Big Data Analytics of Inpatients Flow with Diabetes Mellitus type 1

Revealing new awareness with Advanced Visualization of Medical Information System Data

Olga Kolesnichenko, Elena Marochkina, Roman Komarov
Institute of Social Sciences,
Clinic of Aortic and Cardiovascular Surgery of
I.M. Sechenov First Moscow State Medical University
Moscow, Russia
kolesnichenko.oy@1msmu.ru, doc-lena45@mail.ru,
komarovroman@rambler.ru

Lev Mazelis, Andrey Mazelis, Daniel Soldatov
Department of Mathematics and Modeling
Vladivostok State University of Economics and Service
Vladivostok, Russia
Lev.Mazelis, Andrey.Mazelis@vvsu.ru,
da.soldatov@mail.ru

Larisa Minushkina
Department of Therapy, Cardiology, and Nephrology
CS Medical Academy at the Dep. of Presidential Affairs
Moscow, Russia
Minushkina@mail.ru

Mikhail Chernoskutov, Vladimir Averbukh, Igor Mikhaylov
Institute of Natural Sciences and Mathematics of
Ural Federal University, Computer Visualization
Department of N.N. Krasovskii Institute of Mathematics and Mechanics, RAS Urals Branch
Yekaterinburg, Russia
chernoskutovma@gmail.com, Averbukh@imm.uran.ru,
igormich88@gmail.com

Alexander Martynov, Valeriy Pulit
SP.ARM
St. Petersburg, Russia
Martynov, Valeriy.Pulit@sparm.com

Yuriy Kolesnichenko
Uzgraph
Moscow, Russia
green-apple_2000@mtu-net.ru

Sergey Amelkin, Ivan Grigorevsky
A.K. Aylamazyan Program Systems Institute of RAS
National Supercomputing Technology Platform
Pereslavl Zalessky, Russia
sergey.a.amelkin@gmail.com, gin@nscf.ru

***Abstract* — Big Data Analytics with Advanced Data Visualization of Medical information system qMS records is presented. The inpatients with Diabetes Mellitus type 1 were chosen for analysis. The various methods of analysis and visualization were implemented: Gray reflected binary code (Java), Cluster analysis (iPython), Graph analysis (iPython, Gephi), 3D Visualization (Java), as well as supercomputer "Uran" was used. The connected pathogenetic Continuum of Diabetes Mellitus type 1 was built. The Continuum of Diabetes Mellitus type 1 progression allows assume that Parathyroid hormone-related protein (PTHrP) plays the critical role in multi-organ pathogenetic cascade in this disease, including the development of Lung cancer. Based on our study we suggest considering PTHrP in terms of pharmacological treatment. Big Data Analytics including Cluster and Graph Analysis of Medical information system's data flow can be used to study pathogenesis of the disease and for new drugs creation proposal.**

Keywords — Cluster Analysis; Graph Analysis; Advanced Data Visualization; Big Data Analytics; Diabetes Mellitus type 1; PTHrP

I. INTRODUCTION. DATA BASED MEDICINE

The new era of knowledge and data synthesis has become into medicine. Medical information systems (MIS), Big Data Analytics and Advanced Data Visualization (ADV) help to reveal hidden accents in inpatients management. These hidden accents are not visible at small groups level, until we look at an entire flow of patients. Knowledge synthesis, the integration of different data into a single picture allow to create a connected continuum of the disease. Big Data Analytics approach or Data based approach can be used to study pathogenesis of disease and to search targets for treatment as well as for new drugs creation proposal. As a whole Big Data Analytics of patients flow shows a socio-medical portrait of patient with specific disease. This socio-medical investigation helps to highlight within patients flow the risk groups, such as precancer group, that give new opportunity in cancer prevention and cancer early detection in population.

978-1-5386-5933-5/19/$31.00 ©2019 IEEE

II. Research Methods

A. Medical information system qMS

Internal hospital network (Intranet) or MIS qMS (SP.ARM) collects patients data flow during 2013-2017 years from several Russian hospitals. MIS qMS is the Cloud-based network which brings Big Data into common Data Center and allows analyzing data on SaaS basis (Software as a Service). The patients with Diabetes Mellitus type 1 were chosen for Big Data Analytics and Advanced Data Visualization. Analyzed MIS qMS records include codes of ICD-10 (10th revision of the International Statistical Classification of Diseases and Related Health Problems, World Health Organization), information about patients age, length of stay in hospital, each investigation procedure or surgery, and waiting time of each investigation procedure or surgery. Personal data about patients (304 men and 558 women) was not transferred to research group. Metadata table was created in Microsoft Excel.

