**THIS DOCUMENT IS STILL IN EDITING STAGE, SINCE IT WAS TRANSLATED FROM RUSSIAN INTO ENGLISH.**

**PLEASE PUT YOUR COMMENTS OR TEXT INSIDE THE DOCUMENT IN DIFFERENT COLORS**

**PROJECT INFORMATION FOR INTERNATIONAL GRANTS**

**Project Subject**

***Project Title:***

Development of predictive models for diagnosing early stages of lung, trachea, bronchial, and breast cancer

***End-product description:***

A set of machine learning models for predicting the probability of a user having four types of cancer using his/her medical data (X-rays). In addition to the results of the model's calculation, the user receives an individualized calculation of future cancer probabilities and personalized recommendations for cancer prevention and prevention based on big data from public reliable medical databases. Interaction with the user (individual, medical institutions) is done remotely via mobile and web applications.

***Is Phase 2 (year) R&D required?***

Yes

***Rationale for the need for Phase 2 (year) R&D***

The 2nd stage R&D will be aimed at improving the minimal demonstration version of the developed product, based on the results of the 1st stage of R&D, in order to scale up and commercialize it in the format of sales to patients and medical institutions.

***The main direction of the program***

Medicine and Health Technology.

***Subdirections:***

Artificial Intelligence. Neurocomputer technologies and evolutionary algorithms.

***Focus areas:***

Physician decision support systems

***Priority areas:***

Information and telecommunication systems

***Priority Software Class:***

Decision Support Systems (DSS)

Predictive and augmented analytics, including integration with tools of advanced data processing (Data Science), automatic processing and interpretation of data using AI, including semantic data analysis technologies from various sources

***Direction within the Strategy for Scientific and Technological Development:***

Transition to advanced digital, intelligent production technologies, robotic systems, new materials and methods of construction, creation of systems for big data processing, machine learning and artificial intelligence

***Keywords:***

Computer vision, image recognition, early diagnosis, risk probability calculation, deep learning, artificial intelligence

***Requested grant amount:***

***Timeframe for the project:***

12 months

**INFORMATION ABOUT THE APPLICANT AND PROJECT PARTICIPANTS**

**Background Information**

***Type of applicant:***

Legal entity. InTime BioTech LLC (Delaware, USA)

***Director of the enterprise:***

Oleg Teterin

***Scientific supervisor of the project:***

Abdulina Marina Vadimovna

***Members of the project team:***

| **Employee** | **Position** | **Role in the project** | **Experience in implementing projects on similar topics** |
| --- | --- | --- | --- |
| Abdulina  Marina  Vadimovna | Scientific  project team leader | Preparation of the overall technical solution  (system design  systems), organization of processes  development,  testing | Software development and support, writing user manuals, creating training courses, supporting software users. Substantial experience with data from open and closed sources. Data analyst on longevity markers. |
| Pavlov  Andrey Valeryevich | Leading  Data scientist | Conducting data retrieval work;  Developing  technical  solutions;  Data quality assurance | Substantial experience with building predictive models of diseases (cardiovascular disease, diabetes, chronic kidney disease, prostate cancer, breast cancer) using tabular data. Experience in team management, with 5+ members. |
| Polyansky  Nikolay  Aleksandrovich | Lead  Data scientist developer | Conducting data retrieval work;  Developing  technical  solutions;  Data quality assurance | Experience in NLP and computer vision. Designing predictive models, deep learning models for image segmentation and text generation. |
| Vostretsov  Maxim  Sergeyevich | Lead  Data scientist developer | Conducting data retrieval work;  Developing  technical  solutions;  Data quality assurance | Building computer vision models for diagnosing diseases (Alzheimer's and Parkinson's disease, tuberculosis, oral cavity) and fitness tests. Creation of architecture of complex hardware-software complexes and intelligent diagnostic systems. |
| Bastien Bennetot | Scientific advisor | Reviewing scientific evidence based article, scientific consulting | Bioinformatics, scientific reading and writing and molecular biology |
| Sailaja Surapuraju | Research Analyst | Research Quality Data for respective projects | Found Datasets and evidence based articles for relative projects,and Preparing documentation for presentation |
| Catherine Estibaliz Pabon Perilla MD. | Medical Doctor Science and Healthcare Advisor | Medical and scientific research consulting. Health policies and international advisory standards. | Medical doctor consulting expert & healthcare management advisor. Leading scientific, organizational and healthcare innovation strategies for more than 20 years. |

***Plans to attract new specialists:***

Oleg Teterin has formed an experienced team of highly qualified specialists with experience in scientific, clinical and practical work with data and the construction of disease predictive models, as well as entrepreneurial and managerial activities. There is a group of experts who are ready to provide consulting support for the project on a pro bono basis. Current staff is sufficient for the first year of R&D.

