





دانشگاه علوم پزشکی اسفراین

عنوان: عوامل موثر بر بقای پنج ساله‌ی کودکان مبتلا به سرطان خون با استفاده از مدل‌های یادگیری ماشین

پایان نامه در جهت دریافت درجه کارشناسی

در

فناوری اطلاعات سلامت

اساتید راهنمای

سرکار خانم دکتر مرضیه محمودی منش

جناب آقای دکتر احمد رضا شمس آبادی

استاد مشاور:

سرکار خانم دکتر مرضیه محمودی منش

نگارنده:

مهرداد حسن زاده

اظهار نامه دانشجو

موضوع پایان نامه: عوامل موثر بر بقای پنج ساله‌ی کودکان مبتلا به سرطان خون با استفاده از
مدل‌های یادگیری ماشین

این‌جانب مهرداد حسن‌زاده دانشجوی دوره کارشناسی دانشکده علوم پزشکی اسپرائین گواهی می‌نماییم که تحقیقات ارائه شده در این پایان نامه توسط این‌جانب انجام شده و صحت و اصالت مطالب نگارش شده مورد تایید می‌باشد و در موارد استفاده از کار دیگر محققان به مرجع مورد استفاده اشاره شده است. بعلاوه گواهی می‌نماییم که مطالب مندرج در پایان نامه تاکنون برای دریافت مدرک این‌جانب یا فرد دیگری ارائه نشده است و هیچ بخش آن از کار سایر دانشجویان و محققین کپی نشده است. در تدوین متن پایان نامه دستورالعمل مصوب دانشگاه بطور کامل رعایت کرده‌ام.

امضای دانشجو

تاریخ:

حق چاپ، نشر و مالکیت معنوی پایان نامه

- ۱- هر گونه کپی برداری به صورت کامل از پایان نامه یا بخشی از آن تنها با موافقت استاد راهنمای مجاز میباشد.
- ۲- کلیه حقوق معنوی این اثر متعلق به دانشاه علوم پزشکی تهران می باشد و بدون اجازه کتبی دانشگاه به شخص ثالث قابل واگذاری نیست.
- ۳- استفاده از اطلاعات و نتایج موجود در پایان نامه بدون ذکر مرجع مجاز نیست.

پاسکنزاری:

این اثر به عشق و محبت خانواده ام، به ویره پر و مادرم، تقدیم می شود که بهواره با حیات و دلگرمی هشان به راه من بوده اند. از دوستان و استادی که انقدری که با مشاوره ها و راهنمایی های ارزنده خود مسیر را برایم بهوار کردند، للخصوص جناب دکتر احمد رضا شمس آبادی نزیر سیار پاسکنزارم.

با آرزوی موفقیت و سلامتی برای بمه غریزانم.

با احترام،

مهرداد حسن زاده

تعدیم به

خانواده عزیزم که همواره پسیان من در راه کسرش علم بودند...

عوامل مؤثر بر بقای پنج ساله کودکان مبتلا به سرطان خون با استفاده از مدل های یادگیری ماشین

مقدمه: سرطان خون یکی از بیماری های جدی و شایع در کودکان است که به بافت های خون ساز بدن آسیب می زند. لوسومی، نوعی سرطان خون، شامل تکثیر غیر طبیعی سلول های بنیادی خون ساز در مغز استخوان است. لوسومی لنفو بلاستیک حاد (ALL) شایع ترین نوع لوسومی در کودکان است. با توجه به نقش مهم فناوری در بهبود تشخیص و درمان این بیماری، این پژوهش به بررسی عوامل مؤثر بر بقای پنج ساله کودکان مبتلا به سرطان خون با استفاده از مدل های یادگیری ماشین پرداخته است.

مواد و روش ها: در این پژوهش، ابتدا داده های بالینی مربوط به کودکان مبتلا به سرطان خون جمع آوری و پیش پردازش شد. سپس با استفاده از چهار الگوریتم یادگیری ماشین شامل درخت تصمیمی، جنگل تصادفی، CATBOOST و آزمون کاس، به تحلیل و پیش بینی عوامل مؤثر بر بقای پنج ساله این کودکان پرداخته شد. دقت و کارایی مدل ها با استفاده از معیار های مختلف ارزیابی و مقایسه گردید.

یافته ها: نتایج نشان داد که مدل های یادگیری ماشین در تشخیص و پیش بینی بقای پنج ساله کودکان مبتلا به سرطان خون دقیق تر می باشد. ضریب کاپا برای مدل های مختلف بین 0.77 تا 0.88 به دست آمد که نشان دهنده نزدیکی نتایج مدل ها به واقعیت بالینی است. همچنین مدل CATBOOST با دقت 85 درصد بهترین عملکرد را در بین الگوریتم ها داشت.

بحث و نتیجه گیری: مقایسه نتایج مدل های یادگیری ماشین با داده های بالینی نشان داد که این روش ها می توانند ابزار مؤثری در تشخیص و پیش بینی بقای پنج ساله کودکان مبتلا به سرطان خون باشند. استفاده از این مدل ها می تواند به بهبود فرآیند تشخیص و درمان این بیماری کمک کرده و کیفیت زندگی بیماران را افزایش دهد. این روش ها قابلیت گسترش و کاربرد در سایر انواع سرطان های کودکان را نیز دارا می باشند.

كلمات كليدي: سلطان خون، لوسمي، يادگيري ماشين، پيشبيني بقا، کودکان، درخت تصميم، جنگل تصادفي، CATBOOST، آزمون کاس

فهرست مطالب

	عنوان	
	صفحه	
فصل اول: بیان مسئله		
14	مقدمه:.....	1-1
15	بیان مساله.....	1-2
16	اهمیت و ضرورت اجرای پژوهش.....	1-3
17	هدف کلی.....	1-4
17	اهداف ویژه.....	1-5
18	اهداف کاربردی.....	1-6
18	هدف آرمانی.....	1-7
18	سوالات پژوهش.....	1-8
18	تعریف واژگان.....	1-9
21	وجه تمایز پژوهش با سایر پژوهش های انجام شده.....	1-10
23	ملاحظات اخلاقی.....	1-11
فصل دوم: بررسی متون		
26	1-2 مقدمه.....	1-2
26	2-2 تعریفات و کلیات سرطان خون در کودکان.....	2
26	2-3 عوامل خطر و علل سرطان خون در کودکان.....	3
27	2-4 تشخیص و درمان سرطان خون در کودکان.....	4
27	2-5 تاریخچه یادگیری ماشین.....	5

28.....	2-6 بررسی منابع داخلی
30.....	2-7 بررسی منابع خارجی

فصل سوم: مواد و روش کار

35.....	1-3 مقدمه
36.....	2-3 فاز اول: جمع آوری داده
38.....	3-3 فاز دوم: پیش پردازش داده ها
40.....	4-3 فاز سوم: ایجاد مدل
44.....	4-4 فاز چهارم: ارزیابی مدل
45.....	5-۳ نتیجه گیری فصل

فصل اول:

کلیات و بیان مسئله

عوامل موثر بر بقای پنج ساله‌ی کودکان مبتلا به سرطان خون با استفاده از مدل‌های یادگیری ماشین

۱- مقدمه:

سرطان خون مربوط به بافت‌های خون ساز بدن است. لوسومی یک تکثیر کلونال از سلول‌های بنیادی خونساز در مغز استخوان است. چهار زیرگروه اصلی که احتمالاً توسط پزشکان مراقبت اولیه مشاهده می‌شوند عبارتند از لوسومی لنفوبلاستیک حاد، لوسومی میلوژنوس حاد، لوسومی لنفوسيتی مزمن، و لوسومی میلوژنوس مزمن. لوسومی لنفوبلاستیک حاد بیشتر در کودکان رخ می‌دهد، در حالی که زیرگروه‌های دیگر بیشتر در بزرگسالان شایع هستند. لوسومی حاد میلوئید یک نوع شایع و کشنده آن است.^[1] انجمن سرطان ایتالیا تخمین زده است که در یک دوره پنج ساله 2016 تا 2020، در خط پنج ساله تقریباً 7000 نئوپلاسم در میان کودکان و 4000 نئوپلاسم در میان نوجوانان تشخیص داده شده است. لوسومی حاد لنفوبلاستیک در کشورهای پیشرفته نیز مسئول تقریباً 80 درصد سرطان‌های لوسومی در بین کودکان می‌باشد به طوری که میزان بروز این سرطان 34.3 مورد در هر یک میلیون نفر می‌باشد در حالی که لوسومی حاد میلوبلاستیک در کودکان نادر است و در اروپا میزان بروز آن بین 4 تا 9 در هر یک میلیون نفر در سال می‌باشد.^[2] در نتیجه لوسومی‌ها و تومور‌های مغزی و لنفوم‌ها بیش از دو سوم تمام سرطان‌های کودکان را تشکیل می‌دهند. عوارض طولانی مدت، ارگان‌ها و سیستم‌های مختلف را تحت تاثیر قرار میدهد. با توجه به نقش فناوری در کنترل و پیش‌بینی امراض، مقالات پژوهشی توسط ابونادی و سنان (2020)^[3]، فتحی و همکاران (2020)^[4]، و کرمی و همکاران (2021)^[5] به طور جمعی بینش‌های ارزشمندی را در مورد عوامل موثر بر بقای پنج ساله کودکان مبتلا به لوسومی با استفاده از مدل‌های یادگیری ماشین ارائه می‌دهند. ابونادی و سنان (2022) بر تشخیص زودهنگام لوسومی از طریق توسعه سیستم‌های تشخیصی با استفاده از

