بیمارستان قلب شهید رجایی

طراحی و پیاده سازی بانک یکپارچه زیست_داده و درگاه چندمنظوره محققین حوزه قلب و عروق کشور

شناسنامه طرح

کد رهگیری طرح	401185
تاریخ تصویب پیش پروپوزال	
عنوان طرح	طراحی و پیاده سازی بانک یکپارچه زیست_داده و درگاه چندمنظوره محققین حوزه قلب و عروق کشور
عنوان لاتین طرح	Development and implementation of a national integrated cardiovascular biodata bank AND researchers' multi-purpose portal/gateway
تلفن	09158061511
پست الکترونیکی	smtara@gmail.com
نوع مطالعه	انفورماتیک زیست پزشکی
تاریخ شروع	1402/01/01
تاریخ خاتمه	1404/05/31
آیا طرح چند مرکزی است؟	بله
مرکز/مراکز دیگر	داخل کشور
نام سازمان تصویب کننده اولیه پروپوزال	انستیتو آموزشی، تحقیقاتی، و درمانی قلب و عروق شهیدرجایی
محل اجرای طرح	مرکز فناوری و نوآوری بیمارستان الکترونیک
محل اجرای طرح	بیمارستان قلب شهید رجایی
سازمان مجری	بیمارستان قلب شهید رجایی
سازمان مجری	
دانشکده/محل خدمت	Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences
رشته تخصصی	قلب و عروق
توضیحات	
نوع طرح ها	

مجری همکاران

نام و نام خانوادگی	سمت در طرح	نوع همکاری	توضیحات
محمود تارا	مجری اصلی / نویسنده مقاله	طراحی و تدوین طرح	
فریدون نوحی بزنجانی	همکار طرح	مشاور	عضو کمیته راهبری طرح
مجید ملکی	همکار طرح	مشاور	عضو کمیته راهبری طرح
سعیده مظلوم زاده	همکار طرح	مشاور	عضو کمیته راهبری طرح
علی زاهدمهر	همکار طرح	مشاور	
رضا گل پیرا	همکار طرح	مشاور	عضو کمیته راهبری طرح
هومن بخشنده آبکنار	همکار طرح	مشاوره و آنالیز آماری	
سمیرا کلائی نیا	همکار طرح	مشاور	
پرهام صادقی پور	همکار طرح	مشاور	
حمیدرضا پورعلی اکبر	همکار طرح	مشاور	
لیلا عبدالکریمی	همکار طرح	مشاور	
سید محمد دلیلی	همکار طرح	مشاور	
مهدی دلیری	همکار طرح	مشاور	
مجید حق جو	همکار طرح	مشاور	
احسان بیطرف	همکار طرح	سایر	مدیر فنی طرح
شیوا خالق پرست	همکار طرح	مشاور	
هادی ملک	همکار طرح	مشاور	
مهشید ملکوتیان	همکار طرح	مشاور	
محمد جواد عالم زاده انصاری	همکار طرح	مشاور	
سعید حسینی	همکار طرح	مشاور	
مریم شجاعی فرد	همکار طرح	مشاور	
شبنم مددی	همکار طرح	مشاور	
مازیار غلامپور دهکی	همکار طرح	مشاور	
آذین علیزاده اصل	همکار طرح	مشاور	
شبنم مددی	همکار طرح	مشاور	
مائده عربیان	همکار طرح	مشاور	
آویسا طبیب	همکار طرح	مشاور	

دانشده/مرکز مربوطه

رده	نوع ارتباط با مرکز
گروه داخلی	وارد کننده

اطلاعات تفصیلی

آیتم ها	متن
بیان مسئله	در ایران، شایع ترین علل مرگ از بیماری های عفونی و اسهالی در سال 1960 به بیماری های قلبی عروقی (CVD) در چند دهه پیش منتقل شده است. بیماری های قلبی عروقی، اولین عامل اصلی مرگ و میر بود و یک میلیون سال زندگی تعدیل شده با ناتوانی (DALYs) منجر به 46 درصد از کل مرگ‌ها و 20 تا 23 درصد بار بیماری در ایران شده است. بیماری ایسکمیک قلبی و سکته مغزی به ترتیب اولین و دومین علت مرگ و DALY در ایران محسوب می شوند. اپیدمی افزایش بیماری های قلبی عروقی ممکن است به تغییرات اجتماعی–اقتصادی و فرهنگی، تغذیه، فعالیت بدنی ناکافی ، صنعتی شدن و شهرنشینی و افزایش امید به زندگی، افزایش ریسک فاکتورهای متابولیکی و فیزیکی، دسترسی و مقرون به صرفه بودن نبودن مراقبت‌های اولیه و درمان و انطباق پایین به دلیل مشکلات اقتصادی و روانی مرتبط باشد. از این رو، برنامه ریزی و اجرای راهبردهای پیشگیری و کنترل بیماری و عوامل خطرزای آن در سرلوحه دستور کار وزارت بهداشت در سال های اخیر قرار گرفته است. راهکارهای ارتقای سلامت برای پیشگیری و کنترل عوامل خطر CVD، تشخیص زودهنگام بیماری و درمان حوادث حاد و مزمن بیماری های قلبی عروقی از عناصر ضروری برای کاهش بار بیماری های قلبی عروقی در ایران هستند. برنامه ریزی و اجرای راهبردهای پیشگیری و کنترل بیماری و عوامل خطر آن در رأس دستور کار وزارت بهداشت در سال های اخیر قرار گرفته است..[15] ایران به عنوان یکی از کشور های بزرگ و تاثیر گذار منطقه در خاورمیانه و آسیا علی رغم داشتن تنوع بسیار بالایی از بیماران و جمعیت 85 میلیونی هنوز سیستم بیوبانک و رجیستری منظم و سازمان یافته ای ندارد. در سال های اخیر که داده های الکترونیک سلامت کشور به شدت در حال شکل گیری جزیره ای، پراکنده، و غیر استاندارد، و در عین حال در حجم و سرعت بالا هستند، شاید مهمترین ضرورت اجرایی برای محققین و زیرساخت پژوهش حوزه سلامت، راه اندازی بایودیتابانک هایی هستند که علاوه بر انتظام و یکپارچگی داده ها، بتوانند دو نوع درگاه را برای داده های نظام سلامت محقق سازند: 1. درگاهی برای ورود کلیه انواع داده های چندوجهی سلامت و پزشکی بیماران و بیماریهای قلب و عروق 2. درگاهی برای ارایه خدمات اطلاعاتی و دانشی به محققین حوزه قلب و عروق کشور بایوبانک ها با هدف حل و فصل چالش های فراوان یکپارچگی داده های سلامت و ایجاد درگاه های و بر اساس آخرین فناوری ها و نوآوری های مهندسی داده طراحی می گردند. بایوبانک های قلب و عروق به پژوهشگران این حوزه اجازه می دهد تا نشانگرهای زیستی جدیدی را برای پیش‌بینی پیشرفت یا پیامدهای بیماری شناسایی کنند و به آنها اجازه می دهد تا تأثیر ژنتیک بر خطر بیماری های چند عاملی و اثرات ترکیبی ژنتیک، سبک زندگی (به عنوان مثال، سبک زندگی، سیگار کشیدن، عادات غذایی وفعالیت بدنی) و سایر عوامل محیطی (مانند تحصیلات، آلودگی، فصلی و استرس) در ایجاد بیماری‌های رایج مانند آنژین ناپایدار، انفارکتوس میوکارد، بیماری‌های شریانی محیطی و سکته مغزی به دست آورند. در کنار همه موارد مذکور، شناسایی مسیرهای دخیل در شروع یا پیشرفت بیماری ممکن است منجر به کشف اهداف درمانی جدید شود. همچنین می تواند منجر به توسعه داروهایی متناسب با ویژگی‌های فردی بیماران شود..[14]. مرکز آموزشی، تحقیقاتی و درمانی قلب و عروق شهیدرجایی، اولین و برترین مرکز حوزه قلب و عروق کشور و جزو یکی از معدود بیمارستان های ممتاز در رده بندی کیفی کشور است. این مرکز همواره پیشتاز غالب نوآوری ها و فناوری های حوزه قلب و عروق چه در حوزه آموزشی و در حوزه تحقیقاتی و درمان بوده و میزبان پژوهشکده قلب و عروق ایران نیز می باشد. این مرکز دارای تعداد بسیار بالایی از مراجعات همراه با تنوع بالایی از بیماری های قلبی شایع و ناشایع می باشد. سالانه نزدیک به حداقل 20 هزار بیمار بستری و 200 هزار بیمارسرپایی قلبی در این مرکز مورد مدیریت قرار گرفته و داده های آنها در سامانه های مختلف این مرکز در طول چند دهه گذشته جمع آوری گردیده است. تنها در سیستم پکس این بیمارستان میلیون ها تصویر ثبت شده اعم از تصاویر اکوکاردیوگرافی، CT scan و MRI می باشد که هیچ پروسه اضافه ای اعم از پردازش تصاویر یا Deep learning یا Machine learning بر روی آن ها انجام نشده است. از طرفی، بسیاری از آنومالی های ناهنجاری که در سراسر جهان نایاب هستند، در مرکز قلب رجایی دارای تعداد قابل توجهی هستند که مقالات حاصل از آن ها دارای ارزش بالایی از روایی و پایایی نسبت به سایر مراکز می باشد. همه این ها در حالیست که متاسفانه یک سیستم و زیر ساخت واحد برای اتصال داده های تصویر برداری و پاراکلنیک بیمار به داده های کلینیکی و یا ژنتیکی وجود ندارد و این ثروت عظیم به عنوان یکی از مخازن باارزش بالقوه اطلاعات بیماری های قلب و عروق کشور، بدون استفاده باقی مانده است. از طرف، علاوه بر مرکز قلب شهیدرجایی، مراکز قلب

