

# پروژه درس هوش مصنوعی و محاسبات زیستی

استاد: جناب آقای دکتر رضایی

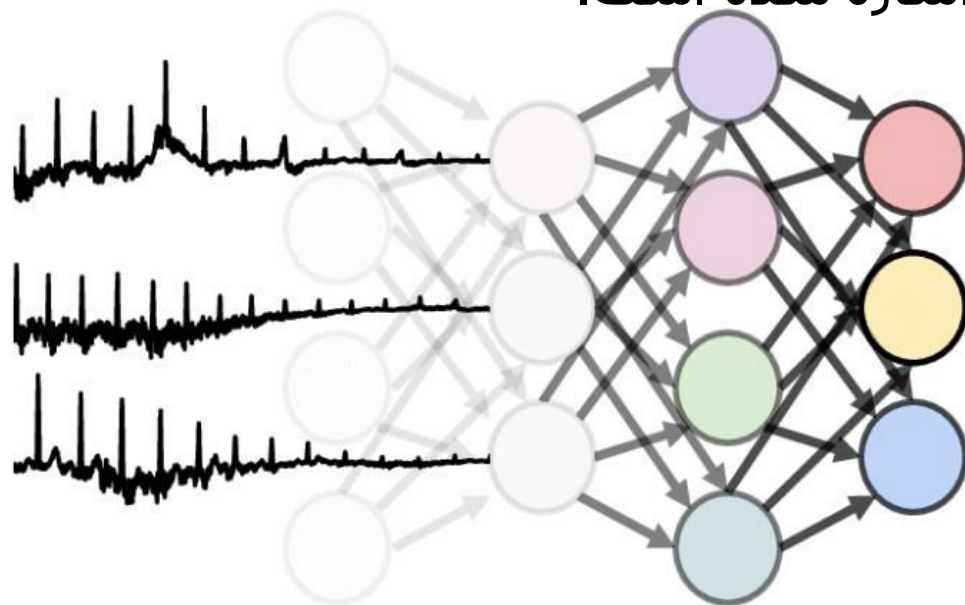
# فهرست

- مقدمه
- هدف
- معرفی روش ثبت سیگنال های قلبی
- دیتاست
- از شما چه انتظاری می رود؟
- تحویل پروژه

# مقدمه

وبسایت فیزیونت در سال 1999 با هدف انجام پروژه های مهندسی پزشکی ایجاد شده است و طی این سال ها با سازمان های مختلفی در ارتباط بوده است.

این وبسایت سعی میکند مشکلاتی که پزشکان در دنیای پزشکی با آن در ارتباط هستند و میتواند توسط مهندسان حل شود، را تحت عنوان چالش در این سایت قرار دهد و از علاقه مندان دعوت میکند تا به حل آن مشکلات پردازند و هر ساله نیز مقالات معتبری در این زمینه چاپ میگردد. همچنین برندگان این چالش از جوایز ویژه ای نیز بهره مند میگردند که در خود وبسایت به آن اشاره شده است.

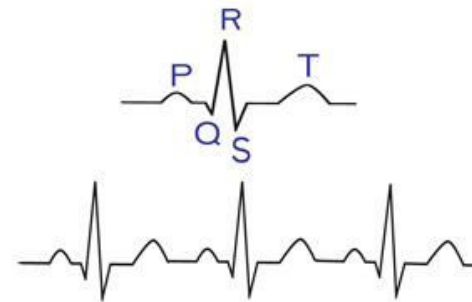


# هدف

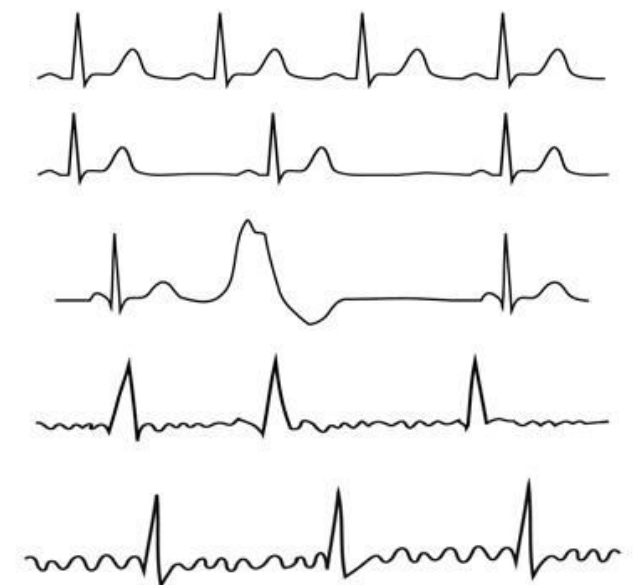
در سایت فیزیونت امسال چالشی برای سال 2020 قرار گذاشته شده است که از داوطلبان خواسته شده است تا با پردازش داده های قلبی افراد بتوانند 8 نوع بیمار قلبی را از یکدیگر و افراد سالم تشخیص دهند. بیماری ها شامل:

- AF - Atrial fibrillation
- I-AVB - First-degree atrioventricular block
- LBBB - Left bundle branch block
- Normal - Normal sinus rhythm
- PAC - Premature atrial complex
- PVC - Premature ventricular complex
- RBBB - Right bundle branch block
- STD - ST-segment depression
- STE - ST-segment elevation

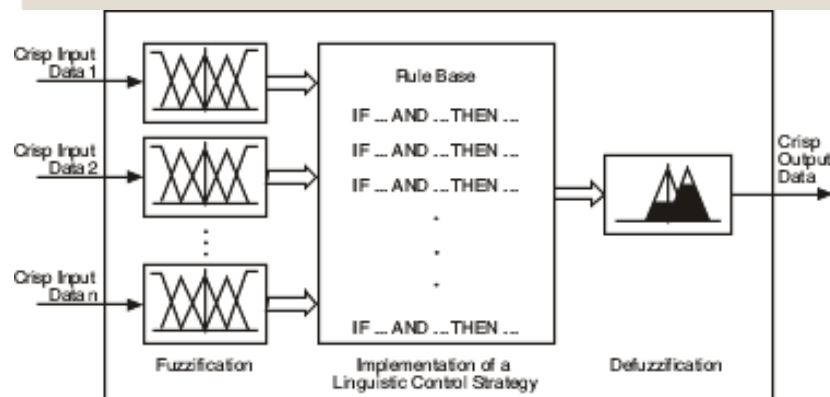
**Normal heart rhythm**



**Irregular heart rhythm**



در راستای رسیدن به این هدف ما قصد داریم که در طی چهار فاز این پروژه ویژگی‌هایی از این داده‌ها استخراج کنیم تا بتوانیم به صورت بهینه داده‌ها را دسته‌بندی کنیم. به همین منظور باید اطلاعاتی را به نوع ثبت سیگنال‌های قلبی و ویژگی‌های بهینه برای استخراج از سیگنال‌های قلبی را بدست آوریم.



## فاز اول:

- 1- استخراج ویژگی‌های سیگنال‌های قلبی
- 2- طبقه‌بندی سیگنال‌های قلبی با سیستم‌های فازی

## فاز دوم:

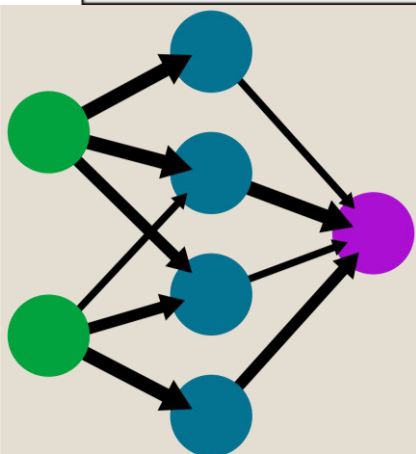
طبقه‌بندی سیگنال‌های قلبی با استفاده از شبکه‌های عصبی

## فاز سوم:

استخراج ویژگی‌های بهینه توسط الگوریتم‌های تکاملی

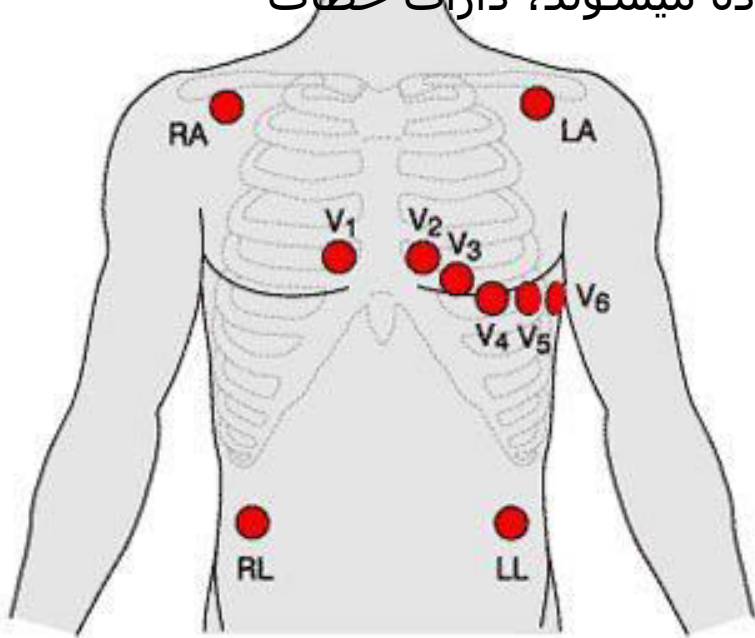
## فاز چهارم:

پیاده‌سازی الگوریتم با استفاده از سیستم‌های هابیرید



# معرفی روش ثبت سیگنال های قلبی

ثبت سیگنال های قلبی یا ECG یک روش غیر تهاجمی برای اندازه گیری سیگنال های قلبی و تشخیص بیماری ها و آریتمی های قلبی از روی شکل سیگنال هاست. معروف ترین روش ثبت این سیگنال ها روش 12 الکترودی است که با استفاده از این تعداد الکتروود میتوان با تقریب خوبی تمام سیگنال قلبی را بدست آورد همچنین برخی از بیماری ها بر روی بعضی کانال ها قابل مشاهده تر هستند و نکته ای که وجود دارد این است که پزشک برای تشخیص بیماری لازم دارد تا تمام کانال ها را بررسی کند که کاری بسیار وقت گیر است و روش هایی نیز که اکنون برای تشخیص این نوع بیماری ها استفاده میشوند، دارای خطای زیادی هستند.



# دیتاست

دیتاستی که در اختیار قرار داریم یک دیتاست سیگنال های قلبی است که در سال 2018 در یک کنفرانس چینی از آن استفاده شده است. افراد شرکت کننده در این آزمایش 6877 نفر بوده اند که 3699 نفر آن ها آقا و 3178 باقی مانده خانم هستند. فرکانس نمونه برداری 500 هرتز و زمان نمونه برداری از 6 ثانیه تا 60 ثانیه متغیر است.

این دیتاستی که در اختیار ما قرار دارد در واقع دیتاست train نامیده میشود چرا که ما علاوه بر سیگنال ها نوع بیماری فرد (از 8 دسته بیماری داده شده) را در اختیار داریم. اما هدف از انجام این چالش پیدا کردن بیماری در افرادی است که تنها سیگنال قلبی آن ها را در اختیار داریم. اما چون این چالش یک مسابقه نیز هست، داده های تست در اختیار ما قرار ندارد.

دیتاست داده شده شامل 6877 فایل با فرمت mat. شامل سیگنال های گرفته شده از هر فرد جهت استفاده در نرم افزار متلب و 6877 فایل با فرمت hea. برای خواندن اطلاعات مربوط به هر بیمار از جمله جنسیت، سن، بیماری، تاریخ ثبت سیگنال و اطلاعات مربوط به هر لید می باشد. این فایل را میتوانید با notepad باز کنید. نمونه ای از این فایل را در زیر مشاهده میکنید:

```
A0001 12 500 7500 05-Feb-2020 11:39:16
A0001.mat 16+24 1000/mV 16 0 28 -1716 0 I
A0001.mat 16+24 1000/mV 16 0 7 2029 0 II
A0001.mat 16+24 1000/mV 16 0 -21 3745 0 III
A0001.mat 16+24 1000/mV 16 0 -17 3680 0 aVR
A0001.mat 16+24 1000/mV 16 0 24 -2664 0 aVL
A0001.mat 16+24 1000/mV 16 0 -7 -1499 0 aVF
A0001.mat 16+24 1000/mV 16 0 -290 390 0 V1
A0001.mat 16+24 1000/mV 16 0 -204 157 0 V2
A0001.mat 16+24 1000/mV 16 0 -96 -2555 0 V3
A0001.mat 16+24 1000/mV 16 0 -112 49 0 V4
A0001.mat 16+24 1000/mV 16 0 -596 -321 0 V5
A0001.mat 16+24 1000/mV 16 0 -16 -3112 0 V6
#Age: 74
#Sex: Male
#Dx: RBBB
#Rx: Unknown
#Hx: Unknown
#Sx: Unknown
```



اطلاعاتی که از این فایل استخراج میشود به صورت زیر است:

A0001 نشان دهنده ی شماره ی داده است که این شماره بین 1 تا 6877 متغیر است. 12 تعداد لید ها و 500 فرکانس نمونه برداری است و 7500 تعداد sample هاست. و در ادامه تاریخ و ساعت ثبت شده است. در 12 خط بعدی اطلاعات لید ها و در سمت راست آن ها نام لید نوشته شده است. و در انتها سن بیمار، جنسیت و بیماری اون نیز ثبت شده است.  
برای اطلاعات بیشتر به لینک زیر مراجعه کنید.