شبکه‌های عصبی مصنوعی، شبکه‌های عصبی پیش‌خور، ماشین‌های بردار پشتیبان و شبکه‌های عصبی پیچشی تمرکز می‌کنند. سیستم‌های ترکیبی آن‌ها به نرخ‌های دقت بالا دست یافتند که نشان‌دهنده پتانسیل یادگیری ماشین در بهبود تشخیص و پیش‌آگهی است. فتحی و همکاران (2020) استفاده از روش‌های نروفازی و روش گروهی مدیریت داده‌ها را برای تشخیص لوسمی حاد در کودکان بررسی می‌کنند و بر اهمیت طبقه‌بندی دقیق برای مداخله به موقع تأکید دارند. در نهایت، کرمی و همکاران (2021) عوامل پیش‌آگهی برای بقا در بیماران مبتلا به لوسمی حاد میلوئیدی را بررسی می‌کنند و به کارایی تکنیک‌های یادگیری ماشین در پیش‌بینی نرخ‌های بقا اشاره دارند. با ادغام روش‌های انتخاب ویژگی و الگوریتم‌های مختلف، آن‌ها نشان می‌دهند که یادگیری ماشین در بهبود نتایج بیماران پتانسیل دارد. این مطالعات به طور جمعی بر اهمیت استفاده از مدل‌های یادگیری ماشین برای بهبود پیش‌آگهی و نتایج درمانی کودکان مبتلا به لوسمی تأکید می‌کنند.

2- بیان مسئله:

سرطان دومین عامل مرگ در بین گروه کودکان صفر تا 14 ساله است. بعد از جراحت‌های غیر عمدی با میزان مرگ 0.41 ، سرطان عامل بعدی مرگ به ازای هر ده هزار کودک تلقی می‌شود. [6] سلول‌های خون سفید (WBCs) سلول‌های خونی هستند که به عنوان بخشی از سیستم ایمنی، در مقابله با عفونت‌ها و بیماری‌ها موثر هستند. آن‌ها همچنین به عنوان "سلول‌های دفاعی" شناخته می‌شوند. اما عدم تعادل در تعداد سلول‌های خون سفید در خون می‌تواند خطرناک باشد. اگرچه احتمال بروز این سرطان پایین است اما لوسمی، شایع‌ترین نوع سرطان خون است که ناشی از افزایش غیرطبیعی سلول‌های خون سفید در سیستم ایمنی است.[7] تا روزهای اخیر، لوسمی میلوئیدی حاد (AML) همچنان چالش‌هایی را در تشخیص و درمان به نمایه کشیده است. حدود 22 درصد از سرطان‌های مربوط به کودکان به این سرطان تشخیص یافته [8] که در این بین شایع‌ترین شکل لوسمی بین کودکان زیر 19 سال ، مربوط به لوسمی حاد لنفوبلاستیک (Acute lymphoblastic leusemia) می‌باشد. امارها نشان میدهد که

این سرطان در بین کودکان ۱-۴ ساله بیش از ۹ برابر این مقدار در افراد ۲۰ - ۲۴ ساله است.^[9] سرطان های اطفال اللخصوص سرطان های خونی کماکان به عنوان عوامل موثر در مرگ و میر اطفال شناخته میشود^[10].

اولین مطالعات درباره تکنیک های یادگیری ماشین در تشخیص بیماری های خونی دو دهه پیش انجام شدند. آنها با شناسایی سلول های لوکمیک از نمونه های خونی، سیتومتری جریانی، و ارزیابی داده های ژنتیکی آغاز شدند، که پایه های روش های یادگیری ماشین در بررسی بیماری های خونی را فراهم کردند.^[11]

در این پژوهش یادگیری ماشین را با روش های انتخاب ویژگی یکپارچه کرده و عملکرد آن ها را برای شناسایی عوامل موثر در ارزیابی بقا بیماران مقایسه کردیم. تشخیص لوسومی در بسیاری از مناطق جهان زمان بر و چالش برانگیز است و استفاده از تکنیک های ML برای تشخیص آن در حال افزایش است، ^[12] در این پژوهش چهار الگوریتم داده کاوی شامل ، درخت تصمیم ، جنگل تصادفی ، CATBOOST، از مون کاکس را برای تشخیص عوامل موثر بر بقای پنج ساله کودکان مبتلا به سرطان خون بکار گرفتیم.^[13] این پژوهش با هدف مقایسه عملکرد روش های یادگیری ماشین با استفاده از الگوریتم های مختلف در تشخیص عوامل مؤثر بر بقا در سرطان خون در کودکان انجام شده است. لذا به علت اهمیت مسئله شانس بقا کودکان مبتلا به این بیماری ، این پژوهش به منظور شناسایی عوامل افزایش شانس بقای بیماران، پرداخته است.

3-1-3- اهمیت ضرورت پژوهش:

در حال حاضر سرطان خون یکی از مهمترین مسائل بهداشتی جهان به ویژه در کودکان است. با وجود پیشرفت های بسیار در سال های اخیر در زمینه تشخیص و درمان سرطان های مختلف ، این بیماری همچنان به عنوان یکی از عوامل اصلی مرگ و میر در جوامع جهانی تلقی میشود.^[14] بررسی ها نشان میدهد که سرطان خون به خصوص لوسومی حاد لنفوبلاستیک در بین کودکان، در بین عوامل مهم مرگ و میر در این گروه سنی قرار دارد. ^[15] با این وجود،

تشخیص زودرس و دقیق این بیماری اغلب چالش بزرگی است، به ویژه در مناطقی که به تجهیزات پزشکی پیشرفته دسترسی محدودی دارند.[16].

در این زمینه، استفاده از روش‌های یادگیری ماشین، به عنوان یک رویکرد نوین، به شدت به منظور تسهیل و تسريع فرآیند تشخیص سرطان خون مورد توجه قرار گرفته است.[17] با توجه به محدودیت منابع و داده‌های پزشکی محدود، استفاده از روش‌های یادگیری ماشین با استفاده از انتقال یادگیری(Transfer Learning)، به‌طور ویژه در مواردی که داده‌های کمی موجود است، از اهمیت ویژه‌ای برخوردار است.

در این روش، با استفاده از تحلیل داده‌های موجود از بیماران سرطان خون کودکان و اعمال روش‌های یادگیری ماشین، می‌توان به شناسایی عوامل مؤثر بر بقا و بهبودی در این گروه از بیماران پرداخت. بنابراین، پژوهش در این زمینه نه تنها می‌تواند به بهبود فرآیند تشخیص و درمان سرطان خون در کودکان کمک کند، بلکه می‌تواند زندگی و کیفیت زندگی این افراد را بهبود بخشد و نیز مراحل درمانی را بهینه‌تر نماید.

4-1 هدف کلی:

تشخیص عوامل موثر بر بقای پنج ساله‌ی کودکان مبتلا به سرطان خون با استفاده از مدل‌های یادگیری ماشین

5-1 هدف ویژه:

1. تعیین الگوریتم‌های یادگیری ماشین مؤثر در پیش‌بینی بقای پنج ساله‌ی کودکان مبتلا به سرطان خون.

2. مقایسه عملکرد الگوریتم‌های مختلف یادگیری ماشین در تشخیص و پیش‌بینی عوامل مؤثر بر بقای بیماران.

3. شناسایی ویژگی‌های کلیدی و متغیرهای مؤثر بر بقای پنج ساله‌ی کودکان مبتلا به سرطان خون.

4. توسعه یک مدل پیش‌بینی کارا و دقیق برای ارزیابی بقای بیماران مبتلا به سرطان خون با استفاده از روش‌های یادگیری ماشین.

5. بررسی تأثیر استفاده از داده‌های بالینی در بهبود دقت مدل‌های یادگیری ماشین.