Future plans:

- In the second year of the project: to hire additional developers-engineers of computer vision, a project manager and two full-stack developers with knowledge of Python and Django, one Android developer (Kotlin) and one iOS developer (Swift) to refine the functionality of the minimum viable product (MVP) in order to enter the industrial production. In the case of more rapid development of the project, this plan can be implemented in the first year of the project.

- In the third year of the project: to form an administrative department: attract a marketing specialist, a specialist in SMM and SEO, as well as five B2B sales specialists to the team. In the case of more rapid development of the project, this plan can be implemented in the second year of the project.

**Applicant Information**

***Applicant:***

Limited Liability Company InTime BioTech LLC

***Company Registration Date:***

07.02.2020

***Region of the applicant:***

Wilmington, Delaware, USA

Moscow, Russia

***Revenue from sales of goods (works, services) during the last calendar year (USD):*** 0

***Average number of employees for the last calendar year, persons:*** 40

***Line of business activity of the enterprise:***

Research and development activities

***To be filled in if you have selected "Other" in the "Company's activity profile" field:***

**Founders**

| **No. n/a** | **Founder** | **Share** |
| --- | --- | --- |
| 1 | Teterin Oleg | 78,00 |

| 2 | Reserved for 40 team members | 22,00 |
| --- | --- | --- |

**PROJECT CONTENT**

**Abstract of the project**

| According to the WHO ([https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of death](https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of)) for the year 2020, cancer is among the top ten leading causes of death. Lung, tracheal and bronchial cancers and breast cancer are among the most common and deadly cancers ([https://www.who.int/news-room/fact sheets/detail/cancer](https://www.who.int/news-room/fact)). Detecting cancer early increases a patient's chances of surviving and living longer than average, with less costly treatments that improve the patient's condition.  On average, it takes up to 20 minutes for a radiologist to describe an X-ray image. Using neural networks allows the doctor to do it in less than 3 seconds with 95-98% accuracy of description. Artificial intelligence helps get more information about the disease and treatment plan (by offering the doctor a second "independent" opinion), increasing the reliability of diagnosis by 48%.  Our project involves the development of a set of machine learning models that, based on the user's medical data (images), predicts the probability of four types of cancer (lung, bronchial, tracheal, and breast cancer). In addition to calculating the risks, the user receives science-based personal recommendations for cancer prevention and prevention. Our solution helps to reduce the mortality rate from certain types of cancer by enabling early or preclinical diagnosis and thus ensuring the safety of human health and increasing the duration of a healthy life.  To implement the project, the team uses artificial intelligence technologies such as deep learning, computer vision algorithms, methods of analysis and construction of predictive models, hybrid system architectures, a structure for storing information about the user (team know-how).  Areas of application of our project: medical (disease diagnosis, medical decision support system), health care, health security of genetic populations, increasing healthy life expectancy, educational, structured storage of personal information.  The solution we offer can be used by individuals, public and private medical institutions, health centers, medical spas, and pharmaceutical companies. |
| --- |

**Belonging to the projects in the field of AI**

***Rationale for the relevance of the project subject in the field of AI:***

The project will create a set of machine learning models to predict the probabilities of diseases: lung, trachea, bronchial, and breast cancer based on human medical data (images), and obtain results comparable to human intelligence when performing these tasks.

***Artificial Intelligence Technology:***

Computer Vision.

Intellectual support for decision making

Autonomous solution of various tasks

***Rationale for the choice of technology:***

As part of the creation of the project, user data sets in the form of images for the specified diseases will be obtained and then machine learning models will be created based on this data.

Also, a system for calculating risks and providing recommendations on risk prevention based on the array of data from public medical databases will be developed. The use of the method implies the creation of a mobile and web application with API-gateways for remote connection to medical institutions.

