



دانشگاه علوم پزشکی و خدمات  
بهداشتی دامانی کرمان

دانشکده بهداشت

پایان نامه مقطع کارشناسی ارشد آمار زیستی

عنوان

مقایسه روش‌های رگرسیون لجستیک لاسو، آداپتیو لاسو و الستیک نت به منظور تعیین ژن‌های  
مهم و اثرگذار در سرطان پستان در زنان

توسط

صادق رؤفی

استاد راهنما

خانم دکتر تانیا دهش

استاد مشاور

خانم دکتر سعیده جعفری نژاد

سال تحصیلی: ۱۴۰۰-۱۳۹۹

شماره پایان‌نامه:



دانشگاه علوم پزشکی گیلان  
تخصصیات تکمیلی دانشکاه

بسمه تعالی

صورتجلسه دفاع از پایان نامه

تاریخ .....  
شماره .....  
پیوست .....

جلسه دفاعیه پایان نامه تحصیلی خواهشمند است نظر خود را در مورد پایان نامه آقای صادق رتوفی دانشجوی کارشناسی ارشد رشته آمار زیستی تحت عنوان "مقایسه کارایی روش های رگرسیون لجستیک لاسو، رگرسیون لجستیک لاسوی انطباقی و رگرسیون لجستیک الستیگ نت به منظور تعیین موثرترین ژن های مرتبط با سرطان پستان در زنان " به راهنمایی خانم دکتر تانیا دهش اعلام نمائید. در ساعت ۱۲ روز شنبه مورخ ۹۹/۴/۲۱ با حضور اعضای محترم هیات داوران متشکل از:

امضا	نام و نام خانوادگی	سمت
	خانم دکتر تانیا دهش	الف: استاد(ان) راهنما
	خانم دکتر سعیده جعفری نژاد	ب: استاد(ان) مشاور
	آقای دکتر عباس بهرام پور	ج: عضو هیات داوران (داخلی)
	آقای دکتر محمدرضا باناشی	د: عضو هیات داوران (خارجی)
	آقای دکتر یونس جهانی	ه: نماینده تخصصیات تکمیلی

تشکیل گردید و ضمن ارزیابی به شرح پیوست با درجه ..... مجید ..... و نمره ..... ۱۸۱۸۴۵ ..... مورد تایید قرار گرفت.



## فهرست مندرجات

.....	فهرست مندرجات
.....	فهرست جداول
.....	فهرست تصاویر و نمودارها
.....	فهرست علائم اختصاری
.....	چکیده

### فصل اول: مقدمه و اهداف

.....	مقدمه	۱-۱
.....	رگرسیون لجستیک و انتخاب متغیر	۱-۱-۱
.....	داده‌های با بُعد بالا	۲-۱-۱
.....	رگرسیون لجستیک لاسو	۱-۳-۱
.....	رگرسیون لجستیک آداپتیو لاسو	۴-۱-۱
.....	رگرسیون لجستیک الستیک نت	۵-۱-۱
.....	سرطان پستان و میکروآرایه‌ها	۱-۱-۶
.....	بیان مساله و اهمیت موضوع	۲-۱
.....	هدف کلی	۳-۱
.....	اهداف جزئی	۴-۱
.....	اهداف کاربردی	۵-۱
.....	فرضیات و سوالات	۶-۱

### فصل دوم: بررسی متون

.....	مقدمه	۱-۲
.....	مروری بر پژوهش‌های پیشین	۲-۲

### فصل سوم: مواد و روش‌ها

.....	ابزار جمع آوری اطلاعات و نحوه جمع آوری آن	۱-۳
.....	پردازش داده‌های میکروآرایه	۲-۳
.....	روش تجزیه و تحلیل داده‌ها	۳-۳
.....	ملاحظات اخلاقی	۴-۳
.....	تصادفی سازی داده‌ها و تعیین گروه آموزش و آزمون	۵-۳
.....	روش‌های ارزیابی و مقایسه مدل‌ها	۶-۳
.....	ماتریس درهم‌ریختگی	۱-۶-۳