متعددی در کشور همچون مرکز قلب دانشگاه تهران، مرکز قلب بیمارستان لواسانی تهران، فاطمه زهرا ساری، امام علی کرمانشاه، چمران اصفهان، سینا اصفهان، شفای گلستان، سید الشهداء ارومیه، مدنی تبریز، فرشچیان همدان، و گاندی تهران نیز وجود دارند که می توانند به عنوان مراکز بالقوه همکار در داده های بایوبانک در نظر گرفته شوند.

به همه دلایل فوق، و مرور تجربیات سایر کشور ها (در بخش مرور متون) لزوم تاسیس یک بایوبانک برای بیماری های قلبی عروقی که جایگاه اول موریتاریتی را دارد کاملا اجتناب ناپذیر است. هر چند راه اندازی بایودیتابانک های چندمرکزه، علاوه بر پیچیدگی های فنی-اجرایی نیازمند زیرساخت های امن پرهزینه ای برای دسترسی پرسرعت و مطمئن توسط محققین هستند، ولی اثرات، آورده ها، و پیامدهای مثبت آن در سطح تحقیقات حوزه سلامت و به ویژه حوزه بیماری های قلب و عروق کشور که مسئول یکی از بزرگترین بار بیماری ها در ایران می باشد می تواند هزینه های اقتصادی آن را کاملا توجیه کند و دریچه ای جدید را به سوی محققین حیطه سلام و بیماری قلب و عروق کشور باز نموده و نتایج آن راهگشای مدیران نظام سلامت کشور باشد.

ضرورت اجرا

در مقطع کنونی کشور که داده های الکترونیک سلامت به شدت در حال شکل گیری جزیره ای، پراکنده، و غیر استاندارد، و در عین حال در حجم و سرعت بالا هستند، شاید مهمترین ضرورت اجرایی برای محققین و زیرساخت پژوهش حوزه سلامت، راه اندازی بایودیتابانک هایی هستند که علاوه بر انتظام و یکپارچگی داده ها، بتوانند دو نوع درگاه را برای داده های نظام سلامت محقق سازند:

- درگاهی برای ورود کلیه انواع داده های چندوجهی سلامت و پزشکی بیماران و بیماریهای قلب و عروق
- درگاهی برای ارایه خدمات اطلاعاتی و دانشی به محققین حوزه قلب و عروق کشور

پنج چالش برای شکل گیری بایودیتابانک های منسجم و کارآمد در کلیه کشورها و به ویژه کشور ما وجود دارند:

- **چالش یکپارچه سازی داده ها:** این چالش یکی از بزرگترین چالش های مهم در علوم داده است و غالبا مرتبط با داده هایی است که یا در سیستم های مختلف و غیر استاندارد (غیر مبتنی بر کدینگ یکسان) تولید می شوند و یا دارای ساختار های پیچیده ای هستند (مانند عدم قطعیت، متن باز، مبهم، و کیفی) که استانداردسازی آنها را دشوار می سازد. لذا قابلیت همخوانی سامانه های متصل به بایودیتابانک ها برای ثبت تمیز داده های سلامت، و توافق بر روی مجموعه داده های مشترک و تعریف استانداردهای کیفیت برای یکپارچه سازی داده های سلامت و نیز رابط های اتصال داده ها در این زمینه بسیار مهم می باشند.
- **چالش داده های بزرگ:** دومین چالش در مدیریت داده های سلامت، چالش بزرگی داده ها (Big Data) است. داده های بزرگ معمولاً با سه V تعریف می شوند که شامل : حجم، سرعت و تنوع می باشند(1). بعد از رفع چالش یکپارچه سازی در حوزه سلامت، معمولا با چالش داده های بزرگ مواجه می شویم. تنوع داده ها در حوزه سلامت بسیار بالا است از داده های متنی تا ساختار یافته تا داده های تصاویر پزشکی و بیوسیگنال های حیاتی به عنوان داده سلامت محسوب می شوند. ژنومیک و بیوانفورماتیک نیز یکی دیگر از موضوعات کلیدی جدید هستند که داده های بزرگ و سلامت را به هم پیوند می دهد (2) که برای پردازش داده‌های توالی‌یابی و بیوانفورماتیک ترجمه‌ای به حجم عظیمی از ذخیره‌سازی و قدرت تحلیلی نیاز دارند.
- **چالش درگاه های کارآمد، امن، و مطمئن دسترسی به داده و اطلاعات:** یکی از بزرگترین چالش هایی که ..... آیا بهترین منافع جامعه در هنگام مذاکره درباره مسیرهای دسترسی به داده در نظر گرفته می‌شود؟ برای مصرف کننده، یک نگرانی اصلی دسترسی شخص ثالث و کنترل داده است. شرکت ها نیز علاقه مند به جمع آوری و استفاده داخلی از اطلاعات هستند، اما نگران افشای مالکیت معنوی نیز هستند. منافع مالی و غیرمالی ارائه دهندگان خدمات بهداشتی و درمانی نیز ممکن است با به اشتراک گذاری اطلاعات به چالش کشیده شود. در نتیجه ممکن است این سازمان‌ها تمایل کمتری به افشای اطلاعات مربوط به عملکرد داشته باشند یا ممکن است فعالانه برای ممانعت از ارسال داده یا داده سازی از بیماران تلاش کنند(3)، که باعث نگرانی‌هایی در مورد اعتبار داده‌ها می‌شود.
- **کیفیت و قابلیت اطمینان داده ها:** قابلیت اطمینان داده ها یکی دیگر از چالش هایی است که اغلب برای پیاده سازی و استفاده از سیستم های کلان داده در سلامت ذکر می شود (4). در سطح عملیاتی، داده‌های سلامت الکترونیکی که به صورت دستی تغذیه می‌شوند ممکن است در معرض خطا و سوگیری از ورود انسان باشند. با این حال، سیستم های منظم همچنین می توانند سوگیری سیستماتیک را در جمع آوری و تجزیه و تحلیل داده

ها وارد کنند (5). برای مثال، آمار بیمارستان های انگلیس (HES)، به عنوان یکی از جامع‌ترین مجموعه داده‌های بهداشتی در جهان در نظر گرفته می‌شود که سالانه بیش از 125 میلیون پرونده بیمار را برای تمام پذیرش‌ها، وزیت های سرپایی و مراجعه اورژانس که در بیمارستان های دولتی انگلیس رخ می دهد را جمع آوری کرده است. علیرغم ارزش فوق العاده ای که برای تحقیقات و عمل بالینی دارد، HES از لحاظ تاریخی به دلیل تشخیص های ثانویه غیرقابل اعتماد مورد انتقاد قرار گرفته است (6).

- **محرمانگی، حریم خصوصی، و امنیت داده ها:** متأسفانه، دسترسی به داده ها و خطرات محرمانه بودن ارتباط مستقیم دارند (7). حریم خصوصی به عنوان یک حقوق اساسی بشر در اعلامیه جهانی حقوق بشر در مجمع عمومی سازمان ملل متحد در سال 1948 تعریف شده است. با این حال، هنوز هیچ اتفاق نظری در مورد اینکه چه چیزی حریم خصوصی است وجود ندارد (8). این موضوع باعث پیچیده شدن اقدامات فنی برای بحث امنیت و حفظ حریم خصوصی خواهد بود. این مسئله در کشور ما از آن جهت بسیار اهمیت دارد که قوانین مشخصی برای تعریف حدود محرمانگی در مورد داده های سلامت و نیز دسترسی به آنها وجود نداشته و یا در مراحل اولیه خود می باشند.