***The technological problem that the project is aimed at solving:***

1. Collect data sets and train classifiers, including for making a diagnosis based on photo and video analysis with a given level of accuracy, as well as for training the system "by the situation".
2. Pattern recognition with training "at the first time" (one or several objects), allowing predictive output of results, including in emergency situations.
3. Preparation of decisions based on open data sources and unstructured information, including for use in intelligent decision support systems to solve strategic issues and/or adaptive dynamic management of complex objects.
4. Providing decision support based on multiyear data, including for calculation of rationing in industries.
5. Autonomous semantic segmentation, classification and identification of objects, partitioning into sub-objects and recognition of individual details, including real time mode.
6. Predictive and prescriptive analysis to predict the development of a situation based on data analysis and automated decision making in real time (including creation of methods and models).
7. Data Partitioning with Artificial Intelligence, including automation of data preparation for application tasks.
8. Combine data-driven models with "classical" models and combine various artificial intelligence methods, including when used in poorly formalized application domains (including the creation of hybrid models).
9. Learning by doing or by analogy (including model building).
10. Data primary processing (verification) and data quality monitoring (including model building).

***Rationale for the choice of technological tasks:***

The project is based on a set of machine learning models and builds on the project team's work in predictive medicine.

Methods such as deep learning, hybrid system architectures, various computer vision algorithms in disease prediction using image (image) analysis, specialized neural network architectures (super precision, ResNet) will be used to train models.

Appropriate methods will be applied for data partitioning.

A special formalized structure that stores patient information (in personalized and depersonalized forms), including historical (static and dynamic) data, is used to form a knowledge base.

***Project Outcome:***

1. Creation and/or development and/or implementation of new technologies, software tools or software and hardware complexes, and their scaling,
2. Creation and processing of data sets, including collection, cleaning, markup, validation, depersonalization, storage, enrichment, auditing, publication, and actualization.

***Rationale for Outcome Selection:***

The outcomes of the project are:

- a set of machine learning models for predicting these diseases; - a knowledge base containing scientifically validated risks and recommendations for the prevention and prevention of these diseases;

- decision-making and response generation modules for issuing risks and recommendations, as well as model results.

**Scientific and technical part of the project**

***Novelty of the solutions proposed in the innovation project:***

1. The use of methods of machine learning and computer vision to diagnose types of cancer at an early stage, whereas at the moment most cancers are detected at stages III and IV.

2. Special formalized structure to store information about users (in personalized and depersonalized forms), including historical data.

3. Analysis of a single image (snapshot) by several predictive models, which will allow excluding or confirming several (hidden) pathologies in the future.

4. Issuing to the user individual recommendations for disease prevention and prevention, based on scientifically proven results of medical research, as well as a list of supporting scientific publications.

***Ways and methods of solving the set R&D tasks:***

Our project involves the creation of web and mobile applications that allow the user to remotely upload images (pictures) and receive an answer about the presence or absence of four types of cancer (lung, bronchial, tracheal and breast cancer), as well as the probability of the user having each type of cancer in the future and scientifically based personal recommendations for the prevention and prevention of the studied diseases. Our solution helps to reduce the mortality rate from certain cancers by enabling early or preclinical cancer diagnosis and thus ensuring the health security of the individual (and the nation) and increasing the length of their healthy life.

To solve this problem we plan to use the following approach:

1. General population data (images on the diseases under study) are collected from public sources, forming a general knowledge base for training predictive models.

2. from the data uploaded by the user (images of the investigated diseases) is formed a base of personal knowledge of the given user.

3. At the same time a scientific database is formed, based on the results of scientific publications and medical research.

4. A set of models is created for diagnostics of the researched diseases, using data from the general knowledge base. As the knowledge base is replenished, each model is retrained or retrained, it is also possible to create a fundamentally new model, which is released into production.

5. Each model analyzes the user's images and sends the result of the analysis to the decision-making module.

6. The decision-making module makes a conclusion about the presence or absence of pathology, and on the basis of the knowledge base of scientific data finds confirmation or refutation of the disease development in the future. The results of the decision-making module are sent to the response generator. In the future, the decision-making module will also analyze the results of model predictions for other pathologies. For example, a chest CT scan was obtained to predict lung cancer, on which the model system found no risk of cancer, but did find a risk of pneumonia.

7. A response generator based on a knowledge base of scientific data generates a response that includes the results of all model complex calculations, the likelihood that the user will develop a disease in the future, ways to prevent and prevent disease, and a list of scientific publications that support the specified risks and recommendations.

8. The cycle is repeated.

***The material and technical base necessary for the implementation of the project (available and/or planned to be attracted:***

The team uses artificial intelligence technologies:

1. Deep Learning - to train predictive models and queries to recognize and partition the resulting data. A specialized neural network architecture will be used for each low-level task, in particular deep convolutional neural networks and residual neural networks (ResNets) will be used for image analysis.