Error! Bookmark not defined.....	۲-۶-۳ شاخص C (C-index)
Error! Bookmark not defined.....	۳-۶-۳ ضریب کاپا
Error! Bookmark not defined.....	۳-۷ رگرسیون لجستیک تاوانیده
Error! Bookmark not defined.....	۱-۷-۳ رگرسیون لجستیک ریج
Error! Bookmark not defined.....	۲-۷-۳ رگرسیون لجستیک لاسو
Error! Bookmark not defined.....	۳-۷-۳ رگرسیون لجستیک آدپتیو لاسو
Error! Bookmark not defined.....	۴-۷-۳ رگرسیون لجستیک الستیک نت

#### فصل چهارم: یافته‌ها

۲۸.....	۴-۱ مقدمه
Error! Bookmark not defined.....	۲-۴ ضرایب رگرسیونی برآورد شده
Error! Bookmark not defined.....	۴-۳ گروه مقایسه‌ای ERBB2 vs. Basal
Error! Bookmark not defined.....	۴-۴ گروه مقایسه‌ای Luminal B vs. Basal
Error! Bookmark not defined.....	۴-۵ گروه مقایسه‌ای Luminal A vs. Basal
Error! Bookmark not defined.....	۴-۶ گروه مقایسه‌ای Normal Like vs. Basal
Error! Bookmark not defined.....	۴-۷ گروه مقایسه‌ای Normal vs. Basal
Error! Bookmark not defined.....	۴-۸ گروه مقایسه‌ای Luminal B vs. ERBB2
Error! Bookmark not defined.....	۴-۹ گروه مقایسه‌ای Luminal A vs. ERBB2
Error! Bookmark not defined.....	۴-۱۰ گروه مقایسه‌ای Normal Like vs. ERBB2
Error! Bookmark not defined.....	۴-۱۱ گروه مقایسه‌ای Normal vs. ERBB2
Error! Bookmark not defined.....	۴-۱۲ گروه مقایسه‌ای Luminal A vs. Luminal B
Error! Bookmark not defined.....	۴-۱۳ گروه مقایسه‌ای Normal Like vs. Luminal B
Error! Bookmark not defined.....	۴-۱۴ گروه مقایسه‌ای Normal vs. Luminal B
Error! Bookmark not defined.....	۴-۱۵ گروه مقایسه‌ای Normal Like vs. Luminal A
Error! Bookmark not defined.....	۴-۱۶ گروه مقایسه‌ای Normal vs. Luminal A
Error! Bookmark not defined.....	۴-۱۷ گروه مقایسه‌ای Normal vs. Normal Like

#### فصل پنجم: بحث و نتیجه‌گیری

Error! Bookmark not defined.....	۵-۱ نتیجه‌گیری
Error! Bookmark not defined.....	۵-۲ نقاط ضعف و قوت مطالعه
Error! Bookmark not defined.....	۳-۵ پیشنهادها

منابع..... ۵۷

پیوست ..... ۶۰

## فهرست جداول

- جدول ۳-۱: حجم نمونه هر زیرگروه سرطان پستان در همه چهار GSE ..... Error! Bookmark not defined.
- جدول ۳-۲: ماتریس درهم‌ریختگی برای دو رده P و N ..... Error! Bookmark not defined.
- جدول ۳-۳: صورت احتمالی نظرات دو فرد در مورد موضوع ..... Error! Bookmark not defined.
- جدول ۴-۱: ضرایب رگرسیونی برآورد شده توسط سه مدل برای زیرگروه‌های سرطان پستان ..... Error! Bookmark not defined.
- جدول ۴-۲: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه ERBB2 vs. Basal ..... Error! Bookmark not defined.
- جدول ۴-۳: معیارهای ارزیابی عملکرد مدل‌ها در گروه ERBB2 vs. Basal ..... Error! Bookmark not defined.
- جدول ۴-۴: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه Luminal B vs. Basal ..... Error! Bookmark not defined.
- جدول ۴-۵: معیارهای ارزیابی عملکرد مدل‌ها در گروه Luminal B vs. Basal ..... Error! Bookmark not defined.
- جدول ۴-۶: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه Luminal A vs. Basal ..... Error! Bookmark not defined.
- جدول ۴-۷: معیارهای ارزیابی عملکرد مدل‌ها در گروه Luminal A vs. Basal ..... Error! Bookmark not defined.
- جدول ۴-۸: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه Normal Like vs. Basal ..... Error! Bookmark not defined.
- جدول ۴-۹: معیارهای ارزیابی عملکرد مدل‌ها در گروه Normal Like vs. Basal ..... Error! Bookmark not defined.
- جدول ۴-۱۰: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه Normal vs. Basal ..... Error! Bookmark not defined.
- جدول ۴-۱۱: معیارهای ارزیابی عملکرد مدل‌ها در گروه Normal vs. Basal ..... Error! Bookmark not defined.
- جدول ۴-۱۲: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه Luminal B vs. ERBB2 ..... Error! Bookmark not defined.
- جدول ۴-۱۳: معیارهای ارزیابی عملکرد مدل‌ها در گروه Luminal B vs. ERBB2 ..... Error! Bookmark not defined.
- جدول ۴-۱۴: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه Luminal A vs. ERBB2 ..... Error! Bookmark not defined.
- جدول ۴-۱۵: معیارهای ارزیابی عملکرد مدل‌ها در گروه Luminal A vs. ERBB2 ..... Error! Bookmark not defined.
- جدول ۴-۱۶: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه Normal Like vs. ERBB2 ..... Error! Bookmark not defined.
- جدول ۴-۱۷: معیارهای ارزیابی عملکرد مدل‌ها در گروه Normal Like vs. ERBB2 ..... Error! Bookmark not defined.
- جدول ۴-۱۸: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه Normal vs. ERBB2 ..... Error! Bookmark not defined.
- جدول ۴-۱۹: معیارهای ارزیابی عملکرد مدل‌ها در گروه Normal vs. ERBB2 ..... Error! Bookmark not defined.
- جدول ۴-۲۰: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه Luminal A vs. Luminal B ..... Error! Bookmark not defined.
- جدول ۴-۲۱: معیارهای ارزیابی عملکرد مدل‌ها در گروه Luminal A vs. Luminal B ..... Error! Bookmark not defined.

