

# سامانه پیش‌بینی هیپوگلیسمی مبتنی بر تحلیل سیگنال ECG با استفاده از یادگیری ماشین

غزل عطاری<sup>1</sup>

1: دانش آموز پایه دوازدهم دبیرستان فرزنانگان 4

## چکیده

اختلالات تنظیم قند خون از جمله هیپوگلیسمی یکی از پیامدهای خطرناک دیابت است که در صورت تشخیص دیرنگام می‌تواند موجب افت هوشیاری و مرگ شود. سامانه‌های مرسوم پایش قند خون مبتنی بر سنسورهای تهاجمی (CGM) یا تست‌های خون هستند که هزینه‌بر و برای برخی گروه‌های جمعیتی غیرقابل‌دسترس‌اند. نیاز به راهکاری غیرتهاجمی، کم‌هزینه و قابل‌استفاده در مقیاس وسیع انگیزه اصلی این پژوهش می‌باشد.

این تحقیق یک چارچوب کامل را معرفی می‌کند که با تحلیل سیگنال ECG به‌طور ویژه شاخص‌های HRV و مورفولوژی موج QRS و به‌کارگیری مدل‌های یادگیری ماشین، اقدام به پیش‌بینی احتمال وقوع هیپوگلیسمی در بازه زمانی کوتاه پیشرو می‌نماید. فرآیند تحقیق شامل تولید داده‌های مصنوعی کنترل‌شده، استخراج ویژگی‌های زمان-فرکانس و مورفولوژی، آموزش و ارزیابی مدل و نهایتاً توسعه یک رابط وب تعاملی برای نمایش پیش‌بینی‌ها است.

انتظار می‌رود مدل پیشنهادی روی داده‌های شبیه‌سازی شده دقت عملیاتی مناسبی (بالاتر از ۸۵٪ با معیارهای precision/recall متوازن) نشان دهد و به‌عنوان یک نمونه اولیه قابل بازتولید برای توسعه‌های آتی با داده بالینی سرویس‌دهی کند. کاربردهای واقعی این سامانه شامل ادغام در پلتفرم‌های تله‌مدیسین، اپلیکیشن‌های موبایل و دستگاه‌های پوشیدنی پزشکی است.

کلمات کلیدی: هیپوگلیسمی، دیابت، یادگیری ماشین، ECG، HRV.

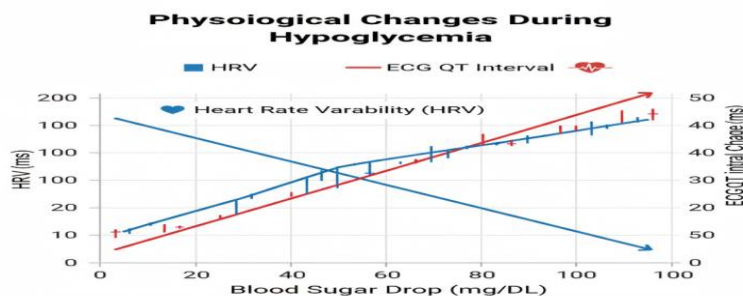
## مقدمه

دیابت نوعی بیماری مزمن با شیوع جهانی بالاست و مدیریت مستمر قند خون بخش مهمی از مراقبت‌های طولانی‌مدت بیماران است. یکی از چالش‌های اصلی مراقبت از افراد دیابتی، شناسایی به‌موقع هیپوگلیسمی است که نه تنها کیفیت زندگی را کاهش می‌دهد، بلکه می‌تواند به عوارض کوتاه‌مدت و بلندمدت جدی منجر شود. از این‌رو، توسعه ابزارهای پیش‌بینی‌کننده و هشداردهنده اهمیت بالایی دارد.

سنسورهای CGM و روش‌های تهاجمی دیگر، هرچند از نظر دقت اندازه‌گیری قند خون مفید هستند، اما هزینه بالا، نیاز به نصب تخصصی و محدودیت‌های نگهداری موجب می‌شود بسیاری از بیماران امکان یا تمایل به استفاده مستمر از آن‌ها را نداشته باشند. این محدودیت‌ها به ویژه در مناطق کم‌درآمد و بین گروه‌های آسیب‌پذیر آشکارتر است.

