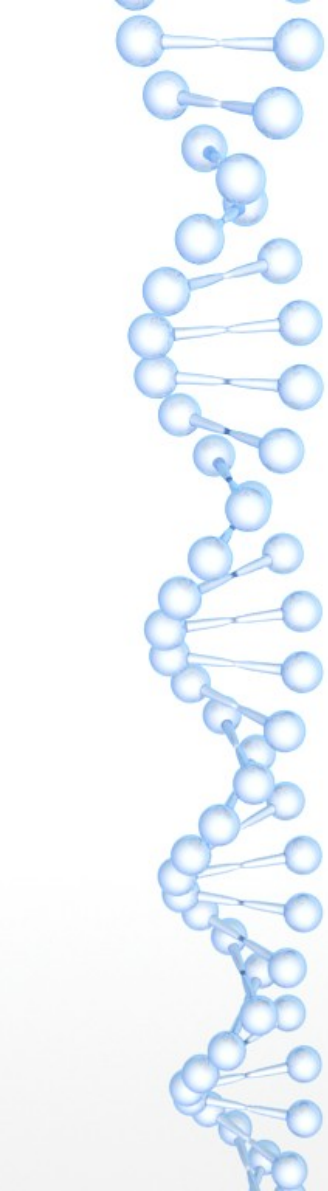


DATA FUSION FOR BETTER DECISION MAKING

PhD student JAROSLAVA ARSENJEVA

Review for 2019/2020 study year

Thesis advisor: Viktor Medvedev, Dr.

Thesis consultant: Gintautas Dzemyda, Prof. Habil. Dr.

Informatics engineering field (T007)

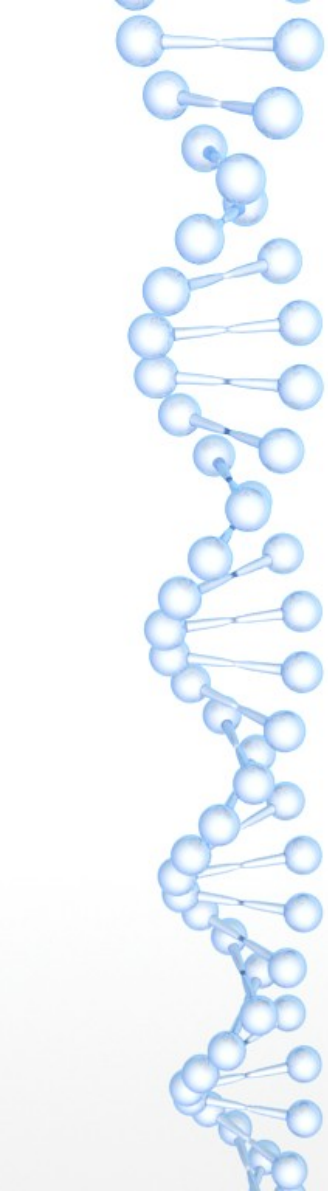
The time of doctorate studies: 2018 – 2022 years

- **Object of research:**

  Data fusion methods.

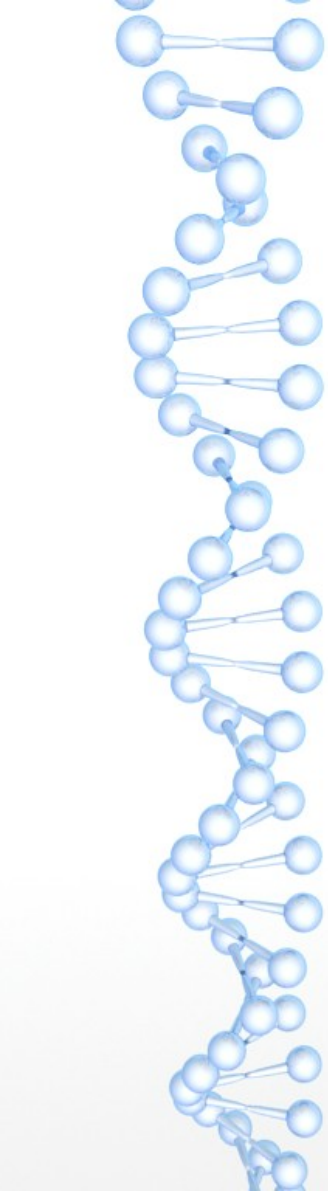
- **Research goal:**

  To propose a data fusion algorithm (framework) to improve decision making in medicine using data mining methods.



