03385 doi ↑29
- [19] Mills Ch. J. *PyTorch Mask R-CNN tutorial*.– GitHub repository.– 2023. URL ↑29
- [20] Redmon J., Divvala S., Girshick R. B., Farhadi A. *You Only Look Once: Unified, real-time object detection*.– Computing Research Repository (CoRR).– 2015.– 10 pp. arXiv:1506.02640 doi ↑29
- [21] Khanam R., Hussain M. *YOLOv11: An overview of the key architectural enhancements*.– 2024.– 9 pp. arXiv:2410.17725 doi ↑29
- [22] Lin T. -Y., P. Dollár, Girshick R. B., He K., Hariharan B., Belongie S. J. *Feature pyramid networks for object detection*.– Computing Research Repository (CoRR).– 2016.– 10 pp. arXiv:1612.03144 doi ↑33
- [23] Waleed A. *Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow*.– GitHub repository.– 2017. URL ↑33
- [24] Stevens E., Antiga L., Viehmann T. *Deep Learning with PyTorch*.– New York: Manning Publications.– 2020.– ISBN 9781617295263.– 520 pp. ↑33
- [25] Sapkota R., Dawood A., Karkee M. *Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments* // Artificial Intelligence in Agriculture.– 2024.– Vol. 13.– No. 1.– Pp. 84–99. doi ↑40
- [26] Булавицкий В. Ф. *Применение беспилотных летательных аппаратов для оперативного получения аэрофотоснимков местности* // Электронное научное издание «Учёные заметки ТОГУ».– 2013.– Т. 4.– № 4.– С. 1747–1755. ✨ ↑40

Поступила в редакцию 21.10.2024;
 одобрена после рецензирования 24.12.2024;
 принята к публикации 11.01.2025;
 опубликована онлайн 31.01.2025.

Рекомендовал к публикации

к.т.н. В. П. Фраленко

Информация об авторе:



Игорь Викторович Винокуров

Кандидат технических наук (PhD), ассоциированный профессор в Финансовом Университете при Правительстве Российской Федерации. Область научных интересов: информационные системы, информационные технологии, технологии обработки данных

id 0000-0001-8697-1032
 e-mail: igvvinokurov@fa.ru

Декларация об отсутствии личной заинтересованности: *благополучие автора не зависит от результатов исследования.*



О реализации QR-разложения на трехмерном систолическом массиве

Виктор Николаевич **Бабенко**^{1✉}, Артём Павлович **Невечеря**²

¹ Кубанский государственный технологический университет, Краснодар, Россия

² Кубанский государственный университет, Краснодар, Россия

Аннотация. Интенсивные потоки данных, формирующихся систем линейных уравнений в режиме реального времени, а также системы линейных уравнений большой размерности обуславливают привлечение систолических массивов для их машинного решения. В представленном систолическом массиве, предназначенном для приведения матриц к треугольному виду, реализация ортогональных преобразований вращения может осуществляться как устройствами двумерного вращения вектора CORDIC, так и его модификациями. Для предложенного систолического массива даны описания его конфигурации, функционирования и технических характеристик, а также структуры входного и выходного потока данных.

Ключевые слова и фразы: систолический массив, функциональное устройство, продолжительность такта, маршрутизация и расписание движения данных, устройство вращения двумерного вектора, прямой ход

Для цитирования: Бабенко В. Н., Невечеря А. П. *О реализации QR-разложения на трехмерном систолическом массиве* // Программные системы: теория и приложения. 2025. Т. 16. № 1(64). С. 45–59. https://psta.psir.as.ru/read/psta2025_1_45-59.pdf

Введение

Непрерывно возрастающие объемы вычислений обуславливают необходимость повышения производительности вычислительных систем. Одно из направлений в этой деятельности – разработка и применение систолических массивов (СМ). Вопросам разработки СМ различного назначения посвящено немало работ, так как они являются одним из наиболее эффективных средств решения указанной проблемы [1–8].

Одной из важнейших задач вычислительной линейной алгебры является решение систем линейных алгебраических уравнений (СЛАУ). Пусть матрица системы

$$Ax = f$$

невырождена, и N – ее порядок. Для осуществления QR-разложения матрицы A мы будем использовать преобразования (матрицы) вращения.

Введем обозначения $A^{(0,0)} = A$, $Q_{i,j}^*$ – преобразование вращения, аннулирующее i -ю компоненту j -ого столбца матрицы.

Для $j = \overline{1, N-1}$, $i = \overline{j, N}$ вычисляем:

$$x = a_j^{(j-1, i-1)}, \quad c = \frac{x_j}{\sqrt{x_j^2 + x_i^2}}, \quad s = \frac{x_i}{\sqrt{x_j^2 + x_i^2}},$$

$$\text{для } k = \overline{j, N}: \quad x = a_k^{(j-1, i-1)}, \quad y = S_{ji}^* x, \quad a_k^{(j, i)} = y.$$

По завершении этой процедуры получим верхнюю треугольную матрицу $R = A^{(N-1, N)}$, связанную с исходной соотношением $B = S^* A$, где

$$S = S_{1,2} \dots S_{1,N} S_{2,3} \dots S_{2,N} \dots S_{N-2,N-1} S_{N-2,N} S_{N-1,N}.$$

Следует сказать, что в литературе для представления матрицы A в виде произведения $A = QR$, где Q и R – ортогональная и верхняя треугольная матрицы соответственно, широко применяется термин QR-разложение. Следуя этой традиции, произведение $Q^* A$ мы будем называть QR-упрощением матрицы A .

Представленный способ QR-разложения матрицы системы уравнений может быть эффективно реализован аппаратно в системах реального времени, формирующих плотные потоки данных для решения СЛАУ с плотными матрицами небольшой размерности.

Второй областью применения QR-разложения матрицы, очевидно, являются задачи решения СЛАУ большой размерности с разреженными матрицами. Для решения таких систем применяются итерационные методы. Пусть x^0 – произвольный вектор, $r^0 = f - Ax^0$ и

$$x^{n+1} = x^n + y^n, \quad n = 0, 1, 2, \dots$$

В проекционных методах вычисление поправки y^n сводится к последовательному решению трех систем линейных уравнений

$$V_n w^n \cong r^n, \quad H_n z^n \cong w^n, \quad V_n^* y^n \cong z^n,$$

где V_n – псевдоортогональная матрица, $r^n = f - Ax^n$ и H_n – некоторая плотная невырожденная матрица порядка m , причем m невелико.

Из трех выше представленных систем наиболее трудоемким является решение второй системы. Системы линейных уравнений $H_n z^n \cong w^n$, $n = 0, 1, 2, \dots$, обозначенного потока могут быть решены с помощью QR-разложения матрицы H_n , реализованного аппаратурно на специально разработанном вычислительном устройстве (систолическая матрица Джентльмена–Кунга [9]).

Обоснование заявленного утверждения упрощенно раскроем следующим образом. Для реализации преобразования вращения вектора разработано специальное вычислительное устройство CORDIC, которое производится в двух вариантах: $R1$, аннулирующее вторую компоненту входного вектора, и $R2$, работающее в режиме «эхо».

На вход $R1$ поступают компоненты x и y двумерного вектора $\begin{bmatrix} x \\ y \end{bmatrix}$. На выходе получают вектор $\begin{bmatrix} z \\ 0 \end{bmatrix}$, где $z = \text{sign}(x)\sqrt{x^2 + y^2}$, и параметры преобразования вращения c и s .

На вход $R2$ поступают компоненты u и v двумерного вектора $\begin{bmatrix} u \\ v \end{bmatrix}$ и параметры преобразования вращения c и s . Соответственного на выходе получают вектор $\begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$ и параметры преобразования вращения c и s .

Устройство CORDIC осуществляет вычисление компонент вектора за $2m$ тактов, где m – количество разрядов, отводимых под мантиссу машинных чисел. Устройства CORDIC получили применение в систолической структуре (матрице Джентльмена–Кунга), применяемой для приведения матрицы к верхнему треугольному виду. Эти устройства в конвейерном варианте способны обрабатывать потоки данных высокой плотности, но размерность систолической структуры Джентльмена–Кунга (она равна 2) препятствует реализации этой способности указанного устройства. Выходом из этого положения может служить переход к применению трехмерных систолических структур.

Следует сказать, что недостатком метода CORDIC является не слишком высокая скорость сходимости (для достижения приемлемой

точности требуется $2m$ итераций). Соответственно этому задержка вычисленного результата на выходе устройства CORDIC составит $2m$ тактов. Одним из авторов этой статьи разработаны модификация метода CORDIC, позволившая сократить число итераций до m , и соответствующее ей устройство, обеспечивающее возможность получить вычисленный результат за m тактов [10–13]).

Выполняя эту работу, мы преследовали цель обеспечить обработку потоков данных высокой плотности при осуществлении решения СЛАУ. Достижение этой цели мы видели в решении задачи разработки систолических структур, использующих в качестве функциональных устройств (ФУ) CORDIC-устройства конвейерного типа или их указанную выше модификацию [10–12].

1. Основные положения

В систолических массивах в состав ФУ входят функциональные элементы (ФЭ), которые под воздействием входных сигналов и синхронизирующих импульсов, называемых тактовыми, переходят из предыдущего состояния в последующее.

Пусть T_m и T_{m+1} – моменты поступления на ФЭ m -го и $m+1$ -го тактовых импульсов. Промежуток времени $\Delta T = T_{m+1} - T_m$ называют m -м тактом.

Символами $I_{ФЭ}$ и $O_{ФЭ}$ обозначают вход и выход ФЭ. Обозначение $I_{ФЭ}(m) = a$ означает, что на вход ФЭ на m -ом такте (обычно в начале промежутка ΔT) поступает сигнал (число) a .

Под воздействием сигнала a и m -го тактового импульса ФЭ с течением времени Δt переходит в новое устойчивое состояние, которое наблюдают на его выходе. При этом пишут $O_{ФЭ}(m) = b$, где b – сигнал (число) на выходе ФЭ на m -ом такте, соответственно Δt называют временем срабатывания ФЭ. Выполнение неравенства $\Delta t \leq \Delta T$ называют условием срабатывания ФЭ.

Переходя к формальному изложению, последнее мы можем записать так.

ЛЕММА 1.1 (о срабатывании ФЭ). Пусть $I_{ФЭ}(m) = a$. Если $\Delta t \leq \Delta T$, то $O_{ФЭ}(m) = b$.

Выражение $I_{ФЭ_2} \Leftarrow O_{ФЭ_1}$ ($O_{ФЭ_1} \Rightarrow I_{ФЭ_2}$) означает, что выход ФЭ₁ соединен с входом ФЭ₂.

ОПРЕДЕЛЕНИЕ 1.1. Если на m -том такте на выходе ФЭ появляется результат b (т.е. $O_{\Phi\mathcal{E}}(m) = b$), то его мы будем называть тактом, порождающим результат b при отображении ФЭ (обозначение $m = \text{ind}_{\Phi\mathcal{E}}(b)$).

ЛЕММА 1.2 (о локальном взаимодействии ФЭ). Пусть $I_{\Phi\mathcal{E}_2} \leq O_{\Phi\mathcal{E}_1}$. Если $I_{\Phi\mathcal{E}_1}(m) = a$ и $O_{\Phi\mathcal{E}_1}(m) = b$, то $I_{\Phi\mathcal{E}_2}(m+1) = b$.

ОПРЕДЕЛЕНИЕ 1.2. Конвейером C мы будем называть функциональное устройство, представляющее собой последовательность ФЭ

$$C = \{F_h, \dots, F_1\} : I_{F_{\rho+1}} \leq O_{F_\rho}, \rho = \overline{1, h-1}.$$

Соответственно $I_C = I_{F_1}$ – вход, $O_C = O_{F_h}$ – выход, и h – длина конвейера C .

Входящие в состав конвейера (СМ) функциональные элементы в общем случае имеют различное время срабатывания. Вопрос о выборе продолжительности такта, общей для конвейера, разрешает следующая

ЛЕММА 1.3 (о необходимом условии функционирования конвейера). Пусть конвейерный вычислитель C с регулярным графом состоит из n функциональных элементов, причем Δt_i , $i = \overline{1, h}$ – время срабатывания i -го элемента, и справедливы соотношения: $a_1 = F_1(a)$, $a_2 = F_2(a_1)$, $\dots$, $a_{h-1} = F_{h-1}(a_{h-2})$, $b = F_h(a_{h-1})$. Если $I_C(m) = a$, и $\Delta t \leq \Delta T$, где $\Delta t \geq \max\{\Delta t_1, \dots, \Delta t_h\}$, то $O_C(m+h-1) = a_h = b$.

Действуя по индукции, очевидно, применяя при этом леммы 1.1 и 1.2, мы докажем утверждение последней леммы.

Если $I_R(m) = a$ и $O_R(m) = a$, то функциональный элемент R называют элементом задержки, а конвейер LD , представляющий собой последовательность элементов задержки $R = \{R_h, \dots, R_1\}$: $I_{R_{\rho+1}} \leq O_{R_\rho}$, $\rho = \overline{1, h-1}$, $I_{LD} = I_{R_1}$, $O_{LD} = O_{R_h}$, называют линией задержки, соответственно I_{LD} – входом, O_{LD} – выходом и h – длиной линии задержки.

2. QR-упрощение матрицы на трехмерном СМ

Поскольку назначением предлагаемого ниже систолического массива прямого хода (СМПХ) является приведение матрицы системы $Ax = y$ к треугольному виду (этап 1) прежде всего для реализации последующего обратного хода (этап 2), и при осуществлении этапа 1 происходит трансформация вектора y , как и всех вектор-столбцов матрицы A , мы

сформируем расширенную матрицу $[A \ y]$. Эту матрицу, как и ранее, мы будем обозначать символом A , а вектор y считать $(N+1)$ -м столбцом расширенной матрицы A .

2.1. Конфигурация СМПХ

Для реализации QR-упрощения мы применим трехмерную матрицу ФУ R , состав элементов которой, описывается соотношениями:

$$R_{ijk} = \begin{cases} LD & \forall j, \text{ если } i = k, \\ R1 & \forall i, \text{ если } j = k, \\ R2 & \text{в остальных случаях,} \end{cases}$$

где

$$k = \overline{1, N-1}, \quad i = \overline{k, N}, \quad j = \overline{k, N+1}.$$

В целях упрощения (единообразия) дальнейшего изложения мы будем придерживаться следующих соглашений:

$$I_{R_{i,j,k}} = I2_{R_{i,j,k}}, \quad O_{R_{i,j,k}} = O1_{R_{i,j,k}}, \quad k = \overline{1, N-1}, \quad i = k, \\ j = \overline{k, N+1}.$$

Функциональные устройства связаны между собой следующим образом.

(1) Одноуровневые (горизонтальные) связи:

$$O3_{R_{i,j,k}^\rho} \Rightarrow I3_{R_{i,j+1,k}^\rho}, \quad k = \overline{1, N-1}, \quad i = \overline{k+1, N}, \quad j = \overline{k, N}, \\ \rho = \overline{1, m}, \quad (\text{направление} - \text{вправо});$$

$$O1_{R_{i,j,k}} \Rightarrow I1_{R_{i+1,j,k}}, \quad k = \overline{1, N-1}, \quad i = \overline{k+1, N-1}, \\ j = \overline{k, N+1}, \quad (\text{направление} - \text{вперед} (\text{«вниз»})).$$

(2) Межуровневые связи:

$$O2_{R_{i,j,k}} \Rightarrow I2_{R_{i,j,k+1}}, \quad k = \overline{1, N-1}, \quad i = \overline{k+1, N}, \\ j = \overline{k+1, N+1}, \quad (\text{направление} - \text{вниз}).$$

Входы СМПХ:

$$I2_{R_{1,j,k}}, \quad i = \overline{1, N}, \quad j = \overline{1, N+1}.$$

Выходы СМПХ:

$$\begin{aligned} O1_{RN,j,k}, \quad k = \overline{1, N-1}, \quad j = \overline{k, N+1}, \\ O2_{RN,N,N-1}, \quad O2_{RN,N+1,N-1}. \end{aligned}$$

2.2. Основные свойства СМПХ (маршрутизация и расписание движения данных)

Пусть требуется осуществить QR-упрощение матриц

$$A^1, \dots, A^l, \dots$$

Описание свойств СМПХ предполагает применение переменных:

$t, \bar{t}$ — номера тактов,

$\tau, \bar{\tau}$ — номера упрощаемых матриц (решаемых систем уравнений).

Для описания свойств (технических характеристик) СМПХ, представляющих нам основными, определим следующие величины:

$\Delta\bar{T}$ — продолжительность порождения упрощенной матрицы, определяющаяся по формуле $\Delta\bar{T} = k_e - k_b + 1$, где k_b — номер такта, на котором на вход СМПХ поступает первый элемент упрощаемой матрицы, k_e — номер такта, на котором порождается последний элемент упрощенной матрицы;

P — производительность, определяемая как количество элементов упрощаемых матриц систем уравнений, появляющихся на выходе (выходах) СМПХ одновременно.

ВЫСКАЗЫВАНИЕ 2.1. Пусть

(i) продолжительность ΔT такта на СМПХ удовлетворяет неравенству

$$\Delta T \geq \Delta t_+,$$

где Δt_+ — продолжительность времени срабатывания сумматора.

(ii) для всех $l = 1, 2, \dots, i = \overline{1, N-1}, j = \overline{1, N+1}$ выполняются соотношения

$$\begin{aligned} (1) \quad I2_{Ri,j,1}(t) &= a_{i,j}^T, \quad \text{где} \\ t &= h(i-1) + l, \\ \tau &= t - h(i-1) - (j-1), \end{aligned}$$

причем соотношения (1) определены для всех $l, i, j : \tau > 0$.

Тогда

1) для всех $k = \overline{1, N-1}$

$$O1_{R_{N,j,k}}(\bar{t}) = b_{k,j}^{\bar{\tau}},$$

$$O2_{R_{N,N+1,N-1}}(h(N + (N-2) + (l-1))) = b_{N,N}^l,$$

$$O2_{R_{N,N+1,N-1}}(h(N + (N-2)) + l) = b_{N,N+1}^{l+1},$$

$$\begin{aligned} \bar{t} &= h(N + (k-1)) + (l-1), \\ \text{где} \quad \bar{\tau} &= \bar{t} - h(N + (k-1)) - (j-1) + 1; \end{aligned}$$

2) продолжительность процедуры QR-упрощения матрицы $\Delta\bar{T}$ равна $2hN + N + 1$ тактам;

3) производительность СМПХ P равна $(N+3)N/2$ слов за такт.

ЗАМЕЧАНИЕ 2.1. Значения величин пунктов 2) и 3) высказывания 2.1 получены применением леммы 1.3.

Заключение

В представленном СМПХ за счет увеличения размерности СМ до трех достигнута максимально возможная плотность потоков движения данных. Она ограничена продолжительностью выполнения операции сложения Δt_+ . Сопоставляя упомянутую во введении матрицу Джентльмена–Кунга с СМПХ, мы видим, что производительность разработанного СМПХ в $n \times 2h$ выше, чем у СМ Джентльмена–Кунга. В n раз выше за счет увеличения размерности массива до трех, и в $2h$ раз – за счет появившейся при этом возможности увеличить плотность потока данных в $2h$ раз, где

$$h = \begin{cases} m & \text{для устройства CORDIC,} \\ m/2 & \text{для модифицированного CORDICa,} \end{cases}$$

m – количество двоичных разрядов, отводимых под мантиссу числа.

Проводя обзор публикаций последних лет по реализации QR-разложения на трехмерных систолических массивах с использованием CORDIC-устройств, мы обнаружили следующее положение дел.

В работе [14] предложена архитектура двумерного систолического массива, который может быть использован в сверхбольших интегральных схемах (СБИС) для аппаратной реализации алгоритма QR-разложения комплексных матриц. Предложенная архитектура является эффективной для реализации в СБИС, так как использует только алгоритм CORDIC и, следовательно, не использует операций умножения. Рассматриваемый систолический массив основан на методе треугольных комплексных

вращений и позволяет получить значительный выигрыш в производительности по сравнению с традиционно используемым методом комплексных вращений Гивенса.

В статье [15] исследуется специфическое исключение элементов для решения линейной алгебраической системы вида $Az = b$, где матрица A задана суммой m парных произведений тёплицевых треугольных (нижней и верхней) матриц порядка n . Предложен алгоритм, требующий лишь $(3m - 2)n^2$ умножений и такого же числа сложений. Для симметричной матрицы A предложен алгоритм, требующий лишь $(2m - 1)n^2$ умножений и такого же числа сложений. Алгоритмы обладают регулярной структурой и допускают параллельную реализацию за $O(nm)$ шагов.

В работах иностранных авторов представлены результаты по практической реализации QR-разложения на двумерных систолических массивах.

В статье [16] дано описание двумерного систолического массива, реализованного на ПЛИС Xilinx Virtex5 и предназначенного для реализации QR-разложения с использованием алгоритма вращения Гивенса.

В работе [17] представлена улучшенная архитектура систолического массива для реализации QR-разложения на основе метода вращения Гивенса (GR) для действительной матрицы 4×4 . Алгоритм цифрового компьютера вращения координат (CORDIC) принят и модифицирован для ускорения и упрощения процесса GR.

В статье [18] представлен улучшенный аппаратный дизайн с фиксированной точкой QR-разложения, специально оптимизированный для ПЛИС Xilinx. Алгоритм вращения Гивенса реализован с использованием свернутого систолического массива и алгоритма CORDIC.

В статье [19] представлена параллельная архитектура систолического массива QR-разложения на основе алгоритма вращения Гивенса на ПЛИС. Предлагаемая архитектура использует прямое отображение 21 процессорных единиц с фиксированной точкой на основе CORDIC, которые могут вычислять QR-разложение для действительной матрицы 4×4 .

По трехмерным систолическим массивам мы обнаружили только три работы

В [20] предложена архитектура трёхмерного систолического вычислителя дискретного преобразования Фурье на основе кронекеровского произведения матриц.

Авторы в статье [21] сообщают, что ими обнаружено следующее. Систолический массив как уровень программной виртуализации может

привести к чрезвычайно масштабируемой парадигме выполнения. Чтобы продемонстрировать эту масштабируемость, они спроектировали и реализовали виртуальный трехмерный систолический массив для реализации QR-разложения. На машине Cray-XT5 продемонстрировано, что виртуальный систолический массив может обеспечить превосходную параллельную производительность.

В работе [22] рассмотрена реализация HLS архитектуры трехмерного систолического массива для умножения матриц, которая нацелена на определенные характеристики ПЛИС Intel Stratix 10, чтобы создавать проекты, которые достигают высокой пропускной способности с плавающей точкой.

Представленный анализ публикаций последних лет позволяет сделать вывод, что полученные в нашей работе результаты являются новыми.

Подведем итоги проделанной работы. Для приведения матрицы СЛАУ к верхнему треугольному виду разработан трехмерный СМ, ориентированный на применение в системах реального времени с жесткими требованиями ко времени реакции, а также в вычислительных системах, осуществляющих решение СЛАУ большой размерности. Выполненная работа имеет очевидное продолжение – проведение проектных работ для последующего изготовления разработанного СМ в «железе».

