

Smart Sign Language Translator «HAMRAZ»

مترجم زبان اشاره هوشمند
«هم راز»

استاد راهنما:

دکتر حسین ابراهیم پور

گردآورنده:

فاطمه احمدی رنانی



چالش اصلی و ضرورت ها

آمارهای نگران کننده

تنها 30 درصد از دانشوایان ایرانی از آموزش زبان اشاره استاندارد برخوردارند

70 درصد از خانواده های دارای عضو ناشنوا با چالش ارتباطی مواجه هستند

نرخ بیکاری در جامعه ناشنوایان 2.5 برابر جامعه عمومی

مسئله مورد بررسی

شکاف ارتباطی عمیق بین جامعه ناشنوایان و جامعه شنوایان

عدم دسترسی به مترجمین زبان اشاره در موقعیت های روزمره و اضطراری

محدودیت ابزارهای موجود که عمدتاً به زبان انگلیسی هستند و با زبان اشاره سازگاری ندارد

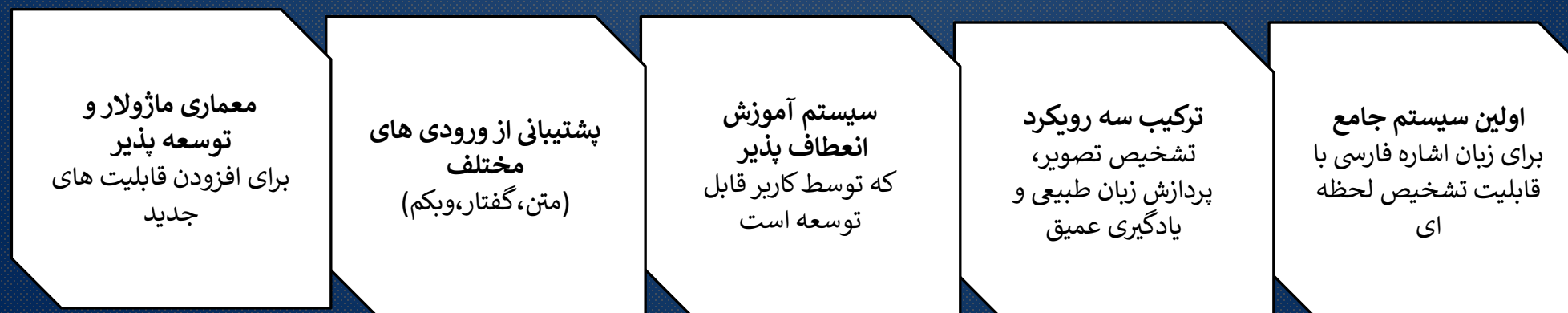
نیاز به راه حل بومی برای فرهنگ و زبان فارسی