6-1 هدف کاربردی:

تعیین الگوریتم موثر و ایجاد مدل مبتنی بر یادگیری ماشین در تشخیص عوامل موثر در بقای پنج ساله کودکان مبتلا به لوسمی

7-1 هدف آرماتی:

ارتقای سلامت بیماران مبتلا به بیماری لوسمی و تشخیص زودرس عواملی که بر بقای آنان موثر است جهت ارتقای کیفیت زندگی

8-1 سوالات پژوهش:

11. کدام یک از الگوریتم‌های یادگیری ماشین در پیش‌بینی بقای پنج ساله‌ی کودکان مبتلا به سرطان خون موثرتر هستند؟

2. کدام ویژگی‌ها و متغیرهای بالینی و ژنتیکی بر بقای پنج ساله‌ی کودکان مبتلا به سرطان خون بیشترین تأثیر را دارند؟

3. چه میزان دقت و کارایی می‌توان از مدل‌های یادگیری ماشین برای پیش‌بینی بقای پنج ساله کودکان مبتلا به سرطان خون انتظار داشت؟

4. چه تفاوت‌هایی در عملکرد الگوریتم‌های مختلف یادگیری ماشین در شناسایی عوامل مؤثر بر بقای بیماران مبتلا به سرطان خون وجود دارد؟

9-1 تعریف واژگان:

(Leukemia) 1. لوسمی:

- یک نوع سرطان خون که از تکثیر غیرطبیعی و غیرقابل کنترل سلول‌های سفید خون در مغز استخوان نشأت می‌گیرد. این بیماری به دو نوع حاد و مزمن و به دو دسته میلوئیدی و لنفوئیدی تقسیم می‌شود.

2. لوسمی لنفوبلاستیک حاد: (Acute Lymphoblastic Leukemia - ALL)

- شایع‌ترین نوع لوسمی در کودکان که در آن لنفوبلاستها (سلول‌های لنفوسيت نابالغ) به طور غیرطبیعی تکثیر می‌شوند.

3. لوسمی میلوژنوس حاد: (Acute Myelogenous Leukemia - AML)

- نوعی لوسمی حاد که در آن میلوبلاستها (سلول‌های میلوئیدی نابالغ) به طور غیرطبیعی در مغز استخوان تکثیر می‌شوند. این نوع لوسمی بیشتر در بزرگسالان شایع است.

4. لوسمی لنفوسيتی مزمن: (Chronic Lymphocytic Leukemia - CLL)

- نوعی لوسمی مزمن که در آن لنفوسيت‌های بالغ به طور غیرطبیعی تکثیر می‌شوند. این نوع لوسمی عمدتاً در افراد مسن دیده می‌شود.

5. لوسمی میلوژنوس مزمن: (Chronic Myelogenous Leukemia - CML)

- نوعی لوسمی مزمن که با تکثیر غیرطبیعی سلول‌های میلوئیدی بالغ همراه است و اغلب با وجود کروموزوم فیلادلفیا مشخص می‌شود.

6. یادگیری ماشین: (Machine Learning)

- شاخه‌ای از هوش مصنوعی که به سیستم‌ها امکان می‌دهد به طور خودکار از داده‌ها یاد بگیرند و بر اساس آن‌ها تصمیم‌گیری یا پیش‌بینی کنند. این روش‌ها شامل الگوریتم‌هایی مانند درخت تصمیم، ماشین بردار پشتیبان و شبکه‌های عصبی مصنوعی می‌باشد.

7. درخت تصمیم: (Decision Tree)

- یک مدل تصمیم‌گیری که از ساختار درختی برای نمایش تصمیمات و نتایج آن‌ها استفاده می‌کند. هر گره داخلی نشان‌دهنده یک ویژگی یا صفت است و هر برگ نشان‌دهنده یک نتیجه است.

8. جنگل تصادفی: (Random Forest)

- یک الگوریتم یادگیری ماشین که از ترکیب چندین درخت تصمیم برای بهبود دقت و کاهش خطر بیش‌برازش (overfitting) استفاده می‌کند.

9. انتخاب ویژگی: (Feature Selection)

- فرآیند انتخاب زیرمجموعه‌ای از ویژگی‌های موجود در داده‌ها که بیشترین تاثیر را بر مدل یادگیری ماشین دارند، به منظور بهبود دقت و کاهش پیچیدگی مدل.

10. پیش‌آگهی: (Prognosis)

- پیش‌بینی پیشرفت و نتیجه نهایی یک بیماری، که می‌تواند شامل احتمال بقا، مدت زمان بقا، و کیفیت زندگی بیمار باشد.

11. کروموزوم فیلادلفیا (Philadelphia Chromosome):

- یک ناهنجاری کروموزومی که در برخی از موارد لوسومی میلوژنوس مزمن دیده می‌شود و شامل ترانسلوکاسیون متقابل بین کروموزوم‌های 9 و 22 است.

12. تحلیل داده‌های بالینی و ژنتیکی:

- فرآیند بررسی و تحلیل اطلاعات حاصل از معاینات بالینی و داده‌های ژنتیکی بیماران به منظور شناسایی الگوها و ارتباطات مهم در بیماری‌ها.

13. انتقال یادگیری (Transfer Learning):

- یک تکنیک یادگیری ماشین که از دانش به دست آمده از یک حوزه یا مجموعه داده برای بهبود عملکرد مدل در حوزه یا مجموعه داده دیگر استفاده می‌کند.

10-1. وجه تمایز با سایر پژوهش‌ها:

این پژوهش در مقایسه با سایر مطالعات، دارای ویژگی‌ها و تمایزات زیر است:

11. استفاده از ترکیب گسترده‌ای از الگوریتم‌های یادگیری ماشین:

- این مطالعه به بررسی و مقایسه عملکرد چهار الگوریتم مختلف یادگیری ماشین شامل درخت تصمیم، جنگل تصادفی، **CATBoost**، آزمون کاکس پرداخته است. چنین ترکیب گسترده‌ای از الگوریتم‌ها در مطالعات مشابه کمتر دیده می‌شود.

2. تمرکز بر شناسایی عوامل مؤثر بر بقای پنج ساله:

- در حالی که بسیاری از پژوهش‌ها تنها به تشخیص لوسمی می‌پردازند، این مطالعه علاوه بر تشخیص، به شناسایی و تحلیل عواملی که بر بقای پنج ساله کودکان مبتلا به سرطان خون تأثیر دارند، نیز پرداخته است. این موضوع به طور ویژه برای بهبود فرآیند درمان و ارتقاء کیفیت زندگی بیماران اهمیت دارد.

3. استفاده از روش‌های انتخاب ویژگی:

- در این پژوهش، روش‌های انتخاب ویژگی برای بهینه‌سازی مدل‌های یادگیری ماشین به کار گرفته شده‌اند. این روش‌ها کمک می‌کنند تا مهم‌ترین ویژگی‌های بالینی و ژنتیکی که بیشترین تأثیر را بر پیش‌بینی بقای بیماران دارند، شناسایی شوند.

4. مقایسه جامع عملکرد الگوریتم‌ها:

- این پژوهش به مقایسه جامع عملکرد الگوریتم‌های مختلف یادگیری ماشین در تشخیص و پیش‌بینی بقای بیماران پرداخته است، که شامل بررسی دقیق، حساسیت، ویژگی و سایر معیارهای عملکرد است. چنین مقایسه جامعی به ندرت در سایر مطالعات دیده می‌شود. این ویژگی‌ها و تمایزات، این پژوهش را به یک منبع ارزشمند برای محققان، پزشکان و سیاست‌گذاران در حوزه بهداشت و درمان تبدیل می‌کند.

1-11 ملاحظات اخلاقی:

این پژوهش با رعایت کامل اصول و ملاحظات اخلاقی انجام شده است. در ادامه، برخی از مهم‌ترین موارد مرتبط با ملاحظات اخلاقی این پژوهش ذکر شده‌اند:

1. رضایت آگاهانه (Informed Consent):

- تمامی شرکت‌کنندگان در این مطالعه (یا والدین و سرپرستان قانونی در مورد کودکان) پس از توضیح کامل اهداف، روش‌ها، مزایا و مخاطرات احتمالی پژوهش، رضایت آگاهانه خود را به صورت کتبی ارائه دادند. همچنین به شرکت‌کنندگان اطمینان داده شد که مشارکت آنها کاملاً داوطلبانه بوده و در هر زمان می‌توانند از ادامه همکاری در پژوهش انصراف دهند.

2. حفظ محترمانگی و حریم خصوصی (Confidentiality and Privacy):

- اطلاعات شخصی و پزشکی شرکت‌کنندگان با رعایت کامل محترمانگی و حریم خصوصی جمع‌آوری و نگهداری شد. تمامی داده‌ها به صورت ناشناس و با استفاده از کدهای شناسایی منحصر به فرد ذخیره شدند تا از شناسایی شرکت‌کنندگان جلوگیری شود.