2. Machine learning methods - to analyze images: radiological images of the user, photos and scanned copies of medical records.

3. Hybrid knowledge-based systems architecture - allows the use of both machine learning and computer vision techniques in general, and for explaining the results of predictive models.

In addition to a set of machine learning models for predicting diseases, the system includes two knowledge bases:

- A user data knowledge base containing digital information about biological markers of patients in the general population obtained from open sources (general knowledge), and historical (static and dynamic) patient-specific data obtained from closed sources (personal knowledge). Models trained on general population data will be used to generate predictions for a particular patient.

- A knowledge base of scientific data that contains information from public medical databases that confirms or disproves the risk of future disease, and provides scientifically proven results of medical research to give the patient personalized recommendations for disease prevention and management.

The system also includes technological services to implement the same functionality for all predictive models, ways of interaction with the user through mobile and web applications: creation of a user account, uploading images, anonymizing user data (if necessary), displaying the results of model calculations, calculating the probability of disease for the user, giving the user personalized recommendations for risk avoidance, providing a list of scientific literature, confirming

***Backlog on the subject matter of the project:***

Terms of reference for a minimally viable product were prepared in January 2020. By September 2021, six predictive models based on public tabular data, five models using computer vision (including early diagnosis of tuberculosis and melanoma), and two fitness tests (Bondarevsky test, lifting from a sitting position without the help of hands) were created. Presentation of the results was presented on September 23, 2021 at the conference "Computational Biology and Artificial Intelligence for Personalized Medicine" (section "AI in Medicine"), organized by the Russian Ministry of Science and Higher Education (<https://cbai.endocrincentr.ru/>).

The architecture of the systems based on big data, used for the project implementation, is reflected in the publication:

Dushkin R. V., Lelekova V. A., Maksimov V. S., Zolman O., Teterin O. О. (2021) A method for early diagnosis of fatal diseases based on AI and Big Data that can significantly prolong active life. PREPRINTS.RU.<https://doi.org/10.24108/preprints-3112261>

In addition, the team extensively researched international materials on the topic of the project: - Çallı E, Sogancioglu E, van Ginneken B, van Leeuwen KG, Murphy K. Deep learning for chest X-ray analysis: A survey. Med Image Anal. 2021 Aug;72:102125. <https://doi.org/10.1016/j.media.2021.102125>.

- Bhandary, A., AnanthPrabhu, G., Basthikodi, M., & ChaitraK., M.. (2021). Early Diagnosis of Lung Cancer Using Computer-Aided Detection via Lung Segmentation Approach. ArXiv, <https://arxiv.org/abs/2107.12205>.

- Fan, S., Xu, R., & Yan, Z. (2021). A Medical Pre-Diagnosis System for Histopathological Image of Breast Cancer. ArXiv, <https://arxiv.org/abs/2109.07878v1>.

- Zhehao He, Wang Lv, Jian Hu, "A Simple Method to Train the AI Diagnosis Model of Pulmonary Nodules," Computational and Mathematical Methods in Medicine, vol. 2020, Article ID 2812874, 6 pages, 2020. <https://doi.org/10.1155/2020/2812874>.

- Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T.L., Zhang, F., & Grundmann, M. (2020). BlazePose: On-device Real-time Body Pose tracking. ArXiv, https://arxiv.org/abs/2006.10204. - Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C., Yong, M.G., Lee, J., Chang, W., Hua, W., Georg, M., & Grundmann, M. (2019). MediaPipe: A Framework for Building Perception Pipelines. ArXiv, <https://arxiv.org/abs/1906.08172>.

- Mercan E, Mehta S, Bartlett J, Shapiro LG, Weaver DL, Elmore JG. Assessment of Machine Learning of Breast Pathology Structures for Automated Differentiation of Breast Cancer and High-Risk Proliferative Lesions. JAMA Netw Open. 2019;2(8):e198777. <https://doi.org/10.1001/jamanetworkopen.2019.8777>.

- Gardezi, S., Elazab, A., Lei, B., & Wang, T. (2019). Breast Cancer Detection and Diagnosis Using Mammographic Data: A Systematic Review. Journal of medical Internet research, 21(7), e14464. <https://doi.org/10.2196/14464>.

- Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. (2018). Artificial intelligence in radiology. Nature reviews. Cancer, 18(8), 500-510. https://doi.org/10.1038/s41568-018-0016-5. - Dholey M. et al. (2018) A Computer Vision Approach for Lung Cancer Classification Using FNAC-Based Cytological Images. In: Chaudhuri B., Kankanhalli M., Raman B. (eds) Proceedings of 2nd International Conference on Computer Vision & Image Processing. Advances in Intelligent Systems and Computing, vol 704. Springer, Singapore. <https://doi.org/10.1007/978-981-10-7898-9_15>.