اهداف و نوآوری ها

اهداف



نوآوری های کلیدی



معماری های سیستم

لایه بندی معماری

لایه اول: رابط کاربری UI

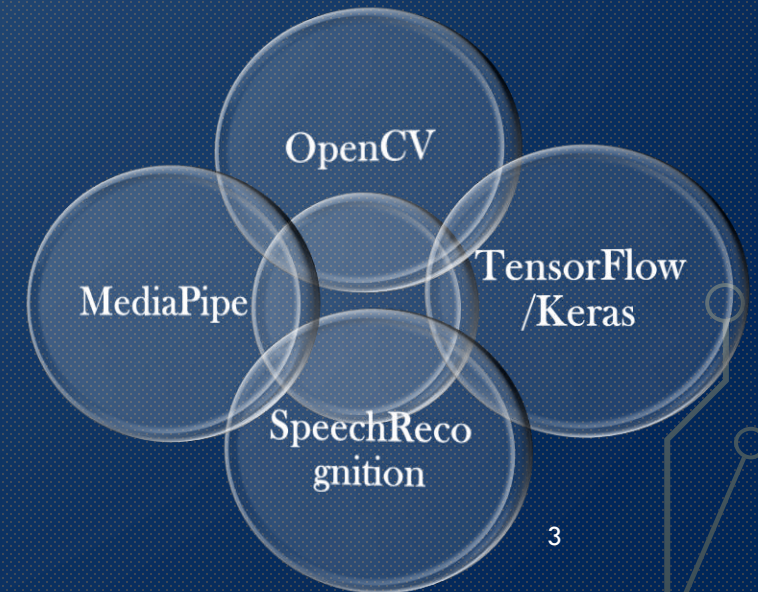
لایه دوم: ماژول های کاربردی

لایه سوم: مدل های یادگیری ماشین

لایه چهارم: کتابخانه های پردازشی

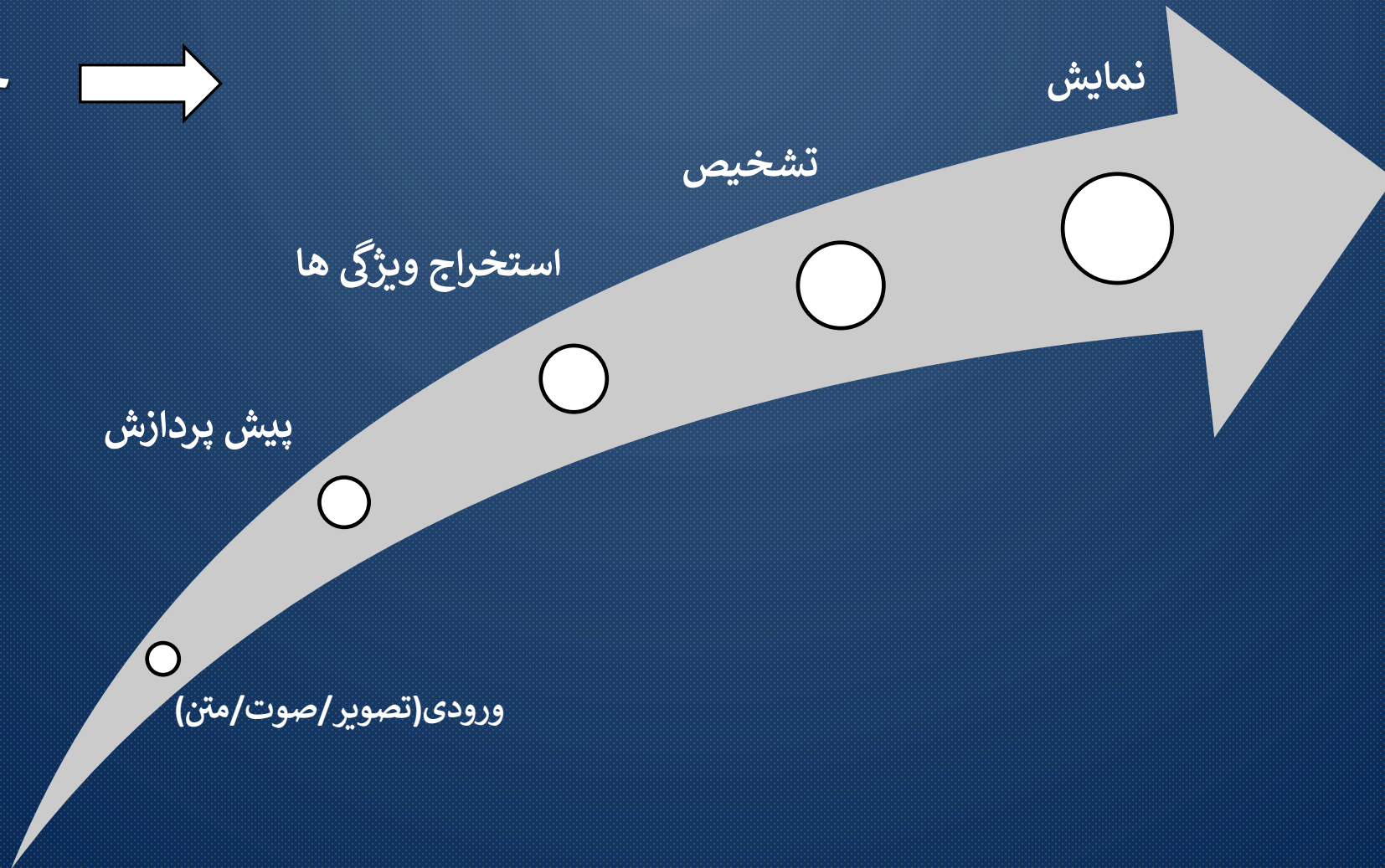
ماژول های

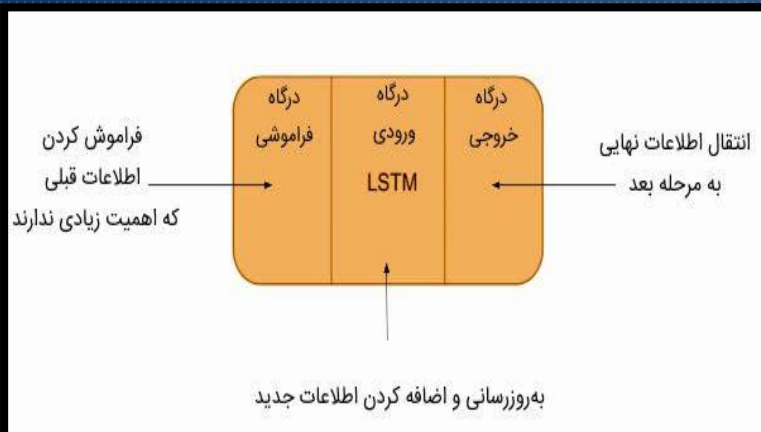
- ❖ تشخیص زنده زبان اشاره
- ❖ تبدیل متن به زبان اشاره
- ❖ تبدیل گفتار به زبان اشاره
- ❖ مدیریت سیستم