3. تصویب کمیته اخلاق (Ethical Committee Approval):

- این پژوهش قبل از شروع توسط کمیته اخلاق پژوهش‌های زیست‌پزشکی تأیید شد. پروپوزال پژوهش به دقت بررسی شد و تأییدیه لازم برای اجرای آن از نظر اخلاقی دریافت شد.

فصل دوم:

بررسی متون

1- مقدمه:

این فصل شامل 2 بخش است. در بخش اول به مبانی نظری پژوهش که شامل کلیاتی در مورد تعریف و کلیات سرطان خون در کودکان، عوامل موثر بر بقا در بیماران مبتلا به سرطان خون، روش‌های یادگیری ماشین در تشخیص و پیش‌بینی سرطان خون، تاریخچه و پیشینه استفاده از یادگیری ماشین و بخش دوم نیز به سابقه پژوهش‌های داخلی و خارجی و مقایسه نتایج آن می‌پردازیم.

2- تعریفات و کلیات سرطان خون در کودکان:

سرطان خون یکی از بیماری‌هایی است که در کودکان نیز می‌تواند رخدهد و اغلب در این گروه سنی مورد توجه قرار می‌گیرد. این بیماری ممکن است به صورت لوسی حاد یا مزمن ظاهر شود و اثرات جانبی جدی بر روی سلامتی و رشد کودکان داشته باشد. با توجه به حساسیت بالای سلول‌های خونساز در دوران کودکی، درمان این بیماری نیازمند رویکردهای خاصی است که باعث کاهش عوارض جانبی و افزایش شанс بقاء بیمار می‌شود [18].

3- عوامل خطر و علل سرطان خون در کودکان:

شناخت عوامل خطر و علل سرطان خون در کودکان از اهمیت بسیاری برخوردار است. عوامل ژنتیکی، محیطی و ترکیبی می‌توانند در بروز این بیماری نقش داشته باشند. برخی از عوامل خطر شامل اختلالات ژنتیکی خاص، مواد شیمیایی، و تابش‌های مضر می‌باشند. همچنین، تغذیه نامناسب و عوامل محیطی دیگر نیز می‌توانند به بروز سرطان خون در کودکان مرتبط باشند [19].

ویژگی‌های مولکولی و ژنتیکی سرطان خون در کودکان:

پیشرفت‌های اخیر در زمینه ژنتیک و مولکولی سرطان خون در کودکان، فرصتی را برای شناخت بهتر بیماری و توسعه رویکردهای درمانی دقیق‌تر فراهم کرده است. تفاوت‌های ژنتیکی در این گروه سنی می‌توانند الگوی درمان و پیش‌بینی نتایج را تغییر دهند. این شناخت عمیق‌تر از ویژگی‌های مولکولی و ژنتیکی به ما امکان می‌دهد تا روش‌های درمانی شخصی‌سازی شده و موثرتری را برای کودکان با سرطان خون ارائه دهیم.

۴- تشخیص و درمان سرطان خون در کودکان:

تشخیص و درمان سرطان خون در کودکان نیازمند رویکردهای خاصی است که به ویژگی‌های فیزیولوژیک و علایم بالینی این گروه سنی توجه دارد. روش‌های درمانی شامل شیمی‌درمانی، پرتو درمانی و پیوند مغز استخوان است. با پیشرفت‌های اخیر در زمینه تکنولوژی‌های پزشکی، درمان‌های جدیدی نظیر درمان‌های هدفمند و انتقال سلول‌های بنیادی نیز در حال توسعه می‌باشند.

(machine learning): یادگیری ماشین:

یادگیری ماشین (**Machine Learning**) یکی از شاخه‌های مهم و کاربردی هوش مصنوعی (**Artificial Intelligence**) است. این شاخه با استفاده از الگوریتم‌های یادگیری ماشین، به یادگیری و کشف الگوهای داده‌ها در داده‌ها می‌پردازد [20].

۵- قاریخچه یادگیری ماشین:

اولین بار توسط پروفسور جان مک‌کارتی در کالج دارتموث در تابستان ۱۹۵۶ پیشنهاد شد، هوش مصنوعی - (AI) هوش انسانی که توسط ماشین‌ها نمایش داده می‌شود - در فرهنگ لغات نسل‌های متوالی دانشمندان کامپیوتر، علاقه‌مندان به داستان‌های علمی تخیلی و محققان پزشکی جایگاه خود را پیدا کرده است. هدف از بی‌شمار شغل‌ها این بوده که ماشین‌های هوشمندی

بسازند که بتوانند دنیا را مانند انسان‌ها تفسیر کنند، زبان را بفهمند و از مثال‌های دنیای واقعی بیاموزند. در اوایل این قرن، دو رویداد هم‌زمان رخ داد که زمینه هوش مصنوعی را تغییر داد. ظهور واحدهای پردازش گرافیکی (GPUs) که به طور گسترده در دسترس بودند، به این معنا بود که پردازش موازی سریع‌تر، ارزان‌تر و قدرتمندتر شد. هم‌زمان، دوران "بیگ دیتا" – تصاویر، متون، بیوانفورماتیک، سوابق پزشکی و تراکنش‌های مالی و غیره – به جریان اصلی حرکت کرد و همراه با آن، فضای ذخیره‌سازی تقریباً نامحدود داده‌ها فراهم شد. این عوامل منجر به افزایش چشمگیر علاقه به هوش مصنوعی در هر دو حوزه دانشگاهی و صنایع خارج از علوم کامپیوتر سنتی شد. بار دیگر، هوش مصنوعی در ذهنیت زمانه قرار گرفته و آماده است تا پزشکی را در سطوح علمی پایه، بالینی، مدیریت بهداشت و مالی دگرگون کند.

ریشه‌های اصلی یادگیری ماشین را می‌توان در دهه 1950 جستجو کرد دانشمندان ایده‌های بنیادی را با الگوریتم‌هایی مانند پرسپترون و مدل نورون معرفی کردند. در دهه‌های 1960 و 1970، الگوریتم‌های کلیدی مانند پروپگاسیون معکوس و یادگیری تقویتی توسعه یافتدند. دهه 1980 و 1990 شاهد ظهور شبکه‌های عصبی عمیق و کاربرد آنها در وظایف تشخیص تصویر بود. دهه 2000 و 2010 نقطه عطفی در این زمینه بود، زیرا مفاهیمی مانند ارائه‌های نمایندگی و مجموعه داده‌های ImageNet انقلابی در یادگیری عمیق ایجاد کردند. در دهه 2020، شاهد نوآوری‌های پیشرفته‌ای مانند AlphaFold هستیم که از یادگیری عمیق برای درک ساختار پروتئین با دقت بسیار بالا استفاده می‌کند [21].

6-2 بررسی منابع داخلی:

"1-پژوهان" و همکاران در مطالعه‌ای با عنوان "تشخیص نوع لوسی می با کمک یادگیری ماشین: کاهش ابعاد و متوازن سازی" که در سال 1397 و به روش‌های یادگیری ماشین با الگوریتم‌های خطی، بیز ساده، فاصله اقلیدسی، نزدیک‌ترین میانگین، تطبیق قالب و دسته‌بندی غیرپارامتری با الگوریتم‌های تخمین‌گرهای پایه، هسته، $-k$ همسایه نزدیک‌تر و k همسایه

نزدیک‌تر مبتنی بر هسته بر روی داده‌های Leukemia1 از دانشگاه واندربریلت آمریکا برای طبقه‌بندی بر اساس سه زیر‌گروه سرطان خون ALL T-cell، ALL B-cell و AML انجام شد، یافته‌ها حاکی از آن است که با در نظر گرفتن الگوریتم‌های طرح شده، بهترین الگوریتم نزدیک‌ترین میانگین بود که به دقت پیش‌بینی 86/92٪ و همچنین با اعمال روش کاوش ویژگی PCA، باز هم بهترین نتیجه مربوط به الگوریتم نزدیک‌ترین میانگین بود و با متوسط تعداد ویژگی 6/8 به دقت 96٪ دست یافتند. در نهایت با متوازن‌سازی داده‌های Leukemia1، متوسط تعداد ویژگی و دقت توسط الگوریتم درجه 2 به ترتیب 5/41 و 59/98 حاصل گردید [22].

