

(Research Article)

Combining pattern recognition and deep-learning-based algorithms to automatically detect commercial quadcopters using audio signals

A. Zarei¹, A. Ghasemi^{*2}, H. Sadeghi³, M. Gholamipour⁴

1. Department of Biomedical Engineering, Tarbiat Modares University

2. Department of Computer and Electrical Engineering, Shahid Beheshti University

3. Infrasound Research Laboratory

4. Department of Computer Engineering, K. N. Toosi University of Technology

Received: 2022/05/26, Accepted: 2022/07/14

DOR 20.1001.1.23455748.1401.10.1.4.3

Abstract

Commercial quadcopters with many private, commercial, and public sector applications are a rapidly advancing technology. Currently, there is no guarantee to facilitate the safe operation of these devices in the community. Three different automatic commercial quadcopters identification methods are presented in this paper. Among these three techniques, two are based on deep neural networks in which all the feature extraction and classification processes are performed automatically. Deep learning-based methods include the convolutional neural network (CNN), LSTM networks, and a combination of those. The third method is presented using cepstral coefficients and support vector machines. In deep learning-based algorithms, the spectral patterns extracted from the commercial quadcopters' sounds are used as input data. The spectral patterns are obtained by applying the short-time Fourier transform method to the acoustic data. Besides, the cepstral coefficients and the support vector machines are used in the third method to identify and classify the received acoustic signals. The performance of the deep learning and cepstrum coefficients-based methods are compared using the acoustic datasets recorded from the commercial quadcopters. The results show that all three presented methods have adequate performance in identifying the quadcopters. However, the LSTM-CNN method had the best performance by providing an average accuracy of 95.31%, average sensitivity of 96.24%, and average specificity of 95.61%.

Keywords: Commercial quadcopters, Deep neural networks, Cepstrum coefficients, Machine learning.

pp. 34-45 (In Persian)

* Corresponding author E-mail: intelligentcontrol.ghasemi@gmail.com

ترکیب الگوریتم‌های مبتنی بر شناسایی الگو و یادگیری عمیق به منظور شناسایی خودکار کوادکوپترهای تجاری با استفاده از علائم صوتی دریافتی

اصغر زارعی^۱، امیر قاسمی^{۲*}، حامد صادقی^۳، مجتبی غلامی پور^۴

۱. دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس

۲. دانشکده برق و کامپیوتر، دانشگاه شهید بهشتی

۳. آزمایشگاه تحقیقاتی فروصدا

۴. دانشکده مهندسی کامپیوتر، دانشگاه خواجه نصیر الدین طوسی

دریافت: ۱۴۰۱/۰۲/۲۶، پذیرش: ۱۴۰۱/۰۴/۲۳

چکیده

استفاده از کوادکوپترهای تجاری یک فن‌آوری به سرعت در حال پیشرفت است که دارای کاربردهای بسیاری در بخش‌های خصوصی، تجاری و دولتی است. در حال حاضر، هیچ تضمینی برای تسهیل عملکرد ایمن این دستگاه‌ها در فضای جامعه وجود ندارد. در این مقاله، سه روش مختلف برای شناسایی خودکار کوادکوپترهای تجاری ارائه می‌شود. از بین سه فن‌آرائه شده، دو روش مبتنی بر شبکه‌های یادگیری عمیق است که در آن‌ها تمام مراحل استخراج ویژگی و طبقه‌بندی به‌صورت خودکار انجام می‌شود. روش‌های مبتنی بر یادگیری عمیق شامل شبکه‌های عصبی کانولوشن (سی‌ان‌ان)، شبکه‌های ال‌اس‌تی‌ام و ترکیب آن‌ها است. روش سوم با استفاده از ضرایب کپسترال (به‌عنوان ویژگی) و ماشین‌های بردار پشتیبان (به‌عنوان طبقه‌بند) ارائه می‌شود. در این مقاله الگوریتم‌های مبتنی بر یادگیری عمیق از الگوهای طیفی منحصر به فرد کوادکوپترهای تجاری به‌منظور استخراج ویژگی استفاده می‌کنند. الگوهای طیفی با اعمال روش تبدیل فوری کوتاه مدت روی داده‌های صوتی به‌دست می‌آیند. هم‌چنین، در روش سوم از ضرایب کپسترال و طبقه‌بند ماشین بردار پشتیبان به‌منظور شناسایی و دسته‌بندی علائم صوتی دریافتی استفاده می‌شود. عملکرد روش‌های مبتنی بر یادگیری عمیق و روش مبتنی بر ضرایب کپستروم با استفاده از مجموعه داده‌های صوتی ثبت‌شده از کوادکوپترهای تجاری مقایسه شده است. نتایج به‌دست آمده نشان می‌دهند که هر سه روش ارائه‌شده دارای عملکرد مناسبی در شناسایی کوادکوپترها هستند. با این حال، روش سی‌ان‌ان-ال‌اس‌تی‌ام با فراهم نمودن میانگین دقت ۹۵/۳۱ درصد، میانگین حساسیت ۹۶/۲۴ درصد و میانگین اختصاصیت ۹۵/۶۱ درصد دارای بهترین عملکرد است.