B. Gray reflected binary code (Boolean algebra)

The Gray reflected binary code (Boolean algebra) in Java was used for search unique matches in the data matrix. All metadata with 141 table's columns were grouped into 8 subgroups, which we called defining points – 8 united columns of table with data for each of 862 patients. Thus 8 united columns have $2^8 = 256$ variants of coincidence or non-coincidence of data for each patient. All found combinations of 2, 3 and more defining points were analyzed. Weighted non-directed graph was drawn with layer-by-layer application in Paint.net. Gray code analysis was made by Yu. Kolesnichenko (Uzgraph).

C. Cluster Analysis

Cluster analysis was implemented in iPython (libraries NumPy, Pandas, Sklearn) using Intel Xeon Servers Network (18,500 cores). Hierarchical and k-means methods were performed. Cluster analysis includes only investigation procedures or surgery for each of 862 patients. Five clusters were determined. Cluster analysis was made by L. Mazelis, A. Mazelis, and D. Soldatov (Vladivostok State University of Economics and Service).

D. Graph Analysis

Graph analysis was implemented in iPython (libraries NetworkX, Pandas, xlrd) and in Gephi software in Java, using supercomputer "Uran" of the Krasovskii Institute of Mathematics and Mechanics. Graph analysis was made by M. Chernoskutov (Institute of Natural Sciences and Mathematics of Ural Federal University, and Krasovskii Institute of Mathematics and Mechanics).

E. 3D Visualization

3D Visualization of patients data flow was created in Java, OpenGL, GLSL, using supercomputer "Uran". The 3D field includes 3 coordinate axes and different shape & color of the markers. This interactive platform has unlimited number of options, which can be changed by user. 3D Visualization software was made by V. Averbukh and I. Mikhaylov (Computer Visualization Department of N.N. Krasovskii Institute of Mathematics and Mechanics).

III. Results of ADV and Discussion

A. Gray binary code match

The combined triple coincidence of defining points was found: diagnosis of bronchial and pulmonary diseases, diagnosis of gastrointestinal tract diseases, diagnosis of osteoporosis (Fig. 1). This triple coincidence indicates the workload on hospital for patients flow with Diabetes Mellitus type 1 and topicality of these three groups of complications.

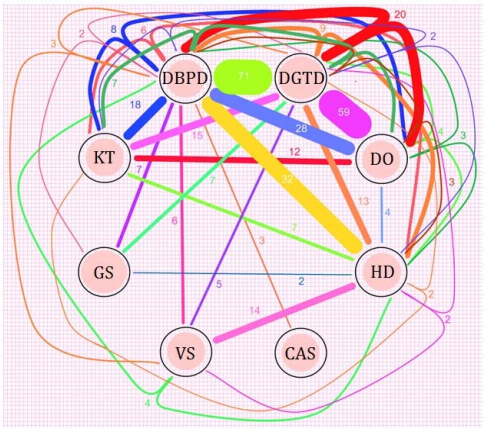

Fig. 1. Visualization of Gray binary code combinations of 8 defining points; first was published in the one of our study articles [1]. Abbreviations: DBPD – diagnosis of bronchial and pulmonary diseases; DGTD – diagnosis of gastrointestinal tract diseases; DO – diagnosis of osteoporosis; HD – hemodialysis; CAS – coronary artery surgery; VS – vascular surgery; GS – gastrointestinal surgery; KT – kidney transplantation.

B. Cluster Analysis

Five clusters were analyzed in terms of distribution of glycated hemoglobin (HbA1) and parathyroid hormone (PTH) blood levels between clusters. Pathogenetic Continuum of Diabetes Mellitus type 1 was built from clusters sequence (Fig. 2). Clusters are indicated by colors.

