راهنماي‌تهيه ‌مقاله‌های فارسی سیزدهمین کنفرانس ملی و سومین کنفرانس بین‌المللی بینایی ماشین و پردازش تصویر ایران

نويسنده اول1، نويسنده دوم2 و نويسنده سوم3

1سازمان متبوع نويسنده اول، FirstAuthor@Email

2سازمان متبوع نويسنده دوم، SecondAuthor@Email

چکيده - تشخیص هوا نوشته با توجه به کاربردهای احتمالی در سیستم‌های هوشمند، توجه گسترده‌ای را به خود جلب کرده است. با وجود این توجه گسترده، چالش‌های بنیادی خاص نوشتن هوا هنوز به نحو کافی حل نشده‌اند. این مقاله یک رویکرد ساده اما کارآمد به تشخیص نوشتن هوا با استفاده از شبکه‌های عصبی کانولوشنی عمیق(CNNs) معرفی می‌کند. یک الگوریتم قوی و ساده شده برای ردیابی دست پیشنهاد شده است تا الگوهای ردیابی شده نوشتن هوا با استفاده از یک دوربین وب تکی ضبط شوند. این الگوریتم به طور خاص برای حل مشکل push-to-write طراحی شده است و نیاز به محدودیت‌های نوشتن توسط کاربر را از بین می‌برد، بدون نیاز به یک جداکننده یا جعبه تصویری. به علاوه، یک طرح پیش‌پردازش نوآورانه ارائه شده است تا الگوهای ردیابی شده نوشتن را به یک فرمت مناسب داده تبدیل کند، که سادگی و کارآمدی CNNs آموزش دیده شده با این داده‌های تغییر یافته را افزایش می‌دهد. نتایج آزمایشی نشان می‌دهد که رویکرد پیشنهادی ما نه تنها دقت تشخیص بالاتری را به دست می‌آورد بلکه همچنین پیچیدگی شبکه را به طور قابل توجهی نسبت به روش‌های مبتنی بر تصویر کاهش می‌دهد.

كليد واژه- تشخیص نوشتن هوا، شبکه عصبی کانولوشن (CNN)، ردیابی دست.

# مقدمه

شکل 1:نمای کلی از مراحل تعامل انسان با کامپیوتر از طریق هوا نویسی

تشخیص حرکات دست، یک جزء کلیدی از تعامل انسان و کامپیوتر (Human-Computer Interaction) است که در سال‌های اخیر توجه زیادی را به خود جلب کرده است، به ویژه در برنامه‌هایی مانند تشخیص نوشتن در هوا. تشخیص هوا نوشته ها، یک نوع از حرکات دست یا انگشت در هوا برای ترسیم حروف یا ارقام است و کاربردهای عملی متعددی دارد، از جمله ورود متن به سیستم های کامپیوتری بدون تماس دست با سخت افزارها، اگر چه مزایای زیادی برای تشخیص هوا نوشته­ها می توان تعریف کرد اما یکی از مزایای آن را می توان پیشگیری از انتشار ویروس‌ها مانند COVID-19 نام بردشکل1.

برخلاف نوشتن سنتی، که در آن حرکت قلم اطلاعات را منتقل می‌کند، نوشتن در هوا بر اساس مسیر حرکت دست یا انگشت برای بیان حروف یا اعداد مورد نظر تکیه دارد [1]. با توجه به اینکه روش‌های ورودی سنتی مانند کیبوردها و صفحه های ‌لمسی محدودیت‌هایی دارند، هر روز HCI نقش بیشتری در دنیای دیجیتال ما ایفا می‌کند، به ویژه برای افراد دارای نقص بینایی و شرایطی نظیر نور کم، ورود کاراکتر دیجیتال به صورت لحظه ای در محیط HCI ضروری است، اما نیاز به رفع چالش‌های تبدیل محتوای دست‌نوشت به فرمت‌های دیجیتال دارد. به همراه داشتن قلم و کاغذ نیز می‌تواند توسط روش‌های دسترسی‌پذیرتر و سریع‌تر کاهش یابد، که نوشتن در هوا یک راه‌حل ساده تر به نظر می­آید[2]. نوشتن در هوا به عنوان یک حرکت پویا برای تعامل انسان و کامپیوتر ظاهر شده است، که ارتباط طبیعی با سیستم‌های هوش مصنوعی، از جمله تلویزیون‌های هوشمند و ربات‌ها، را ممکن می‌کند. برخلاف روش‌های مبتنی بر دستگاه که به دستگاه‌های دستی یا پوشیدنی برای گرفتن حرکات دست وابسته هستند، رویکردهای بدون دستگاه مانند رویکردهای مبتنی بر دید و رادیو، برای کاربران راحتی و دسترسی بیشتری ارائه می‌دهند [3].

نوشتن در هوا، یک زیرشاخه از تشخیص حرکات است که در برنامه‌هایی مانند تعامل با موسیقی، رباتیک و ترجمه زبان اشاره، ارائه وسیله‌ای مناسب و طبیعی برای انتقال اطلاعات در زمینه‌های مرتبط با HCI است. این شامل ایجاد حروف یا کلمات زبانی در یک فضای سه‌بعدی آزاد از طریق حرکات دست یا انگشت است، که با عدم وجود نیاز به تشخیص ضربه‌های گسسته (فشار دست یا انگشت) و وابستگی بین حرکات، نوشتن سنتی را دچار چالش می کند [4].

