

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



Automatic Depression Detection Using Smartphone-Based Text-Dependent Speech Signals: Deep Convolutional Neural Network Approach

تشخیص خودکار افسردگی با استفاده از سیگنال‌های گفتاری وابسته به متن مبتنی بر تلفن هوشمند: رویکرد شبکه عصبی پیچیده عمیق

زهرا کمالی

ژورنال کلاب ۸ خرداد

۱۴۰۲

مشخصات ژورنال





Name:

Journal of Medical
Internet Research



Indexing:

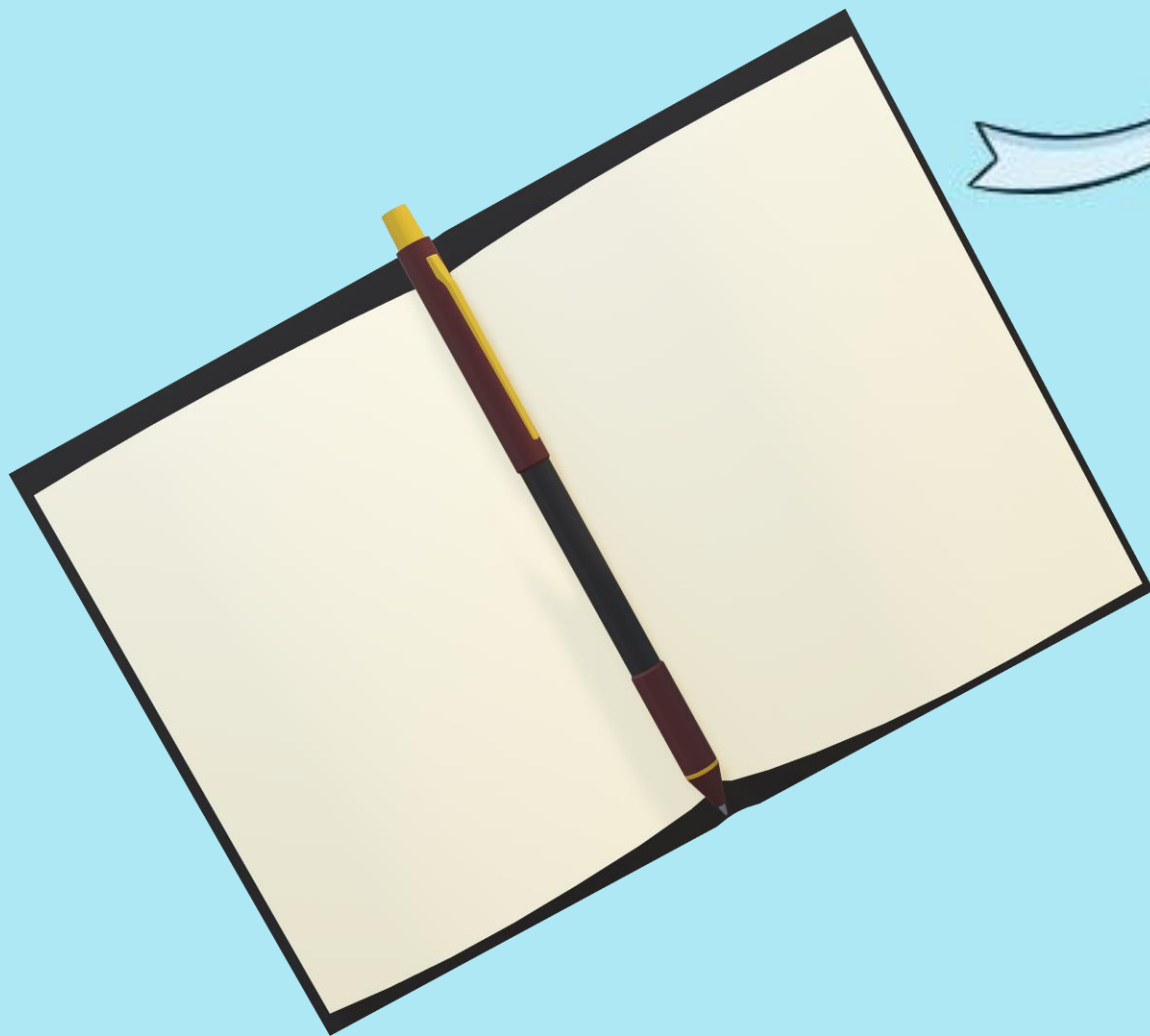
ISI, Scopus, PubMed, DOAJ





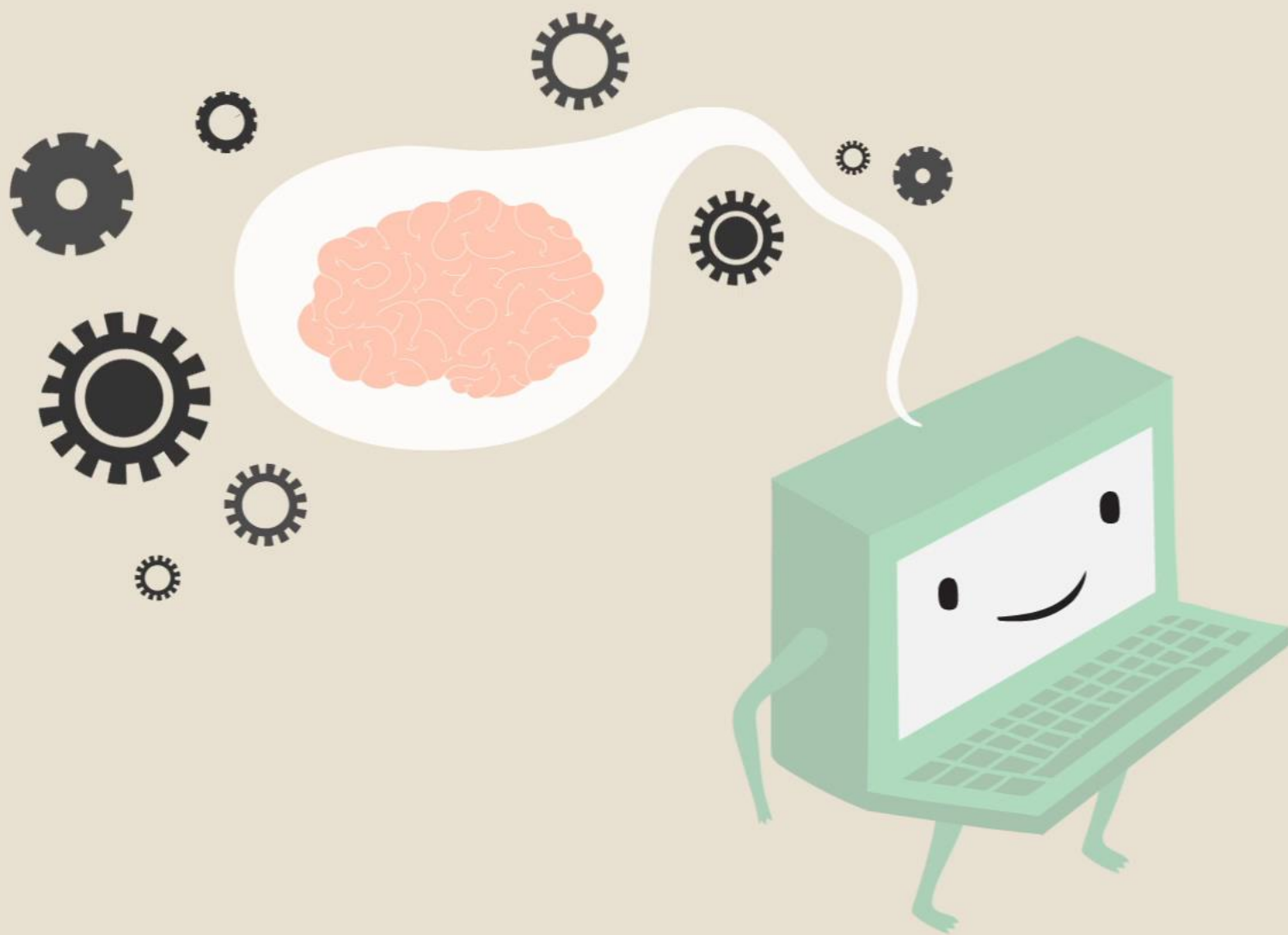
2022 Impact Factor:

7.0[^]



Categories:

Health Informatics: Q1





مقدمه و بیان مسئله



هدف مطالعه



روش اجرا



نتایج



بحث و نتیجه
گیری



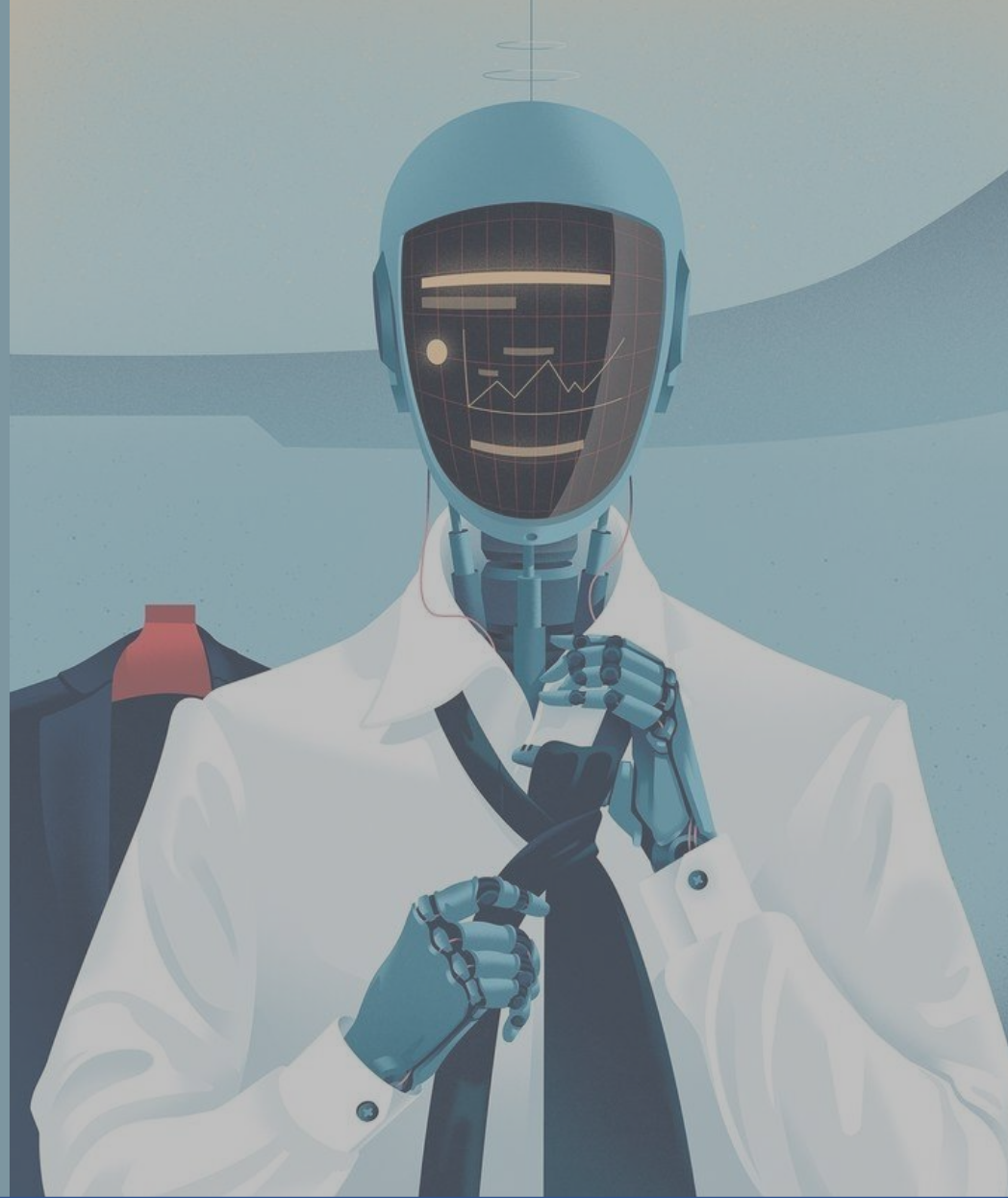
مقدمه و بیان مسئله



مقدمه و بیان مسئله

افسردگی یک بیماری روانی جدی است که بیش از ۳۰۰ میلیون نفر را در سراسر جهان تحت تاثیر قرار می دهد. منجر به انواع پیامدهای منفی سلامتی در افراد منجر می شود. هنگامی که درمان نشود، می تواند بر کیفیت زندگی تاثیر بگذارد، بهره وری کار را کاهش دهد و منجر به خودکشی شود.

مقدمه و بیان مسئله



تشخیص افسردگی خودکار
با ظهور مجموعه داده های
در دسترس عمومی و قدرت
تکنیک های یادگیری ماشین
برای یادگیری الگوهای
پیچیده محبوبیت پیدا کرده
است. این مطالعات نشان
داده اند که ویژگی های
صوتی ارتباط نزدیکی با
افسردگی دارند.

یک رویکرد امیدوارکننده
برای رسیدگی به
مشکلات فوق، شناسایی
نشانگرهای افسردگی
و تکنیک های پیشرفته
یادگیری ماشینی
با استفاده از حسگرهای
قابل دسترس در دنیای
واقعی است.





هدف مطالعه

یک رویکرد سنجش خودکار با استفاده از یادگیری عمیق برای توصیف حالت های خواندن وابسته به متن می تواند کمک قابل توجهی به این زمینه تحقیقاتی کند. بنابراین، در این مطالعه، از گروه گفتاری زبان کره مبتنی بر کوهورت برای تشخیص افسردگی با استفاده از گفتار خواندنی وابسته به متن استفاده شد. سپس یک لایه شبکه عصبی کانولوشنال و لایه متراکم برای مدل سازی حالت افسردگی اعمال کردیم. در نهایت، مدل آموزش دیده برای پیش بینی افسردگی در میان نمونه های داده ناشناخته استفاده شد. عملکرد مدل با رویکردهای درمانی، یادگیری ماشینی و مدل های از پیش آموزش دیده مقایسه شد. نتایج نشان می دهد که رویکرد ما در ارزیابی آسان و خودکار حالات افسردگی در گفتار مؤثر است.





۱۵۳

بیمار



سن بیشتر ۱۹ سال

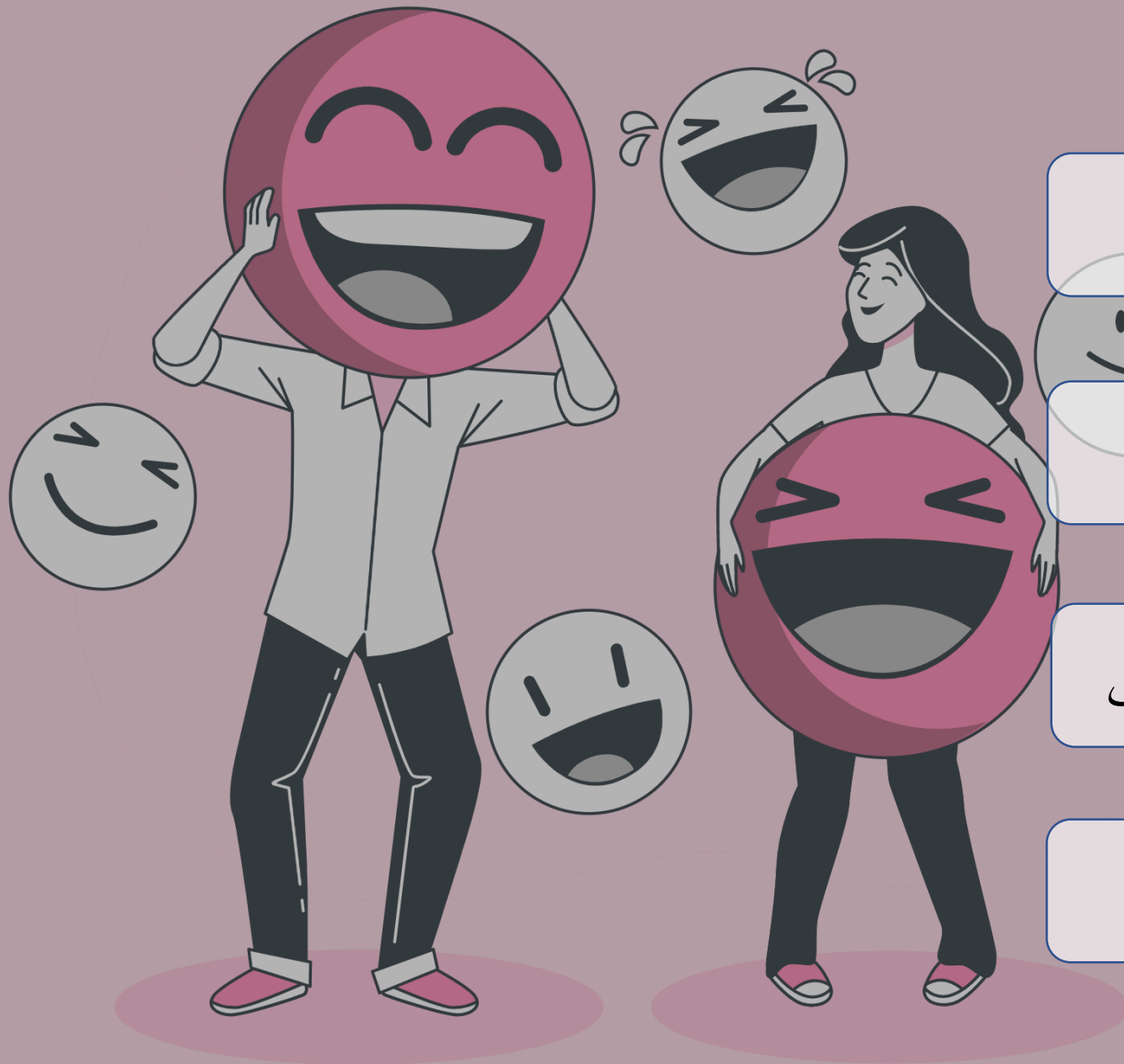


ارزیابی توسط PHQ-9 و HAM-D



رضایت کتبی





۱۶۵ گروه کنترل



سن بیشتر از ۱۹ سال



بدون هیچگونه سابقه اختلالات روانی



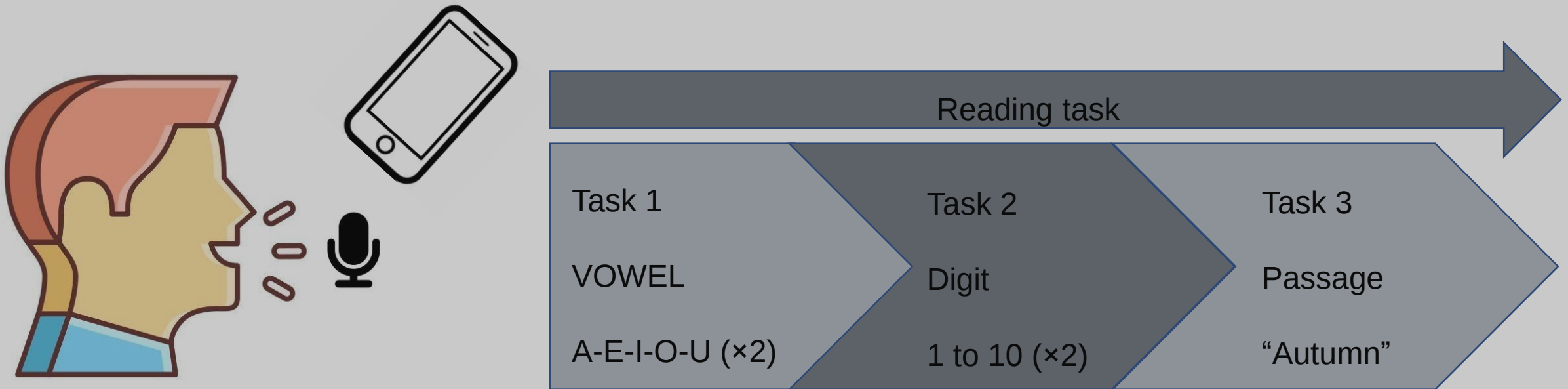
رضایت کتبی



Demographic characteristics of major depressive disorder (MDD) and health controls (HC) groups.

Factors	MDD	HC	<i>P</i> value
Gender, n (%)			.16 ^a
Male	51 (33.3)	42 (25.5)	
Female	102 (66.7)	123 (74.5)	
Age (years), mean (SD)	38.64 (13.26)	37.15 (12.18)	.06
Education (years), mean (SD)	13.10 (2.57)	14.71 (2.19)	<.001 ^b
Nonalcoholic, n (%)	55 (35.95)	83 (50.30)	.008 ^{a,c}
Nonsmoker, n (%)	45 (29.41)	21 (12.73)	.006 ^{a,c}
HAM-D ^d , mean (SD)	22.18 (6.34)	2.07 (4.61)	<.001 ^b
PHQ-9 ^e , mean (SD)	16.92 (6.99)	2.14 (3.44)	<.001 ^b

جمع آوری داده های صوتی

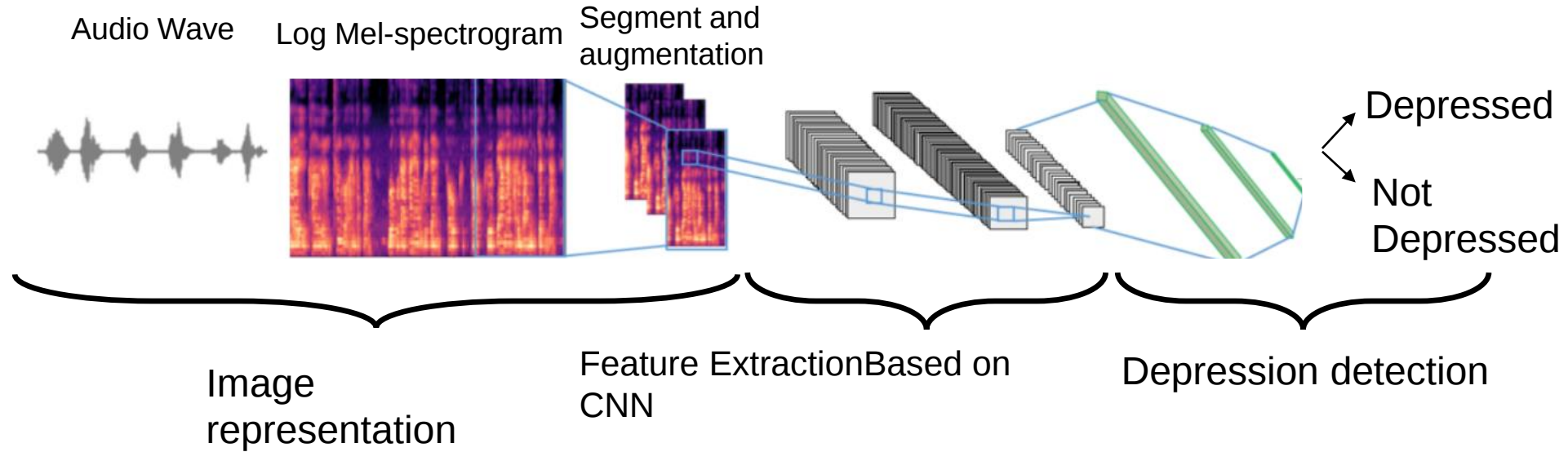




طبقه بندی مدت زمان ضبط نمونه های صوتی پس از پردازش و تقویت سیگنال ها در جدول زیر نمایش داده شده.

	Vowel	Digit	Passage
MDD			
Number of samples, n (%)	153 (48.1)	153 (48.1)	153 (48.1)
Speech time (seconds), mean (SD)	14.67 (4.07)	16.02 (5.39)	91.80 (17.46)
Total speech time (seconds)	2307	2373	14,150
HC			
Number of samples, n (%)	165 (51.9)	165 (51.9)	165 (51.9)
Speech time (seconds), mean (SD)	13.59 (4.06)	14.39 (4.44)	72.71 (7.91)
Total speech time (seconds)	2325	2361	13,110

The pipeline of the proposed method for depression detection





مقایسه ویژگی های جمعیتی و بالینی مطالعه

دموگرافیک و بالینی
عوامل سلامت

تحصیلات

PHQ-9

HAM-D

Mann-whitney
U test



مقایسه ویژگی های جمعیتی و بالینی مطالعه

رابطه جنسی

سابقه مصرف الکل

سابقه مصرف دخانیات

Chi-squared test



نتایج آزمون های مطالعه

وجود تفاوت معنا دار
نتایج آزمون HAM-
D دو گروه

عدم وجود تفاوت معنا دار در
جنس و سن گروه ها.

$P > 0.001$ ✓

وجود تفاوت معنا دار
نتایج آزمون PHQ-
9 دو گروه

وجود تفاوت معنا دار سال
تحصیلی گروه ها.

$P < 0.01$ ✓





معیارهای ارزیابی مدل

Accuracy

precision

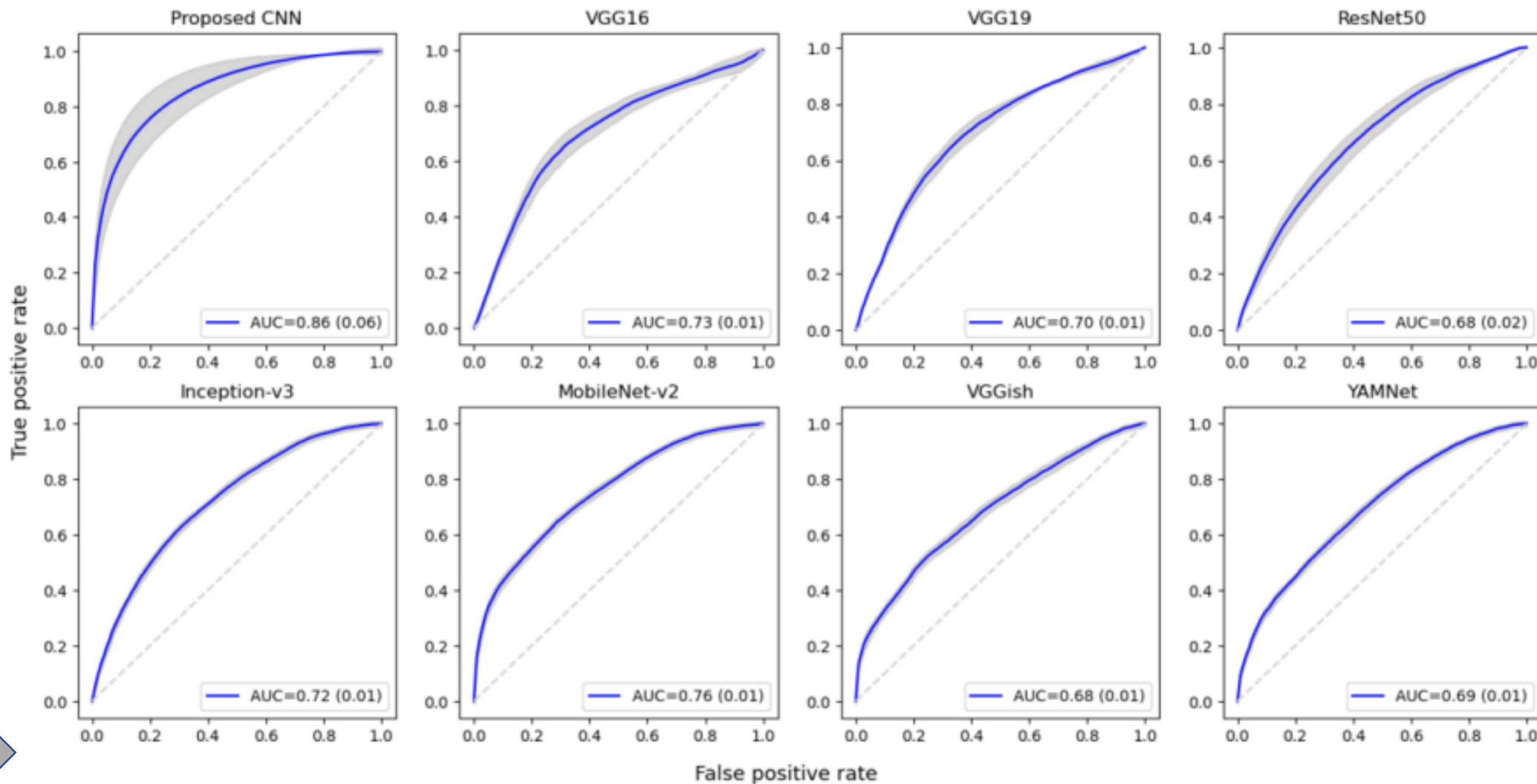
Recall

F1-Score

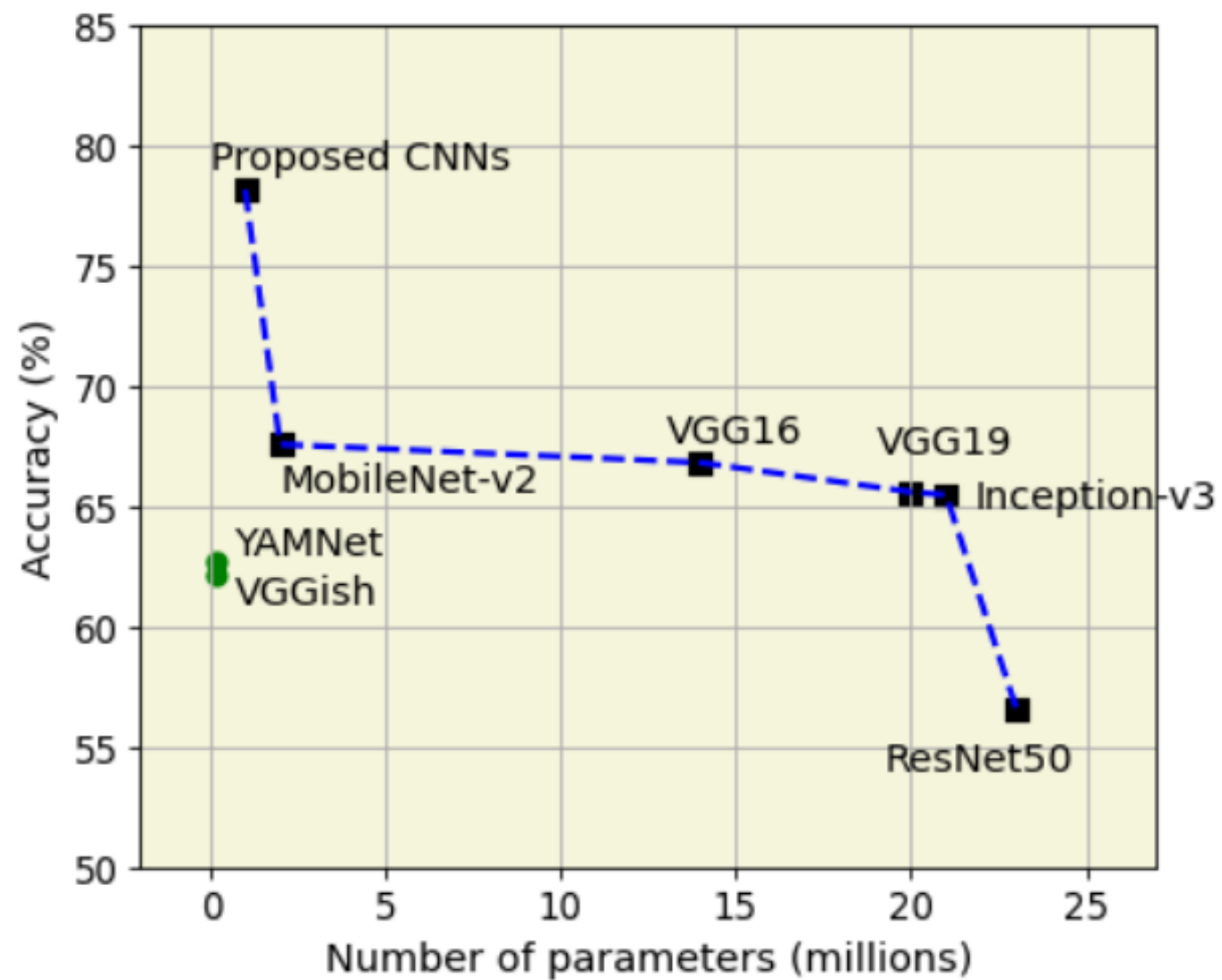
Receiver operating characteristic curve

Receiver operating characteristic curve

Figure 4. Mean and SD of the receiver operating characteristic curve and area under receiver operating characteristic curve (AUC) based on the convolutional neural network (CNN) models with 10-fold cross-validation in the passage task.



Data set and models	Test set			
	ACC ^a (%), mean (SD)	PREC ^b (%), mean (SD)	REC ^c (%), mean (SD)	F ₁ -score ^d (%), mean (SD)
Vowel				
Proposed CNNs	65.44 (6.58)	64.04 (21.56)	46.66 (24.55)	51.13 (24.38)
VGG16	47.70 (1.76)	53.16 (1.26)	41.06 (3.83)	46.24 (2.46)
VGG19	47.97 (0.95)	53.29 (0.83)	43.06 (6.30)	47.37 (4.30)
ResNet50	47.02 (3.98)	10.99 (21.99)	20.00 (40.00)	14.89 (28.38)
Inception-v3	56.83 (1.33)	63.13 (1.81)	51.81 (3.01)	56.84 (1.87)
MoblieNet-v2	57.53 (0.85)	63.34 (0.99)	54.20 (4.66)	58.29 (2.24)
VGGish	55.93 (0.78)	59.93 (0.77)	60.01 (1.16)	59.99 (0.75)
YAMNet	62.54 (0.98)	65.78 (0.82)	66.43 (1.96)	66.09 (1.16)
Digit				
Proposed CNNs	66.60 (7.10)	56.79 (28.83)	47.72 (24.99)	51.53 (26.23)
VGG16	61.69 (1.24)	62.65 (1.19)	48.49 (7.53)	54.32 (5.20)
VGG19	53.20 (1.62)	60.62 (1.66)	35.94 (6.53)	44.79 (5.90)
ResNet50	54.15 (9.25)	11.89 (18.16)	30.00 (45.83)	17.03 (26.00)
Inception-v3	61.69 (1.24)	65.99 (1.50)	58.72 (3.76)	62.06 (2.04)
MoblieNet-v2	61.92 (2.03)	65.59 (1.64)	60.49 (3.81)	62.9 (2.70)
VGGish	46.00 (1.74)	49.83 (1.60)	51.26 (2.94)	50.52 (2.14)
YAMNet	57.37 (0.83)	61.32 (0.73)	55.18 (3.06)	58.04 (1.76)
Passage				
<i>Proposed CNNs^e</i>	<i>78.14 (2.40)</i>	<i>76.83 (3.92)</i>	<i>77.90 (2.73)</i>	<i>77.27 (2.10)</i>
VGG16	66.80 (0.40)	65.97 (1.48)	65.48 (3.11)	65.64 (1.05)
VGG19	65.58 (2.52)	58.14 (2.69)	66.38 (5.16)	61.80 (2.08)
ResNet50	56.62 (4.84)	79.98 (9.15)	17.11 (18.59)	23.54 (21.24)
Inception-v3	65.46 (0.28)	63.00 (0.69)	69.86 (2.04)	66.23 (0.60)
MoblieNet-v2	67.56 (1.28)	65.94 (2.13)	68.98 (3.07)	67.29 (0.97)
VGGish	62.17 (0.69)	59.96 (0.81)	66.70 (1.70)	63.13 (0.75)
YAMNet	62.65 (0.52)	60.38 (0.46)	67.39 (1.46)	63.69 (0.74)



Demographic characteristics of major depressive disorder (MDD) and health controls (HC) groups.

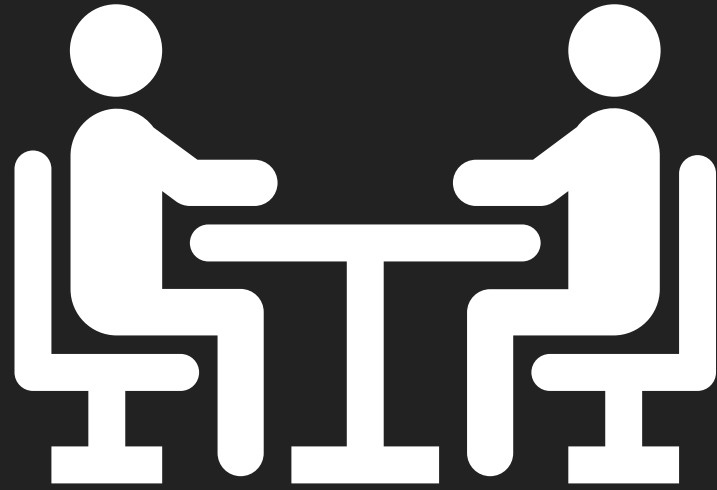
Factors	MDD	HC	<i>P</i> value
Gender, n (%)			.16 ^a
Male	51 (33.3)	42 (25.5)	
Female	102 (66.7)	123 (74.5)	
Age (years), mean (SD)	38.64 (13.26)	37.15 (12.18)	.06
Education (years), mean (SD)	13.10 (2.57)	14.71 (2.19)	<.001 ^b
Nonalcoholic, n (%)	55 (35.95)	83 (50.30)	.008 ^{a,c}
Nonsmoker, n (%)	45 (29.41)	21 (12.73)	.006 ^{a,c}
HAM-D ^d , mean (SD)	22.18 (6.34)	2.07 (4.61)	<.001 ^b
PHQ-9 ^e , mean (SD)	16.92 (6.99)	2.14 (3.44)	<.001 ^b

نتایج تحلیل و مقایسه معیار های ارزیابی مدل

✓ میانگین AUC برای مدل CNN = 0.86

✓ عملکرد دقیق تر در تجزیه و تحلیل با مدل CNN مبتنی بر استخراج ویژگی های داده.

✓ عملکرد طبقه بندی خوب در تشخیص افسردگی با منابع کمتر



بحث و نتیجه
گیری

یافته های اصلی



دستیابی به فواید مربوط یادگیری الگوهای صوتی در راستای تشخیص افسردگی

دستیابی به پتانسیل ارزشمند در جهت کاهش افسردگی

کشف قابلیت موثر و قدرتمند گفتار در افسردگی



نقطه قوت دوم

رویکرد خودکار تشخیص افسردگی فرد با تجزیه و
تحلیل گفتار خود



نقطه قوت اول

ایجاد فرصت های جدید جهت شناسایی نشاگرهای
گفتاری مربوط به افسردگی



محدودیت سوم

عدم در نظر گرفتن صوت و
نویز مکان های پر سروصدا



محدودیت دوم

طیف کمتری از داده ها



محدودیت اول

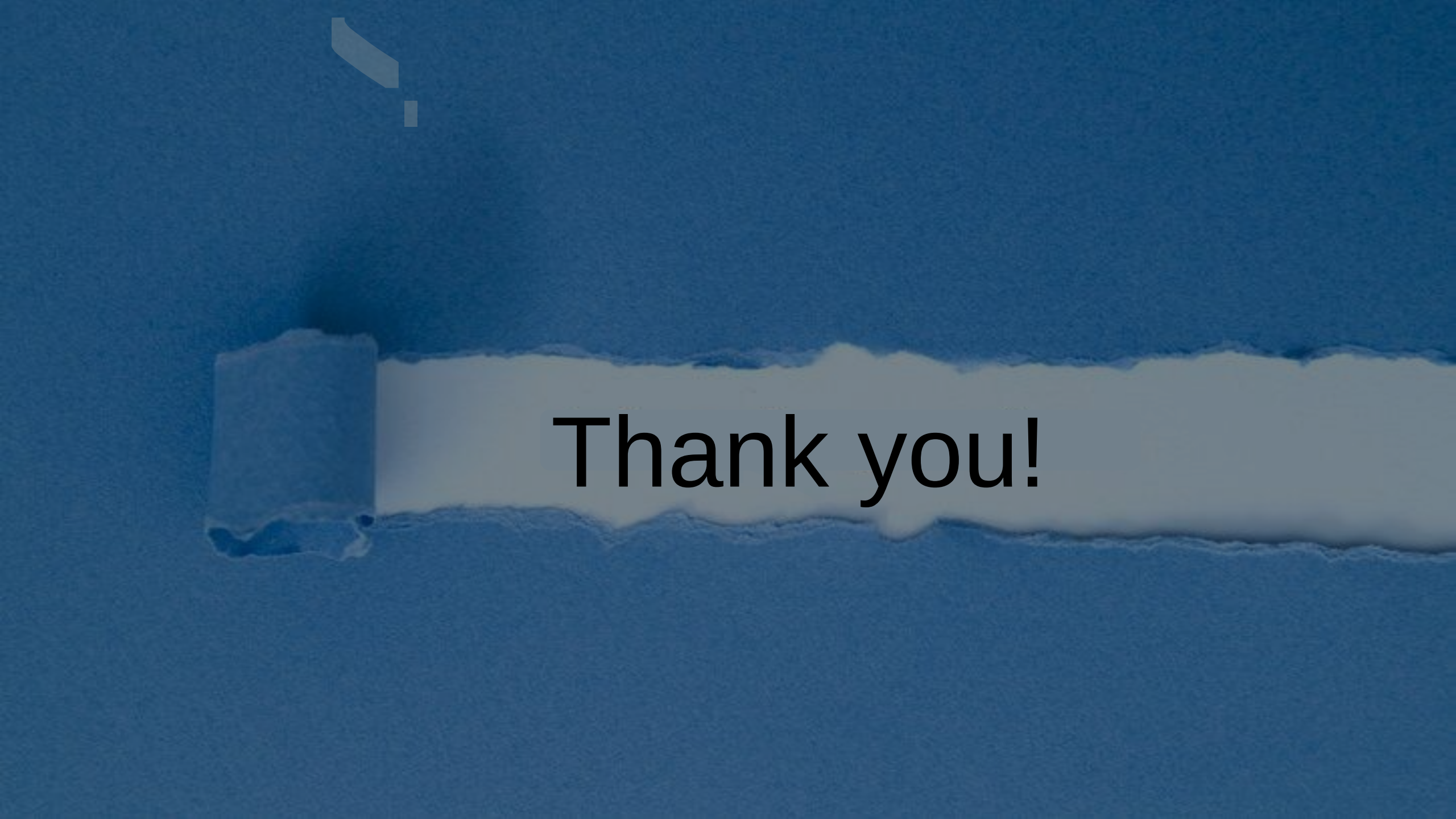
تک زبانه بودن داده های
گفتاری

پیشنهادهای و دیدگاه من

طبقه بندی بیمار افسرده از لحاظ
شدت بیماری برای درمان

جایگزین کردن داده های ویدیویی





Thank you!