کلیدواژه‌ها: کوادکوپترهای تجاری، شبکه‌های یادگیری عمیق، ضرایب کپستروم، یادگیری ماشین.

۱. مقدمه

روزمره، روی آن‌ها مطالعه می‌کنند. از طرف دیگر مشاهده شده است که از پهپادها برای فعالیت‌های مخرب و آسیب رساندن به افراد و مراکز خاص که دارای اهمیت ویژه‌ای هستند، استفاده شده است. علاوه بر این، با توجه به این که ریزپرنده‌ها قابلیت حمل مواد منفجره را دارند، می‌توانند امنیت اجتماعات و گردهمایی‌ها را مختل ساخته و حوادث تلخی رقم بزنند. به همین ترتیب، نگرانی‌های مربوط به حریم خصوصی با توسعه هواپیماهای بدون سرنشین به وجود آمده است. در چندین مورد گزارش شده است که از آن‌ها برای جاسوسی و ضبط کلیپ از

با توجه به توسعه فن‌آوری ریزپرنده، محبوبیت هواپیماهای بدون سرنشین به سرعت در حال افزایش است. زیرا ابعاد آن‌ها دارای قابلیت فشرده‌سازی بوده و به دلیل راحتی کار با آن‌ها و ارزان بودن به راحتی قابل دسترس هستند. بسیاری از افراد هواپیماهای بدون سرنشین را برای مقاصد تفریحی مانند عکاسی و فیلم‌برداری استفاده می‌کنند، برخی دیگر نیز برای استفاده از ریزپرنده‌ها در کاربردهای

* نویسنده پاسخگو: intelligentcontrol.ghasemi@gmail.com

دارد. در عمل این امر یک گام اضافی برای حل مشکل طبقه‌بندی است. با این وجود، استفاده از الگوهای یادگیری عمیق، این مشکل را با اطمینان و مستقل از نحوه آموزش الگو از بین می‌برد [۹]. هم‌چنین، الگوریتم‌های یادگیری عمیق اثربخشی خود را در کاربردهای مختلف پردازش علامت‌های صوتی مانند پردازش گفتار ثابت کرده‌اند [۱، ۱۰-۱۱]. تحقیقات جالب دیگری توسط نویسندگان در منبع [۱۲] انجام شده است. آن‌ها به موضوع استفاده از فن‌های یادگیری عمیق برای شناسایی گونه‌های پرند با استفاده از علامت‌های صوتی پرندگان پرداخته‌اند. مطالعه آن‌ها نشان داده است که استفاده از فن‌های یادگیری عمیق برای شناسایی گونه‌های پرند براساس امضاهای صوتی آن‌ها، نتیجه امیدوارکننده‌ای خواهد داشت.