A set of computer vision models for early disease detection, developed by the team, is currently in the process of patent application preparation.

As for the testing platform Moscow, Russia has been chosen. The team received letters of support from the Agency of Artificial Intelligence in Moscow, Russia and "Center for Sports Innovative Technologies and National Teams Training" of the Moscow City Sports Department (Russia), which confirms the scientific and commercial potential of the project (see the section "Additional Documents on the Project").

**Prospects for commercialization**

***Competitive advantages of the created product, comparison of its technical and economic characteristics with the main analogues, including the world analogues:***

The team analyzed the products available on the Russian and global markets, which are similar to the product developed within the project in terms of their functions and characteristics:

1. Care Mentor AI (<https://carementor.ru/>), a service platform for radiation diagnostics based on artificial intelligence (AI), a participant in the Skolkovo project, offers a service for the detection of COVID-19, lung and breast cancer. The computer vision system analyzes medical radiology images, determines pathology probabilities, and prioritizes cases with pathological findings for physicians. The product has performed well in test operation, but is only presented on the Russian market. Integration of the service with the SberMedI ecosystem is planned.

Our advantage is the greater range of diseases detected by computed tomography (CT) of the lungs. The Care Mentor AI system provides a study protocol as a result of the analysis. Our product additionally calculates the probability of cancer occurrence in the future and gives recommendations on risk prevention and cancer prevention based on the scientifically proven results of medical research. We also offer a mobile app and web platform to interact with individuals (B2C) and with medical institutions (B2B) by connecting through API-gateways.

2. Biopharmaceutical company AstraZeneca (<https://www.astrazeneca.com/>) conducts retrospective analysis of CT lung scans using the Botkin.ai medical information analysis platform (<https://botkin.ai/projects/>). The application of AI in the analysis of CT scans originally performed to diagnose coronavirus infection revealed suspicion of lung cancer. AstraZeneca plans to launch similar projects on breast cancer and chronic obstructive pulmonary disease in the near future.

In AstraZeneca's project, cancer is detected only at the stage of an advanced process. In our proposed solution, cancer can be diagnosed at an early or preclinical stage. Our method shows the probability of developing bronchial and tracheal cancer as well, which AstraZeneca does not deal with. In addition, the use of a third-party platform based on AI technologies for image analysis makes the AstraZeneca project dependent on contract terms.

3. The ChestLink product from the Lithuanian company Oxipit ([www.oxipit.ai](http://www.oxipit.ai)) is a diagnostic aid and case prioritization tool. The purpose of the product is to identify radiographs without abnormalities, allowing the radiologist to avoid wasting time analyzing these images and focus on cases with existing pathology.

Although ChestLink supports more than 70 radiological results covering 90% of respiratory pathology, it does not diagnose cancerous masses. In addition, the results are only reported as "normal" or "abnormal." The product we are developing gives a complete analysis of the image, noting the likelihood of cancer development and ways to prevent it.

4. The VisiRad™ product (<https://imidex.com/solutions/>) by IMIDEX, Inc. of America uses computer vision technology to identify nodules in the lungs on chest x-rays. IMIDEX uses software from another U.S. company, Laurel Bridge Software, Inc. (<https://www.laurelbridge.com/>).

In contrast to the VisiRad™ project, we determine the likelihood of several cancers using our own developed methods rather than third-party software. Also, IMIDEX is not represented on the Russian market.

5. AI Diagnostic service from Intel Diagnostic automatically detects neoplasms in the lungs using chest CT scan and allows to detect cancer at early stages.

The main advantage of our project over AI Diagnostic is a set of machine learning models for cancer diagnosis and product packaging for B2C and B2B clients in global markets.

6. "The cloud service of PhthisisBiomed (<http://ftizisbiomed.ru/>) helps analyze digital fluorographic images and chest x-rays.

The service's pathology classifier is capable of distinguishing and cataloguing 9 types of pathologies, but the results are only indicated as "normal" or "abnormal." Our system provides complete results of image analysis, including the likelihood of developing cancer and ways to prevent it.

7. The Celsus system from Medical Screening Systems (<https://celsus.ai/>) analyzes fluorograms for pathological changes according to codes 1-23 of the Fluoroscopy Numerical Description Codes.