- Error! Bookmark ... **Normal Like vs. Luminal B** جدول ۲۲-۴: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه **Normal Like vs. Luminal B** not defined.
- Error! Bookmark not ..... **Normal Like vs. Luminal B** جدول ۲۳-۴: معیارهای ارزیابی عملکرد مدل‌ها در گروه **Normal Like vs. Luminal B** defined.
- Error! Bookmark not ..... **Normal vs. Luminal B** جدول ۲۴-۴: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه **Normal vs. Luminal B** defined.
- Error! Bookmark not defined. .... **Normal vs. Luminal B** جدول ۲۵-۴: معیارهای ارزیابی عملکرد مدل‌ها در گروه **Normal vs. Luminal B**
- Error! Bookmark ... **Normal Like vs. Luminal A** جدول ۲۶-۴: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه **Normal Like vs. Luminal A** not defined.
- Error! Bookmark not ..... **Normal Like vs. Luminal A** جدول ۲۷-۴: معیارهای ارزیابی عملکرد مدل‌ها در گروه **Normal Like vs. Luminal A** defined.
- Error! Bookmark not ..... **Normal vs. Luminal A** جدول ۲۸-۴: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه **Normal vs. Luminal A** defined.
- Error! Bookmark not defined. .... **Normal vs. Luminal A** جدول ۲۹-۴: معیارهای ارزیابی عملکرد مدل‌ها در گروه **Normal vs. Luminal A**
- Error! Bookmark not .. **Normal vs. Normal Like** جدول ۳۰-۴: تعداد ژن‌های بدست آمده از برازش سه مدل در گروه **Normal vs. Normal Like** defined.
- Error! Bookmark not defined. .... **Normal vs. Normal Like** جدول ۳۱-۴: معیارهای ارزیابی عملکرد مدل‌ها در گروه **Normal vs. Normal Like**

### فهرست تصاویر و نمودارها

- شکل ۳-۱: سه افراز بندی داده‌ها در مدل‌های آماری ..... **Error! Bookmark not defined.**
- شکل ۳-۲: منحنی ROC برای انتخاب بهترین نقطه آستانه‌ای ..... **Error! Bookmark not defined.**

## فهرست علائم اختصاری

<b>Abbreviations</b>	
LASSO	Least Absolute Shrinkage and Selection Operator
OLS	Ordinary Least Square
GEO	Gene Expression Omnibus
GSE	Ground Support Equipment
ROC	Receiver Operating Characteristic
AUC	Area Under ROC Curve
RSS	Residual Sum of Square
LOOCV	Leave-One-Out Cross Validation
MAE	Mean Absolute Error



