

**کالج پروژه**

**[www.collegeprozheh.ir](http://www.collegeprozheh.ir)**



**دانلود پروژه های دانشگاهی**

**بانک موضوعات پایان نامه**

**دانلود مقالات انگلیسی با ترجمه فارسی**

**آموزش نگارش پایان نامه ، مقاله ، پروپوزال**

**دانلود جزوه و نمونه سوالات استعدادی**

**دانلود رایگان پرسشنامه**



## دانشگاه آزاد اسلامی

واحد علوم و تحقیقات (تهران)

Science and Research Branch, Islamic Azad University

### فرم پیشنهاد تحقیق پایان نامه‌ی کارشناسی ارشد

عنوان تحقیق به فارسی:

تشخیص دیابت از داده های بزرگ با رویکرد سیستم مدیریت تصمیم گیری مبتنی بر اصول داده کاوی با شبکه عصبی سلولی

نام دانشجو:

دانشکده:

نام خانوادگی دانشجو:

گروه تخصصی:

رشته تحصیلی:

گرایش:

نیمسال ورود به مقطع جاری:

نیمسال شروع به تحصیل:

نام و نام خانوادگی استاد (اساتید) راهنما:

نام و نام خانوادگی استاد (اساتید) مشاور:

-۱

-۱

تاریخ تصویب در شورای گروه تخصصی:

تاریخ تصویب در شورای پژوهشی دانشکده:

تأیید مدیر پژوهشی دانشکده:

تأیید رئیس دانشکده:

تاریخ ارسال به حوزه پژوهشی واحد:

تأیید کارشناس پژوهشی

تاریخ بررسی و تأیید امور پژوهشی واحد:

تأیید مدیر کل پژوهشی

تأیید معاون پژوهشی واحد

**توجه:** لطفاً این فرم با مساعدت و هدایت استاد راهنما تکمیل شود.

۱- اطلاعات مربوط به دانشجو:

نام: ..... نام خانوادگی: ..... شماره دانشجویی: .....  
مقطع: ..... رشته تحصیلی: ..... گروه تخصصی: .....  
گرایش: ..... نام دانشکده: ..... سال ورود به مقطع جاری: .....  
نیمسال ورودی: .....  
آدرس پستی در تهران: .....  
تلفن ثابت محل سکونت: ..... تلفن همراه: ..... پست الکترونیک: .....  
آدرس پستی در شهرستان: .....  
تلفن ثابت محل سکونت: ..... تلفن محل کار: ..... دورنگار: .....

۲- اطلاعات مربوط به استاد راهنما:

**تذکرات:**

- دانشجویان دوره کارشناسی می توانند یک استاد راهنما و حداکثر دو استاد مشاور و دانشجویان دوره دکتری حداکثر تا دو استاد راهنما و دو استاد مشاور می توانند انتخاب نمایند.
- در صورتی که اساتید راهنما و مشاور **مدعو** می باشند، لازم است سوابق تحصیلی، آموزشی و پژوهشی کامل ایشان (رزومه کامل) شامل فهرست پایان نامه های کارشناسی ارشد و رساله های دکتری دفاع شده و یا در حال انجام که اساتید مدعو، راهنمایی و یا مشاوره آنرا بر عهده داشته اند، به **همراه مدارک مربوطه** و همچنین آخرین حکم کارگزینی (حکم هیأت علمی) ضمیمه گردد.
- اساتید راهنما و مشاور موظف هستند قبل از پذیرش پروپوزال، به سقف ظرفیت پذیرش خود توجه نموده و در صورت تکمیل بودن ظرفیت پذیرش، از ارسال آن به دانشکده و حوزه پژوهشی و یا در نوبت قرارداد و ایجاد وقفه در کار دانشجویان جداً پرهیز نمایند.

اطلاعات مربوط به استاد راهنمای اول:

نام و نام خانوادگی: ..... آخرین مدرک تحصیلی: ..... دانشگاهی: .....  
حوزوی  
عضو هیأت علمی دانشگاه .....  
تخصص اصلی: ..... رتبه دانشگاهی (مرتبه علمی): ..... تلفن همراه: .....  
تلفن منزل یا محل کار: ..... نام و نام خانوادگی به زبان انگلیسی: .....  
نحوه همکاری با واحد علوم و تحقیقات: .....  
☐ تمام وقت    ☐ نیمه وقت    ☐ مدعو

اطلاعات مربوط به استاد راهنمای دوم:

نام و نام خانوادگی: ..... آخرین مدرک تحصیلی: ..... دانشگاهی: .....  
حوزوی  
عضو هیأت علمی دانشگاه .....  
تخصص اصلی: ..... رتبه دانشگاهی (مرتبه علمی): ..... تلفن همراه: .....  
تلفن منزل یا محل کار: ..... نام و نام خانوادگی به زبان انگلیسی: .....  
نحوه همکاری با واحد علوم و تحقیقات: .....  
☐ تمام وقت    ☐ نیمه وقت    ☐ مدعو

اطلاعات مربوط به استاد مشاور:

نام و نام خانوادگی: ..... آخرین مدرک تحصیلی: ..... دانشگاهی: .....  
حوزوی  
عضو هیأت علمی دانشگاه .....  
تخصص اصلی: ..... رتبه دانشگاهی (مرتبه علمی): ..... تلفن همراه: .....  
تلفن منزل یا محل کار: ..... نام و نام خانوادگی به زبان انگلیسی: .....  
نحوه همکاری با واحد علوم و تحقیقات: .....  
☐ تمام وقت    ☐ نیمه وقت    ☐ مدعو

۴- اطلاعات مربوط به پایان نامه:

الف- عنوان تحقیق

۱- عنوان به زبان فارسی:

تشخیص دیابت از داده های بزرگ با رویکرد سیستم مدیریت تصمیم گیری مبتنی بر اصول داده کاوی با شبکه عصبی سلولی

۲- عنوان به زبان انگلیسی/(آلمانی، فرانسه، عربی):

**تذکر:** صرفاً دانشجویان رشته های زبان آلمانی، فرانسه و عربی مجازند عنوان پایان نامه خود را به زبان مربوطه در این بخش درج نمایند و برای بقیه دانشجویان، عنوان بایستی به زبان انگلیسی ذکر شود.

Diabetes Detection from Big Data with Decision Support System based on Data Mining Techniques with Cellular Neural Network

ب - تعداد واحد پایان نامه: ۶

ج- بیان مسأله اساسی تحقیق به طور کلی (شامل تشریح مسأله و معرفی آن، بیان جنبه های مجهول و مبهم، بیان متغیرهای مربوطه و منظور از تحقیق):

دیابت یکی از بیماری های مشترک و رو به رشد در کشورهای مختلف است که محققان دنیا در تلاش هستند تا بتوانند روشی به منظور جلوگیری از این بیماری را ارائه دهند. دیابت یک بیماری مزمن غیر مسری است. رشد این بیماری با افزایش غیر طبیعی سطح گلوکز در خون می باشد که به دو دسته تقسیم می شود که شامل نوع اول دیابت است که با تولید ناکافی انسولین توسط لوزالمعده روی می دهد و دوم، دیابت نوع دو است که شکست سلول ها در پاسخ موثر به تولید انسولین توسط لوزامعده می باشد. این بیماری از جانب سازمان سلامت جهانی و فدراسیون دیابت جهانی از زمان رشد پیوسته دیابت، از نزدیک دنبال و مورد توجه قرار گرفته است. بیان شده است که تا پایان سال ۲۰۱۵ میلادی چیزی نزدیک به ۳۹۲ میلیون نفر با دیابت دست و پنجه نرم خواهند کرد. با توجه به این آمار که روزانه نیز بر اساس گزارشات فدراسیون دیابت جهانی در حال رشد است، کنترل، پیش گیری و شناسایی زودهنگام آن می تواند در رشد این بیماری کمک نماید [۱].

انفورماتیک پزشکی مدیون ساختارهای پردازش، ذخیره سازی و انتشار اطلاعات برای حوزه های گوناگون در زمینه پزشکی است. هدف اصلی در پشت همه تلاش ها به منظور طبقه بندی، خوشه بندی، و استخراج ویژگی از داده های موجود این است که بتوان سیستم پشتیبان تصمیم گیری ساخت که به انسان در زمینه شناسایی و تشخیص بیماری ها کمک کند [۱].

اخیراً روش های داده کاوی به صورت گسترده به منظور تشخیص و توسعه بیماری های پیشرفته به کار گرفته می شوند. تحقیقات اخیر نشان می دهد که سیستم طبقه بند با دقت بالا بر روی داده های مختلف در شرایط یکسان تا به امروز ارائه نشده است. روش های داده کاوی به عنوان ابزاری برجسته در پایگاه داده های پزشکی به کار گرفته می شوند. این روش ها می توانند پنجره ای جدید به سوی شناسایی و تشخیص در بیماری ها را نمایه سازی کنند که به پیشرفت علم و همین طور کاهش بیماری در جوامع مختلف کمک شایانی می نماید [۲].