Cels system provides the user only with an examination report. The product we offer additionally calculates future cancer risks and provides recommendations for their prevention. Besides, the product of the company "Medical Screening Systems" is oriented to the Russian market only.

Thus, the fundamental differences between the solution we offer and similar products are as follows:

- A set of machine learning models for remote diagnostics of different types of cancer;

- Technology for detecting diseases in the early or preclinical stages;

- collection, processing and storage of historical (static and dynamic) data of each user;

- for each user, an individualized calculation of cancer risk and recommendations for prevention and prophylaxis;

- web and mobile applications with an easy-to-use interface allow the uploading of breast and breast organ images (in the perspective of handwritten and typewritten text images);

- B2B partners can be connected to the apps through API gateways and share risk calculations and recommendations with physicians;

- A special place in the development of the product is given to the creation of a digital database of human biological markers for the safety of the genetic population in G-20 countries; - a competitive feature - the potential for scaling our project and entering foreign markets.

***Target consumer segments (markets) of the product, their volume, dynamics and development potential***:

Target consumer segments of computer vision in disease diagnosis:

1. The key consumer segment in the first stage, we see the private individuals (B2C), who thanks to the project will be able to get a second opinion about their health status based on their own accumulated data. As the core of this audience segment, we consider users 34 and older, with higher education and above-average income, interested in high technology and at the same time in metaphysics, life and death, concerned about their health and aging. Our first focus is G-20 countries. The solution will be commercialized through premium and freemium subscriptions. Users will be offered to get acquainted with the product for free for 30 days, and then activate a subscription. Each country will have its own price, depending on the development of the country and the income of its population. In G-20 countries, a subscription will start from $2 (USD) per month. Access to the solution is expected to be sold both directly and via AppStore and Google Play. Our forecast - about 10 000 downloads of the application in the first year after the commercial launch and about 1000 paying clients, with revenue of $2,000 per month only from this segment ($24,000 per year).

2. In parallel it is planned to offer the product to the sector of medical institutions, private and state clinics (B2B, B2B2C), which in recent years are increasingly focused on the introduction of new technological solutions to improve health and interest in mastering information in new ways. According to our estimates, only in Russia the solution can be purchased by more than 5,000 healthcare organizations with a total number of end users of more than 10 million people. Commercialization of the solution will be done based on individual pricing, depending on the number of computer vision models used for the analysis (the pathologies being studied) and the number of end users. Access to our own API-gateway to connect to our server is provided by selling subscriptions and individual orders for the development of additional models depending on the availability of data (images) which can be used to build accurate predictive models. Our calculation assumes cooperation with at least 10 organizations in the first year with more than 20,000 users and revenue of about $50,000 per month ($600,000 in the first year after commercial launch.

3. Similarly assume commercialization of the product in the segment of health centers, medical resorts, fitness centers, etc. (B2B2C). The potential market is about 1,000 organizations with over 1 million end users. We plan to sign at least 5 contracts in the first year for the total sum of $7,000 per month ($84,000 per year)

4. Pharmaceutical companies (B2B) will be able to develop new drug lines for early disease prevention based on predictive computer vision models. Individual pricing with such companies is proposed, as each contract implies individual development for the client's needs. We plan to sign at least 1 contract worth $15,000 in the first year of collaboration.

***Description of the business model and product promotion strategy:***

The business model is a system of subscriptions and direct contracts for B2B clients, which will be sold directly (without intermediaries) and through a mobile app for B2B and B2B2C clients through AppStore and Google Play (with a commission to these sites of 15-30%).

The cost of a monthly subscription for one client will be $2. According to our estimates, by the end of the first year of commercialization of the final solution the revenue will be at least $723,000 on sales of the product with several prediction models.

The cost of subscription is formed on the basis of the following indicators:

- the cost of developing each disease model;

- the cost of leasing the data warehouse;

- cost of querying and extracting data from databases; - cost of project management and personalization capabilities.

