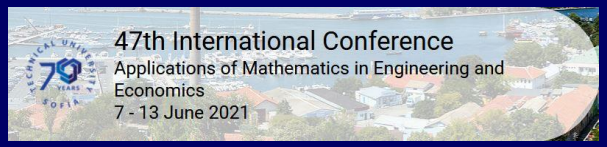



The 47th International Conference
APPLICATIONS OF MATHEMATICS IN ENGINEERING AND ECONOMICS
(AMEE 2021)

# Determination of Accuracy and Probability in the Analysis of Large-Scale Biomedical Data

Stella Vetova

Primary funding for the presented work was provided by the National Science Fund, Ministry of Education and Science, Republic of Bulgaria under contract **KP-06-N37/24**, research project “Innovative Platform for Intelligent Management and Analysis of Big Data Streams Supporting Biomedical Scientific Research”.

## Large Scale Data

1.Generated heterogeneous amount of data used for further processing and analysis.

2.Features of Big Data :

-complexity;

-variety;

-volume ;

-application opportunities.

## Areas of Large-Scale Data Application

-science, business, medicine, biomedicine,
bioinformatics, agriculture, biology;

## Gathering Large-Scale Data Application

-laboratory results, clinical test, patients' exams,
symptoms data captured by the means of telemedicine,
etc.

# Workflow samples

Definition for workflow - a sequence of functions defined to perform a single task;

Advantages:

1. computes complex tasks;
2. includes data visualization and analysis based on the principles of segmentation, diagnosis and therapy;

# Workflow Samples

3. Workflows are applied for complex automatic analysis to improve the interpretation and reporting, reducing time and providing ease of decision making in the process of disease diagnosis and treatment, classification and detection;

# Methods, Algorithms and Tools for Large-Sacle Data Workflow Analysis in Biomedicine

| | | |
|---|---|---|
| Data Classification | Clustering | k-medoids algorithm; |
| | | k-Means partitioning algorithm; |
| | | k-Nearest Neighbor (KNN); |
| | | Unweighted Pair Group Method with Arithmetic Mean (UPGMA); |
| | | Neighbor-Joining (NY) method; |
| | | Fitch and Kitsch method; |
| | Distance computation | Euclidean distance; |
| | | Hamming distance; |
| | | Manhattan distance; |
| | | Minkowski distance; |
| Data Learning | Deep Learning | Convolutional Neuron Networks (CNN); |
| | | Deep Boltzmann Machine; |
| | | Self-Organizing Map (SOM); |