Research tasks

- Perform an analytical review of data fusion existing methods;
- Select appropriate data for further research and data fusion method application;
- Identify main scientific problems (for example, in the medical field) in data fusion tasks (where data from different sensors is aggregated into one framework);
- Select appropriate data fusion methods for previously selected data;
- Develop a data fusion algorithm to improve decision making that would be applied to previously selected data;
- Evaluate the results of the proposed algorithm, make necessary changes.



Planned results

- Identification of proper data fusion methods to improve decision making
- Proposition and further application of a data fusion algorithm (framework), that would improve decision making using data obtained from mining techniques.

Work plan for J.Arsenjeva

	Planned tasks	Comments	1st Year		2nd Year		3rd Year		4th Year	
			1st S	2nd S	1st S	2nd S	1st S	2nd S	1st S	2nd S
Conferences	Participate in a summer/ winter school (3 credits)	TD1403 (26-30.11, Finland)								
	Participate in an international conference	BOS/SOR (24-26.09, Poland) → CIST2021 (30-31.03, Portugal)								
	Participate in a national conference	DAMSS-2019 (November 28-30)								
	Participate in an international conference	RCIS 2021 (May 12-14, Cyprus)								
Papers	Publication in a periodical peer-reviewed scientific journal	CIST2021 (March 30-31, Portugal)								
	Publication in a journal in the Clarivate Analytics Web of Science									
	Publication in a journal in the Clarivate Analytics Web of Science									
Classes	Sprendimų priėmimo strategijos	7 ECTS								
	Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika	8 ECTS								
	Mašininis mokymasis	7 ECTS								
	Fundamentalieji informatikos ir informatikos inžinerijos metodai	8 ECTS								
	LaTeX	1.25 ECTS								
Research stages	Perform an analytical review of data fusion methods									
	Identify scientific problems arising in the tasks related to data fusion									
	Compilation of research methodology for solving data fusion tasks									
	Planning of theoretical and empirical research					+	+			
	Research of data fusion algorithms applications					+	+			
	Improvement of a data fusion algorithm for decision making									
	Investigation of the developed algorithm by solving test problems									
	Identification of application areas of the developed algorithm									
	Analysis of the obtained results, summary, inference preparation									
	Preparation of separate parts of the doctoral dissertation (research methodology, results, defended statements, conclusions)									
	Preparation and discussion of the PhD thesis in the institute									
	Defense of PhD thesis									

Planned task

Completed task

Work plan tasks for 2019/2020

- Create a deep overview of data fusion problems and their solutions in the medical field.
- Select a research methodology for solving the chosen task.
- Investigate data fusion methods used in data research.
- Identify additional scientific issues arising.
- Dissertation preparation stage: conducting research.
- Participate in an international conference in September.
- Pass 2 exams:
  - Machine learning (7 ECTS)
  - Fundamental Methods in Informatics and Computer Engineering (8 ECTS)

Completed tasks in 2019/2020

- Create a deep overview of data fusion problems and their solutions in the medical field.
- Select a research methodology for solving the chosen task.
- Investigate data fusion methods used in data research.
- Identify additional scientific issues arising.
- Dissertation preparation stage: conducting research.
- Participate in a course for LaTeX.
- Pass 2 exams:
  - Machine learning (7 ECTS)
  - Fundamental Methods in Informatics and Computer Engineering (8 ECTS)

Cognitive analysis of the Covid-10 related stress impact on population consumption decisions and forecast of its impact on the Lithuanian economy