چالش‌های مختلفی در روش‌های بینایی ماشین برای ردگیری دست و یا انگشت وجود دارد که در سنسور‌های تصویری دو بعدی و سه بعدی متمایز می‌شوند.]3[ در دنیای سیستم‌های مبتنی بر دوربین 2D، بهبود عملکرد ردگیری معمولاً شامل استفاده از نشانگرهای رنگی بر روی انگشتان می‌شود، زیرا ردگیری انگشتان بدون نشانگر چالش قابل توجهی ایجاد می‌کند. برعکس، سیستم‌های مبتنی بر دوربین 3D مسائل مربوط به ردگیری دست و انگشت را به خوبی با استفاده از اطلاعات عمق ارائه شده توسط سنسور‌های تصویری 3D مانند Kinect، Leap Motion Controller (LMC)، یا دوربین‌های Intel RealSense حل می‌کنن1.

با این حال، نوشتن در هوا به عنوان یک روش ورودی مبتنی بر ژست منحصر به فرد چالش‌های خود را به دلیل عدم وجود قالب مرجع در صفحه نوشتن ایجاد می‌کند که به عدم وجود نقاط شروع و پایان واضح برای خطوط منجر می‌شود. این چالش تقسیم‌بندی را ایجاد می‌کند، یا مشکل "فشار برای نوشتن"، جایی که سیستم باید به طور خودکار مختصات شروع و پایان حروف نوشته شده در هوا را شناسایی کند. برخی از راه‌حل‌های پیشنهادی شامل استفاده از اندام‌های خاص دست برای نمایش مختصر انتهای عمل نوشتن مانند قرار گرفتن دست به صورت مشت است، اما این موضوع پیچیدگی و حفظ موقعیت‌ها را برای کاربران افزایش می‌دهد. برعکس، هنگامی که اطلاعات عمق در دسترس هستند، تقسیم‌بندی اعمال نوشتاری با استفاده از یک آستانه عمق به آسانی قابل انجام است. به طور خلاصه، سیستم‌های مبتنی بر دوربین 3D راه‌حل‌های ظریف‌تری را برای دو مشکل اول در مقایسه با همتایان 2D خود ارائه می‌دهند، هرچند که با افزایش پیچیدگی و هزینه همراه هستند و همچنین استفاده از یک جعبه مجازی به کاهش محدوده نوشتن و کاهش تغییرپذیری در ورودی حروف کمک می‌کند، که پردازش بعدی را ساده‌تر می‌کند اما کار کاربران را سخت تر می‌کند.]4[ شکل شماره 1 ایده کلی این مقاله را نشان می دهد.

در این مقاله، ما یک رویکرد جدید تشخیص هوانویسی مبتنی بر شبکه‌های عصبی کانولوشنال عمیق (CNN) با استفاده از یک دوربین وب (Webcam) دوبعدی مقرون‌به‌صرفه معرفی می‌کنیم که به طور موثر به مشکلات موجود رسیدگی می‌کند و می تواند از برنامه‌های کاربردی بلادرنگ در تلویزیون‌های هوشمند نیز پشتیبانی می‌کند.

مشارکت‌های کلیدی ما شامل:

(الف) یک الگوریتم قوی برای جمع‌آوری مسیر نوشتن در هوا که از یک دوربین وب 2D استفاده می‌کند و مدل تشخیص دست و انگشتان را با استفاده از کتابخانه MediaPipe [[1]](#footnote-1)را به کار می‌گیرد تا مختصات 21 نقطه تفکیک شده و سه بعدی کف دست را به دقت تشخیص و مکان‌یابی کند، از پیچیدگی‌های ردگیری انگشت‌ها عبور کرده و مسئله فشار به نوشتن را بدون وابستگی به محدودکننده‌ها یا جعبه‌های تصوری حل می‌کند، بنابراین به کاربران آزادی نامحدودی در نوشتن در هوا می‌دهد.

(ب) یک طرح پیش‌پردازش داده جدید که توالی مختصات x و y مسیر نوشتن را نرمالیزه می‌کند و آن‌ها را به آرایه‌های یک بعدی و دو بعدی ترکیب می‌کند. این آرایه‌های داده به آموزش مدل‌های 1D-CNN و 2D-CNN کمک می‌کنند. داده‌های EMNIST از شبکه‌های عصبی کانولوشنالی (CNN) استفاده می‌کنند که طراحی شده‌اند تا به صورت خودکار سلسله‌مراتب فضایی ویژگی‌ها را از تصاویر ورودی با استفاده از فیلترهای کانولوشنال یاد بگیرند، که باعث ساده‌تر شدن شبکه‌های عصبی کانولوشنالی و افزایش کارآیی آن‌ها نسبت به تصاویر پیچیده می‌شود.]3[

بخش های باقی‌مانده این مقاله به شرح زیر است. بخش دوم به بررسی کارهای پیشین مرتبط می‌پردازد. بخش سوم روش پیشنهادی را به تفصیل شرح می‌دهد. نتایج آزمایشی در بخش چهارم ارائه شده‌اند. در نهایت، نتیجه‌گیری در بخش پنجم ارائه شده است.

# کارهای پیشین

تشخیص هوانویسی، که همچنین به عنوان تشخیص متن هوایی یا تشخیص نقاشی هوایی شناخته می‌شود، عمل نوشتن یا کشیدن در فضای آزاد با استفاده از حرکات دست و سپس تشخیص حروف یا کلمات نوشته یا کشیده شده است.