معماری های سیستم

جریان داده





Sigmoid

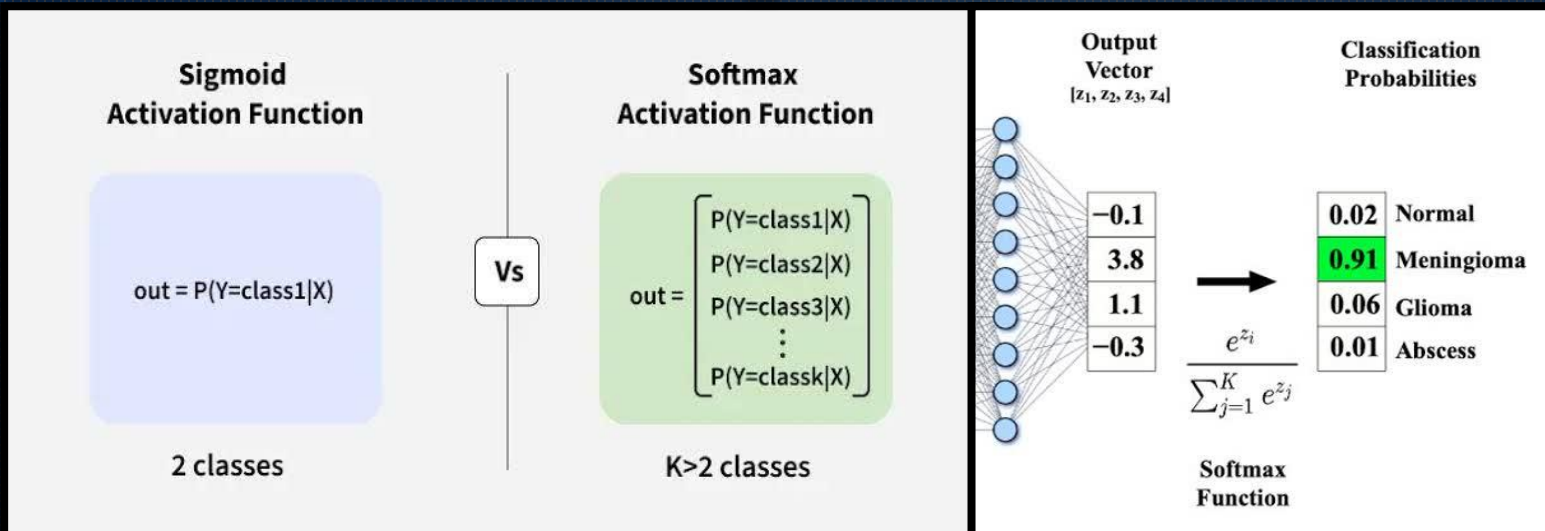
$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

vs.

Softmax

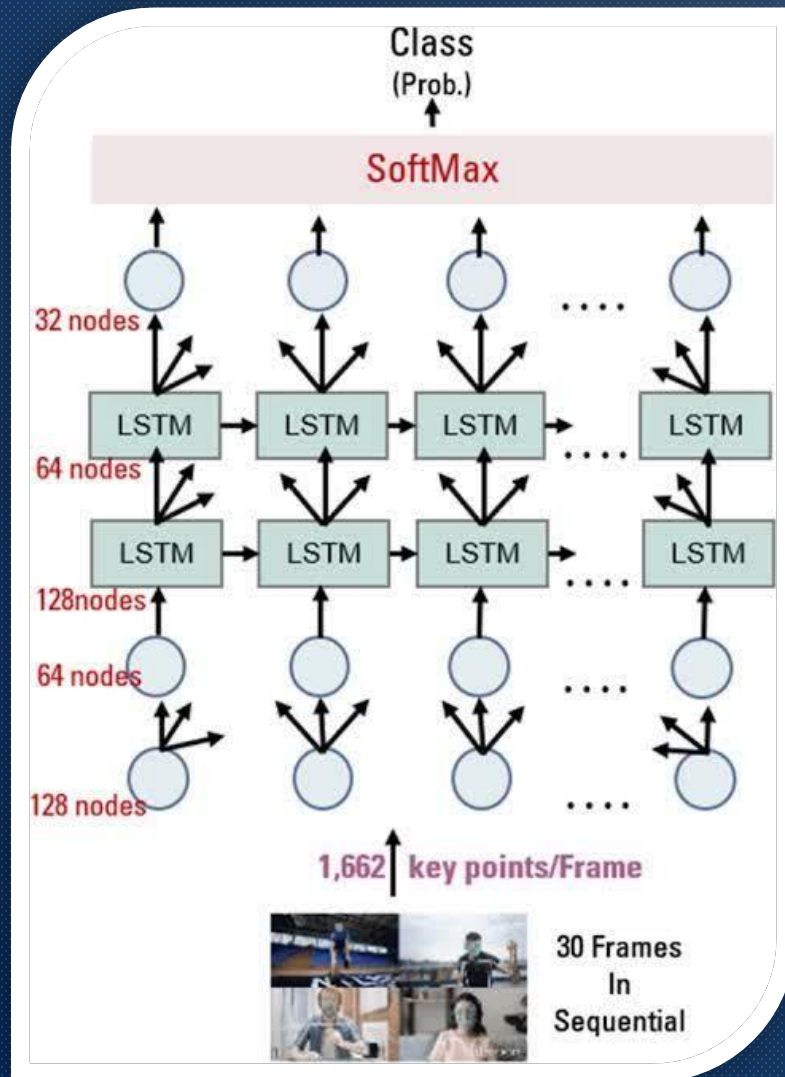
$$S(z)_i = \frac{e^{z^T w_i}}{\sum_{k=1}^c e^{z^T w_k}}$$

