

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ РАДИОФИЗИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ
Кафедра системного анализа и компьютерного моделирования

**РАЗРАБОТКА АЛГОРИТМОВ ПРЕДСКАЗАНИЯ ВЫЖИВАЕМОСТИ
ПАЦИЕНТОВ С ОНКОЛОГИЧЕСКИМИ ЗАБОЛЕВАНИЯМИ**

Аннотация к дипломной работе

Яцкова Владислава Николаевича
студента 4 курса
специальность
«компьютерная безопасность»
Научный руководитель:
ассистент М. К. Чепелева

Минск, 2022

РЕФЕРАТ

Дипломный проект: 55 страницы, 13 иллюстраций, 13 таблиц, 38 источников, 2 приложения.

ПРЕДСКАЗАНИЕ ВЫЖИВАЕМОСТИ, СЛУЧАЙНЫЙ ЛЕС ВЫЖИВАЕМОСТИ, БЛОЧНЫЙ ЛЕС, АЛГОРИТМ РАСЩЕПЛЕНИЯ УЗЛОВ EXPONENTIAL LOG-LIKELIHOOD LOSS, РЕГРЕССИЯ КОКСА, БУСТИНГ РЕГРЕССИИ КОКСА, МЕТОД НЕЗАВИСИМЫХ КОМПОНЕНТ, МЕТОД ГЛАВНЫХ КОМПОНЕНТ, ЗНАЧИМЫЕ ГЕНЫ, ДИФФЕРЕНЦИАЛЬНО ЭКСПРЕССИРОВАННЫЕ ГЕНЫ, С-ИНДЕКС, ОЦЕНКА БРИЕРА, ЭКСПРЕССИЯ ГЕНОВ

Объектом исследования является функции выживаемости пациентов.

Цель работы – разработка алгоритма предсказания выживаемости пациентов с онкологическими заболеваниями.

Методы исследования: алгоритмы случайного леса выживаемости и блочного леса с методом расщепления узлов exponential log-likelihood loss, алгоритм регрессии Кокса для предсказания выживаемости; методы главных и независимых компонент, отбор значимых и дифференциально экспрессированных генов для снижения размерности данных; С-индекс и оценка Бриера для оценивания результатов предсказания выживаемости. Проведен анализ клинических признаков и экспрессий генов пациентов с онкологическими заболеваниями, разработаны алгоритмы на основе деревьев решений и проведено сравнение точности работы с существующими, определена методика подготовки многомерных данных к исследованию.

РЭФЕРАТ

Дыпломны праект: 55 старонкі, 13 ілюстрацый, 13 табліц, 38 крыніц, 2 пракладання.

ПРАДКАЗАННЕ ВЫЖЫВАЛЬНАСЦІ, ВЫПАДКОВЫ ЛЕС ВЫЖЫВАЛЬНАСЦІ, БЛОЧНЫ ЛЕС, АЛГАРЫТМ РАСШЧАПЛЕННЯ ВУЗЛОЎ EXPONENTIAL LOG-LIKELIHOOD LOSS, РЭГРЭСІЯ КОКСА, БУСТЫНГ РЭГРЭСІІ КОКСА, МЕТАД НЕЗАЛЕЖНЫХ КАМПАНАЕНТ, МЕТАД ГАЛОЎНЫХ КАМПАНАЕНТ, ЗНАЧНЫЯ ГЕНЫ, ДЫФЕРЭНЦЫЯЛЬНА ЭКСПРЭСІРАВАНЫЯ ГЕНЫ, С-ІНДЭКС, АЦЭНКА БРЫЕРА, ЭКСПРЭСІЯ ГЕНАЎ

Аб'ектам даследавання з'яўляецца функцыі выжывальнасці пацыентаў.

Мэта працы – распрацоўка алгарытму прадказання выжывальнасці пацыентаў з анкалагічнымі захворваннямі.

Метады даследавання: алгарытмы выпадковага лесу выжывальнасці і блочнага лесу з метадам расшчаПЛення вузлоў exponential log-likelihood loss, алгарытм рэгрэсіі Коксу для прадказання выжывальнасці; метады галоўных і незалежных кампанент, адбор значных і дыферэнцыяльна экспрэсаваных генаў для зніжэння памернасці дадзеных; С-індэкс і ацэнка Брыера для ацэньвання вынікаў прадказання выжывальнасці. Праведзены аналіз клінічных прыкмет і экспрэсій генаў пацыентаў з анкалагічнымі захворваннямі, распрацаваны алгарытмы на аснове дрэў рашэнняў і праведзена параўнанне дакладнасці рацы з існуючымі, вызначана метадыка падрыхтоўкі шматмерных дадзеных да даследавання.

ABSTRACT

Diploma project: 55 pages, 13 illustrations, 13 tables, 38 sources, 2 applications.

SURVIVAL PREDICTION, RANDOM SURVIVAL FOREST, BLOCK FOREST, EXPONENTIAL LOG-LIKELIHOOD LOSS NODE SPLITTING ALGORITHM, COX REGRESSION, COX REGRESSION BOOSTING, INDEPENDENT COMPONENT ANALYSIS, PRINCIPAL COMPONENT ANALYSIS, SIGNIFICANT GENES, DIFFERENTIALLY EXPRESSED GENES, C-INDEX, BRIER SCORE, GENE EXPRESSION

The object of the study is the survival function of patients.

The purpose – to develop an algorithm for survival predicting of cancer patients.

Research methods: random survival forest and block forest algorithms with exponential log-likelihood loss node splitting method, Cox regression algorithm for survival prediction; principal and independent components analysis, selection of significant and differentially expressed genes to reduce data dimension; C-index and Brier score for evaluating survival predictive outcomes.

The analysis of clinical signs and gene expressions of cancer patients was carried out, algorithms based on decision trees were developed and the accuracy of work was compared with existing ones, a methodology for preparing multidimensional data for the study was determined.