تحقیقات قبلی در زمینه تشخیص هوانویسی، زیربنایی را برای رویکرد مورد استفاده در این مطالعه فراهم کرده است. پیشرفت‌های اولیه در تشخیص دست‌نوشته، مانند کار انجام شده توسط SRI در اواخر دهه ۱۹۵۰، زمینه‌ای را برای تشخیص حروف دست‌نوشته ایجاد کرد و در نهایت به تکامل سیستم‌های تشخیص هوانویسی کمک کرد.]5[ چالش‌های تشخیص هوانویسی، از جمله تشخیص و ردیابی حرکات دست، و همچنین تغییرات در حالت، موقعیت، مقیاس و چرخش، در متون تخصصی به رسمیت شناخته شده‌اند.]6[ رویکردهای مختلفی، مانند استفاده از بینایی کامپیوتر، ردیابی شیء و تکنیک‌های یادگیری ماشینی برای تشخیص حروف در سیستم‌های هوانویسی، مورد بررسی قرار گرفته‌اند. علاوه بر این، تحقیقاتی در زمینه مدل‌سازی و تشخیص حروف، کلمات و حرکات پیونددهنده براساس داده‌های حرکات دست انجام شده است که نشان‌دهنده پتانسیل یادگیری ماشینی و شبکه‌های عصبی کانولوشنال در تشخیص هوانویسی است.] 5،7،8[

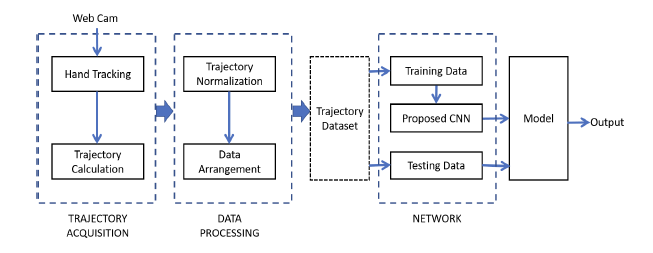
تحقیقات اخیر همچنین با استفاده از تخمین ژست دست و مدل‌های ترکیبی یادگیری عمیق، به حوزه تشخیص دست‌نوشته هوایی مبتنی بر دوربین ۲ بعدی پرداخته‌اند که باعث پیشرفت بیشتر در زمینه تعامل انسان-رایانه از طریق فناوری دست‌ نوشته هوایی شده است. نتیجه‌گیری مقاله فؤاد العبیر و همکارانش این است که روش پیشنهادی تشخیص دست‌نوشته هوایی به طور قابل توجهی از همه روش‌های پیشرفته موجود برای هر دو اصل آموزش وابسته به کاربر و مستقل از کاربر در تمام مجموعه‌های داده پیشی می‌گیرد. و همچنین در مقایسه با روش‌های موجود، مدل پیشنهادی بسته به مجموعه داده و اصل آموزش، بین 0.52٪ تا 3.55٪ افزایش دقت را ارائه می‌دهد.]9[

چالش های موجود در تشخیص هوانوشته، از جمله تشخیص و ردیابی حرکات دست، و همچنین تغییرات در حالت، موقعیت، مقیاس و چرخش، در زبان شناسی به رسمیت شناخته شده است. علاوه بر این، استفاده از تکنیک های یادگیری عمیق و انتخاب روش های درون یابی مناسب برای افزایش دقت و تعمیم پذیری سیستم های تشخیص نگارش هوایی مورد بررسی قرار گرفته است که راه را برای پیاده سازی‌های قوی‌تر و چند منظوره هموار می‌کند. مطالعه تایکی واتانابه و همکارانش بر تشخیص کاراکتر با استفاده از دو مجموعه داده تمرکز کرد: یک مجموعه داده حروف و یک مجموعه داده عددی. ویژگی ها از داده های تصویر با استفاده از CNN و از داده های سری زمانی با استفاده از BiLSTM استخراج شدند. مدل ترکیبی CNN-BiLSTM با استفاده از پنج برابر اعتبارسنجی متقابل آموزش دید. سیستم پیشنهادی به دقت تشخیص بالایی دست یافت، با 99.3٪ برای تشخیص حروف و 99.5٪ برای تشخیص اعداد. مقایسه با مطالعات موجود با استفاده از وب‌کم‌ها نشان داد که سیستم پیشنهادی عملکرد بهتری نسبت به سایرین دارد و موفقیت خود را به استفاده از گرافیتی و ترکیب CNN و BiLSTM نسبت می‌دهد. اعتبارسنجی در مجموعه داده 6DMG عملکرد برتر سیستم را بیش از پیش تأیید کرد و به دقت 99.48٪ برای تشخیص حروف و 99.17٪ برای تشخیص اعداد دست یافت و از روش‌های موجود فراتر رفت. به طور کلی، نتیجه‌گیری به شدت ادعا می‌کند که روش پیشنهادی استاندارد جدیدی برای دقت و تطبیق‌پذیری تشخیص نگارش هوایی تعیین می‌کند.]10[

چندین مطالعه دیگر ارائه شده است که اثربخشی تکنیک های مبتنی بر نشانگر را نشان می دهد. اوکا و همکاران ]11[ از یک دستگاه حسگر پیچیده برای ردیابی نوک انگشت استفاده کرد و به نرخ تشخیص 97.7 درصد دست یافت. روی و همکاران]12[ از یک نشانگر رنگ ثابت استفاده کرد که ردیابی و تقسیم بندی را با نرخ تشخیص 97.7٪ برای اعداد انگلیسی ساده می کند. رحمان و همکاران]13[ ردیابی نشانگر را با کالیبراسیون و شبکه‌های دوگانه RNN-LSTM بهبود بخشید و به نرخ‌های تشخیص بالای 98.75 درصد برای تک رقمی و 85.27 درصد برای تشخیص چند رقمی دست یافت. مصرا و همکاران ]14[ یک طرح تشخیص حرکت دست با استفاده از یک نشانگر نوک انگشت قرمز ایجاد کرد و به نرخ تشخیص 96.95٪ برای 58 حرکت رسید. با این حال، انتقادات ذکر شده است: روش های مبتنی بر نشانگر محدودیت هایی را تحمیل می کنند و کاربرد در دنیای واقعی را محدود می کنند.