Список использованных источников

- [1] Елфимова Л. Д. *Быстрые алгоритмы для базовой операции клеточных методов линейной алгебры* // Кибернетика и системный анализ. – 2015. – Т. 51. – № 6. – С. 35–45.  ^{↑46}
- [2] Горюшкина А. Е., Семенов С. Г. *Применение систолической системы на основе метода моментов для усовершенствования дискретного преобразования Хартли* // Збірник наукових праць Харківського університету Повітряних Сил. – 2016. – № 1(46). – С. 89–91. ^{↑46}
- [3] Azimian A., Dehkordi A. K., Tehrani M. *A novel systolic array architecture for matrix multiplication circuit design using carbon nanotube technology* // International Journal of Computer Applications. – 2017. – Vol. 172. – No. 6. – Pp. 1–4.  ^{↑46}
- [4] Causapruno G., Riente F., Turvani G., Vacca M., Massimo R. R., Zamboni M., Graziano M. *Reconfigurable systolic array: from architecture to physical design for NML* // IEEE Transactions on Very Large Scale Integration (VLSI) Systems. – 2016. – Vol. 24. – No. 11. – Pp. 3208–3217.   ^{↑46}
- [5] Randjelovic B. M., Milovanovic E. I., Milovanovic I. Z. *Systolic algorithms for matrix multiplication on space optimal one-dimensional systolic arrays* // Facta Universitatis (Nis). Series: Mathematics and Informatics. – 2014. – Vol. 29. – No. 3. – Pp. 243–259.  ^{↑46}

- [6] Abdollahi M. M., Tehrani M. *Designing a novel reversible systolic array using QCA* // Italian Journal of Science & Engineering.– 2017.– Vol. 1.– No. 3.– Pp. 158–166. ^{↑46}
- [7] Воеводин В. В. *Математические модели и методы в параллельных процессах.*– М.: Наука.– 1986.– 296 с. ^{↑46}
- [8] Бабенко В. Н., Ивановский О. Я., Тимонов Д. А. *Систолическая структура обратного хода метода Гаусса и ее основные свойства* // Экологический вестник научных центров Черноморского экономического сотрудничества.– 2018.– Т. 15.– № 1.– С. 8–14. ^{↑46}
- [9] С. Гун, Х. Уайтхаус, Т. Кайлат (ред.) *Сверхбольшие интегральные схемы и современная обработка сигналов*, пер. с англ. под ред. В. А. Лексаченко.– М.: Радио и связь.– 1989.– ISBN 5-256-00179-5.– 472 с. ^{↑47}
- [10] Бабенко В. Н. *Устройство вращения вектора*, Пат. 2475830 Российская Федерация, МПК G 06 F 17/16, № 2010134147/08, заявл. 13.08.10, опубл. 20.02.13, Бюл. № 5.– 11 с. ^{↑48}
- [11] Бабенко В. Н. *Устройство нормировки вектора*, Пат. 2449354 Российская Федерация, МПК G 06 F 17/16, № 2010134146/08, заявл. 13.08.10, опубл. 27.04.12, Бюл. № 12.– 7 с. ^{↑48}
- [12] Бабенко В. Н. *Алгоритм инверсии делителя* // Экологический вестник научных центров Черноморского экономического сотрудничества.– 2013.– Т. 10.– № 4-1.– С. 19–25. ^{↑48}
- [13] Бабенко В. Н. *Аспекты машинного решения систем линейных уравнений*, монография.– Краснодар: Экоинвест.– 2020.– ISBN 978-5-94215-531-5.– 307 с. ^{↑48}
- [14] Мальцев А. А., Пестрецов В. А., Хоряев А. В., Масленников Р. О. *Метод QR-разложения комплексной матрицы с помощью треугольного систолического массива* // Сборник научных трудов «Проблемы разработки перспективных микроэлектронных систем» (9–12 октября 2006 г.), ред. А. Л. Стемпковский, М.: ИПИМ РАН.– 2006.– С. 91–96. ^{↑52}
- [15] Тыртышников Е. Е. *Параллельные методы для обобщенно-тёплицевых систем* // Журнал вычислительной математики и математической физики.– 1996.– Т. 36.– № 6.– С. 5–19. ^{↑53}
- [16] Xiaojun W., Miriam L. *A truly two-dimensional systolic array FPGA implementation of QR decomposition* // ACM Transactions on Embedded Computing Systems.– 2009.– Vol. 9.– No. 1.– id. 3.– 17 pp. ^{↑53}
- [17] Jianfeng Z., Paul C., Hengzhu L. *CORDIC-based enhanced systolic array architecture for QR decomposition* // ACM Transactions on Embedded Computing Systems.– 2015.– Vol. 9.– No. 2.– id. 9.– 22 pp. ^{↑53}
- [18] Semih A., Sufeng N., Jafar S. *FPGA implementation of fast QR decomposition based on givens rotation* // 2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS) (05–08 August 2012, Boise, ID, USA).– IEEE.– 2012.– ISBN 9781467325264.– Pp. 470–473. ^{↑53}
- [19] Dongdong C., Mihai S. *Fixed-point CORDIC-based QR decomposition by givens rotations on FPGA* // 2011 International Conference on Reconfigurable Computing and FPGAs (30 November 2011 – 02 December 2011, Cancun, Mexico).– IEEE.– 2012.– ISBN 978-0-7695-4551-6.– Pp. 327–332. ^{↑53}

- [20] Григорьев В. Р. *Методы параллельной цифровой обработки информации в трехмерных оптических интегральных схемах*, Дис. ... канд. техн. наук : 05.13.17. – М. – 2005. – 231 с.  [↑53](#)
- [21] Yamazaki I., Kurzak J., Luszczek P., Dongarra J. *Design and implementation of a large scale tree-based QR decomposition using a 3D virtual systolic array and a lightweight runtime* // *Parallel Processing Letters*. – 2014. – Vol. **24**. – No. 04. – id. 1442004.  [↑53](#)
- [22] Paolo G., Christian P. *High level synthesis implementation of a three-dimensional systolic array architecture for matrix multiplications on Intel Stratix 10 FPGAs*. – 2021. – 9 pp. arXiv:  2110.11521 [cs.AR]  [↑54](#)

Поступила в редакцию 28.11.2024;
 одобрена после рецензирования 04.12.2024;
 принята к публикации 18.02.2025;
 опубликована онлайн 24.02.2025.

Рекомендовал к публикации

к.т.н. С. А. Амелькин

Информация об авторах:



Виктор Николаевич Бабенко

к.ф.-м.н., доц. кафедры общей математики Кубанского государственного технологического университета. Научные интересы: численные методы.

 0009-0002-8341-0170
 e-mail: rnibvd@mail.ru



Артём Павлович Невечеря

ст. преподаватель кафедры математических и компьютерных методов Кубанского государственного университета. Научные интересы: математическое моделирование динамики незамкнутых сложных систем, экстремальные задачи, системы искусственного интеллекта.

 0000-0001-6736-4691
 e-mail: artiom1989@mail.ru

Вклад авторов: В. Н. Бабенко – 75% (идея, методология, написание черновой версии); А. П. Невечеря – 25% (валидация, доработка и редактирование).

Декларация об отсутствии личной заинтересованности: *благополучие авторов не зависит от результатов исследования.*



On the implementation of QR-decomposition on a three-dimensional systolic array

Viktor Nikolaevich **Babenko**¹, Artyom Pavlovich **Nevecherya**²

¹ Kuban State Technological University, Krasnodar, Russia

² Kuban State University, Krasnodar, Russia

Abstract. Intensive data flows, formed systems of linear equations in real time, as well as systems of linear equations of large dimensionality cause the involvement of systolic arrays for their machine solution. In the presented systolic array, designed to reduce matrices to triangular form, the realization of orthogonal rotation transformations can be carried out both by two-dimensional vector rotation devices CORDIC, and its modifications. For the proposed systolic array, descriptions of its configuration, operation and technical characteristics, as well as the structure of input and output data flow are given. (*In Russian*).

Key words and phrases: systolic array, functional unit, cycle duration, data routing and scheduling, two-dimensional vector rotation unit, forward stroke

2020 *Mathematics Subject Classification:* 68M20; 37E45, 68M14

For citation: Viktor N. Babenko, Artyom P. Nevecherya. *On the implementation of QR-decomposition on a three-dimensional systolic array.* Program Systems: Theory and Applications, 2025, **16**:1(64), pp. 45–59. (*In Russ.*). https://psta.psiras.ru/read/psta2025_1_45-59.pdf

References

- [1] L. D. Elfimova. “Fast algorithms for the basic operation of cellular linear algebra methods”, *Kibernetika i sistemnyj analiz*, **51**:6 (2015), pp. 35–45 (in Russian).
- [2] A. E. Goryushkina, S. G. Semenov. “Application of the systolic system based on the method of moments to improve the discrete Hartley transform”, *Zbirnik naukovix prac' Xarkivs'kogo universitetu Povitryanix Sil*, 2016, no. 1(46), pp. 89–91 (in Russian).
- [3] A. Azimian, A. K. Dehkordi, M. Tehrani. “A novel systolic array architecture for matrix multiplication circuit design using carbon nanotube technology”, *International Journal of Computer Applications*, **172**:6 (2017), pp. 1–4. 
- [4] G. Causaprano, F. Riente, G. Turvani, M. Vacca, R. R. Massimo, M. Zamboni, M. Graziano. “Reconfigurable systolic array: from architecture to physical design for NML”, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, **24**:11 (2016), pp. 3208–3217.  
- [5] B. M. Randjelovic, E. I. Milovanovic, I. Z. Milovanovic. “Systolic algorithms for matrix multiplication on space optimal one-dimensional systolic arrays”, *Facta Universitatis (Nis). Series: Mathematics and Informatics*, **29**:3 (2014), pp. 243–259. 
- [6] M. M. Abdollahi, M. Tehrani. “Designing a novel reversible systolic array using QCA”, *Italian Journal of Science & Engineering*, **1**:3 (2017), pp. 158–166. 
- [7] V. V. Voevodin. *Mathematical Models and Methods in Parallel Processes*, Nauka, M., 1986, 296 pp. (in Russian).
- [8] V. N. Babenko, O. Ya. Ivanovskij, D. A. Timonov. “Systolic structure of the Gauss method return stroke and its main properties”, *Ekologicheskij vestnik nauchnyx centrov Chernomorskogo ekonomicheskogo sotrudnichestva*, **15**:1 (2018), pp. 8–14 (in Russian). 
- [9] S. Y. Kung, H. J. Whitehouse. *VLSI and Modern Signal Processing*, ed. T. Kailath, Prentice Hall, Englewood Cliffs, NJ, 1985, ISBN 013942699X, xiii+481 pp.
- [10] V. N. Babenko. *Vector rotation device*, Pat. 2475830 Rossijskaya Federaciya, MPK G 06 F 17/16, No 2010134147/08, zayavl. 13.08.10, opubl. 20.02.13, Byul. No 5, 11 pp. (in Russian).
- [11] V. N. Babenko. *Vector normalization device*, Pat. 2449354 Rossijskaya Federaciya, MPK G 06 F 17/16, No 2010134146/08, zayavl. 13.08.10, opubl. 27.04.12, Byul. No 12, 7 pp. (in Russian).
- [12] V. N. Babenko. “Divisor inversion algorithm”, *Ekologicheskij vestnik nauchnyx centrov Chernomorskogo ekonomicheskogo sotrudnichestva*, **10**:4-1 (2013), pp. 19–25 (in Russian).
- [13] V. N. Babenko. *Aspects of Machine Solution of Systems of Linear Equations*, monograph, Ekoinvest, Krasnodar, 2020, ISBN 978-5-94215-531-5, 307 pp. (in Russian).

- [14] A. A. Mal'cev, V. A. Pestrecov, A. V. Xoryaev, R. O. Maslennikov. "QR decomposition method of complex matrix using triangular systolic array", *Sbornik nauchnyx trudov "Problemy razrabotki perspektivnyx mikroelektronnyx sistem"* (9–12 oktyabrya 2006 g.), ed. A. L. Stempkovskij, IPPM RAN, M., 2006, pp. 91–96 (in Russian). 
- [15] E. E. Tyrtysnikov. "Parallel methods for generalized Toeplitz systems", *Comput. Math. Math. Phys.*, **36**:6 (1996), pp. 691–702.
- [16] W. Xiaojun, L. Miriam. "A truly two-dimensional systolic array FPGA implementation of QR decomposition", *ACM Transactions on Embedded Computing Systems*, **9**:1 (2009), id. 3, 17 pp. 
- [17] Z. Jianfeng, C. Paul, L. Hengzhu. "CORDIC-based enhanced systolic array architecture for QR decomposition", *ACM Transactions on Embedded Computing Systems*, **9**:2 (2015), id. 9, 22 pp. 
- [18] A. Semih, N. Sufeng, S. Jafar. "FPGA implementation of fast QR decomposition based on givens rotation", *2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS)* (05–08 August 2012, Boise, ID, USA), IEEE, 2012, ISBN 9781467325264, pp. 470–473.  
- [19] C. Dongdong, S. Mihai. "Fixed-point CORDIC-based QR decomposition by givens rotations on FPGA", *2011 International Conference on Reconfigurable Computing and FPGAs* (30 November 2011 – 02 December 2011, Cancun, Mexico), IEEE, 2012, ISBN 978-0-7695-4551-6, pp. 327–332. 
- [20] V. R. Grigor'ev. *Methods of parallel digital information processing in three-dimensional optical integrated circuits*, Dis. ... kand. texn. nauk : 05.13.17, M., 2005, 231 pp. (in Russian). 
- [21] I. Yamazaki, J. Kurzak, P. Luszczyk, J. Dongarra. "Design and implementation of a large scale tree-based QR decomposition using a 3D virtual systolic array and a lightweight runtime", *Parallel Processing Letters*, **24**:04 (2014), id. 1442004. 
- [22] G. Paolo, P. Christian. *High level synthesis implementation of a three-dimensional systolic array architecture for matrix multiplications on Intel Stratix 10 FPGAs*, 2021, 9 pp. arXiv:2110.11521~[cs.AR] 

УДК 004.89:[616.83+331.582.2]

 10.25209/2079-3316-2025-16-1-61-82


Комплекс онтологий как модель системы интеллектуальной поддержки в реабилитации пациентов, перенесших инсульт

Валерия Викторовна **Грибова**¹, Валентин Борисович **Шуматов**², Сергей Васильевич **Лебедев**³, Елена Арефьевна **Шалфеева**⁴, Евгений Юрьевич **Шестопалов**⁵, Дмитрий Борисович **Окунь**^{6,8}, Роман Игоревич **Ковалев**⁷, Екатерина Ивановна **Шепета**⁸, Леонид Александрович **Федорищев**⁹, Александр Яковлевич **Лифшиц**¹⁰

^{1, 4, 6, 7, 9, 10} Институт автоматизации и процессов управления ДВО РАН, Владивосток, Россия

^{2, 3, 5, 8} ФГБОУ ВО «Тихоокеанский государственный университет» Министерства здравоохранения Российской Федерации, Владивосток, Россия

Аннотация. Главным стимулом для внедрения компьютерных технологий в систему здравоохранения является стремление к значительному улучшению качества жизни людей. Это включает повышение качества и скорости лечения, уменьшение затрат на медицинские услуги и приобретение эффективных средств для соблюдения нормативных требований.

На современном этапе развития реабилитологии становится очевидной необходимостью активного внедрения систем поддержки принятия врачебных решений и технологий искусственного интеллекта. Эти технологии способны существенно улучшить понимание клинических аспектов нарушений, уровня активности и участия пациентов, перенесших инсульт, в процессе реабилитации. Ключевым компонентом успешного применения этих систем является значимость формализации знаний и создание онтологий, которые обеспечивают структурированное и связанное представление медицинской информации, определяют правила их интерпретации.

В данной работе представлен комплекс взаимосвязанных онтологических моделей, лежащих в основе разрабатываемой интеллектуальной системы поддержки решений в реабилитации пациентов, перенесших инсульт. Для реализации комплекса онтологий используется облачная платформа IASPaas. Онтологии и генерируемые на их основе целевые ресурсы выступают базовыми элементами разрабатываемой системы, которая в ближайшее время будет предоставлена специалистам здравоохранения для решения актуальных вопросов реабилитации. Предусмотрены механизмы планового расширения и уточнения базы знаний, что позволит системе легко адаптироваться к новым результатам медицинских исследований и оптимизировать работу в целом.

Ключевые слова и фразы: система поддержки принятия врачебных решений, интеллектуальный сервис, реабилитация, инсульт, база знаний, онтологический подход, инженерия знаний

Благодарности: Проект реализуется коллективом сотрудников ФГБОУ ВО ТГМУ МЗ РФ и ФГБУН ИАПУ ДВО РАН в рамках субсидии реализации программы стратегического академического лидерства "Приоритет-2030" и при поддержке Минобрнауки России № FFWF-2021-0004

Для цитирования: Грибова В.В., Шуматов В.Б., Лебедев С.В., Шалфеева Е.А., Шестопалов Е.Ю., Окунь Д.Б., Ковалев Р.И., Шепета Е.И., Федорищев Л.А., Лифшиц А.Я. Комплекс онтологий как модель системы интеллектуальной поддержки в реабилитации пациентов, перенесших инсульт // Программные системы: теория и приложения. 2025. Т. 16. № 1(64). С. 61–82. https://psta.psiras.ru/read/psta2025_1_61-82.pdf

Введение

Системы поддержки принятия решений (СППР) используются во многих сферах человеческой деятельности. Их популярность неуклонно возрастает. Это связано с массовой компьютеризацией разных сфер нашей жизни: политика, соцзащита, здравоохранение, образование и прочее. Всё больше людей вовлекаются в процесс принятия важных решений, одновременно растёт ответственность за их последствия. При этом многие области интересов и деятельности (здравоохранение, образование, юриспруденция, культура и др.) слабо формализуемы по своей сути [1–7]. Накопленные в этих областях знания и данные требуют специальных средств для их представления, организации доступа к ним и использования. Владение этими средствами, основанными на знаниях, необходимо для принятия грамотных, взвешенных решений. Поэтому так важны интеллектуальные системы, способные оказать помощь лицам, принимающим решения, в особенности в предметных областях (ПрОбл) с высокой ответственностью за их последствия.

В последние годы развитие медицинской реабилитации привлекает значительное внимание, акцентируясь на важности персонализации и раннего начала этого процесса. Эти принципы отражают глобальные тенденции в предоставлении реабилитационных услуг в различных национальных системах здравоохранения. Они заложены в современные стандарты медицинской реабилитации; активно применяются клинические рекомендации [8, 9], ориентированные на пациента и нацеленные на улучшение результатов реабилитации [10].

Раннее начало реабилитационных мероприятий способствует быстрому и полному восстановлению функций организма. Персонализация процесса реабилитации основана на учете индивидуальных особенностей пациента, обеспечивая максимально эффективное восстановление. Персонализированные программы лечения включают формулирование диагноза, постановку целей и планирование вмешательств, уделяют внимание улучшению взаимодействия между пациентом и терапевтом, – всё это играет ключевую роль в достижении реабилитационных целей [11].

Сообщество разработчиков интеллектуальных систем разрабатывает и апробирует модели и методы для создания систем поддержки реабилитации пациентов, способствующих персонализации данного процесса. Для формирования обоснованных рекомендаций используются знания, соответствующие реальным знаниям специалистов в этой области. В качестве эффективного инструмента для представления сложных знаний и имитации человеческих рассуждений зарекомендовали себя онтологические модели. Такие модели включают в себя правила структурированной организации информации, а также её интерпретации и логического вывода [12].

Цель работы – представить комплекс взаимосвязанных онтологических моделей, формирующих основу для разработки интеллектуальной системы поддержки персонифицированного принятия решений для реабилитации пациентов, перенесших инсульт.

Материалы и методы

Для разработки комплекса онтологических моделей, формирующих основу систем поддержки принятия врачебных решений в реабилитации пациентов, перенесших инсульт, нужно выполнение задач:

- (1) Проанализировать предметную область, выделенные понятия и сгруппировать их по типам.
- (2) Создать комплекс онтологий для представления знаний для реабилитации пациентов, перенесших инсульт.

Анализ предметной области персонифицированной реабилитации пациентов после инсульта проводился непосредственно с экспертами, обладающими практическим опытом реабилитации и применения клинических рекомендаций и стандартов медицинской реабилитации, и стал продолжением работы коллектива по структуризации и формализации знаний о патологиях системы кровообращения.

Создание комплекса онтологических моделей представляет собой один из наиболее интеллектуально сложных и ответственных этапов проектирования систем поддержки принятия врачебных решений (СППВР).

Разработка комплекса онтологий велась на облачной платформе *IACPaaSSM*, реализующей двухуровневый подход для представления и формирования данных и знаний, где уровень, определяющий смысл понятий ПрОбл и их зависимостей, определяет форму изложения информации на уровне данных и знаний. Поскольку при таком подходе онтологическая модель явно отделена от генерируемых на ее основе ресурсов (коллекций данных, знаний, документов), обеспечивается независимое развитие важных компонентов системы (рисунок 1) [13], в частности, это позволяет модифицировать знания, не внося изменений в программный код.

Предлагаемый онтологический подход в сочетании с облачностью платформы *IACPaaS* дают возможность независимого участия в разработке группам специалистов разных компетенций, включая экспертов, предоставляющих и оценивающих знания ПрОбл. Подход обеспечивает связывание, согласованность и повторную используемость онтологических компонентов [13].

Для формального представления онтологических моделей используется язык описания метаинформации *IACPaaS* – орграфовой модели связей понятий [14, 15]. Язык предоставляет средства спецификации моделей

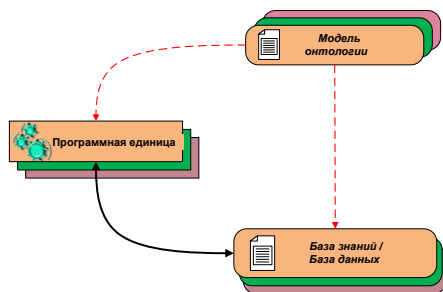


Рисунок 1. Онтологический двухуровневый подход к формированию данных и знаний: отделение предметных знаний от знаний о методах решения задач [13]

онтологий в форме размеченных иерархических орграфов: разметка определяет правила формирования орграфов целевой информации. Формируемая на основе моделей онтологий семантическая информация (например, знания ПрОбл) представляется такого же вида иерархическими орграфами.

Состав комплекса онтологических моделей, достаточного для формирования компонентов систем поддержки решений по реабилитации пациентов, определяется следующими факторами. Прежде всего, необходимо обеспечить возможность структуризации и формализации всей необходимой информации о тактике ведения пациентов, перенесших инсульт, а также о ключевых характеристиках используемых для этого методов исследования. Соответственно, должна быть обеспечена возможность поддержки внесения такой информации о пациенте, сохранения в виде информационных ресурсов и формирования отчетных документов.

Для обеспеченности данных процессов необходимы справочники и «терминологические» базы: уникальные (словарь жалоб пациента, шаблоны осмотра специалистов мультидисциплинарной реабилитационной командой (МДРК), протоколы заключений осмотров и др.) и универсальные (справочник Международной классификации функционирования, ограничений жизнедеятельности и здоровья (МКФ)).

Ресурсы для поддержки внесения клинических данных должны содержать описание шкал и методик осмотра, обследования и опроса пациентов. Ресурсы для документирования и хранения клинических данных должны отражать медицинскую реабилитационную карту. Ресурсы для сотрудничества реабилитационной бригады должны отражать участников и роли МДРК в данном процессе. Ресурсы для интеллектуальной поддержки должны содержать знания ПрОбл.

Выбор и применение облачной платформы IASPaas для создания СППВР дает возможность независимого участия разным группам специалистов как практического, так и научного звеньев здравоохранения. Для создания ресурсов перечисленных типов участвуют эксперты, инженеры, научные кадры.

Результаты

Совместно со специалистами практического здравоохранения проведен системный анализ, формализация и структуризация реабилитационного процесса больных, перенесших инсульт. Неоднократно пересмотрены и уточнены наблюдения, их характеристики с возможными значениями, важные как для описания патологических процессов, так и самой процедуры реабилитации при данной патологии.

Помимо предметно-ориентированных структур для знаний, документов и номенклатуры терминов, применяемых специалистами, были определены структуры для декларативных ресурсов-компонентов, обеспечивающих гибкость и масштабируемость программной «части» (компонентов) СППВР, в частности пользовательского интерфейса (GUI).

В результате был сформирован комплекс моделей онтологий и онтологических баз знаний, данных и терминов, необходимый для интеллектуальной поддержки принятия врачебных решений при реабилитации пациентов, перенесших инсульт. Полный перечень семантических моделей представлен на рисунке 2 (стрелками схематично указаны их связи с ресурсами других уровней).