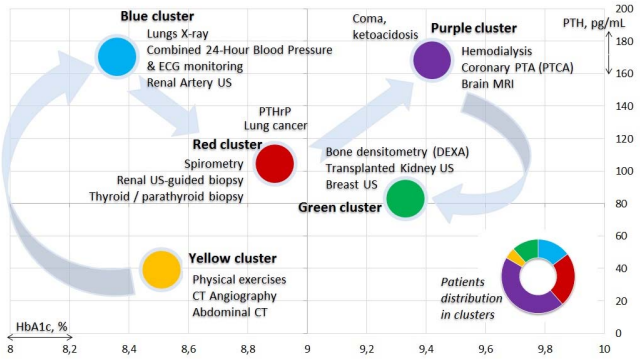

Fig. 2. Pathogenetic Continuum of Diabetes Mellitus type 1 based on Cluster analysis of patients flow; first was published in the one of our study articles [2]. PTHrP – Parathyroid hormone-related protein. Other abbreviations see below in Table I.

The high incidence of Lung cancer was detected in this patients flow (1277 cases per 100,000 patients with Diabetes Mellitus type 1). The most cases of Lung cancer are located in "Red" cluster.

C. Graph Analysis

Graph analysis was made for each of five clusters (Fig. 3, Fig. 4, Fig. 5, Fig. 6, Fig. 7). The points (vertices) and numbers counter-clockwise indicate each investigation procedure or surgery. Graph edges indicate each patient who underwent examination or surgery. Full list of 141 investigation procedures and surgical operations of metadata is shown in Table I. Graph analysis reflects the differences between clusters, which were identified as distribution of procedures & operations to the number of patients in cluster and to the number of matched type of procedures & operations in the entire sample of patients.

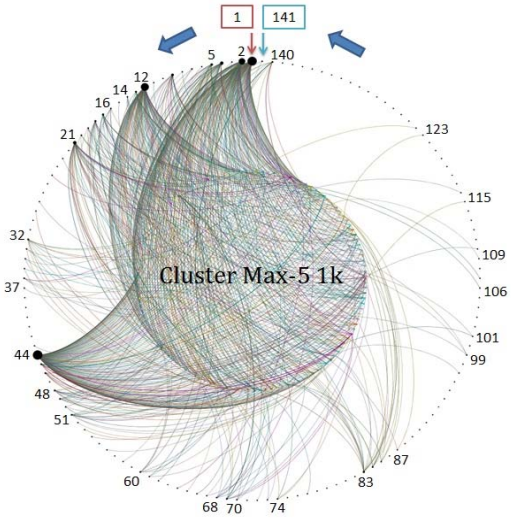

Fig. 3. Graph analysis of data from "Blue" cluster (Max-5 1k).

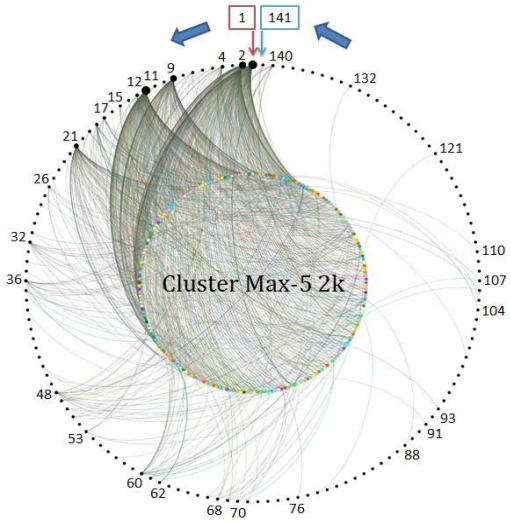

Fig. 4. Graph analysis of data from "Red" cluster (Max-5 2k).

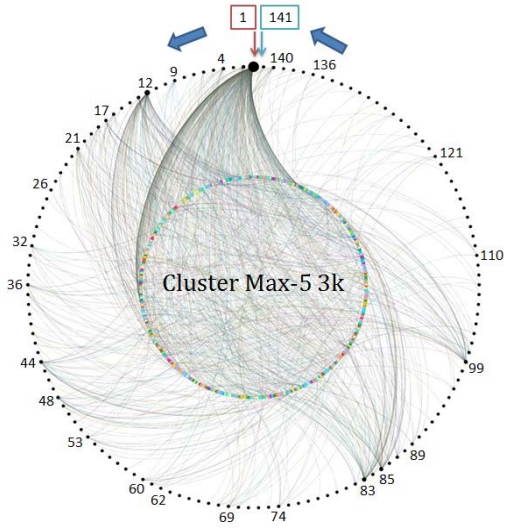

Fig. 5. Graph analysis of data from "Purple" cluster (Max-5 3k).