امروزه طبقه بندی به عنوان ابزاری مناسب در حوزه پزشکی به منظور کشف داده های بیمارانی و استخراج مدلی قابل پیش بینی به کار گرفته می شود. طبقه بندی ابزاری است که در سیستم های تصمیم گیری مورد استفاده قرار می گیرد و تا به امروز روش های طبقه بندی متنوعی در زمینه های گوناگون علمی به منظور کشف دانش، ایجاد شده اند. از آن جایی که داده های دیابت، دارای حجم بالایی هستند و به عنوان داده های بزرگ شناخته می شوند، لذا استخراج و کشف دانش جهت تشخیص و پیش بینی بیماری، دارای ارجحیت می باشد که تا به امروز روشی با دقت بالا در این حوزه ارائه نگردیده است. لذا رویکرد روش پیشنهادی، یک روش مبتنی بر یادگیری ماشین به سبک یادگیری بدون ناظر است که از روش ترکیبی آموزشی-احتمالاتی استفاده می نماید. منظور از یادگیری بدون ناظر، یادگیری ای است که در آن داده های ورودی مشخص هستند، ولی خروجی بر اساس همین داده های ورودی، تولید می شود که این تحقیق، یک پیش بینی بر همین اساس را انجام خواهد داد. داده ها ورودی مدنظر این تحقیق از پیوند <https://data.world/data-society/pima-indians-diabetes-database> و <https://archive.ics.uci.edu/ml/machine-learning-databases/pima-indians-diabetes> قابل دریافت است. درخت تصمیم، می تواند نتیجه هر بخش از عملیات داده کاوی را به صورت بصری نشان دهد. به همین خاطر از درخت تصمیمی به نام CART برای استخراج ویژگی ها استفاده می گردد. همین طور نیاز به سیستم آموزش محور، وجود خواهد داشت. از این رو از شبکه عصبی سلولی<sup>۱</sup> استفاده می شود. این شبکه عصبی جهت طبقه بندی و ایجاد کلاس هایی شامل مشکوک، بیمار و یا عدم بیمار را ایجاد می کند و آن ها را بر اساس ویژگی های استخراج شده، در کلاس های مختلف قرار می دهد.

د - اهمیت و ضرورت انجام تحقیق (شامل اختلاف نظرها و خلاءهای تحقیقاتی موجود، میزان نیاز به موضوع، فواید احتمالی نظری و عملی آن و همچنین مواد، روش و یا فرآیند تحقیقی احتمالاً جدیدی که در این تحقیق مورد استفاده قرار می گیرد):

بیماری قند یا به عبارت دیگر دیابت، بیماری مزمنی است که در نتیجه اختلال در تولید و عملکرد انسولین در بدن به وجود می آید. انسولین هورمونی است که در لوزالمعده تولید می شود و سلول ها را قادر می سازد تا گلوکز را از خون گرفته و برای تولید انرژی استفاده کند. لوزالمعده فرد مبتلا به دیابت، انسولین مورد نیاز بدن

<sup>۱</sup> Cellular Neural Network (CNN)

را تولید نمی‌کند که موجب دیابت نوع یک می‌شود و یا انسولین کارایی لازم را در بدن افراد ندارد که موجب دیابت نوع دو می‌شود. مقاومت به انسولین و همچنین ترشح ناکافی انسولین در پیدایش دیابت نوع دو دخالت دارند. بنابراین افراد دیابتی نسبت به افراد سالم قادر نخواهند بود به خوبی از گلوکز در سوخت و ساز بدنشان استفاده کنند و قند خون در این افراد به طور قابل ملاحظه‌ای افزایش خواهد یافت. دیابت عوارض بسیاری دارد و شایع‌ترین علت قطع اندام، نابینایی و نارسایی مزمن کلیوی و یکی از مهمترین عوامل خطر در ایجاد بیماری‌های قلبی است. اگرچه پیشگیری کامل از عوارض این بیماری امکان پذیر نیست، ولی با کنترل دقیق قند خون می‌توان آن‌ها را به تعویق انداخت. هدف اولیه از درمان دیابت نیز حفظ سطح قند خون نزدیک به محدوده نرمال است. از این رو رعایت رژیم غذایی و انجام فعالیت‌های حرکتی و ورزشی به همراه درمان دارویی سبب بهبود کنترل قند خون و مقاومت انسولینی شده و کیفیت زندگی فرد بیمار را افزایش داده و ابتلا به عوارض این بیماری را به تعویق می‌اندازد. با توجه به شیوع فزاینده این بیماری، به یکی از چالش‌های مهم رو در روی مسئولین بهداشت و درمان کشورهای مختلف دنیا، چه در حال توسعه و چه پیشرفته، تبدیل شده است.

ه- مرور ادبیات و سوابق مربوطه (بیان مختصر پیشینه تحقیقات انجام شده در داخل و خارج کشور پیرامون موضوع تحقیق و نتایج آنها و مرور ادبیات و چارچوب نظری تحقیق):

در [۳] یک طبقه بندی غیرخطی با استفاده از منطق فازی برپایه الگوریتم ژنتیک بر روی داده‌های چندگانه از جمله داده‌های دیابت، فشار خون، سرطان سینه و داده آیریس<sup>۲</sup> ارائه شده است که نتایج حاکی از دقت نسبتاً بالای این پژوهش نسبت به سایر روش‌های پیشین از جمله نایو بیزین<sup>۳</sup>، رگرسیون<sup>۴</sup>، شبکه عصبی تابع پایه شعاعی<sup>۵</sup> و چند روش دیگر است. داده‌های استفاده شده در زمینه دیابت، مجموعه داده PIMA Indian است. در [۴] استفاده از روش لونبرگ مارکارد<sup>۶</sup> به منظور ارزیابی کارایی در کاهش خطا در زمان طبقه بندی داده‌های دیابتی انجام گرفته است. این پژوهش از مجموعه داده‌های PIMA Indian استفاده کرده است. الگوریتم آموزشی به صورت پویا به شبکه عصبی جهت کاهش خطا اعمال شده است و توسط آموزش پیوسته شبکه تا زمان رسیدن به مرحله موثر بهینه، داده شده است. شبکه عصبی به کار گرفته شده در این پژوهش، شبکه عصبی پرسپترون چندلایه<sup>۷</sup> است و از حداقل میانگین خطا<sup>۸</sup> جهت ارزیابی استفاده شده است.

در [۵] یک مقایسه برای طبقه بندی کننده‌های موجود استفاده شده برای پیش بینی دیابت انجام گرفته است. روش‌های پیشینی چون درخت تصمیم<sup>۹</sup>، شبکه عصبی مصنوعی<sup>۱۰</sup>، رگرسیون منطقی<sup>۱۱</sup> و نایو بیزین مورد بررسی واقع

<sup>۲</sup> IRIS Dataset

<sup>۳</sup> Naïve Bayesian

<sup>۴</sup> Regression

<sup>۵</sup> Radial Basis Function (RBF)

<sup>۶</sup> Levenberg-Marquardt

<sup>۷</sup> Multi-Layered Perceptron (MLP)

<sup>۸</sup> Mean Squared Error (MSE)

<sup>۹</sup> Decision Tree

شده اند. سپس روش ارائه شده مقاله به نام بگینگ و بوستینگ<sup>۱۲</sup> برای ارتقای کارایی و پیش بینی مقاوم در داده های دیابت انجام گرفته است. داده های استفاده شده در این پژوهش مربوط به داده های بیماران دیابتی در کشور تایلند است. در [۶] روشی برای پیش بینی دیابت با استفاده از روش طبقه بندی در داده کاوی ارائه شده است که تحلیل ارزیابی نیز انجام می دهد. استفاده از روش آدابوست<sup>۱۳</sup> برپایه درخت تصمیم C4.5 به عنوان بخش آموزش دهنده داده ها مورد استفاده واقع شده است. نتایج به دست آمده نشان می دهد که روش ارائه شده نسبت به روش بوستینگ و بگینگ دارای کارایی بیشتری است. داده استفاده شده نیز CPCSSN<sup>۱۴</sup> می باشد.