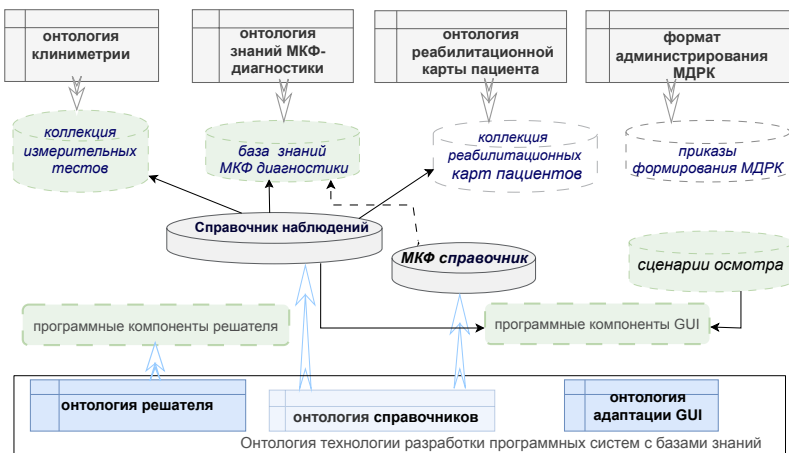


Рисунок 2. Схема комплекса онтологий

о наличии определенного нарушения, кодируемого классификатором МКФ в рамках интеллектуальной поддержки МДРК. Он представляет набор отношений, достаточных для связывания элементов клинической картины пациента с соответствующим кодом МКФ с определителем степени тяжести.

Ресурс группирует знания под узлами-вершинами: *B, D, S, E* – МКФ домены: функции организма, структуры организма, активность и участие, факторы окружающей среды. Информация соответствует элементам структуры:

МКФназвание – код и имя МКФ классификатора;

МКФкод – диагностируемый код МКФ классификатора с соответствующим определителем;

Симптомокомплекс – предназначен для группировки клинически значимых наблюдений, включая результаты тестов.

Факт-наблюдение – единица, отражающая клинически значимые данные пациента. Каждое наблюдение можно выразить как в простом варианте его описания, так и используя множество элементов или характеристик. Результаты проведения тестирования возможно указать как по итоговому балу, так и по результатам отдельных вопросов.

Тип наблюдения – представляет принадлежность к категории, такой как *метрика, наследственность, образ жизни, лабораторное исследование, инструментальное исследование, тест, осмотр, заболевание* и т.д.) для дифференцировки вводимых элементов;

Балл – условная единица, используемая для определения веса результатов тестирования;

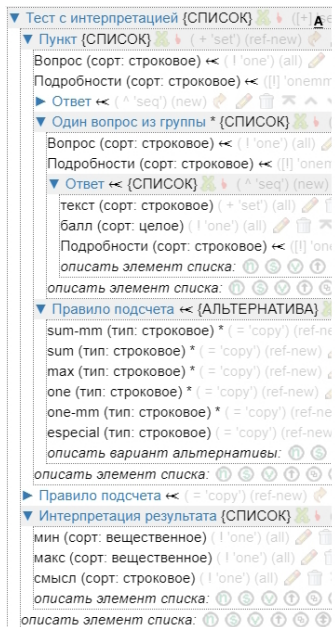
Совместимость элементов – используется для определения полноты диагностируемых состояний через определение сочетаемости наблюдений. Используются следующие определения: *минимальное количество, любой, все*.

Знания о правилах кодирования нарушений организма пациента являются незаменимым ресурсом для поддержки принятых решений в работе МДРК при реабилитации после инсульта. На рисунке 36 представлен фрагмент базы знаний на основе предложенной структуры. Клинические данные, используемые в диагностике определенного кода МКФ формализованы и представлены в формате, привычном специалистам реабилитологам. Каждый тест/шкала имеет строгий порядок исследуемых параметров и данные об интерпретации полученных результатов.

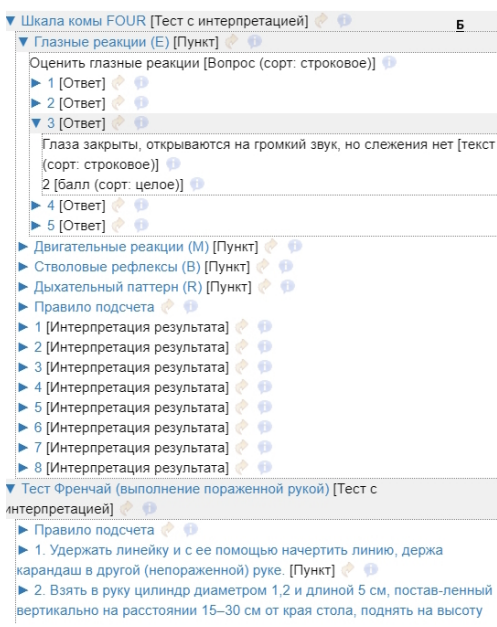
Предусмотрены механизмы планового расширения и уточнения базы знаний. Специалисты этой компетенции и эксперты готовят обновление

знаний о диагностических методах и вносят их в независимые облачные ресурсы платформы IACPaasS. После оценивания и апробации знаний на контрольных тестовых случаях обновленные ресурсы заменяют ранее внедренные компоненты системы.

Сформирован ресурс – онтология клиниметрии – для правил формального описания шкал-тестов с интерпретацией (см. рисунок 4а).



(а) фрагмент онтологии



(б) фрагмент формируемого ресурса

РИСУНОК 4. Фрагменты онтологии МКФ-диагностики и формируемого на этой основе ресурса

Онтология отражает структуру, позволяющую представить тест со всеми возможными вопросами, ключами ответов и интерпретацией результатов. В процессе анализа структуры используемых тестов для диагностирования реабилитационных возможностей пациента были определены следующие ключевые вершины: *Пункт*, *Вопрос*, *Ответ*, *Правило подсчета*.

Кроме того, была выявлена когорта тестов, в которых используется группировка вопросов. Для таких случаев введено понятие *Один вопрос из группы*, которое позволяет определять свою пару *Вопрос* – *Ответ* и величину соответствующих ответу баллов.

Введено понятие *Подробности* для структурирования длинных текстов

вопросов. Вопрос получает расшифровку или уточнение, открывая возможность представления в текстовом варианте, привычном специалистам. При использовании наполненного тестами ресурса баллы используются для дальнейшего расчета результатов теста на основе *Правил подсчета* (например, суммирования всех баллов).

Для поддержки внесения клинических данных процесса осмотра, обследования и опроса пациентов, сформирован справочник наблюдений (для реабилитации), см. рисунок 5. Была использована онтология справоч-

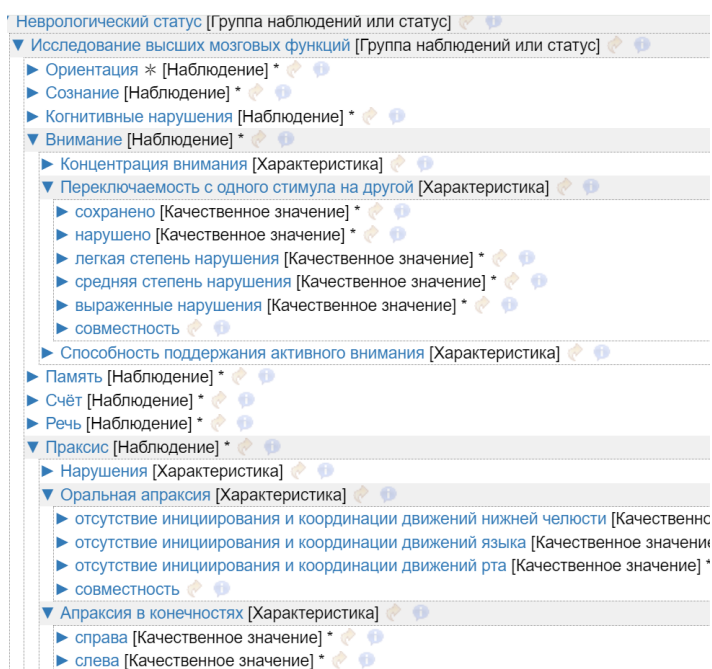


Рисунок 5. Фрагмент справочника наблюдений для реабилитации

ников, апробированная на медицинском портале платформы IASPaas. В нем группируются перечни Наблюдений разного типа. Онтология позволяет описать интересующее *Наблюдение* с использованием *простого* представления или с использованием описания данного наблюдения через множество *Характеристик*. Данный процесс регулируется через управляющий узел *Альтернатива типа наблюдений*. Данный узел в генерируемом целевом ресурсе позволяет выбрать только одно направление для описания наблюдения.

ных осмотров и ведения дневников ежедневных осмотров. Ввод данных в медицинскую карту осуществляется в формализованном виде: каждый раздел медицинской карты структурирован под соответствующий класс данных, используемых в медицинской практике.

Кроме этого, в медицинской карте предусмотрено пространство для внесения и хранения гипотез, генерируемых СПИВР. На рисунке 7

▼ Реабилитационная медицинская карта (история болезни) А

История болезни или наблюдений

- ▶ Сведения исследователя и организации < ([=] 'copymm') (ref-new)
- ▶ Сводная часть < * ([=] 'copy') (new)
- ▶ Осмотры узких специалистов МДРК (СПИСОК) ([=] 'copy') (ref-new)
- ▶ Записи специалиста ([+] 'setmm') (ref-new)
- ▶ Реабилитационная комиссия ([=] 'copymm') (ref-new)
- ▶ Диагноз < ([=] 'copymm') (ref-new)
- ▶ дата обращения (сорт: дата и время) ([one] (new))
- ▶ Эпикриз ([=] 'copymm') (ref-new)
- ▶ История настоящего заболевания * ([=] 'copymm') (ref-new)
- ▶ Жалобы * (СПИСОК) ([=] 'copy') (new)
- ▶ Жалобы при обращении (Прежде всего: признаки болезненного состояния. Кроме того, признаки, важные с точки зрения врача.)
- ▶ Наблюдение < ([+] 'listmm') (ref-new)
- ▶ Объективное состояние < * (СПИСОК) ([=] 'copy') (new)
- ▶ Объективное состояние пациента в день обращения. (Заполняются важные с точки зрения врача наблюдения.)
- ▶ Общий осмотр ([=] 'copy') (new)
- ▶ Группа или система < ([+] 'setmm') (ref-new)
- ▶ Исследования инструментальные * (СПИСОК) ([=] 'copy') (new)
- ▶ Данные инструментальных исследований
- ▶ Вид исследования < ([+] 'setmm') (all)
- ▶ Исследования лабораторные * (СПИСОК) ([=] 'copymm') (new)
- ▶ Данные лабораторных исследований
- ▶ Вид исследования < ([+] 'setmm') (all)
- ▶ Анамнез жизни * ([=] 'copymm') (ref-new)
- ▶ Дополнительная информация ([=] 'copymm') (new)
- ▶ Предварительный диагноз ([=] 'copymm') (ref-new)
- ▶ Дневники (СПИСОК) ([=] 'copymm') (new)
- ▶ Номер записи (СПИСОК) ([*] 'seqmm') (new)
- ▶ дата (сорт: дата и время) ([l] 'onemmm') (all)
- ▶ Жалобы * ([=] 'copymm') (ref-new)
- ▶ Объективное состояние < * ([=] 'copymm') (ref-new)
- ▶ Сведения исследователя и организации < ([=] 'copy') (ref-new)
- ▶ заключение (сорт: строковое) ([+] 'setmm') (all)
- ▶ планируемое обследование (сорт: строковое) ([+] 'setmm') (all)
- ▶ Синдромы и патологические состояния * ([=] 'copymm') (ref-new)
- ▶ Вычисляемые показатели * ([=] 'copymm') (ref-new)
- ▶ Прекращение наблюдения ([=] 'copymm') (ref-new)
- ▶ Лечение, назначенное врачом ([=] 'copymm') (new)
- ▶ Компьютерная диагностика ([=] 'copy') (ref-new)

▼ Осмотры узких специалистов МДРК Б

Логопед [Записи специалиста]

▼ 20.06.2024 [Дата осмотра специалистом]

- ▶ Жалобы
- ▶ Нарушения глотания [Наблюдение]
- ▶ Нарушение речи [Наблюдение]
- ▶ Слюнотечение [Наблюдение]
- ▶ Шкала оценки дисартрии [Наблюдение]
- ▶ Нечеткость речи [Наблюдение]
- ▶ Боль в челюсти [Наблюдение]
- ▶ Сухость во рту [Наблюдение]
- ▶ Боль в челюсти [Наблюдение]
- ▶ Боль в челюсти [Наблюдение]
- ▶ Боль в челюсти [Наблюдение]
- ▶ Поперхивание [Наблюдение]
- ▶ Логопедический статус [Статус специалиста]
- ▶ Запросы на реабилитацию
- ▶ Тесты и опросы
- ▶ Краткая шкала оценки психического статуса (MMSE)
- ▶ Ориентация во времени [Пункт]
- ▶ 1 [Ответ]
- ▶ 3 [балл (сорт: целое)]
- ▶ Шкала Ходкинсона [Тест с интерпретацией]
- ▶ Итог теста
- ▶ 2 [итоговый балл (сорт: целое)]
- ▶ МКФ-домены
- ▶ Дневник
- ▶ 21.07.2024 [Дата осмотра специалистом]
- ▶ 22.07.2024 [Дата осмотра специалистом]
- ▶ 23.07.2024 [Дата осмотра специалистом]
- ▶ Тесты и опросы
- ▶ Жалобы
- ▶ Логопедический статус [Статус специалиста]
- ▶ МКФ-домены
- ▶ 24.07.2024 [Дата осмотра специалистом]
- ▶ 25.07.2024 [Дата осмотра специалистом]
- ▶ Врач ФРМ [Записи специалиста]
- ▶ Объективное состояние *
- ▶ Диагноз
- ▶ Исследования инструментальные *
- ▶ Жалобы *
- ▶ Компьютерная диагностика
- ▶ Логопед [Записи специалиста]
- ▶ Врач ФРМ [Записи специалиста]

(а) Фрагмент онтологии

(б) Фрагмент формируемого ресурса

Рисунок 7. Фрагменты онтологии реабилитационной медицинской карты для ведения пациентов, находящихся на реабилитации и формируемого на этой основе ресурса

демонстрируется фрагмент онтологии для документирования клинических данных в виде медицинской реабилитационной карты. Совокупность заполненных медицинских карт формирует информационный ресурс – коллекцию реабилитационных медицинских карт пациента.

Для сотрудничества реабилитационной бригады предусмотрены сценарии осмотра пациентов и формирования реабилитационной карты, представляемые набором вкладок с определенными типами интерфейсных элементов. Ресурс для описания сценариев сотрудничающих *специалистов* должен отражать их роли в МДРК и в данном процессе. На рисунке 8 демонстрируется фрагмент универсальной онтологии, отвечающей за

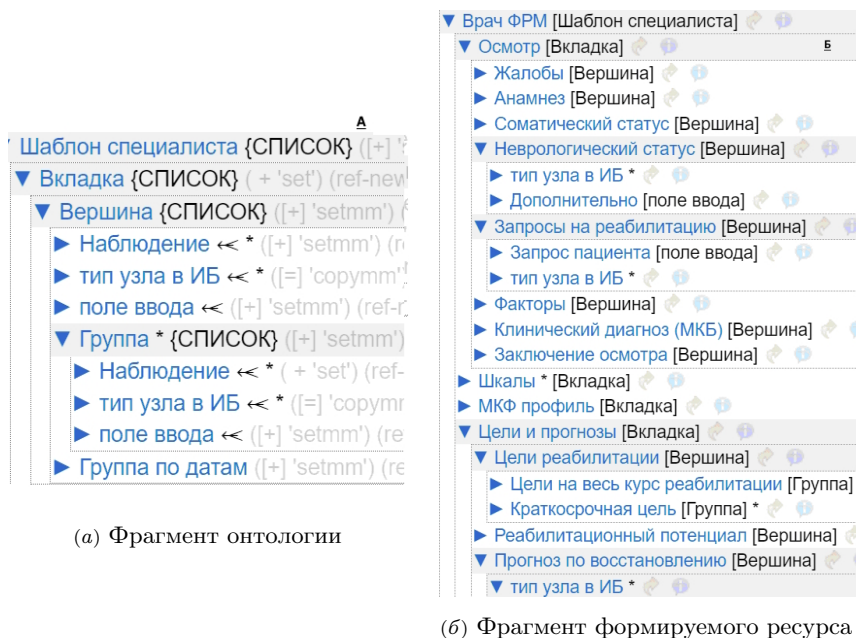
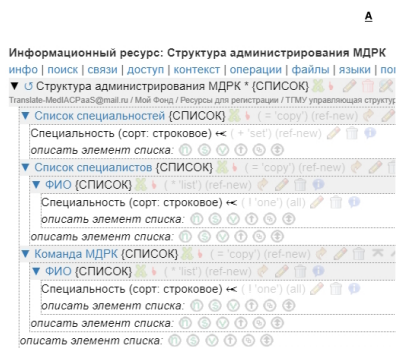


Рисунок 8. Фрагменты онтологии, отвечающей за управление вкладок и меню GUI и формируемого на его основе управляющего ресурса поддержки тактики ведения пациентов

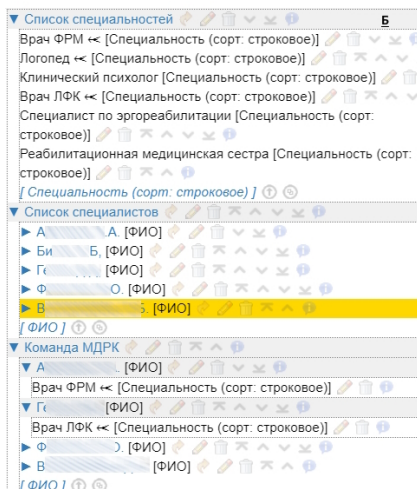
управление вкладок и меню интерфейса – онтология адаптации GUI. Данная онтология формирует правила ввода данных (наблюдений разной сложности) в пользовательском интерфейсе через комплекс *Вкладок* с множеством элементов типа *Вершина*. Такая онтология апробирована и применяется для формирования структурированных *шаблонов* ввода информации разными *специалистами* команды. Онтология также определяет места на вкладках (вершины) для вывода отчетов. Данная

логика позволяет независимо от ИТ-специалистов формировать интерфейс для решения вопросов реабилитации пациентов конкретного класса, в частности, перенесших инсульт.

Для поддержки администрирования создана онтология администрирования МДРК (см. рисунок 9), задающая формат хранения формируемого



(а) Фрагмент онтологии



(б) Фрагмент формируемого ресурса

Рисунок 9. Фрагменты онтологии, отвечающей за формирование МДРК и формируемого на этой основе ресурса

списка участников процесса реабилитации с указанием медицинской специальности и принадлежностью команде МДРК. В данной структуре имеется 3 корневых понятия: *Список специальностей*, *Список специалистов* и *Команда МДРК*. *Список специальностей* - перечень специальностей. *Список специалистов* реализован как множество группировок *ФИО* и *Специальность*. Элемент *Команда МДРК* отвечает за формирование команд МДРК через множество пар *ФИО* и *Специальность*.

Онтология технологии IASPaas для разработки программных систем с базами знаний (как технологии искусственного интеллекта) включает онтологию решателя, онтологию адаптации GUI, инструментарий поддержки сборки программных и информационных компонентов в единую систему. Основа сборки – единый комплекс онтологий для всех формируемых и программно обрабатываемых ресурсов. Существующая технология и новый комплекс онтологий позволил реализовать новую версию СППВР для поддержки реабилитолога на первом этапе (см. рисунок 10).

История болезней / ФИО В.В. (фно_123/852)

Технологи лаб 35 (Врач ФРМ)

ОсмотрШкалыМКФ профильЦели и пронозыИГМРДневникИтог работы

Жалобы

Анамнез

Соматический статус

Неврологический статус

Запросы на реабилитацию

Факторы

Клинический диагноз (МКБ)

Закончение осмотра

Поиск...

Жалоба не предъявляет

Общая слабость

Ограничение поворотов в кровати

Ограничение ходьбы

Ограничение ходьбы по неровной поверхности (на улице)

Ограничение ходьбы по лестнице

Ограничение ходьбы на короткие расстояния

13.01.2025 Внесено

Общая слабость: Выраженность - умеренно выраженная

Ограничение поворотов в кровати: имеется

Ограничение ходьбы: умеренное

Ограничение ходьбы по неровной поверхности (на улице): умеренное

Дополнительно:

История болезней / ФИО В.В. (фно_123/852)

Технологи лаб 35 (Врач ФРМ)

ОсмотрШкалыМКФ профильЦели и пронозыИГМРДневникИтог работы

Шкала реабилитационной маршрутизации (ШРМ)

Модифицированная шкала Ранкин

Тестирование возможно

Оценка нарушений жизнедеятельности

Отсутствие существенных нарушений жизнедеятельности, несмотря на наличие некоторых симптомов болезни, способен выполнять все обычные повседневные обязанности

Легкое нарушение жизнедеятельности, неспособен выполнять некоторые прежние обязанности, но справляется с собственными делами без посторонней помощи

Умеренное нарушение жизнедеятельности, потребность в некоторой помощи, но ходит без посторонней помощи

Выраженное нарушение жизнедеятельности, неспособен ходить без посторонней помощи, неспособен справиться со своими телесными (физическими) потребностями без посторонней помощи

Грубое нарушение жизнедеятельности, прикован к постели, недержание кала и мочи, потребность в постоянной посторонней помощи

13.01.2025 внесено:

Оценка нарушений жизнедеятельности: Умеренное нарушение жизнедеятельности, потребность в некоторой помощи, но ходит без посторонней помощи

Итог: 3 балла/ое (Умеренное нарушение жизнедеятельности: требуется некоторая помощь, однако способен ходить без посторонней помощи.)

История болезней / ФИО В.В. (фно_123/852)

Технологи лаб 35 (Врач ФРМ)

ОсмотрШкалыМКФ профильЦели и пронозыИГМРДневникИтог работы

Результат ИИ-Диагностики

13.01.2025 Внесенные МКФ-коды

Введите код МКФ:

Комментарий:

Внести

Цели реабилитации

Цели на весь курс реабилитации

Краткосрочная цель

Реабилитационный потенциал

Прогноз по восстановлению

Задачи реабилитации

Цель на курс реабилитации

самостоятельно пройти 10 метров используя ходунки без помощи ассистента

сказать простые фразы необходимые для выражения основных потребностей или желаний пациента

одеть самостоятельно рубашку и заменить пуговицы без помощи ассистента

приготовить простые блюда и напитки (бутерброд, чай, кофе)

пациент сидит удерживая гравитационный градиент 90 градусов

Данные истории болезни

Логопед

Врач ФРМ

Клинический психолог

Физический терапевт

Специалист по эргореабилитации

Реабилитационная медицинская сестра

Протоколы консилиума МДРК

Объективизированная оценка состояния пациента по шкалам:

Шкала	Результат расчета (баллы/ое)	Заключение
Шкала реабилитационной маршрутизации (ШРМ)	5	Грубое нарушение процессов жизнедеятельности
Модифицированная шкала Ранкин	3	Умеренное нарушение жизнедеятельности: требуется некоторая помощь, однако способен ходить без посторонней помощи
Шкала тяжести инсульта Национального Института здоровья (NIHSS)	22	Крайне тяжелая степень

Жалобы:

Рисунок 10. СППВР для поддержки реабилитолога на первом этапе

Заключение

Представлен комплекс взаимосвязанных онтологических моделей, лежащих в основе разрабатываемой интеллектуальной системы поддержки принятия врачебных решений для реабилитации пациентов, перенесших инсульт.