نیم‌جیم و همکاران یک روش جدید برای تشخیص ریزپرند با استفاده از صدای آن ارائه دادند. آن‌ها از الگوی مارکوف پنهان برای شناسایی ریزپرند‌های فانتوم نوع ۳ و اف‌پی‌وی ۵۲۵۰ استفاده کرده‌اند [۱۳]. در مطالعه [۱۴] نویسندگان مشکل شناسایی ریزپرند را با استفاده از علامت صوتی ریزپرند و فن‌های یادگیری عمیق مورد مطالعه قرار داده‌اند. در این مطالعه نویسندگان الگوهای مخلوط گاوسی (جی‌ام‌ام) ^۵، شبکه عصبی بازگشتی (آران‌ان) ^۶ و شبکه عصبی کانولوشن سی‌ان‌ان ^۷ را طراحی و آزمایش کرده‌اند. نتایج نشان می‌دهند که آران‌ان بهترین عملکرد را از نظر معیار F۱ در بین دو الگوریتم دیگر دارد. در واقع در منبع [۱۴] آقای جنون و همکاران استفاده از روش‌های الگوی مخلوط گاوسی (جی‌ام‌ام، سی‌ان‌ان)، آران‌ان را برای تشخیص و طبقه‌بندی یک ریزپرند در فاصله ۱۵۰ متری پیشنهاد کردند. در این مطالعه نویسندگان پیشنهاد کردند که با استفاده از روش‌های مختلف نوفه‌های محیطی را به مجموعه داده‌های موجود اضافه کنند. جنبه بسیار جالب این مطالعه، استفاده از ریزپرند‌های مختلف به‌منظور آموزش و آزمون طبقه‌بندی‌های مختلف است. آن‌ها به این نتیجه رسیدند که

افراد در حریم خصوصی آن‌ها استفاده شده است. مانند هر دستگاه دیگری، تصادفات غیر عمدی فراتر از کنترل اپراتور هواپیمای بدون سرنشین ممکن است رخ دهد. یعنی ممکن است ارتباط بین کنترل‌کننده و هواپیمای بدون سرنشین از بین برود و باعث انحراف از مسیر تعریف شده پهباد شود و خسارات جدی به وجود بیاورد [۱].

پرواز ریزپرند موجب تولید صدایی می‌شود که می‌تواند با استفاده از حسگرهای صوتی ضبط شود. صداهای ضبط‌شده با استفاده از روش‌های مختلفی به‌منظور شناسایی ریزپرند‌ها پردازش می‌شوند. در حالت آرمانی فرض می‌شود که می‌توان نوع ریزپرند را با استفاده از صدای منحصر به فرد آن تشخیص داد. در حالت کلی، می‌توان گفت که روش‌های تشخیص ریزپرند مبتنی بر همبستگی (یا خودهمبستگی) علائم دریافتی یا مبتنی بر یادگیری ماشین هستند [۲].

تاکنون تحقیقات قابل توجهی در مورد کاربرد ریزپرند‌ها با استفاده از فن‌آوری‌های مختلف مانند علامت‌های آراف ^۱ [۳، ۴]، سامانه جی‌اس‌ام ^۲ غیرفعال [۵]، رادار استاتیک مبتنی بر تلویزیون رقومی و غیره انجام شده است [۶]. علاوه بر این، محققان معدودی مطالعات خود را روی تشخیص ریزپرند‌ها با استفاده از ویژگی‌های صوتی مانند مطالعات انجام‌شده در منبع [۷] انجام داده‌اند. در مطالعه [۷] نویسندگان روشی جدید را با استفاده از پردازش علامت رقومی (دی‌اس‌پی) ^۳ برای تشخیص حضور ریزپرند‌ها در یک منطقه پیشنهاد داده‌اند. در مطالعه‌ای دیگر نویسندگان الگوریتم مبتنی بر دی‌اس‌پی را با الگوریتم‌های یادگیری ماشین مانند الگوریتم ماشین بردار پشتیبانی (اس‌وی‌ام) ^۴ ترکیب کردند [۸]. نتایج این مطالعه نشان می‌دهد که استفاده از طبقه‌بند اس‌وی‌ام در ردیابی هواپیماهای بدون سرنشین دقت عملکرد بالایی را فراهم می‌کند. با این حال، نتایج این تحقیقات محدود به میزان حضور نوفه پس‌زمینه است. علاوه بر این، دقت عملکرد خوب در اس‌وی‌ام نیاز به استخراج دستی ویژگی‌ها و بهینه‌سازی آن‌ها برای تنظیم دقیق الگوریتم

^۵ FPV 250

^۶ GMM

^۷ RNN

^۸ CNN

^۱ RF

^۲ GSM

^۳ DSP

^۴ SVM

صداهاى ریزپرنده‌ها از فاصله ۱۰ الی ۲۰ متری ضبط شده‌اند [۱۸].