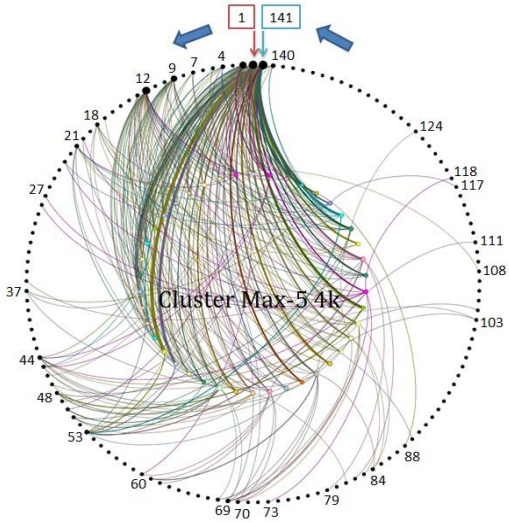

Fig. 6. Graph analysis of data from "Yellow" cluster (Max-5 4k).

Cluster and Graph Analysis revealed *the five models of Diabetes Mellitus type 1*. These models are as follows in the Continuum sequence: Active, predominantly with the course of physical exercises ("Yellow" cluster); Cardiopulmonary, reactive, with strong increase of PTH level ("Blue" cluster); Thyroid and with high risk of Lung cancer ("Red" cluster); Surgical, terminal, with hemodialysis ("Purple" cluster); Densitometric, with insufficient glycemic control ("Green" cluster). Dividing of data into the sequential chain helped us thinking about why Lung cancer occurs very often in patients with Diabetes Mellitus type 1. During development of Diabetes Mellitus hyperglycemia leads to kidney damage [3], [4]; Diabetic Nephropathy and hypocalcemia induce hypersecretion of PTH [5], [6], ("Blue" cluster reflects it). Increased PTH level alongside with Renal failure (uremia and metabolic acidosis) suppress the Parathyroid hormone-related protein (PTHrP) binding to the receptor PTH1R and decrease the

protective, repair effects of PTHrP, which is the most secreted in Lung, in alveoli epithelial cells [7], [8], [9], [10], ("Red" cluster reflects it). PTHrP signaling pathways play the critical role in cancer prevention [11], [12]. Progression of Diabetes Mellitus is accompanied by activation of calcium-sensing receptor (CaSR), which regulates release of PTH with calcium leaching from the bones (Diabetoporosis or Diabetic Osteodystrophy) [13]. Activation of CaSR also leads to decrease the protective effects of PTHrP in Lung, as well as CaSR suppresses PTHrP secretion in normal mammary epithelial cells with promoting breast tissue damage [14], ("Red" cluster, "Green" cluster reflect it), see Fig. 2, Fig. 4.

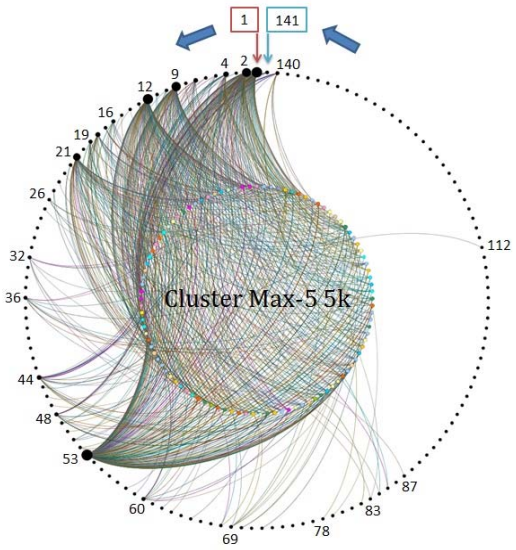

Fig. 7. Graph analysis of data from "Green" cluster (Max-5 5k).

D. 3D Visualization

3D Visualization interactive platform shows the patients data flow for each separate investigation procedure or surgery (Fig. 8, Fig. 9, Fig. 10, Fig. 11, Fig. 12, Fig. 13). Three coordinate axes correspond to length of stay in hospital in days (blue numbers), waiting time of investigation procedure or surgery in days (green numbers), patients age (red numbers).