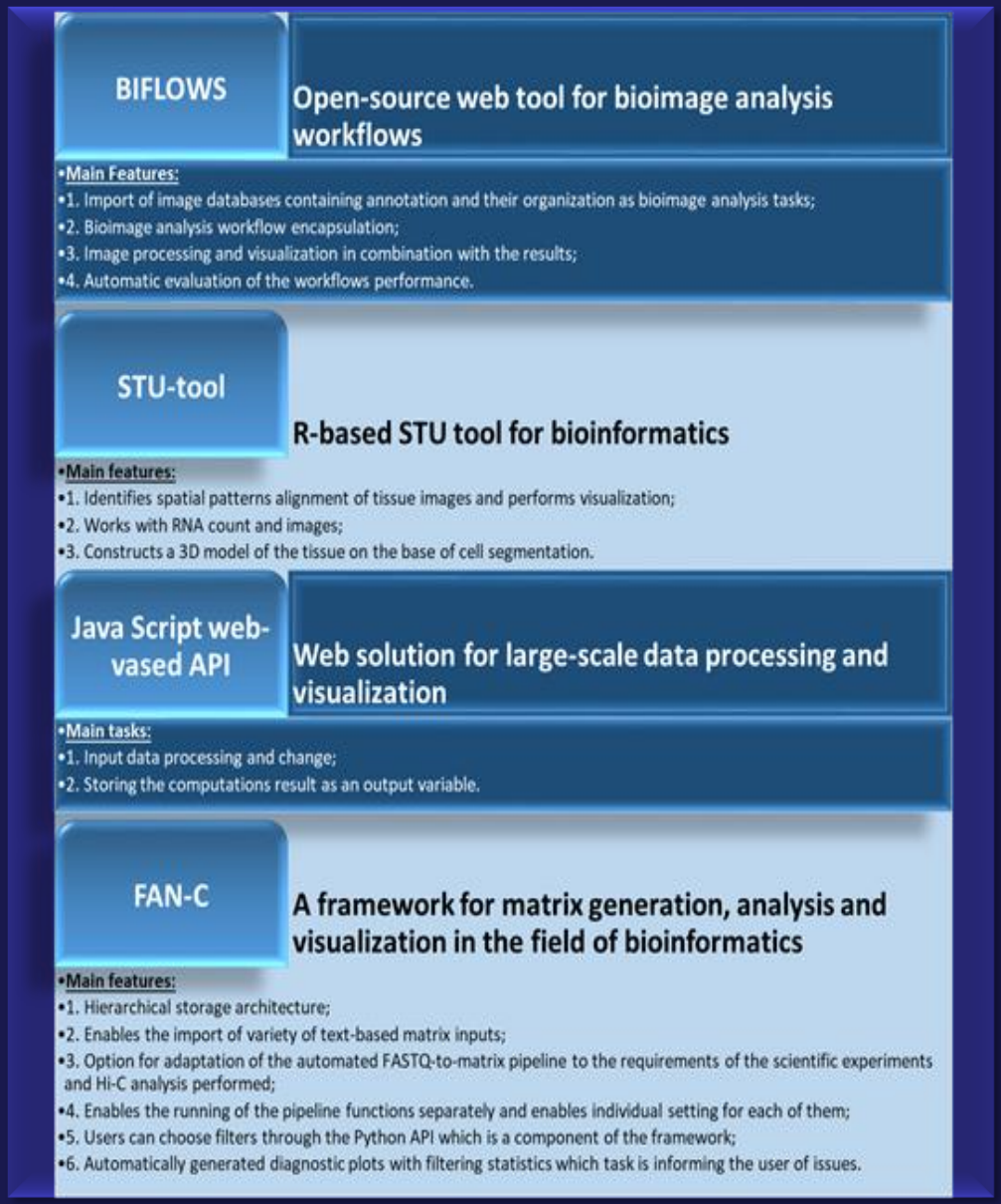
| Issue | Approach | Methods |
|---|---|---|
| Dimensionality reduction | Mapping methods | Principal Component Analysis (PCA); |
| | | Singular value decomposition; |
| | Non-Linear Mapping methods | Kernel Principal Component Analysis; |
| | | Laplasian eigenmaps; |
| | | Sammon's mapping; |