در [۷] تحلیل کارایی برای مدل های طبقه بندی به منظور پیش بینی دیابت ارائه شده است. در این پژوهش روش های مرسوم طبقه بندی یادگیری ماشین<sup>۱۵</sup> به نام های درخت تصمیم C4.5، نزدیک ترین همسایه K<sup>۱۶</sup>، فورست تصادفی<sup>۱۷</sup> و ماشین بردار پشتیبان<sup>۱۸</sup> به منظور طبقه بندی بیماران دارای دیابت انجام گرفته است. نتایج پژوهش یک بار با استفاده از داده های نویز دار و یک بار بدون داده های نویز دار انجام گرفته است. منظور از نویز، داده هایی است که نرمال سازی نشده اند. معیارهای ارزیابی شامل دقت، حساسیت و ویژگی داده می باشد. نتایج به دست آمده حاکی از آن است که روش فورست تصادفی با بیش از ۹۹٪ دارای توانایی بالاتری در عملیات طبقه بندی مجموعه داده های استفاده دیابت نسبت به روش های نام برده می باشد. در [۸] یک سیستم پیش بینی خودکار برای بیماری دیابت برپایه روش ترکیبی ماشین بردار پشتیبان با تحلیل تفکیک کننده خطی<sup>۱۹</sup> برپایه تبدیل موجک<sup>۲۰</sup> ارائه شده است. در واقع به صورت مخفف، نام روش LDA-MWSVM نام گذاری شده است. این کار شامل سه مرحله است: مرحله استخراج ویژگی و مرحله کاهش ابعاد ویژگی با استفاده از روش تحلیل تفکیک کننده خطی و مرحله طبقه بندی با استفاده از مدل ترکیبی ماشین بردار پشتیبان با تبدیل موجک. در گام سوم عملیات پیش بینی نتایج صحیح و کارایی سیستم ارائه شده با استفاده از تحلیل حساسیت، ویژگی های داده، دقت طبقه بندی و ماتریس درهم ریختگی<sup>۲۱</sup> مورد تجزیه و تحلیل قرار گرفته است. دقت طبقه بندی این سیستم نزدیک به ۸۹/۷۴٪ می باشد. مجموعه داده استفاده شده در این پژوهش از داده های معتبر یک وب سایت<sup>۲۲</sup> است.

در [۹] یک سیستم طبقه بندی فازی برپایه الگوریتم بهینه سازی کلونی مورچگان<sup>۲۳</sup> برای پیش بینی بیماری دیابت پیشنهاد شده است. هدف این پژوهش استفاده از سیستم طبقه بندی برپایه الگوریتم بهینه سازی کلونی مورچگان است

<sup>10</sup> Artificial Neural Networks (ANN)

<sup>11</sup> Logistic Regression

<sup>12</sup> Bagging and Boosting

<sup>13</sup> Adaboost

<sup>14</sup> Canadian Primary Care Sentinel Surveillance Network

<sup>15</sup> Machine Learning

<sup>16</sup> K-Nearest Neighbor (KNN)

<sup>17</sup> Random Forest

<sup>18</sup> Support Vector Machine (SVM)

<sup>19</sup> Linear Discriminant Analysis (LDA)

<sup>20</sup> Wavelet Transform

<sup>21</sup> Confusion Matrix

<sup>22</sup> <http://www.phys.uni.torun.pl/kmk/projects/datasets.html>

<sup>23</sup> Ant Colony Optimization (ACO)

که بتواند مجموعه قوانین فازی برای پیش بینی بیماری دیابت را استخراج نماید که این روش به اختصار FCS-ANTMINER نامیده شده است. دقت روش طبقه بندی ۸۴/۲۴٪ تخمین زده شده است. مجموعه داده استفاده شده در این پژوهش، داده های PIMA Indian می باشند. در [۱۰] یک مقایسه بین دو روش طبقه بندی شبکه عصبی پرسپترون چندلایه و رگرسیون منطقی در مجموعه داده های دیابت در افراد امریکایی انجام گرفته است. عملیات استخراج ویژگی یک عملیات مهم و حیاتی در این پژوهش محسوب می شود، زیرا درستی این بخش، می تواند مقایسه ای علمی بین دو روش شبکه عصبی پرسپترون چندلایه و رگرسیون منطقی را به چالش بکشاند. به همین منظور، از الگوریتمی پایا برای استخراج ویژگی استفاده شده است که آن، الگوریتم ژنتیک<sup>۲۴</sup> است. نتایج به دست آمده نشان می دهد که نرخ حساسیت و ویژگی های داده در روش رگرسیون منطقی برابر ۰/۹۹۶۵ و ۰/۹۹۴۶ است و نرخ حساسیت و ویژگی های داده در روش شبکه عصبی پرسپترون چندلایه برابر ۰/۹۹۶۶ و ۰/۹۹۱۸ می باشد که مشخص است که شبکه عصبی پرسپترون چندلایه دارای عملکرد بهتری نسبت به روش رگرسیون منطقی در طبقه بندی داده های دیابتی دارد. مجموعه داده استفاده شده در این پژوهش برای افراد مسن در کشور امریکا است.

در [۱۱] تشخیص زود هنگام بیماری دیابت نوع دوم با استفاده از سیستم طبقه بندی چندگانه به منظور توسعه دقت در تشخیص برای بیماری پیچیده دیابت نوع دوم، انجام گرفته است. یک شماتیک رای گیری وزن دار پویا که ترکیب وزن دار معیارهای چندگانه نامیده می شود، به منظور ترکیب طبقه بندی در تصمیم گیری ارائه گردیده است. این روش شامل فقط دقت محلی یا سراسری نیست، بلکه مشارکت در امر طبقه بندی و خطای تعمیم مکان یابی شده برای هر طبقه بند نیز مد نظر واقع شده است. روش ارائه شده به اختصار MFWC نامیده می شود. ساخت سیستم با استفاده از دو داده دیابت نوع دوم انجام گرفته است. در [۱۲] یک روش طبقه بندی فازی با استفاده از الگوریتم کلونی زنبور مصنوعی<sup>۲۵</sup> برای بیماری دیابت ارائه شده است. در این پژوهش، عملگر جهش<sup>۲۶</sup> از الگوریتم کلونی زنبور مصنوعی برای ارتقای کارایی، اضافه شده است و اگر بهترین نتیجه جاری به دست آمده حاصل از عملیات طبقه بندی نتواند به روز رسانی شود، عملگر مرکب ترکیب<sup>۲۷</sup> به منظور بهسازی عملیات، اضافه می شود. در واقع در این پژوهش از الگوریتم تغییر یافته کلونی زنبور عسل استفاده شده است که این الگوریتم به عنوان ابزاری جدید جهت ایجاد و بهینه سازی خودکار توابع عضویت و قوانین دریافت شده از داده، به کار گرفته می شود. ارزیابی کارایی روش ارائه شده استفاده از نرخ طبقه بندی، حساسیت و ویژگی مقادارهای داده است که از روش 10-Fold-Cross-Vallidation برای این کار استفاده شده است. نرخ طبقه بندی به دست آمده ۸۴/۲۱٪ می باشد. داده استفاده شده در این پژوهش برگرفته شده از <https://archive.ics.uci.edu/ml/machine-learning-databases> است. در [۱۳] پیش بینی دیابت با استفاده از الگوریتم بهینه سازی غذایی باکتری<sup>۲۸</sup> و شبکه عصبی انجام گرفته است. مجموعه داده استفاده شده در این

<sup>24</sup> Genetic Algorithm (GA)

<sup>25</sup> Artificial Bee Colony (ABC)

<sup>26</sup> Mutation

<sup>27</sup> Blended Crossover

<sup>28</sup> Bacterial Foraging Optimization (BFO)

پژوهش داده های دیابتی PIMA India می باشد که شامل ۷۶۸ داده است که دارای هشت ویژگی است که این پژوهش از هر هشت ویژگی استفاده کرده است.

در [۱۴] به استفاده از شبکه عصبی کانولوشن<sup>۲۹</sup> و حافظه طولانی مدت<sup>۳۰</sup> جهت تشخیص بیماری دیابت بر اساس نرخ سیگنال های قلبی از مجموعه داده PIMA INDIA پرداخته شده است که نتیجه دقت روش پیشنهادی تا ۹۳٫۶٪ تخمین زده شده است. در [۱۵] نیز یک سیستم قدرتمند تشخیص هوشمند بیماری دیابت از مجموعه داده های PIMA INDIA ارائه شده است که بر اساس یک الگوریتم ترکیبی به نام سیستم استنتاج فازی مبتنی بر لجستیک تطبیقی<sup>۳۱</sup> می باشد. استفاده از اصول ترکیبی استخراج ویژگی و طبقه بندی در این تحقیق مدنظر واقع شده است. دقت روش پیشنهادی نسبتاً مناسب نیست و حدود ۸۸٫۰۳٪ تخمین زده شده است و پیچیدگی محاسباتی آن نیز بالا است.

به عنوان مقالات مروری نیز می توان به [۱۶] اشاره کرد که روش های تشخیص بیماری دیابت را در زمان استفاده از مجموعه داده PIMA INDIA با هم مقایسه کرده است. روش ها و الگوریتم هایی که مدنظر قرار گرفته اند شامل منطق فازی، FCM، ماشین بردار پشتیبان، الگوریتم ژنتیک، شبکه عصبی مصنوعی و الگوریتم تحلیل مولفه اصلی<sup>۳۲</sup> است. همین طور در [۱۷] تشخیص بیماری دیابت از داده های PIMA INDIA انجام گرفته است که مروری بر روش ها، نقاط ضعف و قوت به همراه نتایج آن ها ارائه گردیده است. ماشین بردار پشتیبان، شبکه عصبی مصنوعی، نایو بیزین<sup>۳۳</sup>، درخت تصمیم J48، روش بگینگ<sup>۳۴</sup> و روش ترکیبی الگوریتم ژنتیک با ماشین بردار پشتیبان مدنظر واقع شده است.

