I INTERNATIONAL CONFERENCE
KRASNOYARSK, RUSSIA
30 July 2020

CAMSTECH
International Conference on Advances in Material Science and Technology

Science and Technology City Hall
KRASNOYARSK, RUSSIA

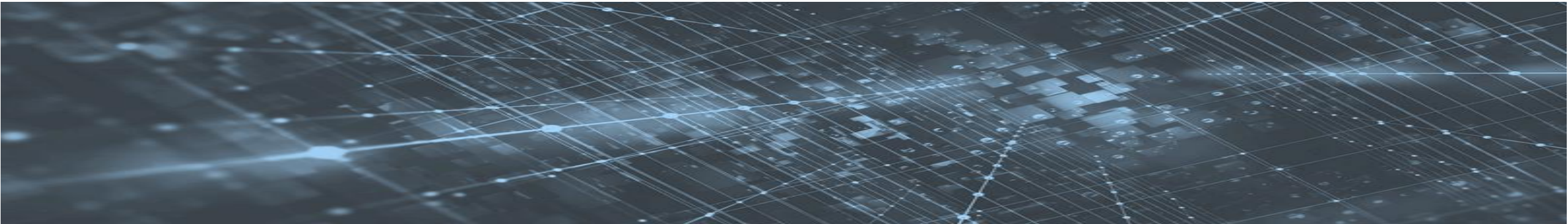
«International Conference on Advances in Material Science and Technology - CAMSTech-2020»

«Additional features implementation in the heart disease classification problem»

D A Petrusevich

Problem statement

- Heart diseases dataset has been used for experiments
- The main goal of the dataset is to build classifiers: is there any heart disease?
- Clustering technique allows understand how many types of diseases are in the dataset
- Subgroups in the dataset can be handled individually (for example, patients that are older than 60)
- Additional features can be used to construct classifiers



Solution methods

- Additional features of the 2nd and 3rd degrees are added
- Principal component analysis method is used to decrease dimensionality (increased with feature addition). Thus, non-linear transformation is made
- Clustering technique can answer the question: how many different diseases are in the dataset?
- Analyzing precision of classifiers, two clusters of ill people older than 60 years are constructed

Algorithm of classification	F_1 value at the enhanced dataset, %	F_1 value at the transformed dataset with PCA, %
Random forest	83.3	79.1
Gradient boosting	83.4	82.4
Support vector machine	84.6	83.4
Logistic regression	85.5	85.7

Conclusions

Results, implementation

After features addition and dimensionality reduction with the PCA method the task becomes 4-dimensional (the original dataset has got 13 columns)

Clustering of patients allows understand how many diseases are in the dataset. Thus, quality of classifiers can be increased

Specialists in medicine are able to explain features in non-linear transformed space of task and clusters that mean various types of heart diseases

Contacts

D A Petrusevich

Russian Technological University (MIREA), Moscow

E-mail: petrdenis@mail.ru