# Methods, Algorithms and Tools for Large-Sacle Data Workflow Analysis in Biomedicine

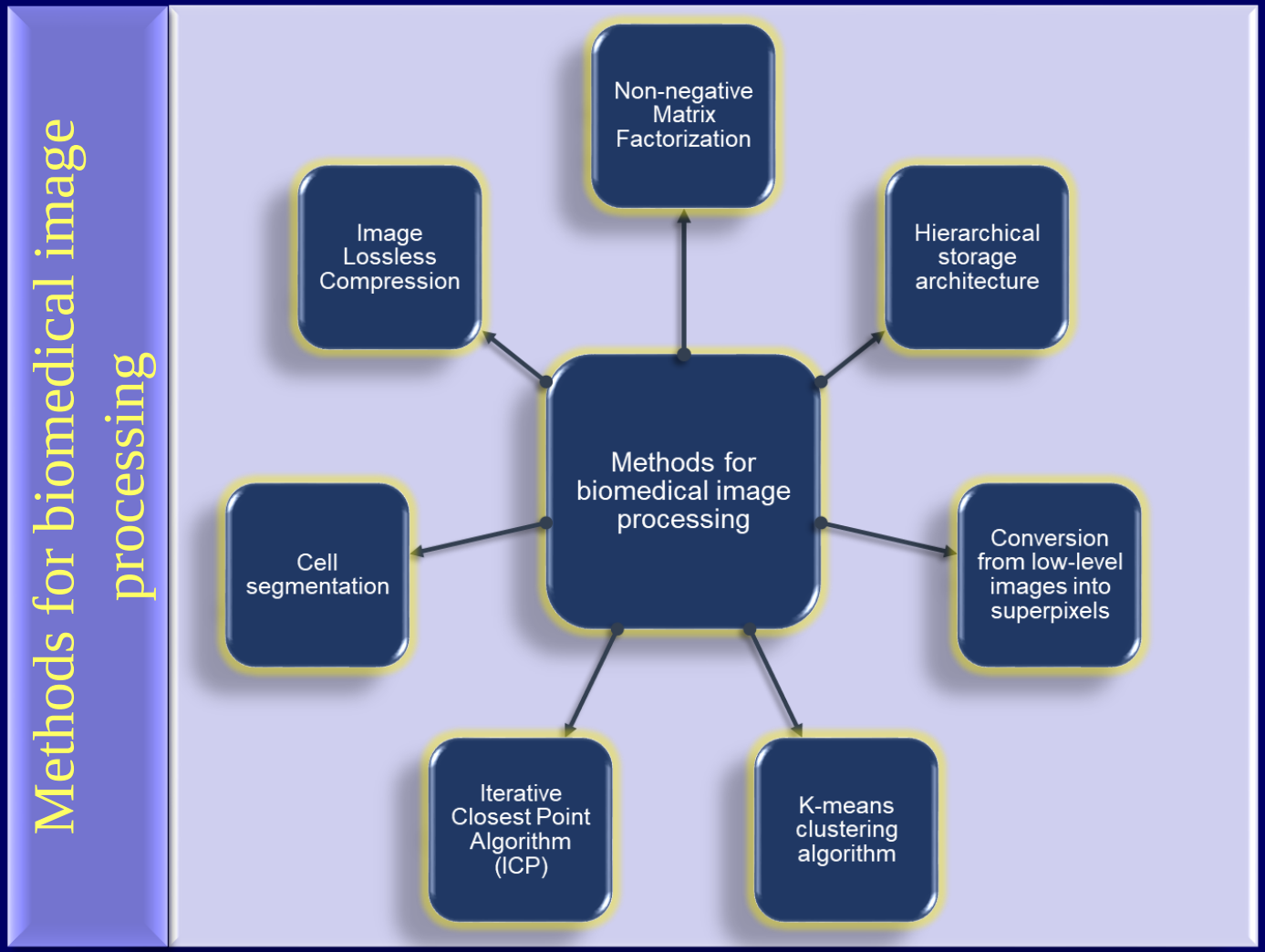
| | |
|---|---|
| **Workflow platforms and design tools** | Galaxy; |
| | myExperiment |
| | Taverna |
| | MapReduce |
| | Hadoop |
| | Spark |
| | GraphLab |
| | Pregel |
| | Closha |
| | Preglix |
| | Pegasus |
| | Kepler |
| | Chipster |
| | Mike |
| | GraphFlow |
| | Gromacs |

| | |
|---|---|
| **Storage Environment** | Cloud storage; |
| | Local storage |
| **Programming Languages** | Bash |
| | R |
| | Perl |
| | Python |