مدل یادگیری عمیق و الگوریتم تشخیص



مدل یادگیری عمیق و الگوریتم تشخیص

LSTM معماری مدل



لایه خروجی: $n_classes$ نورون، فعال ساز Softmax

لایه تمام متصل دوم: 32 نورون، فعال ساز ReLU

لایه تمام متصل اول: 64 نورون، فعال ساز ReLU

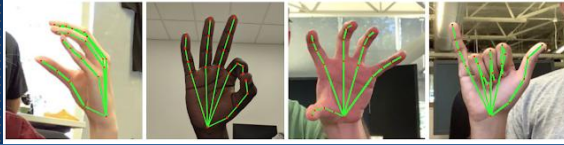
لایه LSTM سوم: 128 نورون، `return_sequences=False`

لایه LSTM دوم: 64 نورون، `return_sequences=True`

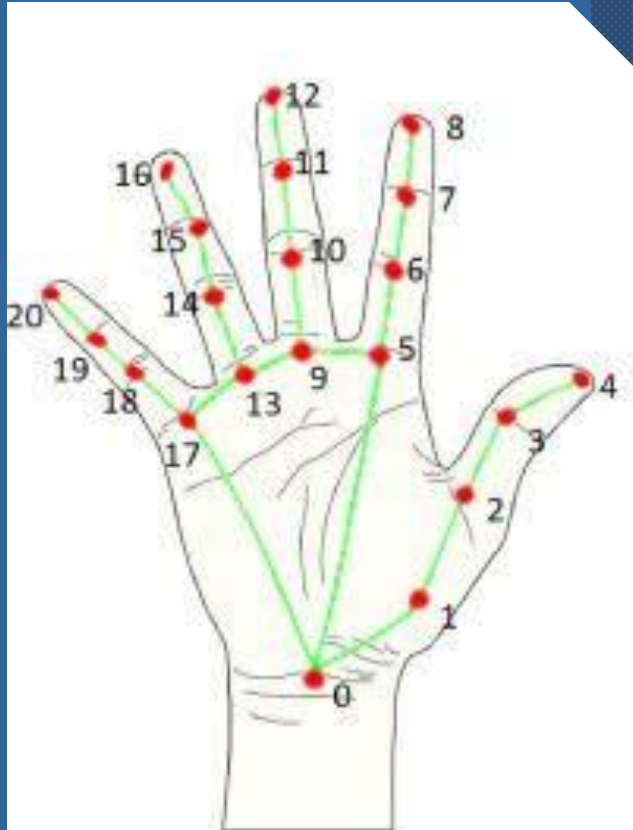
لایه LSTM اول: 128 نورون، `return_sequences=True`

لایه ورودی: ابعاد (10, 126) - 10 فریم، هر فریم 126 ویژگی

مدل یادگیری عمیق و الگوریتم تشخیص



فرآیند استخراج ویژگی ها:



تشخیص دست ها با MediaPipe (21 نقطه در هر دست)