Разработка комплекса моделей базируется на двухуровневом подходе, где уровень онтологических моделей задает правила структурированного формирования и интерпретации целевой семантической информации (в частности, баз знаний). Для реализации комплекса онтологий используется облачная платформа IACPaaS. Онтологии и генерируемые на их основе целевые ресурсы выступают базовыми элементами разрабатываемой СППР, которая в ближайшее время будет предоставлена специалистам здравоохранения для решения актуальных вопросов реабилитации. Целевые ресурсы (знания) не являются статичными элементами системы; предусмотрены механизмы свободного расширения базы знаний, что позволит системе легко адаптироваться к новым результатам медицинских исследований и оптимизировать работу в целом.

Следующий этап предусматривает расширение комплекса онтологических моделей для формализации и применения знаний, относящихся к оценке реабилитационного потенциала, прогнозированию восстановления и разработке персонализированных реабилитационных программ.

Учитывая, что онтологии отделены от целевых ресурсов – баз знаний, баз данных и справочников, они могут повторно использоваться для реализации других систем поддержки принятия решений в медицине (в том числе на других платформах). По запросу к администратору платформы IACPaaS выдается доступ к ним на платформе. Возможен экспорт этих онтологий в формате json.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов в связи с публикацией данной статьи.

Соблюдение этических стандартов. Данный вид исследования не требует прохождения экспертизы локальным этическим комитетом.

Список использованных источников

- [1] Мальцева С. В. *Применение онтологических моделей для решения задач идентификации и мониторинга предметных областей* // Бизнес-информатика.– 2008.– № 3.– С. 18–24.   ^{↑62}
- [2] Соловьев В. Д., Добров Б. В., Иванов В. В., Лукашевич Н. В. *Онтологии и тезаурусы*.– Казань-Москва: МАКС Пресс.– 2006.– 157 с.  ^{↑62}
- [3] Ломов П. А. *Применение паттернов онтологического проектирования для создания и использования онтологий в рамках интегрированного пространства знаний* // Онтология проектирования.– 2015.– Т. 5.– № 2.– С. 233–245.   ^{↑62}
- [4] Плешкова А. Ю. *Онтологии в управлении образовательным процессом* // Онтология проектирования.– 2022.– Т. 12.– № 4.– С. 506–517.    ^{↑62}
- [5] Лебедев С. В., Жукова Н. А. *Слияние медицинских данных на основе онтологий* // Онтология проектирования.– 2017.– Т. 7.– № 2.– С. 145–159.    ^{↑62}
- [6] Волокитин Ю. И., Куприяновский В. П., Гринько О. В., Покусаев О. Н., Синягов С. А. *Проблемы цифровой экономики и формализованные онтологии* // International Journal of Open Information Technologies.– 2018.– Т. 6.– № 6.– С. 87–96.  ^{↑62}
- [7] González C., Blobel B. G. M. E., Lopez D. M. *Ontology-based framework for electronic health records interoperability* // User Centred Networked Health Care, Proceedings of MIE 2011, Studies in Health Technology and Informatics.– vol. 169, eds. A. Moen et al.– IOS Press.– 2011.– ISBN 978-1-60750-805-2.– Pp. 694–698.   ^{↑62}
- [8] *Приказ Минздрава РФ от 31 июля 2020 г. №788Н «Об утверждении порядка организации медицинской реабилитации взрослых»*.– 2021.  ^{↑62}
- [9] Белкин А.А., Алашеев А.М., Белкин В.А., Белкина Ю.Б., Белова А.Н., Бельский Д.В., Бодрова Р.А., Брюсов Г.П., Варако Н.А., Вознюк И.А., Горбачев В.И., Гречко А.В., Гумарова Л.Ш., Даминов В.Д., Ершов В.И., Жигужевский Р.А., Заболотских И.Б., Зайцев О.С., Захаров В.О., Зинченко Ю.П., Иванова Г.Е., Кауркин С.Н., Князенко П.А., Ковязина М.С., Кондратьев А.Н., Кондратьева Е.А., Костенко Е.В., Кулакова А.А., Крылов В.В., Купрейчик В.Л., Лайшева О.А., Лебединский К.М., Левит А.Л., Лейдерман И.Н., Лубнин А.Ю., Лукоянова Е.А., Мальцева М.Н., Макарова М.Р., Мельникова Е.В., Митраков Н.Н., Мишина И.Е., Помешкина С.А., Петриков С.С., Петрова М.В., Пинчук Е.А., Пирадов М.А., Попугаев К.А., Полякова А.В., Погонченкова И.В., Проценко Д.Н., Расева Н.В., Рудник Е.Н., Руднов В.А., Рябинкина Ю.В., Савин И.А., Сафонова Т.Ю., Семенова Ж.Б., Скрипай Е.Ю., Солодов А.А., Суворов А.Ю., Супонева Н.А., Токарева Д.В., Уразов С.П., Усачев Д.Ю., Фадеева А.С., Хасанова Д.Р., Хасаншин В.Т., Царенко С.В., Чуприна С.Е., Шестопапов Е.Ю., Щеголев А.В., Щикота А.М., Энгаус Р.Е., Норвилл С.Н. *Реабилитация в отделении реанимации и интенсивной терапии (РеаБИТ). Методические*

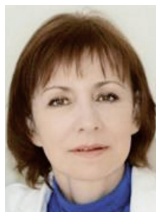
- рекомендации Союза реабилитологов России и Федерации анестезиологов и реаниматологов // Вестник интенсивной терапии имени А.И. Салтанова.– 2022.– № 2.– С. 7–40.    ⁶²
- [10] O’Keeffe M., Cullinane P., Hurley J., Leahy I., Bunzli S., O’Sullivan P. B., O’Sullivan K. *What influences patient-therapist interactions in musculoskeletal physical therapy? Qualitative systematic review and meta-synthesis* // Physical Therapy.– 2016.– Vol. **96**.– No. 5.– Pp. 609–622.   ⁶²
- [11] Mead N., Bower P. *Patient-centredness: a conceptual framework and review of the empirical literature* // Social Science & Medicine.– 2000.– Vol. **51**.– No. 7.– Pp. 1087–1110.   ⁶²
- [12] Грибова В. В., Шалфеева Е. А. *Обеспечение жизнеспособности систем, основанных на знаниях* // Информационные технологии.– 2019.– Т. **25**.– № 12.– С. 738–746.   ⁶²
- [13] Грибова В. В., Кульчин Ю. Н., Никитин А. И., Тимченко В. А. *Ансамбль онтологических моделей для обеспечения интеллектуальной поддержки лазерных аддитивных технологических процессов* // Онтология проектирования.– 2024.– Т. **14**.– № 2.– С. 279–300.     ^{63, 64}
- [14] Gribova V. V., Kleshchev A. S., Moskalenko F. M., Timchenko V. A. *A two-level model of information units with complex structure that correspond to the questioning metaphor* // Automatic Documentation and Mathematical Linguistics.– 2015.– Vol. **49**.– No. 5.– Pp. 172–181.   ⁶³
- [15] Gribova V. V., Kleshchev A. S., Moskalenko F. M., Timchenko V. A. *A model for generation of directed graphs of information by the directed graph of meta-information for a two-level model of information units with a complex structure* // Automatic Documentation and Mathematical Linguistics.– 2015.– Vol. **49**.– No. 6.– Pp. 221–231.   ⁶³
- [16] Иванова Г.Е., Бодрова Р.А., Буйлова Т.В., Каримова Г.М., Комарницкий В.С. *Алгоритм формулирования реабилитационного диагноза с помощью Международной классификации функционирования пациенту, перенесшему инсульт: клинический случай* // Физическая и реабилитационная медицина, медицинская реабилитация.– 2022.– Т. **4**.– № 1.– С. 37–54.   ⁶⁶

Поступила в редакцию	28.11.2024;
одобрена после рецензирования	20.01.2025;
принята к публикации	21.02.2025;
опубликована онлайн	03.03.2025.

Рекомендовал к публикации

к.т.н. С. А. Амелкин

Информация об авторах:



Валерия Викторовна Грибова

ИАПУ ДВО РАН, заместитель директора по научной работе, научный руководитель лаборатории интеллектуальных систем имени А.С. Клещева, д.т.н., чл.-корр. РАН. Научные интересы: онтологии и базы знаний, прикладные и проблемно-ориентированные системы, основанные на знаниях, управление базами знаний



0000-0001-9393-351X

e-mail: gribova@iacp.dvo.ru



Валентин Борисович Шуматов

ТГМУ, чл.-корр. РАН, профессор, доктор медицинских наук, ректор, заведующий кафедрой анестезиологии, реаниматологии, интенсивной терапии и скорой медицинской помощи



0000-0003-3836-390X

e-mail: mail@tgmu.ru



Сергей Васильевич Лебедев

ТГМУ, проректор по развитию, руководитель центра цифровых компетенций, кандидат медицинских наук. Научные интересы: методы управления медицинскими знаниями.



0000-0002-3378-4249

e-mail: lebedev@tgmu.ru



Елена Арефьевна Шалфеева

ИАПУ ДВО РАН, ведущий научный сотрудник лаборатории интеллектуальных систем имени А.С. Клещева, д.т.н. Научные интересы: онтологический инжиниринг, интерпретируемые клинические руководства, технология создания систем с декларативными знаниями, объяснительный искусственный интеллект, управление базами знаний.



0000-0001-5536-2875

e-mail: shalf@dvo.ru



Евгений Юрьевич Шестопалов

ТГМУ, директор института клинической неврологии и реабилитационной медицины, кандидат медицинских наук. Научные интересы: систематизация экспертных медицинских знаний.



0000-0002-7990-3602

e-mail: shestopalov77@mail.ru

**Дмитрий Борисович Окунь**

к.м.н., старший научный сотрудник лаборатории интеллектуальных систем имени А.С. Клещева ИАПУ ДВО РАН. Научные интересы: онтологический инжиниринг, информационные технологии, технологии обработки данных.

 0000-0002-6300-846X

e-mail: okdm@iacp.dvo.ru

**Роман Игоревич Ковалев**

ИАПУ ДВО РАН, научный сотрудник лаборатории интеллектуальных систем. Научные интересы: онтологии и базы знаний, интеллектуальные системы.

 0000-0002-1704-2675

e-mail: koval-995@mail.ru

**Екатерина Ивановна Шепета**

ТГМУ, институт клинической неврологии и реабилитационной медицины, ассистент. Научные интересы: разработка объективных методов оценки состояния пациента;

 0009-0002-0692-5664

e-mail: shepetae@mail.ru

**Леонид Александрович Федорищев**

ИАПУ ДВО РАН, к.т.н., старший научный сотрудник лаборатории интеллектуальных систем имени А.С. Клещева. Научные интересы: инструменты для разработки программного обеспечения, основанные на методах искусственного интеллекта; прикладные интеллектуальные системы на основе веб-технологий; пользовательский интерфейс; программные системы на мобильных устройствах.

 0000-0001-5536-2875

e-mail: fleo1987@mail.ru

**Александр Яковлевич Лифшиц**

ИАПУ ДВО РАН, к.т.н., ведущий инженер-программист лаборатории интеллектуальных систем имени А.С. Клещева. Научные интересы: обработка баз знаний, логический вывод основанный на знаниях.

 0009-0000-5671-2881

e-mail: mmvb@iacp.dvo.ru

Авторы внесли равный вклад в подготовку публикации.

Декларация об отсутствии личной заинтересованности: благополучие авторов не зависит от результатов исследования.



The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke

Valeria Viktorovna **Gribova**¹, Valentin Borisovich **Shumatov**², Sergey Vasilyevich **Lebedev**³, Elena Arefevna **Shalfeeva**⁴, Evgeny Yurievich **Shestopalov**⁵, Dmitri Borisovich **Okun**^{6,8}, Roman Igorevich **Kovalev**⁷, Ekaterina Ivanovna **Shepeta**⁸, Leonid Alexandrovich **Fedorishchev**⁹, Alexander Yakovlevich **Lifshits**¹⁰

^{1,4,6,7,9,10} Institute of Automation and Control Processes of the Far Eastern Branch of the RAS, Vladivostok, Russia
^{2,3,5,8} Federal State Budgetary Educational Institution of Higher Education "Pacific State Medical University" of the Ministry of Healthcare of the Russian Federation, Vladivostok, Russia

Abstract. The main incentive for the introduction of computer technologies into the healthcare system is the desire to significantly improve the quality of life of people. This includes improving the quality and speed of treatment, reducing the cost of medical services and acquiring effective means to comply with regulatory requirements.

At the present stage of rehabilitation development, the need for active implementation of medical decision support systems and artificial intelligence technologies becomes obvious. These technologies can significantly improve the understanding of the clinical aspects of disorders, the level of activity and participation of stroke patients in the rehabilitation process. A key component of the successful application of these systems is the importance of formalizing knowledge and creating ontologies that provide a structured and connected presentation of medical information and define the rules for their interpretation.

This paper presents a set of interrelated ontological models underlying the intellectual decision support system being developed in the rehabilitation of stroke patients. The IACPaaS cloud platform is used to implement the complex of ontologies. Ontologies and the target resources generated on their basis are the basic elements of the system being developed, which will soon be provided to healthcare professionals to solve urgent rehabilitation issues. Mechanisms are provided for the planned expansion and refinement of the knowledge base, which will allow the system to easily adapt to new medical research results and optimize its work as a whole.

(In Russian).

Key words and phrases: medical decision support system, intelligent service, rehabilitation, stroke, knowledge base, ontological approach, knowledge engineering

2020 *Mathematics Subject Classification:* 68P05; 92C50, 68T30

Acknowledgments: The project is being implemented by a team of employees of the Federal State Budgetary Educational Institution of Higher Education Tver State Medical University of the Ministry of Health of the Russian Federation and the Federal State Budgetary Scientific Institution IAPU FEB RAS within the framework of the subsidy for the implementation of the strategic academic leadership program "Priority-2030" and with the support of the Ministry of Education and Science of Russia No. FFWF-2021-0004

For citation: Valeria V. Gribova, Valentin B. Shumatov, Sergey V. Lebedev, Elena A. Shalfeeva, Evgeny Yu. Shestopalov, Dmitri B. Okun, Roman I. Kovalev, Ekaterina I. Shepeta, Leonid A. Fedorishchev, Alexander Ya. Lifshits. *The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke.* Program Systems: Theory and Applications, 2025, **16**:1(64), pp. 61–82. (In Russ.). https://psta.psiras.ru/read/psta2025_1_61-82.pdf

References

- [1] S. V. Mal'ceva. "Application of ontological models for solving problems of identification and monitoring of subject areas", *Biznes-informatika*, 2008, no. 3, pp. 18–24 (in Russian). 
- [2] V. D. Solov'ev, B. V. Dobrov, V. V. Ivanov, N. V. Lukashevich. *Ontologies and Thesauri: models, tools and applications*, Intuit NOU, M., 2016, ISBN 978-5-9963-0007-5, 207 pp. (in Russian).
- [3] P. A. Lomov. "Application of ontology design patterns to development and use of ontologies in an integrated knowledge space", *Ontologiya proektirovaniya*, **5**:2 (2015), pp. 233–245 (in Russian). 
- [4] A. Yu. Pleshkova. "Ontologies in educational process management", *Ontologiya proektirovaniya*, **12**:4 (2022), pp. 506–517 (in Russian).  
- [5] S. V. Lebedev, N. A. Zhukova. "Ontology-driven approach to medical data fusion", *Ontologiya proektirovaniya*, **7**:2 (2017), pp. 145–159 (in Russian).  
- [6] Yu. I. Volokitin, V. P. Kupriyanovskij, O. V. Grin'ko, O. N. Pokusaev, S. A. Sinyagov. "On problems of the digital economy and formalized ontologies", *International Journal of Open Information Technologies*, **6**:6 (2018), pp. 87–96 (in Russian).
- [7] S. González, B. G. M. E. Blobel, D. M. Lopez. "Ontology-based framework for electronic health records interoperability", *User Centred Networked Health Care*, Proceedings of MIE 2011, Studies in Health Technology and Informatics, vol. **169**, eds. A. Moen et al., IOS Press, 2011, ISBN 978-1-60750-805-2, pp. 694–698. 
- [8] *Order of the Ministry of Health of Russia dated July 31, 2020 No. 788n "On approval of the Procedure for organizing medical rehabilitation of adults"*, 2021 (in Russian). 
- [9] A. A. Belkin, A. M. Alasheev, V. A. Belkin, Yu. B. Belkina, A. N. Belova, D. V. Bel'skij, R. A. Bodrova, G. P. Bryusov, N. A. Varako, I. A. Voznyuk, V. I. Gorbachev, A. V. Grechko, L. Sh. Gumarova, V. D. Daminov, V. I. Ershov, R. A. Zhiguzhevskij, I. B. Zabolotskix, O. S. Zajcev, V. O. Zaxarov, Yu. P. Zinchenko, G. E. Ivanova, S. N. Kaurkin, P. A. Knyazenko, M. S. Kovyazina, A. N. Kondrat'ev, E. A. Kondrat'eva, E. V. Kostenko, A. A. Kulakova, V. V. Krylov, V. L. Kuprejchik, O. A. Lajsheva, K. M. Lebedinskij, A. L. Levit, I. N. Lejderman, A. Yu. Lubnin, E. A. Lukoyanova, M. N. Mal'ceva, M. R. Makarova, E. V. Mel'nikova, N. N. Mitrakov, I. E. Mishina, S. A. Pomeschkina, S. S. Petrikov, M. V. Petrova, E. A. Pinchuk, M. A. Piradov, K. A. Popugaev, A. V. Polyakova, I. V. Pogonchenkova, D. N. Procenko, N. V. Raseva, E. N. Rudnik, V. A. Rudnov, Yu. V. Ryabinkina, I. A. Savin, T. Yu. Safonova, Zh. B. Semenova, E. Yu. Skripaj, A. A. Solodov, A. Yu. Suvorov, N. A. Suponeva, D. V. Tokareva, S. P. Urazov, D. Yu. Usachev, A. S. Fadeeva, D. R. Xasanova, V. T. Xasanshin, S. V. Carenko, S. E. Chuprina, E. Yu. Shestopalov, A. V. Shhegolev, A. M. Shhikota, R. E. Engaus, S. N. Norvils. "Rehabilitation in the intensive care unit (RehabICU). Clinical

- practice recommendations of the national Union of Physical and Rehabilitation Medicine Specialists of Russia and of the national Federation of Anesthesiologists and Reanimatologists”, *Vestnik intensivnoj terapii imeni A.I. Saltanova*, 2022, no. 2, pp. 7–40 (in Russian). 
- [10] M. O’Keeffe, P. Cullinane, J. Hurley, I. Leahy, S. Bunzli, O’Sullivan P. B. , O’Sullivan K. . “What influences patient-therapist interactions in musculoskeletal physical therapy? Qualitative systematic review and meta-synthesis”, *Physical Therapy*, **96**:5 (2016), pp. 609–622. 
- [11] N. Mead, P. Bower. “Patient-centredness: a conceptual framework and review of the empirical literature”, *Social Science & Medicine*, **51**:7 (2000), pp. 1087–1110. 
- [12] V. V. Gribova, E. A. Shalfeeva. “Ensuring of viability of systems based on knowledge”, *Informacionnye tekhnologii*, **25**:12 (2019), pp. 738–746 (in Russian). 
- [13] V. V. Gribova, Yu. N. Kul’chin, A. I. Nikitin, V. A. Timchenko. “An ensemble of ontological models for intelligent support of laser additive manufacturing processes”, *Ontologiya proektirovaniya*, **14**:2 (2024), pp. 279–300 (in Russian).  
- [14] V. V. Gribova, A. S. Kleshchev, F. M. Moskalenko, V. A. Timchenko. “A two-level model of information units with complex structure that correspond to the questioning metaphor”, *Automatic Documentation and Mathematical Linguistics*, **49**:5 (2015), pp. 172–181. 
- [15] V. V. Gribova, A. S. Kleshchev, F. M. Moskalenko, V. A. Timchenko. “A model for generation of directed graphs of information by the directed graph of metainformation for a two-level model of information units with a complex structure”, *Automatic Documentation and Mathematical Linguistics*, **49**:6 (2015), pp. 221–231. 
- [16] G.E. Ivanova, R.A. Bodrova, T.V. Bujlova, G.M. Karimova, V.S. Komarnickij. “Algorithm for formulation a rehabilitation diagnosis using the international classification of functioning in a patient with a stroke: clinical case”, *Fizicheskaya i reabilitacionnaya medicina, medicinskaya reabilitaciya*, **4**:1 (2022), pp. 37–54 (in Russian). 



Multimodal Stock Price Prediction: A Case Study of the Russian Securities Market

Kasymkhan Usufovich **Khubiye**¹, Mikhail Evgenyevich **Semenov**²

^{1,2}Sirius University of Science and Technology, Sirius, Russia

Abstract. Classical asset price forecasting methods primarily rely on numerical data, such as price time series, trading volumes, limit order book data, and technical analysis indicators. However, the news flow plays a significant role in price formation, making the development of multimodal approaches that combine textual and numerical data for improved prediction accuracy highly relevant.

This paper addresses the problem of forecasting financial asset prices using the multimodal approach that combines candlestick time series and textual news flow data. A unique dataset was collected for the study, which includes time series for 176 Russian stocks traded on the Moscow Exchange and 79,555 financial news articles in Russian.

For processing textual data, pre-trained models RuBERT and Vikhr-Qwen2.5-0.5b-Instruct (a large language model) were used, while time series and vectorized text data were processed using an LSTM recurrent neural network. The experiments compared models based on a single modality (time series only) and two modalities, as well as various methods for aggregating text vector representations.

Prediction quality was estimated using two key metrics: Accuracy (direction of price movement prediction: up or down) and Mean Absolute Percentage Error (MAPE), which measures the deviation of the predicted price from the true price. The experiments showed that incorporating textual modality reduced the MAPE value by 55%.

The resulting multimodal dataset holds value for the further adaptation of language models in the financial sector. Future research directions include optimizing textual modality parameters, such as the time window, sentiment, and chronological order of news messages. (*Linked article texts in English and in Russian*).

Key words and phrases: Multimodal Forecasting, Quantitative Finance, Machine Learning

2020 *Mathematics Subject Classification:* 68T30; 68T50, 91884

Acknowledgments:

For citation: Kasymkhan U. Khubiye, Mikhail E. Semenov. *Multimodal Stock Price Prediction: A Case Study of the Russian Securities Market*. Program Systems: Theory and Applications, 2025, **16**:1(64), pp. 83–130. (*In English, in Russian*). https://psta.psiras.ru/read/psta2025_1_83-130.pdf

Introduction

Building a price forecast for an asset is a crucial task for financial market participants, as it enables strategic planning, optimal investment portfolio management, and risk assessment. Numerous attempts have been made to apply machine learning methods to construct such forecasts [1–3].

With the growing popularity of deep learning models, researchers have shifted their focus toward the application of neural networks. At the same time, the problem of accurately accounting for the news flow as a key factor influencing market behavior is being reconsidered with the rapid development of generative artificial intelligence models and large language models (LLMs) such as ChatGPT, FinGPT, GigaChat, LLama, and others. In financial economics, LLMs are still rarely used, and their full potential remains untapped.

Researchers are exploring the use of natural language processing models to enhance the accuracy of asset price forecasts and investment portfolio management strategies.

The study [4] describes the use of sentiment analysis of news as an additional parameter. The authors employed the FinBert model, trained on financial data, to assess the sentiment of news articles as positive, negative, or neutral. The study utilized time series data from candlestick charts of the U.S. stock market index, Standard & Poor's 500 (S&P 500). A machine learning model — random forest — was used for price prediction. The study concluded that incorporating sentiment analysis of news flow improves prediction accuracy.

In the study [5], the authors aimed to develop a multimodal artificial intelligence model capable of providing well-founded and accurate forecasts for time series data. They implemented a model that generates predictions of an asset's monthly or weekly returns, accompanied by a textual explanation from a language model based on the user's input query.

The study [6] proposed an approach for fine-tuning instructions to interpret numerical values and contextualize financial data.