# Large-Scale data visualization tools

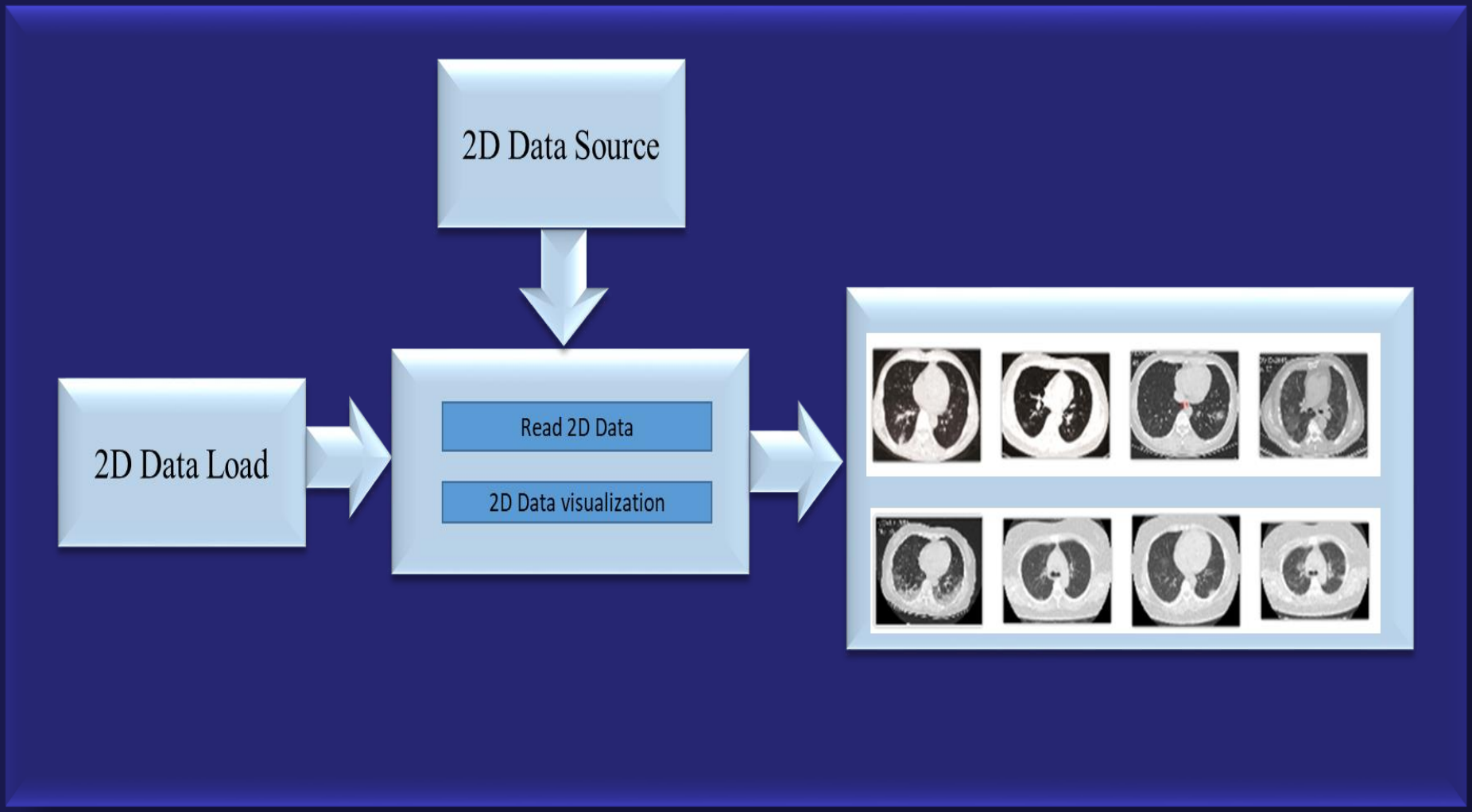


The 47th International Conference
APPLICATIONS OF MATHEMATICS IN ENGINEERING AND ECONOMICS
(AMEE 2021)

