

بررسی کاربرد الگوریتم‌های یادگیری ماشین در تشخیص اختلالات قلبی از سیگنال ECG

مقدمه

الکتروکاردیوگرام یا ECG یکی از مهم‌ترین سیگنال‌های زیستی برای ارزیابی فعالیت الکتریکی قلب است. این سیگنال اطلاعات ارزشمندی درباره ریتم قلب، هدایت الکتریکی، تغییرات قطعه ST، فعالیت دهلیزی و بطنی و برخی شواهد مرتبط با بیماری‌های قلبی در اختیار پزشک قرار می‌دهد. با این حال، تفسیر ECG، به‌ویژه در داده‌های طولانی‌مدت یا در حضور نویز و آرتیفکت، فرایندی زمان‌بر و وابسته به تجربه متخصص است.

پیشرفت روش‌های یادگیری ماشین (Machine Learning) و یادگیری عمیق (Deep Learning) این امکان را ایجاد کرده است که بخش قابل توجهی از فرایند تحلیل ECG به صورت خودکار انجام شود. در این رویکرد، الگوریتم با استفاده از نمونه‌های ECG برچسب‌گذاری شده، الگوهای مرتبط با وضعیت‌های طبیعی یا پاتولوژیک را یاد می‌گیرد و سپس برای طبقه‌بندی نمونه‌های جدید مورد استفاده قرار می‌گیرد.

این موضوع از دیدگاه مهندسی پزشکی اهمیت ویژه‌ای دارد؛ زیرا مسئله صرفاً انتخاب یک الگوریتم طبقه‌بندی نیست. یک سامانه قابل اتکا باید مجموعه‌ای از مراحل شامل دریافت سیگنال، حذف نویز، قطعه‌بندی، استخراج یا یادگیری ویژگی، طراحی مدل، اعتبارسنجی و تفسیر خروجی را به صورت یکپارچه مدیریت کند.

در سال‌های اخیر، معماری‌هایی مانند CNN، RNN و مدل‌های ترکیبی در تشخیص آریتمی‌ها و سایر الگوهای ECG مورد توجه گسترده قرار گرفته‌اند. با این حال، عملکرد بالای گزارش شده در مطالعات پژوهشی لزوماً به معنای آمادگی یک مدل برای استفاده بالینی نیست. کیفیت برچسب‌ها، نحوه تقسیم داده‌ها، تفاوت جمعیت‌های مورد مطالعه، عدم توازن کلاس‌ها، قابلیت تعمیم و تفسیرپذیری از جمله عواملی هستند که باید در ارزیابی چنین سامانه‌هایی در نظر گرفته شوند. انجمن قلب آمریکا نیز بر اهمیت کیفیت داده، تعریف صحیح برچسب مرجع و ارزیابی در محیط‌های بالینی واقعی تأکید کرده است.

۱. سیگنال ECG و اطلاعات قابل استخراج از آن

ECG حاصل ثبت اختلاف پتانسیل‌های الکتریکی ایجاد شده در اثر فعالیت قلب است. در ECG دوازده لید، اطلاعات حاصل از زوایای مختلف نسبت به فعالیت الکتریکی قلب ثبت می‌شود و همین موضوع امکان بررسی جامع‌تر فعالیت قلبی را فراهم می‌کند.

از نظر فیزیولوژیک، بخش‌های مختلف موج ECG اطلاعات متفاوتی دارند:

- موج P: عمدتاً نشان‌دهنده دپولاریزاسیون دهلیزها است.
- کمپلکس QRS: عمدتاً با دپولاریزاسیون بطن‌ها ارتباط دارد.
- موج T: بیانگر رپولاریزاسیون بطنی است.

- **فاصله PR:** اطلاعاتی درباره انتقال تحریک از دهلیز به بطن ارائه می‌کند.
- **فاصله QT:** بیانگر مدت زمان فعالیت الکتریکی بطن‌ها است.
- **قطعه ST:** در ارزیابی برخی تغییرات ایسکمیک اهمیت دارد.
- **فواصل RR:** برای بررسی ضربان و ریتم قلب بسیار مهم هستند.

از دیدگاه پردازش سیگنال، ECG را می‌توان یک سیگنال شبه‌تناوبی دانست که ویژگی‌های زمانی، دامنه‌ای، مورفولوژیک و در برخی کاربردها فرکانسی دارد.

برای مثال، اگر هدف تشخیص آریتمی باشد، تغییرات فاصله RR می‌تواند اطلاعات مهمی فراهم کند؛ در حالی که برای تشخیص برخی اختلالات مورفولوژیک، شکل QRS، ارتفاع و عرض امواج یا تغییرات ST/T ممکن است اهمیت بیشتری داشته باشند.

بنابراین، انتخاب ویژگی‌ها باید بر اساس **ماهیت مسئله تشخیصی** انجام شود، نه صرفاً بر اساس اینکه کدام ویژگی از نظر محاسباتی آسان‌تر استخراج می‌شود.

۲. چرا یادگیری ماشین برای تحلیل ECG اهمیت دارد؟

روش سنتی تحلیل خودکار ECG معمولاً بر پایه استخراج ویژگی‌های مهندسی‌شده و سپس اعمال یک طبقه‌بند بود. در این ساختار، ابتدا سیگنال پردازش می‌شد، ویژگی‌های مناسب استخراج می‌شدند و سپس الگوریتمی مانند SVM یا Random Forest وضعیت سیگنال را طبقه‌بندی می‌کرد.

یادگیری ماشین این فرایند را انعطاف‌پذیرتر کرده است. در روش‌های کلاسیک هنوز طراحی ویژگی نقش مهمی دارد، اما در یادگیری عمیق می‌توان بخشی از ویژگی‌ها را مستقیماً از سیگنال خام یا تقریباً خام یاد گرفت.

به‌صورت ساده می‌توان فرایند را چنین نمایش داد:

→ ECG پیش‌پردازش → استخراج/یادگیری ویژگی → طبقه‌بندی → خروجی تشخیصی

در معماری‌های عمیق‌تر، مرز بین استخراج ویژگی و طبقه‌بندی تا حد زیادی از بین می‌رود:

→ ECG شبکه عصبی → ویژگی‌های آموخته‌شده → احتمال کلاس‌ها

این ویژگی باعث شده است یادگیری عمیق برای داده‌های پیچیده و چندبعدی ECG جذاب باشد.

۳. زنجیره پردازش ECG برای سیستم‌های یادگیری

ماشین

یکی از مهم‌ترین نکات در طراحی سامانه‌های تشخیص خودکار این است که عملکرد مدل را نباید جدا از کیفیت کل زنجیره پردازش بررسی کرد.

۳.۱ دریافت و نمونه‌برداری سیگنال

ECG می‌تواند با دستگاه‌های بالینی، هولتر، تجهیزات پوشیدنی یا سامانه‌های پایش مداوم ثبت شود.

پارامترهایی مانند:

- نرخ نمونه‌برداری،
- تعداد لیدها،
- رزولوشن ADC،
- مدت زمان ثبت،
- کیفیت تماس الکترودها

بر کیفیت داده تأثیر می‌گذارند.

برای مثال، PTB-XL یکی از مجموعه داده‌های شناخته شده در پژوهش‌های ECG است که شامل ECG های بالینی ۱۲ لید با طول ۱۰ ثانیه است. این مجموعه در نسخه‌های منتشر شده شامل بیش از ۲۱ هزار رکورد و اطلاعات مربوط به نزدیک به ۱۹ هزار بیمار است و برای هر رکورد برچسب‌های تشخیصی و فراداده‌های مختلف ارائه می‌کند.

۳.۲ حذف نویز

ECG در محیط واقعی با منابع مختلف نویز مواجه است:

- **Baseline Wander**
- نویز برق شهر
- آرتیفکت حرکتی
- نویز ناشی از تماس الکترودها
- تداخل عضلانی یا EMG
- تغییرات ناشی از تنفس

برای مقابله با این مشکلات می‌توان از روش‌هایی مانند فیلترهای دیجیتال، فیلترهای تطبیقی، تبدیل موجک و روش‌های مبتنی بر تجزیه سیگنال استفاده کرد.

با این حال، فیلترگذاری بیش از حد می‌تواند بخشی از اطلاعات واقعی ECG را حذف کند. این موضوع خصوصاً زمانی مهم است که مدل برای تشخیص تغییرات ظریف ST یا مورفولوژی موج طراحی شده باشد.

بنابراین، هدف پیش‌پردازش صرفاً «تمیز کردن» سیگنال نیست؛ بلکه باید نسبت سیگنال به نویز را بهبود داد بدون اینکه ویژگی تشخیصی سیگنال تخریب شود.

۳.۳ تشخیص ضربان و قطعه‌بندی

در بسیاری از الگوریتم‌های کلاسیک، تشخیص نقاط R و استخراج ضربان‌های منفرد مرحله مهمی است.

با تعیین موقعیت QRS و به‌خصوص قله R می‌توان:

- فاصله RR را محاسبه کرد،
- ضربان‌های منفرد را استخراج کرد،
- HRV را بررسی کرد،
- ضربان‌ها را برای طبقه‌بندی جداگانه آماده کرد.

اما در مدل‌های یادگیری عمیق، الزاماً لازم نیست تمام این مراحل به شکل دستی انجام شوند. یک شبکه می‌تواند بخشی از الگوهای زمانی را مستقیماً از ورودی یاد بگیرد.

۴. استخراج ویژگی در تحلیل کلاسیک ECG

در روش‌های یادگیری ماشین سنتی، انتخاب ویژگی یکی از مهم‌ترین مراحل طراحی سیستم است.

ویژگی‌های زمانی

برخی ویژگی‌های رایج عبارت‌اند از:

- میانگین و انحراف معیار RR
- حداقل و حداکثر RR
- مدت QRS
- فاصله PR
- فاصله QT
- دامنه موج‌ها
- مدت زمان بخش‌های مختلف ECG

ویژگی‌های مورفولوژیک

در این روش، شکل موج‌ها بررسی می‌شود. برای نمونه:

- شکل QRS
- نسبت دامنه R به S
- تغییرات ST
- ویژگی‌های موج T

- شکل و دامنه P

ویژگی های فرکانسی

با استفاده از FFT ، STFT یا روش های مشابه می توان اطلاعات حوزه فرکانس را بررسی کرد.

ویژگی های غیر خطی

ECG یک سیگنال زیستی پیچیده است و در برخی پژوهش ها از روش های تحلیل غیر خطی مانند:

- Entropy
- Sample Entropy
- Approximate Entropy
- Poincaré Plot ویژگی های مبتنی بر

نیز استفاده می شود.

این ویژگی ها می توانند در تحلیل تغییر پذیری ضربان قلب مفید باشند.

۵. الگوریتم های کلاسیک یادگیری ماشین در تشخیص ECG

۵.۱ K-Nearest Neighbors

الگوریتم KNN نمونه جدید را بر اساس نزدیک ترین نمونه های موجود در فضای ویژگی طبقه بندی می کند.

مزیت آن سادگی و پیاده سازی نسبتاً آسان است، اما با افزایش حجم داده و تعداد ویژگی ها، هزینه محاسباتی آن افزایش می یابد. همچنین عملکرد KNN به انتخاب معیار فاصله و مقیاس بندی ویژگی ها وابسته است.

۵.۲ Support Vector Machine

SVM یکی از روش های مهم در طبقه بندی داده های زیستی است.

ایده اصلی SVM یافتن یک مرز تصمیم مناسب میان کلاس ها است. با استفاده از Kernel می توان روابط غیر خطی را نیز مدل کرد.

SVM به ویژه در شرایطی که تعداد نمونه ها نسبت به تعداد ویژگی ها محدود است، می تواند گزینه مناسبی باشد.

در یک پروژه مهندسی پزشکی، برای مثال می‌توان ویژگی‌های مورفولوژیک و زمانی ECG را استخراج کرد و سپس با استفاده از SVM، ضربان‌های طبیعی و غیرطبیعی را تفکیک کرد.

۵.۳ Random Forest و Decision Tree

درخت تصمیم ساختار تصمیم‌گیری قابل تفسیرتری نسبت به بسیاری از مدل‌های پیچیده دارد.

Random Forest با ایجاد مجموعه‌ای از درخت‌ها، محدودیت‌های یک درخت منفرد را کاهش می‌دهد.

مزایای آن شامل:

- امکان مدل‌سازی روابط غیرخطی،
- مقاومت نسبتاً مناسب در برابر نویز،
- قابلیت بررسی اهمیت ویژگی‌ها،
- و عدم نیاز به فرض خطی بودن رابطه میان ویژگی‌ها و کلاس

است.

۵.۴ Logistic Regression

اگر مسئله ساده‌تر باشد، Logistic Regression می‌تواند یک baseline مناسب ایجاد کند.

اهمیت این موضوع در پژوهش‌های یادگیری ماشین بسیار زیاد است. قبل از اینکه یک شبکه عمیق پیچیده طراحی شود، بهتر است عملکرد یک مدل پایه نیز اندازه‌گیری شود تا مشخص شود پیچیدگی مدل واقعاً چه ارزش افزوده‌ای ایجاد کرده است.

۶. نقش یادگیری عمیق در تشخیص ECG

یادگیری عمیق مسیر تحلیل ECG را تا حدی تغییر داده است. در این رویکرد، شبکه به جای اتکا کامل به ویژگی‌های طراحی‌شده توسط پژوهشگر، می‌تواند نمایش مناسب داده را از خود سیگنال یاد بگیرد.

۶.۱ شبکه‌های عصبی کانولوشنی

CNN فقط برای تصاویر کاربرد ندارد. ساختار کانولوشنی یک‌بعدی یا 1D-CNN می‌تواند برای تحلیل سیگنال‌های ECG بسیار مناسب باشد.

در یک CNN یک‌بعدی، فیلترهای کانولوشنی روی محور زمان حرکت می‌کنند و الگوهای محلی را استخراج می‌کنند.

در لایه‌های ابتدایی ممکن است الگوهای ساده‌تر و در لایه‌های عمیق‌تر الگوهای پیچیده‌تر یاد گرفته شوند.

به همین دلیل CNN می‌تواند برای تشخیص:

- ضربان‌های غیرطبیعی،
- آریتمی،
- تغییرات مورفولوژیک،
- و برخی الگوهای تشخیصی ECG

استفاده شود.

۶.۲ شبکه‌های بازگشتی و LSTM

ECG ماهیتی زمانی دارد و ترتیب نمونه‌ها اهمیت دارد.

شبکه‌های RNN برای داده‌های ترتیبی طراحی شده‌اند و LSTM یکی از معماری‌هایی است که برای مدیریت وابستگی‌های زمانی طولانی‌تر توسعه یافته است.

در تحلیل ECG، LSTM می‌تواند برای مدل‌سازی توالی ضربان‌ها یا بخش‌های طولانی‌تر سیگنال به کار رود.

برای مثال، ممکن است یک ضربان منفرد ظاهری نزدیک به حالت طبیعی داشته باشد، اما الگوی چندین ضربان متوالی اطلاعات بیشتری درباره آریتمی ارائه دهد. مدل‌های ترتیبی می‌توانند این وابستگی را بهتر از یک طبقه‌بند مستقل از ضربان در نظر بگیرند.

۶.۳ معماری‌های ترکیبی CNN-LSTM

ترکیب CNN و LSTM یکی از رویکردهای رایج است:

ECG → CNN → استخراج ویژگی محلی → LSTM → مدل‌سازی وابستگی زمانی → طبقه‌بندی

در این ساختار، CNN وظیفه استخراج الگوهای محلی را انجام می‌دهد و LSTM روابط زمانی میان این الگوها را مدل می‌کند.

مرورهای جدید حوزه تشخیص آریتمی نیز استفاده از CNN و معماری‌های ترکیبی CNN/RNN را از روندهای مهم این حوزه معرفی کرده‌اند، در حالی که چالش‌هایی مانند ناهمگونی داده‌ها، تفسیرپذیری و اجرای بلادرنگ همچنان مطرح هستند.

۷. تشخیص چه اختلالاتی با ECG و یادگیری ماشین

امکان‌پذیر است؟

کاربرد یادگیری ماشین به یک بیماری خاص محدود نیست.

۷.۱ آریتمی‌ها

یکی از مهم‌ترین کاربردها، طبقه‌بندی ضربان‌ها و ریتم‌های غیرطبیعی است.

داده‌های MIT-BIH Arrhythmia Database از منابع کلاسیک این حوزه محسوب می‌شوند. این مجموعه شامل ۴۸ قطعه ECG دوکاناله نیم‌ساعته است که از ۴۷ فرد به دست آمده و حدود ۱۱۰ هزار annotation ضربان برای آن ارائه شده است.

این دیتاست سال‌ها برای ارزیابی الگوریتم‌های تشخیص آریتمی مورد استفاده قرار گرفته است.

۷.۲ فیبریلاسیون دهلیزی

Atrial Fibrillation یا **AF** یکی از مهم‌ترین آریتمی‌هایی است که امکان تشخیص خودکار آن از ECG مورد بررسی قرار گرفته است.

در AF، نظم فعالیت دهلیزی و الگوی ضربان قلب تغییر می‌کند و همین ویژگی‌ها می‌توانند توسط الگوریتم‌های پردازش سیگنال و یادگیری ماشین شناسایی شوند.

MIT-BIH Atrial Fibrillation Database شامل ۲۵ ثبت ECG طولانی‌مدت از افراد مبتلا به AF است و برخی از ثبت‌ها عمدتاً شامل AF حمله‌ای هستند.

۷.۳ انفارکتوس میوکارد

ECG می‌تواند دارای تغییراتی مرتبط با انفارکتوس میوکارد باشد. مجموعه‌هایی مانند PTB-XL نیز شامل برچسب‌هایی مربوط به MI هستند و از این نظر برای توسعه مدل‌های طبقه‌بندی چندکلاسه یا چندبرچسبی قابل استفاده‌اند.

البته وجود یک برچسب MI در یک دیتاست پژوهشی به این معنا نیست که مدل می‌تواند به‌تنهایی جایگزین فرایند تشخیص بالینی شود.

۷.۴ اختلالات هدایت و تغییرات ST/T

تشخیص اختلالات هدایت، هیپرتروفی و تغییرات ST/T از دیگر حوزه‌های قابل مطالعه است.

در PTB-XL، علاوه بر حالت طبیعی، کلاس‌های سطح بالایی مانند **ST/T Change**، **Myocardial Infarction** و **Conduction Disturbance** و **Hypertrophy** تعریف شده‌اند.

۸. مقایسه روش‌های کلاسیک و یادگیری عمیق

یادگیری عمیق	روش‌های کلاسیک ML	معیار
کمتر	معمولاً زیاد	نیاز به استخراج ویژگی دستی
معمولاً بیشتر	معمولاً کمتر	نیاز به داده
متوسط تا بسیار بالا	پایین تا متوسط	پیچیدگی مدل
اغلب دشوارتر	معمولاً ساده‌تر	تفسیرپذیری
قابلیت یادگیری خودکار	وابسته به ویژگی‌های ورودی	توانایی یادگیری ویژگی
معمولاً بیشتر	معمولاً کمتر	هزینه محاسباتی آموزش
مناسب‌تر	محدودتر	مناسب برای سیگنال خام
قابل توجه	محدود	نیاز به تنظیم معماری
در داده محدود می‌تواند بالا باشد	وجود دارد	ریسک Overfitting

نباید نتیجه گرفت که یادگیری عمیق همیشه بهتر از الگوریتم‌های کلاسیک است. در یک پروژه واقعی، حجم داده، کیفیت برجسب‌ها، هدف بالینی، منابع محاسباتی و سطح تفسیرپذیری مورد نیاز باید تعیین کند که کدام روش مناسب‌تر است.

۹. مسئله بسیار مهم: عدم توازن کلاس‌ها

در داده‌های پزشکی، توزیع کلاس‌ها معمولاً متوازن نیست.

ممکن است تعداد ECG های طبیعی بسیار بیشتر از یک آریتمی نادر باشد. اگر مدل بدون توجه به این موضوع آموزش داده شود، می‌تواند با پیش‌بینی کلاس غالب به Accuracy ظاهراً بالایی برسد، بدون اینکه در تشخیص کلاس نادر عملکرد مناسبی داشته باشد.

برای مثال، فرض کنید ۹۵ درصد نمونه‌ها سالم و ۵ درصد دارای یک اختلال باشند. مدلی که تمام نمونه‌ها را سالم اعلام کند، Accuracy برابر ۹۵ درصد خواهد داشت، اما از نظر کاربرد تشخیصی تقریباً بی‌فایده است.

به همین دلیل باید معیارهایی مانند:

- Precision
- Recall یا Sensitivity
- Specificity
- F1-score
- ROC-AUC
- PR-AUC
- Confusion Matrix

نیز بررسی شوند.

در مسائل نامتوازن، **PR-AUC** و معیارهای مربوط به کلاس اقلیت ممکن است اطلاعات بیشتری نسبت به **Accuracy** فراهم کنند.

۱۰. جلوگیری از **Data Leakage**؛ مسئله‌ای حیاتی در پژوهش **ECG**

یکی از اشتباهات رایج در پروژه‌های یادگیری ماشین پزشکی، تقسیم تصادفی قطعه‌های **ECG** به مجموعه آموزش و آزمون بدون توجه به شناسه بیمار است.

فرض کنید از یک بیمار چندین **ECG** یا چندین قطعه از یک ثبت طولانی در اختیار داشته باشیم. اگر برخی قطعات در **Training** و برخی دیگر در **Test** قرار گیرند، مدل ممکن است ویژگی‌های اختصاصی همان بیمار را یاد بگیرد.

در چنین شرایطی، **Accuracy** بالا الزاماً نشان‌دهنده توانایی مدل برای تعمیم به بیماران جدید نیست.

بنابراین، در بسیاری از مطالعات **ECG** بهتر است تقسیم داده در سطح **Patient-level** انجام شود.

به عبارت دیگر:

→ **Patient A فقط Train**

→ **Patient B فقط Validation**

→ **Patient C فقط Test**

این موضوع برای ارزیابی واقعی سامانه‌های تشخیص خودکار اهمیت اساسی دارد.

۱۱. انتخاب دیتاست مناسب

انتخاب دیتاست باید متناسب با سؤال پژوهش باشد.

MIT-BIH Arrhythmia Database

برای پژوهش‌های مرتبط با تشخیص آریتمی بسیار شناخته‌شده است. این دیتاست شامل ۴۸ رکورد دوکاناله نیم‌ساعته با **annotation**‌های مرجع است.

PTB-XL

برای مطالعات گسترده تر ECG دوازده لید گزینه مهمی است. این مجموعه شامل بیش از ۲۱ هزار ECG بالینی ۱۲ لید و برچسب‌های متعدد تشخیصی است.

MIT-BIH Atrial Fibrillation Database

برای پژوهش‌های اختصاصی‌تر در زمینه AF کاربرد دارد و شامل ثبت‌های طولانی ECG است.

نکته مهم این است که بزرگ‌تر بودن دیتاست لزوماً به معنای بهتر بودن آن برای هر مسئله نیست. ساختار لیدها، مدت ثبت، کیفیت annotation، جمعیت بیمار، نوع بیماری و نحوه نمونه‌گیری باید با سؤال پژوهش سازگار باشند.

۱۲. تفسیرپذیری مدل؛ حلقه مفقوده بسیاری از سامانه‌های هوشمند

در مهندسی پزشکی، صرفاً رسیدن به یک عدد Accuracy بالا کافی نیست.

اگر مدل اعلام کند:

«این ECG با احتمال بالا متعلق به کلاس X است».

پژوهشگر و پزشک باید تا حد امکان بدانند چرا مدل به این نتیجه رسیده است.

در مدل‌های عمیق می‌توان از روش‌هایی مانند:

- Saliency Map
- Integrated Gradients
- Grad-CAM در معماری‌های مناسب
- Attention Visualization
- SHAP در برخی ساختارها

برای بررسی نواحی مؤثر بر تصمیم مدل استفاده کرد.

برای ECG، یک هدف مهم این است که مشخص شود آیا مدل روی بخش‌های فیزیولوژیک مرتبط با بیماری تمرکز کرده است یا خیر.

این مسئله فقط جنبه نظری ندارد. اگر مدل به دلیل یک آرتیفکت، تفاوت دستگاه ثبت یا ویژگی غیرمرتبط تصمیم بگیرد، عملکرد آن در محیط جدید ممکن است به شدت افت کند.

۱۳. چالش تعمیم پذیری مدل

یکی از مهم ترین فاصله ها میان مقاله پژوهشی و کاربرد بالینی همین موضوع است.

مدلی که روی یک دیتاست خاص عملکرد خوبی دارد، ممکن است روی داده های یک بیمارستان دیگر عملکرد مشابهی نداشته باشد.

دلایل احتمالی عبارت اند از:

- تفاوت دستگاه ECG
- تفاوت نرخ نمونه برداری
- تفاوت جمعیت بیماران
- تفاوت سن و جنس
- تفاوت شیوع بیماری
- تفاوت پروتکل ثبت
- تفاوت کیفیت سیگنال
- تفاوت روش annotation

انجمن قلب آمریکا نیز در بررسی کاربرد هوش مصنوعی در بیماری های قلبی بر ضرورت توجه به تفاوت سناریوهای بالینی، کیفیت Ground Truth و محدودیت شواهد حاصل از کاربردهای واقعی تأکید کرده است.

بنابراین **External Validation** باید بخشی مهم از مسیر توسعه یک سامانه جدی باشد.

۱۴. از Accuracy پژوهشی تا کاربرد بالینی

فرایند توسعه یک سیستم قابل اتکا را می توان در چند مرحله خلاصه کرد:

۱. تعریف دقیق مسئله بالینی
۲. انتخاب دیتاست مناسب
۳. بررسی کیفیت داده
۴. پیش پردازش
۵. تعریف روش برچسب گذاری
۶. تقسیم بیمار محور داده ها
۷. ایجاد baseline
۸. آموزش مدل
۹. تنظیم Hyperparameter ها
۱۰. ارزیابی روی Test Set مستقل

۱۱. اعتبارسنجی خارجی

۱۲. تحلیل خطا

۱۳. بررسی تفسیرپذیری

۱۴. بررسی عملکرد در شرایط واقعی

مرحله «تحلیل خطا» اهمیت زیادی دارد.

به جای اینکه فقط بگوییم مدل ۹۵ درصد Accuracy دارد، باید بررسی کنیم:

- کدام آریتمی‌ها بیشتر با هم اشتباه می‌شوند؟
- False Negative ها چه ویژگی‌هایی دارند؟
- آیا عملکرد مدل در سنین مختلف متفاوت است؟
- آیا نویز باعث خطای سیستم می‌شود؟
- آیا مدل روی داده یک دستگاه خاص بهتر عمل می‌کند؟

این نوع تحلیل اطلاعات بسیار بیشتری برای بهبود مدل فراهم می‌کند.

۱۵. یک معماری پیشنهادی برای پروژه پژوهشی

برای یک پایان‌نامه کارشناسی ارشد یا پروژه دکتری، می‌توان یک چارچوب پژوهشی نسبتاً کامل را به شکل زیر تعریف کرد:

داده ECG



کنترل کیفیت سیگنال



حذف نویز و Baseline Wander



Segmentation / Beat Detection



تقسیم Patient-level



Baseline کلاسیک



مدل یادگیری عمیق



Hyperparameter Tuning



ارزیابی چندمعیاره



Explainability



External Validation

در بخش مدل سازی نیز می توان سه سطح را با یکدیگر مقایسه کرد:

مدل اول: ویژگی مهندسی شده SVM +

این مدل می تواند Baseline پژوهش باشد.

مدل دوم: D-CNN

در این مرحله مدل مستقیماً با توالی ECG کار می کند.

مدل سوم: CNN-LSTM :

در این معماری، ویژگی های محلی و وابستگی زمانی به صورت ترکیبی بررسی می شوند.

چنین طراحی ای امکان یک مقایسه علمی معنادار را فراهم می کند و مشخص می سازد افزایش پیچیدگی مدل تا چه حد موجب بهبود عملکرد شده است.