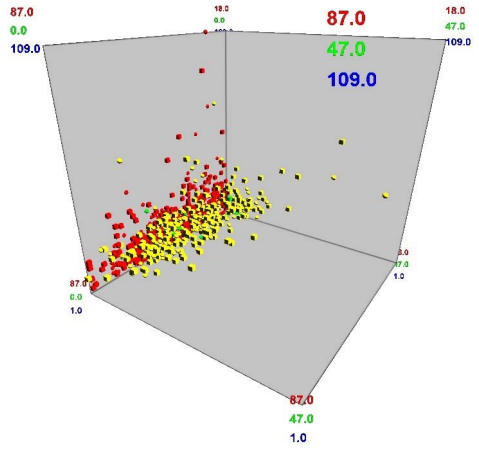

Fig. 8. Echocardiography. Marker-cube – women, marker-ball – men. Red marker – 0, yellow – 1, green – 2 investigation procedures.

TABLE I. INVESTIGATION PROCEDURES AND SURGICAL OPERATIONS FULL LIST OF METADATA

N	Medical information system qMS records	N	Medical information system qMS records
1	Electrocardiography, ECG	41	Lower extremity aa. ANG
2	Echocardiography	42	Intravenous pyelogram
3	Stress Echocardiography	43	Phlebography
4	24-Hour Holter monitoring	44	Lungs X-ray
5	Combined 24-Hour Blood Pressure & ECG monitoring	45	Gastrointestinal tract X-ray
6	Ambulatory Blood Pressure Monitoring (ABPM)	46	Abdominal X-ray
7	Bicycle Ergometry Test	47	Mammography
8	Treadmill Stress Test	48	Extremities X-ray
9	Thyroid US	49	Spine X-ray
10	Thyroid US-guided biopsy	50	Chest X-ray
11	Thyroid / parathyroid biopsy	51	Skull X-ray
12	Abdominal / Renal US	52	Orthopantomography
13	Kidneys, adrenal glands US	53	Bone densitometry (DEXA)
14	Renal Artery US	54	Bone Scintigraphy
15	Renal US-guided biopsy	55	Dynamic renal Scintigraphy
16	Transplanted Kidney US	56	Liver Scintigraphy
17	Abdominal US	57	Parathyroid Scintigraphy
18	Pelvic Floor US	58	Lungs Scintigraphy
19	Breast US	59	Body plethysmography
20	Pleurae US	60	Spirometry
21	Esophagogastroduodeno-scopy (EGD)	61	Bronchoscopy
22	Helicobacter pylori testing during EGD	62	Lungs & bronchi biopsy
23	Fibrocolonoscopy	63	Bone marrow aspiration biopsy
24	EGD with biopsy	64	Bone marrow trepanobiopsy
25	Esophageal biopsy	65	Audiometry
26	Gastric tissue biopsy	66	Electroencephalography
27	Sigmoid biopsy	67	Brain CT
28	Rectal biopsy	68	Chest CT
29	Esophageal endoscopic dilatation	69	Abdominal CT
30	Diagnostic Laparoscopy	70	CT Angiography
31	Abdominal aorta US	71	Eye sockets (orbits) CT
32	Brachiocephalic aa. US	72	Neck CT
33	Jugular, subclavian vv. US	73	Pelvic CT
34	Upper extremity vessels US	74	Brain MRI
35	Iliac arteries US	75	MRI of cerebral aa.
36	Lower extremity vessels US	76	MRI of carotid aa.
37	Coronary ANG	77	Abdominal MRI
38	Renal Arteriography	78	Spine MRI
39	Cerebral ANG	79	Pelvic MRI
40	Aortic, peripheral aa. ANG	80	Joints MRI