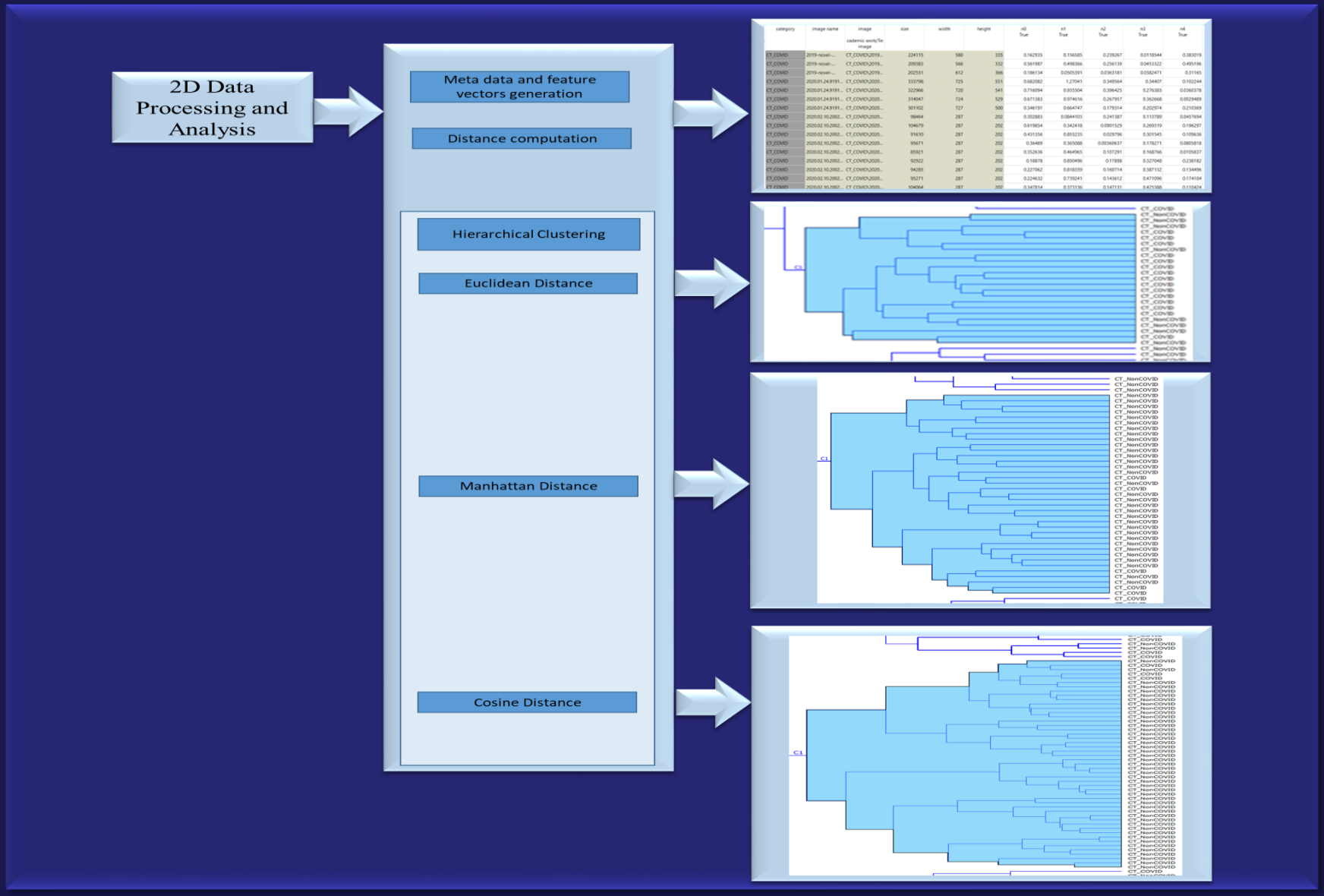



June 7-13, 2021

# Distance-Based Workflow Sample and results



# The 47th International Conference
# APPLICATIONS OF MATHEMATICS IN