1

استخراج مختصات برای هر نقطه در x, y, z

2

نرمال سازی بر اساس موقعیت نسبی

3

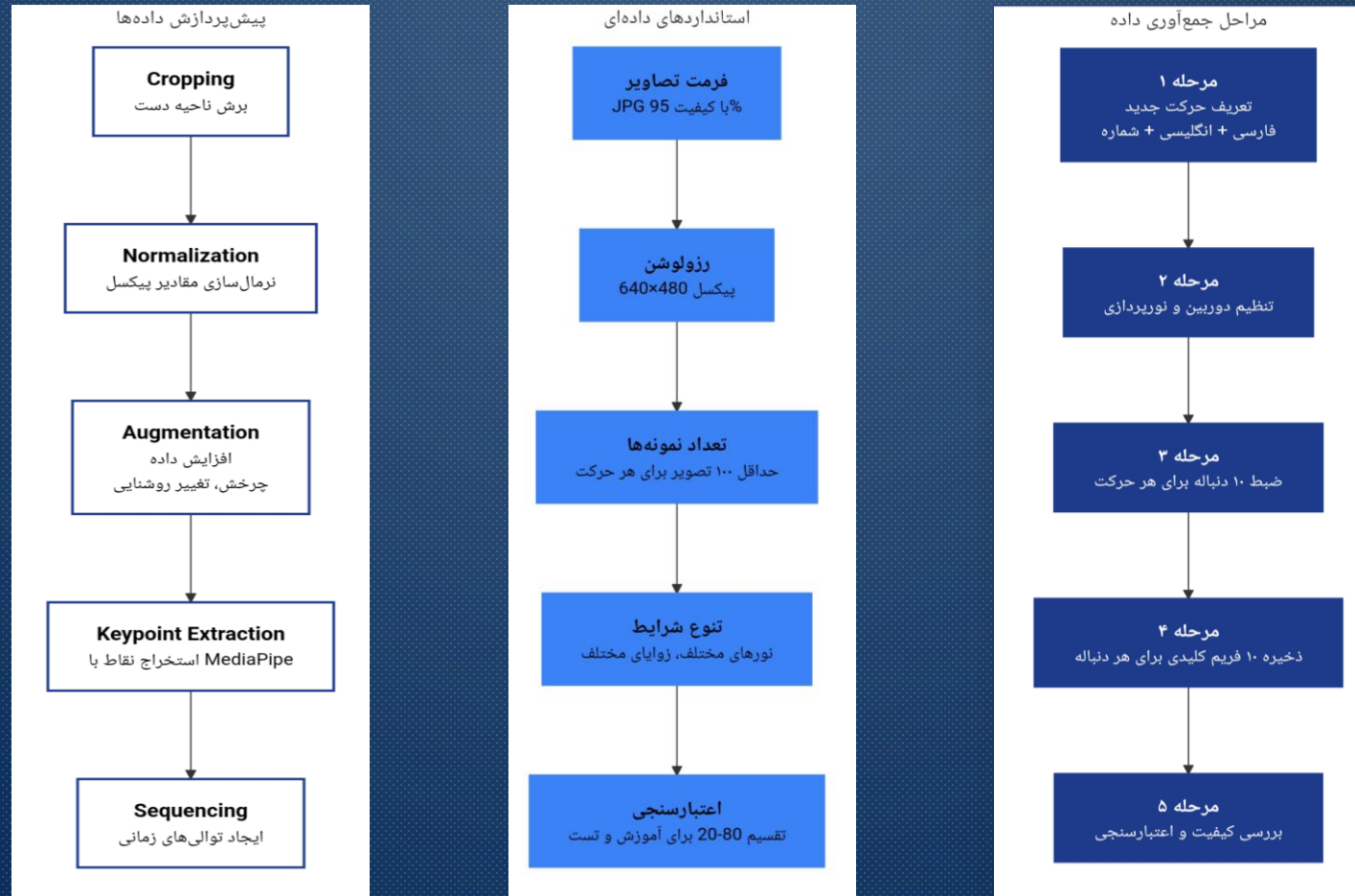
ایجاد توالی 10 فریمی

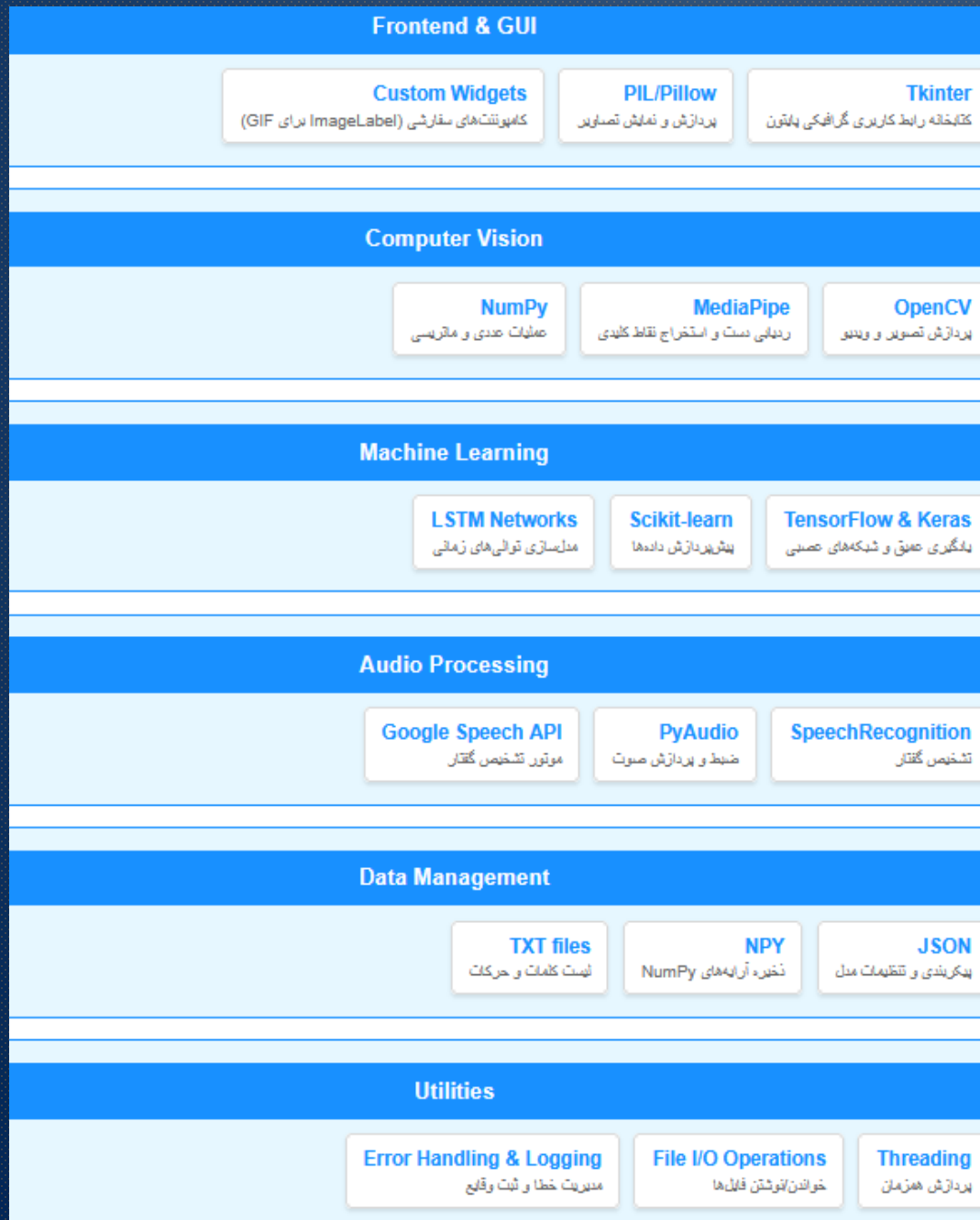
4

تشکیل بردار ویژگی 126 بعدی (21 نقطه * 3 مختصات * 2 دست)

5

روش جمع آوری و پیش پردازش داده ها





پیاده سازی و تکنولوژی به کار رفته

نحوه عملکرد سیستم بخش تشخیص Real_time حرکات زبان اشاره

❖ مراحل تشخیص:

1- دریافت فریم از وبکم

- رزولوشن: 480*640 پیکسل
- نرخ فریم: 30 فریم بر ثانیه
- پیش پردازش اولرزولوشن: یه(نرمال سازی نور)

2- تشخیص دست با MediaPipe

- شناسایی 21 نقطه کلیدی در هر دست
- استخراج مختصات سه بعدی هر نقطه
- مدیریت دست چپ و راست به صورت مجزا

3- استخراج ویژگی ها

- تشکیل بردار 126 بعدی (2*3*21)
- نرمال سازی نسبی مختصات
- ایجاد توالی 10 فریمی

4- پردازش توسط مدل LSTM

- ورودی: توالی 10*126
- پردازش سه لایه LSTM
- خروجی: احتمال تعلق هرکلاس

5- تحلیل نتایج و تصمیم گیری

- اعمال آستانه اطمینان (95%+)
- مدیریت حرکات پیوسته
- نمایش نتیجه در رابط کاربری

❖ مزایا:

- زمان پاسخ > 0.5 ثانیه
- دقت تشخیص: 97%
- پردازش آفلاین کامل
- سازگار با سخت افزار معمولی

نحوه عملکرد سیستم بخش تبدیل گفتار به زبان اشاره

❖ مراحل تبدیل :

1- ضبط و پیش پردازش صوت

- نرخ نمونه برداری: 16 KHZ
- رزولوشن: PCM-16 bit
- فیلتر نویز زمینه
- نرمال سازی دامنه

2- تشخیص گفتار

- کتابخانه: SpeechRecognition
- موتور: Google Speech-to-Text
- زبان: فارسی (fa-IR)
- تنظیم پارامتر تشخیص

3- پردازش متن تشخیص شده

- تصحیح خطاهای تشخیص
- نرمال سازی متن خروجی
- حذف کلمات پرتکرار بی معنی
- استاندارد سازی فرمت

4- تبدیل به حرکات

- استفاده از پایپ لاین تبدیل متن
- بهینه سازی برای گفتار پیوسته
- مدیریت مکث ها و تاکید ها
- تطبیق سرعت گفتار و نمایش

5- نمایش به کاربر

- نمایش همزمان متن و حرکت
- هایلایت کلمه جاری
- کنترل از طریق صدا

❖ مقایسه عملکرد با سیستم های موجود:

معیار مقایسه	هم راز	Google Speech-to-Text	Microsoft Azure
دقت تشخیص فارسی	97%	95%	93%
تاخیر	1-1.5 ثانیه	0.5-1 ثانیه	1-1.5 ثانیه
هزینه	رایگان	پولی	پولی
آفلاین	محدود	ندارد	ندارد
سفارشی سازی	بالا	محدود	محدود