از جمله مشکلاتی که مقالات بررسی شده از مرجع [۳-۱۵] برای پیش بینی، تشخیص، شناسایی و طبقه بندی دیابت دارند، می توان به موارد ذیل اشاره نمود:

- ✓ استفاده نکردن از داده های واقعی و ذکر نشدن نوع داده و تعداد نمونه های موجود
- ✓ دقت روش های ارائه شده نسبتاً پایین است.
- ✓ تعداد روش های ارزیابی برای تضمین روش ارائه شده کم است.
- ✓ یک واحد مشخص غیر از دقت رویکردها برای روش های ارائه شده وجود ندارد تا طبق آن و به درستی مقایسه صورت گیرد.
- ✓ پیچیدگی محاسباتی اکثر روش ها بالاست.
- ✓ اکثر روش ها دارای خروجی در مقاله و پژوهش ها نیستند و فقط یک سری اعداد و ارقام آورده شده است.

<sup>29</sup> Convolution Neural Network (CNN)

<sup>30</sup> Long Short Term Memory (LSTM)

<sup>31</sup> Logistic Adaptive Network-based Fuzzy Inference System (LANFIS)

<sup>32</sup> Principle component Analysis (PCA)

<sup>33</sup> Naïve Bayesian

<sup>34</sup> Bagging

در جدول (۱) مقایسه ای بین روش های موجود انجام گرفته است که می توان نقاط ضعف و قدرت آن ها را مشاهده نمود.

جدول (۱) نقاط ضعف و قدرت روش های ارائه شده

روشن ارائه شده	مجموعه داده و تعداد نمونه ها	مزایا	معایب
استفاده از شبکه عصبی کانولوشن و حافظه طولانی مدت [۱۵]	PIMA India با ۷۶۸ داده	(۱) دقت بالا (۲) مشخص بودن داده ها (۳) سرعت اجرای مناسب و پیچیدگی محاسباتی پایین	(۱) مشخص نکردن روال دقیق اصول طبقه بندی و کلاس ها
استفاده از سیستم استنتاج فازی مبتنی بر لجستیک تطبیقی [۱۴]	PIMA India با ۷۶۸ داده	(۱) مشخص بودن داده ها (۲) ایجاد یک الگوریتم جدید آموزش محور آماری-احتمالاتی	(۱) دقت ناکافی (۲) پیچیدگی محاسباتی (۳) کند بودن اجرا
الگوریتم بهینه سازی غذایابی باکتری و شبکه عصبی پس انتشار [۱۳]	PIMA India با ۷۶۸ داده	(۱) دقت بالا (۲) مشخص بودن داده ها (۳) همگرایی سریع الگوریتم	(۱) پیچیدگی محاسباتی بالا (۲) کند بودن اجرا
طبقه بندی فازی با استفاده از الگوریتم کلونی زنبور مصنوعی [۱۲]	PIMA India با ۷۶۸ داده	(۱) بهبود الگوریتم زنبور مصنوعی (۲) مشخص بودن داده ها	(۱) دقت ناکافی (۲) پیچیدگی محاسباتی (۳) مشخص نبودن قوانین فازی (۴) تعمیم پذیر نبودن روش (۵) کند بودن اجرا
طبقه بندی چندگانه با شمای رای گیری وزن	مشخص نیست	(۱) بهبود طبقه بندی (۲) تعمیم پذیر بودن	(۱) دقت ناکافی (۲) پیچیدگی محاسباتی

دار پویا [۱۱]		روش	بالا (۳) گُند بودن اجرا (۴) مشخص نبودن داده
طبقه بندی شبکه عصبی پرسترون چندلایه و رگرسیون منطقی مبتنی بر الگوریتم ژنتیک [۱۰]	داده دیابت افراد امریکایی از یک موسسه دیابتی	(۱) تعمیم پذیر بودن روش (۲) دقت مناسب (۳) مشخص بودن داده (۴) مناسب بودن نرخ حساسیت	(۱) پیچیدگی محاسباتی (۲) گُند بودن اجرا (۳) مشخص نبودن ویژگی ها
طبقه بندی فازی بر پایه الگوریتم بهینه سازی کلونی مورچگان [۹]	PIMA India با ۷۶۸ داده	(۱) مشخص بودن داده (۲) همگرایی سریع	(۱) پیچیدگی محاسباتی (۲) گُند بودن اجرا (۳) مشخص نبودن توابع عضویت فازی و همچنین قوانین فازی (۴) دقت ناکافی (۵) تعمیم پذیر نبودن
روش ترکیبی ماشین بردار پشتیبان با تحلیل تفکیک کننده خطی بر پایه تبدیل موجک [۸]	مجموعه داده های وب سایتی معتبر <sup>۳۵</sup>	(۱) تعمیم پذیر بودن روش (۲) دقت مناسب (۳) مشخص بودن داده	(۱) پیچیدگی محاسباتی (۲) گُند بودن اجرا
درخت تصمیم C4.5، نزدیک ترین همسایه K، فورست تصادفی و ماشین بردار پشتیبان [۷]	مشخص نیست	(۱) تعمیم پذیر بودن روش (۲) مشخص بودن داده	(۱) پیچیدگی محاسباتی (۲) گُند بودن اجرا (۳) مشخص نبودن خروجی هر الگوریتم نام برده شده (۴) داده سازی عددی

<sup>35</sup> <http://www.phys.uni.torun.pl/kmk/projects/datasets.html>

<p>(۱) پیچیدگی محاسباتی (۲) گُند بودن اجرا (۳) مشخص نبودن حجم داده ها و نمونه های استفاده شده از مجموعه داده ها (۴) مشخص نبودن خروجی ها و ورودی ها</p>	<p>(۱) تعمیم پذیر بودن روش (۲) مشخص بودن داده (۳) بهبود طبقه بندی مبتنی بر روش ترکیبی آدابوست با یک درخت تصمیم</p>	CPCSSN	<p>الگوریتم آدابوست بر پایه درخت تصمیم C4.5 [۶]</p>
<p>(۱) پیچیدگی محاسباتی (۲) گُند بودن اجرا (۳) مشخص نبودن حجم داده ها و نمونه های استفاده شده از مجموعه داده ها</p>	<p>(۱) دقت بالا (۲) مشخص بودن داده (۳) عملکرد سریع نسبت به روش های طبقه بندی پیشین و بهبود آن ها جهت پیش بینی</p>	<p>مجموعه داده های بیماران دیابتی کشور تایلند</p>	<p>استفاده از بگینگ و بوستینگ [۵]</p>
<p>(۱) مشخص نبودن مقدار داده آموزش و آزمون (۲) مشخص نبودن تعداد نرون های لایه ورودی و میانی</p>	<p>(۱) مشخص بودن داده (۲) کاهش پیچیدگی محاسباتی (۳) سرعت اجرای بالا</p>	<p>PIMA India با ۷۶۸ داده</p>	<p>شبکه عصبی پرسپترون چندلایه و لونبرگ مارکارد [۴]</p>
<p>(۱) پیچیدگی محاسباتی (۲) گُند بودن اجرا (۳) مشخص نبودن توابع عضویت فازی و همچنین قوانین فازی (۴) تعمیم پذیر نبودن روش</p>	<p>(۱) دقت بالا (۲) مشخص بودن داده</p>	<p>PIMA India با ۷۶۸ داده</p>	<p>طبقه بندی غیرخطی با استفاده از منطق فازی بر پایه الگوریتم ژنتیک [۳]</p>

و - جنبه جدید بودن و نوآوری در تحقیق:

استفاده از ساختار درخت CART به عنوان روش استخراج ویژگی و شبکه عصبی سلولی برای طبقه بندی داده ها به عنوان یک نوآوری در ساختارهای داده کاوی برای داده های بزرگ در تشخیص دیابت، محسوب می شود. با توجه به این که داده های این تحقیق PIMA INDIA است و این داده ها نرمال سازی شده می باشند، پس می توان مستقیماً روش پیشنهادی را بر روی آن اعمال نمود. عملیات استخراج ویژگی ها مبتنی بر درخت CART<sup>۳۶</sup> انجام می شود و در ادامه درخت ها، به عنوان نمایش دهنده ویژگی ها استفاده می شوند. سپس طبقه بندی با شبکه عصبی سلولی آغاز می گردد و بر اساس ویژگی های استخراج شده، کلاس هایی شامل بیمار، مشکوک، سالم، ایجاد می نماید و داده های هر فرد در آن قرار می گیرد. همچنین بهبود معیارهای ارزیابی شامل دقت<sup>۳۷</sup>، حساسیت<sup>۳۸</sup>، نرخ ویژگی ها<sup>۳۹</sup>، فراخوان<sup>۴۰</sup> و میانگین مربعات خطا<sup>۴۱</sup> به عنوان یک بخش مهم جهت تضمین رویکرد پیشنهادی در مقایسه با سایر روش های پیشین به کار گرفته می شود. به صورت کلی، جنبه های نوآوری تحقیق پیش رو به صورت ذیل است:

- ✓ استفاده از رویکرد ترکیبی درخت تصمیم CART و شبکه عصبی سلولی
- ✓ کاهش پیچیدگی محاسباتی و کاهش مراحل انجام شامل پیش پردازش، استخراج ویژگی ها و طبقه بندی که به سه مرحله کلی کاهش یافته است. این امر، زمان اجرا به همراه دقت را در کارایی، افزایش می دهد.
- ✓ شناسایی و تشخیص دقیق افراد بیمار، سالم و مشکوک به بیماری و قرارگیری آن ها در کلاس های مختلف
- ✓ قابلیت تعمیم پذیری رویکرد پیشنهادی در صورت اضافه شدن حجم داده بیشتر در زمان های بعدی

ز- اهداف مشخص تحقیق (شامل اهداف آرمانی، کلی، اهداف ویژه و کاربردی):

- ✓ ارائه یک رویکرد مبتنی بر داده کاوی جهت کشف دانش از داده های بزرگ دیابتی
- ✓ طبقه بندی و تشخیص داده های افراد بیمار و پیش بینی بیماری آن ها
- ✓ ارائه یک روش موثر و بهینه در تشخیص و پیش بینی بیماری دیابت
- ✓ ارائه یک رویکرد با بالاترین دقت در تشخیص بیماری دیابت مبتنی بر یادگیری بدون نظارت با درخت CART و شبکه عصبی سلولی

<sup>36</sup> Classification And Regression Tree

<sup>37</sup> Accuracy

<sup>38</sup> Sensitivity

<sup>39</sup> Specificity

<sup>40</sup> Recall

<sup>41</sup> Mean Square Error (MSE)

ح - در صورت داشتن هدف کاربردی، نام بهره‌وران (سازمان‌ها، صنایع و یا گروه ذینفعان) ذکر شود (به عبارت دیگر محل اجرای مطالعه موردی):

بیمارستان‌ها، کلینیک‌ها و مراکز دیگر که به بیماران دیابتی کمک می‌کنند.

ط - سؤالات تحقیق:

سؤالات اصلی تحقیق:

✓ مدل سازی رویکرد پیشنهادی جهت تشخیص دیابت به چه صورت انجام می‌شود؟

✓ مزیت رویکرد پیشنهادی نسبت به روش‌های پیشین چیست و تا چه اندازه در تشخیص دیابت موثر واقع شده است؟

سؤالات فرعی تحقیق:

✓ اصول داده کاوی تا چه اندازه در سیستم‌های تشخیص هوشمند پزشکی به خصوص سیستم تشخیص و پیش بینی دیابت کارایی دارند؟

ی - فرضیه‌های تحقیق:

✓ داده کاوی به عنوان یک سیستم مدیریت تصمیم‌گیری با هدف شناسایی و استخراج دانش جدید در سیستم‌های تشخیص هوشمند پزشکی توانایی بالایی را ارائه می‌دهد.

✓ طبقه‌بندی و پیش‌بینی به عنوان سیستم‌های مدیریت تصمیم‌گیری از اصول داده کاوی قابلیت به کارگیری در تشخیص دیابت را دارد می‌باشد.

✓ استفاده از تحلیل‌های هوشمند رویکرد پیشنهادی، قابلیت تشخیص و پیش‌بینی دیابت را برای پردازش‌های آتی فراهم می‌آورد.

✓ تضمین رویکرد پیشنهادی با استفاده از معیارهای ارزیابی و مقایسه آن‌ها با روش‌های پیشین انجام می‌شود.

ک - تعریف واژه‌ها و اصطلاحات فنی و تخصصی (به صورت مفهومی و عملیاتی):

دیابت: بیماری قند یا به عبارت دیگر دیابت، بیماری مزمنی است که در نتیجه اختلال در تولید و عملکرد انسولین در بدن به وجود می‌آید. انسولین هورمونی است که در لوزالمعده تولید می‌شود و سلول‌ها را قادر می‌سازد تا گلوکز را از خون گرفته و برای تولید انرژی استفاده کند. لوزالمعده فرد مبتلا به دیابت، انسولین مورد

نیاز بدن را تولید نمی‌کند که موجب دیابت نوع یک می‌شود و یا انسولین کارایی لازم را در بدن افراد ندارد که موجب دیابت نوع دو می‌شود. مقاومت به انسولین و همچنین ترشح ناکافی انسولین در پیدایش دیابت نوع دو دخالت دارند. بنابراین افراد دیابتی نسبت به افراد سالم قادر نخواهند بود به خوبی از گلوکز در سوخت و ساز بدنشان استفاده کنند و قند خون در این افراد به طور قابل ملاحظه‌ای افزایش خواهد یافت. دیابت عوارض بسیاری دارد و شایع‌ترین علت قطع اندام، نابینایی و نارسایی مزمن کلیوی و یکی از مهمترین عوامل خطر در ایجاد بیماری‌های قلبی است. اگرچه پیشگیری کامل از عوارض این بیماری امکان پذیر نیست، ولی با کنترل دقیق قند خون می‌توان آن‌ها را به تعویق انداخت. هدف اولیه از درمان دیابت نیز حفظ سطح قند خون نزدیک به محدوده نرمال است. از این رو رعایت رژیم غذایی و انجام فعالیت‌های حرکتی و ورزشی به همراه درمان دارویی سبب بهبود کنترل قند خون و مقاومت انسولینی شده و کیفیت زندگی فرد بیمار را افزایش داده و ابتلا به عوارض این بیماری را به تعویق می‌اندازد. با توجه به شیوع فزاینده این بیماری، به یکی از چالش‌های مهم رو در روی مسئولین بهداشت و درمان کشورهای مختلف دنیا، چه در حال توسعه و چه پیشرفته، تبدیل شده است.

داده کاوی: داده کاوی پل ارتباطی میان علم آمار، علم کامپیوتر، هوش مصنوعی، الگوشناسی و فراگیری ماشین داده می‌باشد. داده کاوی فرآیندی پیچیده جهت شناسایی الگوها و مدل‌های صحیح، جدید و به صورت بالقوه مفید، در حجم وسیعی از داده می‌باشد، به طریقی که این الگوها و مدل‌ها برای انسان‌ها قابل درک باشند. داده کاوی به صورت یک محصول قابل خریداری نمی‌باشد، بلکه یک رشته علمی و فرآیندی است که بایستی به صورت یک پروژه پیاده سازی شود.

داده بزرگ: داده بزرگ اصطلاحی است برای مجموعه داده‌های حجیم که بزرگ، متنوع با ساختار پیچیده و با دشواری‌هایی برای ذخیره سازی، تحلیل و تصویرسازی (نمایش)، پردازش‌های بیشتر یا نتایج می‌باشد. پروسه تحقیق بر روی داده‌های حجیم جهت آشکارسازی الگوهای مخفی و راز همبستگی‌ها، تجزیه و تحلیل داده بزرگ نامیده می‌شود. این اطلاعات مفید برای سازمان‌ها و شرکت‌ها در جهت کسب بینش غنی تر و عمیق تر و موفقیت در رقابت کمک می‌کند. به همین دلیل اجرای داده بزرگ نیاز دارند تا در صورت امکان، تحلیل شوند و به طور دقیق اجرا شوند.

درخت تصمیم CART: یکی از محبوب‌ترین و در عین حال از ساده‌ترین الگوریتم‌های درخت‌های تصمیم، درخت تصمیم CART است که کاربردهای زیادی در طبقه بندی و رگرسیون دارد. CART که خود مخفف Classification And Regression Tree است، بر اساس درخت‌های دودویی بنا نهاده شده است. الگوریتم درخت تصمیم CART برای ساخت درخت تصمیم، داده‌ها را به قسمت‌های دوتایی تقسیم کرده و بر اساس آن‌ها درخت دودویی را می‌سازد.

شبکه عصبی: شبکه عصبی، یک سیستم پردازش اطلاعات است که دارای ویژگی‌های مشترکی با شبکه‌های عصبی طبیعی است. شبکه‌های عصبی مصنوعی تعمیم یافته مدل‌های ریاضی تشخیص انسان بر اساس زیست‌شناسی

عصبی هستند. نرون ها، تعداد دور تکرار، لایه های ورودی، پنهان و خروجی به عنوان مسئله مهم در شبکه های عصبی برشمرده می شوند.

#### ۵-روش تحقیق:

الف- شرح کامل روش تحقیق بر حسب هدف، نوع داده ها و نحوه اجراء (شامل مواد، تجهیزات و استانداردهای مورد استفاده در قالب مراحل اجرایی تحقیق به تفکیک):

تذکر: درخصوص تفکیک مراحل اجرایی تحقیق و توضیح آن، از به کار بردن عناوین کلی نظیر، «گردآوری اطلاعات اولیه»، «تهیه نمونه های آزمون»، «انجام آزمایش ها» و غیره خودداری شده و لازم است در هر مورد توضیحات کامل در رابطه با منابع و مراکز تهیه داده ها و ملزومات، نوع فعالیت، مواد، روش ها، استانداردها، تجهیزات و مشخصات هر یک ارائه گردد.