Kulikova et al. [7] examined the effect of classifying news into thematic groups. The authors demonstrated that, in most cases, it is advisable to use a single thematic group of news for the deep learning models considered (Temporal Convolutional Network, D-Linear, Transformer, and Temporal Fusion Transformer). They also determined the probabilities of forecast improvement for the 20 thematic groups analyzed.

In all the aforementioned studies, the models were implemented using a multimodal approach for the U.S. stock market, with English as the modality language. Notably, the news flow was not integrated directly into the predictor’s input vector but rather through a preprocessing block in the form of an additional parameter, such as sentiment analysis, news frequency related to the asset, or news classification, etc.

The objective of the current study is to demonstrate the advantages of a new multimodal method over predictions based solely on numerical data and to present a Russian-language financial news dataset.

To achieve this objective, we formulated the following key tasks:

- (1) Construct a multimodal dataset consisting of time series data and news articles.
- (2) Develop a predictive model capable of utilizing one or two modalities.
- (3) Train the predictive model and analyze the values of accuracy functions and metrics, specifically Accuracy and MAPE.

In this study, we propose a new multimodal approach for integrating news flow into time series numerical data. The text of the news articles is converted into a vector representation and fed into the model alongside the time series vector.

Our hypothesis is that the multimodal approach will enable predictive models to extract semantic information from the text, thereby improving the accuracy of asset price forecasts.

1. Data Collection and Structuring

Multimodality implies the use of more than one data modality, which affects both the data structure and the logic of predictive model development. We utilize two types of modalities:

numerical — time series of stock prices,
textual — news streams.

To train the predictive model and analyze its performance, we collected an original dataset.

The time series, represented as candlestick data with open, close, high, and low prices, were obtained through the Algotrack API of the Moscow Exchange (MOEX). For the numerical experiment, we selected stock time series data spanning from July 7, 2022, to August 30, 2024, covering 176

TABLE 1. Statistical features of the dataset after tokenization, RuBert

Source	Mean	Std	Min	Max	Q25	Q50	Q75
RDV	134	88	8	512	65	123	187
Finam	221	135	18	512	116	178	284
BCS Express	20	10	4	82	13	17	26
BCS Technical Analysis	502	37	29	512	512	512	512
RBC	43	7	16	75	39	44	48
SmartLab	21	8	5	82	15	19	25

companies. During this period, the Russian stock market experienced phases of rapid growth and decline, with the IMOEX index rising from 2,213.81 to 2,650.32 points (+19,72%).

We collected 79,555 news articles from various sources, including the online publication “RBC” (1,823 articles), “BCS Express” (11,331), and “BCS Technical Analysis” (9,670), the investment company website “Finam” (20,647), the trader community website «SmartLab.ru» (30,857), as well as the Telegram channel “RDV” (5,227).

Several factors justify the selection of these sources. First, they provide news coverage for the required time period. Second, the institutional differences between sources, along with variations in writing style and levels of expertise, contribute to a more objective representation of events related to the analyzed time series.

News messages were tokenized using two models: RuBERT [8] and Vikhr-Qwen2.5-0.5b-Instruct [9] (further as Qwen). In the context of tokenized text, a word refers to a token—an element of the vector space represented as an index in the tokenizer’s vocabulary.

Descriptive statistics of the dataset (in tokens), including mean, standard deviation, minimum, maximum word count, and quartiles, are presented in Tables 1 and 2. It is important to note that tokenization can increase the word count in a text, for example, by splitting words into smaller components.

Table 3 provides examples of how a phrase changes after tokenization. For instance, the word «открывает» is split into three subcomponents: «от», «##К», and «##рывает», where the “##” prefix indicates that the token is a continuation of the previous token.

TABLE 2. Statistical features of the dataset after tokenization, Qwen

Source	Mean	Std	Min	Max	Q25	Q50	Q75
RDV	215	157	3	1324	92	187	304
Finam	453	405	35	5732	211	319	501
BCS Express	36	19	5	163	23	32	47
BCS Technical Analysis	1493	310	40	2221	1448	1545	1665
RBC	75	12	28	105	68	77	83
SmartLab	33	12	7	120	25	31	39

TABLE 3. Original and tokenized texts examples

Original text	Tokenized text
Доллар снова ниже 69 рублей	До ##лла ##р снова ниже 69 рублей
Москвич банкрот?	Москви ##ч банк ##рот ?
НПО Наука Отчет РСБУ	Н ##П, ##О Наука От ##чет Р ##С ##Б ##У
Т-банк это желтый банк	Т - банк это же ##лт ##ый банк

News articles characteristics On the “BCS Technical Analysis” platform, news articles tend to be lengthy, which imposes limitations on tokenizers. Specifically, as shown in Table 1 and Table 2, the RuBERT model truncates the tokenized vector for longer texts. Additionally, the average length of tokenized text using the Qwen model exceeds that of RuBERT, indicating that Qwen has a broader vocabulary and a stronger text decomposition capability.

Furthermore, we collected data on 176 companies, forming a dataset consisting of tuples in the format:

(ticker, company name, company activity description).

Such data are essential in our case for:

- extracting keywords from company descriptions,
- improving the language model’s ability to link events described in news articles to specific companies and assess the impact of news on price dynamics.

TABLE 4. Examples of news articles (header snippet) and assigned tags

Source	Article fragment (heading)	Tags
RDV	Сережа (SGZH): таргет 16.2 руб., апсайд +102...	SGZH
RDV	Артген биотех (ABIO) завершил доклинические...	аналитика, ABIO
Finam	Индекс МосБиржи восстанавливает позиции и приб...	ФосАгро, ВСМПО-АВСМ, CNYRUB
Finam	«Ашинский метзавод» назвал АО "Урал-БК" своим ...	АшинскийМЗ
BCS Express	«Восходящее окно»: в каких бумагах замечен это...	Селигдар SELG, ЕвроТранс EUTR
BCS Express	«Сила Сибири» выйдет на максимальную мощность...	Газпром GAZP
BCS Technical Analysis	Мечел. Что ждать от бумаг на следующей неделе	Мечел
BCS Technical Analysis	На предыдущей торговой сессии акции Норникеля ...	ГМК Норникель

The dataset of news articles includes the following parameters: publication date, source, title, article body, and tags (keywords). For sources such as “RDV” and “SmartLab”, article titles are absent, and the corresponding fields are filled with a label: *no title*.

In our case, tags may include the full or abbreviated company name along with the corresponding ticker, the name of the market sector, and similar information. Tags in news articles were assigned by the article authors.

For the “RDV” source, tags were marked by authors in the form of hashtags (e. g. #цифры, #аналитика). In “BCS Express” and “BCS Technical Analysis”, tags were specified in dedicated fields at the beginning or end of the news article (e. g. PhoseAgro, Russian market) and were extracted from the *HTML* code of the page using the corresponding *HTML* tags. When tags were absent (“RBC”, “SmartLab”), the parameter in the dataset remained empty.

Table 4 provides examples of news articles (headline fragments) along with their assigned tags.

2. Methods

To validate our hypothesis regarding the advantages of the multimodal approach, we have planned a series of experiments.

The first series of experiments focused on predicting prices using only numerical time series of candlestick characteristics (close, open, high, and low prices). The quality metrics obtained from this experiment serve as baseline values against which improvements in price prediction accuracy using the proposed multimodal approach will be evaluated.

The second series of experiments aims to generate predictions and compute accuracy metrics (Accuracy, MAPE) using the multimodal approach while exploring different aggregation methods (Sum, Mean) for the vectorized news stream.

2.1. The Single-Modality Approach

We first conducted a series of experiments on asset price prediction using only time series data. For this, we applied classical machine learning models to the daily price values (close, open, high, low), including linear regression (LinReg), k -nearest neighbors (KNN), decision tree (DT), random forest (RF), and the boosting algorithm XGBoost (XGB). Among deep learning models, we utilized a long short-term memory recurrent neural network (LSTM).

Conceptually, the experiment consists of two tasks:

- (a) predicting the price movement direction (increase or decrease), which is a binary classification task;
- (b) predicting the actual price, which is a regression task.

At this stage of the experiment, 176 companies were grouped into 23 industry sectors. We randomly selected 9 economic sectors and, within each sector, randomly chose two companies. Table 5 lists the selected sectors and companies (tickers) that participated in the computational experiment.

Table 7 provides statistical data on the closing price time series of the selected assets. Table 6 shows the distribution of news by companies after filtering. The correlation heat map of the closing price time series is shown in Figure 1. An interesting feature of the examined period is that the market underwent two phase shifts — from a general price decline to growth and back again — as indicated by the vertical lines in Figure 2.

TABLE 5. Economic sectors and companies (tickers) included into the dataset

Sector	Company (ticker)
Metal and Mining	Mechel (MLTR), TMK-Group (TRMK)
Oil and Gas	Surgutneftegas (SNGS), Gazpromneft (SIBN)
Consumer sector	Magnit (MGNT), Lenta (LENT)
Construction	PIK (PIKK), Samolet (SMLT)
Telecommunications	MTS (MTSS), Rostelecom (RTKMP)
Transport	AEROFLOT (AFLT), Sovcomflot (FLOT)
Finance	Bank Saint-Petersburg (BSPB), SFI (SFIN)
Chemical Industry	Phosagro (PHOR), Kazanorgsintez (KZOSP)
Power Engineering	Rushydro (HYDR), Rosseti Center (MRKC)

TABLE 6. Descriptive characteristics for company shares

Company (ticker)	Number of news items
Mechel (MLTR)	4258
Trubnaya Metallurgical Company (TRMK)	11739
Surgutneftegaz (SNGS)	12674
Gazpromneft (SIBN)	11421
Magnit (MGNT)	1236
Lenta (LENT)	311
PIK (PIKK)	897
Samolet (SMLT)	3392
MTS (MTSS)	1101
Rostelecom (RTKMP)	628
Aeroflot (AFLT)	1429
Sovcomflot (FLOT)	14476
Saint Petersburg Exchange (BSPB)	14278
SFAI (SFIN)	1647
PhosAgro (PHOR)	2773
Kazanorgsintez (KZOSP)	168
RusHydro (HYDR)	1921
MRSK Center (MRKC)	1576

TABLE 7. Descriptive characteristics for company shares

Ticker	Mean	Std	Min	Max	Q25	Q50	Q75
MTLR	191.8245	72.5652	81.2800	332.8800	123.8500	187.6700	251.6400
TRMK	153.1245	64.9362	55.8200	271.0000	87.1400	166.4200	218.7800
SNGS	27.0104	4.0119	17.3500	36.9600	23.7750	27.3300	30.0250
SIBN	601.5097	163.9205	335.5500	934.2500	452.0500	582.6500	748.9000
MGNT	5691.6429	1161.7684	4040.0000	8444.0000	4665.0000	5495.0000	6375.0000
LENT	814.3870	154.9502	650.0000	1263.0000	716.5000	749.0000	843.5000
PIKK	732.6617	94.8650	518.0000	955.5000	656.7000	732.9000	811.5000
SMLT	3120.8996	594.1018	1926.5000	4145.5000	2572.0000	3045.0000	3713.0000
MTSS	264.5382	32.0791	183.0000	346.9500	239.0000	266.2500	289.7500
RTKMP	68.1797	9.2753	52.2500	92.1000	60.4500	68.0000	74.7000
AFLT	38.1316	10.3131	22.4400	64.4000	27.9700	38.8800	44.1200
FLOT	88.0111	39.5834	29.9200	149.3000	42.1000	97.2000	124.1800
BSPB	211.1501	101.2533	67.5700	387.6800	100.8400	210.9900	295.3400
SFIN	762.9939	428.5679	425.8000	1975.0000	497.4000	518.0000	992.0000
PHOR	6774.6040	618.1977	4997.0000	8153.0000	6416.0000	6763.0000	7278.0000
KZOSP	25.8603	5.2029	15.3500	40.5700	21.9400	27.0700	29.8500
HYDR	0.7697	0.0810	0.5178	1.0278	0.7318	0.7721	0.8210
MRKS	0.5247	0.2382	0.2025	1.0745	0.2735	0.5550	0.7475

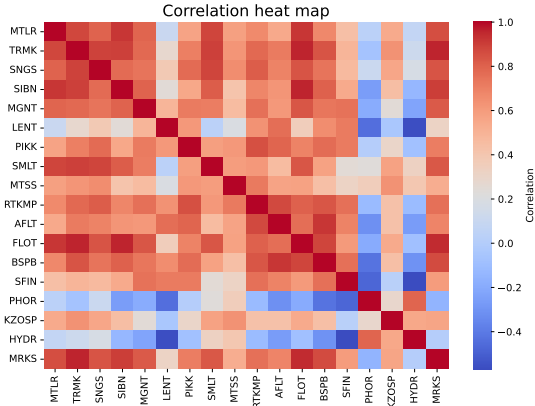


FIGURE 1. The correlations heatmap for 18 assets (close price)

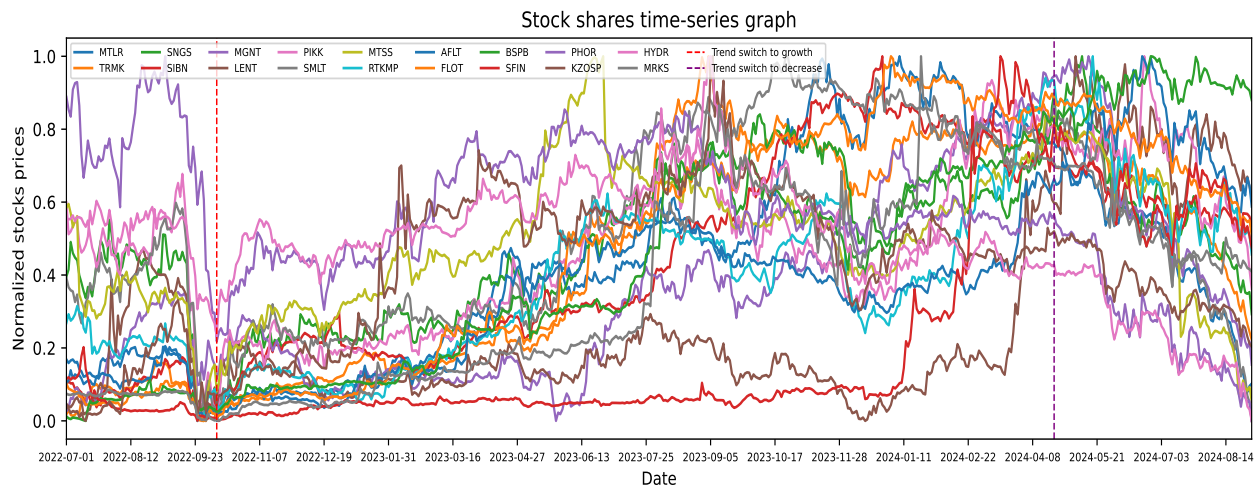


FIGURE 2. Normalized close prices of assets. Market phase transition dates denoted by vertical dashed lines

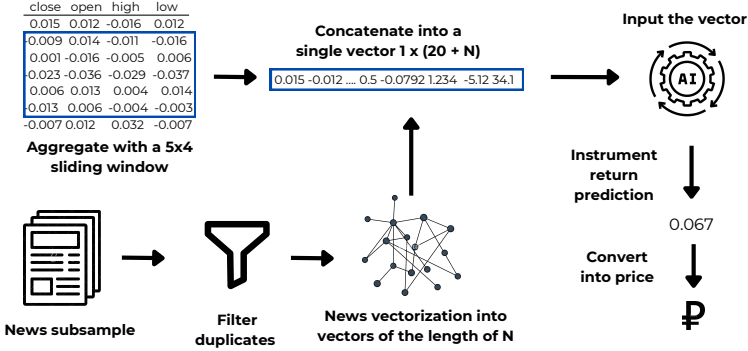


FIGURE 3. Pipeline for a single and dual modalities models

To evaluate prediction quality in the classification task, we used the *Accuracy* metric, while for regression, we employed *MAPE* (Mean Absolute Percentage Error). The choice of these metrics is justified by the nature of the tasks. In classification, the model must accurately predict the price movement direction either an increase (denoted by «+») or a decrease (denoted by «-»). The *MAPE* metric is best suited for assessing regression quality within the financial domain: it represents the average deviation from the asset's actual price in percentage terms, making it easily interpretable in monetary value.

Figure 3 illustrates the model development process for utilizing one and two modalities.

As the input parameter, the model received a return vector of the asset, calculated based on the closing price (close) over the previous five trading sessions:

$$(1) \quad \text{Return}(d+1) = \frac{\text{close}(d+1)}{\text{close}(d)} - 1.$$

The model's output was a prediction for the next trading session.

To assess the accuracy of predicting the price movement direction, the predicted class was determined by the sign ($\pm$) of the forecasted return

value, as the return of an asset represents the relative rate of change. Thus, a positive return indicates a price increase, while a negative return signifies a decline. To evaluate the quality of the asset price forecast, the predicted return vector was converted into price (in Russian rubles):

$$(2) \quad price(d+1) = (Return(d+1) + 1) \cdot price(d).$$

The pointwise predicted price vector, obtained through transformation, was compared to the historical price vector of assets using the *MAPE* metric.

The choice of return (rather than price) as the target variable for the predictive model is justified by the fact that when prices exceed historical highs (or fall below historical lows) during market growth (or decline), the applicability of traditional methods becomes limited.

Based on this reasoning, candlestick characteristics (close, open, high, and low prices) were considered in the form of *relative price changes*, calculated using a formula similar to (1).

Next, a rolling window of five trading days was applied to the relative price changes to form a vector-row, which was then fed into the predictive model. As a result, the model receives a vector of 20 parameters as input and predicts a single output value — the return of the instrument at the end of the next trading session.

2.2. The Dual-Modality Approach

For the experiment involving news flow, we selected news articles relevant to the analyzed assets based on keyword matching (Table 5). The keywords were chosen as the top 30 words extracted using the TF-IDF method. This method determines the importance of words in a text by considering their frequency of occurrence and uniqueness across the entire corpus. An example of keywords extracted using TF-IDF is presented in Table 8.

After obtaining the list of keywords using the TF-IDF method, we further expanded it with the help of the ChatGPT-4o model. This allowed us to increase keyword variability through permutations, letter substitutions, and modifications of word endings (Table 9). The selected news articles for each company (ticker) were converted into vectors and filtered to remove duplicates.

TABLE 8. Keywords by companies extracted from their descriptions

Ticker	Keywords
MTLR	mechel, mining, ore, raw materials, energy, ferroalloys, coal
SNGS	gas, geological exploration, oil, Surgutneftegas, petroleum products, electricity, drilling
SMLT	rent, development, developer, real estate, construction, Moscow region, residential areas
MTSS	subscriber, automation, internet, mobile communications, provider, communications
BSPB	bank, deposit, dividends, financial services, kaliningrad, sbank, saint-petersburg

TABLE 9. Complementary generated keywords

Ticker	Keywords
MTLR	мечел, метчел, мечал, mechel, Mchel, ферросплавы, фурросплав
SNGS	сургутнефтегаз, surgutneftegaz, surgut, сурнефтегаз, суграз, сургут, сур-нфтгз
SMLT	самолет, smlt, samolet, samalet, Самлет
RTKMP	ростелеком, телеком, rostelecom, telecom, rtkm, ртк, r-telecom, растелком
HYDR	русгидро, rushydro, rshydro, r-gidro, гидпорус, гидра, русгидра

Figure 4 presents a distribution chart of the news articles for the companies after filtration.

As a vectorizer for the Russian language news stream, we employed two models: RuBERT [8] and Qwen [9].

While working with the news stream, we encountered two main challenges. The first challenge is the problem of news rewriting, which necessitates filtering out duplicate articles. To ensure that our model accounts for each news article only once, it is essential to implement a duplicate identification algorithm.

The second challenge is to determinate an asset on which is affected the news article. This problem can be framed as a classification task, where

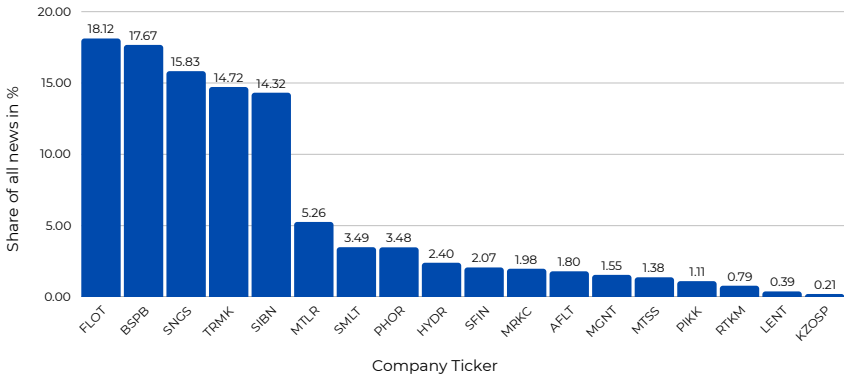


FIGURE 4. The distribution of news articles by company after filtration (Numbers on the diagram show percentage of news about the company in the dataset)

tickers serve as class labels.

To address the issue of news rewriting, we designed a Siamese neural network. We constructed a training dataset using the GigaChat API as follows: for each article, three paraphrased versions of both the title and body were generated. Then, pairs were randomly formed in equal proportion from the original and paraphrased news articles and their titles.

The Siamese neural network was designed as follows: a pair of news articles is fed as input, and vector representations of the articles are extracted using the RuBERT model [8]. The two vectors are then concatenated, and the resulting vector is passed through a fully connected neural network (MLP). To determine the optimal depth of the MLP model, we conducted a series of experiments, evaluating both prediction accuracy and news stream processing time. Based on the results, we selected the MLP architecture with three layers.

The filtered news articles are then converted into vectors so that duplicate classification can be performed in a one-shot mode when new articles arrive. This approach reduces both the processing time of the news stream and the computational resources required (in our case, a GPU V100).

To address the second challenge—matching news article samples by date and utilizing them for price forecasting—it is essential to formalize the data selection and prediction process. We assume that the closing price prediction for an asset is made for each trading day at the market opening. In this case, only news articles published before the start of the current trading day are included in the dataset.

The dataset is formed by grouping news articles based on their publication date. For predicting the price on a given day, only articles published on the previous trading day are used. For example, analytical articles such as those under the “Technical Analysis” section from the “BCS” source, which are published daily before the market opens, are included in the dataset for forecasting the prices of assets analyzed in those reports. This approach ensures that the most relevant information is considered, thereby improving prediction accuracy.

For the dual-modality approach, training sequences were formed by concatenating price return vectors from the previous five days with news stream vectors. The relative price return vectors were constructed similarly to the single-modality experiment, while news articles were selected from the previous trading day based on the chosen asset. These news articles were then transformed into vectors and aggregated.

If no publications were available on the previous day or before the market opened on the current day, a zero vector was concatenated with the relative price return vector of length 768 for the RuBERT model and 896 for the Vikhr-Qwen2.5-0.5b-Instruct (Qwen) model. Otherwise, the aggregated news vector of the same length was appended. These final vector lengths correspond to the output sizes of the pretrained RuBERT and Qwen models.

In this study, we explored two approaches for aggregating news vectors: vector summation (Sum) and averaged summation (Mean). By vector summation, we mean summing the values of corresponding vector coordinates. In the averaged summation approach, each coordinate of the aggregated vector is assigned the arithmetic mean of the corresponding coordinates across all aggregated vectors.

The baseline RuBERT model has a limited context window of 512 tokens. As a result, articles exceeding this limit were either truncated or

split for separate processing, meaning that a single news article could correspond to multiple vectors. In contrast, the Qwen model has a significantly larger context window of 32,768 tokens (64 times larger), allowing it to process entire articles without truncation. Next, we compare how different news vectorization methods impact the accuracy of price predictions.

The pointwise predicted return vectors were converted into asset prices using equation (2). The prediction quality was evaluated using two metrics: Accuracy and Mean Absolute Percentage Error (MAPE). Accuracy was measured as the proportion of correctly predicted signs of the return vector elements—either positive or negative. The MAPE metric indicates the average percentage deviation of the predicted price from the actual value. This allows us to assess the prediction quality not only in relative terms but also in absolute monetary units (rubles).

