



دفاع

طراحی و ارائه مدلی هوشمند جهت پیش بینی خطر پر فشاری خون مبتنی بر SNP

Design and presenting an intelligent model for predicting
SNP-based hypertension

استاد راهنما: دکتر مهرداد کارگری

استاد راهنمای دوم: دکتر مریم سادات دانشپور

استاد مشاور: دکتر مهدی اکبرزاده

دانشجو: سید علی لاجوردی

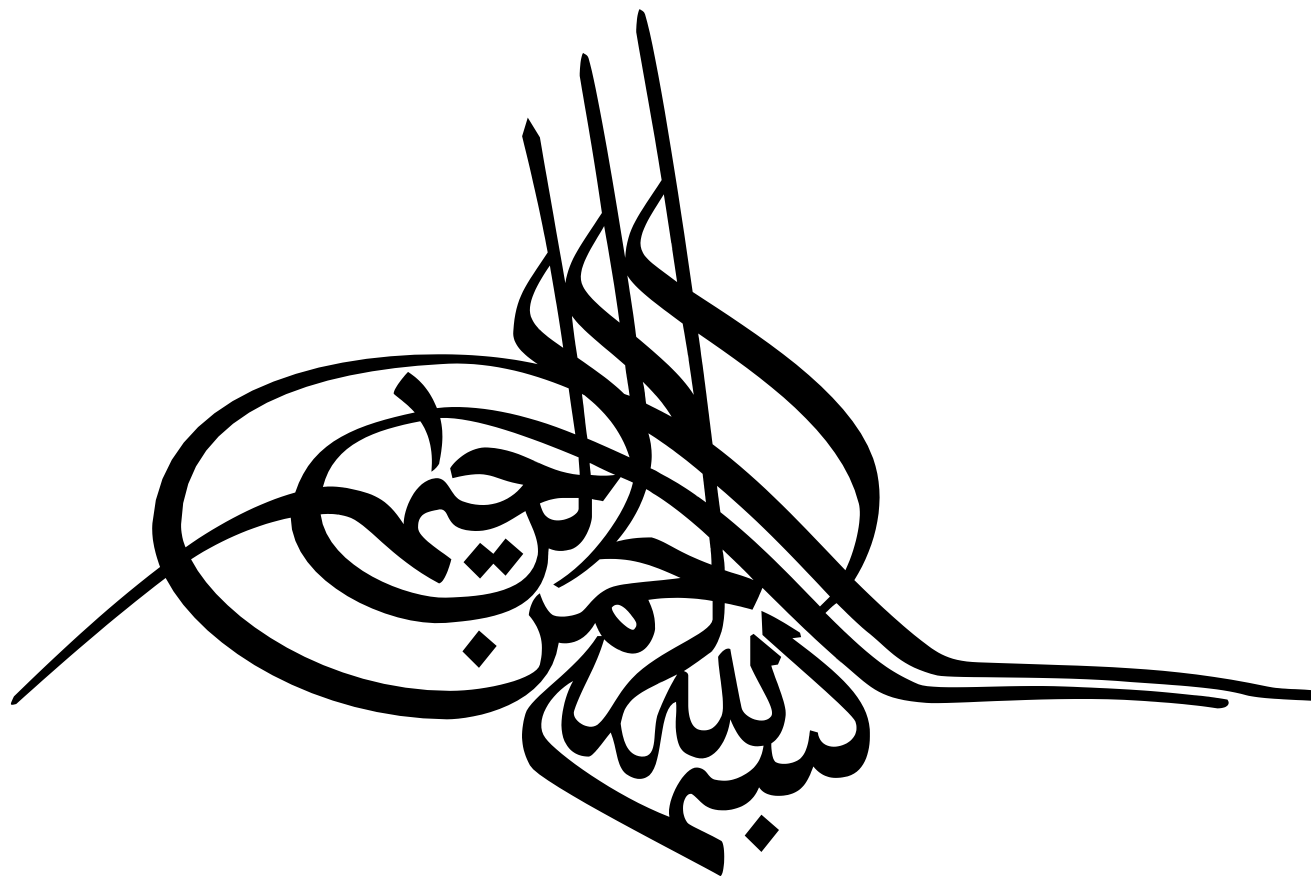
مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی



بحث و جمع بندی

نتایج

روش شناسی

مرور ادبیات

مقدمه

پزشکی شخصی

ژنوم انسان

GWAS

MSA

فشار خون

حوزه پژوهش

ضرورت و نوآوری

رویکرد آماری

رویکرد فراتحلیل

رویکرد یادگیری ماشین

پیش بینی امتیاز خطر

پر فشاری خون

ساختار علمی

یادگیری ماشین

یادگیری عمیق

عدم قطعیت

داده های تحقیق

ارزیابی مدل

روش اجرا

کنترل کیفی داده

اجرای شبکه عمیق

اجرای مبتنی بر SNP

پیش بینی مبتنی بر PRS

مدل سازی فازی

شناسایی SNP مهم

ابزار سنجش خطر HTN

بررسی مدل ها

رتبه بندی SNP

محدودیت ها و پیشنهادات

مقدمه

مرور ادبیات

روش شناسی

نتایج

بحث و جمع بندی

Introduction

مقدمه



پزشکی شخصی

© تاثیر خصوصیات ژنتیکی در کیفیت درمان Rafiei and Amjadi, 2013

© پیشرفت‌های چشمگیر در تشخیص و درمان بیماری بر اساس

تفاوت‌های ژنوم Strachan and Read, 2018

© پیش بینی مبتنی بر یادگیری ماشین راهی برای افزایش دقت

HO, et al., 2019

© پزشکی دقیق

- شخصی سازی
- مراقبت‌های پیشگیرانه
- کارایی درمان
- کاهش هزینه

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

ژنوم انسان

⊙ دنا DNA

⊙ توالی نوکلئوتیدها: AAACCGGT

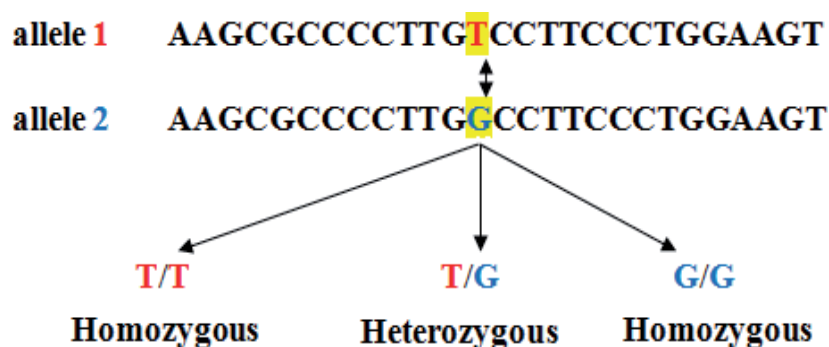
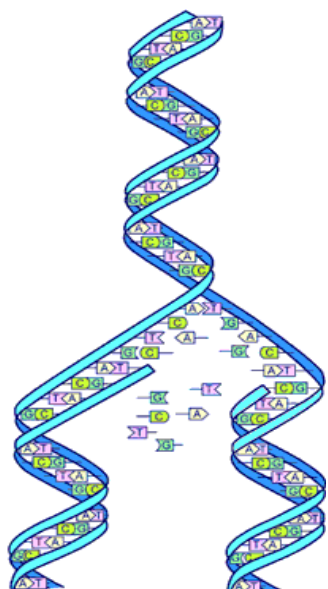
■ گروه فسفات و گروه قند

■ باز نوکلئوتیدی: تشکیل پیوند شیمیایی

■ رشته مکمل معکوس

⊙ صفات دو آللی

⊙ چند ریختی تک نوکلئوتید SNP



Restante. 2018 Çalışkan, Erol and Öz, 2020

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

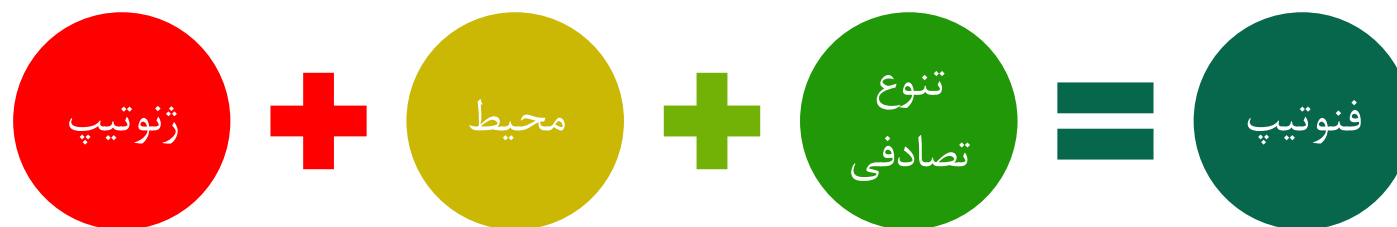
ژن نمود و رخ نمود

◎ ژن نمود: Genotype

▪ اطلاعات حضور آلی یک تغییر ژنتیکی

◎ رخ نمود: Phenotype

▪ خصوصیات قابل مشاهده یا اندازه گیری



Ziegler et al., 2012

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

مطالعه ارتباط گسترده ژنوم GWAS

◎ روش های اولیه کشف ارتباط بین ژنوتیپ و فنوتیپ

■ توارث در خانواده Inheritance

■ مطالعات ژنتیک پیوستگی Genetic linkage

◎ بررسی ژنتیکی بیماری های چند عاملی
Risch & Merikangas, 1996

◎ مطالعات اغلب از نوع مورد شاهدهی

■ تعیین ژنوتیپ

■ میزان فراوانی الی مارکرها

■ گزارش میزان اثر بر مبنای محاسبه نسبت شانس

■ آزمون Chi squared جهت محاسبه p-value
Restante. 2018

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

توالی‌یابی، همترازی و تحلیل توالی

- ◉ فرآیند تعیین دنباله نوکلئوتیدی DNA
- ◉ ژن‌های مؤثر در بیماری‌های ژنتیکی و نحوه فعال شدن آنها
- ◉ میزان مستعد بودن یک فرد را برای یک بیماری
- ◉ اندازه‌گیری شباهت دو دنباله DNA، RNA یا پروتئین
- ◉ هم‌ترازی توالی چندگانه

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع‌بندی



Developing new genetic algorithm based on integer programming for multiple sequence alignment

S. Ali Lajevardy¹ · Mehrdad Kargari¹

Accepted: 11 January 2022 / Published online: 7 March 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Molecular biology advances in the past few decades have contributed to the rapid increase in genome sequencing of various organisms; sequence alignment is usually considered the first step in understanding a sequence's molecular function. Understanding such a function is made possible by aligning an unknown sequence with known sequences based on evolution. An optimal alignment adjusts two or more sequences in a way that it could compare the maximum number of identical or similar residues. The two sequence alignments types are Pairwise Sequence Alignment (PSA) and Multiple Sequence Alignment (MSA). The MSA enjoys a higher advantage than PSA since it predicts the similarity in a family of similar sequences, providing more biological data. Moreover, while the dynamic programming (DP) technique is used in PSA to provide the optimal method, it will lead to more complexity if used in MSA. So, the MSA mainly uses heuristic and approximation methods. Such methods include progressive alignment, iterative alignment, Hidden Markov Model (HMM), and Metaheuristic algorithms. This paper presents a genetic algorithm and chromosome design to solve a mathematical model for MSA. This model uses a basis for an optimal solution in different ways, and an X-mediated matrix with binary elements is used to model the sequence. Then, the model is implemented using the Genetic Algorithm method on the web, and the final results indicate the success of GA for the MSA approach.