با توجه به این که داده های این تحقیق <https://archive.ics.uci.edu/ml/machine-learning-databases> است و این داده ها نرمال سازی شده می باشند، می توان مستقیماً رویکرد پیشنهادی را بر روی آن اعمال نمود. عملیات استخراج ویژگی ها مبتنی بر درخت تصمیم CART انجام می شود و در ادامه درخت ها، به عنوان نمایش دهنده ویژگی ها استفاده می شوند. سپس طبقه بندی با شبکه عصبی سلولی آغاز می گردد و بر اساس ویژگی های استخراج شده، کلاس هایی شامل بیمار، مشکوک، سالم، ایجاد می نماید و داده های هر فرد در آن قرار می گیرد. در انتها تضمین رویکرد پیشنهادی با مدنظر قرار دادن معیارهای ارزیابی شامل دقت، حساسیت، نرخ ویژگی ها، فراخوان و میانگین مربعات خطا مدنظر واقع می گردد که یک مقایسه بین روش های پیشین نیز انجام می شود. مراحل انجام به صورت کلی به شکل ذیل است:

- (۱) مطالعه مقالات پیشین در زمینه تشخیص دیابت به صورت هوشمند و بررسی نقاط ضعف و قوت آن ها
- (۲) ارائه یک مدل جدید جهت رفع چالش های روش های قبلی در تشخیص دیابت به صورت هوشمند

- ۳) انتخاب یک مجموعه داده با ویژگی های مشخص و شناسایی آن ها
- ۴) مدل سازی و ارائه شبیه سازی جهت تضمین رویکرد پیشنهادی
- ۵) ارزیابی و اعتبارسنجی نتایج به همراه مقایسه با روش های مشابه پیشین در شرایط کیسان (استفاده از داده های مشابه و یکسان)
- ۶) نتیجه گیری و جمع بندی نهایی

ب- متغیرهای مورد بررسی در قالب یک مدل مفهومی و شرح چگونگی بررسی و اندازه گیری متغیرها:

از جمله متغیرهایی که مورد بررسی قرار می گیرد، ویژگی های موجود در مجموعه داده استاندارد هست که شامل تعداد افراد باردار، سطح قند خون، سیستولیک، پوسته پوسته شدن، انسولین، سن، شاخص توده بدن و عوامل ارثی بودن در یک خانواده است. بر اساس مقادیر عددی که در مقابل هر کدام از این ویژگی ها ذکر شده است، می توان تشخیص دیابت را انجام داد. همین طور در نظر گرفتن پارامترهای درخت تصمیم CART شامل برگ، ریشه و فرزندان به عنوان یک مسئله مهم است. در شبکه عصبی سلولی نیز می بایست کلاس های بیمار، عدم بیمار و مشکوک را ایجاد نمود که در این بین در کلاس بیماری، دو زیرمجموعه شامل بدخیم و خوش خیم نیز وجود دارد. تنظیم پارامترهای شبکه عصبی سلولی شامل تعداد دور تکرار، هسته آموزش<sup>۴۲</sup>، الگوریتم یادگیر، مجموع بخش آموزش و آزمون و سایر موارد، مدنظر قرار می گیرد. همچنین یک سری متغیر برای ارزیابی وجود دارد که شامل میانگین مربعات خطا، حساسیت، دقت، فراخوان و نرخ ویژگی ها می باشد که بر اساس این معیارها، مقایسه با روش های دیگر صورت می پذیرد. به صورت کلی، می توان متغیرهای تحقیق را در دو دسته قرار داد که به صورت جدول ذیل است:

متغیرهای تحقیق برای بیماری دیابت از مجموعه داده (متغیرهای وابسته)	متغیرهای تحقیق برای الگوریتم و ارزیابی (متغیرهای مستقل)
تعداد افراد باردار	پارامترهای شبکه عصبی سلولی شامل تعداد دور تکرار، هسته آموزش، الگوریتم یادگیر، مجموع بخش آموزش و آزمون و سایر موارد
سطح قند خون	میانگین مربعات خطا
سیستولیک	حساسیت
پوسته پوسته شدن	دقت
انسولین	نرخ ویژگی ها
سن	فراخوان

شاخص توده بدن	منحنی ROC و AUC
عوامل ارثی بودن در یک خانواده	ایجاد کلاس یا طبقه خوش خیم، بدخیم به همراه کلاس های کلی بیمار، سالم و مشکوک

ج - شرح کامل روش (میدانی، کتابخانه‌ای) و ابزار (مشاهده و آزمون، پرسشنامه، مصاحبه، فیش‌برداری و غیره) گردآوری داده‌ها:

مطالعه کتاب‌ها، مقالات و پایان نامه‌های معتبر در زمینه تشخیص دیابت به صورت هوشمند و در کل مطالعه از نوع کتابخانه‌ای است.



د - جامعه آماری، روش نمونه‌گیری و حجم نمونه (در صورت وجود و امکان):

داده این تحقیق، استفاده از <https://archive.ics.uci.edu/ml/machine-learning-databases> با حدود ۶ میلیون داده است که مقالات پایه این تحقیق نیز از آن‌ها استفاده نموده‌اند. نمایی از بخش اولیه این مجموعه داده به صورت شکل (۱) است.

	A	B	C	D	E	F	G	H	I
747	12	100	84	33	105	30	0.488	46	0
748	1	147	94	41	0	49.3	0.358	27	1
749	1	81	74	41	57	46.3	1.096	32	0
750	3	187	70	22	200	36.4	0.408	36	1
751	6	162	62	0	0	24.3	0.178	50	1
752	4	136	70	0	0	31.2	1.182	22	1
753	1	121	78	39	74	39	0.261	28	0
754	3	108	62	24	0	26	0.223	25	0
755	0	181	88	44	510	43.3	0.222	26	1
756	8	154	78	32	0	32.4	0.443	45	1
757	1	128	88	39	110	36.5	1.057	37	1
758	7	137	90	41	0	32	0.391	39	0
759	0	123	72	0	0	36.3	0.258	52	1
760	1	106	76	0	0	37.5	0.197	26	0
761	6	190	92	0	0	35.5	0.278	66	1
762	2	88	58	26	16	28.4	0.766	22	0
763	9	170	74	31	0	44	0.403	43	1
764	9	89	62	0	0	22.5	0.142	33	0
765	10	101	76	48	180	32.9	0.171	63	0
766	2	122	70	27	0	36.8	0.34	27	0
767	5	121	72	23	112	26.2	0.245	30	0
768	1	126	60	0	0	30.1	0.349	47	1
769	1	93	70	31	0	30.4	0.315	23	0

نمایی از بخش اولیه مجموعه داده مدنظر

همین طور نمایی از پوشه هایی که می توان برای این داده بارگذاری کرد، به صورت شکل (۲) است که پیوند دسترسی آن <https://archive.ics.uci.edu/ml/machine-learning-databases> می باشد.

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 <a href="#">Parent Directory</a>		-	
 <a href="#">00000/</a>	22-Mar-2018 14:43	-	
 <a href="#">00192/</a>	11-Sep-2010 18:05	-	
 <a href="#">00193/</a>	11-Sep-2010 18:38	-	
 <a href="#">00194/</a>	11-Sep-2010 18:44	-	
 <a href="#">00195/</a>	11-Oct-2010 10:19	-	
 <a href="#">00196/</a>	03-Nov-2010 22:59	-	
 <a href="#">00197/</a>	03-Nov-2010 22:57	-	
 <a href="#">00198/</a>	20-May-2011 15:36	-	
 <a href="#">00199/</a>	05-Jan-2011 10:07	-	
 <a href="#">00200/</a>	07-Feb-2011 11:33	-	
 <a href="#">00201/</a>	07-Feb-2011 11:33	-	
 <a href="#">00202/</a>	07-Feb-2011 11:33	-	
 <a href="#">00203/</a>	07-Feb-2011 11:52	-	
 <a href="#">00204/</a>	01-Jun-2011 10:03	-	
 <a href="#">00205/</a>	26-Jul-2011 23:10	-	
 <a href="#">00206/</a>	26-Jul-2011 23:05	-	
 <a href="#">00207/</a>	26-Jul-2011 22:35	-	
 <a href="#">00208/</a>	26-Jul-2011 22:53	-	
 <a href="#">00209/</a>	26-Jul-2011 23:07	-	

شکل (۲) نمایی از پوشه های مجموعه داده مدنظر

پارامترهایی که از این مجموعه داده به عنوان عوامل تشخیص بیماری دیابت مدنظر واقع شده است، شامل تعداد افراد باردار، سطح قند خون، سیستمیک، پوسته پوسته شدن، انسولین، سن، شاخص توده بدن و عوامل ارثی بودن در یک خانواده است. قابل ذکر است که سه ویژگی شامل سطح قند خون، میزان مصرف انسولین و سیستمیک، به عنوان سه ویژگی اصلی در شناسایی و تشخیص دیابت در این تحقیق، متصور واقع شده اند. ستون اول، شناسه هر بیمار است، در ستون دوم، M نماینده بدخیم یا Malingal و B نماینده خوش خیم یا Benign است. همین طور S یا Suspicious برای حالت مشکوک به دیابت مدنظر قرار می گیرد. موارد دیگر در ستون ها نیز ویژگی های دیگر بیماری دیابت است که بر اساس این ویژگی ها و بازه های آن، به روش طبقه بندی و استخراج ویژگی مقادیری داده می شود که بتواند امر تشخیص دیابت را به درستی انجام دهد. این مقادیر به صورت دستی انجام نمی شود، زیرا هدف داده سازی عددی نیست، بلکه استفاده از مراجع در این زمینه امری ضروری است. قابل ذکر است که تحقیقات دیگر اصولاً از یک پوشه از مجموعه داده استفاده کرده اند، اما رویکرد پیشنهادی سعی در استفاده از تمامی مجموعه داده به عنوان یک داده بزرگ را دارد که با استفاده از تمامی پوشه های مجموعه داده مورد نظر، اعمال خواهد شد که چیزی حدود ۵ میلیون داده است.