3. Computational experiment

In this section, we present the results of computational experiments for two predictive models (single- and dual-modalities). The predictive model was developed using the Transformers framework from the Hugging Face platform. All computations were performed on an NVIDIA V100 GPU.

3.1. The Single-Modality Approach Performance

The results of the experiment on predicting return vectors using only time series data for classical and deep learning models are presented in a Table 10.

Table 11 provides the averaged prediction quality metrics for all models, sorted in ascending order of the mean absolute percentage error (MAPE) (column “Deviation”).

From the experiment results, it is evident that the recurrent model LSTM achieves the best classification performance (predicting upward or downward trends) and regression accuracy (smallest deviation of the predicted price from the actual price). However, it lags slightly in terms of the mean absolute error metric.

TABLE 10. Results of forecasting return vectors using only time series. Accuracy (left) and deviation (right) in percent

	Source	LSTM		XGB		KNN		RF		LinReg		DT	
Metals and Mining	MTLR	56.364	0.410	40.000	2.089	42.273	2.050	50.909	2.020	50.000	2.029	42.727	2.679
	TRMK	56.364	0.362	40.909	2.105	38.182	2.167	47.273	2.154	49.091	2.114	52.727	2.308
Oil and Gas	SNGS	50.303	0.352	49.091	1.776	48.182	1.775	50.000	1.735	60.909	1.744	52.727	1.857
	SIBN	58.182	0.341	40.000	1.766	58.182	1.746	46.364	1.788	41.818	1.839	51.818	1.813
Consumer Sector	MGNT	46.667	0.331	39.091	1.517	43.636	1.493	49.091	1.519	40.000	1.709	60.000	1.672
	LENT	56.364	0.371	54.546	2.202	39.091	2.178	52.723	2.145	51.818	2.220	51.818	2.589
Construction	PIKK	49.091	0.484	40.909	1.565	50.909	1.563	50.000	1.558	44.545	1.637	51.818	1.592
	SMLT	53.939	0.328	42.727	1.577	38.182	1.552	46.364	1.539	49.091	1.536	41.818	1.683
Telecommunications	MTSS	56.970	0.541	42.727	1.290	40.000	1.306	45.455	1.520	53.636	1.419	50.000	1.395
	RTKMP	55.152	0.246	45.455	1.299	42.723	1.303	42.727	1.335	50.909	1.355	48.182	1.411
Transport	AFLT	55.152	0.419	46.364	2.079	57.273	2.017	52.727	2.062	60.909	1.976	51.818	2.194
	FLOT	47.273	0.258	43.637	2.116	38.182	2.124	42.727	2.104	45.454	2.074	49.091	2.294
Finance	BSPB	46.061	0.410	49.091	1.612	50.909	1.695	50.909	1.598	54.545	1.602	45.455	1.829
	SFIN	49.697	0.447	40.000	1.603	30.909	1.647	39.091	1.743	48.182	1.960	41.818	1.959
Chemical Industry	PHOR	41.818	0.231	42.727	1.194	52.723	1.149	48.182	1.168	50.000	1.227	45.455	1.218
	KZOSP	57.576	0.458	49.091	1.198	42.723	1.237	49.091	1.210	46.364	1.217	54.545	1.581
Power Engineering	HYDR	59.394	0.380	51.182	1.124	60.000	1.130	48.182	1.214	45.455	1.151	49.091	1.355
	MRKC	40.000	0.768	51.182	1.182	49.091	1.225	54.545	1.214	50.000	1.224	55.455	1.403

TABLE 11. The Single-Modality approach forecast (time-series) inference metrics: Accuracy and MAPE in percentage

Model	Accuracy, %	MAPE, %
LSTM	52.020	0.397
XGB	45.000	1.627
KNN	46.010	1.631
RF	48.384	1.646
LinReg	50.152	1.669
DT	49.798	1.824

3.2. The Dual-Modality Approach Performance

The results of the second experiment, which involved merging the news stream with numerical time series data and comparing the proposed multimodal approach with a forecast based solely on candlestick time series, are presented in the Table 12.

The Table 13 provides the averaged prediction quality metrics for the considered models. The data in this table is sorted by the “Deviation” column in ascending order, reflecting the mean absolute percentage error (MAPE) of the predicted price deviations.

In this second experiment, the LSTM neural network was chosen as the baseline model. We compared different vectorization methods (RuBert, Qwen) and aggregation techniques (Sum, Mean) to evaluate their impact on prediction performance.

Figure 5 shows the dependence of the mean squared error (MSE Loss) function values on the number of training iterations for different models, based on the training set (from July 7, 2022, to March 27, 2024) and the test set (from March 28 to August 30, 2024). The graph indicates that after 30 training epochs, the curves reach a stationary value.

TABLE 12. The Dual-Modality returns vector forecasting metrics. Accuracy (the upper row), MAPE (the lower row) in percentage

	Source	vanilla LSTM	LSTM_RuBert_SUM	LSTM_RuBert_MEAN	LSTM_QWEN_SUM	LSTM_QWEN_MEAN
Metals and Mining	MTLR	56.364 0.410	39.394 0.409	38.788 0.410	45.455 0.522	52.121 0.246
	TRMK	56.364 0.362	35.152 0.392	42.424 0.192	36.364 0.504	35.758 0.419
Oil and Gas	SNGS	50.303 0.352	53.939 0.865	58.182 1.824	44.848 0.307	49.697 0.106
	SIBN	58.182 0.341	58.182 0.265	58.182 0.216	39.394 0.368	47.879 0.165
Consumer Sector	MGNT	46.667 0.331	53.333 0.417	47.879 0.299	46.061 0.307	48.485 0.235
	LENT	56.364 0.371	49.091 0.400	50.909 0.359	53.333 0.346	52.121 0.331
Construction	PIKK	49.091 0.484	50.303 0.462	57.576 0.436	47.273 0.529	53.333 0.322
	SMLT	53.939 0.328	38.788 0.200	46.061 0.270	36.364 0.311	43.030 0.241
Telecommunications	MTSS	56.970 0.541	53.939 0.473	55.152 0.368	47.879 0.316	45.455 0.193
	RTKMP	55.152 0.246	49.697 0.274	45.455 0.271	44.848 0.171	44.242 0.178
Transport	AFLT	55.152 0.419	51.515 0.641	50.303 0.348	45.455 0.259	52.121 0.182
	FLOT	47.273 0.258	43.636 0.532	52.121 0.262	43.636 0.392	43.636 0.345
Finance	BSPB	46.061 0.410	47.879 0.406	50.909 0.326	47.879 0.369	52.121 0.227
	SFIN	49.697 0.447	44.848 0.445	47.273 0.390	56.970 0.195	56.970 0.272
Chemical Industry	PHOR	41.818 0.231	53.333 0.264	55.152 0.238	60.000 0.354	44.848 0.219
	KZOSP	57.576 0.458	42.424 0.492	41.212 0.491	48.485 0.369	49.697 0.352
Power Engineering	HYDR	59.394 0.380	58.788 0.326	55.758 0.321	47.879 0.292	61.212 0.178
	MRKC	40.000 0.768	42.424 0.742	43.030 0.839	42.424 0.660	41.818 0.543

TABLE 13. The Dual-Modality Approach forecast: Accuracy, MAPE

Model	Accuracy, %	MAPE, %
LSTM-Qwen-Mean	48.552	0.256
LSTM-Qwen-Sum	46.970	0.367
LSTM	52.020	0.397
LSTM-RuBert-Mean	49.798	0.437
LSTM-RuBert-Sum	48.148	0.445

The results from the tables implies that the forecast based on the vectorized news stream using a large language model outperforms the forecast built solely on candlestick data of assets, demonstrating the smallest deviation of the pointwise price prediction from the actual price vector. Additionally, averaging the vectors (Mean) provides the best results.

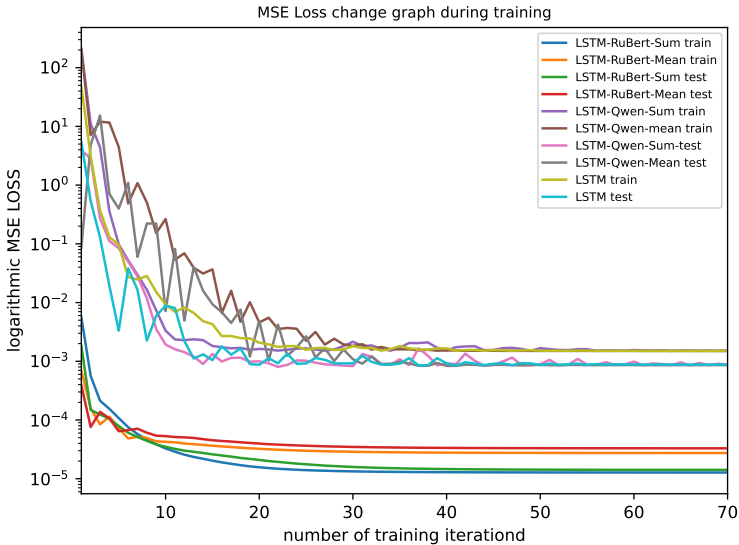


FIGURE 5. Dependence of the mean squared error function values on the number of training iterations for different models. Training and test sets

The dataset (176 stocks of Russian companies traded on the Moscow Exchange and 79,555 Russian-language financial news articles) collected for the study is available at [11].

Conclusion

As a result of the conducted experiments, we demonstrated that adding a textual modality—analyzing the news stream—positively impacts the accuracy of price prediction. On average, the MAPE metric (the deviation of the predicted price from the actual price) decreases by 55%: from 0.397 (LSTM model) to 0.256 (LSTM-Qwen-Mean model). Additionally, predictions based on vectors obtained using the large language model Vikhr-Qwen2.5-0.5b-Instruct outperformed those based on RuBert. This can be partly attributed to the fact that the Qwen model has a significantly larger context window and is trained on a larger text corpus with support for «Chain-of-Thought» (CoT) reasoning. This enhances the model’s ability to reason and capture complex semantic dependencies within the text. The experimental results indicate that the averaging method (Mean) performed better than summation (Sum) and is the preferred method for aggregating news stream vectors.

At the same time, it is important to note that the test data, on which the final metric values were calculated, covers the period from March 28 to August 30, 2024. During this period, the Russian securities market exhibited a general downward trend. The presence of a clear trend is a significant factor that simplifies the prediction task. However, even in this setting, the proposed multimodal approach proved to be the best among those considered.

The training and validation of the model for the rewriting task were conducted on news articles whose length did not exceed the context window of the RuBert model. As a result, artifacts related to the context window size only became apparent during the forecasting phase when the news dataset included articles averaging around 290 words in length. For future improvements in news filtering and classification by company, it is necessary to utilize models with a larger context window, such as Qwen.

The collected dataset [11] demonstrates good structuring and can be used for fine-tuning large language models in Russian or adapted for the Russian language for applications in the financial sector.

TABLE 14. Multimodal approach forecasting metrics in comparison with the approach based on news sentiment score (Baseline) offered by [7]

Model	Ticker	R2	MAPE, %	MAE
LSTM-Qwen-Mean	AAPL	0.989	0.628	0.003
Baseline	AAPL	0.947	2.333	0.018
LSTM-Qwen-Mean	AMZN	0.968	1.601	0.013
Baseline	AMZN	0.870	1.730	0.015
LSTM-Qwen-Mean	GOOGL	0.935	1.394	0.008
Baseline	GOOGL	0.788	2.286	0.020
LSTM-Qwen-Mean	NFLX	0.955	2.361	0.076
Baseline	NFLX	0.919	2.512	0.019
LSTM-Qwen-Mean	TSLA	0.915	3.206	0.006
Baseline	TSLA	0.930	7.423	0.034

For a quantitative comparison of the proposed model, we conducted a computational experiment based on the approach and metrics from the study [7]. Following the methodology of [7], we used time series data of stock prices from five major American companies: AAPL, AMZN, GOOGL, NFLX, and TSLA, along with a dataset of English-language news articles labeled by company for the period from October 12, 2012 to January 31, 2020 (Table 14).

It is worth noting that the dataset used includes text data in English; therefore, we utilized the original Qwen2.5-0.5b-Instruct model [10] for news vectorization. To generate forecasts, we selected and trained the *LSTM-Qwen-Mean* model, as it demonstrated the best overall performance in our study. For evaluation, we used the coefficient of determination (R^2), mean absolute error (MAE), and mean absolute percentage error ($MAPE$).

Thus, we worked with the same time series and evaluation metrics. Across all metrics, except for MAE on NFLX and R^2 on TSLA, the proposed multimodal approach with vector averaging outperformed the best-performing results from the approach in [7]. Based on our computational experiments, we conclude that the proposed multimodal approach demonstrated superior forecasting quality and greater adaptability to both Russian and international markets.

In the future, it is necessary to explore how to incorporate the incoming news stream into the predictive model—specifically, the optimal time window for using news data and the best approach for weighting news messages (e.g. adjusting the weight of a news article based on its chronological position in the dataset).

References

- [1] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. Chitkushev, D. Trajanov. “Evaluation of sentiment analysis in finance: from lexicons to transformers”, *IEEE Access*, **8** (2020), pp. 131662–131682. ^{↑84}
- [2] T.-T. Ho, Y. Huang. “Stock price movement prediction using sentiment analysis and CandleStick chart representation”, *Sensors*, **21**:23 (2021), id. 7957, 18 pp. ^{↑84}
- [3] M. Jaggi, P. Mandal, S. Narang, U. Naseem, M. Khushi. “Text mining of stocktwits data for predicting stock prices”, *Applied System Innovation*, **4**:1 (2021), id. 13, 22 pp. ^{↑84}
- [4] B. Fazlija, P. Harder. “Using financial news sentiment for stock price direction prediction”, *Mathematics*, **10**:13 (2022), id. 2156, 20 pp. ^{↑84}
- [5] Y. Xinli, Ch. Zheng, L. Yuan, D. Shujing, L. Zongyi, L. Yanbin. *Temporal data meets LLM — Explainable financial time series forecasting*, 2023, 13 pp. ^{↑84} 2306.11025 [cs.LG]
- [6] Zh. Boyu, Y. Hongyang, X. -Y. Liu. *Instruct-FinGPT: Financial sentiment analysis by instruction tuning of general-purpose large language models*, 2023, 7 pp. ^{↑84} 2306.12659 [cs.CL]
- [7] T. D. Kulikova, E. Y. Kovtun, S. A. Budennyy. “Do we benefit from the categorization of the news flow in the stock price prediction problem?”, *Dokl. Math.*, **108**, Suppl. 2 (2023), pp. S503–S510. ^{↑84, 104}
- [8] Y. Kuratov, M. Arkhipov. *Adaptation of deep bidirectional multilingual transformers for Russian language*, 2019, 8 pp. ^{↑86, 95, 96} 1905.07213 [cs.CL]
- [9] A. Nikolich, K. Korolev, A. Shelmanov, I. Kiselev. *Vikhr: The family of open-source instruction-tuned large language models for Russian*, 2024, 8 pp. ^{↑86, 95} 2405.13929 [cs.CL]
- [10] A. Yang, B. Yang, B. Hui, B. Zheng, B. Yu, Ch. Zhou, Ch. Li, Ch. Li, D. Liu, F. Huang, G. Dong, H. Wei, H. Lin, J. Tang, J. Wang, J. Yang, J. Tu, J. Zhang, J. Ma, J. Yang, J. Xu, J. Zhou, J. Bai, J. He, J. Lin, K. Dang, K. Lu, K. Chen, K. Yang, M. Li, M. Xue, N. Ni, P. Zhang, P. Wang, R. Peng, R. Men, R. Gao, R. Lin, Sh. Wang, Sh. Bai, S. Tan, T. Zhu, T. Li, T. Liu, W. Ge, X. Deng,

- X. Zhou, X. Ren, X. Zhang, X. Wei, X. Ren, X. Liu, Y. Fan, Y. Yao, Y. Zhang, Y. Wan, Y. Chu, Y. Liu, Z. Cui, Zh. Zhang, Zh. Guo, Zh. Fan. *Qwen2 Technical Report*, 2024, 26 pp. [arXiv:2407.10671](https://arxiv.org/abs/2407.10671) [cs.CL] ↑¹⁰⁴
- [11] K. Khubiev. *Russian financial news dataset*, Kaggle Platform, 2025. [URL](#) [DOI](#) ↑¹⁰³

<i>Received</i>	24.12.2024;
<i>approved after reviewing</i>	30.01.2025;
<i>accepted for publication</i>	27.02.2025;
<i>published online</i>	11.03.2025.

Information about the authors:

Kasymkhan Usufovich Khubiyev



Researcher, Center of Social and Economic Forecasting; Master's Student of "Financial Mathematics and Financial Technologies", Sirius University of Science and Technology, Sirius, Russia. Research interests: artificial intelligence and its application in science, finance, industry, and business



0009-0007-1719-1455

e-mail: kasymkhankhubievnis@gmail.com

Mikhail Evgenyevich Semenov



PhD in Physics and Mathematics, Scientific Supervisor of the "Financial Mathematics and Financial Technologies" direction, Sirius University of Science and Technology, Sirius, Russia. Research interests: Information technology, intelligent data processing and analysis technologies.



0000-0002-0716-5065

e-mail: semenov.me@talantiuspeh.ru

The authors contributed equally to this article.

The authors declare no conflicts of interests.



Мультимодальное предсказание цен акций на примере российского рынка ценных бумаг

Касымхан Юсуфович **Хубиев**^{1,2}, Михаил Евгеньевич **Семенов**²

^{1,2} Университет «Сириус», «Сириус», Россия

Аннотация. Классические методы прогнозирования цен активов в основном опираются на числовые данные, такие как временные ряды цен, объемы торгов, распределение лимитированных ордеров и индикаторы технического анализа. Однако новостной фон играет существенную роль в формировании цен, что делает актуальным развитие мультимодальных подходов, объединяющих текстовые и числовые данные для повышения точности предсказаний.

В данной работе решается задача прогнозирования цен финансовых активов с использованием мультимодального подхода, объединяющего временные ряды цен и текстовую модальность новостного потока. Для исследований был собран уникальный набор данных, включающий временные ряды для 176 акций российских компаний, торгуемых на Московской бирже, и 79555 русскоязычных финансовых новостей.

Для обработки текстовых данных использовались предобученные модели RuBERT и Vikhr-Qwen2.5-0.5b-Instruct (большая языковая модель), временные ряды и векторизованная текстовая модальность обрабатывались рекуррентной нейронной сетью LSTM. В ходе экспериментов сравнивались модели с одной модальностью и двумя модальностями, а также различные методы агрегации векторных представлений текстов.

Качество прогнозов оценивалось по двум ключевым метрикам: точности (ассигасу) предсказания направления изменения цены (рост/снижение) и средней абсолютной процентной ошибке (MAPE) отклонения предсказанной цены от истинной. Эксперименты показали, что добавление текстовой модальности позволяет уменьшить значение MAPE на 55%.

Полученный мультимодальный набор данных представляет ценность для дальнейшей адаптации языковых моделей в финансовой сфере. Перспективные направления исследований включают оптимизацию параметров текстовой модальности, таких как временное окно, тональность и хронологический порядок новостных сообщений. (*Связанные тексты статьи на английском и на русском языках*)

Ключевые слова и фразы: мультимодальная предсказательная модель, количественные финансы, машинное обучение

Благодарности:

Для цитирования: Хубиев К.Ю., Семенов М.Е. Мультимодальное предсказание цен акций на примере российского рынка ценных бумаг // Программные системы: теория и приложения. 2025. Т. 16. № 1(64). С. 83–130. (Англ.+русс.) https://psta.psir.ru/read/psta2025_1_83-130.pdf

Введение

Построение прогноза цены актива является важной задачей для участников финансового рынка для стратегического планирования, оптимального управления инвестиционным портфелем и учета рисков. Существует множество попыток применения методов машинного обучения для построения таких прогнозов.

В связи с ростом популярности моделей глубокого обучения исследователи сместили свой фокус в сторону применения нейронных сетей [1–3]. При этом проблема учета новостного потока, как важного фактора влияния на поведение рынка, переосмысливается с бурным развитием генеративных моделей искусственного интеллекта и больших языковых моделей (ChatGPT, FinGPT, GigaChat, LLama и другие). В финансовой экономике большие языковые модели применяются достаточно редко и их потенциал еще не раскрыт.

Исследователи предпринимают попытки применения моделей обработки естественного языка для улучшения качества прогноза цен активов и стратегий для управления инвестиционным портфелем.

В статье [4] описывается использование оценки тональности новостей в качестве дополнительного параметра. Авторы использовали модель FinBert, обученную на финансовых данных, для оценки тональности (положительная, негативная или нейтральная) новостей. В работе использовались временные ряды свечных данных индекса американского рынка ценных бумаг Standard and Poor's 500 (S&P500). Для предсказания цены использовалась модель машинного обучения – случайный лес. Результатом исследования стал вывод – учёт тональности новостного потока улучшает качество предсказания.

В статье [5] авторы обозначили цель создать мультимодальную модель искусственного интеллекта, способную предоставлять обоснованный и точный прогноз по временным рядам. В ходе работы была реализована модель, которая генерирует прогноз месячной или недельной доходности актива, сопровождаемый текстовым пояснением языковой модели по введенному пользователем запросу. В статье [6] предложен подход к настройке инструкций для интерпретации числовых значений и трактовки финансового контекста.

В исследовании Куликовой с коллегами [7] изучен эффект от учета разделения новостей на группы по тематическому признаку. Авторы продемонстрировали, что в большинстве случаев целесообразно использовать одну тематическую группу новостей для рассмотренных моделей глубокого обучения (темпоральная сверточная сеть, D-Linear, трансформатор, и трансформатор темпорального слияния), а также определили вероятности улучшения прогнозов для рассмотренных 20 тематических групп.

Во всех выше перечисленных исследованиях модели были имплементированы с помощью мультимодального подхода для рынка ценных бумаг США, язык модальности – английский. При этом новостной поток был интегрирован в входной вектор предсказателя не напрямую, а через блок предобработки в виде дополнительного параметра, например, оценки тональности, частоты новостей по активу, класс новости.

Целью данной работы является демонстрация преимущества нового мультимодального метода над предсказаниями, построенными только на числовых данных, и представление русскоязычного набора данных финансовых новостей.

Для достижения поставленной цели сформулированы следующие основные задачи.

- (1) Сформировать мультимодальный набор данных из временных рядов и новостных сообщений.
- (2) Создать предсказательную модель – для использования одной и двух модальностей.
- (3) Провести обучение предсказательной модели и анализ значений функций и метрик точности, Ассигасу и МАРЕ.

В данной работе мы предлагаем новый мультимодальный подход для интеграции новостного потока во временной ряд числовых данных. Текст новостей отображается в векторное представление и подается в модель наряду с вектором временных рядов. Наша гипотеза заключается в следующем: мультимодальный подход позволит предсказательным моделям извлечь семантическую информацию из текста, что улучшит качество предсказания цены актива.

1. Сбор и структурирование данных

Мультимодальность подразумевает применение более одной модальности данных, что влияет на структуру данных и логику разработки предсказательной модели. Мы используем два вида модальности:

числовую – временные ряды цен на акции,

текстовую – новостной поток.

Для обучения предсказательной модели и исследования её работы был сформирован оригинальный набор данных.

Временные ряды в виде свечей с указанием цен открытия (open), закрытия (close), максимальной (high) и минимальной цен (low) мы получили с помощью программного интерфейса Algorack (API) Московской биржи (МОЕХ). Для проведения численного эксперимента мы выбрали временные ряды акций с 7 июля 2022 года по 30 августа 2024 года для 176

компаний. В указанный временной интервал российский фондовый рынок продемонстрировал фазы быстрого роста и падения, индекс ИМОЕХ вырос за этот период с 2213,81 до 2650,32 пунктов (+19,72%).