"-2زارعی فر" و همکاران، در مطالعه با عنوان، "میزان بقای پنج ساله و عوامل مؤثر آن در لوسومی های اطفال" با مطالعه‌ای گذشته نگر، 280 بیمار مبتلا به سرطان‌های ALL، AML به منظور محاسبه میزان بقا وارد مطالعه شدند. از روش‌های کاپلاین‌مایر و آزمون لگاریتم رتبه‌ای استفاده کردند. مدل سازی نیز با استفاده از مدل رگرسیون کاکس انجام دادند. یافته‌ها حاکی از آن است که، "میانگین دوره پی‌گیری بیماران 28.2±16.1 (دامنه: 6.4–67.4) ماه بوده که طی این دوره 60 مورد (درصد) از بیماران 47 مورد ALL و 15 مورد AML فوت کردند. میزان بقا تجمی 5 ساله سرطان در این مطالعه 0.055±53.3 درصد بدست آمد. مدل رگرسیون کاکس نشان داد که بین متغیرهای پلاکت و تعداد موارد عود بیماری، ارتباط معنا داری با میزان بقا سرطان وجود دارد [23]."

"-3اسماعیلی" و همکاران، در مطالعه‌ای با عنوان، "عیین عوامل مؤثر بر بقای پنج ساله کودکان مبتلا به لوسومی لنفوبلاستیک حاد بر اساس مدل بقای درختی در حضور مخاطرات رقیب" داده‌های مورد نیاز از ۲۵۵ کودک زیر ۱۵ سال که مبتلا به لوسومی لنفوبلاستیک حاد بودند و در سال‌های ۷-۲۰۰۱ به بیمارستان دکتر شیخ در مشهد، ایران مراجعه کرده بودند، جمع‌آوری شد. سپس بقای بیمار تا پایان مارس ۱۵ ۲۰ بررسی شد. در این راستا، متغیرهای مختلفی مانند جنسیت، سن، دوره درمان، شمارش گلبول‌های سفید خون، هموگلوبین، شمارش

پلاکت، سطح LDH ، در گیری CNS ، توده مدياستين، علائم روماتولوژيك و غيره نيز در نظر گرفته شد. عود بيماري به عنوان رويداد مطلوب در نظر گرفته شد، در حالى که مرگ بدون عود به عنوان رويداد رقابتی شناخته شد. زمان بقای بيماران از تاريخ تشخيص تا تاريخ رويداد (سانسور) به صورت ماهانه محاسبه شد. برازش مدل بر اساس حداکثر همگنی درون گرهای، که به نوبه خود بر اساس تابع تقسيم تغييرات باقیمانده مارتینگل خاص رويداد است، انجام می شود[24].

مطالعه حاکی از آن است که، ميانگين زمان بقای تخمينی در دوره‌های عود و مرگ بدون عود و همچنان در حضور هر دو رویداد به ترتیب ۵۵.۵۱، ۴۷.۵۳ و ۴۴.۲۰ ماه به دست آمد که نشان‌دهنده کاهش ميانگين زمان بقا در حضور ريسک‌های رقابتی است. شمارش گلبول‌های سفید خون و شمارش پلاکت به عنوان مؤثرترین عوامل مؤثر بر عود یا بقا در نظر گرفته شدند. سه زيرگروه از بيماران در معرض خطر شناسايي شدند و بيماراني که شمارش گلبول‌های سفید آن‌ها ≤ 50000 بود، به عنوان کسانی با کمترین ميانگين زمان بقا شناخته شدند.

7-بررسی منابع خارجی:

Liyan Pan و همکاران در مطالعه‌ای با عنوان "كاربردهای يادگیری ماشین برای پيش‌بياني عود در لوسمي لنفوبلاستيک حاد دوران کودکی" که در سال ۲۰۱۷ انجام دادند، به بررسی استفاده از الگوريتم‌های يادگیری ماشین برای پيش‌بياني عود در کودکان مبتلا به لوسمي لنفوبلاستيک حاد (ALL) پرداختند. اين مطالعه شامل ۶۶۱ کودک زير ۱۶ سال بود که در بازه زمانی ژانويه ۲۰۰۸ تا دسامبر ۲۰۱۵ در مرکز پزشكی زنان و کودکان گوانگزو، چين، تازه تشخيص داده شده و پس از درمان القايی به بهبودی كامل رسیده بودند[25].

محققان برای پيش‌بياني عود از يك استراتژي تحليل چند متغيره مبتنی بر داده استفاده کردند و تمام متغيرهای بيوشيميايی، مورفولوژيکی، ايمنی‌شناختی و سیتوژنتیکی مرتبط را که در اولین تشخيص یا در بهبودی كامل اندازه‌گيري شده بودند، جمع‌آوری نمودند. آن‌ها برای بهبود کيفيت داده‌ها از روش‌های جايگزينی مقادير گمشده و نرمال‌سازی استفاده کردند. همچنان، برای حل مشكل عدم تعادل داده‌ها، از روش نمونه‌برداری بيش از حد SMOTE استفاده کردند. برای

انتخاب مدل مناسب، چهار الگوریتم طبقه‌بندی شامل SVM، رگرسیون لجستیک (LR)، درخت تصمیم (DT) و جنگل تصادفی (RF) مورد استفاده قرار گرفتند. به منظور بهبود کارایی و عملکرد مدل‌ها، از روش انتخاب ویژگی جلو روشه استفاده شد تا مجموعه‌ای بهینه از ویژگی‌ها با کمترین تعداد و بالاترین دقت پیش‌بینی انتخاب شود [26].

در نهایت، مدل‌ها با استفاده از بسته Scikit-learn در پایتون توسعه یافتند و عملکرد آن‌ها با استفاده از معیارهایی مانند دقت، حساسیت، ویژگی، ارزش پیش‌بینی مثبت (PPV) و ارزش پیش‌بینی منفی (NPV) ارزیابی شد. این مطالعه نشان داد که الگوریتم‌های یادگیری ماشین می‌توانند به طور موثری برای پیش‌بینی عود در کودکان مبتلا به ALL استفاده شوند و به بهبود فرآیندهای تصمیم‌گیری بالینی کمک کنند.

2. کیوان کرمی و همکاران در مطالعه‌ای با عنوان "عوامل پیش آگهی بقا در بیماران مبتلا به لوسمی میلوئید حاد با استفاده از تکنیک‌های یادگیری ماشینی" که در سال 2021 به روش الگوریتم‌های یادگیری ماشین انجام شد، یافته‌های آن حاکی از آن است که چندین متغیر برای بیماران در پایگاه داده فعلی ثبت شده است که شامل متغیرهای عددی و دسته‌ای می‌شود. بیماران بر اساس سیستم فرانسوی-آمریکایی-بریتانیایی (FAB) طبقه‌بندی شدند. داده‌ها با استفاده از ابزارها و الگوریتم‌های داده کاوی برای تجزیه و تحلیل آماده شدند. انتخاب ویژگی از طریق روش‌های وزن‌دهی ویژگی انجام شد و پس از این بخش، 25 مورد از ویژگی‌های با وزن بالا برای ادامه تحلیل انتخاب شدند. همانطور که در بخش‌های زیر ذکر شد، طبقه‌بندی کننده‌های متعدد برای توانایی آنها در پیش‌بینی بقای بیماران آموزش دیده و ارزیابی شدند [27].

3. Yidong Zhu و همکاران در مطالعه‌ای با عنوان "شناسایی شش نشانگر زیستی تشخیصی برای لوسمی لنفوسيتی مزمن بر اساس الگوریتم‌های یادگیری ماشینی"، طبق مطالعات آن‌ها شش مجموعه داده شامل نمونه‌های CLL (لوسمی لنفوسيتی مزمن) و نمونه‌های کنترل از