<https://physionetchallenges.github.io/2020/>

# از شما چه انتظاری می‌رود؟

از آنجایی که دیتاست این چالش دیتاست سنگینی است، برای سرعت بخشیدن به کار، همه ی دیتاست در اختیار شما قرار نگرفته است و **هر کدام از شما قرار است روی بخشی از دیتاست کار کنید** که این بخش برای شما انتخاب شده است و فایل آن در اختیارتان قرار خواهد گرفت.

همچنین یک فایل به شما داده خواهد شد که در آن تعدادی ویژگی که قابل استخراج از سیگنال های قلبی میباشند به شما داده شده است که برای مراحل ابتدایی پروژه مناسب است اما کافی نیست و فقط جهت ایده دادن به شما آورده شده است. از شما انتظار می رود که ابتدا با مطالعه ی منابعی که در انتهای این فایل آورده شده است و سایر منابعی که میتوانید از اینترنت پیدا کنید، تعدادی ویژگی علاوه بر ویژگی های داده شده استخراج کنید و در گزارش خود بنویسید. معمولا 64 ویژگی برای سیگنال های قلبی استخراج میشود. (اجباری بر روی تعداد نداریم اما هرچه بیشتر، بهتر. ضمن این که اگر لازم داشتید که ویژگی ها را بعدا تغییر دهید مجازید).

در ادامه شما می توانید با استفاده از تولباکس های **anfis** و **fuzzy logic** متلب این دیتاست را آموزش دهید. اما از آنجایی که داده های تست در اختیار نداریم برای آموزش این داده ها باید از روش 5-fold cross-validation استفاده کنید. (برای جلوگیری از overfitting).

فرایند overfitting است که در طی آن شما از یک سری ویژگی منحصر به فرد این دیتاست برای آموزش و train کردن استفاده کنید که قابل تعمیم به سایر دیتاست ها نیست.

اما جهت ساده کردن کار آموزش برای شما ابتدا به جای آن که داده هارا به 9 دسته طبقه بندی کنید، ابتدا به دو دسته طبقه بندی کنید و افراد دارای برچسب normal را از سایر افراد بیمار جدا کنید. در ادامه ی پروژه شما باید برچسب های افراد بیمار را نیز از هم تفکیک کنید و به 8 دسته تقسیم کنید.

در هر مرحله گزارشی که باید بنویسید باید علاوه بر ویژگی های استفاده شده و توضیح روش استفاده شده، شامل درصد صحت طبقه بندی نیز باشد. روش اندازه گیری درصد صحت به این صورت است که شما باید برچسب هایی که برای داده های تست بدست آورده اید را با برچسب هایی که در اختیار دارید مقایسه کنید و نسبت تعداد برچسب های صحیح داده های تست به کل را بدست آورید و درصد بگیرید و گزارش کنید. از آنجایی که روش 5-fold را استفاده میکنید، 5 تا درصد باید گزارش کنید و در انتها میانگین این درصدها را نیز گزارش کنید.

+ برای اینکه حالت رقابتی حفظ شود، به فردی که درصد صحت بیشتری کسب کند امتیاز ویژه تعلق میگیرد.  
++ از آنجایی که تولباکس های متلب جهت استفاده به شما معرفی شدند، برای استفاده از سایر موتورهای استنتاج فازی (غیر از mamdani و TSK) که به صورت دستی کد شوند یا دارای خلاقیت باشند، امتیاز در نظر گرفته شده است.

!! توجه کنید که استفاده از نرم افزار متلب اجباری نیست و میتوانید با استفاده از پایتون و R نیز این پروژه را انجام دهید.

مراحلی که در بالا توضیح داده شده اند همگی فاز 1 پروژه هستند که مهمترین بخش پروژه به حساب می آیند. بنابراین سعی کنید با دقت و حوصله ی کافی این فاز را انجام دهید تا فازهای بعدی برایتان سبک تر شود.

# تحويل پروژه

## زمان:


موعد تحويل پروژه از روز اپلود فایل پروژه و دیتاست ها به مدت **21 روز** خواهد بود و با توجه به تعطیلات، شما فرصت کافی برای انجام آن خواهید داشت. اما دوستانی که قصد شرکت در مسابقه را دارند، مهلت ثبت نام و ارسال چکیده کار و بخش ابتدایی کار برای فیزیونت تاریخ **15 آپریل** است. لذا اگر قصد شرکت در چالش را دارید این تاریخ را مد نظر قرار دهید.

## نوع تحويل:

فایل تحویلی شما باید شامل دیتاست اختصاص داده شده به شما، گزارش پروژه و کدهایتان باشد.