Мы собрали 79555 новостных сообщений из различных источников, в том числе – сетевое издание «РБК» (1823), «БКС Экспресс» (11331) и «БКС Теханализ» (9670), сайт инвестиционной компании «Финам» (20647), сайт сообщества трейдеров «SmartLab.ru» (30857), а также телеграм-канал «РДВ» (5227).

Выбор указанных источников аргументирован несколькими причинами. Во-первых, на этих источниках опубликованы новости за необходимый временной интервал. Во-вторых, институциональное различие источников, стиль изложения с разной степенью экспертности – позволит сформировать более объективное освещение событий, связанных с используемыми временными рядами.

Новостные сообщения были токенизированы с использованием двух моделей RuBert [8] и Vikhr-Qwen2.5-0.5b-Instruct [9] (далее Qwen). Под словом в контексте токенизированного текста подразумевается токен – элемент векторного пространства в виде индекса словаря токенизатора.

Описательные характеристики (среднее, отклонение, минимальное, максимальное количество слов, а также квантили) набора данных приведены в таблицах 1 и 2 (токены). Необходимо отметить, что токенизация

Таблица 1. Статистические характеристики набора данных после токенизации, RuBert

Источник	Mean	Std	Min	Max	Q25	Q50	Q75
РДВ	134	88	8	512	65	123	187
Финам	221	135	18	512	116	178	284
БКС Экспресс	20	10	4	82	13	17	26
БКС Теханализ	502	37	29	512	512	512	512
РБК	43	7	16	75	39	44	48
SmartLab	21	8	5	82	15	19	25

может привести к увеличению количества слов в тексте (например, за счет разделения слова на составные части).

В таблице 3 приведены примеры того, как изменяется фраза после токенизации. Например, слово «открывает» разделяется на три составных элемента: «от», «##к», «##рывает», где префикс «##» означает, что токен является продолжением предыдущего токена.

Таблица 2. Статистические характеристики набора данных после и токенизации, Qwen

Источник	Mean	Std	Min	Max	Q25	Q50	Q75
РДВ	215	157	3	1324	92	187	304
Финам	453	405	35	5732	211	319	501
БКС Экспресс	36	19	5	163	23	32	47
БКС Теханализ	1493	310	40	2221	1448	1545	1665
РБК	75	12	28	105	68	77	83
SmartLab	33	12	7	120	25	31	39

Таблица 3. Примеры оригинального и токенизированного текста

Оригинальный текст	Токенизированный текст
Доллар снова ниже 69 рублей	До ##лла ##р снова ниже 69 рублей
Москвич банкрот?	Москви ##ч банк ##рот ?
НПО Наука Отчет РСБУ	Н ##П, ##О Наука От ##чет Р ##С ##Б ##У
Т-банк это желтый банк	Т - банк это же ##лт ##ый банк

Особенности новостных сообщений. На ресурсе «БКС Теханализ» новостные статьи часто большие по объему, что накладывает ограничение на использование токенизаторов. В частности, из таблиц 1 и 2 видно, модель RuBert на больших текстах усекает токенизированный вектор. К тому же средняя длина токенизированного текста с помощью модели Qwen превосходит среднюю длину токенизированного текста RuBert, что говорит о том, что модель Qwen обладает более широким словарем и более сильной способностью декомпозиции текста.

Дополнительно мы собрали данные о 176 компаниях, сформировав набор данных из кортежей вида:

(тикер, наименование компании, описание деятельности компании).

Такие данные необходимы в нашем случае для:

- извлечения ключевых слов из описания,
- улучшения способности языковой модели связывать события, описываемые в новостном сообщении, с конкретной компанией и оценивать влияние новости на динамику цены.

Набор данных с новостными статьями содержит следующие параметры: дата публикации, источник, заголовок и тело статьи, теги (ключевые слова). Для источников РДВ, SmartLab заголовок отсутствует, соответствующие поля заполнены специальным словом: *no title*.

В нашем случае теги могут содержать полное или краткое название компании и соответствующий тикер, наименование сектора рынка и т. п. Теги в новостных сообщениях были установлены авторами статей. В случае источника «РДВ» теги отмечались авторами в виде хештегов (например, #цифры, #аналитика), «БКС Экспресс» и «БКС Теханализ» теги обозначались в специальных полях в начале или конце новостной статьи (например, ФосАгро, Российский рынок), и извлекались из *HTML* кода страницы по соответствующим *HTML* тегам. При отсутствии тегов (РБК, SmartLab) параметр в наборе данных остается пустым.

В таблице 4 приведены примеры новостной статьи (фрагмент заголовка) и приписанные к ней теги.

Таблица 4. Примеры новостных статей (фрагмент заголовка) и приписанные теги

Источник	Фрагмент статьи (заголовок)	Теги
РДВ	Сегежа (SGZH): таргет 16.2 руб., апсайд +102...	SGZH
РДВ	Артген биотех (ABIO) завершил доклинические...	аналитика, ABIO
Финам	Индекс МосБиржи восстанавливает позиции и приб...	ФосАгро, ВСМПО-ABCM, CNYRUB
Финам	«Ашинский метзавод» назвал АО "Урал-ВК" своим ...	АшинскийМЗ
БКС Экспресс	«Восходящее окно»: в каких бумагах замечен это...	Селигдар SELG, ЕвроТранс EUTR
БКС Экспресс	«Сила Сибири» выйдет на максимальную мощность...	Газпром GAZP
БКС Теханализ	Мечел. Что ждать от бумаг на следующей неделе	Мечел
БКС Теханализ	На предыдущей торговой сессии акции Норникеля ...	ГМК Норникель

2. Методология

Для проверки нашей гипотезы о преимуществе мультимодального подхода мы запланировали серию экспериментов.

Первая серия экспериментов была направлена на прогнозирование цен с использованием только числовых рядов свечных характеристик актива (цены close, open, high, low). Метрики качества этого эксперимента будут являться базовыми значениями, относительно которых будет оцениваться прирост качества предсказания цен с применением предложенного мультимодального подхода.

Вторая серия экспериментов направлена на получение предсказаний и вычисления метрик точности (Assigasy, MAPE) с применением мультимодального подхода, рассмотрение разных методов агрегации (Sum, Mean) векторизованного новостного потока.

2.1. Использование одной модальности

Сначала мы провели серию экспериментов по прогнозированию цены активов только на временных рядах. Для этого к ежедневным значениям цен (close, open, high, low) мы применили модели классического машинного обучения: линейная регрессия (LinReg), k -ближайших соседей (KNN), решающее дерево (DT), случайный лес (RF) и бустинговый алгоритм XGBoost (XGB), среди моделей глубокого обучения мы использовали – рекуррентную нейронную сеть долгой и короткой памяти (LSTM).

Концептуально эксперимент состоит из двух задач:

- (а) предсказание направления движения цены (рост или падение) – задача бинарной классификации
- (б) предсказание цены – задача регрессии.

На данном этапе эксперимента 176 компаний были сгруппированы по 23 секторам деятельности. Мы случайным образом выбрали 9 секторов экономики, а затем внутри секторов выбрали случайным образом по 2 компании. В таблице 5 перечислены секторы и компании (тикер), которые участвовали в вычислительном эксперименте. В таблице 6 показано распределение новостей по компаниям после фильтрации. В таблице 7 приведены статистические данные о временных рядах цен закрытия выбранных активов. Тепловая карта корреляций временного ряда цен закрытия активов приведена на рисунке 1. Интересная особенность рассматриваемого интервала – рынок претерпел две смены фазы – с общего снижение цен к росту и обратно, как показано на рисунке 2 вертикальными линиями.

Таблица 5. Секторы экономики и компании (тикер), включенные в набор данных

Сектор	Компания (тикер)
Металлы и добыча	Мечел (MLTR), Трубная металлургическая компания (TRMK)
Нефть и газ	Сургутнефтегаз (SNGS), Газпромнефть (SIBN)
Потребительский сектор	Магнит (MGNT), Лента (LENT)
Строительство	ПИК (PIKK), Самолет (SMLT)
Телекоммуникации	МТС (MTSS), Ростелеком (RTKMP)
Транспорт	Аэрофлот (AFLT), Совкомфлот (FLOT)
Финансы	Банк Санкт-Петербург (BSPB), ЭсЭфАй (SFIN)
Химия	ФосАгро (PHOR), Казаньоргсинтез (KZOSP)
Электроэнергетика	РусГидро (HYDR), МРСК Центра (MRKC)

Таблица 6. Распределение новостей по компаниям после фильтрации

Компания (тикер)	Количество новостей
Мечел (MLTR)	4258
Трубная металлургическая компания (TRMK)	11739
Сургутнефтегаз (SNGS)	12674
Газпромнефть (SIBN)	11421
Магнит (MGNT)	1236
Лента (LENT)	311
ПИК (PIKK)	897
Самолет (SMLT)	3392
МТС (MTSS)	1101
Ростелеком (RTKMP)	628
Аэрофлот (AFLT)	1429
Совкомфлот (FLOT)	14476
Санкт-Петербургская биржа (BSPB)	14278
ЭсЭфАй (SFIN)	1647
ФосАгро (PHOR)	2773
Казаньоргсинтез (KZOSP)	168
РусГидро (HYDR)	1921
МРСК Центра (MRKC)	1576

Таблица 7. Описательные характеристики для акций компаний

Тикер	Mean	Std	Min	Max	Q25	Q50	Q75
MTLR	191.8245	72.5652	81.2800	332.8800	123.8500	187.6700	251.6400
TRMK	153.1245	64.9362	55.8200	271.0000	87.1400	166.4200	218.7800
SNGS	27.0104	4.0119	17.3500	36.9600	23.7750	27.3300	30.0250
SIBN	601.5097	163.9205	335.5500	934.2500	452.0500	582.6500	748.9000
MGNT	5691.6429	1161.7684	4040.0000	8444.0000	4665.0000	5495.0000	6375.0000
LENT	814.3870	154.9502	650.0000	1263.0000	716.5000	749.0000	843.5000
PIKK	732.6617	94.8650	518.0000	955.5000	656.7000	732.9000	811.5000
SMLT	3120.8996	594.1018	1926.5000	4145.5000	2572.0000	3045.0000	3713.0000
MTSS	264.5382	32.0791	183.0000	346.9500	239.0000	266.2500	289.7500
RTKMP	68.1797	9.2753	52.2500	92.1000	60.4500	68.0000	74.7000
AFLT	38.1316	10.3131	22.4400	64.4000	27.9700	38.8800	44.1200
FLOT	88.0111	39.5834	29.9200	149.3000	42.1000	97.2000	124.1800
BSPB	211.1501	101.2533	67.5700	387.6800	100.8400	210.9900	295.3400
SFIN	762.9939	428.5679	425.8000	1975.0000	497.4000	518.0000	992.0000
PHOR	6774.6040	618.1977	4997.0000	8153.0000	6416.0000	6763.0000	7278.0000
KZOSP	25.8603	5.2029	15.3500	40.5700	21.9400	27.0700	29.8500
HYDR	0.7697	0.0810	0.5178	1.0278	0.7318	0.7721	0.8210
MRKS	0.5247	0.2382	0.2025	1.0745	0.2735	0.5550	0.7475

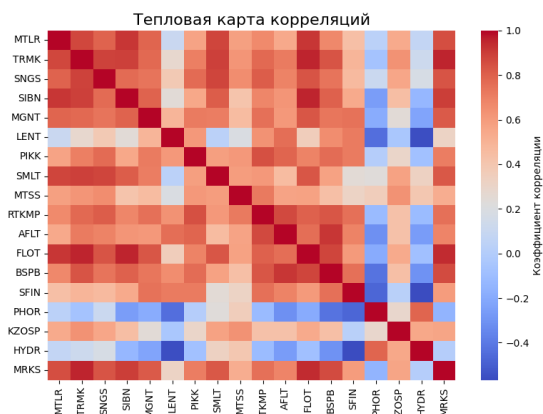


Рисунок 1. Тепловая карта корреляций цен закрытия активов



Рисунок 2. Нормированные исторические цены закрытия активов с указанием точек смены фазы рынка (вертикальные пунктирные линии)

Для оценки качества предсказания в задаче классификации использовалась метрика *Accuracy* (точность), для регрессии – *MAPE* (средняя абсолютная ошибка в процентах). Выбор этих метрик аргументируется постановкой задач. В случае классификации модель должна наиболее точно предсказать направление движение цены – рост (знак плюс «+») или падение (знак минус «-»). Метрика *MAPE* наилучшим образом подходит для оценки качества задачи регрессии с точки зрения доменной области – финансов: *MAPE* демонстрирует среднее отклонение от цены актива в процентах, что легко переводится в денежный эквивалент.

На рисунке 3 приведен процесс разработки модели для использования одной и двух модальностей.

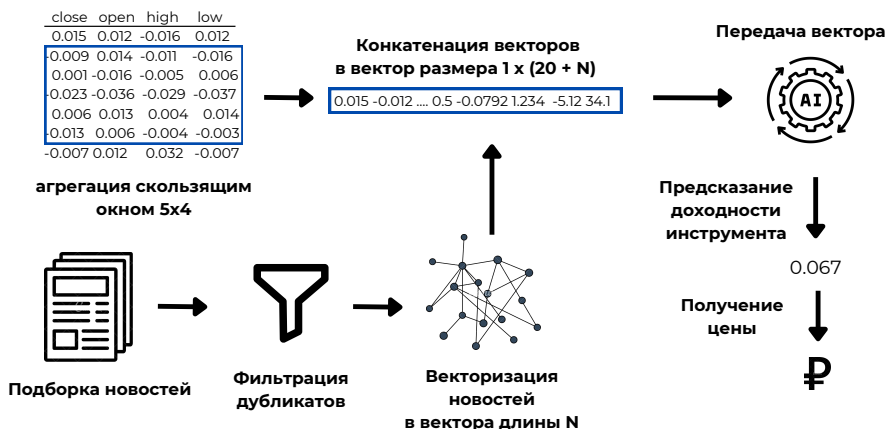


Рисунок 3. Процесс разработки модели для использования одной и двух модальностей

В качестве входного параметра в модель передавался вектор доходности инструмента, рассчитанный для цены закрытия (close) за предыдущие пять торговых сессий:

$$(1) \quad \text{Return}(d+1) = \frac{\text{close}(d+1)}{\text{close}(d)} - 1.$$

На выходе модели – предсказание на следующую торговую сессию.

Для оценки качества предсказания направления движения цены в качестве предсказанного класса использовался знак ($\pm$) предсказанного значения доходности, так как физический смысл доходности инструмента

это величина относительного прироста. Таким образом, положительное значение доходности говорит о росте цены, отрицательное – падении. Для оценки качества прогноза цены актива предсказанный вектор доходности инструмента преобразовывался в цену (в рублях):

$$(2) \quad price(d + 1) = (Return(d + 1) + 1) \cdot price(d).$$

Полученный в результате преобразования поточечно спрогнозированный вектор цен сравнивался с историческим вектором цен активов по метрике *MAPE*.

Выбор величины доходности инструмента (а не цены инструмента) в качестве целевого значения для предсказательной модели обоснован тем, что при выходе цены при росте (падении) рынка за пределы исторического максимума (минимума) возможность использования методов ограничена.

Исходя из этого рассуждения свечные характеристики (цены *close*, *open*, *high*, *low*) учитывались в виде значений *относительных приростов цен*, рассчитанных по формуле аналогично (1).

Далее из значений относительных приростов цен скользящим окном в пять торговых дней формируется вектор-строка и он подается в предсказательную модель. Таким образом, на вход модель получает вектор из двадцати параметров и на выходе предсказывает одно значение – величину доходности инструмента на конец следующей торговой сессии.

2.2. Использование двух модальностей

Для проведения эксперимента с применением новостного потока мы отобрали по ключевым словам новости, которые соответствуют анализируемым активам (таблица 5). Ключевые слова были выбраны как топ-30 слов извлеченных методом TF-IDF. Данный метод вычисляет важность слов в тексте относительно частоты появления слова и его уникальность во всем тексте. Пример извлеченных ключевых слов методом TF-IDF приведен в таблице 8.

Получив список ключевых слов с помощью метода TF-IDF, мы дополнительно добавили ключевые слова с помощью модели ChatGPT-4o, увеличив тем самым вариабельность ключевых слов с помощью перестановок, замен букв, изменения окончаний (таблица 9). Отобранные новости для каждой компании (тикера) были отображены в векторы и отфильтрованы на предмет дубликатов.

Таблица 8. Ключевые слова, полученные из описаний компаний

Тикер	Ключевые слова
MTLR	мечел, горнодобывающей, руда, сырье, энергия, ферросплавы, уголь
SNGS	газ, геологоразведка, нефть, сургутнефтегаз, нефтепродукты, электро-энергия, бурение
SMLT	аренда, девелопмент, девелопер, недвижимость, строительство, московский регион, жилые кварталы
MTSS	абонент, автоматизация, интернет, мобильной связи, провайдер, коммуникационных
BSPB	банк, вклад, дивиденды, финансовые услуги, калининград, спбанк, Санкт-Петербург

Таблица 9. Ключевые слова, полученные дополнительно

Тикер	Ключевые слова
MTLR	мечел, метчел, мечал, mechel, Mchel, ферросплавы, фурросплав
SNGS	сургутнефтегаз, surgutneftegaz, surgut, сурнефтегаз, супраз, сургут, сур-нфтгз
SMLT	самолет, smlt, samolet, samalet, Самлет
RTKMP	постелеком, телеком, rostelecom, telecom, rtkm, ртк, r-telecom, пастелком
HYDR	русгидро, rushydro, rshydro, r-gidro, гидпорус, гидра, русгидра

На рисунке 4 приведена диаграмма распределения новостных статей для компаний после фильтрации.

В качестве векторизатора новостного потока на русском языке мы применили две модели: RuBert [8] и Qwen [9].

Работая с новостным потоком мы столкнулись с двумя проблемами. Первая проблема – это проблема ререйтинга, поэтому необходима фильтрация дублирующих новостей. Для того чтобы наша модель учитывала новость только один раз, необходимо реализовать алгоритм идентификации дубликатов.

Вторая проблема – выявление активнов, на которые оказывает влияние конкретная новость. Данную проблему можно интерпретировать как

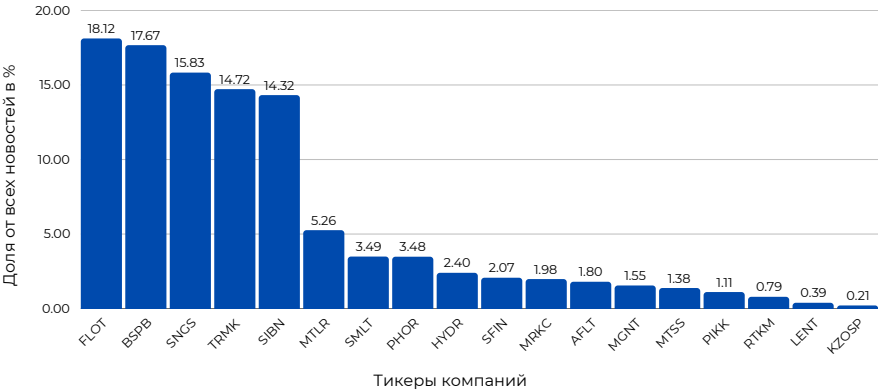


Рисунок 4. Распределение новостей по компаниям после фильтрации (цифры на диаграмме – процент новостей про компанию в наборе данных)

задачу классификации новости, для которой тикеры выступают в качестве меток классов.

Для решения проблемы ререйтинга мы спроектировали сиамскую нейронную сеть. Для этого мы составили обучающий набор данных с помощью API GigaChat следующим образом: для каждой статьи генерировались три перефразированных заголовка и тела статьи, далее случайным образом составлялись в равном соотношении пары из оригинальной и перефразированных новостей и их заголовков.

Сиамская нейронная сеть спроектирована следующим образом: на вход подается пара новостей, с помощью модели RuBert [8] извлекаются векторные представления новостей, над обоими векторами выполняется операция конкатенации, итоговый вектор передается далее в полносвязанную нейронную сеть (MLP). Для того чтобы определить оптимальную глубину модели MLP мы провели серию экспериментов, в ходе которой оценивались качество предсказания и время обработки новостного потока. В результате проведенных экспериментов мы выбрали глубину MLP в 3 слоя.

Затем отфильтрованные новостные статьи переводятся в векторы, чтобы при поступлении новых новостей классификацию дубликатов проводить в One-Shot режиме, что позволяет снизить затраты на время обработки новостного потока и вычислительные ресурсы (в нашем случае GPU V100).

Для решения второй проблемы – сопоставления выборок новостных статей по датам и их последующего использования в прогнозировании цен – необходимо формализовать процесс отбора данных и генерации прогнозов. Мы предполагаем, что предсказание цены закрытия актива производится для каждого торгового дня на момент открытия торгов. При этом в выборку включаются только те новости, которые были опубликованы до начала текущего торгового дня.

Формирование выборок осуществляется путем группировки новостей по дате публикации. Для прогноза цены на заданный день используются материалы, опубликованные в предшествующий торговый день. Например, статьи аналитического характера, такие как материалы под рубрикой «технический анализ» от источника «БКС», публикуемые ежедневно до начала торгов, включаются в выборку для прогнозирования цен активов, которые проанализированы в сообщении. Такой подход позволяет учитывать наиболее актуальную информацию и улучшить качество предсказания.

Для использования двух модальностей обучающие серии сформированы через конкатенацию векторов приростов цен за предыдущие пять дней и векторов новостного потока. Векторы относительных приростов цен конструировались аналогично эксперименту с одной модальностью, а новости выбирались за предыдущий торговый день в соответствии с выбранным активом. Далее эти новости отображались в векторы и агрегировались.

Если за предыдущий день или до открытия торгов текущего дня публикаций не было, то к вектору относительных приростов конкатенируется нулевой вектор длины 768 для модели RuBert и 896 для модели Vikhr-Qwen2.5-0.5b-Instruct (Qwen), иначе добавляется агрегированный вектор новостного потока той же длины. Такие длины конечных векторов соответствуют размерам оригинального выходного слоя предобученных моделей RuBert и Qwen.

В исследовании мы рассмотрели два варианта агрегации новостных векторов: сумма векторов (Sum) и усредненная сумма (Mean). Под суммой векторов мы подразумеваем суммирование значений соответствующих координат векторов. Под усредненной суммой мы подразумеваем, что в соответствующих координатах вектора выставляется среднее арифметическое значение координат агрегируемых векторов.

Базовая модель RuBert имеет ограниченный размер контекстного окна, равный 512 токенам. По этой причине, статьи, превышающие лимит

контекстного окна, усекались или дробились для отдельной обработки, поэтому одной новостной статье могли соответствовать более одного вектора. Модель Qwen имеет размер контекстного окна – 32768 токенов (в 64 раза больше), это достаточно для обработки статей без усечений. Далее мы сравниваем, как влияет на качество предсказания цен векторизатор новостного потока.

Поточечно предсказанные вектора доходностей переводились в цены активов по формуле (2). Качество предсказания оценивалось по двум метрикам: точность (Accuracy) и величина среднего абсолютного отклонения в процентах (MAPE). Точность оценивалась как доля верно предсказанных знаков значений элементов вектора доходностей: положительный или отрицательный. Метрика MAPE демонстрирует, насколько процентов в среднем предсказанная стоимость отличается от истинного значения. Таким образом, мы можем оценить качество предсказания в денежных единицах (рублях).

3. Вычислительный эксперимент

В данном разделе приведем результаты вычислительных экспериментов для двух предсказательных моделей (одна и две модальности). Для разработки предсказательной модели мы использовали фреймворк Transformers (платформа Hugging Face). Для проведения вычислений была использована видеокарта V100.

3.1. Результаты использования одной модальности

Результаты эксперимента предсказания векторов доходностей инструментов только на временных рядах для моделей классического и глубокого машинного обучения представлены в таблице 10.