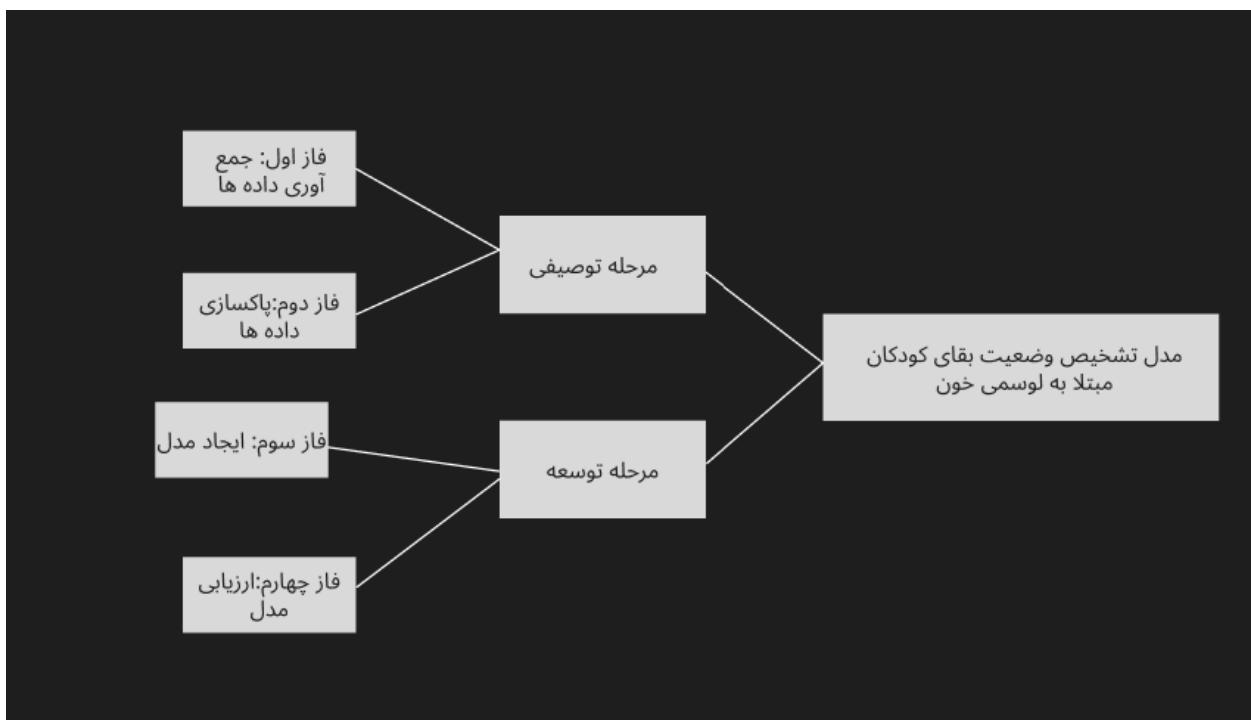
پایگاه داده Gene Expression Omnibus دانلود شد. تحلیل بیان دیفرانسیلی ژن‌ها، تحلیل شبکه همبیانی وزنی ژن‌ها (WGCNA) و رگرسیون LASSO (کمترین انقباض مطلق و اپراتور انتخاب) برای شناسایی نشانگرهای زیستی تشخیصی بالقوه برای CLL با استفاده از نرم‌افزار R اعمال شد. عملکرد تشخیصی ژن‌های محوری سپس با استفاده از تحلیل منحنی ROC (ویژگی عملیاتی گیرنده) اندازه‌گیری شد. تحلیل عملکردی برای کشف مکانیسم‌های زیرین اجرا شد. علاوه بر این، تحلیل همبستگی برای ارزیابی رابطه بین ژن‌های محوری و ویژگی‌های ایمنی انجام شد. نتیجه حاکی از آن است که تعداد کل 47 ژن با بیان دیفرانسیلی (DEGs) و 25 ژن محوری کاندید از طریق تحلیل بیان دیفرانسیلی ژن‌ها و WGCNA به ترتیب استخراج شد. بر اساس 14 ژن مشترک بین DEGs و ژن‌های محوری کاندید، تحلیل رگرسیون LASSO انجام شد که در نهایت شش ژن محوری به عنوان نشانگرهای زیستی بالقوه برای CLL شناسایی شدند: ABCA6، TEAD2، FILIP1L، EBF1، PMEPA1، CCDC88A و ROC. این شش ژن توانایی پیش‌بینی قابل اطمینانی را در گروه‌های آموزشی و اعتبارسنجی نشان دادند، به طوری که همه مقادیر AUC بالای 0.80 بودند. تحلیل عملکردی وضعیت ایمنی غیرعادی در بیماران CLL را نشان داد. همبستگی قابل توجهی بین ژن‌های محوری و مسیرهای مرتبط با ایمنی یافت شد که نشان‌دهنده ارتباط نزدیک بین ژن‌های محوری و ایمنی تومور در CLL است [28].

فصل سوم

مواد و روش کار

1-3 مقدمه

در این بخش نوع و محیط پژوهه، روش‌ها و ابزار تحلیل داده‌ها شرح داده شده است. کلیه مراحلی که برای انجام پژوهش طی شده است به ترتیب انجام توصیف گردیده است. این یک مطالعه توصیفی بالینی با هدف کشف عوامل موثر بر بقای کودکان مبتلا به لوسومی بوسیله مدل‌های یادگیری ماشین برای استخراج ویژگی‌ها و تحلیل و تفسیر آن ویژگی‌ها است.



شکل 1-3: نقشه کلی ایجاد فرایند

۲-۳ فاز اول: جمع آوری داده

۱-۲-۳ نوع پژوهش

۱. از نظر هدف

داده ها و اطلاعات از آزمایشگاه های مختلف با استفاده از مشاهدات ثبت شده مراجعان در محل پژوهش بدست آمده است که نوعی تحقیق مشاهده ای تلقی میشود. این داده ها خود به دو روش توصیفی و تحلیلی تقسیم میشوند. لذا با توجه به اینکه این پژوهش به بررسی و تحلیل متغیر های بدست آمده پرداخته است، یک پژوهش مشاهده ای - توصیفی است.

۲. از نظر زمان

این فاز در یک بازه چند ساله برای ذخیره و جمع آوری داده ها انجام گردید است لذا این قسمت از پژوهش مقطوعی - گذشته نگر می باشد.

۳. از نظر شیوه اجرا

در این فاز پژوهش یک شیوه اجرای کیفی جهت تفسیر و تحلیل داده ها نموده ایم.

۲-۲-۳ محیط انجام پژوهش

دانشکده علوم پزشکی اسپرائین بخش تحقیقات.

۳-۲-۳ ابزار گردآوری و تحلیل

داده ها بصورت کیفی و کمی از خروجی نهایی آزمایشات مربوطه، گردآوری شده اند.

۴-۲-۳ حجم نمونه گیری

در مرحله تعداد 260 نمونه گردآوری شده و بعد از بررسی های لازم و تمیز دادن دیتا ها برای تحلیل، داده های نامعتبر حذف گردید. به دلیل اطلاعات ناقص در برخی فیلد ها از روش های آماری با حفظ اثر بخشی در نتیجه، (مثل میانگین گیری) برای برخی داده های از دست رفته، جایگزین داده در نظر گرفته شد.

۵-۲-۳ روش اجرا و پیاده سازی

در این مطالعه تعداد 260 داده با ویژگی ها مربوطه از آزمایشگاه ها جمع آوری شده و بعد از بررسی های داده ای، حدود 12 داده به صورت میسینگ (Invalid data) تشخیص داده شد که با استفاده از روش های آماری مثل جایگزین میانگین اصلاح شده بجای داده های از دست رفته، داده ها اصلاح و برای فاز پیش پردازش و توسعه مدل آماده شد.

۳-۳ فاز دوم: پیش پردازش داده ها

۱-۳-۳ نوع پژوهش

این فاز پژوهش در یک بازه یک ماهه به منظور پیش پردازش داده ها مورد نیاز انجام گردید لذا این قسمت نیز یک پژوهش مقطعی-گذشته نگر و توصیفی می باشد.

۲. از نظر شیوه اجرا

در این فاز پژوهش یک شیوه اجرای کمی جهت پیش پردازش داده ها استفاده نمودیم. به این صورت که داده ها بصورت عددی توسط کتابخانه های تحلیل داده پایتون، مورد بررسی قرار گرفت.

۲-۳-۳ محیط وابزار انجام پژوهش

در این فاز پژوهش در نرم افزار **jupyter – notebook** توسط زبان پایتون و کتابخانه های آماری و داده کاوی آن مورد بررسی کاوش قرار گرفت.

۳-۳-۳ روش اجرا و پیاده سازی

با بررسی مقالات قبلی، جزئیات روش‌ها و نتایج آنها استخراج و در فرم‌های طراحی شده مخصوص ثبت شدند. دو پژوهشگر داده‌ها را استخراج کردند و اختلاف نظرها با بحث و گفتگو حل شد. کیفیت مطالعات واحد شرایط بر اساس معیارهای پیشنهادی ارزیابی شد. [29] ارزیابی بر اساس پنج دسته‌بندی انجام شد: نیاز برآورده نشده، محدودیت‌های روش‌های غیر ML کنونی، قابلیت باز تولید (روش‌های مهندسی ویژگی‌ها، پلتفرم‌ها/پکیج‌ها، هایپرپارامترها)، پایداری (روش‌های معتبر برای غلبه بر بیش‌برازش، ثبات نتایج)، قابلیت تعمیم (اعتبار سنجی داده‌های خارجی) و اهمیت بالینی (توضیح پیش‌بینی کننده‌ها و کاربرد بالینی پیشنهادی). جدول ارزیابی کیفیت با ذکر "بله" یا "خیر" برای موارد مربوطه در هر دسته‌بندی ارائه شد. از بین متغیر‌های جمع آوری شده، بر اساس همبستگی پیرسون، متغیر‌ها با همبستگی بالا برای یادگیری ماشین استخراج شد. به دلیل از دست دادن برخی داده‌ها، مقادیر از دست رفته نیز با میانگین اصلاح شده با ضریب 20 درصد، ترمیم شد [30].