این مطالعه با تکیه بر این تحقیقات قبلی، با ادغام کتابخانه Media Pipe برای تشخیص حرکات دست و داده‌های EMNIST برای تشخیص حروف در اشکال نوشته‌شده در هوا، هدف دارد زمینه تشخیص هوانویسی را به جلوتر ببرد که در ادامه در مورد آن صحبت خواهیم کرد.

# روش پیشنهادی

روش هوانویسی که در شکل 2 مشخص شده است شامل سه مرحله کلیدی است: اکتساب مسیر، پردازش داده و شبکه. با استفاده از یک وب کم، یک دنباله تصویر گرفته می شود، و یک الگوریتم جدید ردیابی دست برای محاسبه مسیر حرکت انگشت در هوا استفاده می شود.

شکل 2:مراحل تشخیص هوانویسی

متعاقبا، داده های مسیر تحت پردازش و تبدیل به آرایه های دو بعدی قرار می گیرند. این نوع داده ها در مجموعه داده های مسیر برای آموزش آفلاین مدل های شبکه عصبی کانولوشنال (CNN)ادغام می شوند.

در مرحله پیش‌بینی، سیستم داده‌های بلادرنگ را از وب‌کم دریافت می‌کند، سعی می‌کند از داده‌های دریافتی تصویری ترسیم کند، و سپس رقم (یا نماد) را که کاربر با استفاده از مدل‌های آموخته شده قبلی نوشته است، پیش‌بینی می‌کند. سه مرحله اولیه سیستم پیشنهادی به شرح زیر است.

## مسیر‌یابی یا اکتساب مسیر

این واحد دو هدف را دنبال می کند: اول، گرفتن صفحه دو بعدی ایجاد شده توسط کاربر هنگام نوشتن در هوا، و دوم، مستندسازی توالی موقعیت یک خط، که به عنوان جهت نوشتن شناخته می شود. جهت با تغییر موقعیت انگشت اشاره در حال حرکت تعریف می شود. در نتیجه، شناسایی و دنبال کردن انشگت اشاره در حال حرکت در داخل توالی صفحه دو بعدی، عملکردهای حیاتی این واحد هستند.

مطالعه تشخیص و ردیابی انگشت اشاره تاریخچه‌ای طولانی دارد، اما دستیابی به استحکام و دقت مناسب، همچنان یک چالش باقی مانده است. در این مقاله، ما از کتابخانه MediaPipe برای تشخیص حرکت انگشت اشاره استفاده می‌کنیم و با ذخیره مقادیر مختصات X و Y، ردیابی مؤثر دست در حال حرکت را انجام می‌دهیم. کتابخانه پیشنهادی با استحکام و قابلیت عملکرد بلادرنگ خود مشخص می‌شود.

تشخیص دستی با استفاده از MediaPipe یک روش ساده و سریع است. اما به راحتی به دلیل تداخل تغییرات نور و تغییرات کاربران مختلف مستعد خطا است. برای حل تداخل شرایط نور، پیچیدگی دست، پیچیدگی پس زمینه، کیفیت دوربین، منابع محاسباتی، یک ویژگی متحرک در الگوریتم تشخیص دست MediaPipe گنجانده شده است]15[.

کتابخانه MediaPipe از ترکیبی از مدل‌ها و الگوریتم‌ها برای تشخیص حالت انگشت استفاده می‌کند. مدل اول یک مدل تشخیص کف دست است که حضور دست‌ها و چند نشانه کلیدی را در تصویر مشخص می‌کند. مدل دوم یک مدل نقطه عطف دست است که مختصات سه بعدی 21 نقظه مشخص دست را برای هر دست شناسایی شده پیش بینی می کند. مؤلفه سوم یک الگوریتم تشخیص ژست است که جهت و پیچش هر انگشت را بر اساس نشانه های دست تخمین می زند. الگوریتم تشخیص ژست نیز نتیجه را با مجموعه ای از توصیفات از پیش تعریف شده ژست مقایسه می کند]16[.

کتابخانه MediaPipe مشکلات تداخلی مانند انسداد، خود انسداد و تاری حرکت را با استفاده از یک الگوریتم ردیابی دست سبک که مکان دست‌ها را برای تشخیص نشانه‌های بعدی تعیین می‌کند، کنترل می‌کند. این الگوریتم از آستانه جعبه محدودIoU بین دست ها در فریم فعلی و آخرین فریم برای ردیابی حرکات دست استفاده می کند. اگر نمره اطمینان حضور دست از مدل نقطه عطف دست کمتر از یک آستانه خاص باشد، مدل تشخیص کف دست دوباره فعال می شود ]17[.