مقدمه

مرور
ادبیات

روش
شناسی

نتایج

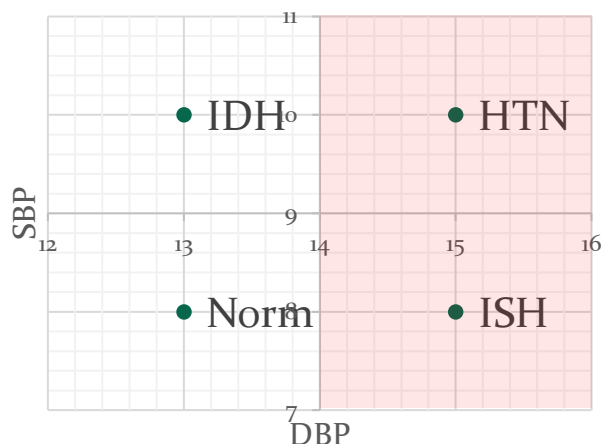
بحث و
جمع بندی

فشار خون

⊙ فشار مایع درون رگ های خون

⊙ فشار خون سیستولیک SBP و فشار خون دیاستولیک DBP

⊙ جدول پرفشاری خون



Yarahmadian, 2018

Guzmán, 2019

Mattoo, 2019

Basile, 2019

پرفشاری خون	سیستولی (mmHg)	دیاستولی (mmHg)
عادی	کمتر از ۱۲۰	کمتر از ۸۰
مرتفع	۱۲۰-۱۲۹	کمتر از ۸۰
سطح ۱	۱۳۰-۱۳۹	۸۰-۹۰
سطح ۲	۱۴۰ و بالاتر	۹۰ یا بالاتر
بحرانی	بیشتر از ۱۸۰	بیشتر از ۱۲۰

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

بیماری پُرفشاری خون

⊙ بیماری های قلبی – عروقی مهم ترین علت مرگ و میر در اکثر نقاط

دنیای و ایران است. نجفی پور و همکاران ۱۳۹۵

⊙ علت دوم مرگ در ایران

Hadaegh, 2009

Loganathan *et al.*, 2019

⊙ عامل مهم در بیماری قلبی عروقی

⊙ تشخیص به هنگام فشارخون برای درمان زودرس Nour and Polat, 2020

⊙ فشارخون سیستولیک

▪ سن، BMI، دور میچ، فعالیت فیزیکی، سابقه خانوادگی و SNP

⊙ فشارخون دیاستولیک

▪ وزن، مصرف الکل، فعالیت فیزیکی، تری گلیسیرید، سابقه خانوادگی و SNP

Li *et al.*, 2019

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

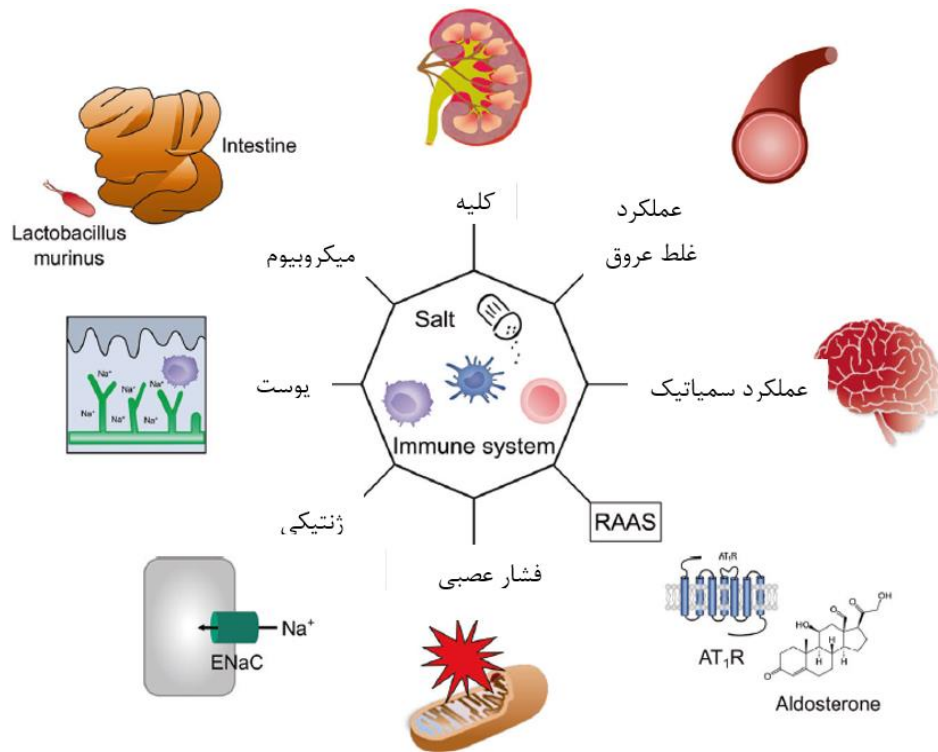
تئوری موازی یک بیماری پر فشاری خون

● شبکه ای متقابل از موارد ایجاد کننده فشار خون

● تاثیر سن: رسوب نمک در پوست

● تاثیر جنسیت: بدون تاثیر مستقیم

● تاثیر سیستم ایمنی بدن



Hengel, Benitah and Wenzel, 2022

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

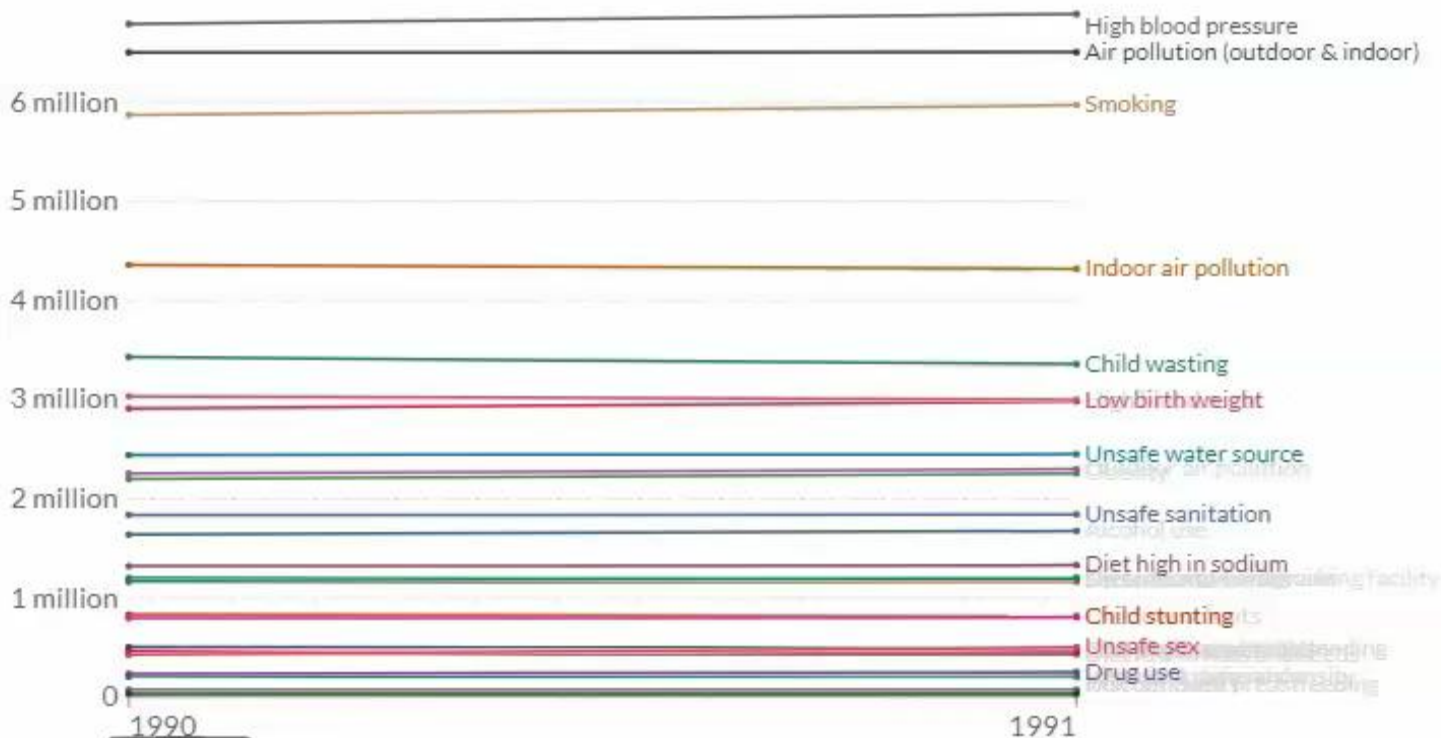
بحث و
جمع بندی

Number of deaths by risk factor, World, 1990 to 1991

Total annual number of deaths by risk factor, measured across all age groups and both sexes.

Our World
in Data

Change country Relative change



Source: IHME, Institute of Health Metrics and Evaluation, Global Burden of Disease (GBD)