هـ - روش‌ها و ابزار تجزیه و تحلیل داده‌ها:

استفاده از شبیه ساز و زبان MATLAB به عنوان سکو یا پلتفرم شبیه سازی جهت ایجاد رویکرد پیشنهادی بر روی مجموعه داده استاندارد و در انتها استفاده از یک سری معیارهای ارزیابی از جمله دقت، حساسیت، نرخ ویژگی ها، فراخوان، میانگین مربعات خطا و موارد مشابه که در مقالات پایه استفاده شده اند.

#### مراجع

- [1] Tora Grauers Willadsen, Volkert Siersma, Anni Brit Sternhagen Nielsen, Rasmus Køster-Rasmussen, and Niels de Fine Olivarius. "The effect of structured personal care on diabetes symptoms and self-rated health over 14 years after diabetes diagnosis." *Primary Care Diabetes*, Vol. 12, Issue 4, pp. 354-363, August 2018.
- [2] Juan Salazar, Cristobal Espinoza, Andres Mindiola, and Valmore Bermudez. "Data Mining and Endocrine Diseases: A New Way to Classify?" *Archives of Medical Research*, in press, corrected proof, Available online 14 August 2018.
- [3] Hua Fang, Maria L. Rizzo, Honggang Wang, Kimberly Andrews Espy, and Zhenyuan Wang. "A new nonlinear classifier with a penalized signed fuzzy measure using effective genetic algorithm." *Pattern Recognition*, Vol. 43, pp. 1393-1401, 2010.
- [4] Nawaz Khan, Dhara Gaurav, and Thomas Kandl. "Performance Evaluation of Levenberg-Marquardt Technique in Error Reduction for Diabetes Condition Classification." *International Conference on Computational Science (ICCS)*, *Procedia Computer Science*, Vol. 18, pp. 2629-2637, 2013.
- [5] Nongyao Nai-arun, and Rungruttikarn Mounngmai. "Comparison of Classifiers for the Risk of Diabetes Prediction." *7th International Conference on Advances in Information Technology*, *Procedia Computer Science*, Vol. 69, pp. 132-142, 2015.
- [6] Sajida Perveen, Muhammad Shahbaz, Aziz Guergachi, and Karim Keshavjee. "Performance Analysis of Data Mining Classification Techniques to Predict Diabetes." *Symposium on Data Mining Applications (SDMA)*, Riyadh, Saudi Arabia, *Procedia Computer Science*, Vol. 82, pp. 115-121, 2016.
- [7] J. Pradeep Kandhasamy, and S. Balamurali. "Performance Analysis of Classifier Models to Predict Diabetes Mellitus." *Procedia Computer Science*, Vol. 47, pp. 45-51, 2015.
- [8] Duygu Calisir, and Esin Dogantekin. "An automatic diabetes diagnosis system based on LDA-Wavelet Support Vector Machine Classifier." *Expert Systems with Applications*, Vol. 48, pp. 8311-8315, 2011.
- [9] Mostafa Fathi Ganji, and Mohammad Saniee Abadeh. "A fuzzy classification system based on Ant Colony Optimization for diabetes disease diagnosis." *Expert Systems with Applications*, Vol. 38, pp. 14650-14659, 2011.

- [10] S. Upadhyaya, K. Farahmand, and T. Baker-Demaray. "Comparison of NN and LR classifiers in the context of screening Native American elders with diabetes." *Expert Systems with Applications*, Vol. 40, pp. 5830-5838, 2013.
- [11] Jia Zhu, Qing Xie, and Kai Zheng. "An improved early detection method of type-2 diabetes mellitus using multiple classifier system." *Information Sciences*, Vol. 292, pp. 1-14, 2015.
- [12] Fayssal Beloufa, and M.A. Chikh. "Design of fuzzy classifier for diabetes disease using Modified Artificial Bee Colony algorithm." *Computer Methods and Program in Biomedicine*, Vol. 112, pp. 92-103, 2013.
- [13] Rabina, and Prof. Meenakshi Sharma. "Diabetes Prediction by using Bacterial Foraging Optimization Algorithm and Artificial Neural Network." *International Journal of Computer Science and Information Technology & Security (IJCSITS)*, Vol.6, No.1, 2016.
- [14] Swapna G, Soman Kp, and Vinayakumar R. "Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals." *Procedia Computer Science*, Vol. 132, pp. 1253-1262, 2018.
- [15] Rohollah Ramezani, Mansoureh Maadi, and Seyedeh Malihe Khatami. "A novel hybrid intelligent system with missing value imputation for diabetes diagnosis." *Alexandria Engineering Journal*, In press, corrected proof, Available online 19 April 2017.
- [16] Sakshi Gujral. "Early Diabetes Detection using Machine Learning: A Review." *IJIRST–International Journal for Innovative Research in Science &Technology*, Vol. 3, Issue 10, March 2017.
- [17] Meherwar Fatima, and Maruf Pasha. "Survey of Machine Learning Algorithms for Disease Diagnostic." *Journal of Intelligent Learning Systems and Applications*, Vol. 9, pp. 1-16, 2017.

## ۶- استفاده از امکانات آزمایشگاهی واحد:

- آیا برای انجام تحقیقات نیاز به استفاده از امکانات آزمایشگاهی واحد علوم و تحقیقات می باشد؟ بلی

☐ خیر ☐

در صورت نیاز به امکانات آزمایشگاهی لازم است نوع آزمایشگاه، تجهیزات، مواد و وسایل مورد نیاز در این قسمت مشخص گردد.

نوع آزمایشگاه	تجهیزات مورد نیاز	مواد و وسایل	مقدار مورد نیاز

- آیا برای انجام تحقیقات نیاز به حمایت از سایر مراکز خارج از واحد علوم و تحقیقات می باشید؟

بلی ☐ خیر ☐

در صورت نیاز نام مراکز و نحوه حمایت (مالی، امکانات و تجهیزات و ..) مشخص گردد.

امضاء استاد راهنما:

امضاء مدیر گروه تخصصی:

۷- زمان بندی انجام تحقیق:

الف- تاریخ شروع: ..... ب- مدت زمان انجام تحقیق: ..... ج- تاریخ اتمام: .....

تذکر: لازم است کلیه فعالیت ها و مراحل اجرایی تحقیق (شامل زمان ارائه گزارشات دوره ای) و مدت زمان مورد نیاز برای هر یک، به تفکیک پیش بینی و در جدول مربوطه درج گردیده و در هنگام انجام عملی تحقیق، حتی الامکان رعایت گردد.



پیش‌بینی زمان‌بندی فعالیت‌ها و مراحل اجرایی تحقیق و ارائه گزارش پیشرفت کار

	شرح فعالیت	زمان کل (ماه)	زمان اجرا به ماه											
			۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
۱														
۲														
۳														
۴														
۵														
۶														
۷														
۸														
۹														
۱۰														
۱۱														
۱۲														

توجه: ۱- زمان و نوع فعالیت‌های اجرایی پایان‌نامه، حتی‌الامکان باید با مندرجات جدول منطبق باشد.

۲- حداقل زمان قابل قبول برای پیش‌بینی مراحل مطالعاتی و اجرایی پایان‌نامه کارشناسی ارشد ۶ ماه و حداکثر ۱۲ ماه می‌باشد.

**تذکر:** اساتید راهنما و مشاور موظف هستند قبل از پذیرش پروپوزال، به سقف ظرفیت راهنمایی و مشاوره خود توجه نموده و در صورت تکمیل بودن ظرفیت پذیرش، از امضاء این فرم و یا در نوبت قرار دادن آن و ایجاد وقفه در کار دانشجویان جداً پرهیز نمایند. بدیهی است در صورت عدم رعایت موازین مربوطه، مسئولیت تأخیر در ارائه پروپوزال و عواقب کار، متوجه گروه تخصصی خواهد بود.

#### ۸- صورت جلسه گروه تخصصی

نام و نام خانوادگی دانشجو:	امضاء	تاریخ
نام و نام خانوادگی استاد راهنما	امضاء	تاریخ
۱-	(عضو هیأت علمی دانشگاه .....	
نام و نام خانوادگی استاد مشاور	امضاء	تاریخ
۱-	(عضو هیأت علمی دانشگاه .....	
نام و نام خانوادگی داوران	امضاء	تاریخ
۱-	(داور خارجی عضو هیأت علمی دانشگاه .....	
۲-	(داور داخلی)	
شورای گروه تخصصی ..... در تاریخ ..... در محل ..... با حضور اعضای مربوطه		
تشکیل و موضوع پایان نامه خانم		
..... با عنوان ..... آقای		
.....		