Continuation of Table I

N	Medical information system qMS records	N	Medical information system qMS records
81	Lungs SPECT	111	Hemihepatectomy
82	Screening examination of fetus	112	Laparoscopic surgery
83	Hemodialysis	113	Resection of the stomach
84	Central Venous Catheterization (Subclavian)	114	Endoscopic bleeding stop in gastrointestinal tract
85	Native arteriovenous fistula surgery	115	Intestinal obstruction surgery
86	Prosthetic arteriovenous fistula graft	116	Abdominal abscess surgery
87	Peritoneal Dialysis	117	Colonectomy
88	Coronary bypass surgery	118	Rectal resection
89	Coronary PTA (PTCA)	119	Rectal extirpation
90	Coronary thrombosis removal	120	Hemorrhoids surgery
91	Carotid PTA	121	Hernia surgery
92	Subclavian PTA	122	Percutaneous nephrolithotripsy
93	Aortic bifemoral bypass	123	Kidney transplantation
94	PTA of superficial femoral artery	124	Nephrectomy
95	PTA of lower extremity aa. without stenting	125	Foot necrosis surgery
96	PTA of lower extremity aa. with stenting	126	Hip Endoprosthetics
97	Ureteral stent placement	127	Thyroidectomy
98	PTA of other aa.	128	Thyroid resection
99	Venous Arterialization	129	Thyroid tumor surgery
100	Arterial thrombosis removal	130	Parathyroid tumor surgery
101	Venous thrombosis removal	131	Parathyroidectomy
102	Phlebectomy in lower extremity	132	Mastectomy
103	Pleural puncture	133	Mammary gland resection
104	Lungs surgery	134	Skin tumor surgery
105	Exploratory thoracotomy	135	Lymphadenectomy
106	Endoscopic APC of tumor in Lung tissue	136	Vitrectomy
107	Endobronchial photo-dynamic therapy of tumor	137	Retinal detachment surgery
108	Intestinal anastomosis	138	Endovitreal Surgery with Vitreal Cavity Tamponade
109	Choledocho-duodeno anastomosis	139	Phacoemulsification
110	Cholecystectomy	140	Physiotherapy
		141	Physical exercises

Abbreviations: US – Ultrasound; vv. – veins; aa. – arteries; ANG – Angiography; X-ray – Radiography; DEXA – dual-energy X-ray absorptiometry; CT – Computed Tomography; MRI – Magnetic Resonance Imaging; SPECT – Single photon emission computed tomography; PTA – Percutaneous transluminal balloon angioplasty and stenting; APC – Argon plasma coagulation.

3D Visualization of patients data flow helps to plan the hospital time workload. We can see that Echocardiography, Thyroid Ultrasound, and Bone densitometry are widespread among patients with Diabetes Mellitus type 1; Bone densitometry was carried out mainly for young patients, and Coronary Angiography was carried out mainly for elderly patients. The situation was different for patients with transplanted kidney, they are few with different ages, and they didn't stay in hospital for a long time. This type of Big Data visualization reveals the socio-medical differences of patients, helping to plan inpatient & outpatient time management.

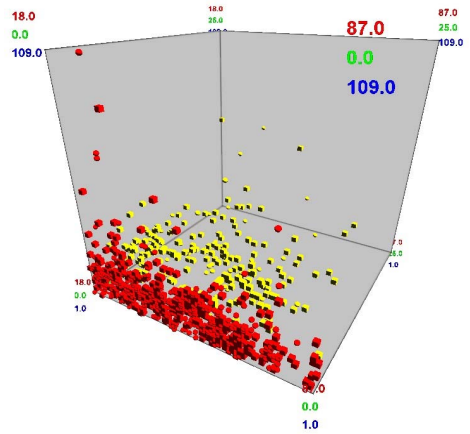

Fig. 9. Thyroid Ultrasound. Marker-cube – women, marker-ball – men. Red marker – 0, yellow – 1, green – 2 investigation procedures.

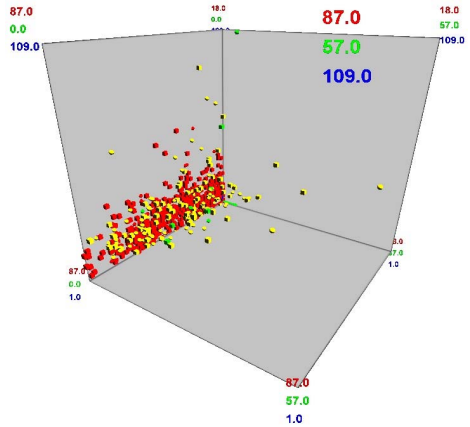

Fig. 10. Lungs Radiography. Marker-cube – women, marker-ball – men. Red marker – 0, yellow – 1, green – 2 investigation procedures.