OurWorldInData.org/causes-of-death • CC BY

1990 1991 2019

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

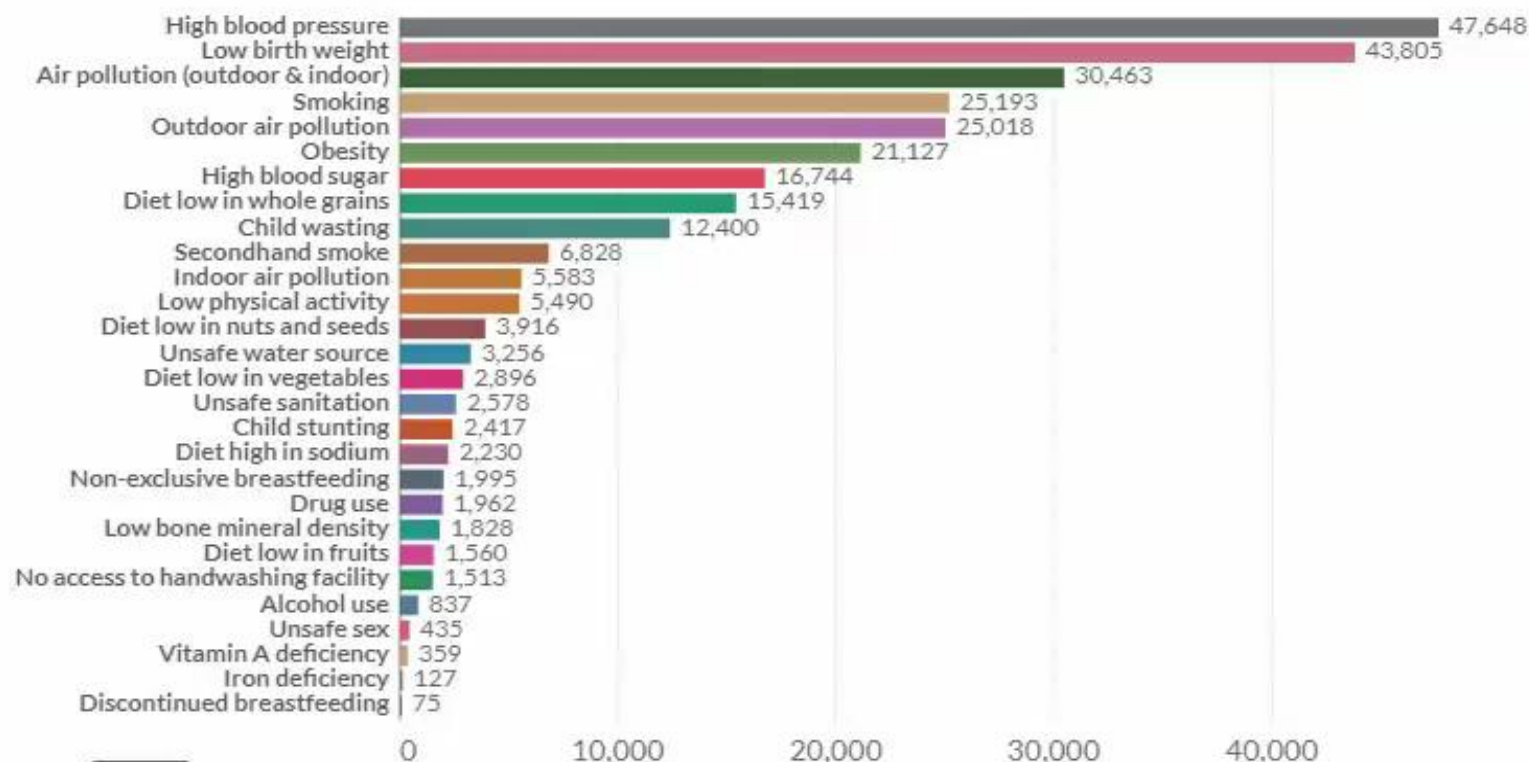
بحث و
جمع بندی

Number of deaths by risk factor, Iran, 1990

Total annual number of deaths by risk factor, measured across all age groups and both sexes.

Our World
in Data

↔ Change country



Source: IHME Global Burden of Disease (GBD)

OurWorldInData.org/causes-of-death • CC BY

1990 2019



مقدمه

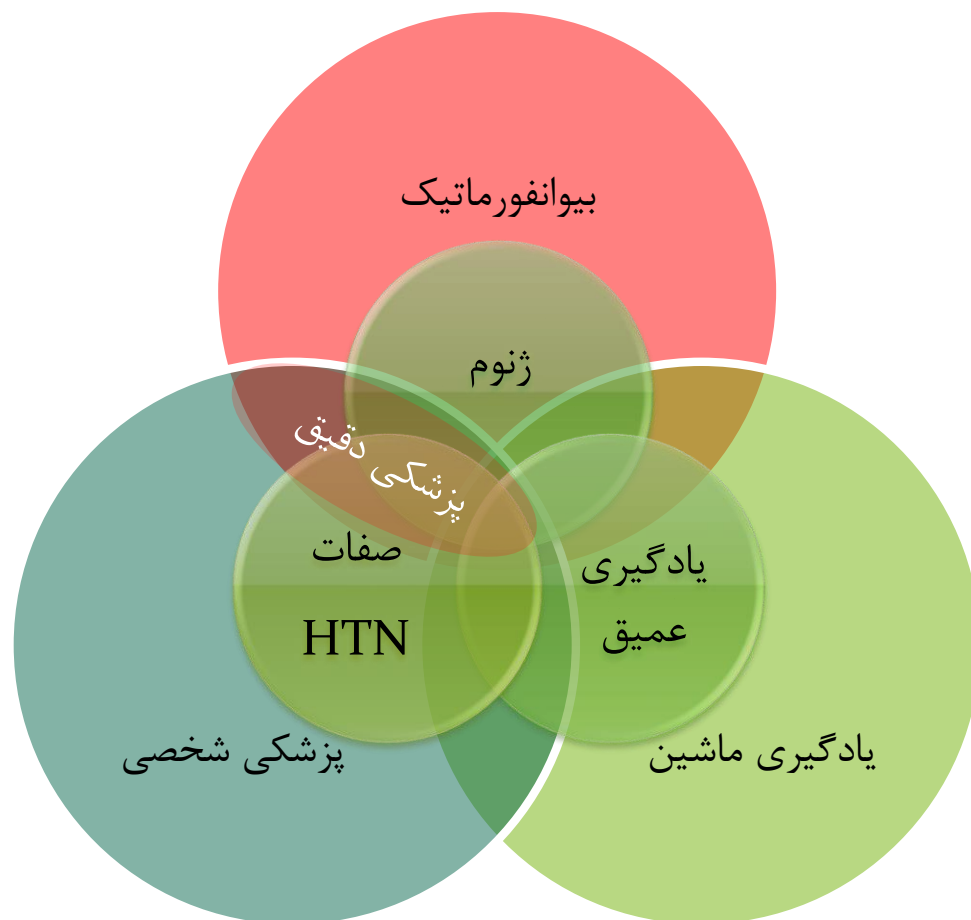
مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

حوزه پژوهش



مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

ضرورت تحقیق

○ وابستگی به واریانس فنوتیپ در نمونه

○ عدم شناسایی تغییرات نادر و گسترده در ژنوم

Korte & Farlow, 2013

○ مشکلات طبقه بندی جمعیت

Visscher. 2014

○ مشکلات اندازه نمونه

Yang. 2017

○ خطای پیش بینی در نمونه

Restante. 2018

○ خطای اشتباه در دقت پیش بینی بیماری و ارتباط ژن-صفت

○ مشکلات محاسباتی

Runcie and Crawford, 2018

○ خطای تاثیر کوچک

Runcie. 2019

○ وابستگی های ژنومی (LD)