علاوه بر این، MediaPipe برای انطباق با ویژگی دست، در درجه اول از ترکیبی از تکنیک‌های بینایی کامپیوتری، از جمله یادگیری عمیق، برای ردیابی دست و تخمین وضعیت انگشت استفاده می‌کند. الگوریتم‌ها و مدل‌های دقیق استفاده شده توسط MediaPipe ممکن است در طول زمان با توسعه و بهبود کتابخانه فعال شوند. MediaPipe اغلب از یک معماری شبکه عصبی عمیق برای ردیابی دست استفاده می‌کند که می‌تواند موقعیت و جهت دست را در یک فریم ویدیویی مشخص تشخیص دهد. علاوه بر این، برای تخمین موقعیت انگشت، شبکه‌های عصبی کانولوشنال (CNN) معمولاً برای پیش‌بینی موقعیت یا زوایای مفاصل انگشتان استفاده می‌شوند. برای کاهش مشکلات تداخل و افزایش استحکام ردیابی دست و تخمین موقعیت انگشت، ممکن است از تکنیک‌هایی مانند تقویت داده‌ها، بهینه‌سازی مدل و الگوریتم‌های پس از پردازش استفاده شود. MediaPipe همچنین ممکن است از مجموعه داده های بزرگ و متنوعی برای آموزش شبکه های عصبی استفاده کند تا از تعمیم بهتر به اشکال، اندازه ها و جهت گیری های مختلف دست اطمینان حاصل کند.]18،20[

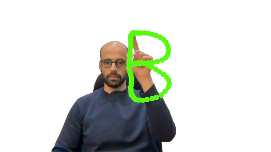
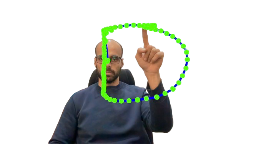
فرآیند ثبت تصویر اعداد یا حروف نوشته شده در هوا به شرح زیر است:

1. کاربر خود را در مقابل دوربین وب قرار می دهد. همانطور که سیستم کادر دوربین را برای تشخیص دست آماده می کند، مراحل نوشتن در هوا شروع می شود (شکل 3).
2. با بالا بردن دست و رسم کردن حروف یا اعداد با حرکت انگشت، سیستم، دست در حال حرکت را شناسایی می‌کند. قاب بلافاصله پس از تشخیص دست متحرک، شروع ترسیم یک خط بین نقاط X و Y انگشت اشاره را نشان می‌دهد.
3. پس از ناپدید شدن دست (بیرون بردن دست از کادر) و یا با استفاد از یک کلید میانبر مراحل ترسیم حرف یا عدد مورد نظر به پایان می رسد و تصویر نویسه نوشته شده نیز ثبت می‌شود.

در تمام فریم‌های بین شروع و پایان حرکت دست در کادر، مختصات مرکز انگشت اشاره ثبت می‌شود که مسیر حرف یا رقم را تشکیل می‌دهد (شکل ۳). این روش چالش فشار برای نوشتن را بدون نیاز به جداکننده برطرف می کند.



شکل 3:آماده سازی تشخیص کامل دست و فیلتر آن به انگشت اشاره



شکل 4: مختصات مرکز انگشت اشاره ثبت می‌شود

## پردازش داده

همانطور که قبلاً ذکر شد، داده‌های دست‌نویس را در یک تصویر ۲ بعدی (با اندازه اصلی ثبت شده) ذخیره می‌کنیم. کاربر به احتمال زیاد دستورات را در موقعیت‌های مختلف در هوا می‌نویسد. برای مقابله با تغییر مکان، تصویر ثبت شده را به تصویری با اندازه ۳۶۰ در ۳۶۰ تبدیل می‌کنیم که در وسط یک پنجره قرار دارد. همچنین برای خوانایی بهتر، ضخامت خطوط کشیده شده توسط کاربر را افزایش می دهیم. تصویر حاصل در شکل 4 نشان داده شده است.

تبدیل داده‌ها با استفاده از روابط زیر صورت می‌گیرد:

در معادلات فوق، و به ترتیب مختصات اصلی در محور x و y و همچنین و مختصات تبدیل شده هستند. xmax و ymax به ترتیب بیشینه مقادیر مختصات در محور x و y می باشند. هدف از ضریب 1.4\*r در معادلات ایجاد حاشیه‌های r0.2 در لبه‌های چپ و راست تصویر نهایی پس از ترسیم مختصات است. برای کاهش بار محاسباتی در مرحله آموزش، تصویر نرمال‌شده با اندازه ۳۶۰ در ۳۶۰ با استفاده از درون‌یابی خطی مجدداً به اندازه ۳۶ در ۳۶ تغییر اندازه داده می‌شود. مجموعه تصاویر تغییر اندازه داده شده برای پیاده‌سازی و مقایسه با رویکردهای رایج موجود استفاده می‌شود.