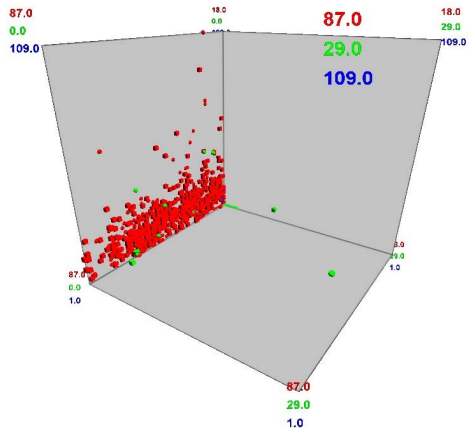

Fig. 11. Transplanted Kidney Ultrasound. Marker-cube – women, marker-ball – men. Red marker – 0, green – 1 investigation procedure.

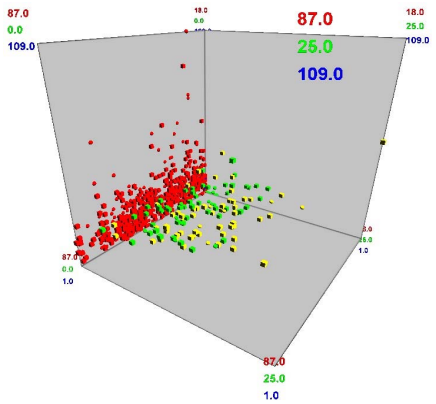

Fig. 12. Bone densitometry (dual-energy X-ray absorptiometry, DEXA). Marker-cube – women, marker-ball – men. Red marker – 0, yellow – 1, green – 2 investigation procedures.

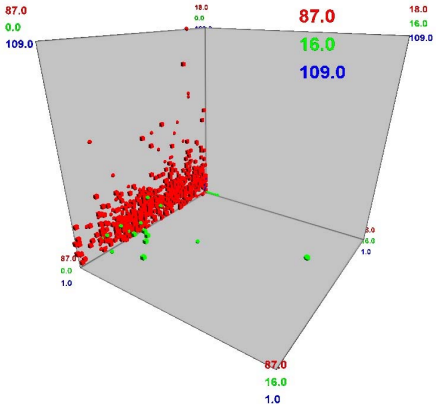

Fig. 13. Coronary Angiography. Marker-cube – women, marker-ball – men. Red marker – 0, green – 1 investigation procedure.

IV. Conclusion

Thus in our study the varied ADV of inpatient flow is shown. We can state that Big Data Analytics & ADV of MIS records can reveal new groups of patients and can give new awareness about disease development and complications causes. This study correlates with recent published results of the Cluster analysis of patients with Diabetes Mellitus (types 1 & 2) by Lund University (Sweden) [15], [16]. The difference is that in our study the patients with only Diabetes Mellitus types 1 were studied. The connected Continuum of Diabetes Mellitus type 1 progression, which was created in our Big Data Analytics study, allows assume that PTHrP plays the critical role in multi-organ pathogenetic cascade in this disease, including the development of Lung cancer. PTHrP has already appeared as approved drug, for example subcutaneous injections of Abaloparatide treat osteoporosis in postmenopausal women; as well as the anti-PTHrP antibody has been investigated for treatment of humoral hypercalcemia of malignancy [17]. Based on our study we suggest to deeply elaborate the role of PTHrP in Diabetes Mellitus as well as to consider PTHrP in terms of pharmacological treatment. The study is carried out under the National Supercomputer Technology Platform and Project 5-100 (5top100.ru).

References

[1] O.Yu. Kolesnichenko, Yu.Yu. Kolesnichenko, L.O. Minushkina, G.N. Smorodin, A.V. Martynov, V.V. Pulit, and A.N. Dolzhenkov, "The possibility of Gray Binary Code application for Big Data Analytics: Medical Information System records of patients with Diabetes Mellitus type 1," in Remedium J., vol. 10, pp. 38-47, 2017.

[2] L.O. Minushkina, O.Yu. Kolesnichenko, A.L. Mazelis, L.S. Mazelis, D.A. Soldatov, Yu.Yu. Kolesnichenko, A.V. Martynov, V.V. Pulit, A.N. Dolzhenkov, and I.N. Grigorevsky, "Medical information system's Big Data Analytics based on Cluster analysis: patients with Diabetes Mellitus type 1," in Remedium J., vol. 7-8, pp. 30-40, 2018.