Khan *et al.*, 2020

○ وابستگی داده ها

Wu *et al.*, 2020

○ طبقه بندی ناکافی خطر

Phulka *et al.*, 2023

○ خطر بیماری در جوانی

○ کاهش هزینه

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

اهداف پژوهش

◎ ارائه مدل مبتنی بر یادگیری ماشین در داده های حجیم جهت پیش بینی

فنوتیپ پُر فشاری خون بر اساس SNP

▪ فشار سیستولیک SBP و فشار دیاستولیک DBP

▪ پیش آگهی خطر پُر فشاری خون

◎ بهبود مدل با استفاده از داده های طولی

◎ انتخاب SNP های با اهمیت بر اساس دقت و تعداد

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

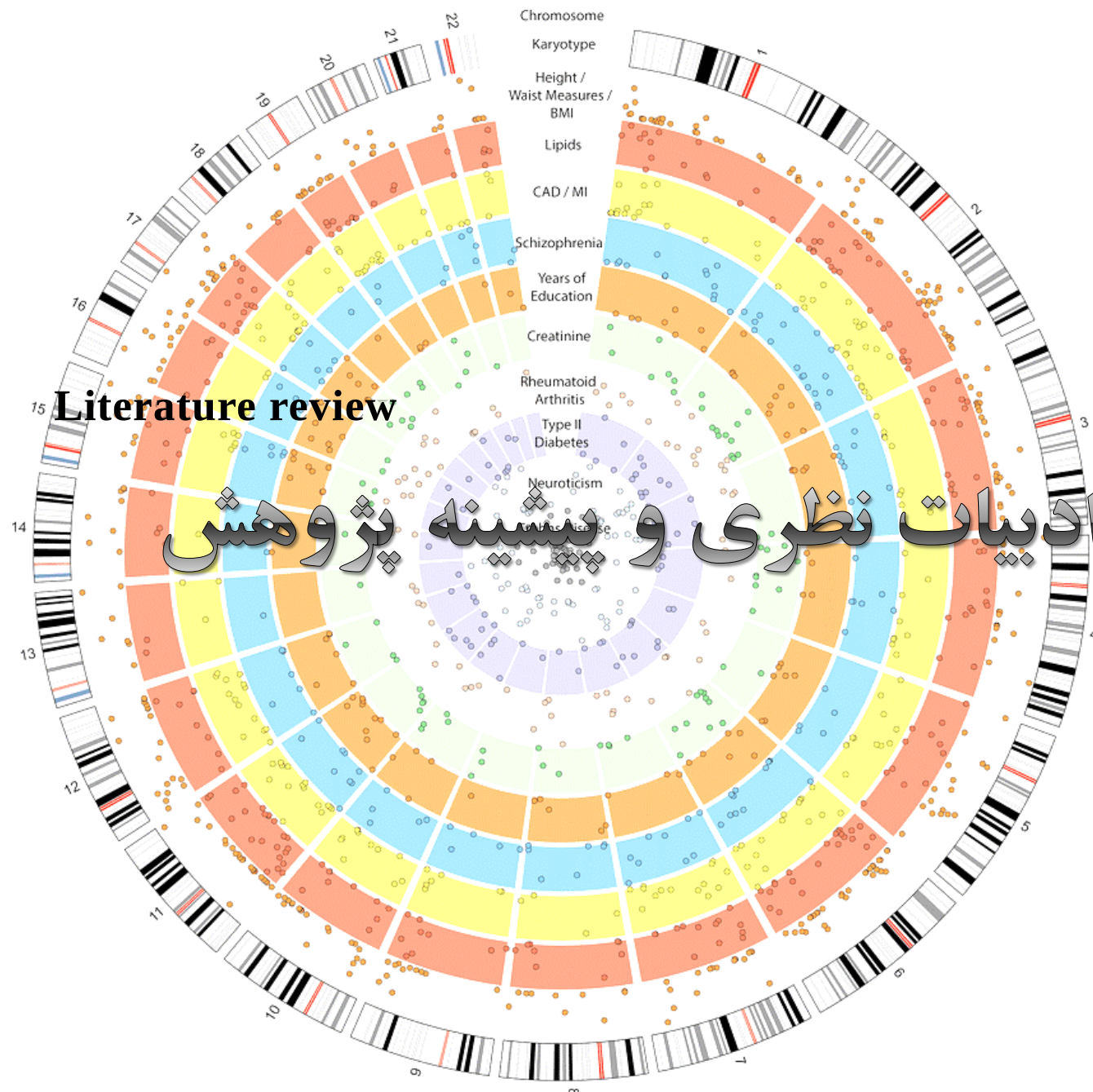
مقدمه

مرور ادبیات

روش شناسی

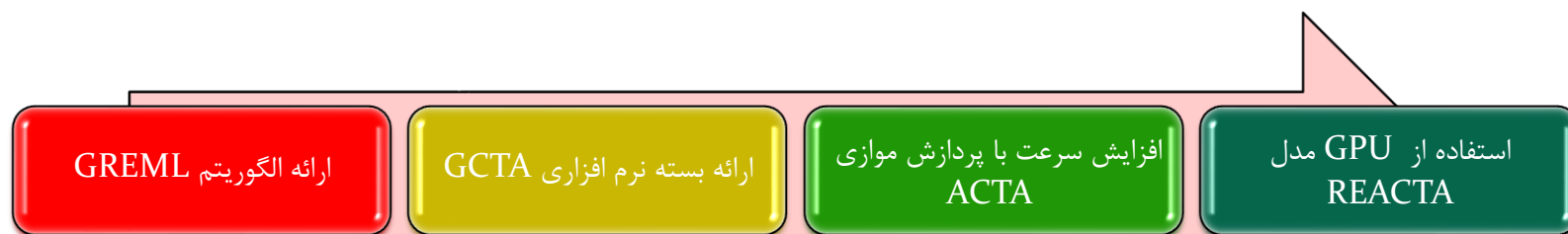
نتایج

بحث و جمع بندی



رویکرد مدل های پیش بینی خطر بیماری

◎ رویکرد آماری



A. Gray, 2012

J. Yang, 2017

Evangelou and Ioannidis, 2013

Willer, Li and Abecasis, 2010

S. Szymczak, 2009

C. J. Willer, 2010

X. Wu, 2020

◎ رویکرد فراتحلیل

◎ رویکرد یادگیری ماشین

◎ رویکرد بالینی

مقدمه

مرور ادبیات

روش شناسی

نتایج

بحث و جمع بندی

پیش بینی امتیاز خطر

Mosley et al., 2020

⊙ خطر بیماری بر اساس همه گیرشناسی

■ عامل های شیوه زندگی

■ سابقه خانوادگی

⊙ امتیازدهی خطر چندرئی PRS

■ کشف (وزنی - غیر وزنی)

■ اعتبار سنجی (آستانه p-value)

■ خطا در بیماری های پیچیده

⊙ یادگیری ماشین

Ho et al., 2019

امتیاز خطر چندرئی	مدل یادگیری ماشین	
✓ سادگی اجرا ✓ سادگی تفسیر	✓ موثر برای داده های چند بعدی ✓ قابل استفاده در داده های پیچیده مرتبط ✓ عدم نیاز به فرض نرمال بودن	نقاط قوت
⊗ تاثیر افزایشی و مستقل فرض نرمال بودن بی استفاده در داده پیچیده مرتبط	⊗ سختی در اجرا ⊗ سختی در تفسیر اجزای موثر ⊗ نیاز به داده زیاد	نقاط ضعف

مقدمه

مرور ادبیات

روش شناسی

نتایج

بحث و جمع بندی

تحلیل ساختار علمی موضوع امتیاز خطر ژنتیکی

© هم رخدادی واژگان

© بررسی ۱۱۷۵ مقاله

■ کلمه کلیدی: Genetic Risk Score

■ پایگاه داده Science Direct

■ بازه سال‌های ۲۰۱۶ تا ۲۰۲۰

ردیف	نام مجله	تعداد مقالات چاپ شده	Impact Factor	5-Year Impact Factor
۱	Gastroenterology	۸۵	۱۹,۲۳۳	۱۸,۸۴۴
۲	European Neuropsychopharmacology	۵۱	۴,۴۶۸	۴,۳۹۴
۳	Journal of the American College of Cardiology	۴۳	۱۸,۶۳۹	۱۹,۰۶۸
۴	Atherosclerosis	۴۱	۳,۹۱۹	۴,۶۳۹
۵	Ophthalmology	۲۱	۷,۷۳۲	۷,۸۴
۶	Alzheimer's & Dementia	۲۱	۱۴,۴۲۳	۱۳,۵۳۳
۷	The American Journal of Human Genetics	۱۷	۱۰,۵۰۲	۱۰,۳۴۴
۸	The Lancet Diabetes & Endocrinology	۱۶	۲۴,۵۴۰	۲۳,۰۱۳
۹	Journal of Clinical Lipidology	۱۶	۳,۸۶۰	۴,۰۸۰
۱۰	Journal of Allergy and Clinical Immunology	۱۵	۱۰,۲۲۸	۱۱,۵۹۲

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

شبکه علمی

genome wide associated studies
single nucleotide polymorphism

atrial fibrillation

body mass

type diabetes

insulin receptor

interleukin

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی



6th International Conference on Industrial and Systems Engineering
9 & 10 Sep. 2020 • Ferdowsi University of Mashhad

2020

ششمین کنفرانس بین المللی
مهندسی صنایع و سیستم‌ها

۱۹ و ۲۰ شهریور ۱۳۹۹ • دانشگاه فردوسی مشهد



مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

تحلیل ساختار علمی موضوع امتیاز خطر ژنتیکی با استفاده از روش تحلیل هم‌رخدادی

واژگان مبتنی بر مقالات پایگاه ScienceDirect

سیدعلی لاجوردی^۱، سیدمیثم علوی^۲، مهرداد کارگری^۳

^۱ دپارتمان مهندسی فناوری اطلاعات دانشگاه تربیت مدرس؛ Sayedali.lajevardy@modares.ac.ir

^۲ دپارتمان مهندسی فناوری اطلاعات دانشگاه تربیت مدرس؛ Meysamalavi@modares.ac.ir

^۳ دپارتمان مهندسی فناوری اطلاعات دانشگاه تربیت مدرس؛ M_kargari@modares.ac.ir

* نویسنده مسئول: مهرداد کارگری

چکیده

در سال‌های اخیر، تلاش برای کشف و شناسایی تغییرات ژنتیکی و تأثیر آن بر بروز و پیشگیری بیماری‌ها یکی از اهداف اساسی علم پزشکی بوده است. با توجه به اینکه توصیه‌ها و دستورات پزشکی عموماً بر اساس خصوصیات میانگین افراد تنظیم می‌شود، نمی‌تواند برای همه افراد کاربرد داشته باشد. امروزه مقوله پزشکی شخصی‌سازی شده برای طراحی توصیه و درمان‌های پزشکی بر پایه داده‌های بالینی و ژنتیکی افراد مورد توجه فراوان قرار گرفته است. یکی از موضوعات بسیار مهم در این خصوص معیار «امتیاز خطر ژنتیکی» است. که کاربرد زیادی در زمینه پزشکی شخصی‌سازی شده دارد. یکی از مسائل و دغدغه‌های محققین این حوزه شناخت حوزه‌های پژوهشی فعال در این خصوص و یافتن حوزه‌های پژوهشی آتی است. لذا داشتن درک و نمایی کلی از چارچوب علمی حوزه مورد نظر، یکی از راهپای است که محققین را در نیل به اهداف پژوهشی آن حوزه کمک می‌کند. در این راستا ترسیم نقشه یا ساختار علمی حوزه مورد نظر، امری ضروری به نظر می‌رسد. از سوی دیگر تحلیل شبکه‌های اجتماعی در سال‌های اخیر به عنوان ابزاری جهت بررسی روابط حال و آینده بین موجودیت‌های یک ساختار شبکه‌ای نظر محققین علوم مختلف را برای تحلیل این روابط و ترسیم نقشه علمی یک حوزه از علم به خود معطوف کرده است. در این مقاله با استفاده از هم‌رخدادی واژگان و تحلیل شبکه‌های اجتماعی، نقشه و ساختار علمی گستره «امتیاز خطر ژنتیکی» بر اساس مقالات بازه پنج ساله ۲۰۱۶ تا ۲۰۲۰ نمایه شده در پایگاه ScienceDirect، ترسیم و روند موضوعی حاکم بر این حوزه بررسی شده است. نتایج تحلیل نشان می‌دهد بیماری‌های مرتبط با انواع دیابت (۲۶،۳۲٪)، بیماری‌های قلبی (۱۳،۲۹٪)، بیماری‌های خونی (۶،۳۱٪) و فشار خون (۵،۳۴٪) از منظر مسائل خطر ژنتیکی بیشترین سهم تحقیقات را به خود اختصاص داده‌اند.

پیش بینی خطر پرفشاری خون

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

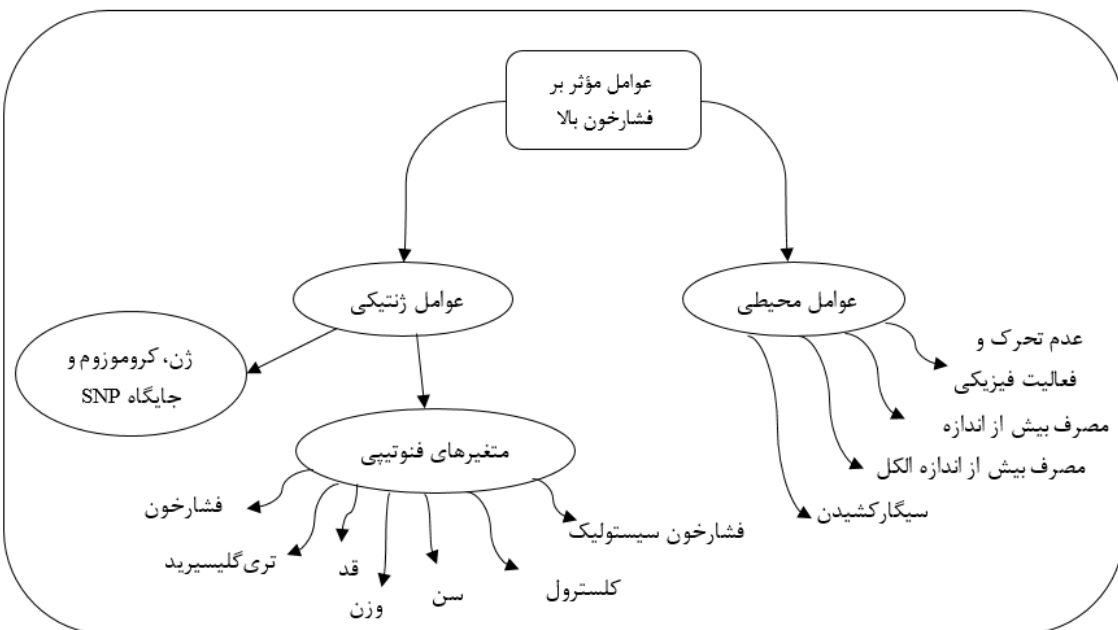
25

◎ پارامترهای محیطی

◎ پارامترهای فنوتیپی

◎ اطلاعات ژنتیکی

■ چند SNP موثر



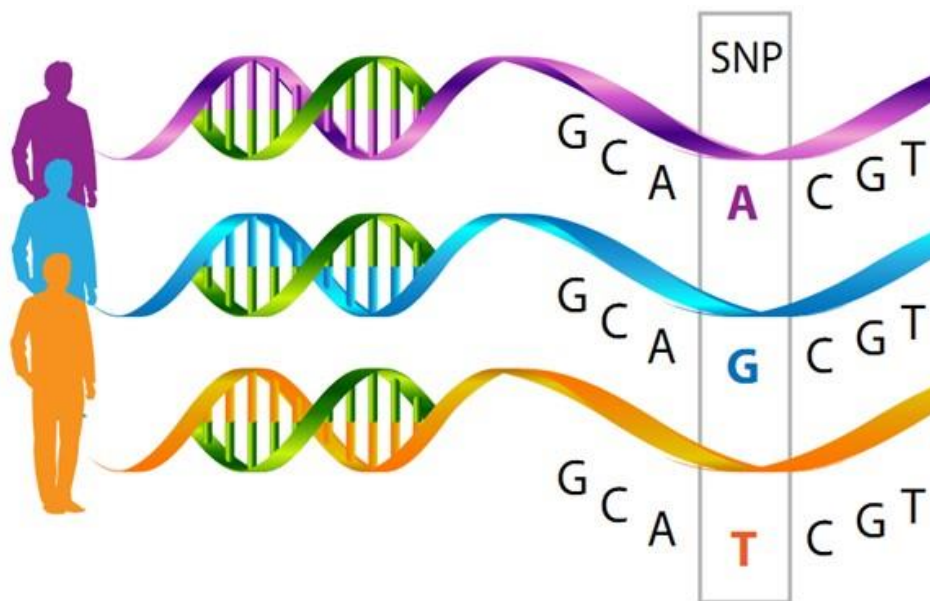
Giontella et al., 2021; Kurniansyah et al., 2022; Maj et al., 2022