بایوبانک ها با هدف حل و فصل چالش های پنج گانه فوق و بر اساس آخرین فناوری ها و نوآوری های مهندسی داده طراحی می گردند. بایوبانک های قلب و عروق به پژوهشگران این حوزه اجازه می دهد تا نشانگرهای زیستی جدیدی را برای پیش‌بینی پیشرفت یا پیامدهای بیماری شناسایی کنند و به آنها اجازه می دهد تا تأثیر ژنتیک بر خطر بیماری های چند عاملی و اثرات ترکیبی ژنتیک، سبک زندگی (به عنوان مثال، سبک زندگی، سیگار کشیدن، عادات غذایی وفعالیت بدنی) و سایر عوامل محیطی (مانند تحصیلات، آلودگی، فصلی و استرس) در ایجاد بیماری‌های رایج مانند آنژین ناپایدار، انفارکتوس میوکارد، بیماری‌های شریانی محیطی و سکته مغزی به دست آورند. در کنار همه موارد مذکور، شناسایی مسیرهای دخیل در شروع یا پیشرفت بیماری ممکن است منجر به کشف اهداف درمانی جدید شود. همچنین می تواند منجر به توسعه داروهایی متناسب با ویژگی‌های فردی بیماران شود..[14]. در انگلستان UKbiobank امکان دسترسی به پژوهشگران سایر کشور ها را می دهد تا با استفاده از داده های بیوبانک، بسیاری از مطالعات اپیدیمیولوژیک را در سطح بسیار وسیع تری انجام دهند.

در ایران، شایع ترین علل مرگ از بیماری های عفونی و اسهالی در سال 1960 به بیماری های قلبی عروقی (CVD) در چند دهه پیش منتقل شده است. بیماری های قلبی عروقی، اولین عامل اصلی مرگ و میر بود و یک میلیون سال زندگی تعدیل شده با ناتوانی (DALYs) منجر به 46 درصد از کل مرگ‌ها و 20 تا 23 درصد بار بیماری در ایران شده است. بیماری ایسکمیک قلبی و سکته مغزی به ترتیب اولین و دومین علت مرگ و DALY در ایران محسوب می شوند. اپیدمی افزایش بیماری های قلبی عروقی ممکن است به تغییرات اجتماعی–اقتصادی و فرهنگی، تغذیه، فعالیت بدنی ناکافی ، صنعتی شدن و شهرنشینی و افزایش امید به زندگی، افزایش ریسک فاکتورهای متابولیکی و فیزیکی، دسترسی و مقرون به صرفه بودن نبودن مراقبت‌های اولیه و درمان و انطباق پایین به دلیل مشکلات اقتصادی و روانی مرتبط باشد. از این رو، برنامه ریزی و اجرای راهبردهای پیشگیری و کنترل بیماری و عوامل خطرزای آن در سرلوحه دستور کار وزارت بهداشت در سال های اخیر قرار گرفته است. راهکارهای ارتقای سلامت برای پیشگیری و کنترل عوامل خطر CVD، تشخیص زودهنگام بیماری و درمان حوادث حاد و مزمن بیماری های قلبی عروقی از عناصر ضروری برای کاهش باربیماری های قلبی عروقی در ایران هستند. برنامه ریزی و اجرای راهبردهای پیشگیری و کنترل بیماری و عوامل خطر آن در رأس دستور کار وزارت بهداشت در سال های اخیر قرار گرفته است..[15]

ایران به عنوان یکی از کشور های بزرگ و تاثیر گذار منطقه در خاورمیانه و آسیا علی رغم داشتن تنوع بسیار بالایی از بیماران و جمعیت 85 میلیونی هنوز سیستم بیوبانک و رجیستری منظم و سازمان یافته ای ندارد. مرکز آموزشی، تحقیقاتی و درمانی قلب و عروق شهیدرجایی، اولین و برترین مرکز حوزه قلب و عروق کشور و جزو یکی از معدود بیمارستان های ممتاز در رده بندی کیفی کشور است. این مرکز همواره پیشتاز غالب نوآوری ها و فناوری های حوزه قلب و عروق چه در حوزه آموزشی و در حوزه تحقیقاتی و درمان بوده و میزبان پژوهشکده قلب و عروق ایران نیز می باشد. این مرکز دارای تعداد بسیار بالایی از مراجعات همراه با تنوع بالایی از بیماری های قلبی شایع و ناشایع می باشد. سالانه نزدیک به حداقل 20 هزار بیمار بستری و 200 هزار بیمارسرپایی قلبی در این مرکز مورد مدیریت قرار گرفته و داده های آنها در سامانه های مختلف این مرکز در طول چند دهه گذشته جمع آوری گردیده است. تنها در سیستم پکس این بیمارستان میلیون ها تصویر ثبت شده اعم از تصاویر اکوکاردیوگرافی، CT scan و MRI می باشد که هیچ پروسه اضافه ای اعم از پردازش تصاویر یا Deep learning یا Machine learning بر روی آن ها انجام نشده است. از طرفی، بسیاری از آنومالی های ناهنجاری که در سراسر جهان نایاب هستند، در مرکز قلب رجایی دارای تعداد قابل

توجهی هستند که مقالات حاصل از آن ها دارای ارزش بالایی از روایی و پایایی نسبت به سایر مراکز می باشد. همه این ها در حالیست که متاسفانه یک سیستم و زیر ساخت واحد برای اتصال داده های تصویر برداری و پاراکلنیک بیمار به داده های کلینیکی و یا ژنتیکی وجود ندارد و این ثروت عظیم به عنوان یکی از مخازن باارزش بالقوه اطلاعات بیماری های قلب و عروق کشور، بدون استفاده باقی مانده است. از طرف، علاوه بر مرکز قلب شهیدرجایی، مراکز قلب متعددی در کشور همچون مرکز قلب دانشگاه تهران، مرکز قلب بیمارستان لواسانی تهران، فاطمه زهرا ساری، امام علی کرمانشاه، چمران اصفهان، سینا اصفهان، شفای گلستان، سید الشهداء ارومیه، مدنی تبریز، فرشچیان همدان، و گاندی تهران نیز وجود دارند که می توانند به عنوان مراکز بالقوه همکار در داده های بایوبانک در نظر گرفته شوند.

به همه دلایل فوق، و مرور تجربیات سایر کشور ها (در بخش مرور متون) لزوم تاسیس یک بایوبانک برای بیماری های قلبی عروقی که جابگاه اول مورتاریتی را دارد کاملا اجتناب ناپذیر است. هر چند راه اندازی بایودیتابانک های چندمرکزه، علاوه بر پیچیدگی های فنی-اجرایی نیازمند زیرساخت های امن پرهزینه ای برای دسترسی پرسرعت و مطمئن توسط محققین هستند، ولی اثرات، آورده ها، و پیامدهای مثبت آن در سطح تحقیقات حوزه سلامت و به ویژه حوزه بیماری های قلب و عروق کشور که مسئول یکی از بزرگترین بار بیماری ها در ایران می باشد می تواند هزینه های اقتصادی آن را کاملا توجیه کند و دریچه ای جدید را به سوی محققین حیطه سلام و بیماری قلب و عروق کشور باز نموده و نتایج آن راهگشای مدیران نظام سلامت کشور باشد.

بررسی متون

بایوبانک های مرتبط با جمعیت در مقیاس بزرگ در چندین کشور در حال راه‌اندازی هستند و نه تنها امکان تحقیق در مورد بیماری‌های فردی را فراهم می‌کنند، بلکه رویکردهایی به طیف گسترده‌ای از مسائل مرتبط با سلامت، مانند فعالیت بدنی، خوردن، آشامیدن یا آموزش برخی از پروژه‌های تحقیقاتی ژنومیک که شامل کل جمعیت می شوند را نیز شامل می شوند [1،2]. بایوبانک کشور انگلستان، به عنوان یک نمونه ملی بایوبانک ها، یک منبع داده های سلامت ملی است که هدف آن بهبود پیشگیری، تشخیص و درمان طیف گسترده ای از بیماری های جدی و تهدید کننده زندگی از جمله سرطان، بیماری های قلبی، سکته مغزی، دیابت، پوکی استخوان، اختلالات چشمی، افسردگی و اشکال زوال عقل است [3, 4]. این بایوبانک در سال های 2006 تا 2010 در سراسر کشور ، حدود 500000 نفر از 40 تا 69 سال را به کار گرفته است تا نمونه های خون، ادرار و بزاق را برای تجزیه و تحلیل آینده، اطلاعات دقیق در مورد خود ارائه نموده و موافقت نمایند که سلامت خود را پیگیری کنند. بایو بانک انگستان توسط خیریه پزشکیWellcome Trust، شورای تحقیقات پزشکی، وزارت بهداشت، دولت اسکاتلند و آژانس توسعه منطقه ای شمال غربی تاسیس شده و توسط دولت و مجلس ولز تامین مالی می شود.