## چکیده

**مقدمه و اهداف:** سرطان دومین علت مرگ و میر در سطح جهان است، به طوریکه ۹/۶ میلیون نفر در سال ۲۰۱۸ به علت سرطان جان خود را از دست داده‌اند. طبق گزارش سازمان بهداشت جهانی، از هر ۶ مرگ و میر یکی از آن‌ها ناشی از سرطان است. سرطان پستان یک سرطان شایع بخصوص در زنان در سراسر جهان است به طوریکه ۱۸ درصد از انواع سرطان‌ها را در زنان شامل می‌شود و پنجمین علت مرگ و میر در دنیا است. داده‌های ریزآرایه میزان بیان هزاران ژن را به طور همزمان، جهت مطالعه رفتار و عملکرد سلول مورد بررسی قرار می‌دهند، در واقع با بررسی تعداد زیادی ژن در کنار یکدیگر انتظار می‌رود تغییرات حاصل از بیماری را دقیق‌تر بتوان در علم بیولوژیک مورد بررسی قرار داد. داده‌های با بُعد بالا که اغلب به داده‌های  $n$  کوچک (حجم نمونه یا تعداد مشاهدات کم) و  $p$  بزرگ (تعداد متغیرهای پیشگو) معروف هستند، یکی از بزرگترین چالش‌های ممکن برای محققان می‌باشند. وجود داده‌های با بُعد بالا مشکلات بسیار زیادی در برآزش مدل‌های رگرسیون بوجود می‌آورد. انتخاب موثرترین متغیرها از بین تعداد زیادی متغیر چالش اصلی در مواجهه با این داده‌ها است. انتخاب متغیرهای بهینه تفسیر پذیری مدل را راحت‌تر و از اتلاف وقت و هزینه جلوگیری می‌کند. روش‌های معمول برآوردیابی مانند کمترین توان‌های دوم و ماکزیمم لاکلیهود در مواجهه با داده‌هایی با بُعد بالا عملکرد قابل اطمینانی از خود نشان نمی‌دهند. روش‌های رگرسیونی توانانیده (Penalized

Regression) در دهه‌های اخیر به عنوان روشی کارآمد در مواجهه با داده‌های بُعد بالا بشمار می‌روند. این روش‌ها، ضرایب رگرسیونی را با اعمال محدودیت روی دامنه تغییرات آنها برآورد می‌کنند. از جمله این روش‌ها می‌توان به روش لاسو، روش آدپتیو لاسو و روش الستیک نت اشاره نمود.

**روش‌ها:** چهار مجموعه داده (Gene Expression Omnibus) GEO از مطالعات مستقل برای این مطالعه از طریق سایت PubMed جمع‌آوری شده‌اند و هر چهار مطالعه شامل مجموعه داده میکروآرایه از سرطان پستان برای زیرگروه‌های مختلف می‌باشند. پنج زیرگروه سرطان پستان نیز شامل Basal ، Luminal A ، Luminal B ، ERBB2 و Normal like است. همچنین متغیرهای پیشگوی مورد مطالعه در این پژوهش، ژن‌های اندازه‌گیری شده می‌باشند. بعد از نرمال‌سازی داده‌ها و حذف اثرات دسته‌ای، چهار دیتاست تجمیع شدند و فایل داده نهایی شامل ۳۲۴ نفر سرطانی و ۱۲ نفر نرمال می‌باشد.

**یافته‌ها:** در مجموع ۱۵ مقایسه دو به دو بین زیرگروه‌های مختلف و افراد نرمال، در سه گروه مقایسه‌ای مدل لاسو با توجه به شاخص‌های مقایسه، بهترین عملکرد را نشان داد، در یک گروه مقایسه‌ای مدل آدپتیو لاسو با توجه به شاخص‌های مقایسه، بهترین عملکرد را نشان داد، در چهار گروه مقایسه‌ای مدل الستیک نت با توجه به شاخص‌های مقایسه، بهترین عملکرد را نشان داد، در پنج گروه مقایسه‌ای دو مدل الستیک نت و آدپتیو لاسو دقتی یکسان و بالاتر از لاسو داشتند و در دو گروه مقایسه‌ای عملکرد سه مدل یکسان گزارش گردید.

