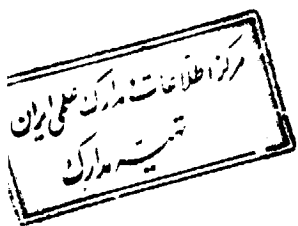


۲۴.۹.



۱۳۷۹ / ۸ / ۲۰



دانشگاه تربیت مدرس

دانشگاه تربیت مدرس

دانشکده علوم پزشکی

### پایان نامه

جهت دریافت درجه کارشناسی ارشد رشته آمار حیاتی

موضوع :

کاربرد تحلیل ممیزی برای شناسایی بیماران دیابتی تهران

۱۱۱۱۹۲

نگارش :

مرجان رکابدار

استاد راهنما:

جناب آقای دکتر محمود محمودی

استاد مشاور:

جناب آقای دکتر انوشیروان کاظم نژاد

تابستان ۱۳۷۹

۳۴۰۹۰

«فرم تأییدیه اعضای هیأت داوران مندرج در پایان نامه کارشناسی ارشد»

بدینوسیله پایان نامه کارشناسی ارشد خانم / آقای مرجان رکابدار  
تحت عنوان کاربرد تحلیل ممیزی برای شناسایی بیماران دیابتی شهر تهران  
تقدیم می شود. اینجانبان نسخه نهائی این پایان نامه را از نظر فرم و محتوی بررسی و تأیید کرده و پذیرش آنرا برای  
تکمیل درجه کارشناسی ارشد پیشنهاد می کنیم.

نام و نام خانوادگی و امضاء اعضای هیأت داوران:

آقای دکتر محمود محمودی (استاد راهنما)

آقای دکتر انوشیروان کاظم نژاد (استاد مشاور)

آقای دکتر غلامرضا بابایی (نماینده تحصیلات تکمیلی)

آقای دکتر یداله محرابی (استاد ناظر)

آقای دکتر سقراط فقیه زاده (استاد ناظر)



بسمه تعالی

## آیین نامه چاپ پایان نامه (رساله) های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان نامه (رساله) های تحصیلی دانشجویان دانشگاه تربیت مدرس، مبین بخشی از فعالیتهای علمی - پژوهشی دانشگاه است بنابراین به منظور آگاهی و رعایت حقوق دانشگاه، دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می شوند:

ماده ۱ در صورت اقدام به چاپ پایان نامه (رساله) ی خود، مراتب را قبلاً به طور کتبی به «دفتر نشر آثار علمی» دانشگاه اطلاع دهد.

ماده ۲ در صفحه سوم کتاب (پس از برگ شناسنامه)، عبارت ذیل را چاپ کند:

کتاب حاضر، حاصل پایان نامه کارشناسی ارشد / رساله دکتری نگارنده در رشته آمار حیاتی است که در سال ۱۳۷۹ در دانشکده علوم پزشکی دانشگاه تربیت مدرس به راهنمایی سرکار خانم / جناب آقای دکتر محمد محمودی، مشاوره سرکار خانم / جناب آقای دکتر انزلیان سلم ترادو مشاوره سرکار خانم / جناب آقای دکتر — از آن دفاع شده است.

ماده ۳ به منظور جبران بخشی از هزینه های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت چاپ) را به «دفتر نشر آثار علمی» دانشگاه اهدا کند. دانشگاه می تواند مازاد نیاز خود را به نفع مرکز نشر در معرض فروش قرار دهد.

ماده ۴ در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ شده را به عنوان خسارت به دانشگاه تربیت مدرس، تأدیه کند.

ماده ۵ دانشجو تعهد و قبول می کند در صورت خودداری از پرداخت بهای خسارت، دانشگاه می تواند خسارت مذکور را از طریق مراجع قضایی مطالبه و وصول کند؛ به علاوه به دانشگاه حق می دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده ۴ را از محل توقیف کتابهای عرضه شده نگارنده برای فروش، تأمین نماید.