Niu et al., 2021; Chowdhury et al., 2022; Wu et al., 2020;

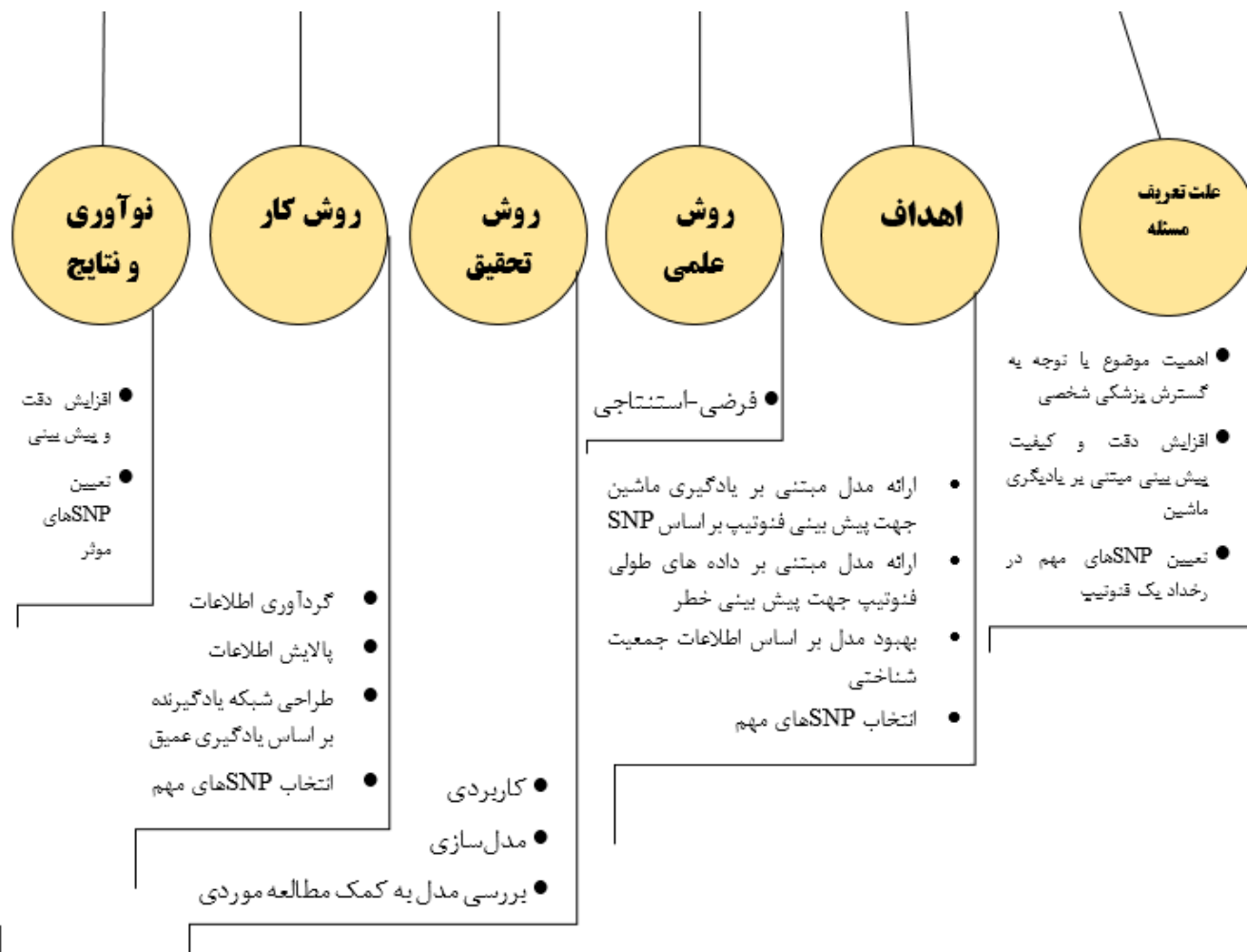
Li et al., 2019 Nour and Polat, 2020 Kim and Andrade, 2019 Wang et al., 2020

Methodology

روش شناسی



آونگ تحقیق



مقدمه

مرور ادبیات

روش شناسی

نتایج

بحث و جمع بندی

مدل های یادگیری ماشین

متغیر های مدل

■ برچسب فشار خون

■ SNP های مرتبط مبتنی بر GWAS: ۳۷۸ SNP

○ GRCh38.p13

○ $MLOG \geq 6$

○ EFO_1001132

■ سن دوره ای

○ مدل های استفاده شده

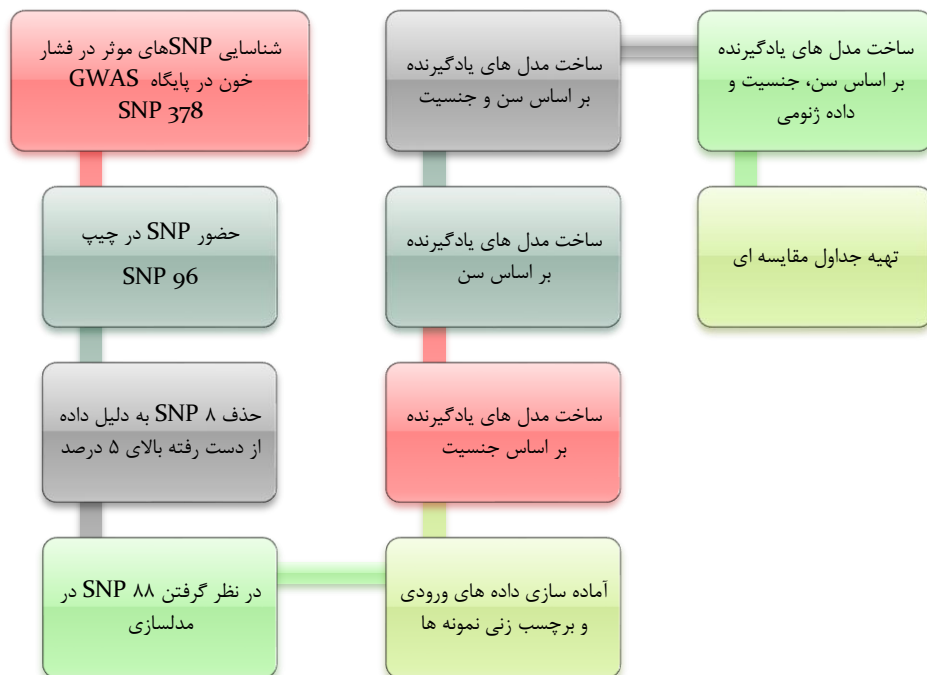
■ شبکه عصبی

■ رگرسیون خطی

■ جنگل تصادفی

■ درخت تصمیم

■ ماشین بردار پشتیبان



مقدمه

مرور ادبیات

روش شناسی

نتایج

بحث و جمع بندی

شناسایی SNP مهم

◎ روش ها

Information gain ■

Gain ratio ■

Gini index ■

χ^2 ■

Relief ■

FCBF ■

◎ میانگین رتبه به عنوان امتیاز هر SNP

◎ مرتب سازی بر مبنای امتیاز و شناسایی Top10

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

مسیر اجرایی یادگیری ماشین

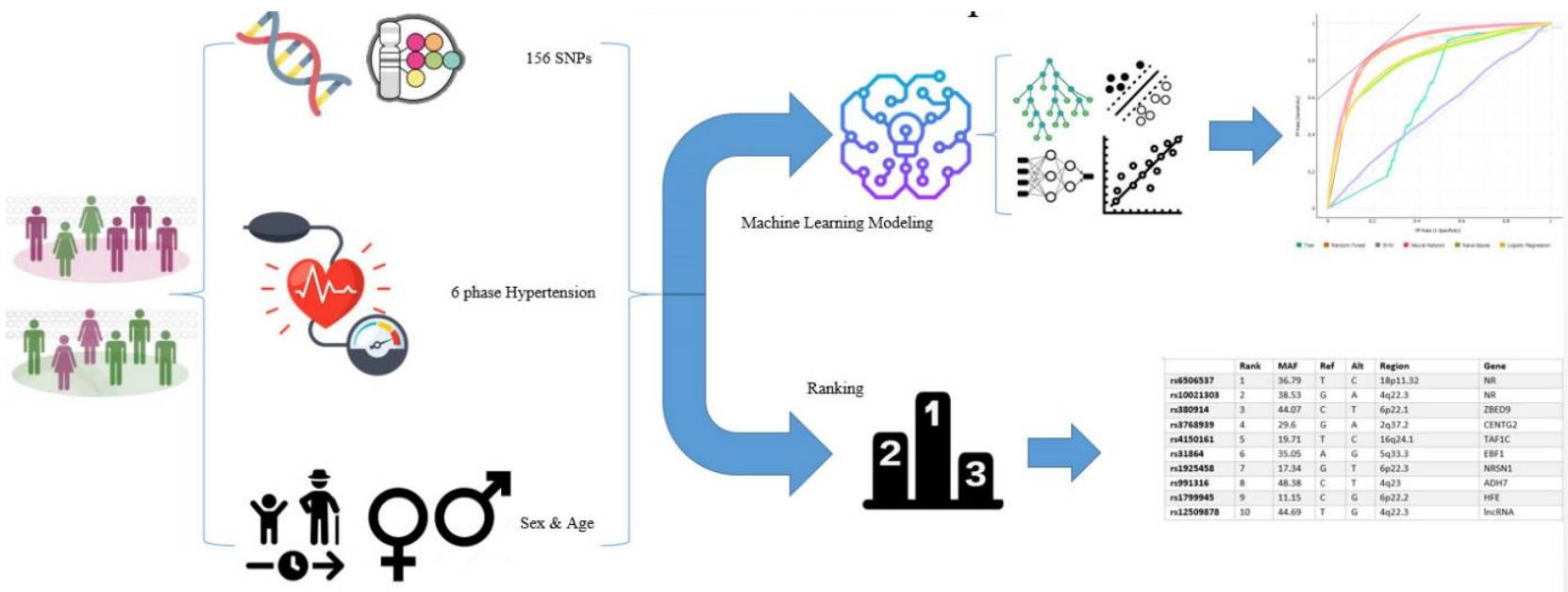
مقدمه

مرور ادبیات

روش شناسی

نتایج

بحث و جمع بندی



RESEARCH ARTICLE

Hypertension Risk Prediction Based on SNPs by Machine Learning Models

S. Ali Lajevardi¹, Mehrdad Kargari^{1,*}, Maryam S. Daneshpour² and Mahdi Akbarzadeh²

¹Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran; ²Cellular and Molecular Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

ARTICLE HISTORY

Received: March 29, 2022
Revised: July 22, 2022
Accepted: September 09, 2022

DOI:
10.2174/1574893617666221011093322

Abstract: Background: Hypertension is one of the most significant underlying ailments of cardiovascular disease; hence, methods that can accurately reveal the risk of hypertension at an early age are essential. Also, one of the most critical personal health objectives is to improve disease prediction accuracy by examining genetic variants.