[3] Endocrinology. National Manual, I.I. Dedov and G.A. Melnichenko ed., Moscow, Russia: Publishing House GEOTAR-Media, 2013, p. 752.

[4] V.V. Serov and M.A. Paltsev, Kidney and Arterial Hypertension. Moscow, Russia: Publishing House Medicine, 1993, p. 256.

[5] E.A. Hantakova, L.Yu. Hamnueva, and G.M. Orlova, "Analysis of phosphorus-calcium metabolism and parathyroid function in patients with Diabetes Mellitus type 1," in Diabetes Mellitus J., vol. 4, pp. 33-37, 2013.

[6] A.M. Mkrtumyan, Osteoporosis is an undetected disease in Diabetes Mellitus, Open lecture, Moscow City Scientific Society of Therapists, I.M. Sechenov First Moscow State Medical University, Moscow, Russia, 2018.

[7] M.L. Picton, P.R. Moore, E.B. Mawer, D. Houghton, A.J. Freemont, A.J. Hutchison, R. Gokal, and J.A. Hoyland, "Down-regulation of human osteoblast PTH/PTHrP receptor mRNA in end-stage renal failure," in Kidney International J., vol. 58 (4), pp. 1440-1449, 2000.

[8] P. Urena, M. Mannstadt, M. Hruby, A. Ferreira, G.V. Segre, and T. Drueke, "Down-regulation of the PTH/PTHrP receptor in uremia," in J. of Bone and Mineral Metabolism, vol. 12 (1), S87–S90, 1994.

[9] S. Disthabanchong, K.J. Martin, C.L. McConkey, and E.A. Gonzalez, "Metabolic acidosis up-regulates PTH/PTHrP receptors in UMR 106-01 osteoblast-like cells," in Kidney International J., vol. 62 (4), pp. 1171-1177, 2002.

[10] R.H. Hastings and L.J. Deftos, "Parathyroid Hormone-Related Protein and Lung Injury," in Chest J., vol. 122 (4), pp. 1494-1495, 2002.

[11] P.R. Montgrain, J. Phun, R.V. Werff, R.A. Quintana, A.J. Davani, and R.H. Hastings, "Parathyroid-hormone-related protein signaling mechanisms in lung carcinoma growth inhibition," in Springer Plus J., vol. 4, pp. 268, 2015.

[12] R.H. Hastings, P.R. Montgrain, R. Quintana, Y. Rascon, L.J. Deftos, and E. Healy, "Cell cycle actions of parathyroid hormone-related protein in non-small cell lung carcinoma," in The American J. of Physiology - Lung Cellular and Molecular Physiology, vol. 297 (4), L578-L585, 2009.

[13] A.Yu. Babenko, V.S. Nikitin, and T.L. Karonova, "Diabetes and bone," in Medical Council J., vol. 17, pp. 108-113, 2015.

[14] W. Kim and J.J. Wysolmerski, "Calcium-Sensing Receptor in Breast Physiology and Cancer," in Frontiers in Physiology J., vol. 7, Article 440, pp. 1-11, 2016.

[15] E. Ahlqvist, P. Storm, A. Karajamaki, M. Martinell, M. Dorkhan, A. Carlsson, P. Vikman, R.B. Prasad, D.M. Aly, P. Almgren, Y. Wessman, N. Shaat, P. Spegel, H. Mulder, E. Lindholm, O. Melander, O. Hansson, U. Malmqvist, A. Lernmark, K. Lahti, T. Forsen, T. Tuomi, A.H. Rosengren, and L. Groop, "Novel subgroups of adult-onset Diabetes and their association with outcomes: a data-driven Cluster analysis of six variables," in The Lancet Diabetes & Endocrinology, vol. 6 (5), pp. 361-369, 2018.

[16] E. Ahlqvist and L. Groop, "Paradigm shift in the diagnosis of Diabetes," Lund University, Available: https://www.lunduniversity.lu.se/article/paradigm-shift-in-the-diagnosis-of-diabetes.

[17] K. Sato, E. Onuma, R.C. Yocum, and E. Ogata, "Treatment of malignancy-associated hypercalcemia and cachexia with humanized anti-parathyroid hormone-related protein antibody," in Seminars in Oncology J., vol. 30 (Suppl. 16), pp. 167-173, 2003.