ماده ۶ اینجانب رجان طایب دانشجوی رشته آمار حیاتی مقطع کارشناسی ارشد تعهد فوق و ضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

نام و نام خانوادگی: رجان طایب

تاریخ و امضا:

۱۱/۴/۸۸

تقديم به

تمامی عزیزانم

## تقدیر و تشکر

" من لم يشكر المخلوق لم يشكر الخالق "

با سپاس و ستایش به درگاه ذات احدیت که توفیق علم اندوزی و انجام تحقیق حاضر را فراهم نمود، لازم است مراتب سپاسگزاری خویش را از کلیه خانمها و آقایانی که به نحوی در طول انجام تحقیق با اینجانب مساعدت، همکاری و همیاری داشته‌اند، اعلام نمایم.

سپاس و قدردانی خویش را از استاد گرانقدر جناب آقای دکتر محمود محمودی استاد راهنما که در طول تحصیل و تحقیق افتخار شاگردی ایشان را داشته‌ام و از راهنماییها و ارشادات ایشان بهره جسته‌ام، اعلام می‌دارم و توفیقات روزافزون ایشان را از درگاه ایزدمنان خواستارم.

در جریان تحقیق از هدایتهای استاد مشاور جناب آقای دکتر انوشیروان کاظم‌نژاد بهره‌مند بودم و بدین وسیله صمیمانه سپاس و امتنان خویش را از ایشان ابراز می‌دارم.

همچنین از هیأت داوران جناب آقای دکتر بابایی و جناب آقای دکتر فقیه‌زاده و جناب آقای دکتر محرابی که زحمت مطالعه تحقیق و ارزشیابی آن را متقبل شدند از صمیم قلب سپاسگزاری می‌نمایم.

در پایان شایسته است از سرکار خانم گلزار که زحمت حروفچینی پایان نامه را متقبل شدند و همه دوستان و افرادی که به نحوی در انجام تحقیق حاضر اینجانب را یاری و مساعدت نمودند صادقانه قدردانی نمایم و توفیقات روزافزون ایشان را از خداوند متعال مسئلت نمایم.

مرجان رکابدار

تابستان ۱۳۷۹

## چکیده

دیابت یا بیماری قند، بیماری است که سالیان دراز گریبانگیر بشر می‌باشد و در سراسر دنیا پراکنده است و در تمام سنین دیده می‌شود و انواع گوناگونی دارد که در صورت شناخت و کنترل و درمان، بیماری کنترل‌پذیر بوده و در افراد شناخته شده مشکلات عمده‌ای ایجاد نمی‌کند. هم‌اینک بیش از ۱۰۰ میلیون نفر در جهان مبتلا به دیابت هستند که از سال ۱۹۸۷ میلادی آمار افراد دیابتی سه برابر افزایش یافته است که به پیش‌بینی فدراسیون بین‌المللی دیابت تا سال ۲۰۱۰ به دو برابر کنونی می‌رسد. یکی از اهداف آمارحیاتی، شناسایی مشکلات و نیازهای بهداشتی جامعه و ارائه روشهای قابل اجرا برای حل آن می‌باشد و هدف دیگر آن استقرار سلامتی و حفظ آن است. با توجه به اینکه برای تشخیص دیابت بخصوص حالتی که فرد در جامعه هیچگونه علائم بیماری را ندارد، تحقیق گسترده‌ای انجام پذیرفته است و مطالعات گذشته صرفاً طرحهای مقطعی بودند که هدفی جز تعیین ریسک فاکتورهای بیماری دیابت را نداشتند، برآن شدیم که با همکاری مرکز تحقیقات نمود بیمارستان دکتر شریعتی و به کمک مدل‌های آماری تابع تشخیص بهترین روش در راستای شناسایی این دسته از بیماران را پیدا نماییم.