The product will be marketed through several channels:

1. Participation in profile events, thematic exhibitions and conferences to present the principal idea, functionality, and advantages of the developed product: - CES 2022 Health Protocols (January 5-8, 2022, USA, Las Vegas);

- Arab Health (January 24-27, 2022, UAE, Dubai);

- Ai4 2022 Healthcare Summit (February 23-24, 2022);

- The AAAI Conference on Artificial Intelligence (February 22-March 1, 2022, Canada, Vancouver);

- HIMSS Global Health Conference & Exhibition (March 14-18, 2022, Orlando, USA);

- ICAIMLHA 2022: International Conference on Artificial Intelligence and Machine Learning for Healthcare Applications (June 2-3, 2022, San Francisco, USA; November 18-19, 2022, Singapore);

- AI World Conference & Expo (June 2-3, 2022, San Francisco, USA);

- ODSC Europe Virtual Conference (June 8-10, 2022, Boston, San Francisco, Brazil, London, and India);

- Machine Learning Week (June 19-24, 2022, USA, Las Vegas and Germany, Munich);

- European Health Tech Innovation Week (2022, UK).

2. Targeted advertising in social networks and internet resources dedicated to high technology, health, medicine, etc. Creation of a community of the project in the main social media, promotion through opinion leaders.

3. Work with the media, including thematic scientific publications, to increase the project's visibility. By the end of the first year it is planned to receive not less than 50 mentions of the project in various editions, including international.

4. "Sarafanov radio" and controlled PR, feedback from loyal users, including on the sites AppStore and Google Play.

5. Interaction with state structures to ensure support for the project as part of the implementation of the national programs of G-20 countries regarding integration into the digital economy.

6. Creation of accounts on the main crowdfunding platforms: [www.kickstarter.com](http://www.kickstarter.com), [www.patreon.com](http://www.patreon.com), [www.gofundme.com](http://www.gofundme.com), [www.rockethub.com](http://www.rockethub.com), etc. Participation in some platforms will require the creation of a branch in another jurisdiction and/or the inclusion of foreign employees in the team.

7. Establishing direct contacts and forming partnerships with medical institutions (public and private), pharmaceutical companies.

Planned indicators:

- by the end of the first year: awareness of the project among technology experts - coverage of at least 25%, among the mass consumer - coverage of at least 3%. - By the end of the second year: market coverage - at least 7%.

- In the future we plan to take a place among the top 3 companies in the industry of disease diagnostics using computer vision.

**R&D Terms of Reference R&D Terms of Reference**

***The purpose of R&D work***

The purpose of R&D is to develop predictive models of computer vision for early diagnosis of lung, trachea, bronchial and mammary gland cancer.

During the first year of R&D, the goal is to develop a minimum viable product (MVP), a set of computer vision models for early disease diagnosis, viz:

- lung cancer diagnosis;

- diagnosis of tracheal cancer;

- diagnosis of bronchial cancer;

- diagnosis of breast cancer;

- Elaboration of a machine learning model for image recognition and diagnosis formulation;

- development of a method for obtaining scientifically based calculations of individual risks of cancer development from public medical databases and recommendations for preventing the risks found;

- development of a personal account of the user within the MVP functions;

- implementation of MVP interaction with the user (individual and medical institution) in the format of uploading images;

- testing the MVP functionality.

As part of the second year of R&D, it is planned to refine MVP functionality by developing additional predictive models for other diseases, including stomach cancer, skin cancer, etc. It is also planned to prepare the project for scaling and commercial implementation.

***Purpose of the scientific and technical product (product, etc.)***

The project is aimed at diagnosing four types of cancer at an early (or preclinical) stage using computer vision techniques, as well as creating a digital database of human biological markers accumulated over 3-5 years of each patient. The technology offers a medical decision support system, providing a "second opinion" based on data sets from around the world.

Areas of application of our project: medical (disease diagnosis, medical decision support system), health conservation, health security of genetic populations, increase in healthy life expectancy, educational, structured storage of personal information.

**Technical requirements to the scientific and technical product (prototype, prototype) to be developed within the current stage of R&D performance**

**The main technical parameters defining functional, quantitative (numerical) and qualitative characteristics of the scientific and technical product obtained as a result of the current R&D stage**

***Functions, which the developed scientific and technical product should provide***

The product under development must provide the following functions when answering "is or is not a disease in healthy or sick patients":

1. Training of computer vision models on large amounts of open data to obtain 80-90% accuracy of predictive predictions

2. Accumulating an array of historical data from a particular user (from closed sources, such as medical records) to further train the models and increase the accuracy of predictive predictions to 95-99%.

3. Obtaining calculations of future probability of disease.

**Quantitative parameters determining how the scientific and technical product performs its functions**

We assume that for a model built on open data (without using historical user data), the accuracy of responses about the presence or absence of disease will be 80-90%.

When building a predictive model based on historical user data (assuming this data is available for a period of 3 to 5 years), the accuracy of responses would be as high as 95-99%.