۴-۳ فاز سوم: ایجاد مدل

۱-۴-۳ نوع پژوهش

۱. از نظر هدف

داده ها و اطلاعات برای طراحی مدل ماشینی با استفاده از روش های داده کاوی در استخراج و انتخاب ویژگی به دست آمد. لذا این فاز پژوهش از نظر هدف یک پژوهش توسعه ای می باشد.

۲. از نظر زمان

از فاز پژوهش در یک بازه یک ماهه به منظور ارزیابی کیفیت داده و آموزش مدل های مختلف، انجام گردید لذا این قسمت یک پژوهش مقطعی - گذشته نگر می باشد.

۳. از نظر شیوه اجرا

این فاز پژوهش یک شیوه اجرای کمی جهت ساخت مدل و ایجاد سیستم تحلیلی مبتنی بر مدل استفاده نمودیم.

۴. از نظر نتایج

نتایج حاصل از این فاز برای ارزیابی در فاز پایانی مورد استفاده قرار گرفت، لذا این فاز از نظر نتایج یک پژوهش کابردی - توسعه ای می باشد.

۲-۴-۳ روش اجرا و پیاده سازی

در این پژوهش، هدف اصلی ما توسعه یک مدل یادگیری ماشین برای شناسایی عوامل موثر بر بقای پنج ساله کودکان مبتلا به لوسومی است. به منظور دستیابی به این هدف، مراحل زیر را طی کردیم:

3.2.1 گردآوری داده‌ها

ابتدا داده‌های مربوط به بیماران مبتلا به لوسومی از چندین پایگاه داده معتبر پزشکی جمع‌آوری شد. این داده‌ها شامل اطلاعات دموگرافیک، بالینی و نتایج آزمایشگاهی بودند. همچنین، اطلاعاتی نظیر نوع لوسومی، روش‌های درمانی استفاده شده و نتایج بقا نیز جمع‌آوری گردید.

3.2.2 پیش‌پردازش داده‌ها

داده‌های جمع‌آوری شده نیازمند پیش‌پردازش بودند تا برای استفاده در مدل‌های یادگیری ماشین آماده شوند. این مرحله شامل موارد زیر بود:

- پاک‌سازی داده‌ها: حذف داده‌های ناقص یا اشتباه و تصحیح داده‌های نادرست.
- نرمال‌سازی داده‌ها: مقیاس‌بندی داده‌ها به یک محدوده مشخص برای بهبود عملکرد مدل.
- تبدیل دسته‌ها به مقادیر عددی: تبدیل داده‌های دسته‌ای به مقادیر عددی قابل استفاده در مدل‌های یادگیری ماشین.

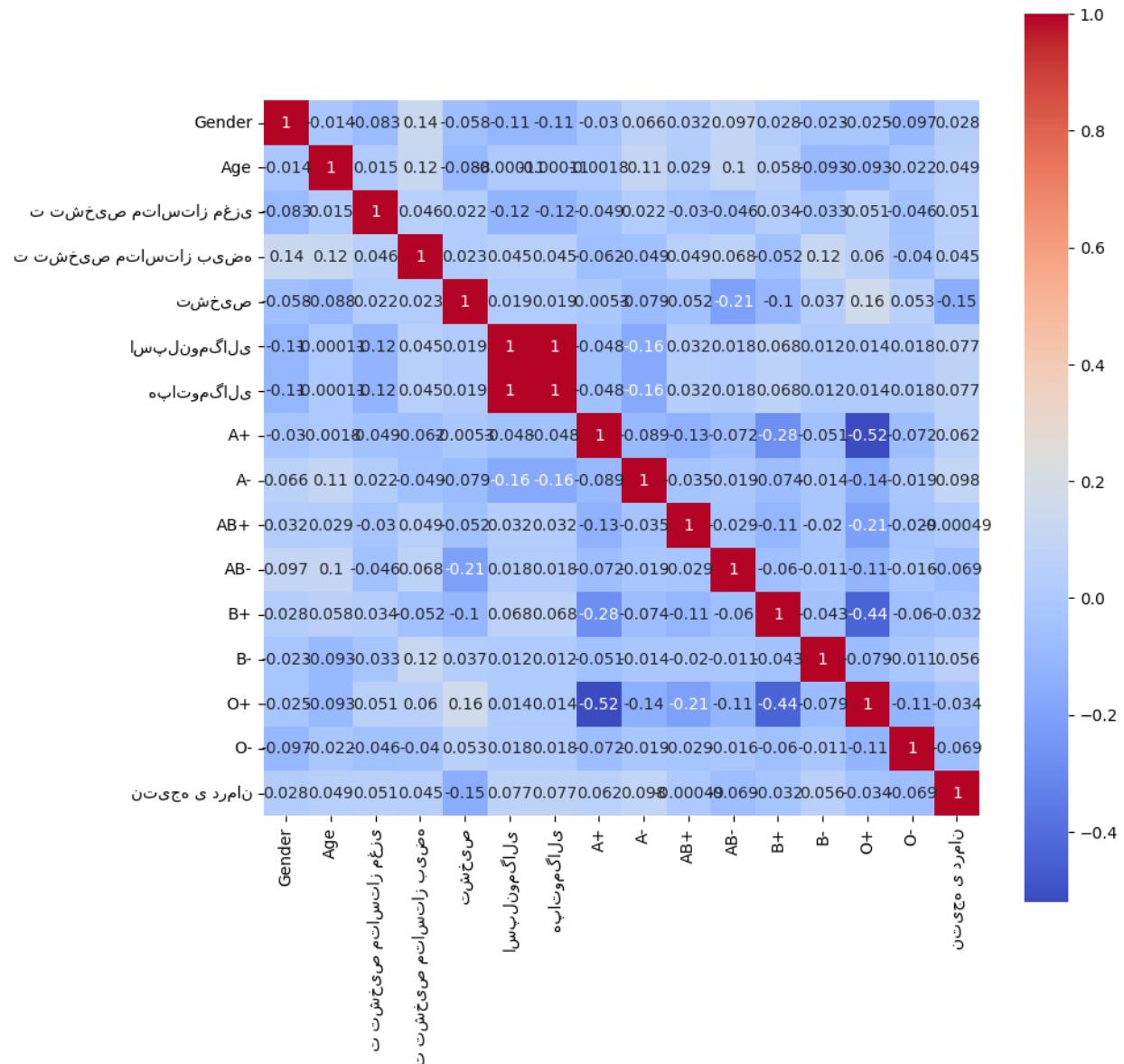
3.2.3 انتخاب ویژگی‌ها

برای بهبود دقت مدل و کاهش پیچیدگی آن، انتخاب ویژگی‌های مهم از داده‌ها ضروری بود. از روش‌های مختلف انتخاب ویژگی نظیر تحلیل مؤلفه‌های اصلی (PCA) و الگوریتم‌های انتخاب ویژگی مبتنی بر آمار استفاده شد.

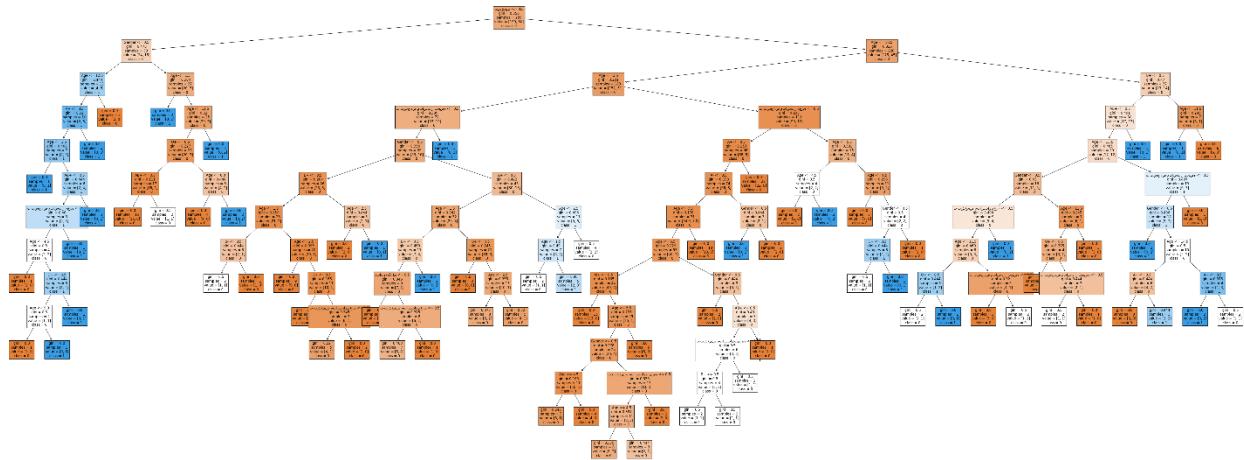
3.2.4 توسعه مدل‌های یادگیری ماشین

4 الگوریتم مختلف یادگیری ماشین برای شناسایی عوامل موثر بر بقای پنج ساله کودکان مبتلا به لوسمی مورد استفاده قرار گرفتند. این الگوریتم‌ها عبارتند از:

- درخت تصمیم (Decision Tree)
- جنگل تصادفی (Random Forest)
- catBoost
- مدل ارزیابی بقا کاکس



شکل 3-2: ارتباط ویژگی متغیر ها بر اساس همبستگی پیرسون



شکل 3-3: درخت تصمیم پیاده شده بر مدل

3.2.5 ارزیابی مدل‌ها

مدل‌های توسعه یافته با استفاده از مجموعه داده‌های تست و معیارهای ارزیابی مختلف نظیر دقت، حساسیت، ویژگی، و F1-score ارزیابی شدند. همچنین، برای اطمینان از پایداری مدل‌ها از روش‌های اعتبارسنجی متقاطع استفاده شد.

	precision	recall	f1-score	support
0	0.85	1.00	0.92	33
1	0.00	0.00	0.00	6
accuracy			0.85	39
macro avg	0.42	0.50	0.46	39
weighted avg	0.72	0.85	0.78	39
دقت مدل :	0.846			

شکل 4-3: ارزیابی دقت مدل با معیارهای مطرح شده

3.2.6 تحلیل نتایج

نتایج به دست آمده از ارزیابی مدل‌ها تحلیل و بررسی شد. مدل‌های مختلف با یکدیگر مقایسه شده و بهترین مدل بر اساس دقت و سایر معیارهای ارزیابی انتخاب گردید. بر اساس فاکتورهای ارزیابی، مدل **catboost** مورد پذیرش قرار گرفت، همچنین، تاثیر هر ویژگی بر پیش‌بینی بقا تحلیل شد تا بتوان عوامل موثر بر بقای پنج ساله کودکان مبتلا به لوسومی را شناسایی کرد.

3.2.7 پیاده‌سازی و استفاده از مدل در سیستم‌های بالینی

در نهایت، مدل انتخاب شده به منظور استفاده در سیستم‌های تصمیم‌یار بالینی پیاده‌سازی شد. این سیستم‌ها می‌توانند به پزشکان در تشخیص زودهنگام لوسومی و پیش‌بینی بقا کمک کرده و در نتیجه بهبود نتایج درمانی کودکان مبتلا به لوسومی را فراهم کنند.

این مراحل به طور جمعی منجر به توسعه یک مدل یادگیری ماشین کارآمد و قابل اطمینان برای شناسایی عوامل موثر بر بقای پنج ساله کودکان مبتلا به لوسومی شده است. در ادامه، جزئیات بیشتری از نتایج به دست آمده و تحلیل‌های انجام شده ارائه خواهد شد.

۵-۳ نتیجه گیری فصل

در این فصل به محیط انجام پژوهش و همچنین بررسی روند کار، ابزار گرد آوری داده و نحوه تحلیل داده‌ها، ساخت مدل‌ها و ارزیابی آن‌ها جهت انتخاب بهترین مدل و ویژگی‌های استفاده شده برای این منظور، معرفی گردیدند.

رفرنس ها تا فصل سوم

[1]: <https://pubmed.ncbi.nlm.nih.gov/24784336/>

[2]: <https://www.sid.ir/paper/10274/fa>

[3]: Gatta G, Luksch R, Coleman MP, Corazziari I. Survival from acute non-lymphocytic leukaemia (ANLL) and chronic myeloid leukaemia (CML) in European children since 1978: a population-based study. Eur J Cancer 2001; 37: 695-702.

[4]: 1. Abunadi, I., & Senan, E. M. (2022, February 19). Multi-Method Diagnosis of Blood Microscopic Sample for Early Detection of Acute Lymphoblastic Leukemia Based on Deep Learning and Hybrid Techniques. *< i> Sensors*. MDPI AG. <http://doi.org/10.3390/s22041629>

[5]: 2. Fathi, E., Rezaee, M. J., Tavakkoli-Moghaddam, R., Alizadeh, A., & Montazer, A. (2020, July 7). Design of an integrated model for diagnosis and classification of pediatric acute leukemia using machine learning. *< i> Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*. SAGE Publications.

<http://doi.org/10.1177/0954411920938567>

[6]: 3. Karami, K., Akbari, M., Moradi, M.-T., Soleymani, B., & Fallahi, H. (2021, July 21). Survival prognostic factors in patients with acute myeloid leukemia using machine learning techniques. (S. Palaniyandi, Ed.), *< i> Plos One*. Public Library of Science (PLoS).
<http://doi.org/10.1371/journal.pone.0254976>

[7]: [1] Pession A, Dama E, Rondelli R, Magnani C, De Rosa M, Locatelli F, et al. Survival of children with cancer in Italy, 1989–98. A report from the hospital based registry of the Italian association of paediatric haematology and oncology (AIEOP). Eur J Cancer 2008; 44: 1282-1289.

[8]: <https://pubmed.ncbi.nlm.nih.gov/35528341/>

[1] Pession A, Dama E, Rondelli R, Magnani C, De Rosa M, Locatelli F, et al. Survival of children with cancer in Italy, 1989–98. A report from the hospital based registry of the Italian association of paediatric haematology and oncology (AIEOP). Eur J Cancer 2008; 44: 1282-1289.

[9]: [2] Desandes E, Berger C, Tron I, Demeocq F, Bellec S, Blouin P, et al. Childhood cancer survival in France, 1990–1999. Eur J Cancer 2008; 44: 205-215.

[10]: Karimi M, Mehrabani D, Yarmohammadi H, Jahromi FS. The prevalence of signs and symptoms of childhood leukemia and lymphoma in Fars Province, Southern Iran. Cancer Detect Prev 2008; 32: 178-183.

[11]: Pui CH, Schrappe M, Ribeiro RC, Niemeyer CM. Childhood and adolescent lymphoid and myeloid leukemia. Hematology Am Soc Hematol Educ Program 2004; 118-145.

[12]: <https://ashpublications.org/bloodadvances/article/4/23/6077/474420/Application-of-machine-learning-in-the-management>

[13]: [Keyvan Karami^{1,2}](#), [Mahboubeh Akbari³](#), [Mohammad-Taher Moradi¹](#), [Bijan Soleymani¹](#), [Hossein Fallahi⁴](#)

<https://pubmed.ncbi.nlm.nih.gov/34288963/>

[14]: <https://onlinelibrary.wiley.com/doi/full/10.1111/ijlh.13089>

[15]: <https://pubmed.ncbi.nlm.nih.gov/25680783/>

[16]: <https://pubmed.ncbi.nlm.nih.gov/25680783/>

[17]: <https://www.ncbi.nlm.nih.gov/books/NBK560490/>
<https://jamanetwork.com/journals/jama/fullarticle/198779>

[18]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10531920/>

[19]: <https://www.ncbi.nlm.nih.gov/books/NBK65763/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8880753/>

<https://pubmed.ncbi.nlm.nih.gov/35053543/>

[20]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3734905/>

<https://pubmed.ncbi.nlm.nih.gov/35053543/>

[22]: <https://link.springer.com/article/10.1007/BF02478259>

[23]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5987686/>

[24]: <https://www.sid.ir/paper/132525/fa>

[25]:

<https://www.sid.ir/paper/37399/fa#:~:text=%DB%8C%D8%A7%D9%81%D8%AA%D9%87%20%D9%87%D8%A7%3A%20%D9%85%DB%8C%D8%A7%D9%86%DA%AF%DB%8C%D9%86%20%D8%AF%D9%88%D8%B1%D9%87%20%D9%BE%DB%8C,0.055%20%D8%AF%D8%B1%D8%B5%D8%AF%20%D8%A8%D9%87%20%D8%AF%D8%B3%D8%AA%20%D8%A2%D9%85%D8%AF.>

[26]: https://goums.ac.ir/jorjanijournal/browse.php?a_id=583&slc_lang=fa&sid=1&printcase=1&hbnr=1&hmb=1

[27]: <https://www.nature.com/articles/s41598-017-07408-0>

[28]: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0254976>

[29]: <https://pubmed.ncbi.nlm.nih.gov/36467501/>

[30]: 15 Qiao N., A systematic review on machine learning in sellar region diseases: quality and reporting items, *Endocrine Connections*. (2019) **8**, no. 7, 952–960, <https://doi.org/10.1530/ec-19-0156>, 2-s2.0-85069698199.

[31]: <https://onlinelibrary.wiley.com/doi/full/10.1155/2021/9933481>