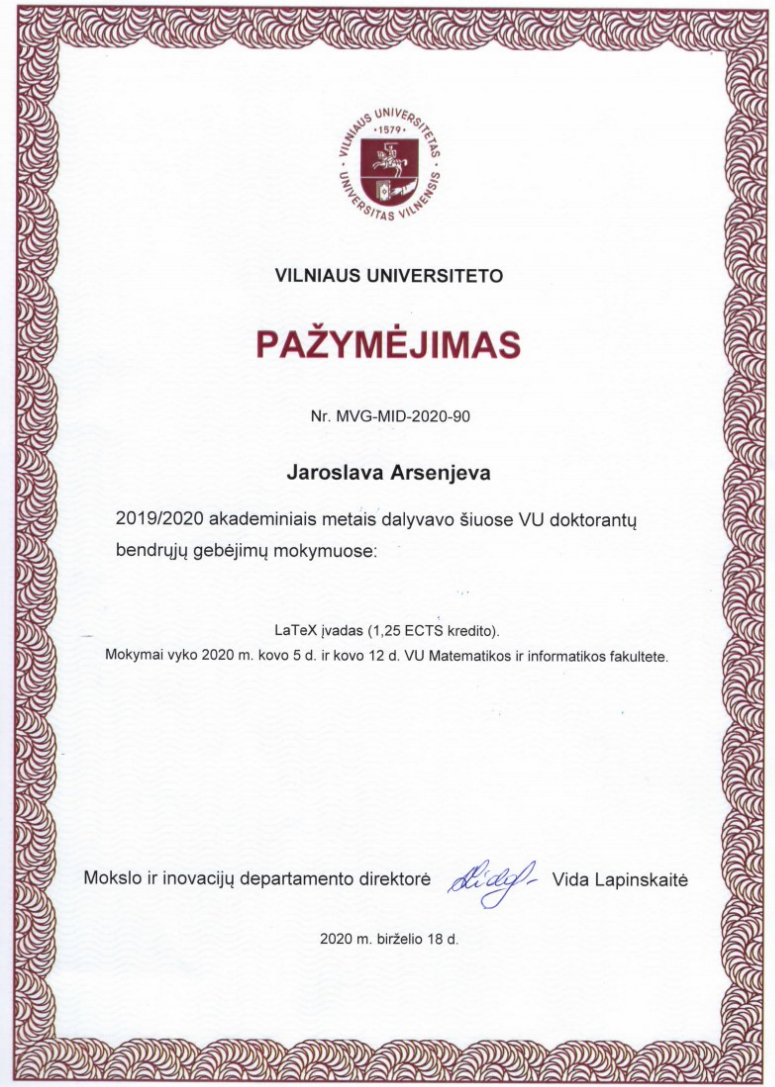
(Research Council of Lithuania,
Vadove – Dr. Gerda Ana Melnik)

The object of research: the correlation between levels of stress due to the pandemic and consumption behavior of Lithuanian population.

Tasks: Using experimental methods of cognitive psychology study how stress affects people's behavior; with data provided by the Department of Statistics model how the pandemic affected consumption.

Result: Provide recommendations for reports and further actions of business owners/ government in different economic and epidemic scenarios, taking into account the general mood and expectations of the population.

LaTex certificate



VILNIAUS UNIVERSITETO

PAŽYMĖJIMAS

Nr. MVG-MID-2020-90

Jaroslava Arsenjeva

2019/2020 akademiniais metais dalyvavo šiuose VU doktorantų bendrųjų gebėjimų mokymuose:

LaTeX įvadas (1,25 ECTS kredito).
Mokymai vyko 2020 m. kovo 5 d. ir kovo 12 d. VU Matematikos ir informatikos fakultete.

Mokslo ir inovacijų departamento direktorė Vida Lapinskaitė

2020 m. birželio 18 d.