هدف اصلی این مقاله معرفی یک سامانه خودکار است که هواپیماهای بدون سرنشین را براساس امضاهای صوتی آن‌ها شناسایی کند. شناسایی هواپیماهای بدون سرنشین به‌منظور تعیین نوع آن‌ها برای طبقه‌بندی اینکه اجازه حضور در یک منطقه پرواز خاص را دارند یا نه، ضروری است. این مقاله، یک رویکرد جدید برای شناسایی ریزپرنده‌ها در یک محیط با انواع نوفه‌های پس‌زمینه با استفاده از فن‌های قدرتمند یادگیری عمیق و ضرایب کپسترال به‌صورت زمان واقعی ارائه می‌دهد.

بلوک‌دیگرام روش‌های پیشنهادی در شکل نمایش داده شده است. همان‌طور که در این شکل مشاهده می‌شود، دستاوردهای اصلی این مقاله شامل دو بخش است: (۱) توانایی الگوریتم‌های یادگیری عمیق و ضرایب کپسترال در تشخیص و شناسایی ریزپرنده‌ها براساس معیارهای دقت، حساسیت و اختصاصیت بررسی و ارزیابی می‌شوند. (۲) نتایج به‌دست آمده توسط روش‌های مختلف با هم مقایسه می‌شوند تا دید بهتری در مورد استفاده از الگوهای یادگیری عمیق و ضرایب کپسترال در تشخیص ریزپرنده‌ها ارائه شود.