بررسی و به تصویب رسید.

نام و نام خانوادگی اعضای شورا	امضاء	تاریخ
۱-		
۲-		
۳-		
۴-		
۵-		
۶-		

نام و نام خانوادگی مدیر گروه:	امضاء	تاریخ
-------------------------------	-------	-------

**تذکر:** لازم است پروپوزال دانشجویان از تاریخ تأیید در شورای گروه تخصصی تا زمان طرح در شورای پژوهشی دانشکده بیشتر از یکماه نگذرد.

**تذکر:** لازم است قبل از تصویب پروپوزال در شورای پژوهشی دانشکده، شرایط احراز و ظرفیت پذیرش اساتید راهنما و مشاور مطابق بخشنامه‌های مربوطه توسط پژوهش دانشکده کنترل شود.

۹- صورتجلسه شورای (پژوهشی) دانشکده:

موضوع و طرح تحقیق پایان‌نامه ..... خانم  
آقای ..... دانشجوی مقطع کارشناسی ارشد گروه .....  
گرایش ..... که به تصویب کمیته گروه تخصصی مربوطه رسیده است، در جلسه مورخ .....  
شورای (پژوهشی) دانشکده طرح شد و پس از بحث و تبادل نظر مورد تصویب اکثریت اعضا قرار گرفت.

ردیف	نام و نام خانوادگی	نوع رأی (موافق یا مخالف)	محل امضاء	توضیحات
۱				
۲				
۳				
۴				
۵				
۶				
۷				

نام و نام خانوادگی مدیر/کارشناس پژوهشی دانشکده:

امضاء تاریخ

نام و نام خانوادگی ریاست دانشکده

امضاء تاریخ

این فرم باید توسط دانشجو تکمیل شود

فرم سازمان مرکزی دانشگاه آزاد اسلامی

فرم الف - فرم اطلاعات پایان نامه کارشناسی ارشد

نام واحد دانشگاهی: واحد علوم و تحقیقات	
عنوان پایان نامه کارشناسی ارشد:	
نام و نام خانوادگی دانشجو:	نیمسال تحصیلی:
شماره دانشجویی:	تعداد واحد پایان نامه:
رشته تحصیلی:	کد رشته:
<input type="checkbox"/> فنی و مهندسی	<input type="checkbox"/> علوم پایه
<input type="checkbox"/> کشاورزی	<input type="checkbox"/> هنر
<input type="checkbox"/> علوم انسانی	گرایش:
نام و نام خانوادگی استاد راهنما ۱:	
مرتبه علمی: استادیار <input type="checkbox"/> دانشیار <input type="checkbox"/> استاد <input type="checkbox"/>	رشته تحصیلی:
کد شناسایی استاد راهنما:	
نام و نام خانوادگی استاد مشاور ۱:	
مرتبه علمی: استادیار <input type="checkbox"/> دانشیار <input type="checkbox"/> استاد <input type="checkbox"/> مربی <input type="checkbox"/>	رشته تحصیلی:
کد شناسایی استاد راهنما:	

دستاوردهای حاصل از نتایج حاصل از پایان نامه کارشناسی ارشد یا رساله دکتری

تمامی تولیدات علمی مستخرج از پایان نامه/رساله (در قالب کتاب، مقاله، طرح تحقیقاتی، اختراع، اکتشاف و ...) با رعایت موارد ذیل قابل انتشار است.

الف) نویسنده اول مقاله باید به نام دانشجو و به عنوان تنها آدرس دانشگاه آزاد اسلامی واحد علوم و تحقیقات باشد.

ب) عهده دار مکاتبات (Corresponding Author) می تواند استاد راهنما با آدرس مؤسسه محل خدمت خود یا دانشجو با نشانی دانشگاه آزاد اسلامی واحد علوم و تحقیقات باشد.

ج) فقط اسامی افراد مرتبط با پایان نامه یا رساله (دانشجو -استاد راهنما-استاد مشاور) در مستخرجات ذکر شود.

د) آدرس دقیق دانشگاه و واحد و رعایت ترتیب و توالی آن باید به صورت زیر آورده شود:

به انگلیسی:

Department of تخصصی , Science and Research branch, Islamic Azad University, Tehran, Iran.

به فارسی:

دانشگاه آزاد اسلامی، واحد علوم و تحقیقات ، گروه .....، تهران، ایران.

اینجانبان اساتید راهنما و مشاور پایان نامه / رساله، خانم/آقای .....، دانشجوی مقطع

..... رشته ..... گرایش..... با عنوان

.....»

«.....»

موارد فوق را می پذیریم

نام و نام خانوادگی استاد راهنما:

تاریخ و امضاء

نام و نام خانوادگی استاد راهنما:

تاریخ و امضاء

نام و نام خانوادگی استاد مشاور:

تاریخ و امضاء

نام و نام خانوادگی استاد مشاور:

تاریخ و امضاء

در تاریخ ..... فرم مزبور که به امضاء اساتید محترم راهنما و مشاور رسیده است، دریافت گردید.

امضاء مدیر پژوهش دانشکده

تذکر: لازم است اساتید راهنما و مشاور انتخابی پس از مطالعه فرم مذکور، نسبت به تکمیل مشخصات و

توشیح آن شخصاً اقدام نمایند تا از هر گونه تخلفات احتمالی جلوگیری گردد.

دستورالعمل نحوه تدوین و استفاده از نتایج حاصل از پایان نامه کارشناسی ارشد یا رساله دکتری

الف) کلیه مطالب و مندرجات پایان نامه/ رساله بر اساس اصول علمی و حاصل از تحقیقات خودم تهیه شود و در صورت استفاده از مطالب، نتایج تحقیقات، نقل قول ها، جداول و نمودارهای دیگران در پایان نامه/رساله، منابع و مآخذ آن به نحوی که قابل تشخیص و تفکیک از متن اصلی باشد قید گردد.

ب) در صورتیکه از نتایج تحقیقاتم علاوه بر پایان نامه / رساله، کتاب، مقاله، اختراع، اکتشاف و هر گونه تولیدات علمی حاصل شود، صرفاً بنام دانشگاه آزاد اسلامی واحد علوم و تحقیقات تهران بوده و این موضوع صراحتاً در مکاتبات و تولیدات اینجانب درج و بر اساس ضوابط دانشگاه اقدام نمایم.

ج) در صورت استفاده از کمکهای مالی و غیر مالی نهادهای دولتی و غیر دولتی از موضوع تحقیق اینجانب مراتب را کتباً به دانشکده اطلاع دهم در غیر اینصورت دانشکده مجاز به تغییر عنوان پایان نامه یا سایر اقدامات حقوقی می باشد.

د) آدرس دقیق دانشگاه و واحد و رعایت ترتیب و توالی آن به عنوان تنها آدرس در تولیدات علمی مستخرج از پایان نامه/رساله باید به صورت زیر آورده شود:

آدرس دانشگاه و واحد به فارسی: دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران، گروه .....، تهران، ایران.

آدرس دانشگاه و واحد به انگلیسی:

Department of تخصصی , Science and Research branch, Islamic Azad University, Tehran, Iran.

اینجانب ..... دانشجوی ورودی ..... مقطع ..... رشته .....  
گرایش ..... که موضوع پایان نامه/رساله ام تحت عنوان:  
.....»

«.....»

در شورای گروه تخصصی مطرح و به تصویب رسیده است موارد فوق را مطالعه کرده و پذیرفتم

نام و نام خانوادگی دانشجو

تاریخ و امضاء

تذکر: بدیهی است چنانچه تحت هر شرایطی و در هر زمان، دانشگاه خلاف موارد ذکر شده را مشاهده نماید نسبت به تصمیم اتخاذ شده هیچگونه ادعایی نداشته و حق هرگونه اعتراضی را از خود سلب و ساقط می نمایم.

در تاریخ ..... فرم مزبور که توسط آقای / خانم ..... به امضاء رسیده است، دریافت گردید.

امضاء مدیر پژوهش دانشکده

فرم تأیید استعمال کتابخانه مرکزی و پژوهشگاه علوم و فناوری اطلاعات ایران  
(ایران داک)

اینجانبان اساتید راهنما پایان نامه / رساله آقای / خانم ..... دانشجوی مقطع  
دکتری تخصصی / کارشناسی ارشد / دکتری حرفه‌ای رشته ..... با  
عنوان ..... استعمال‌های اخذ شده کتابخانه مرکزی  
و پژوهشگاه علوم و فناوری اطلاعات ایران (ایران داک) مبنی بر عناوین مشابه عنوان فوق را مطالعه نموده و  
با عنایت به کلمات کلیدی عنوان پروپوزال در سایت‌های [www.irandoc.ac.ir](http://www.irandoc.ac.ir) و [sika.iau.ir](http://sika.iau.ir)  
تکراری نبودن عنوان پروپوزال مذکور مور تأیید می‌باشد.

نام و نام خانوادگی استاد راهنمای اول:

نام و نام خانوادگی استاد راهنمای دوم: