

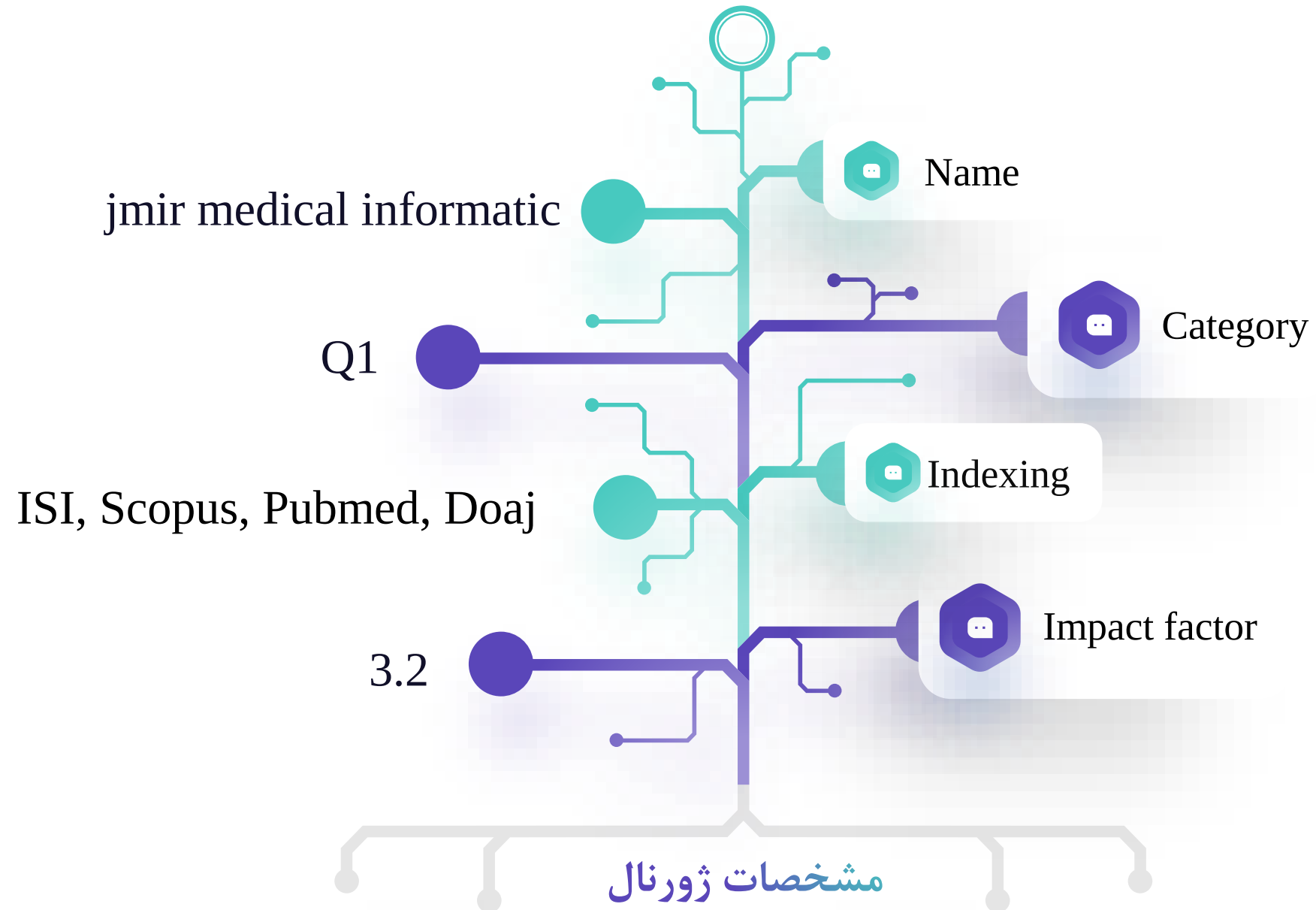
الحمد لله

Identifying Asthma-Related Symptoms From Electronic Health Records Using a Hybrid Natural Language Processing Approach Within a Large Integrated Health Care System: Retrospective Study

شناسایی علائم مرتبط با آسم از طریق الکترونیک سوابق سلامت با استفاده از زبان طبیعی ترکیبی رویکرد پردازش در یک سلامت یکپارچه بزرگ سیستم مراقبت: مطالعه گذشته نگر

تاریخ ارائه: 1404/3/19

ارائه دهنده: نیلوفر جانباز



AUC	area under the receiver operating characteristic curve	مساحت زیر منحنی مشخصه عملکرد گیرنده
BERT	Bidirectional Encoder Representations from Transformers	نمایش‌های رمزگذار دوطرفه از ترنسفورمرها
EHR	electronic health record	پرونده الکترونیکی سلامت
GPU	graphics processing unit	واحد پردازش گرافیکی

KPSC	Kaiser Permanente Southern California	مرکز کایزر پرمننته در جنوب کالیفرنیا
NLP	natural language processing	پردازش زبان طبیعی (شاخه‌ای از هوش مصنوعی برای درک و تولید زبان انسانی)
PPV	positive predictive value	ارزش پیش‌بینی مثبت
MLM	Masked language modeling	مدل سازی زبان با کلمات مخفی شده



بخش اول : مقدمه



بار جهانی و چالش‌های تشخیص آسم

✓ کنترل مؤثر آسم نیازمند شناسایی دقیق علائم است

✓ علائم کلیدی: سرفه، تنگی نفس، خس‌خس، فشار قفسه سینه

✓ چالش: علائم اغلب در یادداشت‌های آزاد EHR ثبت می‌شوند، نه در قالب‌های کدگذاری شده

✓ آسم یک بیماری مزمن التهابی تنفسی است

✓ در سال 2019، حدود 262 میلیون نفر در جهان مبتلا بودند

✓ شیوع در آمریکا: 7.8% در سال 2020

✓ آسم کنترل نشده منجر به خطرات جدی و بار اقتصادی می‌شود



نقش NLP در شناسایی علائم آسم

✓ موفقیت NLP در کودکان آسمی
(تشخیص، کنترل، پیش‌آگهی، درمان)

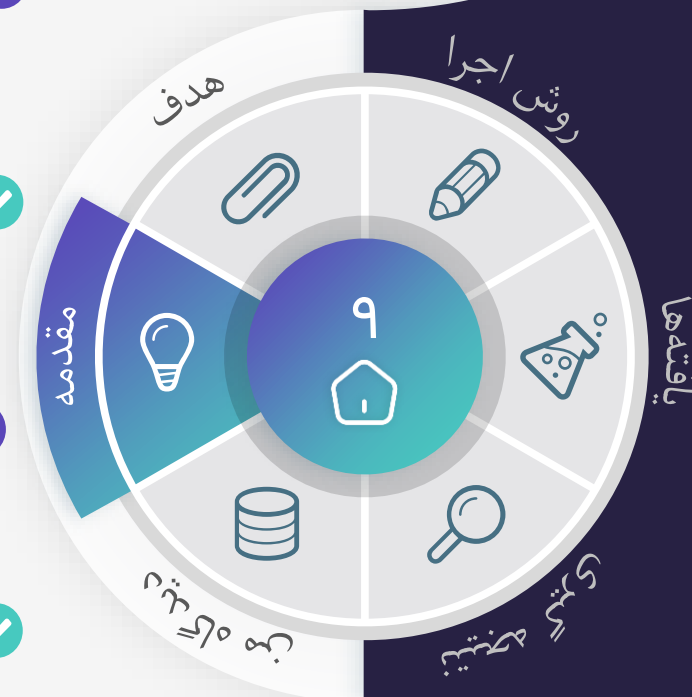
✓ NLP ابزاری برای پردازش متن‌های غیر ساختاریافته

✓ خلأ پژوهشی: عدم تحلیل سیستماتیک علائم
در بزرگسالان

✓ قابلیت استخراج، طبقه‌بندی و پیش‌بینی علائم از
یادداشت‌های بالینی

✓ روش‌های اولیه: مبتنی بر قوانین

✓ روش‌های جدیدتر: یادگیری عمیق و مدل‌های
ترنسفورمری مثل BERT

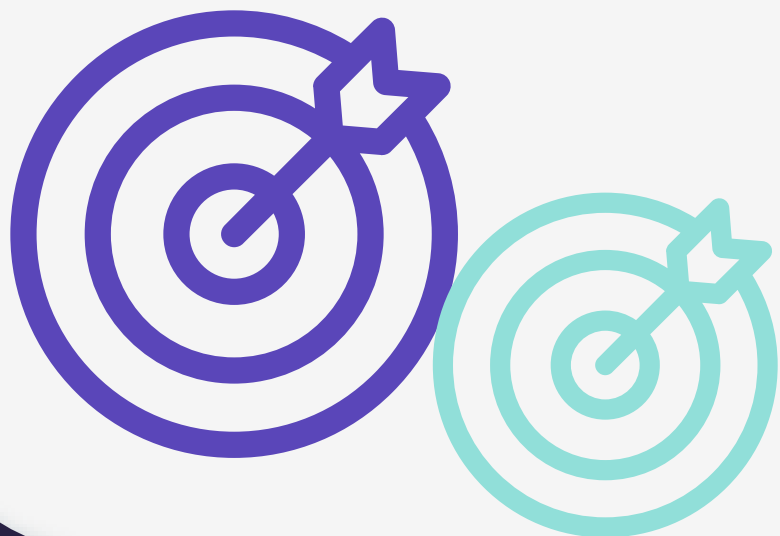


بخش دوم : هدف

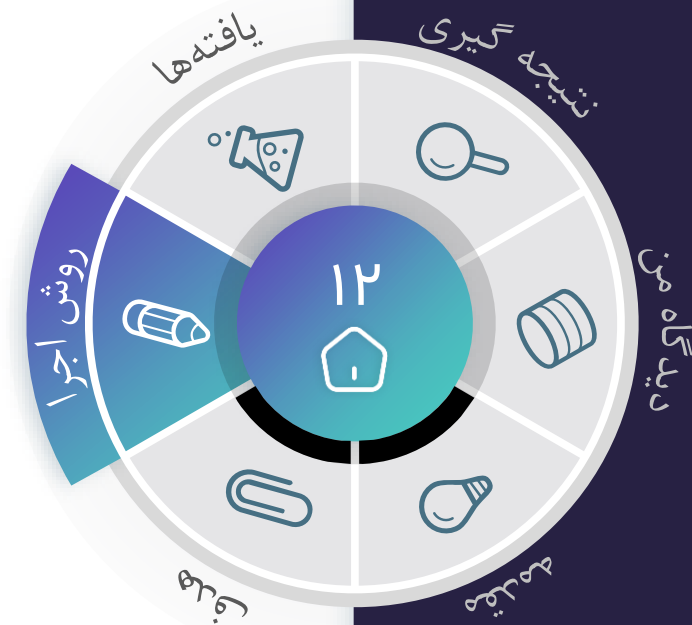


هدف اصلی

توسعه و اعتبارسنجی یک الگوریتم ترکیبی پردازش زبان طبیعی (NLP) برای شناسایی 4 علامت شایع آسم در بین بزرگسالان، با استفاده از داده‌های پرونده الکترونیکی سلامت (HER) در یک سیستم بهداشتی یکپارچه



بخش سوم : روش اجرا



ملاحظات اخلاقی

✓ تأیید مطالعه توسط کمیته بررسی مؤسسه KPSC

✓ معافیت از رضایت نامه آگاهانه به دلیل ماهیت داده ها

✓ سازگاری کامل با قانون HIPAA
(قانون قابلیت حمل و پاسخگویی بیمه سلامت)

✓ محدود بودن دسترسی به داده ها فقط برای پرسنل مجاز



نوع مطالعه:

گذشته نگر

سال های بررسی:

- 2013 تا 2018
- 2021 تا 2022



موسسه شرکت کننده:

کایزر پرممنته ساترن کالیفرنیا (KPSC)

محل انجام مطالعه:

امریکا - کالیفرنیا



معیارهای ورود به مطالعه

سن بین 18 تا 85 سال

1

داشتن ویزیت با تشخیص آسم

2

در بازه یک ساله قبل از تاریخ مرجع (index date):

- دریافت حداکثر یک داروی کنترل آسم یا حداکثر 2 اسپری سریع الاثر (SABA)
- حداکثر 1 حمله حاد آسمی
- عدم بستری به دلیل آسم یا بیماری‌های مرتبط مانند: COPD، ایدز، نقص ایمنی، سرطان‌های ریوی، حنجره

3

عضویت پیوسته در بیمه درمانی:
حداقل 1 سال قبل و 1 سال بعد از تاریخ مرجع

4





1 انتخاب کلیدواژه‌ها

2 استخراج و پیش پردازش داده‌ها

3 تقسیم داده‌ها

4 حاشیه نویسی دستی

5 توسعه الگوریتم قاعده بنیاد

6 توسعه الگوریتم ترنسفورمری (BERT)

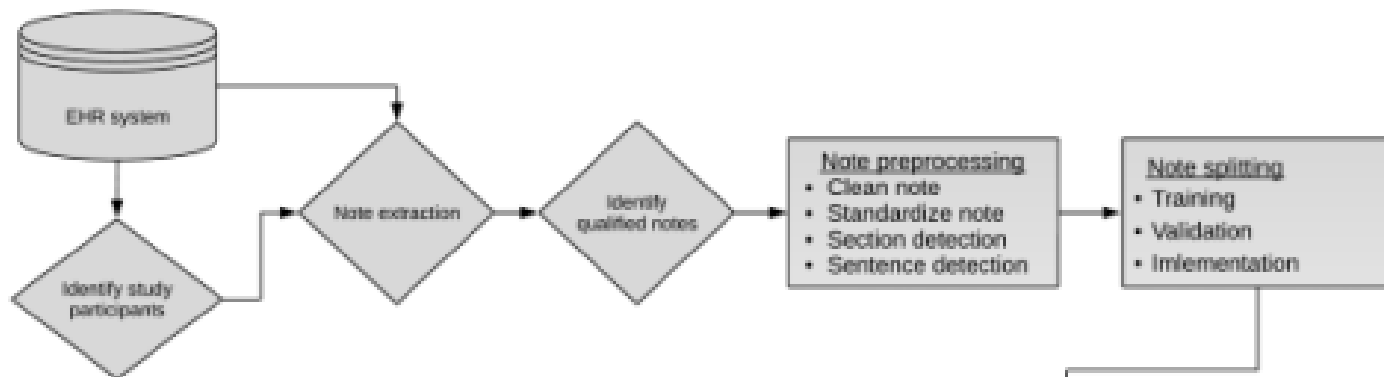
7 الگوریتم ترکیبی

8 ارزیابی عملکرد

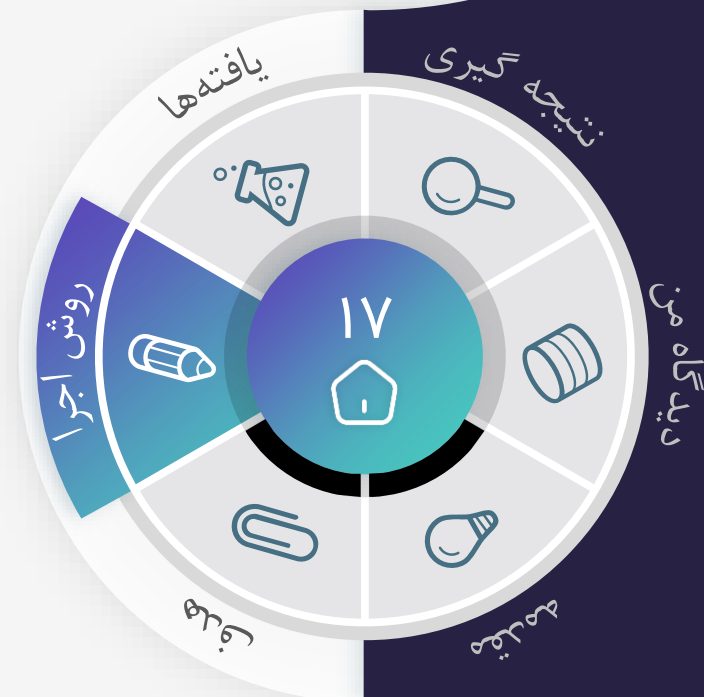
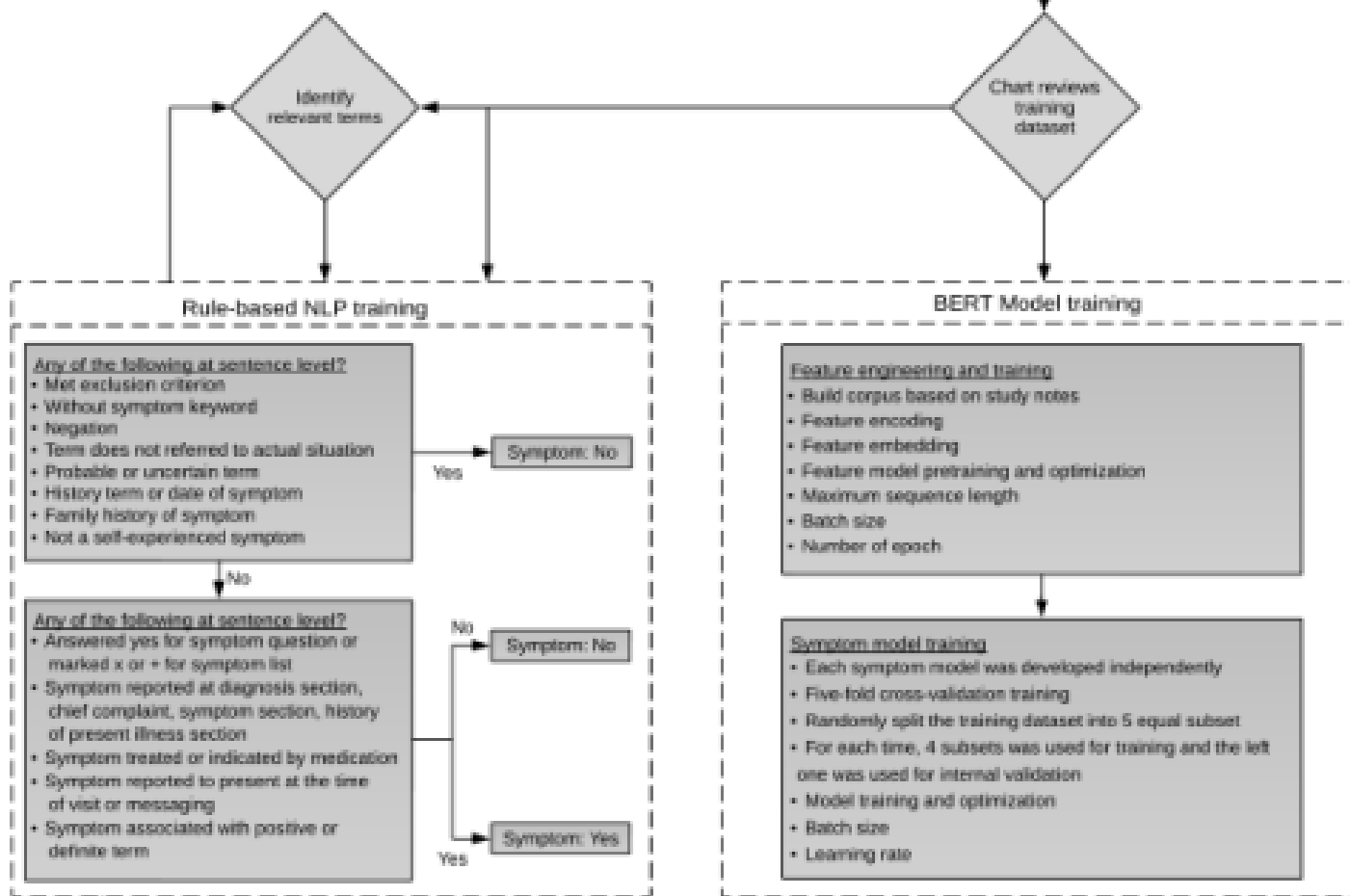
9 تجزیه و تحلیل ناسازگاریها

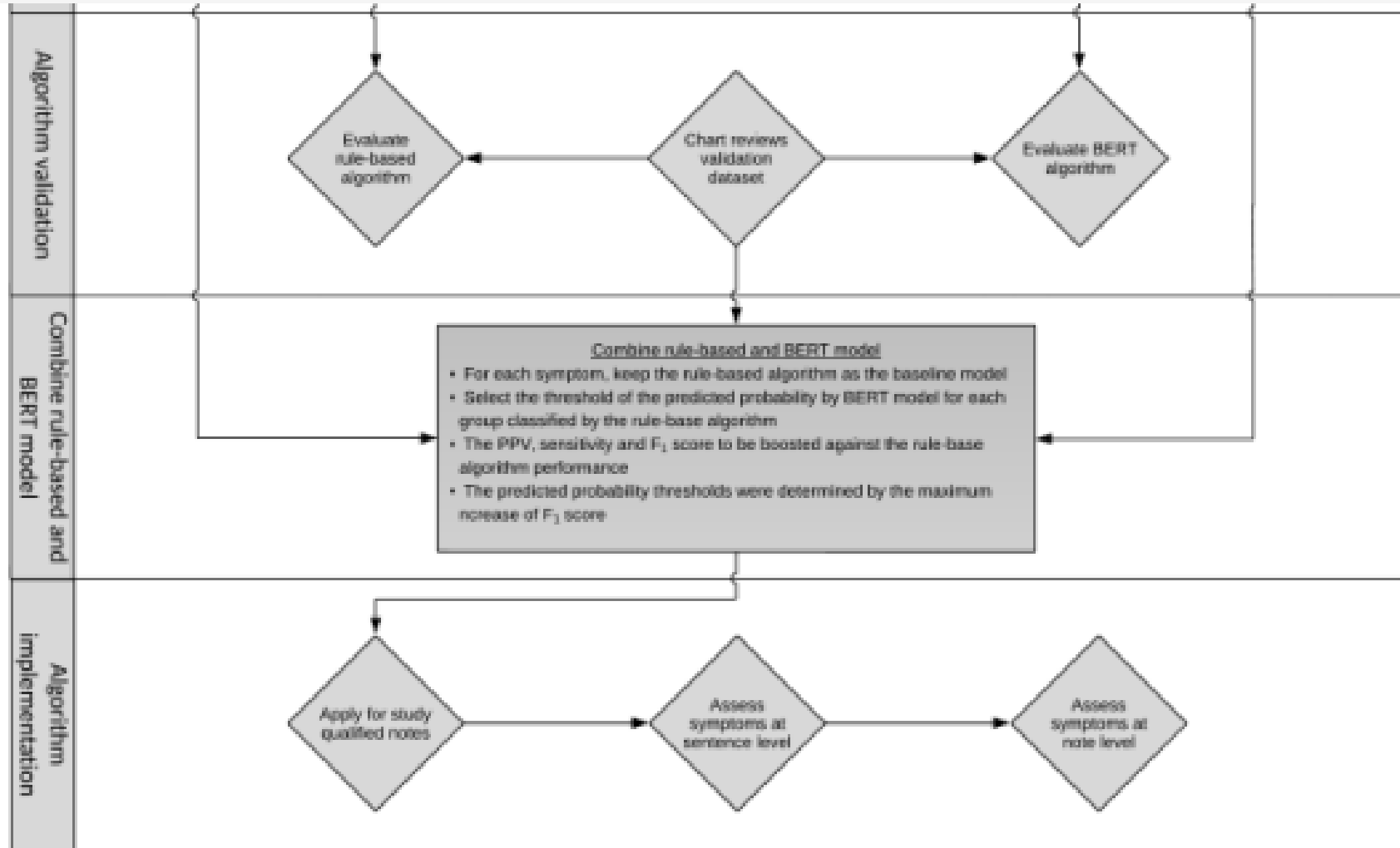
10 پیاده سازی نهایی

Data preparation and preprocessing



Algorithm training and development





بخش چهارم : یافته‌ها



Datasets	Total notes, n	Total sentences, n	Sentences per note, mean (SD)	Words per note, mean (SD)	Words per sentence, mean (SD)
Training	8000	380,363	48.4 (51.8)	612.1 (560.8)	12.6 (21.9)
Validation	1600	68,344	42.8 (37)	684.9 (601.4)	16 (27.5)
Implementation	11,364,952	127,763,086	11.2 (18.4)	183.6 (303.8)	16.3 (25.2)



Symptom	AUC		
	Training, mean (SD)	Internal validation, mean (SD)	Validation, mean (SD)
Cough	0.9989 (0.0002)	0.9975 (0.0008)	0.9986
Dyspnea	0.9963 (0.0013)	0.9935 (0.0021)	0.9973
Wheezing	0.9974 (0.0025)	0.9957 (0.0025)	0.997
Chest tightness	0.9988 (0.0007)	0.9969 (0.0026)	0.9971



Symptom	PPV ^a , %	Sensitivity, %	F_1 -score
Sentence level			
Rule-based			
Cough	96.95	93.74	0.953
Dyspnea	96.55	93.75	0.951
Wheezing	96.52	94.69	0.956
Chest tightness	97.41	93.89	0.956
BERT ^b			
Cough	95.9	94.7	0.953
Dyspnea	91.65	90.4	0.910
Wheezing	93.06	94.12	0.935
Chest tightness	93.73	95.56	0.946
Hybrid			
Cough	97.17	95.95	0.966
Dyspnea	96.86	93.9	0.954
Wheezing	96.53	94.88	0.957
Chest tightness	97.42	94.72	0.961
Note level			
Rule-based			
Cough	96.9	97.2	0.970
Dyspnea	97.15	97.15	0.972
Wheezing	96.44	96.44	0.964
Chest tightness	97.77	95.64	0.966
BERT ^b			
Cough	96.32	97.67	0.970
Dyspnea	91.64	93.73	0.927
Wheezing	92.5	95.79	0.941
Chest tightness	93.66	96.73	0.952
Hybrid			
Cough	97.7	99.07	0.984
Dyspnea	97.71	97.15	0.974
Wheezing	96.76	96.76	0.968
Chest tightness	97.42	96	0.967

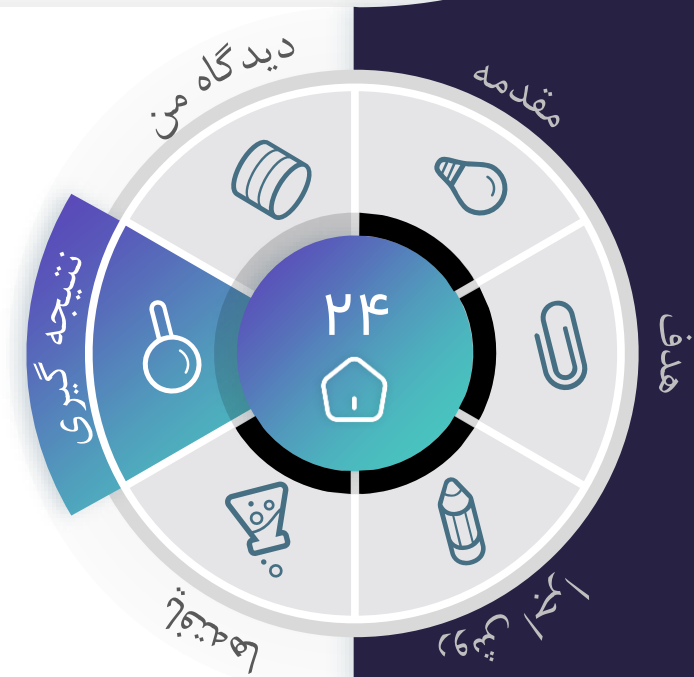


	Sentence level (n=127,763,086), n (%)	Note level (n=11,364,952), n (%)
Symptom		
Cough	1,363,713 (1.07)	660,685 (5.81)
Dyspnea	678,778 (0.53)	312,703 (2.75)
Wheezing	554,679 (0.43)	224,918 (1.98)
Chest tightness	141,733 (0.11)	64,251 (0.57)
Any of above symptoms	1,663,450 (1.3)	858,350 (7.55)
Number of symptoms ^a		
1	901,727 (0.71)	556,821 (5)
2	484,050 (0.38)	209,805 (1.85)
3	241,616 (0.19)	80,770 (0.71)
4	36,057 (0.03)	10,954 (0.1)

^aThe number of mutual symptoms present.



بخش پنجم : نتیجه گیری

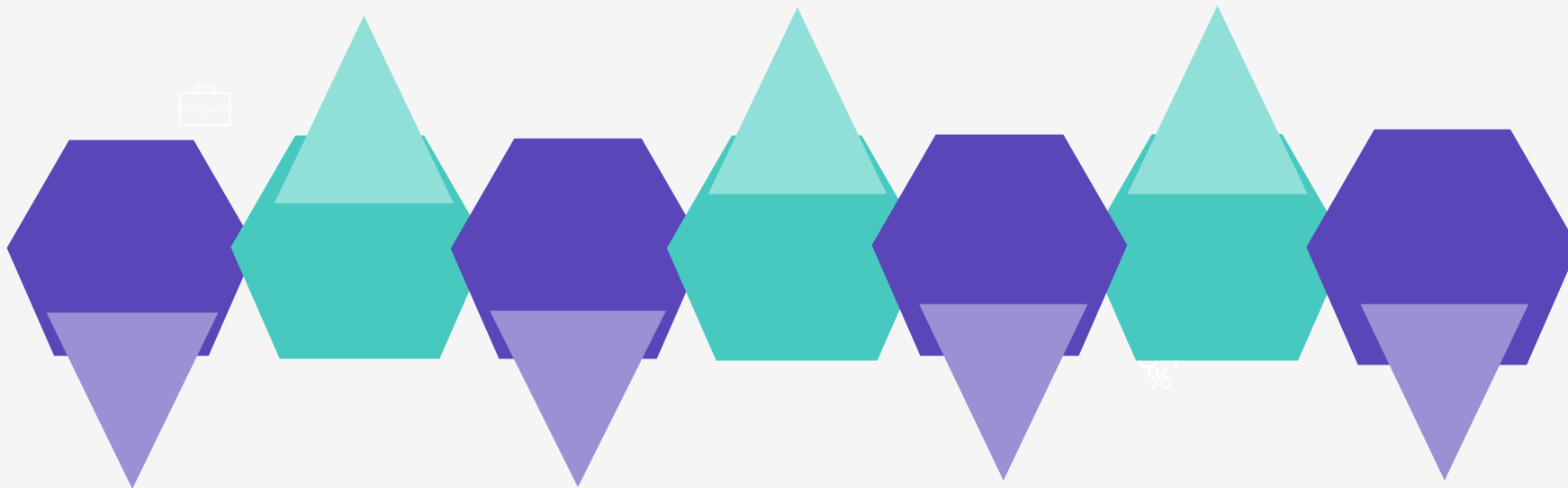


مزایای رویکرد ترکیبی

ترکیب NLP مبتنی بر قواعد و مدل ترنسفورمر

مناسب بودن rule-based برای کاربردهای حساس

تعادل بین دقت و پوشش

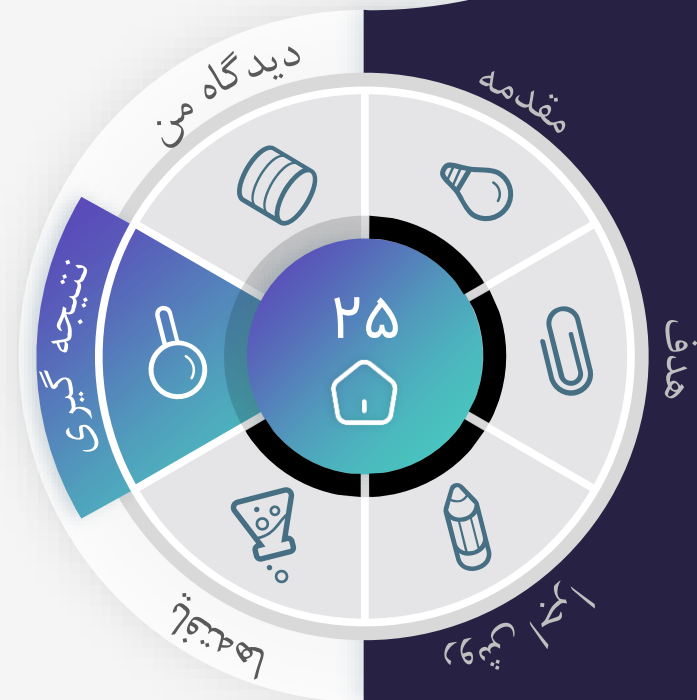


برتری دقت در مدل rule-based

قابلیت تعمیم به کاربردهای دیگر بالینی

دقت و حساسیت بالای 90%

برتری پوشش (recall) در مدل transformer



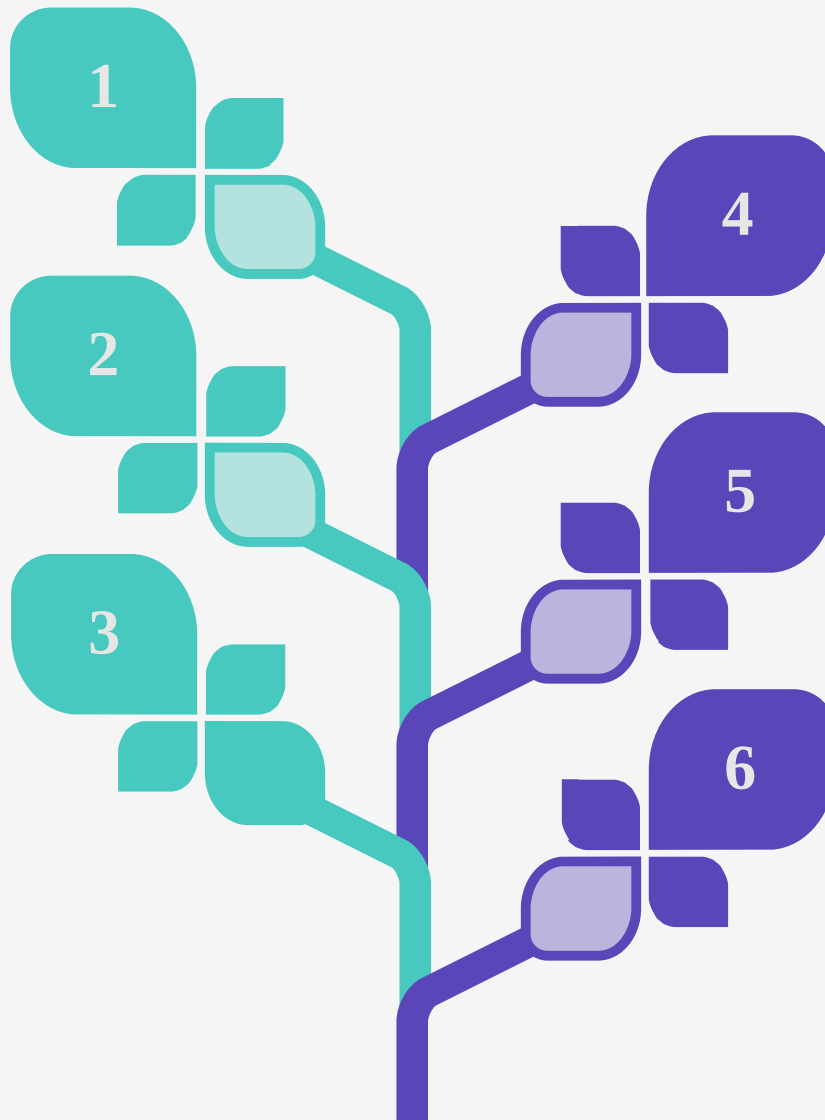
چالش ها

پراکندگی علائم در بخش های

مختلف یادداشت

اطلاعات منسوخ در لیست مشکلات

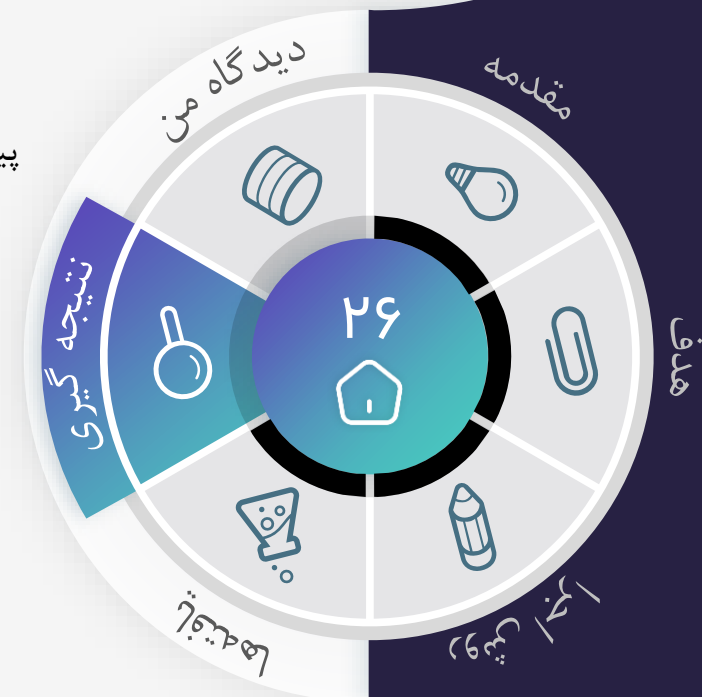
ابهام در دامنه نفی جملات



پیچیدگی در تشخیص نفی علائم

نیاز به داده ی آموزشی با کیفیت

حساسیت ساختار جمله ها در عملکرد مدل



محدودیت‌ها

وابستگی به کیفیت مستندسازی در EHR

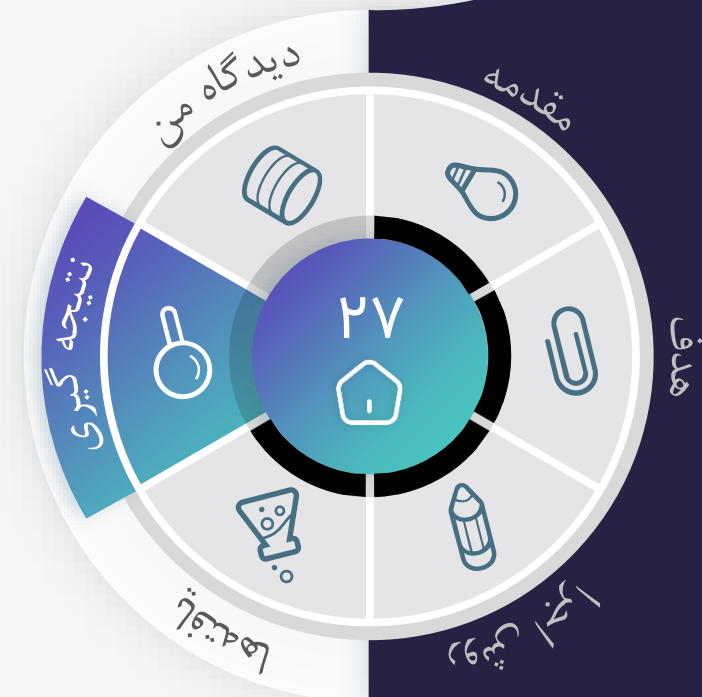
حذف علائم در حضور اضطراب

پوشش محدود واژگان در rule-based

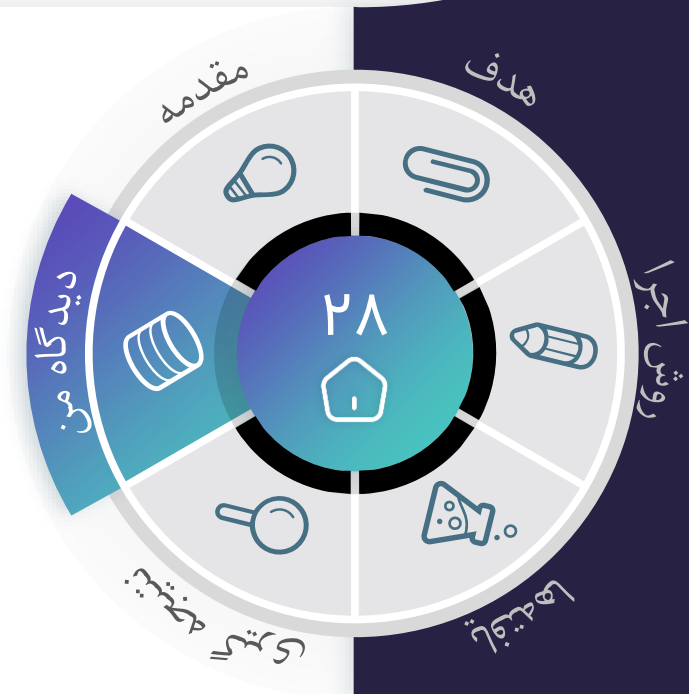
احتمال حذف توصیف‌های غیرکلیشه‌ای

آموزش مدل BERT فقط روی داده‌های برچسب‌خورده

محدودیت فاصله‌ی کلمات در الگوریتم



بخش ششم : دیدگاه من





افزودن قابلیت پیش‌بینی روند بیماری



تحلیل یادداشت‌های چندزبانه

بررسی روند زمانی علائم
(temporal tracking)

سپاس از توجه و نگاه زیباتون