A proprietary knowledge base will be used to answer questions about the likelihood of future disease and the study of symptoms. It is not yet possible to estimate the accuracy of such answers. Based on the information that is stored in public medical databases and with access to closed data in medical institutions, accuracy will be close to 80%.

***The input influences necessary for the scientific and technical product to perform the given functions***

The method of diagnosing diseases based on computer vision implements several modes of interaction:

- Data collection

- Data exchange with a user (individual, medical institution)

- Execution of server commands

- Data exchange with public medical databases

Input impacts in data collection mode: an array of user data transmitted to train (train) machine learning models.

Input impacts in user data exchange mode: X-rays, photos or scanned copies of medical records, symptom descriptions, requests and commands.

Server command execution mode inputs: unified server commands.

Inputs in the mode of data exchange with public medical databases: questions posed by the server related to the identified diseases/probability of disease/symptoms.

Output reactions provided by the scientific and technical product as a result of performing its functions

Output reactions in the data collection mode: a populated digital database of human biological markers.

Output reactions in the mode of data exchange with the user: responses in the form of texts or images.

Output reactions in the mode of server commands execution: command execution.

Output reactions in the mode of data exchange with public medical databases: answers about risks and recommendations for the given questions.

**Design requirements to the scientific-technical product to be obtained as a result of the current stage of R&D**

***Design requirements and constituent parts of the scientific and technical product***

The product to be developed consists of:

- four predictive machine learning models;

- interface of communication with user (loading of images, output of models calculation results, probability of disease development, text of recommendations on development risk prevention, list of scientific publications)

- a module for image recognition and markup;

- a decision making module;

- two knowledge bases (user and scientific data);

- response generator.

***Type of design, product forms***

The first version of a minimally viable product - a mobile application and web platform, the interface of which is designed for interaction with both an individual user and a medical institution.

The final (commercial) version of the product is a platform that operates on the Premium and Freemium models and allows the collection and storage of historical user data in order to assess the likelihood of a user having diagnosable types of cancer in the future (lung, bronchial, tracheal and breast cancer) and to give scientifically based personal recommendations on the prevention and prevention of cancer.

**Other requirements for scientific and technical product (prototype, prototype) that shall be developed within the current R&D stage**

***Requirements to obtain patent protection***

It is planned to file a patent application for the company's invention - a system for early cancer diagnosis using computer vision, calculating the probability of cancer for a particular user and giving him individual recommendations for avoiding these risks.

***The list of the main categories of components and materials (included into the developed product (article) or used in the process of its development and manufacturing)***

Purchase of materials, components or licensed software at the expense of the Foundation is not expected for the project implementation.

***Reporting on R&D (list of technical documentation developed in the process of implementation of the current stage of R&D)***

Scientific and technical reports (intermediate and final) Software complex algorithms

Description of the software complex

Instructions for the user (patient, medical institution, pharmaceutical company)

Instructions for the system programmer

Software package testing methodology

Protocol of Software Package Testing

**SCHEDULE AND COST ESTIMATE**

**Timetable**

***Research and Development (R&D) Implementation Schedule. The 1st annual stage of the project:***

| №  stage | **Name of the stage of the calendar plan** | **Duration of stage, months** | **Cost, USD.** |
| --- | --- | --- | --- |
| 1 | 1. Data collection and analysis for early cancer diagnosis using computer vision techniques.  2. Development of a method for early diagnosis of cancer using computer vision, creation of predictive models for lung, bronchial, and tracheal cancers.  3. Creation of a personal user account within the MVP functions.  4. Implementation of MVP interaction with the user. Preparation of terms of reference for web platform and mobile application development. | 6,00 |  |
| 2 | 1. Creating a predictive model for breast cancer.  2. Creation of beta versions of web platform and mobile application for disease diagnostics.  3. testing of the finished product, improvement of models. | 6,00 |  |
|  | TOTAL: |  |  |

**Estimate**

***Estimate of the cost of implementing the project:***

| **No. n/a** | **Name of expense items:** |
| --- | --- |
| 1 | Wages and salaries |
| 2 | Payroll accrual |
| 3 | Data sets |
| 4 | Materials |
| 5 | Rental of cloud services |
| 6 | Payment for the work of co-executors and third-party organizations |
| 7 | Other general expenses |

| **No. n/a** | **Payment for the work of co-executors and third parties** |
| --- | --- |
| 1 | Partitioning the database of received images |
| 2 | Developing the functionality of the user's personal account |
| 3 | Developing the functionality of the web platform and mobile application |