**نتیجه‌گیری:** روش‌های توانانیده در انتخاب ژن‌های موثر در سرطان پستان عملکرد قابل قبولی ارائه کردند و بر اساس مطالعات انجام شده مشخص گردید ژن‌های مشترک به دست آمده توسط سه مدل در سرطان‌های دیگر نیز دخیل هستند و جز مهمترین ژن‌ها در سرطان پستان در سایر مطالعات نیز گزارش شده‌اند.

**کلمات کلیدی:** سرطان پستان، رگرسیون توانانیده، رگرسیون لجستیک لاسو، رگرسیون لجستیک آدپتیو لاسو، رگرسیون لجستیک الستیک نت.

## Abstract

**Background and Objectives:** Cancer is the second leading cause of death globally, accounting for an estimated 9.6 million deaths, or one in six deaths, in ۲۰۱۸. Breast cancer (BC) is the most common cancer worldwide among women. So that comprises for about 18% of all types of cancers in women and the fifth reason for death around the world. Microarray data examines the expression of thousands of genes simultaneously to study cell behavior and function. In fact, by examining a large number of genes together, it is expected that the changes in the disease will be more accurately studied in biological science. High dimensional data which is called data with small number of sample size and high number of variables, one of the biggest challenge for researchers. There is problem with high dimensional data for fitting of regression models. Selecting the most effective variables from a large number of variables is the main challenge in dealing with this data. Optimal variables selection makes it easier to interpret the model and avoids wasting time and money. Typical estimation methods like least square estimation and maximum likelihood don't show Reliable performance in presence of high dimensional data. Recently, penalized regression methods have been used as an effective way to deal with high dimensional data.

These methods estimated regression coefficients by applying restrictions on the scope of their changes. These methods include the Lasso method, the Adaptive Lasso method and the elastic net method.

**Methods:** Four datasets GEO (Gene Expression Omnibus) of independent studies were downloaded from Pubmed website and all four studies included a microarray data of breast cancer for different subtypes. The five molecular subtypes in breast cancer in order of disease severity contain: Basal, ERBB2, Luminal B, Luminal A and Normal like respectively. Also in this study genes be as predictive variables. After normalization and remove batch effects, four datasets were aggregated and final data file was consisted of 324 patients with breast cancer in five molecular subtypes and 12 normal tissues.

**Results:** The results of this study showed that among fifteen comparative groups: in the three groups, lasso model best performance than adaptive lasso and elastic net models. In the one group, adaptive lasso model best performance than lasso and elastic net models. In the four groups, elastic net model best performance than lasso and adaptive lasso models. In the five groups, elastic net and adaptive lasso models performed the same and both best performance than lasso model. In the two groups, three models had an equal performance.

**Conclusion:** Penalized logistic methods presented well performance in breast cancer and according to studies, it was determined common genes obtained from three models are also involved in other cancers and the most important genes in breast cancer have been reported in other studies.

**Keywords:** Breast Cancer, Penalized Regression, Lasso Logistic Regression, Adaptive Lasso Logistic Regression, Elastic net





**Kerman University of Medical Sciences**

**School of Public Health**

In Partial Fulfillment of the requirements for the Degree MSc in Biostatistics

Title

**The Comparison of Lasso logistic Regression, adaptive lasso logistic regression, elastic net logistic regression methods in order to determine the effective and important genes of breast cancer in women**

