



# The Open Bioinformatics Journal

Content list available at: <https://openbioinformaticsjournal.com>



## EDITORIAL

### Advances in Text and Data Mining of Biological Data: Models, Methods and Applications

Ivan Izonin<sup>1,\*</sup> and Sergii Babichev<sup>2,3</sup>

<sup>1</sup>Department of Artificial Intelligence, Lviv Polytechnic National University, S. Bandera str., 12, 79013, Lviv, Ukraine

<sup>2</sup>Jan Evangelista Purkyně University in Ústí nad Labem, Pasteurova 3632/15 400 96 Ústí nad Labem, Czech Republic

<sup>3</sup>Kherson State University, University Street, 27, 73003, Kherson, Ukraine

The development of biological systems over billions of years has made them very difficult to understand. Biologists and clinical scientists try to understand various biological processes using different tools. However, vast amounts of data for analysis, complex multi-parameter interconnections between the data of a particular set, and hidden relationships between them significantly affect its processing and analysis. The latest advances in Artificial Intelligence (AI), mainly text mining, data mining, artificial neural networks, fuzzy logic, machine learning, and others, can significantly improve the processing of such data. In particular, it creates potential opportunities for doing high-impact investigations that can solve real-world tasks in the system biology branch. The peculiarities of biological data are that it has different types, formats, structures, and huge volumes, which significantly complicates its processing and analysis. Such processing should include models, methods, and tools for efficient storage and retrieval of various types of data; an effective conversion and consolidation of the data of multiple formats; fast optimization and transfer; reliable intellectual analysis to obtain valuable information; as well as informative data visualizations for future visual analysis or better human perception. All this necessitates combining existing and developing new, faster, and precision AI techniques for future information discovery and knowledge engineering from such data.

This special issue is dedicated to covering up-to-date text- and data mining models, methods, and application of biological data processing and analysis to transform such data into useful information and knowledge.

This special issue covered only the best papers from the 3<sup>rd</sup> International Conference on Informatics & Data-Driven

\* Address correspondence to this author at the Department of Artificial Intelligence, Lviv Polytechnic National University, S. Bandera str., 12, 79013, Lviv, Ukraine; Tel: +38 098 888 96 87; E-mails: [ivanizonin@gmail.com](mailto:ivanizonin@gmail.com), [ivan.v.izonin@lpnu.ua](mailto:ivan.v.izonin@lpnu.ua)

Medicine (IDDM-2020) [1]. This conference is indexed by Scopus, Web of Science, and dblp databases. In addition, according to the CORE list of conferences 2021, it has rank C.

We have accepted only extended versions of the conference's papers, modified more than 70%, which contained science-intensive solutions that have a solid theoretical basis and demonstrate readiness for practical application in the field of Mining bioinformatics. The IDDM 2020 Program Committee recommended it based on the scientific novelty, reasonable value, and prospects for further research.

The main included themes in this Special Issue are the following:

- Database and data warehousing for biological data storage
- Biological data retrieval and optimization
- Data integration strategies in bioinformatics
- Machine learning for biomedical information extraction and analysis
- Ensemble-based methods for classification in bioinformatics
- Frequent patterns algorithms for biological data sequences
- Text mining methods in the biomedical domain
- Cluster analysis of biological datasets
- Graph theory for analysis of biological networks
- Gene regulatory networks reconstruction and simulation
- Information visualization techniques in bioinformatics
- Organization of health systems for catastrophes

The fifteen submissions from Germany, Egypt, Canada, and Ukraine were received. After the review process, only

eleven papers were accepted for publication.

The paper “Information System for Screening and Automation of Document Management in Oncological Clinics” [2] is dedicated to the problem of effective document management in a modern clinic. The effectiveness of medical services depends significantly on the effective solution to this problem. The authors have developed a specialized application that provides efficient interaction between distributed and heterogeneous applications and the devices. This approach has improved the efficiency of information flow processing for treating patients with breast cancer.

The paper “Development of a Genetic Method for X-Ray Images Analysis Based on a Neural Network Model” [3] discusses the problems of automatically determining the size of tumors in the human body. This task significantly affects the diagnosis and treatment of various diseases, including surgery. The authors proposed a new method of identification of abnormal formations in the lungs based on radiological images. To do this, the authors improved the procedure for selecting the deep convolutional neural network parameters, which provided:

- A significant reduction in its training.
- Reduced the complexity of calculations.
- Increased the accuracy of its work.

The article [4], “Unsupervised Clustering in Epidemiological Factor Analysis,” is devoted to solving the problems of early analysis of epidemiological data. The author shows that this problem is relevant in cases where the correlation between the factors for future data mining has not yet been established. In such cases, practical intellectual analysis is not possible. The author proposes an approach, which consists of applying the Principal Components Analysis to reduce the dimension of the problem space with the subsequent use of deep convolutional neural networks. It is shown that this approach justifies itself in the early stages of analysis.

The study “Modeling and Methods of Statistical Processing of a Vector Rhythmic Signal” [5] presented a new mathematical model of a vector rhythmic signal to analyze human heart rate. This task is relevant not only in the field of cardiology but also in various areas of biometrics. Also, the authors proposed new methods for statistical estimation of spectral-correlation characteristics of increased heart rate. Based on experimental modeling, the efficiency of using the proposed approach is shown.

The article [6] “Mathematical model of the process of ultrasonic wave propagation in a relax environment with its given profiles at three time moments” considers the problem of building mathematical models that are used in ultrasound diagnostics. The authors propose an equation that describes the process of ultrasonic oscillations propagation in a relaxed environment. The advantage of such a solution - the profile of the ultrasonic wave is known at any time and an arbitrary point of space.