## تشخیص نویسه

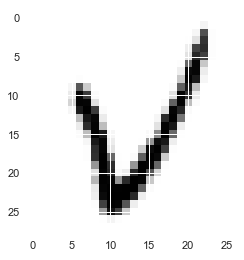
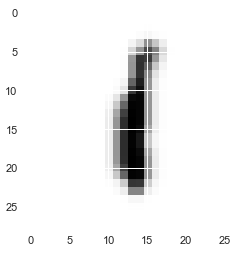
این پژوهش از پایگاه‌ داده‌ی تصویری حروف EMNIST استفاده می‌کند. این پایگاه‌ داده، توسعه‌ای از MNIST، پایگاه‌ داده‌ی مشهور ارقام دست‌نویس، است. مجموعه داده EMNIST-Balanced شامل مجموعه‌هایی از حروف با تعداد نمونه‌ی برابر در هر کلاس است و همچنین مجموعه داده‌ی EMNIST Letters حروف بزرگ و کوچک را به‌صورت متوازن در یک دسته‌بندی ۲۶ کلاسه ادغام می‌کند. قابل ذکر است مجموعه‌های داده‌ی EMNIST Digits و EMNIST MNIST شامل مجموعه‌های متوازنی از ارقام دست‌نویس هستند که با MNIST اصلی سازگاری دارند.

مجموعه داده‌ی EMNIST Letters با ادغام تمام کلاس‌های حروف بزرگ و کوچک به یک دسته‌بندی ۲۶ کلاسه‌ی متوازن، در تلاش است تا خطاهای ناشی از اشتباه در تشخیص حروف کوچک و بزرگ را کاهش دهد. به همین ترتیب، کلاس EMNIST Digits شامل یک زیرمجموعه‌ی متوازن از مجموعه داده‌های ارقام است که حاوی ۲۸۰۰۰ نمونه از هر رقم می‌باشد[20].

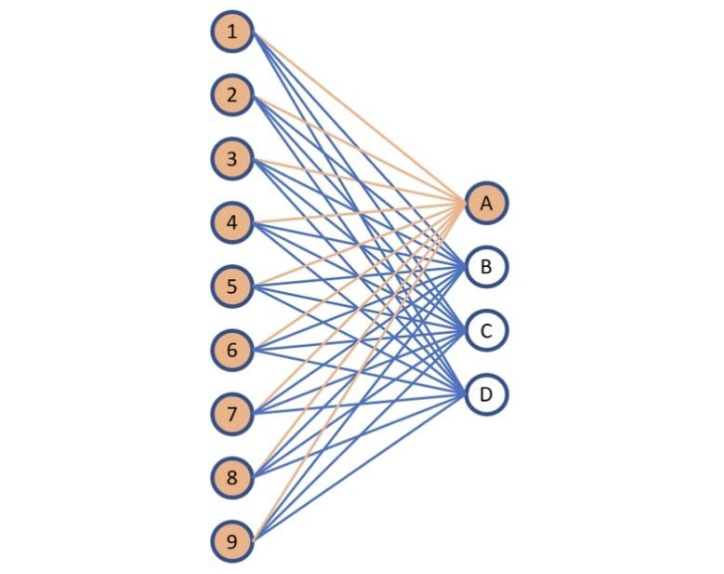
داده‌های تصویر صاف می‌شوند، به این معنا که هر تصویر ۲۸x۲۸ به یک آرایه ۷۸۴ (۲۸\*۲۸) عنصری تبدیل می‌شود شکل 5. این امر ضروری است زیرا از لایه‌های کاملاً متصل به جای لایه‌های کانولوشنال (Convolutional) که می‌توانند ورودی‌های ۲ بعدی را پردازش کنند استفاده می‌کنیم شکل 6.

داده‌ها برای کمک به فرآیند آموزش با تقسیم بر بیشینه مقدار پیکسل (۲۵۵) نرمال‌سازی می‌شوند. One-hot Encoding بر روی برچسب‌ها اعمال می‌شود و آن‌ها را به یک ماتریس باینری تبدیل می‌کند که برای طبقه‌بندی چند کلاسه ضروری است.

در نهایت یک مدل Keras برای چیدن لایه‌ها در شبکه ایجاد می‌شود. دو لایه متراکم (کاملاً متصل) با ۵۰۰ واحد/نورون هر کدام، هر دو با استفاده از توابع فعال‌سازی ReLU اضافه می‌شوند. یک لایه متراکم نهایی با تعداد واحدهایی که برابر با num\_classes است با یک تابع فعال‌سازی softmax برای خروجی استفاده می‌شود که برای طبقه‌بندی چند کلاسه مناسب است. مدل با استفاده از انتروپی متقاطع طبقه‌ای (cross-entropy) به عنوان تابع از دست رفته (loss function)، بهینه‌ساز ‘rmsprop’ و accuracy به عنوان معیار، کامپایل می‌شود شکل 7. ما مدل را برای 10 مرتبه با سایز دسته ای 128 آموزش داده ایم که در نهایت مدل بر روی مجموعه داده‌ی تست ارزیابی می‌شود.



شکل5: نمایش حروف در کادر

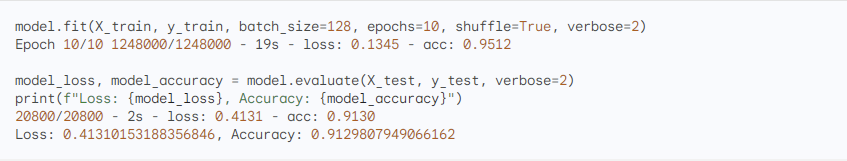


شکل 6(تصویر لایه های متراکم متصل (لایه های کاملاً متصل).)