By

**Sadegh Raoufi**

Supervisor

**Dr. Tania Dehesh**

Advisor

**Dr. Saeideh Jafari Nejad**

Thesis No

Date: July 2020

- 
1. Sperandei S. Understanding logistic regression analysis. *Biochemia medica: Biochemia medica*. 2014;24(1):12-8.
  2. Breiman L. Heuristics of instability and stabilization in model selection. *The annals of statistics*. 1996;24(7):2350-83.
  3. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996;58(1):267-88.
  4. Liu C, Jiang J, Gu J, Yu Z, Wang T, Lu H. High-dimensional omics data analysis using a variable screening protocol with prior knowledge integration (SKI). *BMC systems biology*. 2016;10(4):118.
  5. Fonti V, Belitser E. Feature selection using lasso. *VU Amsterdam Research Paper in Business Analytics*. 2017;1-25.
  6. Zou H. The adaptive lasso and its oracle properties. *Journal of the American statistical association*. 2007;102(477):1478-29.
  7. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*. 2005;67(2):301-20.
  8. Azamjah N, Soltan-Zadeh Y, Zayeri F. Global Trend of Breast Cancer Mortality Rate: A 25-Year Study. *Asian Pacific journal of cancer prevention: APJCP*. 2019;20(7):2015.
  9. Zhang Y, Zhang B, Lu W. Breast cancer histological image classification with multiple features and random subspace classifier ensemble. *Knowledge-Based Systems in Biomedicine and Computational Life Science: Springer*; 2013. p. 27-42.
  10. Tekmal RR, Keshava N. Role of MMTV integration locus cellular genes in breast cancer. *Front Biosci*. 1997;2:d51.27-9
  11. Lotfi E, Keshavarz A. Gene expression microarray classification using PCA-BEL. *Computers in biology and medicine*. 2014;54:180-7.
  12. Perou CM, Sørlie T, Eisen MB, Van De Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *nature*. 2000;406(6797):747.
  13. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron J, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(14):8418-23.
  14. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*. 2001;96(456):1348-60.
  15. Frank LE, Friedman JH. A statistical view of some chemometrics regression tools. *Technometrics*. 1993;35(2):109-35.
  16. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2011;73(3):273-82.
  17. Saleh AME. A ridge regression estimation approach to the measurement error model. *Journal of Multivariate Analysis*. 2014;123:68-84.
  18. Algamal ZY, Lee MH. Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. *Expert Systems with Applications*. 2015;42(23):9326-32.
  19. Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, et al. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. *Journal of translational medicine*. 2019;17(1):119.
  20. Mostafaei S, Kazemnejad A, Azimzadeh Jamalkandi S, Amirhashchi S, Donnelly SC, Armstrong ME, et al. Identification of Novel Genes in Human Airway Epithelial Cells associated with Chronic Obstructive Pulmonary Disease (COPD) using Machine-Based Learning Algorithms. *Scientific reports*. 2018;8(1):10775.

۲۱. Min W, Liu J, Zhang S. Network-regularized sparse logistic regression models for clinical risk prediction and biomarker discovery. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*. 2018;15(3):944-53.
۲۲. Majd TM, Kalantari S, Shahraki HR, Nafar M, Almasi A, Samavat S, et al. Application of sparse linear discriminant analysis and elastic net for diagnosis of IgA nephropathy: Statistical and biological viewpoints. *Iranian biomedical journal*. 2018;22(6):374.
۲۳. Hashemian AH, Asl MG, Shahsavari S, Rezaei M, Shahraki HR. Adaptive LASSO Logistic Regression applied on gene expression of prostate Cancer. *WORLD FAMILY MEDICINE*. 2017;15(7):136.۴۱-
۲۴. Vasquez MM, Hu C, Roe DJ, Chen Z, Halonen M, Guerra S. Least absolute shrinkage and selection operator type methods for the identification of serum biomarkers of overweight and obesity: simulation and application. *BMC medical research methodology*. ۱۵۴:(۱)۱۶;۲۰۱۶.
۲۵. Shahraki HR, Jaberipoor M, Zare N, Hosseini A. The role of 22 genes expression in bladder cancer by adaptive LASSO. *Iranian Journal of Cancer Prevention*. 2016;9.(۶)
۲۶. Raeisi Shahraki H, Pourahmad S, Ayatollahi SMT. Identifying the prognosis factors in death after liver transplantation via adaptive LASSO in Iran. *Journal of Environmental and Public Health*. 2016;2016.
۲۷. Wu S, Jiang H, Shen H, Yang Z. Gene Selection in Cancer Classification Using Sparse Logistic Regression with L1/2 Regularization. *Applied Sciences*. 2018;8(9):1569.
۲۸. Chang C, Sung C-Y, Hsiao H, Chen J, Chen I-H, Kuo W-T, et al. HDMAC: A Web-Based Interactive Program for High-Dimensional Analysis of Molecular Alterations in Cancer. *Scientific reports*. ۲۰۲۰;۱۰.(۱):۱-۱۰.
۲۹. Wen Y, Pi F-H, Guo P, Dong W-Y, Xie Y-Q, Wang X-Y, et al. Sleep duration, daytime napping, markers of obstructive sleep apnea and stroke in a population of southern China. *Scientific reports*. ۲۰۱۶;۶:۳۴۶۸۹.
۳۰. Kim SM, Kim Y, Jeong K, Jeong H, Kim J. Logistic LASSO regression for the diagnosis of breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography. *Ultrasonography*. ۲۰۱۸;۳۷(۱):۳۶.
۳۱. Kohannim O, Hibar DP, Stein JL, Jahanshad N, Hua X, Rajagopalan P, et al. Discovery and replication of gene influences on brain structure using LASSO regression. *Frontiers in neuroscience*. 2012;6:115.
۳۲. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*: Springer Science & Business Media; 2009.
۳۳. Hoerl AE, Kennard RW. Ridge regression—1980: Advances, algorithms, and applications. *American Journal of Mathematical and Management Sciences*. 1981;1(1):5-83.
۳۴. Wang F, Mukherjee S, Richardson S, Hill SM. High-dimensional regression in practice: an empirical study of finite-sample prediction, variable selection and ranking. *Statistics and computing*. ۲۰۲۰;۳۰.(۳):۶۹۷-۷۱۹.
۳۵. Lima E, Davies P, Kaler J, Lovatt F, Green M. Variable selection for inferential models with relatively high-dimensional data: Between method heterogeneity and covariate stability as adjuncts to robust selection. *Scientific reports*. 2020;10(1):1-11.
۳۶. Mostafaei S, Kazemnejad A, Jamalkandi SA, Amirhashchi S, Donnelly SC, Armstrong ME, et al. Identification of Novel Genes in Human Airway Epithelial Cells associated with Chronic Obstructive Pulmonary Disease (COPD) using Machine-Based Learning Algorithms. *Scientific reports*. 2018;8(1):1-۲۰.
۳۷. Samimi G, Ring BZ, Ross DT, Seitz RS, Sutherland RL, O'Brien PM, et al. TLE3 expression is associated with sensitivity to taxane treatment in ovarian carcinoma. *Cancer Epidemiology and Prevention Biomarkers*. 2012;21(2):273-9.