۱۶. خطاهای رایج در طراحی پژوهش

استفاده بیش از حد از Accuracy

Accuracy به تنهایی معیار مناسبی برای بسیاری از مسائل پزشکی نیست.

تقسیم تصادفی قطعات ECG

اگر بیمار در هر دو مجموعه Train و Test حضور داشته باشد، احتمال Data Leakage وجود دارد.

استفاده از دیتاست کوچک و ادعای تعمیم گسترده

عملکرد روی یک دیتاست عمومی را نباید بدون اعتبارسنجی خارجی به کل جمعیت تعمیم داد.

نادیده گرفتن کیفیت برچسب‌ها

برچسب پزشکی همیشه حقیقت مطلق نیست. ممکن است بین متخصصان نیز اختلاف نظر وجود داشته باشد.

مقایسه ناعادلانه مدل‌ها

اگر دو مدل با پیش‌پردازش، تقسیم داده یا معیارهای متفاوت ارزیابی شوند، مقایسه عملکرد آنها معتبر نخواهد بود.

تمرکز صرف بر پیچیدگی مدل

یک شبکه بسیار عمیق لزوماً بهترین انتخاب نیست. اگر مدل ساده‌تر عملکرد مشابهی داشته باشد، از نظر محاسباتی و تفسیرپذیری ممکن است انتخاب عملی‌تری باشد.

۱۷. آینده یادگیری ماشین در تحلیل ECG

مسیر آینده این حوزه صرفاً به ساخت مدل‌های بزرگ‌تر محدود نمی‌شود.

یکی از جهت‌های مهم، یادگیری خودنظارتی (Self-Supervised Learning) است. در این روش می‌توان از حجم بزرگی از ECG های بدون برچسب برای یادگیری نمایش مناسب سیگنال استفاده کرد و سپس مدل را برای یک وظیفه مشخص Fine-tune کرد.

جهت مهم دیگر، استفاده از Multimodal AI است. در این رویکرد ECG می‌تواند همراه با اطلاعاتی مانند سن، علائم بالینی، فشار خون، آزمایش‌ها و سایر داده‌های بیمار تحلیل شود.

همچنین مدل‌های کوچک‌تر و بهینه‌شده می‌توانند امکان اجرای الگوریتم‌ها را روی تجهیزات قابل حمل و پوشیدنی افزایش دهند.

از سوی دیگر، **Federated Learning** می‌تواند در شرایطی که انتقال داده خام بیمارستان‌ها به یک مرکز مشترک دشوار است، مورد توجه قرار گیرد؛ هرچند این رویکرد نیز مسائل فنی، آماری و امنیتی خاص خود را دارد.

۱۸. نقش مهندس پزشکی در این فرایند

این حوزه نمونه خوبی از ماهیت میان‌رشته‌ای مهندسی پزشکی است.

یک پروژه موفق نیازمند ترکیبی از دانش‌های زیر است:

- فیزیولوژی قلب
- الکتروکاردیوگرافی
- پردازش سیگنال دیجیتال
- یادگیری ماشین
- یادگیری عمیق
- آمار و طراحی آزمایش
- برنامه‌نویسی
- دانش بالینی
- مدیریت داده‌های پزشکی
- ارزیابی سیستم‌های تشخیصی

مهندس پزشکی نباید صرفاً مدل را اجرا کند. مهم‌تر از اجرای یک الگوریتم آماده، توانایی تعریف مسئله درست، انتخاب داده مناسب، طراحی پروتکل ارزیابی و تفسیر نتیجه در زمینه فیزیولوژیک و بالینی است.

این تفاوت میان یک پروژه صرفاً محاسباتی و یک پژوهش واقعی مهندسی پزشکی است.

۱۹. جمع‌بندی

یادگیری ماشین و به‌ویژه یادگیری عمیق، ظرفیت قابل‌توجهی برای تحلیل خودکار ECG و پشتیبانی از تشخیص اختلالات قلبی ایجاد کرده‌اند. الگوریتم‌های کلاسیک مانند SVM، Random Forest و KNN در شرایطی که ویژگی‌های مناسبی در اختیار باشد همچنان کاربرد دارند؛ در مقابل، مدل‌هایی مانند ۱ D-CNN و CNN-LSTM امکان یادگیری الگوهای پیچیده‌تر از توالی ECG را فراهم می‌کنند.

با این حال، مسئله اصلی در پژوهش‌های پیشرفته ECG صرفاً دستیابی به بالاترین Accuracy نیست. کیفیت داده، Ground Truth، عدم توازن کلاس‌ها، جلوگیری از Data Leakage، تقسیم بیمارمحور، تفسیرپذیری و اعتبارسنجی خارجی عواملی هستند که ارزش واقعی یک مدل را مشخص می‌کنند.