شکل 7 مدل استفاده شده

این مدل با دقت ۹۵٪ آموزش دیده شده و در تست به دقت ۹۱٪ رسیده است. شکل 8



شکل 8:دقت مدل استفاده شده

# نتایج تجربی

در طول آزمایش‌ها، اولین تصویر به‌اشتباه پیش‌بینی شد، اما همان حرف با فونتی متفاوت به‌درستی پیش‌ بینی شد. به نظر می‌رسد تصویر وارد شده‌ی دوم شباهت بیشتری به مجموعه داده‌ی EMNIST دارد. این مدل سعی کرد تا با خواندن داده‌های دست‌نویس جدید (نوشته شده و وارد شده توسط نویسنده)، دقت مدل EMNIST ساخته‌شده و دست‌نویس واقعی را آزمایش کند. نتایج آنطور که در نظر گرفته شده بود، مطلوب نبودند. تصویری از هر حرف الفبا (فقط حروف بزرگ) گرفته و بارگذاری شد. چهار مجموعه وجود داشت: مداد، خودکار، ماژیک وهایلایتر. هدف این کار تعیین اینکه کدام ابزار نوشتاری بیشترین دقت را با مجموعه داده‌ی EMNIST خواهد داشت، بود؛ با این حال، تمام تصاویر وارد شده، کج بودند و چرخاندن آن‌ها باعث می‌شد تصاویر بسیار روشن‌تر و غیرقابل خواندن شوند و در نتیجه پیش‌بینی برای همه آن‌ها موفق انجام می‌شد . برای دستیابی به نتایجی دقیق‌تر، نیاز به گرفتن عکس‌های بیشتر و تبدیل پیکسل‌های بهتر از ۴ بعدی به ۲ بعدی می‌باشد. مدل EMNIST به طور بسیار خوبی کار می‌کند. مدل با تصاویر نوشته‌شده در هوا نیاز به آزمایش‌های بیشتری داشت، اما نتایج درستی را پیش‌بینی کرد.

|  |  |  |  |
| --- | --- | --- | --- |
| Trainable Parameters | Runtime | Accuracy | Models |
| 94,344 | 0.72 s | 98.5% | CNN |
| 726,310 | 0.94 s | 97.5% | CNN-LSTM |
| 367,810 | 0.82 s | 99.5% | LSTM |
| 367,810 | 0.85 s | 99% | BLSTM |
| 49,410 | 0.07 s | 97.5% | TCN-Dynamic |
| 66,910 | 0.09 s | 97% | TCN-Static |
| 827,374 | 1.21 s | 98% | LSTM&CNN |
| 814,255 | 1.12 s | 77% | MediaPipe & EMNIST |

# نتيجه‌گيري

# نتایج آزمایش‌ها بر ضرورت انجام آزمایش‌های گسترده و دقیق هنگام اعمال مدل‌های یادگیری ماشین به مسائل دنیای واقعی تأکید می‌کند. مدل فعلی عملکرد خوبی روی مجموعه داده‌ی EMNIST نشان می‌دهد، اما هنگامی که با داده‌هایی روبرو می‌شود که به طور قابل توجهی از ویژگی‌های مجموعه داده‌ی آموزشی منحرف می‌شوند، محدودیت‌هایی دارد. برای استقرار عملی، این مدل به یک خط لوله پیش‌پردازش جامع‌تر برای رسیدگی به مسائل کیفیت تصویر نیاز دارد و باید روی مجموعه داده‌ای که تنوع مشاهده‌شده در دست‌نویس دنیای واقعی را بهتر نشان می‌دهد، از جمله دست‌خطی که با ابزارهای نوشتاری مختلف در شرایط مختلف ایجاد شده است، آموزش داده شود یا حداقل اعتبارسنجی شود.

# در نتیجه، حرکت از آزمایشگاه به کاربرد، پیچیدگی‌های مختلفی را به وظیفه‌ی تشخیص دست‌نویس وارد می‌کند که کمترین آن‌ها، تغییرات دست‌خط انسان و شرایطی است که در آن ثبت می‌شود، نیست. آزمایش‌های ما نشان می‌دهد که در حالی که مدل سطح پایه خوبی از عملکرد را روی داده‌های استاندارد دارد، کار قابل توجهی برای آماده کردن آن برای طیف گسترده‌ای از ورودی‌هایی که خارج از مجموعه داده‌های کنترل‌شده مانند EMNIST با آن‌ها مواجه خواهد شد، باید انجام شود. برای رسیدگی به این موارد، یک رویکرد سیستماتیک شامل بازتعریف استراتژی‌های پیش‌پردازش، افزایش داده، پیچیدگی مدل و جمع‌آوری داده‌های دنیای واقعی برای اینکه مدل در کاربردهای عملی واقعاً مؤثر شود، ضروری است.

# این مطالعه به اهمیت آزمایش دقیق و جامع مدل‌های یادگیری ماشین قبل از استقرار آن‌ها در دنیای واقعی اشاره می‌کند. با توجه به اینکه دست‌خط انسان می‌تواند به طور قابل توجهی از داده‌های آموزشی استاندارد منحرف شود، مدل‌های یادگیری ماشین برای عملکرد مؤثر در شرایط واقعی نیاز به خطوط لوله پیش‌پردازش قوی‌تر و مجموعه داده‌های آموزشی جامع‌تر دارند. نتایج این مطالعه همچنین اهمیت تمرکز بر چالش‌های خاص دنیای واقعی مانند کیفیت تصویر و تنوع ابزارهای نوشتاری را برای ایجاد مدل‌های تشخیص دست‌نویس قوی و قابل اطمینان‌تر برجسته می‌کند.

# مراجع

[1] Shahzad Ahmed, Wancheol Kim, Junbyung Park, and Sung Ho Cho. Radar-based air-writing gesture recognition using a novel multistream cnn approach. IEEE Internet of Things Journal, 9(23):23869–23880, 2022.