- 38. Susini T, Berti B, Carriero C, Tavella K, Nori J, Vanzi E, et al. Topoisomerase II alpha and TLE3 as predictive markers of response to anthracycline and taxane-containing regimens for neoadjuvant chemotherapy in breast cancer. *OncoTargets and therapy*. 2014;7:2111.
- 39. Yang R-W, Zeng Y-Y, Wei W-T, Cui Y-M, Sun H-Y, Cai Y-L, et al. TLE3 represses colorectal cancer proliferation by inhibiting MAPK and AKT signaling pathways. *Journal of Experimental & Clinical Cancer Research*. 2016;35(1):152.
- 40. Stifani S, Blaumueller CM, Redhead NJ, Hill RE, Artavanis-Tsakonas S. Human homologs of a *Drosophila* Enhancer of split gene product define a novel family of nuclear proteins. *Nature genetics*. 1992;2(2):119-27.
- 41. Kim D-W, Min H, Lee K-H, Kim YJ, Oh D-Y, Jeon YK, et al. High tumour islet macrophage infiltration correlates with improved patient survival but not with EGFR mutations, gene copy number or protein expression in resected non-small cell lung cancer. *British journal of cancer*. 2008;98(6):1118-24.
- 42. Bhargava R, Gerald WL, Li AR, Pan Q, Lal P, Ladanyi M, et al. EGFR gene amplification in breast cancer: correlation with epidermal growth factor receptor mRNA and protein expression and HER-2 status and absence of EGFR-activating mutations. *Modern pathology*. 2005;18(8):1027-33.
- 43. Moroni M, Veronese S, Benvenuti S, Marrapese G, Sartore-Bianchi A, Di Nicolantonio F, et al. Gene copy number for epidermal growth factor receptor (EGFR) and clinical response to antiEGFR treatment in colorectal cancer: a cohort study. *The lancet oncology*. 2005;6(5):279-86.
- 44. Mitsudomi T, Yatabe Y. Epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. *The FEBS journal*. 2010;277(2):301-8.
- 45. Tan X, Chen M. MYLK and MYL9 expression in non-small cell lung cancer identified by bioinformatics analysis of public expression data. *Tumor Biology*. 2014;35(12):12189-200.
- 46. Chen L, Su L, Li J, Zheng Y, Yu B, Yu Y, et al. Hypermethylated FAM5C and MYLK in serum as diagnosis and pre-warning markers for gastric cancer. *Disease markers*. 2012;32(3):195-202.
- 47. Kim S, Kon M, DeLisi C. Pathway-based classification of cancer subtypes. *Biology direct*. 2012;7(1):21.