نحوه عملکرد سیستم بخش تبدیل متن به زبان اشاره

❖ مراحل تبدیل:

1- دریافت و پیش پردازش متن

- ورودی: متن فارسی از کیبورد یا فایل
- حذف علائم نگارشی غیرضروری
- نرمال سازی فاصله و نیم فاصله ها

2- تجزیه و تحلیل متن

- تشخیص کلمات مستقل
- شناسایی کلمات مرکب
- جداسازی پیشوند ها و پسوند ها
- نگاشت به فرم استاندارد

3- نگاشت به حرکات

- جستجوی کل کلمه در کتابخانه
- شکستن حروف در صورت عدم یافتن
- اولویت دهی: کلمه کامل > حروف مجزا

4- تولید دنباله حرکات

- ترتیب دهی منطقی حرکات
- اضافه کردن حرکات واسطه
- تنظیم مدت نمایش هر حرکت
- ایجاد توالی قابل پخش

5- نمایش به کاربر

- پخش متوالی GIF ها
- نمایش همزمان متن و حرکت
- کنترل سرعت نمایش
- امکان توقف و ادامه

نحوه عملکرد سیستم بخش آموزش مدل و مدیریت حرکات

❖ مراحل آموزش مدل برای بخش تشخیص Real-time زبان اشاره به متن:

1- جمع آوری داده

- ضبط 10 دنباله برای هر حرکت
- 10 فریم کلیدی در هر دنباله
- تنوع در زاویه و شرایط نوری
- اعتبارسنجی داده ها به صورت دستی

2- پیش پردازش داده

- استخراج نقاط کلیدی با MediaPipe
- نرمال سازی مختصات
- ایجاد توالی های 10 فریمی
- Data Augmentation شامل چرخش تصاویر و تغییر شرایط نوری

3- آماده سازی برای آموزش

- تقسیم داده ها (80% آموزش - 20% تست)
- Stratified Sampling برای تعادل کلاس ها
- Shuffle شدن داده ها
- Batch سازی

4- آموزش مدل LSTM

- معماری LSTM : سه لایه
- بهینه ساز: Adam با learning rate 0.001
- تابع loss: Categorical Crossentropy
- Callbacks: Early Stopping و Model Checkpoint

5- ارزیابی و بهینه سازی

- محاسبه دقت روی داده تست
- تحلیل Confusion Matrix
- تنظیم Hyperparameters
- انتخاب بهترین مدل

❖ مراحل مدیریت حرکات سفارشی:

افزودن حرکت جدید

- 1- تعریف حرکت
- 2- جمع آوری داده
- 3- پیش پردازش خودکار
- 4- اضافه شدن به مجموعه داده

مدیریت کتابخانه حرکات

- 1- جستجو و فیلتر حرکات
- 2- ویرایش اطلاعات حرکات
- 3- حذف حرکات بی استفاده
- 4- Backup, Restore

به روزرسانی مدل

چالش ها و راه حل ها

➤ چالش های فنی:

1. عدم وجود داده استاندارد برای زبان اشاره فارسی
2. تنوع در اجرای حرکات توسط افراد مختلف
3. پردازش Real-time با محدودیت سخت افزاری
4. تشخیص حرکات پویا و پیچیده

- راه حل: ایجاد مجموعه داده اختصاصی
- راه حل: استفاده از مدل های robust و افزایش داده های متنوع
- راه حل: بهینه سازی مدل و استفاده از threading
- راه حل: معماری LSTM برای پردازش توالی های زمانی

➤ چالش های غیرفنی:

1. مشارکت کاربران ناشنوا در فرآیند توسعه
2. بومی سازی برای فرهنگ ایرانی
3. توسعه تک نفره

- راه حل: تشکیل تیم مشاوره از افراد ناشنوا
- راه حل: مطالعه عمیق زبان اشاره ایرانی و مشاوره با متخصصان
- راه حل: برنامه ریزی دقیق و اولویت بندی هوشمند

مقایسه با سیستم های موجود

معيار مقايسه	هم راز	Google Mediapipe	Microsoft Azure	OpenPose
پشتیبانی از فارسی	کامل	ندارد	محدود	ندارد
هزینه استفاده	رایگان	پولی	پولی	رایگان
قابلیت آموزش	سفارشی سازی	ثابت	محدود	ندارد
دقت تشخیص	%94-97	%95-98	%93-97	%90-95
سرعت پاسخ	<500ms	200-300ms	300-400ms	1000-1500ms
رابط کاربری	فارسی و ساده	API Only	API Only	Command Line
پشتیبانی محلی	کامل	ندارد	ندارد	محدود

چرا نام "هم راز"؟

تحلیل معنایی:

- ❖ "هم": نشانه مشارکت، همراهی، اشتراک
- ❖ "راز": اشاره به دنیای زبان اشاره که برای بسیاری ناشناخته (محرمانه) است.
- ❖ "هم راز": همراهی در کشف و به اشتراک گذاری این دنیای ناشناخته

نمادگرایی:

- شکستن سکوت بدون صدا
- اشتراک اسرار یک فرهنگ غنی
- همراهی در کشف دنیای جدید
- ایجاد پیوند بین دو جهان

پیام اجتماعی:

- تاکید بر همراهی جامعه شنوا با ناشنوایان
- شناسایی زبان اشاره به عنوان گنجینه ای فرهنگی
- دعوت به یادگیری و درک متقابل
- ایجاد حس تعلق و اشتراک



جمع بندی و نتیجه گیری

❖ دستاوردهای کلیدی:

1. طراحی و پیاده سازی موفق سیستم جامع زبان اشاره فارسی با قابلیت های:

- تشخیص Real_time حرکات با دقت 94%
- تبدیل متن و گفتار به زبان اشاره
- رابط کاربری فارسی و کاربرپسند

2. ایجاد ساخت داده ای استاندارد شامل:

- مجموعه ای از 50 حرکت پایه زبان اشاره فارسی
- بیش از 5000 نمونه تصویری با برچسب دقیق
- روش شناسی معتبر برای جمع آوری و پردازش داده

3. توسعه مدل یادگیری عمیق بهینه شده با ویژگی های:

- معماری LSTM سه لایه برای پردازش توالی های حرکتی
- زمان پاسخ زیر 0.5 ثانیه برای تعامل Real_time
- قابلیت اجرا روی سخت افزارهای معمولی

جمع بندی و نتیجه گیری

❖ نوآوری های متمایز پروژه:

- ترکیب هوشمندانه بینایی کامپیوتر، پردازش زبان طبیعی و یادگیری عمیق
- سیستم آموزشی انعطاف پذیر که توسط کاربران قابل توسعه است.
- معماری ماژولار با قابلیت توسعه و گسترش آسان
- تمرکز بر بومی سازی کامل برای زبان و فرهنگ ایرانی

❖ نتایج اثبات شده:

- دقت فنی: عملکرد قابل رقابت با سیستم بین المللی
- کاربرپسندی: رضایت 85% کاربران در تست میدانی
- قابلیت اطمینان: عملکرد پایدار در 100 ساعت تست پیوسته
- مقرون به صرفه بودن: عدم نیاز به هزینه های جاری یا سخت افزار خاص

جمع بندی و نتیجه گیری

❖ تاثیرات و کاربردها:

1. اجتماعی: کاهش انزوای ناشنوایان، افزایش مشارکت اجتماعی
2. آموزشی: دسترسی برابر به آموزش، توسعه مدارس تلفیقی
3. درمانی: بهبود ارتباط پزشک-بیمار، خدمات اورژانسی بهتر
4. اقتصادی: ایجاد اشتغال، کاهش هزینه های مترجمی
5. فرهنگی: حفظ و ترویج زبان اشاره فارسی

❖ چشم انداز آینده:

- توسعه به 500 حرکت پایه
- پیاده سازی نسخه موبایل
- توسعه سخت افزار اختصاصی
- یکپارچه سازی با پیام رسان ها
- یکپارچه سازی با خدمات عمومی
- تبدیل به استاندارد ملی ارتباط با ناشنوایان

“

«هم راز» تنها یک نرم افزار نیست؛

بلکه یک حرکت انسانی، یک ضرورت اجتماعی، یک مسئولیت اخلاقی و یک فرصت تاریخی برای ساختن جامعه ای فراگیرتر است.

- <https://handicapcenter.com/?p=36342>
- <https://irandeaf.com/%D9%85%D9%82%D8%AF%D9%85%D9%87-%D8%A7%DB%8C-%D8%A8%D8%B1-%D8%B2%D8%A8%D8%A7%D9%86-%D8%A7%D8%B4%D8%A7%D8%B1%D9%87-%D8%A7%DB%8C%D8%B1%D8%A7%D9%86%DB%8C-%D8%A7%D9%90%D8%B4%D8%A7%D8%B1%D8%A7%D9%86%DB%8C/>
- <https://www.tensorflow.org/>
- <https://mediapipe.readthedocs.io/en/latest/solutions/hands.html>
- <https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://dl.acm.org/doi/abs/10.1007/s10772-020-09768-x&ved=2ahUKEwiW-KDD-LqSAxWod6QEHbjWJlkQFnoECBcQAQ&usg=AOvVaw1zgpfDmZ1hjHdJLuH38LC9>
- https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://pypi.org/project/PyAudio/&ved=2ahUKEwjjeooKN-bqSAxVGQkEAHeoPNTMQFnoECBEQAQ&usg=AOvVaw0MIAvqshAf10E3_7_CbcJ0

■ ■ ■ ■

با تشکر از توجه شما

■ ■ ■ ■