Work plan for 2020/2021

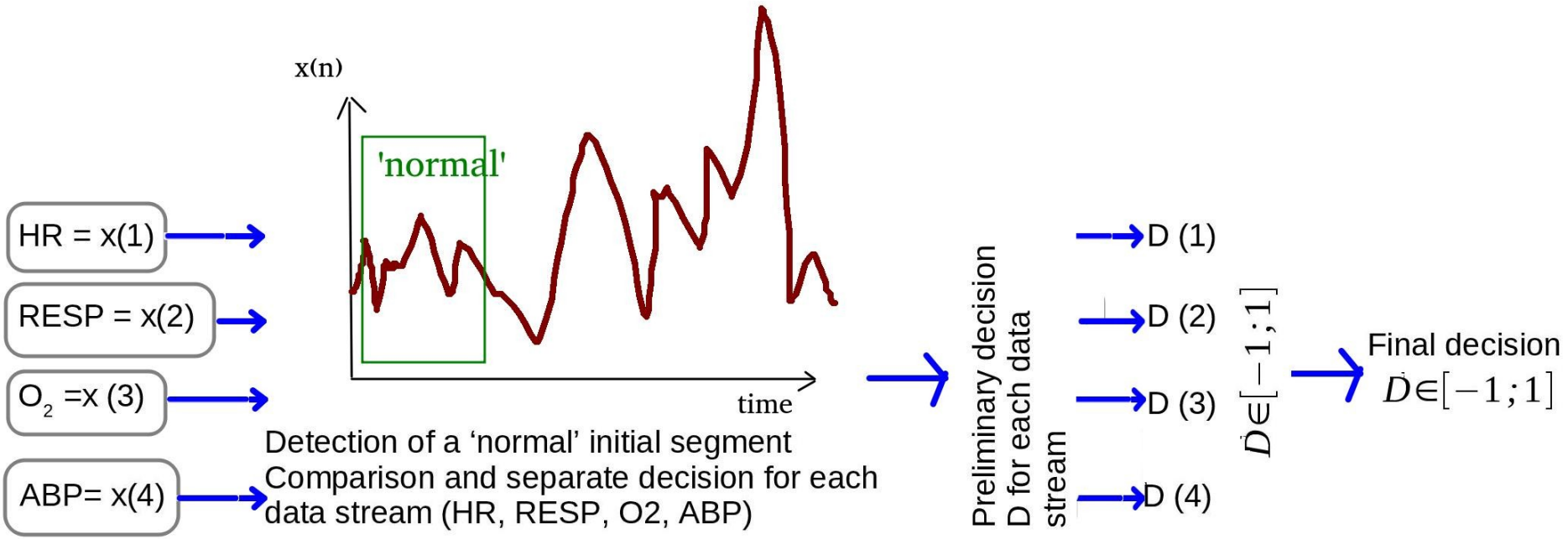
- Find a way to improve an existing data fusion algorithm using data analysis techniques.
- Investigate the developed algorithm by solving several test problems;
- Identify ways of implementing the created algorithm to solve practical problems.
- Analyze and make conclusion about the application of created algorithm.
- Participate and present research results at an international scientific conference in Lithuania or abroad. (2 conferences because of cancelled one last year).
- Publish a paper in a peer-reviewed periodical journal.

Data stream for prediction

- Acute respiratory distress syndrome
- Data streams: Heart rate (HR)
- Respiratory rate (RESP)
- Peripheral arterial oxygen saturation (O_2)
- Mean airway blood pressure (ABP_{Mean})

Taoum A, Mourad-Chehade F, Amoud H. Early-warning of ARDS using novelt
data fusion. Comput Biol Med. 2018 Nov 1;102:191-199.

Model



Taoum A, Mourad-Chehade F, Amoud H. Early-warning of ARDS using novelty data fusion. Comput Biol Med. 2018 Nov 1;102:191-199.

Inference fusion

- Outliers for ‘normal’ segment (3 standart deviations)
- More outliers in latter segment than normal → D=1 (likely to develop ARDS)
- Normal segment and deviations optimized using the test dataset

Taoum A, Mourad-Chehade F, Amoud H. Early-warning of ARDS using novelty data fusion. Comput Biol Med. 2018 Nov 1;102:191-199.

Final decision fusion

Final decision = $D_0 + \sum \bar{D}_i W_i$

$D_0 = \log\left(\left(Amount\ of\ stable\right) / \left(Amount\ of\ unstable\right)\right)$

W_i for (stable, D=-1) is calculated as follows:
log((sensitivity)/(1-specificity))

W_i for (unstable, D=1) is calculated as follows:
log((specificity)/(1-sensitivity))

Taoum A, Mourad-Chehade F, Amoud H. Early-warning of ARDS using novelt
data fusion. Comput Biol Med. 2018 Nov 1;102:191-199.