بایوبانک کشور استونی، زیست بانک مبتنی بر جمعیت در مرکز ژنوم استونی است که توسط دانشگاه تارتو (استونی) در حال اجرا است [5]. اندازه گروه در حال حاضر 51535 نفر از 18 سال و بالاتراست که در آن اطلاعاتی در مورد داده های شخصی (به عنوان مثال، سن و جنس)، داده های شجره نامه (خانواده) ساختار، سابقه تحصیلی و شغلی، داده های سبک زندگی (به عنوان مثال، سیگار کشیدن، رژیم غذایی و فعالیت فیزیکی)، سابقه پزشکی، وضعیت سلامت فعلی (به عنوان مثال، فشار خون بالا، دیابت، کلسترول خون بالا و بیماری های قلبی عروقی ، دارو و همچنین اندازه گیری های آنتروپومتریک، فشار خون و ضربان قلب در حالت استراحت در طول دوره اندازه گیری شده و جمع آوری می شود. [6]

deCODE genetics یک شرکت خصوصی است که در ایسلند مستقر است و بین سال‌های 1998 و در سال 2000، مجوز ساخت و اجرای پایگاه داده بخش سلامت ایسلند را دریافت کرد[5, 7]. آنها داده های ژنوتیپی و پزشکی بیش از 140000 شرکت کننده داوطلب را جمع آوری کرده اند. بیش از نیمی از جمعیت بزرگسال ایسلند را به خود اختصاص داده است. با استفاده از جامع منحصر به فرد ایسلند سوابق شجره نامه، deCODE همچنین یک پایگاه داده شجره نامه ای را گردآوری کرده است که کل آن را پوشش می دهد در حال حاضر deCODE یک بانک زیستی از نمونه های ژنتیکی از 100000 جلد ایسلندی ایجاد کرده است که حاصل موارد پژوهشی آن این بوده است که ژن های دخیل در بیش از 50 بیماری رایج و همچنین چندین ژن خاص ایجاد کننده بیماری، از جمله ژن های مربوط به سکته قلبی، سکته مغزی و و آنوریسم عروقی شناسایی شده است [7، 8].

پروژه CARTaGENE یک پروژه Centre Hospitalier Universitaire Sainte-Justine (کبک، کانادا) است که زیرساختی شامل پایگاه داده سلامت و بانکی از نمونه های بیولوژیکی را با هدف هدایت در درازمدت به پیشگیری بهتر، تشخیص و درمان بیماری‌های مزمن مانند بیماری‌های قلبی، دیابت، سکته مغزی و سرطان و در نتیجه بهبود سیستم

مراقبت‌های بهداشتی کبک ایجاد شده است. [18،19]. بیش از 20000 زن و مرد 40-69 ساله ساکن کبک در مرحله اولیه CARTaGENE فاز A از ژوئن 2009 تا اکتبر 2010 شرکت کردند. هر یک از شرکت کنندگان اطلاعاتی در مورد سلامت و شیوه زندگی خود ارائه کردند و نمونه خون و ادرار را ارائه کردند. برخی از شرکت کنندگان نیز شرکت در یک گزینه شجره نامه ای را که با همکاری پروژه BALSAC انجام شد، انتخاب کردند[5, 10]

17000 شرکت کننده برنامه ملی بانکداری زیستی سوئد یک برنامه ملی مشترک از دو طرح سوئدی در زمینه ژنومیک عملکردی، سوئجین و کنسرسیوم والنبرگ شمال است که توسط بنیاد والنبرگ و دانشگاه های بزرگ سوئد حمایت می شود[11] برنامه ملی Biobank از سال 2002 تا 2005 با بودجه کلی بیش از 51 میلیون کرون سوئد اجرا شده است. بانک زیستی نمونه های جمع آوری شده را برای تحقیقات یا در طول مراقبت های معمول پزشکی ذخیره می کند. تعداد کل نمونه ها در بانک های زیستی سیستم مراقبت های بهداشتی سوئد تقریباً 50 تا 100 میلیون نمونه انسانی تخمین زده می شود که نرخ افزایش تقریباً 3 تا 4 میلیون نمونه در سال است. به طور خاص، بانک زیستی تحقیقاتی مبتنی بر جمعیت دارای مجموعه ای جامع از اطلاعات فنوتیپی و محیطی، و همچنین بانک زیستی پوشش های بافی، پلاسما و گلبول های قرمز است که در دمای 80- درجه سانتی گراد یا پایین تر ذخیره شده اند. چهار شرکت کننده بانک‌های زیستی گروه تحقیقاتی (Preventive Medicine، Malmö Diet & Cancer، Malmö Medical Biobank، Umeå و Twin Gene Biobank) که همگی در سوئد واقع شده اند این بایوبانک را توسعه می دهند.

بایوبانک های کشورهای پیشرو در زمینه مطالعات پژوهشی حوزه قلب و عروق

در بررسی مقالات منتشر شده در حوزه قلب و عروق در سال های اخیر و مقایسه بین کشور های مختلف، از سال 2000 تا کنون، امریکا در جایگاه اول بوده است و سپس آلمان و بریتانیا قرار داشته اند. [13] این در حالی است که انگستان تا سال 2006 در این جمع حضور نداشته است و جایگاه آن را ژاپن در اختیار خود داشت اما از سال 2006 وارد این جمع شده است و این موضوع زمانی به وقوع پیوست که UK biobank وارد فاز اجرایی شد. حال وقتی به سازمان های رجیستری و ایجاد کننده بایوبانک در این کشور ها می پردازیم متوجه می شویم که قوی ترین بیوبانک ها در این کشور ها قرار دارند.

کشور امریکا با داشتن بایوبانک های بزرگی همچون بایوبانک استندفورد (SCVI Biobank) ، بایوبانک مایوکلینیک ، سری بایوبانک های دانشگاه های دوک و همچنین کلیولند دارای سیستم رجیستری بسیار گسترده ای است. آلمان با داشتن چندین بایوبانک مانند (KaBi-DHM Cardiovascular Biobank of the German Heart Centre Munich) و همچنین DZHK نیز درارای سیستم رجیستری کاملی است. از سوی دیگر انگلستان نیز با داشتن UK biobank به عنوان بزرگ ترین بزرگ ترین بیوبانک دنیا با مطالعه روی نیم میلیون فرد از جمله بیماران قلبی، نقش قابل توجهی در انجام پژوهش دارد.

PhysioNet، نام مستعار منبع داده تحقیقاتی برای سیگنال‌های فیزیولوژیکی پیچیده است که در سال 1999 تحت نظارت مؤسسه ملی بهداشت (NIH) تأسیس شد، همانطور که در ادامه توضیح داده می‌شود. مأموریت اصلی PhysioNet انجام و تسریع تحقیقات و آموزش پزشکی، تا حدی با ارائه دسترسی رایگان به مجموعه‌های بزرگی از داده‌های فیزیولوژیکی و بالینی و نرم‌افزار متن باز مرتبط می باشد. در همکاری با کنفرانس سالانه محاسبات در قلب و عروق[1]، PhysioNet میزبان یک سری چالش های سالانه است که بر روی مشکلات حل نشده در علوم بالینی و پایه تمرکز دارد. پلت فرم PhysioNet توسط اعضای آزمایشگاه فیزیولوژی محاسباتی MIT مدیریت می شود[12]

بانک اطلاعاتی دوک برای بیماری های قلبی عروقی[2] (DDCD) یک بانک اطلاعاتی است که شامل تمام سوابق کاتتریزاسیون قلبی برای بیمارانی است که بین 2 ژوئیه 1969 تا 17 مارس 2015 در مرکز پزشکی دانشگاه دوک تحت درمان قرار گرفته اند. DDCD بزرگترین و قدیمی ترین بانک اطلاعات قلب و عروق در جهان است. این مجموعه داده ها پس از تایید کمیته ای مستقل در اختیار محققان قرار می گیرد[13]. مجموعه داده های تحقیقاتی DukeCath: استخراج شده از DDCD، این مجموعه حاوی یک پرونده برای فردی هست که به هر روشی کاتتریزاسیون شده است و شامل ویژگی های بیمار، نتایج کاتتریزاسیون، درمان های مداخله ای و داده های پیامدهای طولانی مدت است که امکان ارزیابی روابط متغیرها با پیامدها را فراهم می کند. مجموعه داده آموزشی DukeCathR: این مجموعه داده آموزشی به طور خاص برای خدمت به عنوان کمک آموزشی برای استادان ایجاد شده است. بی نام شده و عمداً اصلاح شده است تا هرگونه تفسیر بالینی معنی دار یا قابل انتشار غیرممکن باشد.[13].