در این مطالعه که به صورت مورد-شاهدی انجام شده است از هر فرد ۶ متغیر شامل جنس، سن، سابقه فامیلی، شاخص حجم بدن (BMI)، شاخص چاقی شکمی (WHI) و تعداد موارد حاملگی جمع‌آوری گردید. افرادی که در آنها قند خون ناشتا بزرگتر یا مساوی  $140 \frac{mg}{dl}$  و قند خون ساعت دوم آنها بزرگتر یا مساوی  $200 \frac{mg}{dl}$  باشد فرد دیابتی شناخته شده (گروه مورد) و درغیراین صورت به عنوان گروه شاهد تعریف گردیده است. کل افراد مورد بررسی ۲۸۲ نفر بوده‌اند که از این تعداد ۸۲ نفر مبتلا به دیابت تیپ II بوده‌اند. در بین ۶ متغیر اندازه‌گیری شده از دو گروه مورد و شاهد دو متغیر جنسیت و سابقه فامیلی از نظر آماری ارتباط معنی‌داری با بروز بیماری نداشته است.

در آنالیز چندمتغیره تابع تشخیص مدل نهایی استاندارد شده با استفاده از ۶ متغیر پیشگو به صورت برآورد گردید.

$$Z = 0.54729age + 0.38588BMI + 0.41985WHR + 0.46109pregnancy - 0.29376X_7$$

که نشان می دهد متغیرهای سن، شاخص حجم بدن و تعداد موارد حاملگی و شاخص چاقی شکمی و سابقه فامیلی به ترتیب دارای بیشترین نقش در ممیزی هستند.

ضریب همبستگی مثبت بین متغیرهای سن، BMI، WHR و pregnancy حاکی از آن است که افزایش در مقدار این متغیرها شانس قرارگرفتن فرد در گروه مورد(بیمار) را بیشتر می کند. و برعکس ضریب همبستگی  $r = -0.20415$  برای متغیر سابقه فامیلی ( $X_7$ ) نشان می دهد که افرادی که در خویشاوندان خود افراد دیابتی را ندارند شانس قرارگرفتن آنها در گروه شاهد(سالم) بیشتر است.

درصد خطای طبقه بندی برای این مدل زمانی که احتمالهای پیشین متناسب با اندازه نمونه است مینیمم مقدار خود یعنی ۸۲/۹٪ است. اما از آنجایی که هدف شناسایی بیماران و به عبارتی دیگر پیدا کردن مدلی با حساسیت بالاست که بیماران را درست شناسایی نماید بنابراین نقطه بهینه در احتمالات پیشین مساوی رخ می دهد یعنی جایی که درصد طبقه بندی درست در کل نمونه تقریباً نزدیک به حالت قبل است (۸۱/۱٪) و از طرفی حساسیت و ویژگی مدل تقریباً با یکدیگر مساوی شده و افزایش یافته است.

گل واژه : تابع تشخیص - بروز - شیوع - خطر نسبی - حساسیت - ویژگی.



## فهرست مطالب

صفحه	عنوان
۱	فصل اول: کلیات
۲	۱-۱- مقدمه
۳	۲-۱- اهداف پژوهش
۳	۳-۱- فرضیه‌هایی تحقیق
۳	۴-۱- سؤالات پژوهشی
۴	۵-۱- اهمیت پژوهش
۵	۶-۱- نوع مطالعه
۶	۷-۱- جمعیت و واحد مورد بررسی
۶	۸-۱- محاسبه حجم نمونه و روش جمع‌آوری اطلاعات
۸	۹-۱- روش تجزیه و تحلیل اطلاعات
۸	۱۰-۱- تعریف واژه‌ها
۱۲	۱۱-۱- تعریف علمی متغیرها
۱۴	فصل دوم
۱۵	بخش اول: بیماری دیابت
۱۵	۱-۲- دیابت قندی
۱۶	۲-۲- انواع بیماری قند
۱۷	۳-۲- دیابت و حاملگی
۱۸	۴-۲- عوارض طولانی مدت
۱۸	۱-۴-۲- عوارض چشمی
۱۹	۲-۴-۲- نوروپاتی
۱۹	۳-۴-۲- نوروپاتی
۲۰	۴-۴-۲- بیماریهای قلبی و عروقی
۲۱	۵-۴-۲- افزایش فشارخون
۲۲	۵-۲- اهداف کلی اقدامات درمانی در دیابت
۲۲	۶-۲- تشخیص

۲۴	بخش دوم : مروری بر تحقیقات انجام گرفته
۲۵	الف: تحقیقات انجام گرفته در ایران
۲۹	ب: تحقیقات انجام شده در سایر کشورها
۳۲	فهرست منابع فصل دوم
۳۵	فصل سوم: چارچوب نظری
۳۶	۳-۱- مقدمه
۳۷	۳-۲- مفهوم تابع تشخیص
۳۸	۳-۳- تعریف تحلیل ممیزی
۳۹	۳-۴- مراحل تحلیل ممیزی
۳۹	۳-۴-۱- مشخص نمودن جوامع یا "تحلیل خوشه‌ای"
۳۹	۳-۴-۲- جمع‌آوری اطلاعات
۳۹	۳-۴-۳- اعمال الگوریتم‌های آماری
۴۰	۳-۴-۴- ارائه و تفسیر صحیح نتایج
۴۰	۳-۵- روشهای غیررسمی ممیزی
۴۱	۳-۶- نظریه تابع ممیز خطی
۴۸	۳-۶-۱- احتمالات خطا و محاسبه آنها
۴۹	۳-۶-۲- معایب و مزایای تابع ممیز خطی
۵۰	۳-۷- بهینه سازی تابع ممیز
۵۰	۳-۷-۱- روش مینیم کردن احتمال کل رده بندی نادرست
۵۳	۳-۷-۲- ممیزی به روش مینیماکس
۵۵	۳-۷-۳- قاعده ممیزی نیمن- پیرسن
۵۶	۳-۷-۴- ممیزی به روش بیز
	۳-۷-۵- ممیزی با معیار مینیمم کردن امید ریاضی یا متوسط زیان
۵۷	رده بندی نادرست
	۳-۷-۶- قاعده ممیزی مینیماکس با در نظر گرفتن زیانهای رده بندی
۵۸	نادرست
۵۹	۳-۸- ممیزی درجه دوم
۶۱	۳-۸-۱- محاسن و معایب ممیز درجه دوم

۶۱	۹-۳- معیارهای عملی برآورد خطا
۶۲	۳-۹-۱- میزان خطای آشکار
۶۳	۳-۹-۲- روش کنار گذاشتن یک مشاهده (جک نایف)
۶۴	۳-۹-۳- روش بوت استرپ
۶۵	۳-۹-۴- روش اعتبار مقاطع
۶۶	۳-۱۰- بهینه‌سازی تابع ممیز خطی در دو جامعه
۶۷	۳-۱۱- سایر روشهای ممیزی
۶۷	۳-۱۱-۱- تحلیل ممیزی دنباله‌ای
۶۸	۳-۱۱-۲- تحلیل ممیزی لجستیک
۶۹	۳-۱۱-۲-۱- الگوی رگرسیون لجستیک
۷۱	۳-۱۱-۲-۲- برآورد پارامترها
۷۲	۳-۱۱-۲-۳- گسترش توابع لجستیک چندمتغیره
۷۴	۳-۱۱-۲-۴- نقاط ضعف و قوت
۷۴	۳-۱۱-۳- ممیزی داده‌های کیفی
۷۵	۳-۱۲- وجه تشابه تابع ممیز خطی و الگوهای خطی
۷۶	۳-۱۲-۱- ارتباط تحلیل ممیزی و مسأله مقایسه میانگین‌های دو جامعه
۷۸	۳-۱۲-۲- ارتباط $R^2, D^2$
۷۹	فهرست منابع فصل سوم
۸۱	فصل چهارم : نتیجه‌گیری
۸۲	۴-۱- توصیف متغیرها
۸۳	۴-۲- برآورد تابع تشخیص
۸۸	۴-۳- مقایسه رگرسیون لجستیک و تابع تشخیص
۹۰	۴-۴- مقایسه تابع تشخیص و رگرسیون چند متغیره
۹۱	۴-۵- خلاصه و بحث
۹۸	۴-۶- پیشنهادات برای پژوهشهای بعدی
۹۹	فهرست منابع فصل چهارم
۱۰۲	پیوست فصل چهارم

# فصل اول

کلیات

## ۱-۱- مقدمه

دیابت شیرین یکی از شایعترین بیماریهای متابولیک است و در کشورهای پیشرفته و نیز در حال پیشرفت یک مشکل بهداشتی در حال گسترش می‌باشد که منجر به ناتوانی و مرگ زودرس می‌گردد. درمان ناتوانی‌ها و مرگ زودرس منجر به تحمیل هزینه‌های زیادی چه بصورت مستقیم و چه به صورت غیرمستقیم می‌گردد. چهل و دومین نشست "WHO" تأکید بر بررسی وضعیت دیابت در سطح ملی و نیز اجرای برنامه‌هایی در سطح جامعه در جهت پیشگیری و کنترل این بیماری دارد. دیابت بیماری است متأثر از فاکتورهای ژنتیک و محیطی که از نقطه نظر ایتولوژی و پیش آگهی و شیوع در جوامع مختلف متفاوت است. دیابت تیپ II که شایعترین نوع دیابت می‌باشد از سومین دهه زندگی شیوعی رو به افزایش داشته و در هفتمین دهه عمر، ۳ تا ۴ برابر شیوع آن در بالغین است چنانچه چاره‌ای اندیشیده نشود، شیوع دیابت تیپ II در تمامی جوامع به دلیل عواملی چون افزایش دوره عمر، کاهش مورتالیتی ناشی از بیماری عفونی و افزایش شیوع فاکتورهای مساعدکننده چون چاقی، از دست دادن فعالیت بدنی منظم، رژیم نامناسب رو به افزایش خواهد گذاشت.

با توجه به این نکته که کشور ما کشوری است در حال توسعه با منابع اقتصادی محدود و جمعیت رو به تزاید، تحقیق در زمینه راههای کنترل بیماری دیابت تیپ II از اهمیت ویژه‌ای برخوردار است که یکی از آنها شناسایی در بیماران فاقد علائم کلاسیک بیماری است. شناسایی بیماری از آن جهت مثمرتر است که شناخت فاکتورهای مؤثر بر بیماری نقش پیشگیری و کنترل از آن را نمایان می‌سازد و همچنین در مواردی که تستهای آزمایشگاهی محدودیت دارند و لازم است نتایج خیلی سریع در اختیار پزشکان قرار گیرد می‌توان از این روش استفاده نمود. مطالعات گذشته صرفاً طرحهای مقطعی بودند که هدفی جز تعیین ریسک فاکتورهای بیماری دیابت را نداشتند، اما در تحقیق حاضر سعی داریم به انتخاب یک قاعده تصمیم‌گیری بهینه به منظور شناسایی هر چه بهتر بیماری با استفاده از مدل‌های تابع تشخیص برسیم.

## ۱-۲- اهداف پژوهش

### هدف کلی:

- برآورد مدل‌های آماری با استفاده از تابع تشخیص برای شناسایی بیماران دیابتی.
- برای دستیابی به هدف کلی، اهداف جزئی زیر را در نظر می‌گیریم.

### اهداف جزئی:

- تعیین متغیرهایی که بین افراد سالم و بیمار تفاوت معنی‌دار داشته‌اند.
- تعیین مجموعه‌ای از متغیرها با کمترین خطای پیش‌بینی
- تعیین حساسیت و ویژگی مدل تشخیص
- ساخت یک قاعده تصمیم‌گیری بهینه در زمینه تشخیص دیابت

## ۱-۳- فرضیه‌های تحقیق

- ۱- استفاده از مدل‌های تابع تشخیص، در صورتیکه احتمال‌های پیشین دو جامعه تغییر کند، درصد خطای طبقه‌بندی را تغییر می‌دهند. اما تغییری در متغیرهای اثرگذار و ضرایب آنها نمی‌دهد.
- ۲- بین ابتلا به دیابت و متغیرهای سن و جنس و سابقه فامیلی ارتباط وجود دارد.
- ۳- بین ابتلا به دیابت و شاخص‌های آنتروپومتریک (BMI و WHR) ارتباط وجود دارد.
- ۴- بین ابتلا به دیابت و تعداد موارد حاملگی ارتباط وجود دارد.

## ۱-۴- سؤالات پژوهش

در این پایان‌نامه سعی خواهد شد که به صورت علمی به سؤالات زیر پاسخ داده شود:

- ۱- کدام متغیرها، بین گروهی که دیابت دارند و افراد سالم، متفاوت است؟
- ۲- ارزش هر یک از متغیرها برای شناسایی فرد مبتلا به دیابت چقدر است؟

۳- با استفاده از کدام مجموعه متغیرها و با چه احتمالات پیشین می‌توان مدل‌هایی تعیین نمود که

وضعیت سلامتی یا بیماری فرد را دقیقتر معین نماییم؟

۴- خطای این مدلها چقدر است؟

## ۱-۵- اهمیت پژوهش

یکی از عواملی که باعث پیشگیری از بروز بیماریها و ارتقاء سلامت جامعه و در پی آن کاهش هزینه‌های اقتصادی و بروز مسائل اجتماعی متعاقب بیماری و مشکلات پیامد آن می‌گردد، شناخت مشکلات و آشنایی با وضعیت آن موضوع و به عبارتی بیماری در جامعه مورد نظر است. با توجه به وضعیت اقلیمی، نژادی، اقتصادی و اجتماعی ایران لزوم شناخت اپیدمیولوژیک بیماریها جهت کنترل، و پیشگیری و برنامه ریزی مناسب جهت برخورد با این مسائل کاملاً مشهود می‌باشد.

در این میان بیماری دیابت نوع بزرگسالان شایعترین نوع بیماری، و با توجه به اینکه اغلب بدون علامت می‌باشد، در مناطق در حال توسعه رو به فزونی نهاده است. بنابراین بکارگیری راههای پیشگیری و کنترل سریع بیماری دیابت با توجه به عوارض و صدمات وسیع آن امری ضروری است. هرچند که بیماری ممکن است با علائم واضحی همچون لاغری، چاقی، تمایل به آب خوردن، ادرارکردن زیاد یا عوارض ناشی از بیماری بروز کند ولی در عده زیادی از بیماران بدون اینکه علائم واضحی داشته باشد، مشکلات زیادی برای فرد مبتلا ایجاد می‌کند.

در این راستا یکی از اهداف این تحقیق شناسایی بیماران بدون علائم در مواردیست که تستهای آزمایشگاهی محدودیت دارند و لازم است نتایج خیلی سریع در اختیار پزشکان قرار گیرد. هم اینک بیش از ۱۴۰ میلیون نفر دیابتی وجود دارند که این عده تا سال ۲۰۲۵ به ۳۰۰ میلیون نفر افزایش می‌یابند که عمدتاً مربوط به رشد جمعیت، مسن شدن جمعیت و زندگی شهرنشینی است. علاوه بر این مطالعات اپیدمیولوژیک در خصوص شیوع NIDDM در دنیا در طی دهه گذشته نشان‌دهنده ارتباط استعداد ژنتیک و عوامل خطر محیطی در بروز بیماری می‌باشند. شیوع بیماری در جوامع مدرنیزه و شهری و در حال تکامل بسیار بالاتر از جوامع سنتی و روستایی می‌باشد و