مطالعات بالینی و آزمایشگاهی نشان داده‌اند که تغییرات متابولیک مانند هیپوگلیسمی بر فعالیت سیستم عصبی خودکار تأثیر گذاشته و شاخص‌های HRV و پارامترهای ECG را دچار تغییر می‌کند. این شواهد علمی، امکان استخراج ویژگی‌های مرتبط از ECG و کاربرد یادگیری ماشین برای طبقه‌بندی حالات متابولیک را پشتیبانی می‌کند.

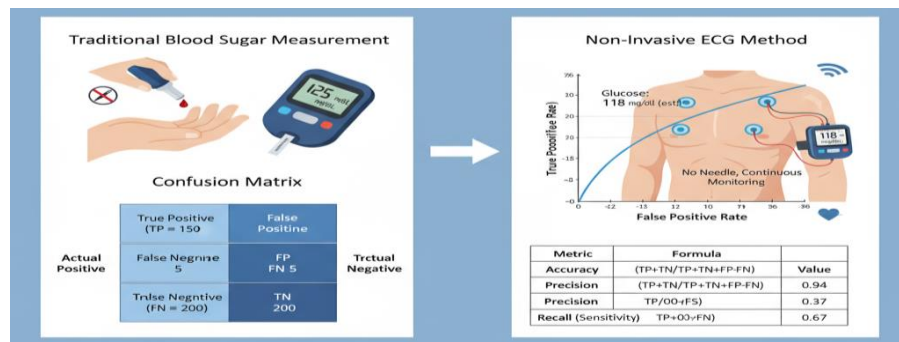
پیشنهاد این پروژه ساخت یک سامانه قابل بازتولید و قابل گسترش است که بدون استفاده از حسگر گلوکز، با تحلیل جریان داده ECG و مدل‌های یادگیری ماشین، هشدار زودهنگام هیپوگلیسمی را فراهم کند. هدف پژوهش دستیابی به چارچوبی است که هم از منظر فنی قابل‌اعتماد و هم از منظر پیاده‌سازی در محیط‌های واقعی اقتصادی باشد.



نمودار 1: ارتباط بین کاهش قند خون و تغییرات HRV/ECG

## بیان مسئله

در بسیاری از سناریوهای بالینی و غیربالینی، تشخیص به‌موقع هیپوگلیسمی دشوار است؛ زیرا بیمار ممکن است قبل از تشخیص توسط سیستم کنترل، علائم را تجربه نکرده یا قادر به اعلان وضعیت نباشد. این مشکل در میان کودکان، سالمندان و افراد دارای اختلال بینایی یا ذهنی برجسته‌تر است. به علاوه، هزینه نصب و نگهداری CGM برای کاربران کم‌درآمد یک مانع بزرگ در دسترسی جهانی به پایش مستمر است. یک راهکار مطلوب باید غیرتهاجمی، دارای مصرف انرژی پایین، قابل تعویض با دستگاه‌های موجود (مانند ساعت‌های هوشمند یا پیچ‌های ECG) و دارای زمان پاسخ سریع باشد. از لحاظ الگوریتمی، مدل باید بتواند با داده‌های نویزی و تغییرپذیر حاصل از شرایط فیزیولوژیک گوناگون کار کند و همزمان خطر **overfitting** و حساسیت به تغییرات محیطی را کاهش دهد.



تصویر 1: تفاوت بین روش سنتی اندازه‌گیری قند خون و روش غیرتهاجمی ECG

## اهداف پژوهش

هدف کلی: توسعه و ارزیابی سامانه‌ای نرم‌افزاری-تحقیقاتی برای پیش‌بینی زودهنگام هیپوگلیسمی با استفاده از سیگنال ECG و الگوریتم‌های یادگیری ماشین.

## اهداف جزئی

1. طراحی و پیاده‌سازی ماژولی برای تولید داده ECG مصنوعی که حالات نرمال و هیپوگلیسمی را شبیه‌سازی کند.
2. استخراج ویژگی‌های زمانی، فرکانسی و مورفولوژیکی از سیگنال‌های ECG با روش‌های استاندارد پردازش سیگنال.
3. طراحی و آموزش مدل‌های طبقه‌بندی (شامل Random Forest و مقایسه با SVM و Logistic Regression) و انتخاب مدل بهینه بر اساس معیارهای آماری.
4. توسعه یک رابط وب تعاملی (Streamlit) برای دریافت ورودی ECG و نمایش پیش‌بینی در زمان واقعی.
5. مستندسازی فرآیند برای بازتولیدپذیری و آماده‌سازی بستر آزمایشی برای گام‌های بعدی با داده بالینی.

## روش تحقیق

روش تحقیق یک زنجیره مهندسی داده-مدل-استقرار است که شامل تولید/جمع‌آوری داده، پیش‌پردازش، استخراج ویژگی، ساخت مدل، ارزیابی و پیاده‌سازی رابط کاربری می‌شود. در ادامه هر مرحله به تفصیل آمده است.

## محیط و ابزارها

زبان: Python 3.9

محیط توسعه/آزمایش: Google Colab (برای تسهیل دسترسی و توان محاسباتی)

کتابخانه‌ها: NumPy, Pandas, SciPy, scikit-learn, joblib, matplotlib, Streamlit

کنترل نسخه: Git و GitHub

فایل‌های اصلی: generate\_data.py, ecg\_utils.py, train\_model.py, app.py, colab\_demo.ipynb

## تولید داده مصنوعی

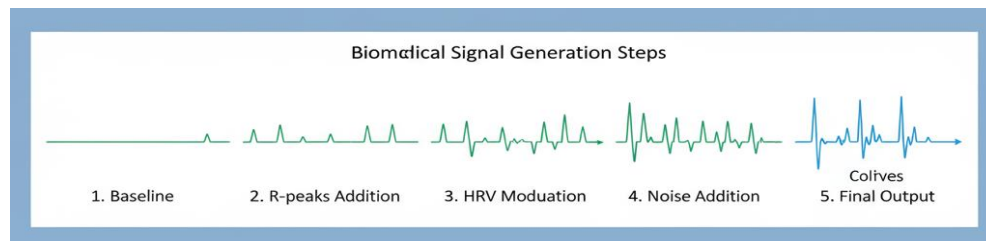
1. تعریف پارامترهای پایه: نرخ نمونه‌برداری ( $f_s = 250, \text{Hz}$ )، طول سیگنال ( $T = 60, \text{sec}$ )، تعداد نمونه‌ها ( $N$ ) (به عنوان مثال 1000 برای فاز توسعه، قابل افزایش).
2. تولید پایه‌ای موج ECG با ترکیب سینوسی‌های هارمونیک برای مدل‌سازی موج P-QRS-T: برای هر ضربان، موج R با یک تابع گاوسی (Gaussian) یا تابع منفی‌شده اضافه می‌شود تا قله R شبیه‌سازی شود.

3. افزودن نوسانات: HRV فواصل RR براساس توزیع نرمال با میانگین ( $\mu = 60/\text{HR}$ ) و واریانس متناسب با سطح HRV نمونه تولید می‌شوند.

4. افزودن پارامترهای ویژه حالت هیپوگلیسمی: کاهش HRV (کاهش دامنه نوسانات)، افزایش/کاهش جزئی در پهنای QRS یا کشیدگی QT، و تغییر در دامنه موج T.

5. افزودن نویز سفید و نویز با طیف مشخص (برای شبیه‌سازی نویزهای دستگاه و حرکت) تا مدل نسبت به نویز مقاوم باشد.

6. هر نمونه به‌صورت یک فایل CSV با دو ستون time و value ذخیره می‌شود؛ متادیتای مربوط (file, label) در data\_samples/labels.csv نگهداری می‌شود.



تصویر 2: مراحل تولید سیگنال

### پیش‌پردازش سیگنال

1. فیلتر پایین‌گذر و بالاگذر (bandpass) برای حذف نویز DC و فرکانس‌های خیلی بالا؛ به‌عنوان مثال فیلتر Butterworth باند 0.5–40 Hz

2. نرمال‌سازی دامنه (z-score یا min-max) برای مقایسه بهتر بین نمونه‌ها.

3. تشخیص قله‌های R (R-peak detection): از الگوریتم Pan–Tompkins یا نسخه ساده‌شده بر پایه مشتق و آستانه‌گذاری استفاده می‌شود.

4. حذف بخش‌های پرت و نمونه‌های خراب (artifact rejection) با استفاده از معیارهایی مانند تعداد قله‌های غیرمعمول در بازه مشخص.

### استخراج ویژگی‌ها (feature engineering)

برای هر نمونه، مجموعه ویژگی‌های زیر استخراج می‌شود:

الف) ویژگی‌های زمانی (Time-domain):

- میانگین RR (mean RR)
- SDNN (standard deviation of NN intervals)
- RMSSD (root mean square of successive differences)
- pNN50 (درصد اختلافات متوالی بیش از 50 ms)

ب) ویژگی‌های فرکانسی (Frequency-domain):

طیف توان با استفاده از Welch's method

LF (low frequency power: 0.04–0.15 Hz) و HF (high frequency power: 0.15–0.4 Hz) و نسبت LF/HF

ج) ویژگی‌های مورفولوژیک (Morphological):

- پهنای QRS (ms)

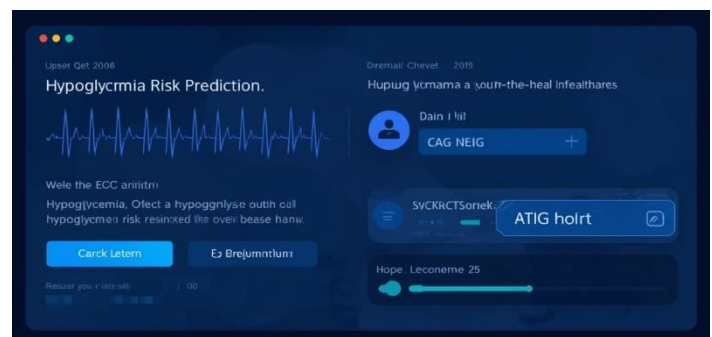
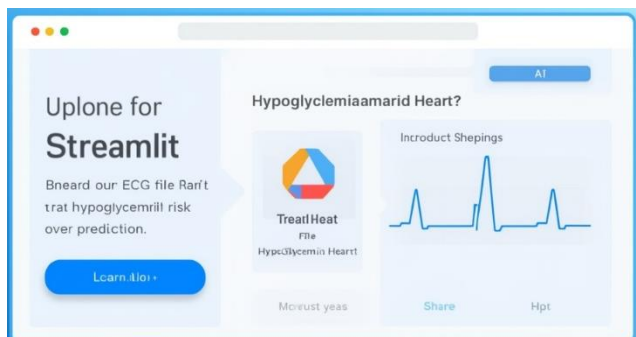
- ارتفاع موج  $T$  و  $R$
- شیب‌های صعودی/نزولی در قسمت‌های مشخص موج
- (د) ویژگی‌های آماری و زمان-فرکانس:
- انرژی سیگنال، (Approximate entropy یا Sample entropy)
- ویژگی‌های کوتاه‌مدت با STFT یا Continuous Wavelet Transform
- خروجی: ماتریس ویژگی  $(X)$  به شکل  $(N \times p)$  که  $(N)$  تعداد نمونه‌ها و  $(p)$  تعداد ویژگی‌ها می‌باشد. برچسب‌ها  $(y)$  عددی ( $=0$  نرمال،  $=1$  هیپوگلیسمی).

### ساخت مدل و آموزش

1. تقسیم داده stratified split: با نسبت 80٪ آموزش و 20٪ آزمون.
  2. نرمال‌سازی/استانداردسازی ویژگی‌ها (StandardScaler) براساس مجموعه آموزش و اعمال آن روی مجموعه آزمون.
  3. انتخاب مدل‌ها: Random Forest (پیشنهادی)، SVM، Logistic Regression.
  4. تنظیم ابرپارامترها: GridSearchCV یا RandomizedSearchCV با cross-validation ( $k=5$  یا 10). مثال پارامترها برای
- RF:  $n\_estimators \in \{100, 200, 500\}$ ,  $max\_depth \in \{None, 10, 20, 30\}$ ,  $min\_samples\_split \in \{2, 5, 10\}$ .
5. آموزش مدل نهایی روی داده آموزش با پارامترهای بهینه و ذخیره مدل با joblib به صورت model.pkl.
  6. ارزیابی: محاسبه AUC، precision، ROC curve، confusion matrix، F1-score، recall. همچنین محاسبه حساسیت و اختصاصی بودن (sensitivity and specificity) مهم است.

### اعتبارسنجی داخلی و آزمون پایداری

1. انجام cross-validation نشکن (nested CV) برای برآورد غیرمغرضانه از عملکرد.
  2. تست پایداری نسبت به نویز: اضافه کردن سطوح مختلف نویز به نمونه‌ها و ارزیابی افت عملکرد.
  3. تست عمومی‌سازی: تقسیم داده بر اساس سوژه‌های مختلف (subject-wise split) در صورت وجود داده واقعی.
- استقرار و رابط کاربری
1. توسعه رابط کاربری با Streamlit (app.py) که API ساده‌ای شامل بارگذاری CSV، نمایش موج، استخراج ویژگی و فراخوانی مدل دارد.
  2. آماده‌سازی مدل برای بارگذاری سریع در حافظه و پاسخ‌دهی زیر ۱-۲ ثانیه برای تجربه بلادرنگ.
  3. ارائه نسخه قابل اجرا در Colab (برای داوران) و مستندسازی اجرای محلی با requirements.txt.



تصویر 3: پیشبینی اپ Streamlit

## تولید مستندات بازتولیدپذیری

1. تهیه نوت‌بوک Colab (colab\_demo.ipynb) که مراحل کل را با سلول‌های قابل اجرا دارد.
2. ثبت نسخه کتابخانه‌ها (requirements.txt > pip freeze یا environment.yml).
3. توصیه به استفاده از Docker برای تولید محیط یکسان (در صورت نیاز).

## ابزارها و محیط تولید

- Python 3.9 یا 3.10
- NumPy >=1.21، Pandas >=1.4، SciPy >=1.8، scikit-learn >=1.0
- joblib برای ذخیره مدل
- matplotlib و seaborn برای مصورسازی
- Streamlit >=1.0 برای وب اپ
- Google Colab (برای اجرای سریع و اشتراک‌پذیری)
- Git و GitHub برای کنترل نسخه

ecg-hypoglycemia-predictor / requirements.txt

```
Code Blame
1 numpy==1.24.3
2 pandas==2.2.2
3 scipy==1.10.1
4 matplotlib
5 scikit-learn==1.2.2
6 joblib==1.2.0
7 streamlit==1.32.0
```

تصویر 4: requirements.txt

## دستاوردها و خروجی‌های مورد انتظار

الف) خروجی‌های فنی (Deliverables)

1. کد منبع کامل شامل generate\_data.py (تولید داده)، ecg\_utils.py (استخراج ویژگی)، train\_model.py (آموزش مدل)، app.py (رابط کاربری)
2. فایل مدل آموزش‌دیده در قالب model.pkl با متادیتای پارامترها و تاریخ آموزش.
3. Colab demo notebook (colab\_demo.ipynb) با سلول‌های قابل اجرا برای بازتولید کامل نتایج.
4. مستندات شامل README، راهنمای اجرای سریع (Quick start) و گزارش فنی.
5. مجموعه نمونه داده کوچک (demo\_data/) شامل چند فایل CSV برای تست سریع رابط.
6. مجموعه تصاویر برای گزارش و ارائه (Confusion Matrix, ROC, Streamlit UI, ECG examples).

ب) خروجی‌های آزمایشی و علمی

1. گزارش آماری کامل شامل accuracy, precision, recall, F1-score، AUC برای مجموعه آزمون.
2. تحلیل خطا: بررسی موارد False Positive / False Negative و عوامل زمینه‌ای مربوط (نویز، HRهای غیرمعمول و غیره).
3. ارزیابی پایداری مدل در مقابل نویز و تغییرات پارامترهای سیگنال.

ج) خروجی‌های کاربردی و تجاری بالقوه

1. نسخه اولیه وب اپ برای دمو در نمایشگاه‌ها.
2. طرح آغازین برای یک API ساده که قادر به دریافت فایل ECG و بازگشت پیش‌بینی باشد.
3. برآورد هزینه تولید و نصب در حد نمونه اولیه. (Prototyping budget)

### نوآوری‌ها و اهمیت پژوهشی

1. نوآوری علمی: کاربرد مستقیم تحلیل ECG برای پیش‌بینی رویداد متابولیک (هیپوگلیسمی) به جای تمرکز صرف بر آریتمی یا نارسایی قلبی، که حوزه تحقیقاتی جدیدی است.
2. نوآوری فنی: چارچوب تمام‌عیار از تولید داده مصنوعی واقع‌نما تا آموزش و استقرار مدل، که امکان بازتولید و توسعه سریع را فراهم می‌کند.
3. نوآوری کاربردی: توسعه نمونه نرم‌افزاری کم‌هزینه و سریع که قابلیت یکپارچه‌سازی با دستگاه‌های پوشیدنی و پلتفرم‌های تله‌مدیسن را دارد.
4. نوآوری آموزشی و زیست‌محیطی: فراهم‌سازی یک پلتفرم آموزشی و متن‌باز برای پژوهشگران جوان و دانشجویان که می‌خواهند وارد حوزه پزشکی-داده‌ای شوند.

### ملاحظات اخلاقی، قانونی و ایمنی

1. حریم خصوصی و داده بالینی: در این فاز از داده مصنوعی استفاده شده است تا از خطرات مرتبط با حریم خصوصی جلوگیری شود. در هر مرحله آتی که داده بالینی استفاده شود، پروتکل‌های اخلاقی، حصول رضایت آگاهانه و تاییدیه IRB باید دریافت شود.
2. محدودیت پزشکی: نتایج کنونی به‌عنوان ابزار تحقیقاتی و دمو فنی مطرح است و نباید به‌عنوان تشخیص پزشکی استفاده شود مگر پس از ارزیابی و گواهی‌نامه‌های بالینی.
3. ملاحظات فنی-ایمنی: در پیاده‌سازی نهایی روی دستگاه‌های پوشیدنی باید محدودیت‌های مصرف انرژی و تاخیر زمانی مدنظر قرار گیرد تا هشدارها به‌موقع و ایمن ارسال شوند.

### تحلیل اقتصادی و امکان‌سنجی بازار

خلاصه بازار هدف

- بازار اولیه: بیماران دیابتی وابسته به مانیتورینگ مکرر، سناریوهای خانگی و مراکز مراقبت از سالمندان.
- بازار ثانویه: تولیدکنندگان پوشیدنی، اپلیکیشن‌های سلامت و پلتفرم‌های تله‌مدیسن.

تخمین هزینه و صرفه‌جویی

- هزینه متوسط یک دستگاه CGM: چند صد تا هزار دلار در سال با هزینه سنسورها.
- هزینه پیشنهادی سامانه ECG-based: در صورت بهره‌گیری از یک سنسور ECG ساده یا استفاده از داده ساعت هوشمند، هزینه سخت‌افزار می‌تواند به‌طور بالقوه کمتر از ۳۰-۵۰ دلار برای قطعات پایه باشد.
- با یک سناریوی محافظه‌کارانه، پیاده‌سازی اولیه و نگهداری سرویس مبتنی بر ابر برای هر کاربر می‌تواند هزینه سالانه را به میزان بیش از ۵۰٪ کاهش دهد (نیاز به تحلیل بازار محلی و هزینه‌های استقرار دارد).

مدل کسب‌وکار پیشنهادی

1. اشتراک نرم‌افزار (SaaS): ارائه API و اپلیکیشن تحت اشتراک ماهانه برای کاربران و مراکز درمانی کوچک.

2. لایسنس برای تولیدکنندگان پوشیدنی: ارائه SDK/firmware جهت ادغام مدل با دستگاه‌های سخت‌افزاری.
3. خدمات B2B برای مراکز درمانی و بیمه‌ها: ارائه سیستم برای پایش جمعیتی و بهبود پیش‌بینی حوادث درمانی.

#### تحلیل ریسک اقتصادی

- نیاز به تایید بالینی می‌تواند زمان ورود به بازار را افزایش دهد.
- پذیرش بازار به شدت وابسته به اعتبار علمی و قوانین محلی است.
- مزیت رقابتی در کاهش هزینه و پذیرش آسان‌تر در محیط‌های کم‌درآمد است.

این پژوهش دارای محدودیت‌های شناخته‌شده‌ای است: اولاً، مجموعه داده مورد استفاده در فاز کنونی مصنوعی است و بازتولید نتایج در محیط بالینی نیازمند جمع‌آوری داده از بیماران واقعی و تاییدیه اخلاقی است. ثانیاً، به دلیل تمرکز بر ECG، اطلاعات تک‌حسی ممکن است در برخی موارد به‌تنهایی کافی نباشد؛ ترکیب با داده‌های تکمیلی (مانند فعالیت بدنی، دمای پوست،  $SpO_2$ ) می‌تواند دقت را افزایش دهد. علاوه بر این، پیاده‌سازی در دستگاه‌های با محدودیت منابع (باتری و پردازش) نیازمند بهینه‌سازی مدل و کاهش مصرف انرژی است. در مسیر آینده، جمع‌آوری داده واقعی، رگرسیون و طبقه‌بندی مبتنی بر معماری‌های عمیق (CNN/LSTM)، و اجرای آزمایش‌های میدان واقعی با گروه‌های داوطلب ترجیح داده می‌شود.

#### نتیجه‌گیری

این پژوهش نشان می‌دهد که تحلیل سیگنال ECG، همراه با استخراج ویژگی‌های مناسب و استفاده از الگوریتم‌های یادگیری ماشین، می‌تواند پایه یک سامانه هشداردهنده غیرتهاجمی برای هیپوگلیسمی را فراهم آورد. چارچوب فنی ارائه‌شده امکان بازتولید و توسعه سریع را فراهم کرده و به‌عنوان یک نقطه شروع برای مطالعات بالینی و توسعه محصول عمل می‌کند. با توجه به مزیت‌های اقتصادی بالقوه و قابلیت یکپارچه‌سازی با دستگاه‌های پوشیدنی و پلتفرم‌های تله‌مدیسن، این پروژه می‌تواند در مسیر تجاری‌سازی و کاربرد در جوامع کم‌منبع نقشی مؤثر ایفا کند؛ اما برای ورود به حوزه بالینی و تجاری نیاز به جمع‌آوری داده واقعی، بهینه‌سازی مدل و اخذ مجوزهای قانونی و اخلاقی دارد.

#### یادداشت اخلاقی و نحوه توسعه پروژه

این پروژه توسط نگارنده طراحی، پیاده‌سازی و اجرا شده است. در فرایند توسعه‌ی کدها و نگارش مستندات، از ابزار هوش مصنوعی ChatGPT (OpenAI) صرفاً به‌عنوان یک دستیار آموزشی و راهنمای فنی برای تولید ایده‌ها، رفع خطاهای برنامه‌نویسی و بهبود ساختار مستندات استفاده شده است. تمامی مراحل اجرای پروژه، از جمله طراحی مدل، تولید داده‌ها، آموزش، ارزیابی، و استقرار نهایی توسط نگارنده در محیط‌های برنامه‌نویسی و آزمایشگاهی انجام گرفته است. این پروژه هیچ‌گونه کپی‌برداری یا استفاده‌ی غیرمجاز از آثار دیگران ندارد و کلیه‌ی کدها و مستندات با درک و اجرای مستقیم نگارنده تولید و بررسی شده‌اند.

#### مراجع

1. Nguyen et al., \*ECG-based Detection of Hypoglycemia Using Machine Learning\*, IEEE Trans. Biomed. Eng., 2021.
2. Goldberger AL, \*PhysioNet: Physiologic Signal Archives\*, Circulation, 2000.
3. American Diabetes Association, \*Standards of Medical Care in Diabetes\*, 2023.
4. Pedregosa et al., \*Scikit-learn: Machine Learning in Python\*, JMLR, 2011.