## فرمت تحويل:

فایل های خود را به صورت زیپ با نام **CI\_Ph1\_Student#** بر روی **cw** آپلود کنید و از ایمیل کردن آن به استاد و تی ای ها خودداری کنید.



فاز دوم

# در این فاز از شما چه انتظاری می رود؟

- (1) با استفاده از ویژگی هایی که در فاز قبلی به دست آوردید، این بار می خواهیم به جای یک سیستم فازی یک شبکه ی عصبی پیاده سازی کنیم تا بتوانیم با استفاده از آن label های داده های تست را تشخیص دهیم. به همین جهت با استفاده از تولباکس nntool متلب یا هر تولباکس دیگری که صلاح میدانید، یک شبکه عصبی multilayer perceptron طراحی کنید و درصد دقت شبکه ی طراحی شده را به دست آورید. (مشابه فاز قبل با استفاده از روش 5-fold) در این مرحله تعداد لایه های میانی را آنقدر تغییر دهید تا به بهترین دقت برسید. دقت های به دست آمده در هر مرحله را در گزارش خود بنویسید. حال بهترین دقتی که به دست آورده اید را با فاز قبلی مقایسه کرده و در گزارش خود بنویسید که کدام یک از شبکه های فازی یا عصبی در تقسیم بندی داده ها بهتر عمل میکند.
- (2) حال مراحل بالا را با استفاده از یک شبکه ی عصبی rbf انجام دهید. اینبار علاوه بر این که دقت شبکه ی عصبی rbf را با شبکه ی فازی مقایسه میکنید، با شبکه ی عصبی multilayer perceptron بخش (1) نیز مقایسه کنید.

# در این فاز از شما چه انتظاری می رود؟

- (3) در فاز قبلی تعدادی ویژگی بدست آورده بودید که به علت محدودیت حافظی کامپیوتر نتوانستید از همه ی آن ها در طراحی شبکه فازی استفاده کنید و در بخش قبلی این فاز نیز از همان ویژگی های محدود استفاده کردید تا بتوانید داده ها را با فاز قبل مقایسه کنید. به همین دلیل در این بخش فاز دوم از شما میخواهیم که با استفاده از ویژگی هایی که در فاز اول از آن ها استفاده نکردید و فقط آن ها را شناسایی کرده بودید ( یا حتی شناسایی نکردید و به تازگی به ذهنتان رسیده است) شبکه ی عصبی multilayer طراحی کنید. در این بخش نیز تعداد لایه ها را آنقدر تغییر دهید که بهترین صحت را بگیرید. (حتما از روش 5-fold استفاده کنید چون در شبکه عصبی احتمال وقوع overfitting خیلی زیاد است). حال شبکه ی به دست آمده را با شبکه ی بخش قبل (1) مقایسه کنید.
- (4) اکنون یک شبکه ی rbf با ویژگی هایی که در بخش قبل استفاده کردید طراحی کنید و با شبکه ی rbf قبلی (2) و شبکه های multilayer perceptron قبلی (1و3) نیز مقایسه کنید.
- نکته: برای پیاده سازی این فاز از هر زبان برنامه نویسی (متلب، پایتون، R و...) میتوانید استفاده کنید.

# بخش امتیازی

- در این فاز نیز مانند فاز قبل برای حفظ شدن خاصیت رقابتی پروژه به کسی که بیشترین درصد صحت را داشته باشد، نمره امتیازی تعلق خواهد گرفت
- اگر یک شبکه ی عصبی عمیق به صورت آنلاین پیاده سازی کنید، نمره امتیازی خواهد داشت.
- اگر یک شبکه ی عصبی multilayer یا rbf را خودتان پیاده کنید ( بدون استفاده از تولباکس متلب یا توابع پایتون) که به صورت خلاقانه پیاده شده باشد و از تولباکس های آماده درصد صحت بیشتری داشته باشد، نمره امتیازی خواهد داشت.



# تحويل پروژه

## زمان:

موعد تحويل پروژه روز **4 شنبه 31 اردیبهشت ماه ساعت 23:55** خواهد بود و با توجه به تعطیلات، شما فرصت کافی برای انجام آن خواهید داشت.

## نوع تحويل:

فایل تحويلی شما باید شامل دیتاست اختصاص داده شده به شما، گزارش پروژه و کدهایتان باشد.

## فرمت تحويل:

فایل های خود را به صورت زیپ با نام **CI\_Ph2\_Student#** بر روی **cw** آپلود کنید و از ایمیل کردن آن به استاد و تی ای ها خودداری کنید.