Objective: Therefore, various clinical and genetically based methods are used to predict the disease; however, the critical issue with these methods is the high number of input variables as genetic markers with small samples. One approach that can be used to solve this problem is machine learning.

Methods: This study was conducted on the participants' genetic markers in the 20-year research of cardiometabolic genetics in Tehran (TCGS). Various machine learning methods were used, including linear regression, neural network, random forest, decision tree, and support vector machine. The top ten genetic markers were identified using importance-based ranking methods, including information gain, gain ratio, Gini index, χ^2 , relief, and FCBF.

Results: A model based on a neural network with AUC of 89% was presented. This model has an accuracy and an f-measure of 0.89, which shows the quality. The final results indicate the success of the machine learning approach.

Conclusion: Study shows machine learning approach helps predict the risk of hypertension at a young age and finds significant SNPs that affect HTN.

Keywords: Hypertension risk, machine learning, SNP markers, TCGS, cardiovascular disease, genetic markers.

مقدمه

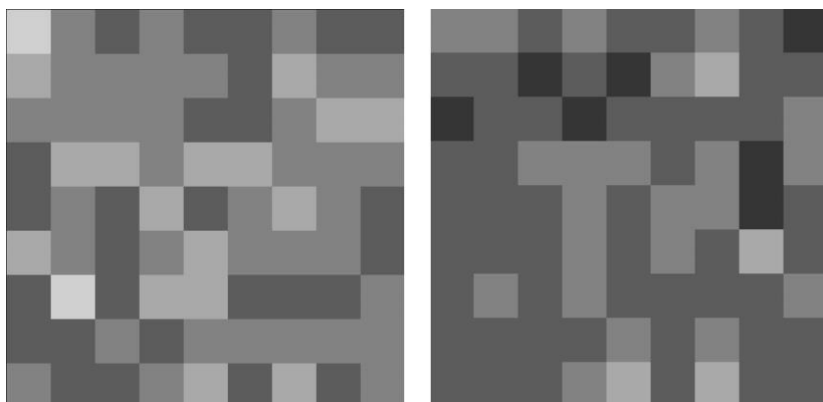
مرور
ادبیاتروش
شناسی

نتایج

بحث و
جمع بندی

شبکه عمیق CNN/MAXP/FCN

◎ تبدیل داده ژنومی (SNP 600k) به یک تصویر فشرده



◎ شبکه CNN/MaxP

- فیلتر ۲۴ گانه 4×4
- MaxPool 2×2
- اجرای Normalize داده

◎ مدل دسته بند

- شبکه عصبی، رگرسیون خطی، جنگل تصادفی
- درخت تصمیم، ماشین بردار پشتیبان

مقدمه

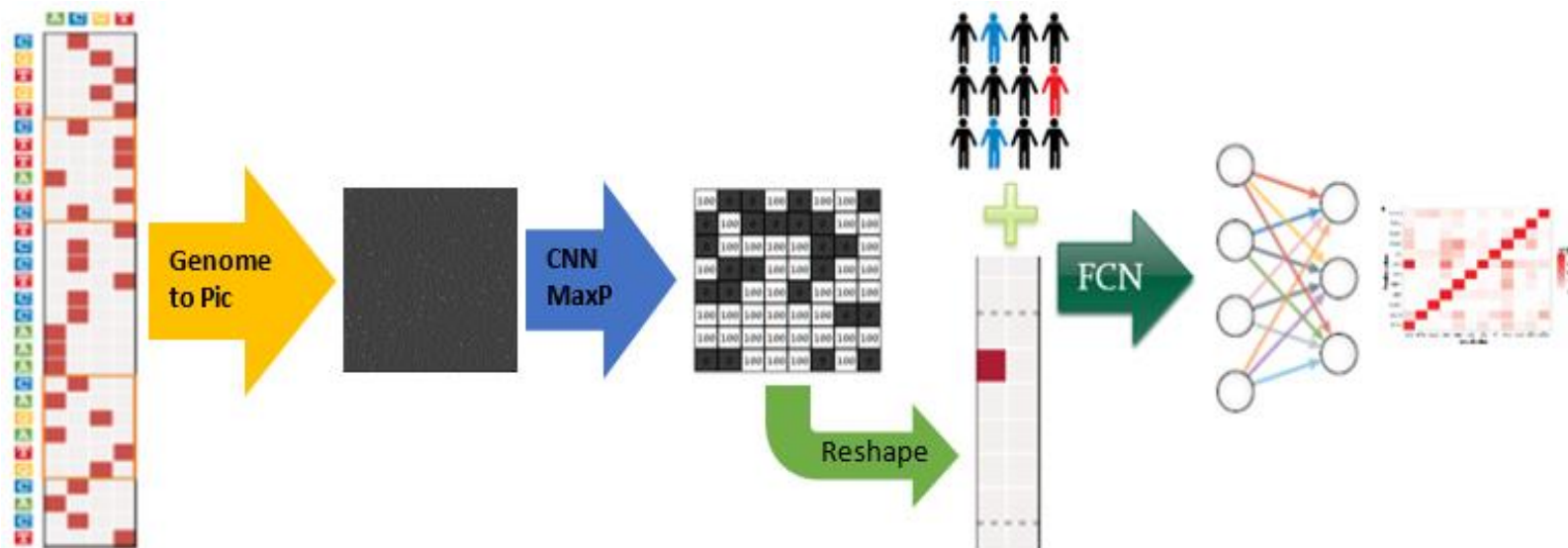
مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

مسیر اجرایی یادگیری عمیق



مقدمه

مرور ادبیات

روش شناسی

نتایج

بحث و جمع بندی

محاسبه امتیاز خطر ژنتیکی مبتنی بر PRS

◎ با ابزارهای GCTA و PLINK

◎ ۱۰ قسمت (فولد) تصادفی

```
plink.exe --ped test.ped --map test.map --noweb --make-bed --out test  
gcta64 --bfile test --make-grm --out test --thread-num 10 --reml --pheno test.phen  
gcta64 --reml --grm test --pheno test.phen --reml-pred-rand --out test  
gcta64 --bfile test --blup-snp test.indi.blp --out test  
plink --score test.snp.blp --noweb --bfile test --pheno test.phen --out test  
gcta64 --reml --grm test --pheno test.phen --out test --cvblup
```

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی



ارائه مدل یادگیری عمیق CNN با هدف بهبود دقت پیش بینی خطر بیماری پرفشاری خون بر مبنای SNP

سیدعلی لاجوردی^۱، مهرداد کارگری^۲، مریم السادات دانشپور^۳، مهدی اکبرزاده^۴

۱- دانشگاه تربیت مدرس - دانشکده صنایع و سیستم ها

۲- دانشگاه تربیت مدرس - دانشکده صنایع و سیستم ها

۳- دانشگاه علوم پزشکی شهید بهشتی - پژوهشکده علوم غدد درون ریز و متابولیسم

۴- دانشگاه علوم پزشکی شهید بهشتی - پژوهشکده علوم غدد درون ریز و متابولیسم

چکیده

فشارخون عامل قابل اصلاح برای بیماری های قلبی عروقی مانند بیماری اسکمیک قلب است که از دلایلی اصلی مرگ در سراسر جهان است که با نام کشنده خاموش معرفی می شود. از این رو، روش هایی که می توانند به طور دقیق خطر ابتلا به فشار خون بالا را در سنین پایین آشکار کنند، ضروری هستند. مدل های معمول پیش بینی خطر بیماری در درجه اول مبتنی بر عامل های شیوه زندگی است. اخیراً، در نظر گرفتن خطر عامل های ژنتیکی از جمله SNP های مرتبط با بیماری همراه با مدل سازی خطر، دقت پیش بینی فردی بیماری ها را بهبود بخ شیده است. SNP یک تغییر ژنتیکی کوچک در DNA است و رایج ترین نوع تنوع ژنتیکی است که در میان افراد رخ می دهد. مدل های مختلفی که از نشانگر های ژنومی جهت پیش بینی بیماری فشار خون استفاده می کنند شامل چهار رویکرد آماری، فرا تحلیل، یادگیری ماشین و مدل بالینی هستند. مهم ترین م ساله در این مدل ها تعداد بالای SNP های ورودی و ارتباط این SNP ها با

داده های تحقیق

⊙ داده های ژنومی (Chip 600K SNP)