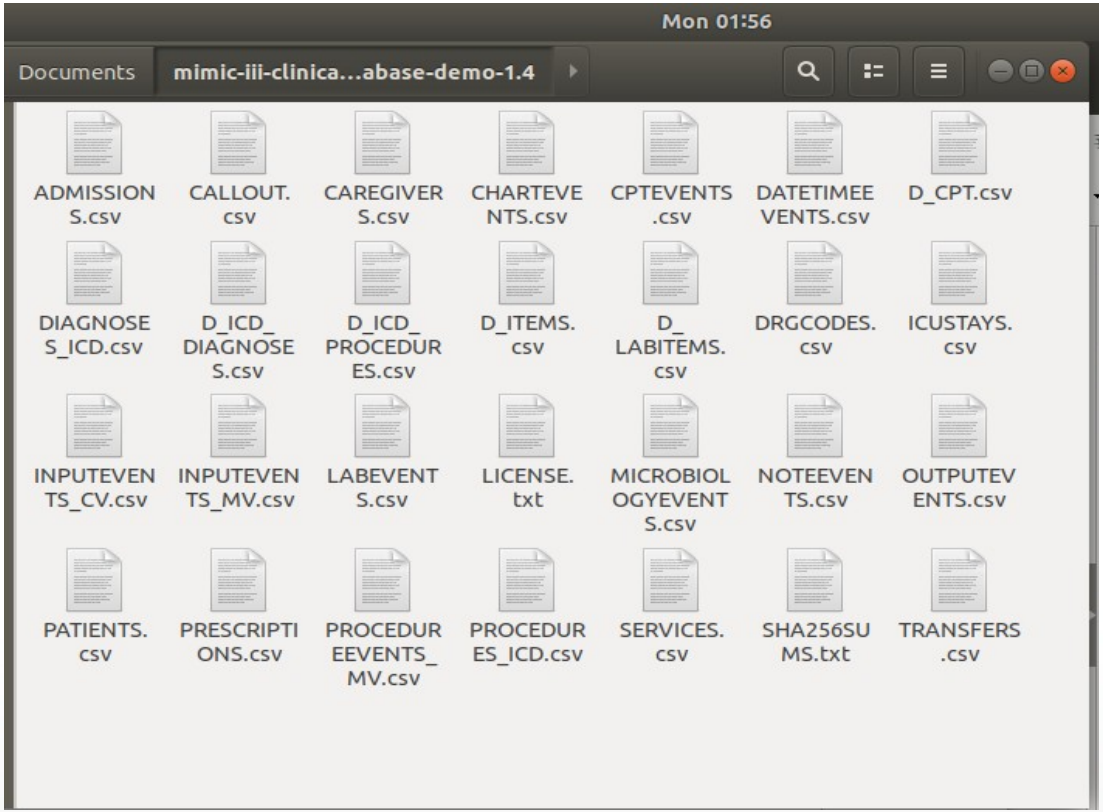
Results

Lead time	Sensitivity (%)	Specificity (%)
0h	85.71	50
6h	85.71	50
12h	85.71	50
24h	78.57	50

Taoum A, Mourad-Chehade F, Amoud H. Early-warning of ARDS using novelt
data fusion. Comput Biol Med. 2018 Nov 1;102:191-199.

•MIMIC database issues

- Separate files
- Some patients were not
- given ARDS related tests
- Less healthy patients
- (can be duplicated)



Common algorithms

AVG/ weighted AVG :The average of temperature measurements from different body parts
Kalman filter: Acceleometer + gyroscope (detecting if a person is standing or leaning on a wall)
Particle filter: Detecting the position of a human; segmentation of noisy medical images
ANN: Activity recognition, natural language processing, cancer detection, identification and diagnosis of micro-calcification, remote medical diagnosis, image processing (convolution layers)
Decision trees: Mostly activity recognition, condition surveillance; heath system improvement
Kmeans: Activity recognition, cancer diagnosis and treatment, healthy/ unhealthy cell classification
KNN: Activity classification; healthy/unhealthy cell classification
SVM: Activity classification; cancer diagnosis; tumor classification, gene classification
Bayesian inference: Activity recognition, reverse engineering of networks from genomics; cancer diagnosis; gene system modeling and analyzing
Fuzzy logic: Image based decisions (cancer cells identification for example); diagnosis of brain conditions, cancer treatment; ovarian cancer detection and diagnosis, gene model reconstruction