ENGINEERING AND ECONOMICS
# (AMEE 2021)




June 7-13, 2021

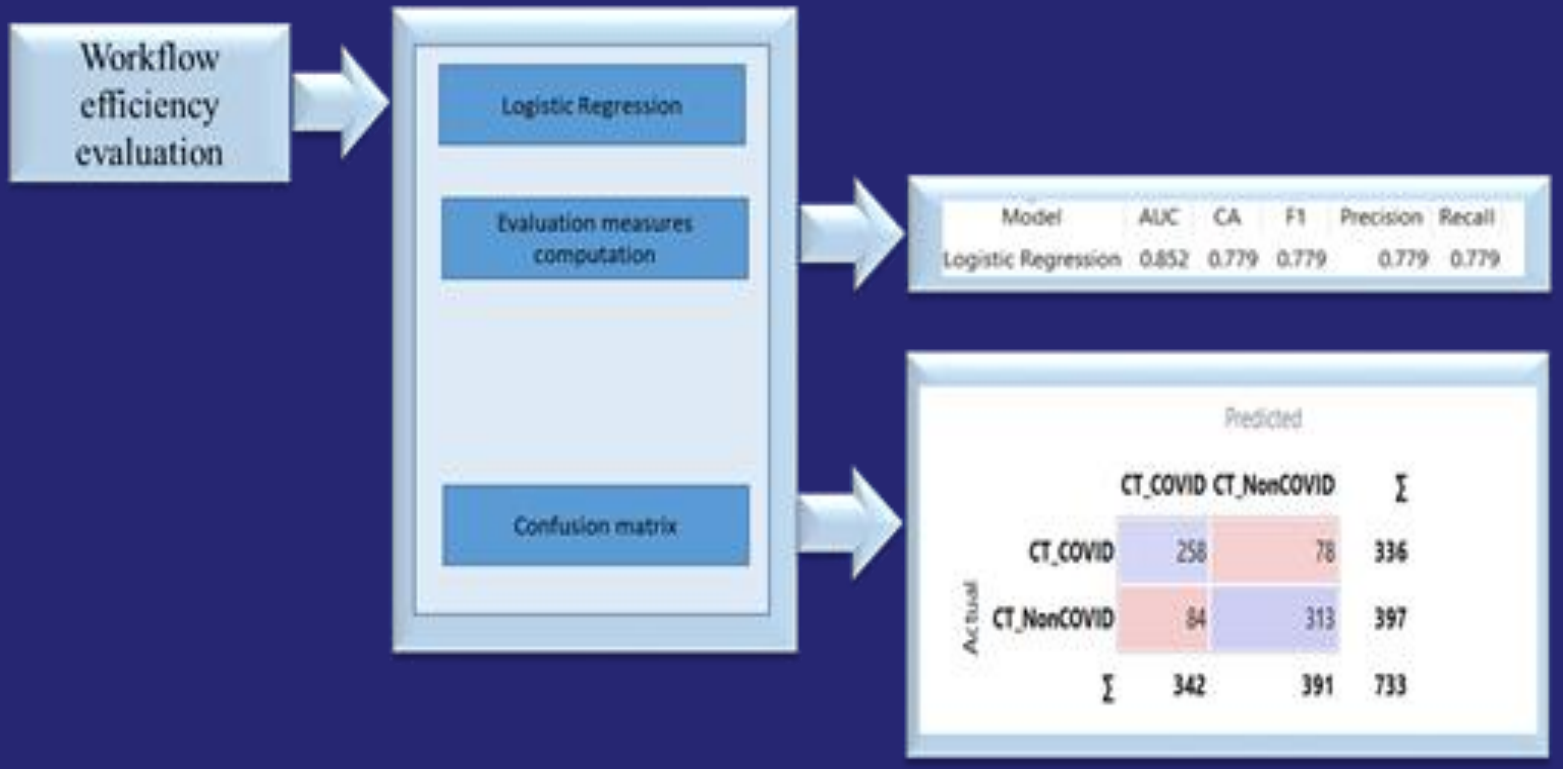



Image meta data and feature vectors

# Thank You!

Stella Vetova
vetova.bas@gmail.com