[2] Hongyu Zhang, Lichang Chen, Yunhao Zhang, Renjie Hu, Chunjuan He,Yaqing Tan, Jiajin Zhang, et al. A wearable real-time character recognition system based on edge computing-enabled deep learning for air-writing. Journal of Sensors, 2022, 2022.

[3] Chaur-Heh Hsieh, You-Shen Lo, Jen-Yang Chen, and Sheng-Kai Tang. Air-writing recognition based on deep convolutional neural networks. IEEE Access, 9:142827–142836, 2021.

[4] Grigoris Bastas, Kosmas Kritsis, and Vassilis Katsouros. Air-writing recognition using deep convolutional and recurrent neural network architectures. In 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR), pages 7–12. IEEE, 2020.

[5] M. Chen, G. AlRegib and B. -H. Juang, "Air-Writing Recognition—Part I: Modeling and Recognition of Characters, Words, and Connecting Motions," in IEEE Transactions on Human-Machine Systems, vol. 46, no. 3, pp. 403-413, June 2016, doi: 10.1109/THMS.2015.2492598.

[6] Mukherjee, Sohom & Sekh, Arif Ahmed & Dogra, Debi & Kar, Samarjit & Roy, Partha. (2019). Fingertip Detection and Tracking for Recognition of Air-Writing in Videos. Expert Systems with Applications. 136. 10.1016/j.eswa.2019.06.034.

[7] Nahar KMO, Alsmadi I, Al Mamlook RE, Nasayreh A, Gharaibeh H, Almuflih AS, Alasim F. Recognition of Arabic Air-Written Letters: Machine Learning, Convolutional Neural Networks, and Optical Character Recognition (OCR) Techniques. Sensors (Basel). 2023 Nov 28;23(23):9475. doi: 10.3390/s23239475. PMID: 38067848; PMCID: PMC10708688.

[8] Zare, M., Jampour, M., Arezoomand, A. S., & Sabouri, M. (2019, March). Handwritten recognition based on hand gesture recognition using deterministic finite automata and fuzzy logic. In 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA) (pp. 93-99). IEEE.

[9] Watanabe, T.; Maniruzzaman, M.; Hasan, M.A.M.; Lee, H.-S.; Jang, S.-W.; Shin, J. 2D Camera-Based Air-Writing Recognition Using Hand Pose Estimation and Hybrid Deep Learning Model. Electronics 2023, 12, 995. https://doi.org/10.3390/electronics12040995

[10] Abir FA, Siam MA, Sayeed A, Hasan MAM, Shin J. Deep Learning Based Air-Writing Recognition with the Choice of Proper Interpolation Technique. Sensors (Basel). 2021 Dec 16;21(24):8407. doi: 10.3390/s21248407. PMID: 34960499; PMCID: PMC8705512.

[11] Oka, Kenji & Sato, Yoichi & Koike, Hideki. (2002). Real-time fingertip tracking and gesture recognition. Computer Graphics and Applications, IEEE. 22. 64- 71. 10.1109/MCG.2002.1046630.

[12] Roy, Prasun & Ghosh, Subhankar & Pal, Umapada. (2018). A CNN Based Framework for Unistroke Numeral Recognition in Air-Writing. 404-409. 10.1109/ICFHR-2018.2018.00077.

[13] Rahman, Adil & Roy, Prasun & Pal, Umapada. (2021). Air Writing: Recognizing Multi-Digit Numeral String Traced in Air Using RNN-LSTM Architecture. SN Computer Science. 2. 10.1007/s42979-020-00384-9.

[14] Misra, Songhita & Singha, Joyeeta & Laskar, Rabul. (2018). Vision based hand gesture recognition of alphabets, numbers, arithmetic operators and ASCII characters in order to develop a virtual text-entry interface system. Neural Computing and Applications. 29. 10.1007/s00521-017-2838-6.

[15] C. Gunda, M. Maddelabanda and H. Shanmugasundaram, "Free Hand Text Displaying Through Hand Gestures Using MediaPipe," 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), Kannur, India, 2022, pp. 996-1000, doi: 10.1109/ICICICT54557.2022.9917991.

[16] B Sundar, T Bagyammal, American Sign Language Recognition for Alphabets Using MediaPipe and LSTM, Procedia Computer Science, Volume 215, 2022, Pages 642-651, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.12.066.

[17] S. Salim, M. M. A. Jamil, R. Ambar, R. Roslan and M. G. Kamardan, "Sign Language Digit Detection with MediaPipe and Machine Learning Algorithm," 2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2022, pp. 180-184, doi: 10.1109/ICCSCE54767.2022.9935659.

[18] Kavana, K. M., and N. R. Suma. "Recognization of Hand Gestures Using Mediapipe Hands." International Research Journal of Modernization in Engineering Technology and Science 4.06 (2022).

[19] Zhang, Fan, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang and Matthias Grundmann. “MediaPipe Hands: On-device Real-time Hand Tracking.” ArXiv abs/2006.10214 (2020): n. pa

[20] V. Jayasundara, S. Jayasekara, H. Jayasekara, J. Rajasegaran, S. Seneviratne and R. Rodrigo, "TextCaps: Handwritten Character Recognition With Very Small Datasets," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2019, pp. 254-262, doi: 10.1109/WACV.2019.00033.

1. https://github.com/google/mediapipe [↑](#footnote-ref-1)