پایگاه داده جراحی قلب بزرگسالان[3] برترین رجیستری نتایج بالینی در جهان برای جراحی قلب بزرگسالان است. این

پایگاه داده که در سال 1989 راه اندازی شد، حاوی بیش از 7.5 میلیون پرونده جراحی قلب است و در حال حاضر نزدیک به 3800 پزشک شرکت کننده، از جمله جراحان و متخصصان بیهوشی، دارد (12).

CardioNet یک پایگاه داده جامع است که می تواند به عنوان یک مجموعه آموزشی برای مدل های هوش مصنوعی عمل کند و در تمام جنبه های مدیریت بالینی CVD ها کمک کند. این شامل اطلاعات استخراج شده از پرونده الکترونیک سلامت و نتایج آزمایشات مربوط به CVD است[14].

پروژه شبکه تحقیقات قلبی عروقی (CVRG) در حال ایجاد زیرساختی برای دسترسی یکپارچه ایمن به مطالعه داده ها و ابزارهای تجزیه و تحلیل است. ابزارهای CVRG با استفاده از مدل نرم‌افزار به‌عنوان سرویس توسعه داده می‌شوند و به کاربران اجازه می‌دهند از طریق مرورگر خود به ابزارها دسترسی داشته باشند، بنابراین نیازی به نصب و نگهداری نرم‌افزارهای پیچیده را از بین می‌برند. پروژه CVRG توسط موسسه ملی قلب ریه و خون پشتیبانی می شود. این پروژه در موسسه پزشکی محاسباتی در دانشگاه جان هاپکینز، گروه انفورماتیک زیست پزشکی در مرکز پزشکی دانشگاه واندربیلت، کالج محاسبات و انفورماتیک در دانشکده شارلوت، مرکز انفورماتیک جامع در دانشگاه اموری، کالج مهندسی و علوم کاربردی در دانشگاه استونی بروک مدیریت می شود(14).

بانک اطلاعات قلب سنگاپور (SCDB)، یک ثبت ملی که اطلاعات اپیدمیولوژیک و بالینی را در مورد بیماری‌های قلبی عروقی و روش‌هایی مانند مداخلات عروق کرونر از راه پوست (PCI)، جراحی بای پس عروق کرونر، الکتروفیزیولوژی و ضربان‌سازی و همچنین پذیرش نارسایی قلبی جمع‌آوری می‌کند. تمام بیمارستان های بازسازی شده در سنگاپور به SCDB[4] متصل می باشند.

[1] Computing in Cardiology Conference

[2] Duke Databank for Cardiovascular Disease

[3] Adult Cardiac Surgery Database

[4] Singapore Cardiac Data bank

منابع

1. Kopelman, L.M., Philosophy and medical education. Acad Med, 1995. **70**(9): p. 795-805.

2. Pasteur, L., The History of a Mind. Philadelphia, Pennsylvania: W.B. Saunders Company.

3. Ludmerer, K., Learning to heal: The development of American medical education. 1985: New York: Basic Books;.

4. Salehi, A.M., et al., Move to the Fourth-Generation Universities: A Systematic Scoping Review of Educational and Management Strategies. Strides in Development of Medical Education, 2021. **18**(1): p. 1-9.

5. Salas-Vega, S., A. Haimann, and E. Mossialos, Big Data and Health Care: Challenges and Opportunities for Coordinated Policy Development in the EU. Health Syst Reform, 2015. **1**(4): p. 285-300.

6. Phillips, K.A., et al., Genomic sequencing: assessing the health care system, policy, and big-data implications. Health Aff (Millwood), 2014. **33**(7): p. 1246-53.

.Fallik, D., For big data, big questions remain. Health Aff (Millwood), 2014. **33**(7): p. 1111–4 .7

.Fihn, S.D., et al., Insights from advanced analytics at the Veterans Health Administration .8
.Health Aff (Millwood), 2014. **33**(7): p. 1203–11

.Spencer, S.A. and M.P. Davies, Hospital episode statistics: improving the quality and value of .9
.(hospital data: a national internet e-survey of hospital consultants. BMJ Open, 2012. 2(6

.Kum, H.C. and S. Ahalt, Privacy-by-Design: Understanding Data Access Models for .10
.Secondary Data. AMIA Jt Summits Transl Sci Proc, 2013. **2013**: p. 126–30

.Kayaalp, M., Patient Privacy in the Era of Big Data. Balkan Med J, 2018. **35**(1): p. 8–17 .11

.Goldberger, A.L., et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new .12
.research resource for complex physiologic signals. Circulation, 2000. **101**(23): p. E215–20

.Available from: https://exhibits.mclibrary.duke.edu/databank/overview.html ; .13

.Ahn, I., et al., CardioNet: a manually curated database for artificial intelligence-based .14
.research on cardiovascular diseases. BMC Med Inform Decis Mak, 2021. **21**(1): p. 29

اهداف: هدف اصلی، اهداف اختصاصی، هدف کاربردی

اهداف اصلي بنيادی-کاربردی طرح:

- طراحی مدل یکپارچه سازی داده های حاصل از دیتابیس های مختلف شامل ثبت اطلاعات مربوط به بیماری های قلب و عروق در سطح مراکز قلب و عروق کشور شامل اقدامات درمانی، تجویز های دارویی، اطلاعات پاراکلینیک شامل آزمایشات بیوشیمی، خونی، کشت و هیستوپاتولوژی، ژنتیک و تصاویر CT scan، اکوکاردیوگرافی و MRI، و نگهداری داده های نمونه-محور آن
- راه اندازی درگاه عملیاتی ورود و یا اتصال کلیه داده های چندوجهی سلامت و پزشکی بیماران و بیماریهای قلب و عروق
- راه اندازی بایودیتابانک بیماری های قلب و عروق مبتنی بر کلیه داده های مرتبط با بیماری های قلب و عروق
- طراحی و عملیاتی سازی درگاهی برای ارایه خدمات اطلاعاتی و دانشی به محققین حوزه قلب و عروق کشور

اهداف تخصصی و کاربردی طرح:

- راه اندازی پلتفرم ثبت ملی داده های مورتالیتی و موربیدیتی در مبتلایان به بیماری های قلبی عروقی و ایجاد زیرساخت مطالعات برای شناسایی عوامل خطر موثر در ایجاد و پیشروی بیماری های قلب و عروق کشور
- راه اندازی پلتفرم های دیجیتال مطالعه بار بیماری ها و اثرات آن بر نظام سلامت بر اساس یافته های حاصل از رجیستری های انبوه داده ها در بیوبانک جهت تصمیم گیری در مقیاس کلان و سیاست گذاری برای تدوین یا اصلاح گایدلاین ها
- طراحی و اجرای کلینیکال ترایال های انسانی با نسل های جدید دارویی به صورت چند مرکزی با قابلیت فالو آپ طولی بیماران به صورت مطالعات کوهورت آینده نگر
- طراحی و مطالعه انجام مداخلات مقرون به صرفه در زمان درست با جمعیت مناسب و پایش کوتاه تا دراز مدت تاثیر مداخلات از طریق پیگیری و فالوآپ طولی بیماران جهت بررسی پیامد های مداخلات تشخیصی درمانی در بلند مدت و بهینه سازی آن ها
- تشکیل تیم تخصصی تحلیل داده ها، اطلاعات و دانش بیماری های قلب و عروق به منظور انجام تحلیل های پیشرفته چند داده ای با مطابقت داده های چند وجهی اعم از آزمایش، تصویر، نمونه های بافتی، و مطالعات ژنتیکی و ....و ایجاد دیتاست های سودمند ترکیبی برای محققین ملی و بین المللی و مطابقت یافته های مطالعات سلولی مولکولی با یافته های بالینی بیمار و توسعه دیتاست های پروتئومیکس و ژنومیکس
- طراحی اپلیکیشن های هوشمند سلامت بر اساس تحلیل های آماری انجام شده حاصل از خروجی بایوبانک
- راه اندازی داشبورد رصد اپیدمیولوژیک وقوع یک همه گیری یا افزایش شیوع بیماری های خاص در یک بازه