В таблице 11 приведены усредненные оценки качества предсказания по моделям, данные в этой таблице отсортированы в порядке возрастания средней величины абсолютной ошибки отклонения предсказания от цены актива в процентах (столбец «Отклонение»).

Из результатов эксперимента видно, что рекуррентная модель LSTM демонстрирует наилучшее качество классификации, то есть предсказывает рост или падение, и регрессии – наименьшее среднее отклонение прогнозируемой цены от истинной, но при этом отстает по метрике средней абсолютной ошибки.

Таблица 10. Результаты предсказания векторов доходности с использованием только временных рядов. Точность (слева) и отклонение (справа) в процентах

	Источник	LSTM		XGB		KNN		RF		LinReg		DT	
Металлы и добыча	MTLR	56.364	0.410	40.000	2.089	42.273	2.050	50.909	2.020	50.000	2.029	42.727	2.679
	TRMK	56.364	0.362	40.909	2.105	38.182	2.167	47.273	2.154	49.091	2.114	52.727	2.308
Нефть и газ	SNGS	50.303	0.352	49.091	1.776	48.182	1.775	50.000	1.735	60.909	1.744	52.727	1.857
	SIBN	58.182	0.341	40.000	1.766	58.182	1.746	46.364	1.788	41.818	1.839	51.818	1.813
Потребительский сектор	MGNT	46.667	0.331	39.091	1.517	43.636	1.493	49.091	1.519	40.000	1.709	60.000	1.672
	LENT	56.364	0.371	54.546	2.202	39.091	2.178	52.723	2.145	51.818	2.220	51.818	2.589
Строительство	PIKK	49.091	0.484	40.909	1.565	50.909	1.563	50.000	1.558	44.545	1.637	51.818	1.592
	SMLT	53.939	0.328	42.727	1.577	38.182	1.552	46.364	1.539	49.091	1.536	41.818	1.683
Телекоммуникации	MTSS	56.970	0.541	42.727	1.290	40.000	1.306	45.455	1.520	53.636	1.419	50.000	1.395
	RTKMP	55.152	0.246	45.455	1.299	42.723	1.303	42.727	1.335	50.909	1.355	48.182	1.411
Транспорт	AFLT	55.152	0.419	46.364	2.079	57.273	2.017	52.727	2.062	60.909	1.976	51.818	2.194
	FLOT	47.273	0.258	43.637	2.116	38.182	2.124	42.727	2.104	45.454	2.074	49.091	2.294
Финансы	BSPB	46.061	0.410	49.091	1.612	50.909	1.695	50.909	1.598	54.545	1.602	45.455	1.829
	SFIN	49.697	0.447	40.000	1.603	30.909	1.647	39.091	1.743	48.182	1.960	41.818	1.959
Химическая промышленность	PHOR	41.818	0.231	42.727	1.194	52.723	1.149	48.182	1.168	50.000	1.227	45.455	1.218
	KZOSP	57.576	0.458	49.091	1.198	42.723	1.237	49.091	1.210	46.364	1.217	54.545	1.581
Электроэнергетика	HYDR	59.394	0.380	51.182	1.124	60.000	1.130	48.182	1.214	45.455	1.151	49.091	1.355
	MRKC	40.000	0.768	51.182	1.182	49.091	1.225	54.545	1.214	50.000	1.224	55.455	1.403

Таблица 11. Результаты предсказания с использованием одной модальности (временные ряды)

Модель	Точность, %	Отклонение, %
LSTM	52.020	0.397
XGB	45.000	1.627
KNN	46.010	1.631
RF	48.384	1.646
LinReg	50.152	1.669
DT	49.798	1.824

3.2. Результаты использования двух модальностей

Результаты второго эксперимента, состоявшего в слиянии новостного потока и числовых временных рядов и сравнении предложенного мультимодального подхода с прогнозом построенным исключительно на временных ряда свечей активов, представлены в таблице 12. В таблице 13 представлены усредненные значения метрик предсказания по рассмотренным моделям, данные в таблице отсортированы по столбцу «Отклонение» – в порядке возрастания средней величины абсолютной ошибки отклонения предсказания от цены актива в процентах.

Базовой моделью во втором эксперименте была выбрана нейронная сеть LSTM. С ней мы провели сравнение различных векторизаторов (RuBert, Qwen) и методов агрегации векторов (Sum, Mean).

На рисунке 5 приведена зависимость величин функции ошибки среднеквадратичного отклонения (MSE Loss) от количества итераций обучения для различных моделей для тренировочного (с 7 июля 2022 года по 27 марта 2024 года) и тестового наборов (с 28 марта по 30 августа 2024 года). По графику видно, что после 30 эпох обучения кривые выходят на стационарное значение.

Таблица 12. Результаты предсказания векторов доходности с использованием двух модальностей.
Точность (слева) и отклонение (справа) в процентах

	Источник	vanilla LSTM	LSTM_RuBert_SUM	LSTM_RuBert_MEAN	LSTM_QWEN_SUM	LSTM_QWEN_MEAN
Металлы и добыча	MTLR	56.364 0.410	39.394 0.409	38.788 0.410	45.455 0.522	52.121 0.246
	TRMK	56.364 0.362	35.152 0.392	42.424 0.192	36.364 0.504	35.758 0.419
Нефть и газ	SNGS	50.303 0.352	53.939 0.865	58.182 1.824	44.848 0.307	49.697 0.106
	SIBN	58.182 0.341	58.182 0.265	58.182 0.216	39.394 0.368	47.879 0.165
Потребительский сектор	MGNT	46.667 0.331	53.333 0.417	47.879 0.299	46.061 0.307	48.485 0.235
	LENT	56.364 0.371	49.091 0.400	50.909 0.359	53.333 0.346	52.121 0.331
Строительство	PIKK	49.091 0.484	50.303 0.462	57.576 0.436	47.273 0.529	53.333 0.322
	SMLT	53.939 0.328	38.788 0.200	46.061 0.270	36.364 0.311	43.030 0.241
Телекоммуникации	MTSS	56.970 0.541	53.939 0.473	55.152 0.368	47.879 0.316	45.455 0.193
	RTKMP	55.152 0.246	49.697 0.274	45.455 0.271	44.848 0.171	44.242 0.178
Транспорт	AFLT	55.152 0.419	51.515 0.641	50.303 0.348	45.455 0.259	52.121 0.182
	FLOT	47.273 0.258	43.636 0.532	52.121 0.262	43.636 0.392	43.636 0.345
Финансы	BSPB	46.061 0.410	47.879 0.406	50.909 0.326	47.879 0.369	52.121 0.227
	SFIN	49.697 0.447	44.848 0.445	47.273 0.390	56.970 0.195	56.970 0.272
Химическая промышленность	PHOR	41.818 0.231	53.333 0.264	55.152 0.238	60.000 0.354	44.848 0.219
	KZOSP	57.576 0.458	42.424 0.492	41.212 0.491	48.485 0.369	49.697 0.352
Электроэнергетика	HYDR	59.394 0.380	58.788 0.326	55.758 0.321	47.879 0.292	61.212 0.178
	MRKC	40.000 0.768	42.424 0.742	43.030 0.839	42.424 0.660	41.818 0.543

Таблица 13. Результаты предсказания с использованием двух модальностей

Модель	Точность, %	Отклонение, %
LSTM-Qwen-Mean	48.552	0.256
LSTM-Qwen-Sum	46.970	0.367
LSTM	52.020	0.397
LSTM-RuBert-Mean	49.798	0.437
LSTM-RuBert-Sum	48.148	0.445

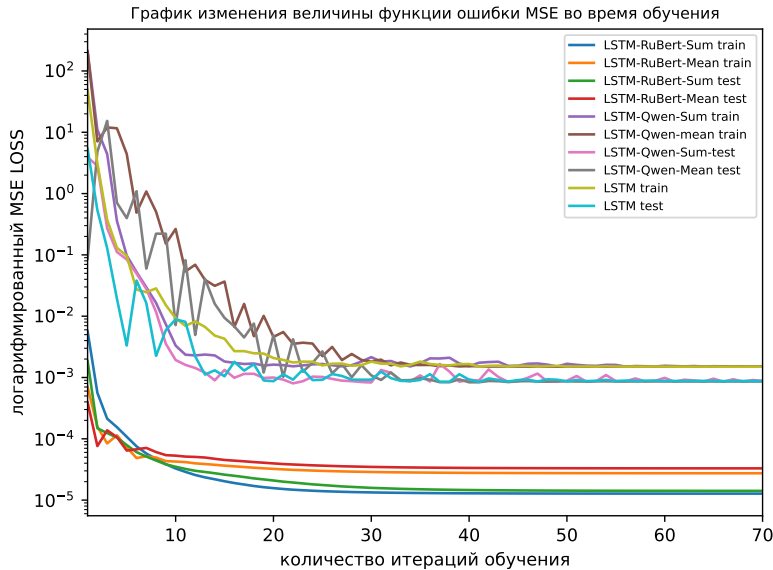


Рисунок 5. Зависимость величин функции ошибки средне-квадратичного отклонения от количества итераций обучения для различных моделей. Тренировочный и тестовой наборы

Из таблиц результатов следует, что прогноз построенный на векторизованном новостном потоке с помощью большой языковой модели превосходит прогноз, построенный исключительно на свечных данных активов, демонстрируя наименьшее значение отклонения поточечного прогноза цены от истинного вектора цен. При этом, усреднение векторов (Mean) дает наилучший результат.

Набор данных (176 акций российских компаний, торгуемых на Московской бирже, и 79555 русскоязычных финансовых новостей), собранный для проведения исследований, доступен по ссылке [11].

Заключение

В результате проведенных экспериментов мы продемонстрировали, что добавление текстовой модальности — анализ новостного потока — положительно влияет на качество предсказания цены. В среднем значение метрики MAPE (отклонение прогнозируемой цены от истинной) уменьшается на 55%: с 0.397 (модель LSTM) до 0.256 (модель LSTM-Qwen-Mean). К тому же качество предсказания на основе векторов, полученных с помощью большой языковой модели Vikhr-Qwen2.5-0.5b-Instruct, оказалось лучше, чем RuBert. От части это следует из того факта, что модель Qwen обладает большим контекстным окном и обучена на большем корпусе текста с поддержкой «цепочки размышлений» (Chain-of-Thoughts, CoT), что улучшает способность модели размышлять и улавливать сложные семантические зависимости внутри текста. Из результатов эксперимента следует, что метод усреднения векторов (Mean) оказался лучше, чем суммирование (Sum), и является предпочтительным методом агрегации векторов новостного потока.

При этом, важно заметить, что тестовые данные, на основе которых рассчитывались финальные значения метрик, включают интервал с 28 марта по 30 августа 2024 года. На этот временной интервал у рынка российских ценных бумаг наблюдается общая тенденция к спаду. Наличие явного тренда является важным фактором, упрощающим задачу предсказания. Однако, даже в этой постановке, предложенный мультимодальный подход оказался наилучшим среди рассмотренных.

Обучение и валидация модели для решения задачи ререйтинга происходило на новостных статьях, объем которых не превосходил контекстное окно модели RuBert, потому артефакты, связанные с величиной контекстного окна, выявились только во время этапа построения прогнозов, когда в новостную выборку попадали статьи длиной в среднем около 290 слов. Потому на будущее, для улучшения модели фильтрации и классификации новостей по компаниям необходимо воспользоваться моделями с большим контекстным окном, например Qwen.

Собранный набор данных [11] демонстрирует хорошую структурированность и может быть применен для тонкой настройки русскоязычных и адаптированных под русский язык больших языковых моделей для применения в финансовой сфере.

Для количественного сравнения предложенной модели мы провели вычислительный эксперимент, в котором за основу взяли подход и метрики статьи [7]. Следуя работе [7], в качестве набора данных мы использовали временные ряды цен пяти крупных американских компаний: AAPL, AMZN, GOOGL NFLX, TSLA и набор англоязычных новостей с разметкой по компаниям за период с 12 октября 2012 года по 31 января 2020 года. (таблица 14).

Таблица 14. Сравнение метрик прогнозирования мультимодального подхода (LSTM-Qwen-Mean) с подходом, основанным на учете оценки тональности новостей (Baseline) [7]

Модель	Тикер	R2	MAPE, %	MAE
LSTM-Qwen-Mean	AAPL	0.989	0.628	0.003
Baseline	AAPL	0.947	2.333	0.018
LSTM-Qwen-Mean	AMZN	0.968	1.601	0.013
Baseline	AMZN	0.870	1.730	0.015
LSTM-Qwen-Mean	GOOGL	0.935	1.394	0.008
Baseline	GOOGL	0.788	2.286	0.020
LSTM-Qwen-Mean	NFLX	0.955	2.361	0.076
Baseline	NFLX	0.919	2.512	0.019
LSTM-Qwen-Mean	TSLA	0.915	3.206	0.006
Baseline	TSLA	0.930	7.423	0.034

Заметим, что использованный набор данных включает текстовую составляющую на английском языке, поэтому для векторизации новостей мы использовали оригинальную модель Qwen2.5-0.5b-Instruct [10]. Для построения прогнозов мы выбрали и обучили модель *LSTM-Qwen-Mean*, так как она показала в среднем лучшее качество в нашем исследовании. В качестве метрик мы использовали коэффициент детерминации (*R2*), среднюю абсолютную ошибку (*MAE*) и среднюю абсолютную ошибку в процентах (*MAPE*).

Таким образом, мы работали с одними и теми же временными рядами и метриками. По всем метрикам, за исключением *MAE* для NFLX и *R2* для TSLA, предложенный мультимодальный подход с усреднением векторов превосходит метрики наилучшего запуска подхода [7]. На основании проведенных вычислений можно сделать вывод, что предложенный мультимодальный подход продемонстрировал лучшее качество прогноза и универсальность применительно к российскому и зарубежному рынку.

В дальнейшем, необходимо исследовать каким образом учитывать входной новостной поток в предсказательной модели – за какой период времени необходимо использовать новости и как учитывать новостные сообщения (например, варьировать вес новости в зависимости от ее хронологического порядка в выборке).

Список использованных источников

- [1] Mishev K., Gjorgjevikj A., Vodenska I., Chitkushev L., Trajanov D. *Evaluation of sentiment analysis in finance: from lexicons to transformers* // IEEE Access.– 2020.– Vol. 8.– Pp. 131662–131682. [↑₁₀₈](#)
- [2] Ho T.-T., Huang Y. *Stock price movement prediction using sentiment analysis and CandleStick chart representation* // Sensors.– 2021.– Vol. 21.– No. 23.– id. 7957.– 18 pp. [↑₁₀₈](#)
- [3] Jaggi M., Mandal P., Narang S., Naseem U., Khushi M. *Text mining of stocktwits data for predicting stock prices* // Applied System Innovation.– 2021.– Vol. 4.– No. 1.– id. 13.– 22 pp. [↑₁₀₈](#)
- [4] Fazlija B., Harder P. *Using financial news sentiment for stock price direction prediction* // Mathematics.– 2022.– Vol. 10.– No. 13.– id. 2156.– 20 pp. [↑₁₀₈](#)
- [5] Xinli Y., Zheng Ch., Yuan L., Shujing D., Zongyi L., Yanbin L. *Temporal data meets LLM — Explainable financial time series forecasting.*– 2023.– 13 pp. [2306.11025 \[cs.LG\]](#) [↑₁₀₈](#)
- [6] Boyu Zh., Hongyang Y., Liu X.-Y. *Instruct-FinGPT: Financial sentiment analysis by instruction tuning of general-purpose large language models.*– 2023.– 7 pp. [2306.12659 \[cs.CL\]](#) [↑₁₀₈](#)
- [7] Куликова Т. Д., Ковтун Е. Ю., Буденный С. А. *Получаем ли мы пользу от категоризации потока новостей в задаче прогнозирования цен акций? // Доклады Российской академии наук. Математика, информатика, процессы управления.*– 2023.– Т. 514.– № 2.– С. 385–394. [↑_{108, 128}](#)
- [8] Kuratov Y., Arkhipov M. *Adaptation of deep bidirectional multilingual transformers for Russian language.*– 2019.– 8 pp. [1905.07213 \[cs.CL\]](#) [↑_{110, 119, 120}](#)
- [9] Nikolich A., Korolev K., Shelmanov A., Kiselev I. *Vikhr: The family of open-source instruction-tuned large language models for Russian.*– 2024.– 8 pp. [2405.13929 \[cs.CL\]](#) [↑_{110, 119}](#)
- [10] Yang A., Yang B., Hui B., Zheng B., Yu B., Zhou Ch., Li Ch., Li Ch., Liu D., Huang F., Dong G., Wei H., Lin H., J. Tang, J. Wang, J. Yang, J. Tu, J. Zhang, J. Ma, J. Yang, J. Xu, J. Zhou, J. Bai, J. He, J. Lin, K. Dang, K. Lu, K. Chen, K. Yang, M. Li, M. Xue, N. Ni, P. Zhang, P. Wang, R. Peng, R. Men, R. Gao, R. Lin, Sh. Wang, Sh. Bai, S. Tan, T. Zhu, T. Li, T. Liu, W. Ge, X. Deng, X. Zhou,

X. Ren, X. Zhang, X. Wei, X. Ren, X. Liu, Y. Fan, Y. Yao, Y. Zhang, Y. Wan, Y. Chu, Y. Liu, Z. Cui, Zh. Zhang, Zh. Guo, Zh. Fan *Qwen2 Technical Report*.– 2024.– 26 pp. arXiv:  2407.10671 [cs.CL] ↑¹²⁸

[11] Khubiev K. *Russian financial news dataset*.– Kaggle Platform.– 2025.   ↑¹²⁷

Поступила в редакцию 24.12.2024;
одобрена после рецензирования 30.01.2025;
принята к публикации 27.02.2025;
опубликована онлайн 11.03.2025.

Рекомендовал к публикации

к.т.н. Е. П. Куршев

Информация об авторах:



Касымхан Юсуфович Хубиев

исследователь в центре социально-экономического прогнозирования, магистрант направления «Финансовая математика и финансовые технологии», Университет «Сириус». Научные интересы: искусственный интеллект и его приложения в науке, финансах, промышленности и бизнесе.



0009-0007-1719-1455

e-mail: kasymkhankhubievnis@gmail.com



Михаил Евгеньевич Семенов

к.ф.-м.н., научный руководитель направления «Финансовая математика и финансовые технологии», Университет «Сириус», Научные интересы: информационные технологии, интеллектуальные технологии обработки и анализа данных.



0000-0002-0716-5065

e-mail: semenov.me@talantiuspeh.ru

Авторы внесли равный вклад в подготовку публикации.

Декларация об отсутствии личной заинтересованности: *благополучие авторов не зависит от результатов исследования.*

Том 16

Выпуск 1(64)

2025 г.

АВТОРСКИЙ УКАЗАТЕЛЬ

Бабенко Виктор Николаевич

О реализации QR-разложения на трехмерном систолическом массиве **45, 56**

Винокуров Игорь Викторович

Использование модели Mask R-CNN для сегментации объектов недвижимости на аэрофотоснимках **3, 44**

Грибова Валерия Викторовна

Комплекс онтологий как модель системы интеллектуальной поддержки в реабилитации пациентов, перенесших инсульт **61, 79**

Ковалев Роман Игоревич

Комплекс онтологий как модель системы интеллектуальной поддержки в реабилитации пациентов, перенесших инсульт **61, 79**

Лебедев Сергей Васильевич

Комплекс онтологий как модель системы интеллектуальной поддержки в реабилитации пациентов, перенесших инсульт **61, 79**

Лифшиц Александр Яковлевич

Комплекс онтологий как модель системы интеллектуальной поддержки в реабилитации пациентов, перенесших инсульт **61, 79**

Невечеря Артём Павлович

О реализации QR-разложения на трехмерном систолическом массиве **45, 56**

Окунь Дмитрий Борисович

Комплекс онтологий как модель системы интеллектуальной поддержки в реабилитации пациентов, перенесших инсульт **61, 79**

Семенов Михаил Евгеньевич

Мультимодальное предсказание цен акций на примере российского рынка ценных бумаг **83, 130**

Федорищев Леонид Александрович

Комплекс онтологий как модель системы интеллектуальной поддержки в реабилитации пациентов, перенесших инсульт **61, 79**

Хубиев Касымхан Юсуфович

Мультимодальное предсказание цен акций на примере российского рынка ценных бумаг **83, 130**

Шалфеева Елена Арефьевна

Комплекс онтологий как модель системы интеллектуальной поддержки в реабилитации пациентов, перенесших инсульт **61, 79**

Шепета Екатерина Ивановна

*Комплекс онтологий как модель системы интеллектуальной поддержки
в реабилитации пациентов, перенесших инсульт 61, 79*

Шестопалов Евгений Юрьевич

*Комплекс онтологий как модель системы интеллектуальной поддержки
в реабилитации пациентов, перенесших инсульт 61, 79*

Шуматов Валентин Борисович

*Комплекс онтологий как модель системы интеллектуальной поддержки
в реабилитации пациентов, перенесших инсульт 61, 79*

VOL. 16 ISSUE 1(64) 2025

AUTHOR INDEX

Babenko Viktor Nikolaevich

On the implementation of QR-decomposition on a three-dimensional systolic array **45**, 56

Fedorishchev Leonid Alexandrovich

The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke **61**, 79

Gribova Valeria Viktorovna

The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke **61**, 79

Khubiyeu Kasymkhan Usufovich

Multimodal Stock Price Prediction: A Case Study of the Russian Securities Market **83**, 106

Kovalev Roman Igorevich

The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke **61**, 79

Lebedev Sergey Vasilyevich

The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke **61**, 79

Lifshits Alexander Yakovlevich

The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke **61**, 79

Nevecherya Artyom Pavlovich

On the implementation of QR-decomposition on a three-dimensional systolic array **45**, 56

Okun Dmitri Borisovich

The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke **61**, 79

Semenov Mikhail Evgenyevich

Multimodal Stock Price Prediction: A Case Study of the Russian Securities Market **83**, 106

Shalfeeva Elena Aref'evna

The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke **61**, 79

Shepeta Ekaterina Ivanovna
The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke **61**, 79

Shestopalov Evgenij Yur’evich
The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke **61**, 79

Shumatov Valentin Borisovich
The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke **61**, 79

Vinokurov Igor Victorovich
Using the Mask R-CNN model for segmentation of real estate objects in aerial photographs **3**, 23

VOL. 16 ISSUE 1(64) 2025

CONTENTS

Research Article

APPLIED SOFTWARE SYSTEMS

IGOR V. VINOKUROV[✉]. *Using the Mask R-CNN model for segmentation of real estate objects in aerial photographs (In Engl., In Russ.)* ... **3–23**, 24–44

Research Article

HARDWARE AND SOFTWARE FOR DISTRIBUTED AND SUPERCOMPUTER SYSTEMS

VIKTOR N. BABENKO[✉], ARTYOM P. NEVECHERYA. *On the implementation of QR-decomposition on a three-dimensional systolic array (In Russ.)* 45–56, **57–59**

Research Article

MEDICAL INFORMATICS

VALERIA V. GRIBOVA, VALENTIN B. SHUMATOV, SERGEY V. LEBEDEV, ELENA A. SHALFEEVA, EVGENIJ YU. SHESTOPALOV, DMITRI B. OKUN[✉], ROMAN I. KOVALEV, EKATERINA I. SHEPETA, LEONID A. FEDORISHCHEV, ALEXANDER YA. LIFSHITS. *The ontology complex as the model of intelligent system to support rehabilitation of patients after stroke (In Russ.)* 61–79, **80–82**

Research Article

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

KASYMKHAN U. KHUBIYEV[✉], MIKHAIL E. SEMENOV. *Multimodal Stock Price Prediction: A Case Study of the Russian Securities Market (In Engl., In Russ.)* **83–106**, 107–130

Author index **133**

Чтобы сменить язык страницы, кликните, пожалуйста, флаг в верхнем углу

Авторский указатель 131

Содержание 2