⊙ داده های بالینی سن، جنسیت، مصرف دارو

⊙ داده های طولی فنوتیپ فشار خون

❖ طرح ژنتیک کاردیو متابولیک تهران پژوهشکده علوم غدد درون ریز و

متابولیسم دانشگاه علوم پزشکی شهید بهشتی

Azizi, et al., 2009

Azizi, et al., 1396

Kheradmand et al., 2015

❖ در مجموع، ۳۹۵۹ خانواده شامل ۱۳۸۸۷ نفر

❖ IR.SBMU.ENDOCRINE.REC.1400.008

مقدمه

مرور
ادبیات

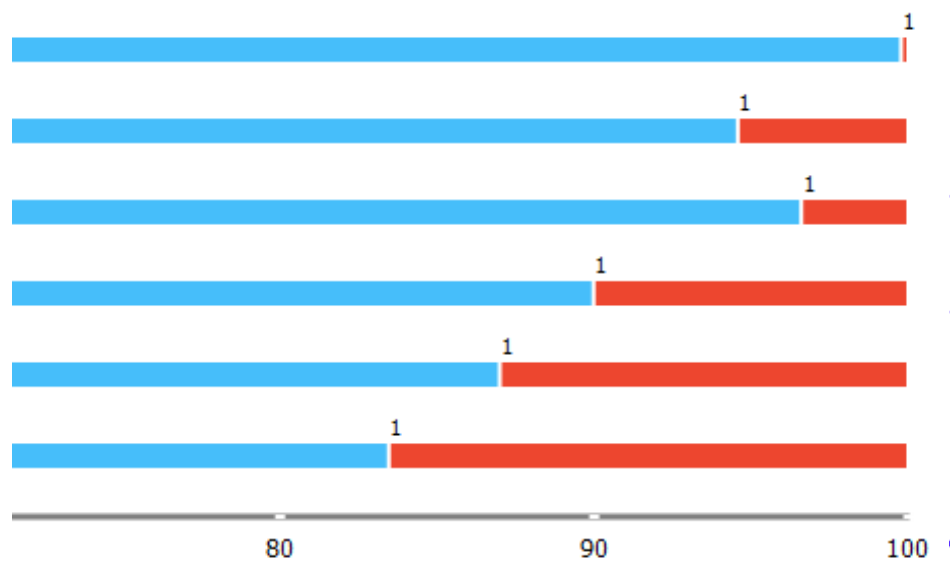
روش
شناسی

نتایج

بحث و
جمع بندی

توصیف داده در دوره ها

3: 111,63 ± 18.95



○ سن

○ جنسیت

○ DBP

○ SBP

○ مصرف دارو

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

ارزیابی مدل

⊙ ماتریس درهم ریختگی

		برچسب پیش بینی شده	
		مثبت	منفی
برچسب شناخته شده	مثبت	TP	FN
	منفی	FP	TN

Powers and Ailab, 2011

Fawcett, 2006

- دقت
- حساسیت
- خاصیت
- معیار اف

⊙ منحنی مشخصه عملکرد سیستم

- منحنی ROC
- مقدار AUC

⊙ روش های نظام بخشی

- روش k-fold
- تمامی فازهای هر فرد در یک fold

مقدمه

مرور ادبیات

روش شناسی

نتایج

بحث و جمع بندی

روش اجرا

◎ پیاده سازی کد

Shell ■

PHP ■

Python ■

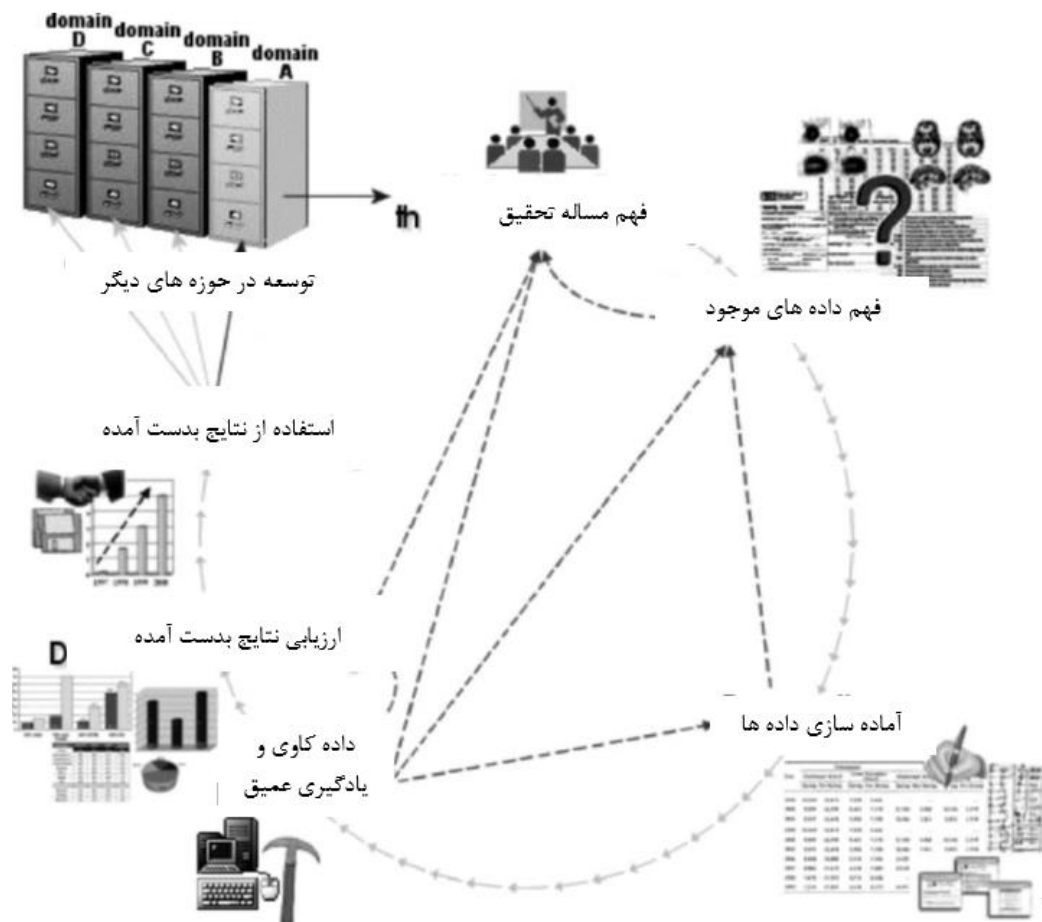
◎ آماده سازی و

مصور سازی

Orange ■

Excel ■

SPSS ■



مقدمه

مرور ادبیات

روش شناسی

نتایج

بحث و جمع بندی

مقدمه

مرور
ادبیات

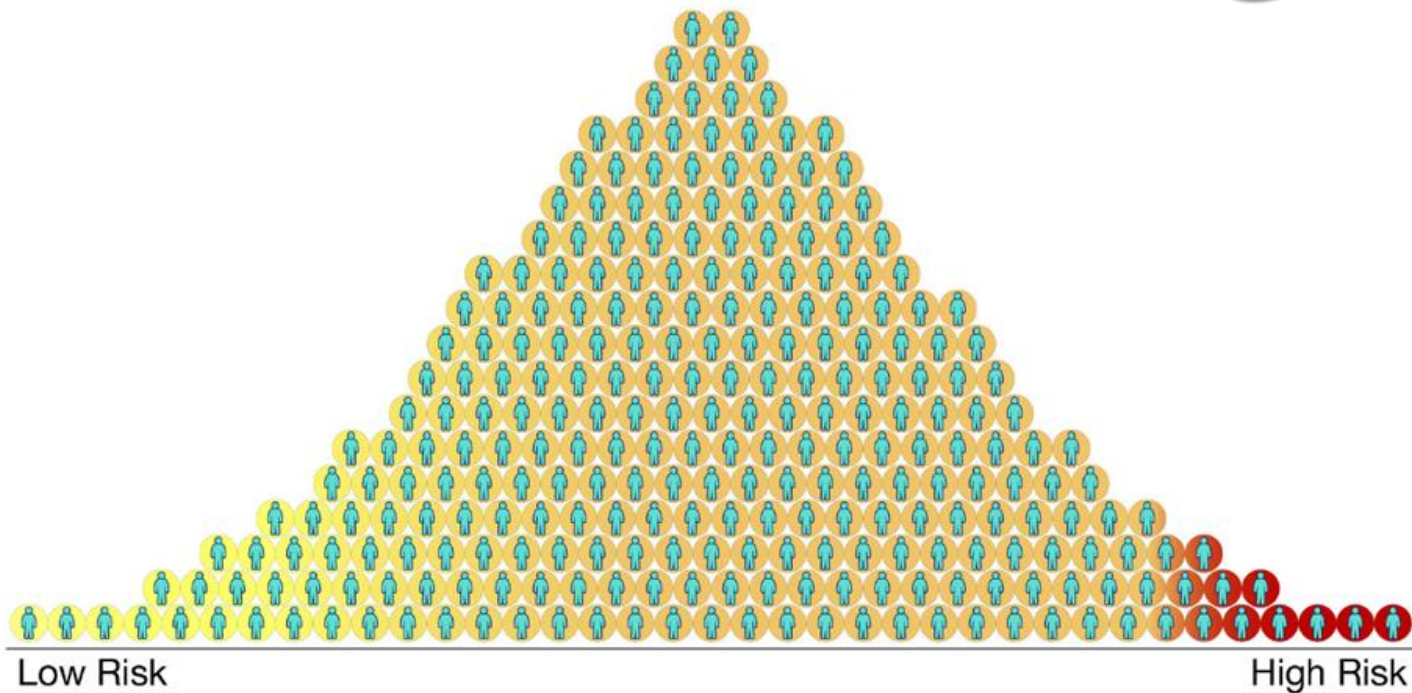
روش
شناسی

نتایج

بحث و
جمع بندی

Result

نتایج



آماده سازی داده

- بررسی فایل SNP ها شامل بیش از ۶۰۰ هزار SNP
- بررسی فایل مطالعه کوهورت شامل اطلاعات سن و جنس
- گسسته سازی داده ها

مجموع	652919	100.0%
Miss<5	647357	99.1%
MAF>5	571860	87.6%

- سن
- جنس

- برچسب زنی داده بر اساس مقادیر فشار خون
- حذف داده های دارای بیش از ۵ درصد گمشدگی
- حذف SNP با MAF کمتر از ۵ درصد

مقدمه

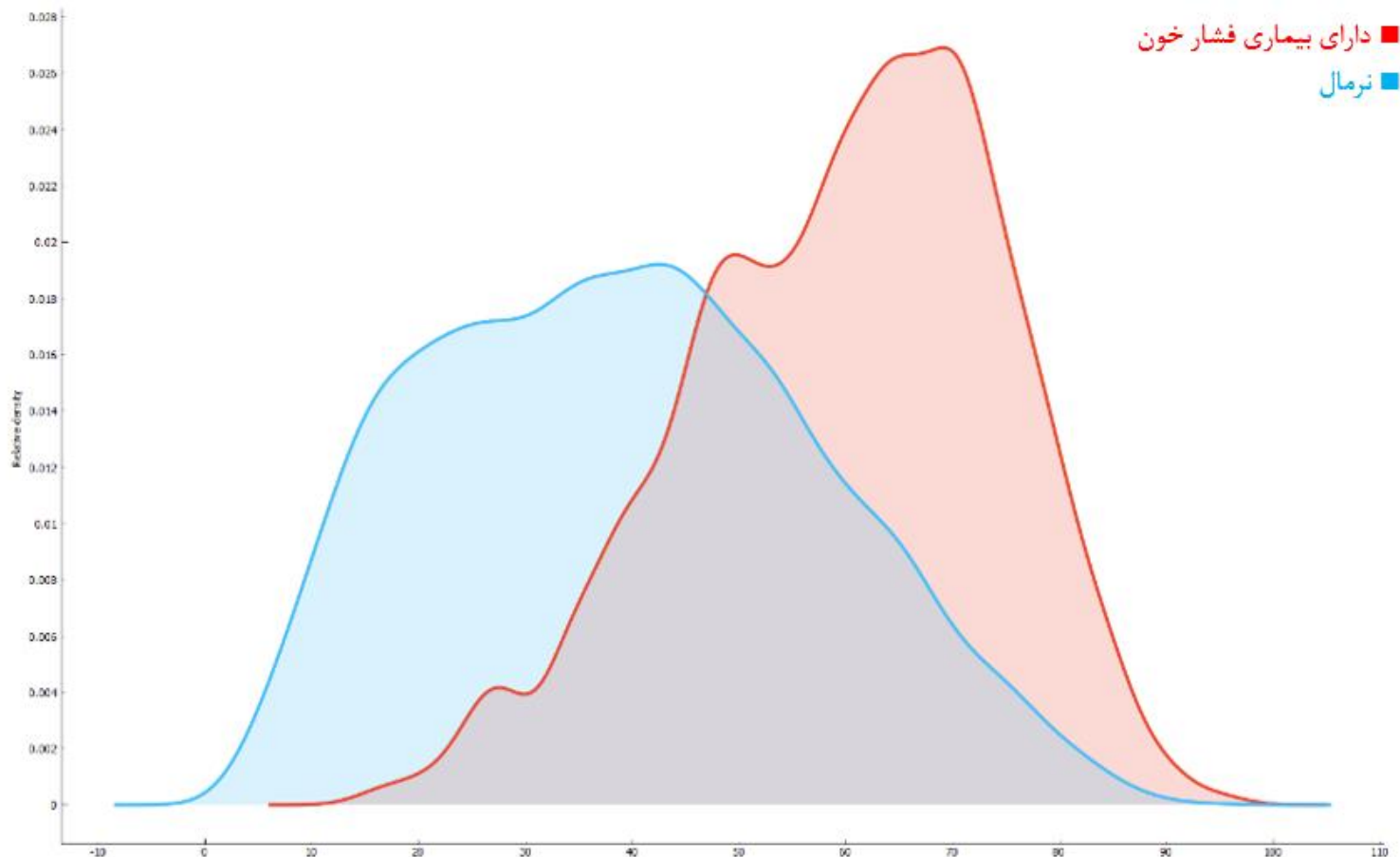
مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