	زمانی جهت شناسایی یک عامل خطر ایجاد شده و راه های پیشگیری جغرافیایی و غیر جغرافیایی • نشان دادن تشخیص های محتمل و اقدامات درمانی پیشنهادی برای بیمار با مجموعه علایم و نشانه های مشترک به صورت گراف جهت اهداف آموزشی برای دانشجویان پزشکی به صورت چند مرکزی • راه اندازی پرونده الکترونیک با قابلیت فالو طولی بیماران و نشان دادن سیر درمان و نتیجه همه اقدامات پاراکلینیک و تسهیل دسترسی به سوابق تشخیصی و درمانی بیمار برای پزشک معالج
فرضیات یا سوالات پژوهشی	از آنجا که طرح از نوع بنیادی-کاربردی است فرضیه یا سوالات پژوهشی در طرح وجود ندارد.
روش اجرا	این مطالعه، از نوع مطالعات مدیریت سیستم بهداشتی، و با هدف بررسی چالش های جاری دسترسی به داده های پزشکی و سلامت حوزه قلب و عروق از طرف محققین و ارایه راه حلی بنیادی-کاربردی برای حل این مشکلات در قالب راه اندازی یک بایودیتابانک چند مرکزه داده های سلامت و پزشکی بیماری های قلب و عروق و ایجاد درگاه خدمات محققین می باشد. این مطالعه در سه فاز انجام می گردد: **فاز اول – مطالعه، بسترسازی و طراحی پلتفرم اولیه درگاه محققین:** این فاز شامل مراحل فنی-اجرایی-مطالعاتی زیر خواهد بود: در این فاز، هدف اصلی ما آماده‌سازی زیرساخت‌های لازم برای راه‌اندازی پلتفرم اولیه بایودیتابانک قلب و عروق است. در اینجا، جزئیات فنی، اجرایی و عملیاتی کار بررسی می‌شود: ۱. تعیین نیازمندی‌های داده‌های قلب و عروق: برای شروع، نیازمندی‌های دقیق برای داده‌های قلب و عروق را بررسی می‌کنیم. این شامل نوع داده‌ها، ویژگی‌های مورد نیاز، استانداردهای مربوطه و ساختار داده‌ها می‌شود. خروجی نهایی این مرحله، تدوین اولین کاتالوگ داده سلامت و بیماری قلب و عروق است. ۲. پاکسازی، بهینه‌سازی و استانداردسازی داده‌ها: در این مرحله، استانداردهای جهانی و ملی مرتبط با داده‌های قلب و عروق را مورد بررسی قرار می‌دهیم. سپس، فرآیندهای پاکسازی، بهینه‌سازی و استانداردسازی داده‌ها را برای حذف نویز، افزایش دقت و قابلیت استفاده بهینه مورد اعمال قرار می‌دهیم. خروجی این مرحله، استانداردهای کمی و کیفی داده های قلب و عروق بانک به عنوان مرکز رفرانس می باشد. ۳. طراحی و پیاده‌سازی زیرساخت‌های مرکز داده بایودیتابانک: در این مرحله، پلتفرم نرم افزاری و سخت افزاری اولیه مرکز داده برای بایوبانک قلب و عروق طراحی و پیاده‌سازی می‌شود. این زیرساخت شامل موارد زیر است: • زیرساخت پشتیبان‌گیری: سیستمی برای پشتیبان‌گیری از داده‌ها و حفظ یک نسخه پشتیبان ایمن از آنها ایجاد می‌شود. • زیرساخت مانیتورینگ: سیستمی برای مانیتورینگ و نظارت بر عملکرد سرورها، شبکه و زیرساخت‌های فیزیکی داده‌ها ایجاد می‌شود. • زیرساخت پردازشی مبتنی بر موتورهای پردازشی گرافیکی (GPU): این زیرساخت برای فعالیت های مرتبط با هوش مصنوعی، یادگیری پیشرفته و یادگیری عمیق ضروری است و قسمت عمده هزینه پروژه مربوط به فراهم سازی این بخش است. • زیرساخت مخزن کد (Source Control): یک سیستم مدیریت نسخه برای کنترل و ردیابی تغییرات در کدهای بایوبانک تعبیه می‌شود ۴. طراحی آزمایشی پلتفرم اولیه درگاه محققین قلب و عروق: یک پلتفرم درگاه محققین قلب و عروق طراحی و پیاده‌سازی می‌شود. این پلتفرم ارتباط محققان با بایوبانک را فراهم می‌کند و قابلیت جمع‌آوری، مدیریت و به اشتراک گذاری داده‌های قلب و عروق را فراهم می‌کند. در این فاز، ما از فناوری‌های مختلف برای طراحی و پیاده‌سازی زیرساخت‌ها استفاده خواهیم کرد:

ـ ماشین‌های مجازی (Virtual Machines): استفاده از ماشین‌های مجازی به ما امکان می‌دهد تا سرعت اجرا و مقیاس‌پذیری سرورها را افزایش دهیم و محیطی جداگانه برای هر محقق فراهم کنیم.

ـ فناوری‌های محاسبات ابری (Cloud Computing): استفاده از ابرها به ما کمک می‌کند تا منابع محاسباتی قابل مقیاس‌پذیری را برای ذخیره و پردازش داده‌ها فراهم کنیم.

ـ ابزارهای مانیتورینگ سلامت (Health Monitoring Tools): استفاده از ابزارهای مانیتورینگ سلامت مانند سیستم‌های پیش‌بینی عیب‌ها (Predictive Failure Systems) و سیستم‌های زمانبندی تعمیر و نگهداری (Maintenance Scheduling Systems)، امکان پیش‌بینی و رفع مشکلات سرورها و زیرساخت‌های داده را فراهم می‌کند.

ـ فناوری های تحلیل داده (Data Analytics): استفاده از فناوری‌های تحلیل داده مانند الگوریتم‌های یادگیری ماشین و کاوش داده (Data Mining) به ما کمک می‌کند تا از داده‌های قلب و عروق بهره‌وری بیشتری کسب کنیم و الگوها و اطلاعات مفید را استخراج کنیم.

ـ سیستم‌های پایگاه داده (Database Systems): استفاده از سیستم‌های پایگاه داده قوی و قابل اطمینان برای ذخیره‌سازی و مدیریت داده‌های قلب و عروق ضروری است. می‌توان از پایگاه داده‌های رابطه‌ای مانند MySQL یا پایگاه داده‌های بزرگتر و مقیاس‌پذیر مانند Apache Cassandra استفاده کرد.

ـ فناوری‌های امنیتی (Security Technologies): با توجه به حساسیت داده‌های پزشکی، امنیت اطلاعات بسیار حائز اهمیت است. استفاده از روش‌های رمزنگاری قوی، سیستم‌های تشخیص نفوذ و سیستم‌های کنترل دسترسی مورد نیاز است.

در این فاز، نیازمند توجه به جزئیات فنی و عملیاتی هستیم تا زیرساخت‌های لازم برای راه‌اندازی بایوبانک قلب و عروق را به درستی طراحی و پیاده‌سازی کنیم. تأکید بر امنیت، دقت و سرعت عملکرد این زیرساخت‌ها بسیار مهم است.

فاز دوم – استقرار آزمایشی بایودیتابانک قلب و عروق:

استقرار آزمایشی بایودیتابانک قلب و عروق شامل شکل گیری و استقرار فنی پلتفرمی مبتنی بر یک معماری هفت لایه در مرکز داده بایودیتابانک به شرح ذیل است:

1- لایه رابط داده یا Data Connector: این لایه برای اتصال به داده های اصلی و دریافت آن ایجاد شده است. سیستم های اطلاعات سلامت مانند HIS و یا LIS و یا PACS سیستم هایی هستند که داده های تولید شده در خود را مدیریت می کنند و می بایست به زیر ساخت CBDB منتقل شده و به روز نگه داشته شوند. این لایه می تواند با یک زیر ساخت مشابه نیز متصل شود و کاربرد آن ایجاد شبکه های تحقیقاتی در کشور می باشد. به عنوان مثال می تواند به یک بایوبانک یا به یک بیمارستان تخصصی قلب دیگر متصل گردد. یکی از استفاده های این زیرساخت ایجاد رجیستری بیماری و یا کمک به مطالعات ک.وهورت می باشد که با استفاده از لایه ارتباطی این مورد محقق می شود.

2- لایه دریاچه داده یا Data Lake و مخزن فراداده یا Meta Data Repository: دریاچه داده، مخزنی از اطلاعات می باشد که داده های برای مرحله پردازش در آن ذخیره می گردند. این مخزن شامل تمامی داده هایی است که ممکن است در کاربردهای زیر ساخت استفاده شود. در دریاچه داده، داده ها به همان شکل اولیه ای که در سیستم های اصلی تولید شدن قرار داده می شوند. به عنوان مثال داده های HIS به همان شکل اولیه قرار داده می شوند. این مخزن باید قابلیت این را داشته باشد که داده ها را با کمترین زمان ذخیره نماید و بتواند حجم زیادی از داده را در خود ذخیره سازی نماید.

در دریاچه داده، داده ها به مدل های مختلفی دسته بندی و در دسترس قرار می گیرند. هر یک از این نوع داده ها ماهیت ذخیره سازی خاص خود را دارند:

- داده های متنی: این داده ها شامل داده های ساختار یافته و غیر ساختار یافته می شوند و می تواند از عدد، متن و سایر الگو های داده اولیه باشند.
- داده های تصویری: شامل داده های تصویر برداری های پزشکی که معمولا از استاندارد DICOM استفاده می

کنند و تصاویر اسکن شده پرونده ها و سایر تصاویری که در مراقبت ها استفاده می گردد
- داده های سیگنال های پزشکی: این داده ها ماهیت سیگنال دارند مانند نوار قلب (ECG)
- داده های ژنتیک

مخزن فرا داده برای ذخیره سازی دیتا دیکشنری که در دریاچه داده ذخیره شده است و دیتا دیکشنری مطلوب برای CDBD که در مخزن اصلی داده ذخیره شده است، مورد استفاده قرار می گیرد. فرا داده های مربوط به شاخص ها (KPI) نیز در این مخزن داده قرار می گیرد. مخزن اصلی داده شامل یک مدل استاندارد داده برای یکپارچه سازی تمام داده هایی که در دریاچه داده ذخیره شده است می باشد. برای اینکه بتوان از دریاچه داده مخزن اصلی داده را ایجاد کرد نیاز به فرایند های مختلف بر روی داده می باشد. به عنوان مثال فرایند ETL شامل استخراج داده از دریاچه داده، تغییر شکل داده به مدل استاندارد و ذخیره سازی در Master Data می باشد که با استفاده از لایه Data Preparation Infrastructure صورت می گیرد.

3- لایه زیرساخت آماده سازی/پاکسازی داده یا Data Preparation Infrastructure: زیر ساخت آماده سازی داده، زیر ساخت است که برای کار با داده ها و مهندسی داده طراحی شده است. مهندسی داده مجموعه ای از عملیات با هدف ایجاد رابط ها و مکانیسم هایی برای جریان و دسترسی به اطلاعات است. به متخصصان خاص ـ مهندسین داده ـ نیاز است تا داده ها را به گونه ای نگهداری و مدیریت کنند که در دسترس و قابل استفاده دیگران باقی بماند. به طور خلاصه، مهندسان داده زیرساخت داده سازمان را راه اندازی کرده و آن را برای تجزیه و تحلیل بیشتر توسط تحلیلگران و محققان داده آماده می کنند. اقدامات مورد انجام توسط مهندسین داده شامل ایجاد جریان داده، حذف داده های تکراری، ترمیم داده ها (پر کردن مقادیر ناقص در صورت امکان)، تبدیل کردن داده به یک مدل داده‌ای مشخص، پاکسازی داده ها و حذف داده خراب و غیرقابل استفاده، محاسبه و تجمیع داده ها، و در نهایت بی نام سازی داده هاست.

از آنجایی که داده های سلامت داده های محرمانه ای می باشند زیر ساخت آماده سازی داده ها به گونه ای طراحی شده است که امنیت را برای داده ها در شرایطی که مهندسین داده بر روی آن می خواهند کار کنند، فراهم می سازد. ابزارهای مختلفی در این زیر ساخت برای کار با داده ها قرار داده شده است. در این فضا ممکن است برای یادگیری ماشین و پیش تحلیل ها یا اجرای الگوریتم های محاسباتی بر روی داده ها نیاز به توان پردازشی بالا باشد. همچنین این زیر ساخت می تواند به صورت سرویس ارائه شود. برای مواردی که محقق می خواهد بر روی داده های خام کار کنند و یا داده هایی را مورد جستجو (navigation) قرار دهد. زیر ساخت هایی با توان پردازشی بالا نیز می تواند در اختیار محققانی که نیاز به عملیات یادگیری ماشین بر روی داده ها یا مجموعه داده ها (Dataset) انجام دهند. سرویس های قابل ارائه از این زیر ساخت شامل :

- DSVM : یا ماشین های مجازی علوم داده است (Data Science Virtual Machine). این ماشین ها به مهندس داده یا محقق این امکان را می دهد که در یک محیط امن به داده های مد نظر خود دسترسی پیدا نماید و با ابزارهایی که در این محیط دارد بتواند با داده ها کار کند.
- HPCVM : ماشین مجازی با قدرت پردازش بالا (High Performance Computing Virtual Machine) ماشین مجازی می باشد که مانند DSVM عمل می کند با این تفاوت که امکان استفاده از قدرت پردازش بالا را دارند به عبارت دیگر به غیر از استفاده از CPU در پردازش امکان استفاده از GPU در پردازش ها را نیز خواهند داشت.
- SafeZone : قسمت فیزیکی محافظت شده ای است که در آن محقق یا مهندس داده می تواند در شرایط کنترل شده به سرویس های DSVM و HPCVM متصل گردد.

4- لایه ابرمخزن داده و انباره داده ها Data Warehouse / Data Mart: زیر ساخت انباره داده حاوی چینش جدید داده ها بر اساس موضوعات و نیازمندیهای مطرح شده است. برای هر مارت یکسری فرا داده تولید می گردد که خود در مخزن فرا داده ذخیره می گردد و مستند می گردد. چینش داده ها در دیتا مارت ها به گونه ای است که تحلیل داده را سریعتر و راحتر می نماید در نتیجه استخراج داده بهتر صورت می گیرد. مسیر های داده ای در زیر ساخت آماده سازی داده ها برای بروز نگهداری انباره داده و دیتا مارت ها طراحی می گردد و بر اساس نیاز در زمانبندیهای مشخص اجرا میگردند. خروجی این لایه هم برای تولید مجموعه داده ها استفاده می شود و هم برای مصورسازی در لایه داشبورد مورد استفاده قرار خواهد گرفت.

5- لایه مخزن دیتاست یا Dataset Repository و شاخص های کلیدی KPI Repository: مخزن دیتاست به

عنوان مخزنی برای ذخیره سازی مجموعه داده های آماده شده و اهداء شده در نظر گرفته می شود در حالیکه مخزن شاخص های کلیدی خوراک سیستم های مدیریت داده را از لحاظ تامین داده های شاخص های کلیدی بیماری و سلامت و مدیریت نظام های سلامت فراهم می سازد. مخزن شاخص های کلیدی شامل فراداده های تعریف شاخص ها و اقلام اطلاعاتی می باشد. امکان تعریف شاخص های جدید و مستند سازی اقلام داده جدید در این مخزن بر اساس سامانه ای که آن را مدیریت می نماید فراهم می شود.

6- لایه داشبوردهای داده ای و سامانه های مدیریت داده بایوبانک: داشبورد اطلاعاتی شامل مجموعه نرم افزارهای مصور سازی داده می باشد داشبورد اطلاعاتی اطلاعات خود را از انباره داده یا مخزن مجموعه داده دریافت می نماید. از خروجی های داشبورد اطلاعاتی می توان برای تولید مجموعه داده های محدود استفاده شود. سیستم مدیریت داده ها، سامانه ای می باشد که مجموعه داده ها را برای استفاده محققات و ذینفعان ارائه می دهد. این سامانه فراداده مجموعه داده ها را مدیریت می کند و قابلیت جستجو بر روی فرا داده ها دارد.

7- لایه درگاه محققین یا Researchers' Gateway: درگاه واحد ارائه خدمات به محققان در این حوزه است علاوه بر سرویس های مدیریت داده، دریافت داده و امکان تعریف ماشین مجازی برای علوم داده یا ماشین مجازی با قدرت پردازشی بالا را فراهم می نماید. امکان ارتباط محققان در این فضا به صورت یک رویداد اجتماعی خاص فراهم می گردد. این لایه پس از طراحی لایه های شش گانه پیش نیاز، در فاز سوم، راه اندازی و سرویس های آن به تدریج پیاده سازی خواهد شد.

فاز سوم: ـ راه اندازی سرویس های درگاه خدمات محققین و مراکز قلب و عروق کشور:

درگاه خدمات محققین، درگاهی آنلاین برای دریافت خدمات بایوبانک ، و رجیسترکردن و اتصال بانک های سایر مراکز می باشد. بایودیتابانک قلب و عروق ایران، به تدریج و با توسعه تیم تخصصی مهندسی علوم داده قلب و عروق خود (Cardioclinical Data Scientists) در مرکز فناوری و نوآوری بایوبانک، خدماتی را نیز به صورت حضوری و در قالب تشکیل تیم های مشاوره تحلیل و پشتیبان پژوهش های ملی ارایه خواهد داد.

از جمله خدمات درگاه محققین بایوبانک، عبارت است از:

1. **سفارش ایجاد یک مجموعه داده خاص (Cardio-dataset by Special Order):** یکی از کاربردهای این زیر ساخت ایجاد مجموعه دادهای خاص می باشد که برای فعالیت های تحقیقاتی و پژوهشی با موضوعات مشخص استفاده می گردد. روش های گوناگونی برای ایجاد یک مجموعه داده وجود دارد. یکی از این روش ها در شکل زیر آورده شده است. استفاده از کاردیوبایودیتابانک برای ایجاد مجموعه داده ها این امکان را به ما می دهد که مجموعه داده های مرتبط به هم را بتوانیم ایجاد کنیم و به راحتی قابل نگهداری و بروز رسانی می باشد. برای مثال، در یک سفارش می تواند، اطلاعات مربوط به HIS یک مرکز وارد دریاچه داده ای شود و سپس داده های آن به فرمت و مدل استانداردی برای پرونده بیماران قلبی و عروقی همان مرکز شده و این مجموعه داده در مخزن داده اصلی (Master Data) ذخیره گردد. حال محققین مرکز سفارش دهنده می توانند از روی مخزن اطلاعات یک Data Mart از بیماران دارای CHF خود طراحی نموده و آن را در بایوبانک به نام خود نگاه دارند که این Data Mart به صورت اتوماتیک به روز می گردد. سپس در یک بازه یکساله از داده های مارت مربوط به بیماران CHF مجموعه داده ای شامل اطلاعات دموگرافیک بیماران، میزان EF بیمار و Troponin بیمار تهیه می گردد و در مخزن دیتاست ها قرار داده می شود و این مجموعه داده ها از طریق درگاه برای محققان و پژوهش گران همکار قابل استفاده است.
2. **خدمت سفارش طراحی داشبورد اطلاعاتی:** داشبوردها هم برای استفاده های مدیریتی و هم استفاده های علمی-پژوهشی کاربرد دارند. گردش (Navigation) در داده ها از طریق داشبوردها اتفاق می افتد. داشبوردها معمولا قسمتی از مفهوم هوش سازمانی یا BI[1] می باشند. داشبورد اطلاعاتی می تواند برای استخراج مجموعه داده ها نیز استفاده شود و در فرایند تولید مجموعه داده ها استفاده گردد.
3. **سفارش ایجاد رجیستری بیماری های قلب و عروق:** ایجاد سیستم های ثبت بیماریها همواره در مسیر تحقیقات پزشکی اهمیت داشته است. معمولا ایجاد یک سیستم ثبت در کنار سیستم های روتین ورود اطلاعات بیمار مانند سیستم اطلاعات بیمارستانی ممکن است دوباره کاری در ورود داده ها را باعث گردد. از زیر ساخت موجود به گونه ای می شود استفاده نمود که داده های روتین بر اساس فیلتر ها و شرط های خاص به یک بانک اطلاعاتی رجیستری به صورت اتوماتیک هدایت شوند و با این روش مورد یابی (case finding) مناسبی

	اتفاق افتد. دز این مرحله یک زیرساخت نرم افزاری برای رجیستری می توان بر روی این بانک اطلاعاتی قرار گرفته و علاوه بر تایید مواردی که توسط فرایند اتوماتیک وارد سیستم شده است می تواند موارد کاملا جدید را از طریق سیستم نرم افزاری وارد نماید. در این موارد پرونده الکترونیکی سلامت و سیستم های اطلاعاتی معمول در ایجاد رجیستری ها استفاده موثر دارند(15). Business intelligence [1]
ملاحظات اخلاقی	سه ملاحظه اخلاقی زیر در طراحی و راه اندازی بانک زیست_داده و درگاه محققین مورد عنایت قرار خواهد گرفت: • حفظ حریم خصوصی افرادی که اطلاعات آن‌ها در بانک قرار می‌گیرد (از طریق بی هویت سازی قانونمند داده ها و غیرقابل ردگیری کردن داده ها به صاحبان آنها (در کنار الزام رعایت رعایت اصول اخلاقی در جمع‌آوری و استفاده از اطلاعات بیماران) • اطمینان از اینکه کاربران غیرمجاز به اطلاعات بیماران دسترسی پیدا نمی‌کنند. • اطمینان از اینکه اطلاعات بیماران به کاربران مجاز فقط برای مقاصد پژوهشی تعریف شده و توافق شده ارائه می‌شود.

جدول متغیرها

نام متغیر	نقش متغیر	نوع متغیر کمی ـ پیوسته است؟	نوع متغیر کمی ـ گسسته است؟	نوع متغیر کیفی ـ رتبه ای است؟	نوع متغیر کیفی ـ اسمی است؟	واحد اندازه گیری	تعریف کاربردی	نحوه اندازه گیری
متغیری وجود ندارد. کار ایجاد یک بایوبانک کاربردی است		☐	☐	☐	☐	--	--	--

زمانبندی و اجرا

شرح مختصر مرحله	درصد مرحله	مدت زمان اجرا ـ ماه	از تاریخ	تا تاریخ
زمانبندی انجام کل پروژه	100	28	1402/01/01	1404/04/31
فاز اول – مطالعه، بسترسازی و طراحی پلتفرم	30	8	1402/01/01	1402/08/30
فاز دوم – استقرار آزمایشی بایودیتابانک قلب و عروق	30	8	1402/09/01	1403/04/31
فاز سوم: ـ راه اندازی سرویس های درگاه خدمات محققین و مراکز قلب و عروق کشور	40	12	1403/04/31	1404/04/31

هزینه وسایل و مواد مورد نیاز

نوع	نان دستگاه / مواد اولیه	تعداد مورد نیاز	قیمت دستگاه/ وسیله/ مواد ـ ریال	کشور سازنده	شرکت سازنده	شرکت فروشنده	محل تامین اعتبار	جمع کل هزینه به ریال
غیر مصرفی – سرمایه ای	سرور GPU	2	30,000,000,000					60,000,000,000

هزینه پرسنلی

نام و نام خانوادگی	توصیف دقیق فعالیتی که فرد باید در این تحقیق انجام دهد	کل حق الزحمه ـ ریال
محمود تارا	مجموعه فعالیت های نیروهای انسانی پروژه شامل: مهندسی داده، تحلیل هوشمند داده ها، معماری داده ها و پایپ لاین های بایوبانک، و مهندسی ومعماری دیتاست و رجیستری و کوهورت های مورد تقاضا، و داشبورد های مدیریتی خواهد بود.	26,500,000,000

هزینه آزمایشات و خدمات تخصصی

نام خدمت	نام مؤسسه ارائه کننده خدمت	تعداد یا مقدار لازم	قیمت واحد ـ ریال	قیمت کل ـ ریال
شامل سه دسته خدمت: خدمات مهندسی بخش اکتیو و پسیو زیرساخت بانک زیست دادها طراحی و پیاده سازی و نگهداری پلتفرم بانک زیست_داده خدمات هوشمندسازی بانک زیست_داده مبتنی بر هوش عمومی و تخصصی	پیمانکاران مختلف خدمات تخصصی	1	21,500,000,000	21,500,000,000

جمع کل هزینه های طرح

هزینه پرسنلی (هیات علمی و غیر هیات علمی)	هزینه مواد مصرفی	هزینه مواد غیر مصرفی	هزینه تجهیزات،موادوخدمات موجوددر مرکز	هزینه مسافرت	هزینه چاپ و تکثیر	سایر هزینه ها	جمع کل هزینه ـ ریال
26,500,000,000	0	60,000,000,000	21,500,000,000	0	0	0	108,000,000,000