The paper “An Approach to Early Diagnosis of Pneumonia on Individual Radiographs Based on the CNN Information Technology” [7] is dedicated to the problem of early detection

of pneumonia in X-rays. The authors have developed information technology to solve this problem. The authors used deep convolutional neural networks to classify radiographs with mild signs of pneumonia. To overcome the lack of observations for training, the authors used transfer training. Experimental modeling confirmed the high efficiency of using the developed information technology to solve the stated task.

Professor A.M. Saleh and his research team dealt with the problems of predicting blood infections. The issue of highly accurate prediction of sepsis is very relevant in various fields of medicine. In the paper [8] “Predicting Sepsis in the Intensive Care Unit (ICU) through Vital Signs using Support Vector Machine (SVM),” the authors proposed to use a feature selection algorithm and SVM-based classifier. Experimental modeling of this approach was based on using a set of medical data collected by the authors. The relatively high accuracy of the proposed approach in comparison with similar methods is established.

The study “Finite Element Calculation of the Linear Elasticity Problem For Biomaterials with Fractal Structure” [9] presented a new model and software tool for calculating the components of the stress-strain state of biomaterials with fractal structure. The authors used the finite element method with a piecewise linear basis to solve the constructed model. The proposed approach provides an opportunity to analyze the rheological behavior of biomaterials.

Professor Ye. Bodyanskiy and his research team solved the problem of medical diagnostics for small data sets with overlapping classes. This task is critical in various fields of medicine. In the paper [10] “Adaptive Probabilistic Neuro-Fuzzy System and Its Hybrid Learning in Medical Diagnostics Task,” the authors developed a hybrid neuro-fuzzy system based on a probabilistic neural network and adaptive neuro-fuzzy interference system. The developed method is relatively easy to implement, does not require lengthy training, and provides high classification accuracy.

The study “Complex automatic determination of morphological parameters for bone tissue in human paranasal sinuses” [11] presented a new complex automatic approach to determine morphological parameters of bone tissue in human paranasal sinuses. The authors collected a set of medical images. They also proposed a step-by-step algorithm for using several image processing techniques to obtain the desired result.

The paper “A Method for Assessing the Risks of Complications in Chemoradiation Treatment of Squamous Cell Carcinoma of the Head and Neck” [12] discusses the problems of squamous cell carcinoma of the head and neck. The authors proposed using RBF ANN to determine the effectiveness of chemoradiation treatment with cisplatin and 5-fluorouracil in the treatment of patients with this disease. The results of classification using a neural network of this type showed satisfactory results. It provided an opportunity for the authors to choose the methods of treatment of the above disease.

The proposed authors of this Special Issue methods, models and software tools provide original solutions for complex problems in Medicine and Biology, which have a high practical value and can be used in practice.

Guest Editors would like to express their sincere gratitude to the anonymous reviewers for their hard work. In addition, we would like to thank the entire team of the Open Bio-

informatics Journal for the opportunity to publish materials of Special Issue free of charge.

## REFERENCES

- [1] Shakhovska N, Campos J, Melnykova N, Izonin I. IDDM 2020: Proceedings of the 3<sup>rd</sup> international conference on informatics & data-driven medicine, växjö CEUR-WSorg 2020; 2753: 482. Available from: <http://ceur-ws.org/Vol-2753/>
- [2] Sheketa V, Pasiaka M, Chupakhina S, *et al.* Information system for screening and automation of document management in oncological clinics In: Open Bioinform J. 2021.
- [3] Fedorchenko I, Oliinyk A, Stepanenko A, Fedoronchak T, Kharchenko A, Goncharenko D. Development of a genetic method for X-Ray images analysis based on a neural network model. In: Open Bioinform J. 2021.
- [4] Dolgikh S. Unsupervised clustering in epidemiological factor analysis. In: Open Bioinform J. 2021.
- [5] Lytvynenko I, Lupenko S, Onyskiv P, Zozulia A. Modeling and methods of statistical processing of a vector rhythmocardiogram. In: Open Bioinform J. 2021.
- [6] Nytrebych Z, Il'kiv V, Malanchuk O. Mathematical model of the process of ultrasonic wave propagation in a relax environment with its given profiles at three time moments. In: Open Bioinform J. 2021.
- [7] Radiuk P, Barmak O, Krak I. An approach to early diagnosis of pneumonia on individual radiographs based on the CNN information technology. In: Open Bioinform J. 2021.
- [8] Rayan Z, Alfonse M, Salem AM. Predicting Sepsis in the Intensive Care Unit (ICU) through Vital Signs using Support Vector Machine (SVM) In: Open Bioinform J. 2021.
- [9] Shymanskyi V, Sokolovskyy Y. Finite element calculation of the linear elasticity problem for biomaterials with fractal structure. In: Open Bioinform J. 2021.
- [10] Bodyanskiy Ye, Pliss I, Deineko A, Chala O. Adaptive probabilistic neuro-fuzzy system and its hybrid learning in medical diagnostics task. In: Open Bioinform J. 2021.
- [11] Nechyporenko AS, Radutny R, Alekseeva VV, Titova G, Gargin VV. Complex automatic determination of morphological parameters for bone tissue in human paranasal sinuses. In: Open Bioinform J. 2021.
- [12] Starenkiy V, Artiukh S, Ugryumov M, Strilets V, Chernysh S, Chumachenko D. A method for assessing the risks of complications in chemoradiation treatment of squamous cell carcinoma of the head and neck. In: Open Bioinform J. 2021.

© 2021 Izonin and Babichev

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.