توزیع سنی بر اساس برچسب فشارخون



مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

نتایج مدل سازی یادگیری ماشین

روش	مدل	AUC	CA	F ₁	Precision	Recall
Random Forest	جنسیت	0.507	0.507	0.505	0.507	0.507
	سن	0.806	0.741	0.741	0.742	0.741
	سن و جنسیت	0.809	0.74	0.74	0.743	0.74
	جنسیت، سن و داده ژنومی	0.868	0.888	0.881	0.878	0.888
Neural Network	جنسیت	0.504	0.505	0.505	0.505	0.505
	سن	0.808	0.743	0.742	0.746	0.743
	سن و جنسیت	0.814	0.742	0.741	0.745	0.742
	جنسیت، سن و داده ژنومی	0.885	0.891	0.889	0.887	0.891
Logistic Regression	جنسیت	0.507	0.507	0.505	0.507	0.507
	سن	0.808	0.743	0.743	0.743	0.743
	سن و جنسیت	0.808	0.743	0.743	0.744	0.743
	جنسیت، سن و داده ژنومی	0.811	0.853	0.819	0.817	0.853
Tree	جنسیت	0.507	0.507	0.505	0.507	0.507
	سن	0.806	0.741	0.74	0.742	0.741
	سن و جنسیت	0.81	0.741	0.741	0.743	0.741
	جنسیت، سن و داده ژنومی	0.610	0.852	0.844	0.838	0.852
SVM	جنسیت	0.5	0.503	0.498	0.503	0.503
	سن	0.397	0.303	0.303	0.303	0.303
	سن و جنسیت	0.549	0.53	0.498	0.54	0.53
	جنسیت، سن و داده ژنومی	0.565	0.810	0.780	0.759	0.810

مقدمه

مرور ادبیات

روش شناسی

نتایج

بحث و جمع بندی

نتایج نهایی

مقدمه

مرور ادبیات

روش شناسی

نتایج

بحث و جمع بندی

ML (NN Best)						Top10	Fuzzy	PRS	CNN	تعریف	داده	نوع
	*	*	*	*	*	*	*	*	*	مرد/زن	جنسیت	متغیر
*		*	*	*		*			*	سال دوره ای	سن	
							*			سن فازی		
	*							*		سن تعیین شده		
									*	تصویر 600K	ژنوم	
								*		600K GRM		
				*		*	*			GWAS SNP		
			*									
*	*	*	*	*	*	*	*	*	*	HTN	فشار خون	برچسب
0.804	0.726	0.81	0.885	0.817	0.542	0.83	0.866	0.71	0.877	AUC		نتایج
0.883	0.747	0.884	0.891	0.771	0.884	0.861	0.88	0.74	0.901	CA		
0.838	0.639	0.834	0.889	0.768	0.829	0.832	0.878	0.65	0.896	F1		
0.839	0.6	0.841	0.887	0.767	0.781	0.834	0.876	0.65	0.893	Precision		
0.883	0.747	0.884	0.891	0.771	0.884	0.861	0.88	0.74	0.901	Recall		

لیست TOP10

		TCGS			GWAS				
		MAF	Ref	Alt	TRAIT	POS	Gene	RISK ALLELE	P-VALUE
rs6506537	1	36.79	T	C	Hypertension	18p11.32	NR	rs6506537-T	4.00E-09
rs10021303	2	38.53	G	A	Hypertension	4q22.3	NR	rs10021303-A	7.00E-06
rs380914	3	44.07	C	T					
rs3768939	4	29.6	G	A	Cardiovascular disease in hypertension (calcium channel blocker interaction)	2q37.2	CENTG2	rs3768939-A	9.00E-07
rs4150161	5	19.71	T	C	Systolic blood pressure response to hydrochlorothiazide in hypertension	16q24.1	TAF1C	rs4150161-T	1.00E-06
rs31864	6	35.05	A	G	Hypertension	5q33.3	EBF1	rs31864-A	3.00E-06
rs1925458	7	17.34	G	T	Alzheimer's disease in hypertension	6p22.3	NRSN1	rs1925458-?	1.00E-06
rs991316	8	48.38	C	T	Hypertension	4q23	ADH7	rs991316-T	5.00E-06
rs1799945	9	11.15	C	G	Hypertension	6p22.2	HFE	rs1799945-G	2.00E-10
rs12509878	10	44.69	T	G	Diastolic blood pressure night-to-day ratio in hypertension	4q22.3	lncRNA	rs12509878-C	6.00E-06

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

ابزار پیش بینی خطر TOP10

● بر اساس ضرایب ماشین مدلسازی Top10

● با دقت ۸۶ درصد

● محاسبه خطر بیماری با لغزنده سن و تعیین وضعیت آلل

RS	Status	Age
rs6506537	☒ Ref ☐ Het ☐ Hom	 50 Calculated HTN Risk 61.3%
rs1021303	☐ Ref ☒ Het ☐ Hom	
rs380914	☐ Ref ☒ Het ☐ Hom	
rs3768939	☐ Ref ☒ Het ☐ Hom	
rs4150161	☐ Ref ☐ Het ☒ Hom	
rs31864	☐ Ref ☒ Het ☐ Hom	
rs1925458	☒ Ref ☐ Het ☐ Hom	
rs991316	☐ Ref ☐ Het ☒ Hom	
rs1799945	☐ Ref ☒ Het ☐ Hom	
rs12509878	☒ Ref ☐ Het ☐ Hom	

مقدمه

مرور
ادبیات

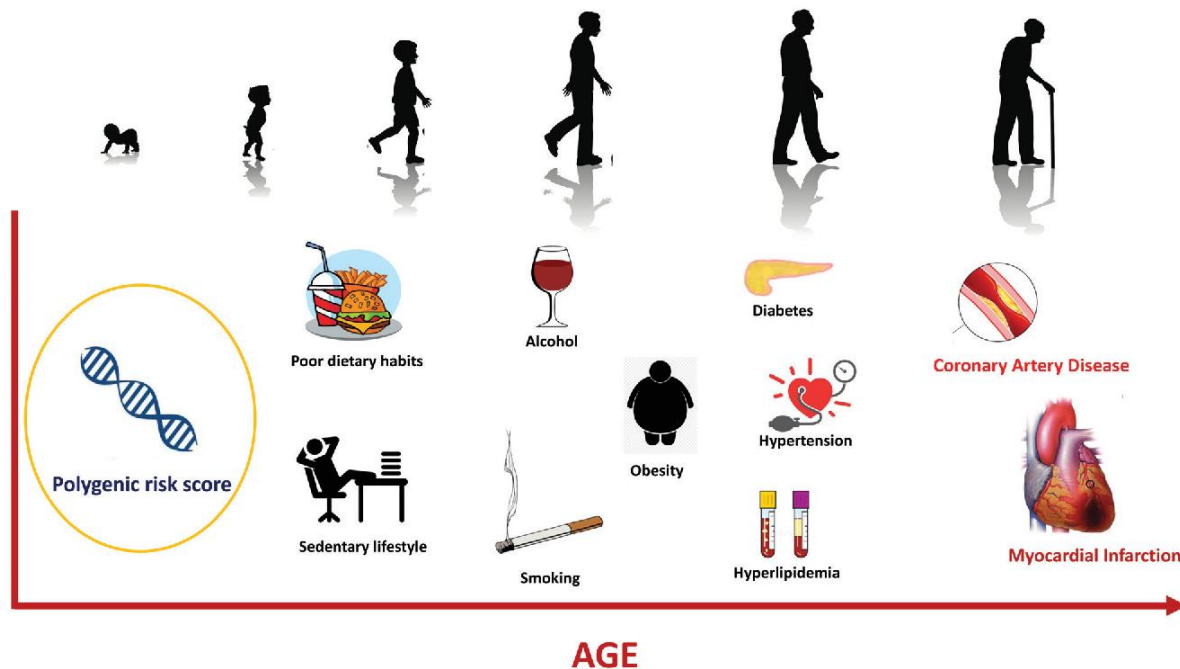
روش
شناسی

نتایج

بحث و
جمع بندی

Discussion and conclusion

بحث و نتیجه گیری



بررسی مدل ها

⊙ در نظر گرفتن نشانگرهای ژنومی، جنسیت و سن

⊙ عدم تاثیر گذاری جنسیت در پرفشاری خون

⊙ پیش بینی زود هنگام خطر در جوانی

⊙ درخت تصمیم

■ انشعاب اول: سن ۵۰ سالگی

■ انشعاب دوم در شاخه کم خطر: سن ۱۸ سالگی

■ انشعاب دوم در شاخه پر خطر: سن ۵۹ سالگی

■ انشعاب تا سطح ۴ بر اساس سن

■ عدم تفاوت معنی دار SNP های موثر پرفشاری خون در زن و مرد

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

رتبه بندی SNP

● عدد p-value معیار مناسبی برای رتبه نیست
Sullivan and Feinn, 2012

● لیست TOP10 و ارتباط با p-value

● سطح معنی داری مناسب برای SNP

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

محدودیت ها و پیشنهادات آتی

◎ محدودیت ها

- اعتبار سنجی خارجی
- ترکیب قومی کم

◎ پیشنهادات

- تاثیر نژاد و دیگر پارامترهای مستقل از سن
- مدل سازی IDH و ISH
- استفاده از الگوریتم های LightGBM و XGBoost و Ensemble های دیگر

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

مقدمه

مرور
ادبیات

روش
شناسی

نتایج

بحث و
جمع بندی

خرداد ۱۴۰۲

با تشکر از حسن توجهتان