برای دانشجوی کارشناسی ارشد یا دکتری مهندسی پزشکی، مسیر پژوهشی مناسب می‌تواند مقایسه یک مدل کلاسیک با یک یا چند معماری عمیق، همراه با تحلیل دقیق خطا و ارزیابی تعمیم‌پذیری باشد. دیتاست‌هایی مانند MIT-BIH و PTB-XL نیز امکان طراحی آزمایش‌های قابل تکرار و مقایسه‌ای را فراهم می‌کنند.

در نهایت، هدف یک سامانه هوشمند ECG نباید صرفاً «طبقه‌بندی سیگنال» باشد؛ بلکه باید بتواند در یک چارچوب معتبر و قابل تفسیر، اطلاعاتی قابل اتکا برای پشتیبانی از تصمیم‌گیری بالینی فراهم کند. فاصله میان یک مدل موفق آزمایشگاهی و یک ابزار قابل استفاده در پزشکی دقیقاً در همین مرحله مشخص می‌شود.

سوالات متداول

آیا یادگیری ماشین می‌تواند جایگزین پزشک در تفسیر ECG شود؟

خیر. مدل‌های یادگیری ماشین و یادگیری عمیق در درجه اول به‌عنوان ابزارهای تحلیل و پشتیبان تصمیم بررسی می‌شوند. عملکرد مدل به داده آموزشی، جمعیت مورد مطالعه، کیفیت برچسب‌ها و شرایط ثبت ECG وابسته است و نتیجه الگوریتم باید در چارچوب بالینی مناسب تفسیر شود.

برای شروع پژوهش ECG کدام دیتاست مناسب است؟

انتخاب دیتاست به سؤال پژوهش بستگی دارد. MIT-BIH Arrhythmia برای مطالعات آریتمی یکی از منابع کلاسیک است، در حالی که PTB-XL برای پژوهش‌های گسترده‌تر روی ECG دوازده‌لید و چندین گروه تشخیصی گزینه مهمی محسوب می‌شود.

آیا برای تحلیل ECG حتماً باید ویژگی‌ها را به‌صورت دستی استخراج کرد؟

خیر. در روش‌های کلاسیک معمولاً استخراج ویژگی اهمیت زیادی دارد، اما شبکه‌های عمیق می‌توانند بخشی از نمایش ویژگی را مستقیماً از سیگنال یاد بگیرند. با این حال، این موضوع به معنای بی‌اهمیت شدن پردازش سیگنال نیست.

CNN برای ECG بهتر است یا LSTM؟

هیچ پاسخ عمومی و مستقیمی وجود ندارد. CNN برای استخراج الگوهای محلی و مورفولوژیک بسیار مناسب است، در حالی که LSTM برای مدل‌سازی وابستگی‌های زمانی طراحی شده است. انتخاب باید بر اساس طول سیگنال، نوع مسئله و ساختار داده انجام شود. در برخی پروژه‌ها ترکیب CNN و LSTM منطقی است.

چرا Accuracy بالا به تنهایی کافی نیست؟

زیرا داده‌های پزشکی ممکن است نامتوازن باشند. در یک دیتاست با کلاس غالب، حتی مدلی که کلاس اقلیت را به‌خوبی تشخیص نمی‌دهد می‌تواند Accuracy بالایی داشته باشد. بنابراین Sensitivity، Specificity، F1-score، ROC-AUC و در شرایط مناسب PR-AUC نیز باید بررسی شوند.

Data Leakage در پروژه‌های ECG چیست؟

Data Leakage زمانی رخ می‌دهد که اطلاعاتی از داده آزمون، مستقیم یا غیرمستقیم، در فرایند آموزش وارد شود. در ECG یکی از نمونه‌های مهم آن قرار گرفتن قطعات مختلف مربوط به یک بیمار در مجموعه‌های آموزش و آزمون است.

آیا مدل آموزش‌دیده روی MIT-BIH را می‌توان مستقیماً در بیمارستان استفاده کرد؟

خیر، چنین نتیجه‌ای بدون اعتبارسنجی مستقل قابل دفاع نیست. تفاوت جمعیت، دستگاه ثبت، کیفیت سیگنال، پروتکل‌ها و شیوع بیماری می‌تواند عملکرد مدل را تغییر دهد. اعتبارسنجی خارجی یکی از مراحل مهم پیش از کاربرد عملی است.

مهم‌ترین چالش آینده این حوزه چیست؟

یکی از چالش‌های اصلی تبدیل مدل‌های موفق پژوهشی به سامانه‌هایی است که در جمعیت‌ها و محیط‌های متفاوت عملکرد پایدار، قابل تفسیر و قابل اعتماد داشته باشند. کیفیت Ground Truth، تعمیم‌پذیری، تفسیرپذیری، حریم خصوصی داده و ارزیابی بالینی از مسائل مهم این مسیر هستند.