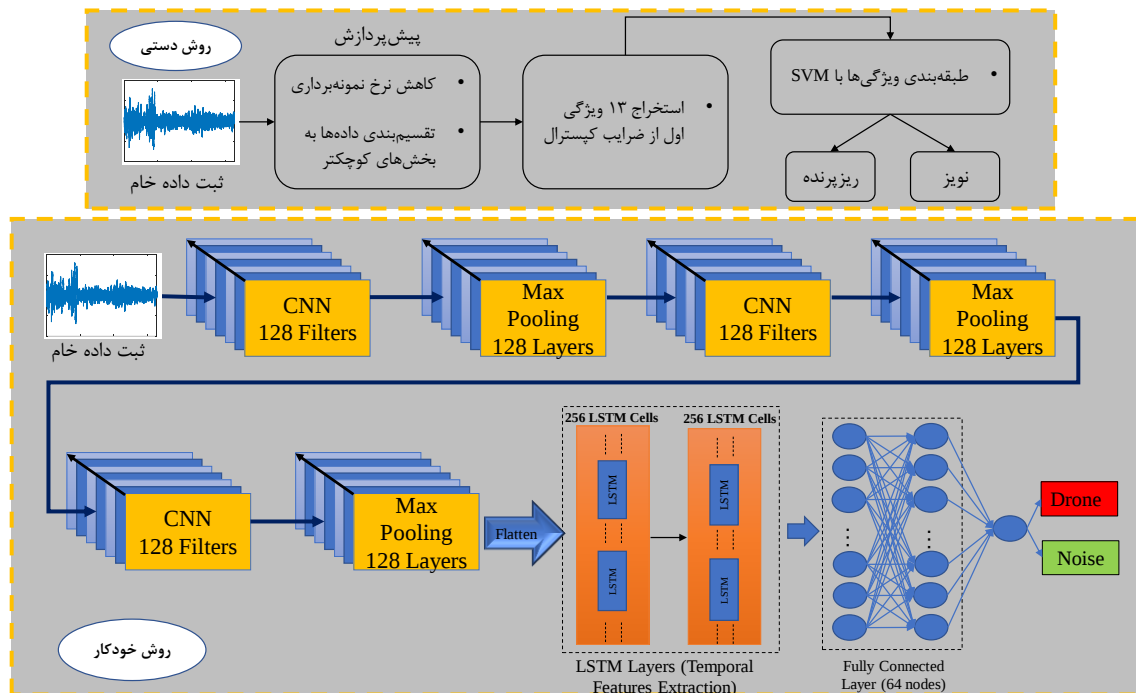
۲. مواد و روش‌ها

۲-۱. پایگاه داده

داده‌های استفاده‌شده در این مطالعه به‌صورت رایگان در دسترس عموم قرار داده شده‌اند [۱]. مجموعه داده‌ی مذکور دارای بیش از ۱۳۰۰ کلیپ صوتی از کوادکوپترهای تجاری هستند. علاوه بر این، برای شبیه‌سازی شرایط عملی از مجموعه داده‌های نوفه‌ای که به‌صورت رایگان قابل دانلود هستند، استفاده شده است [۱۹-۲۰]. هدف اصلی استفاده از داده‌های نوفه‌ای مختلف، امکان‌سنجی سطح عملکرد الگوریتم در شرایط نوفه‌ای مختلف است. هم‌چنین بخشی از داده‌ها که شامل نوفه‌های سکوت، نوفه خالص و صداهاى هواپیماهای بدون سرنشین بود، برای آزمون کنار گذاشته شد تا اطمینان حاصل شود که این سامانه قادر به شناسایی صدای ریزپرنده در یک محیط با شرایط نوفه‌ای مشابه است.

طبقه‌بند آران‌ان دارای بهترین عملکرد (دارای دقت عملکرد ۸۰ درصد) و طبقه‌بند سی‌ان‌ان دارای بدترین عملکرد (با فراهم کردن دقت ۵۸ درصد) است. با این وجود، به هنگام استفاده از یک ریزپرنده جدید عملکرد روش‌های مذکور به‌صورت قابل توجهی کاهش می‌یابد [۱۴]. آقای برناردینی و همکاران از طبقه‌بند اِس‌وی‌ام چندکلاسه به‌منظور شناسایی علائم صوتی ریزپرنده و سایر نوفه‌ها مانند نوفه محیط شهری استفاده کردند. در مطالعه مذکور از پنج مجموعه داده ۷۰ دقیقه‌ای از پرواز ریزپرنده و فایل‌های صوتی اینترنتی برای نوفه استفاده کردند. در مطالعه مذکور نویسندگان از پنجره‌های ۵ ثانیه‌ای و ۲۰ میلی‌ثانیه‌ای با میزان هم‌پوشانی ۱۰ میلی‌ثانیه استفاده کرده‌اند. نویسندگان ویژگی‌های انرژی، نرخ عبور از صفر، مرکز طیف، ضرایب کپسترال و طبقه‌بند اِس‌وی‌ام استفاده کردند. نتیجه تشخیص صدای ریزپرنده از سایر صداهاى نوفه‌ای منجر به فراهم شدن دقت ۹۶/۴ درصد شد [۱۵]. آقای کیم و همکاران استفاده از تصاویر اسپکتروگرام از علامت‌های صوتی و طبقه‌بند کی‌ان‌ان^۱ برای تشخیص ریزپرنده‌های فانتوم نوع ۱ و ۲ را پیشنهاد دادند. صداهاى ریزپرنده در محیط‌های سرپوشیده و محیط بیرون ضبط شده‌اند. هم‌چنین از کلیپ‌های اینترنتی به‌عنوان صداهاى نوفه‌ای استفاده کردند. همه صداهاى ضبط‌شده ابتدا به سگمنت‌های یک ثانیه‌ای تقسیم‌بندی شدند. در نهایت دقت ۸۳ درصد با استفاده از طبقه‌بند کی‌ان‌ان به‌دست آمد [۱۶]. آقایا سِنو و همکاران از روش تبدیل فوریه کوتاه مدت برای ساخت تصاویر دو-بعدی با استفاده از علائم صوتی ریزپرنده‌ها استفاده کردند. در مطالعه مذکور علامت‌های دریافتی ابتدا به سگمنت‌های ۲۰ میلی‌ثانیه‌ای با هم‌پوشانی ۵۰ درصد تقسیم‌بندی شدند. سپس، تصاویر اسپکتروگرام به ورودی طبقه‌بند سی‌ان‌ان اعمال شدند. در این مطالعه دقت ۹۸/۹۷ درصد به‌دست آمد [۱۷]. آقایا مَتسون و همکاران از روش تبدیل فوریه کوتاه مدت و ضرایب کپسترال برای استخراج ویژگی استفاده کردند. سپس، ویژگی‌های استخراج‌شده به ورودی طبقه‌بند اِس‌وی‌ام و سی‌ان‌ان اعمال شدند. در این مطالعه،

^۱ KNN



شکل ۱: بلوک دیاگرام روش‌های پیشنهادی؛ ردیف اول الگوریتم استخراج ویژگی و طبقه‌بندی دستی و ردیف دوم الگوریتم شناسایی خودکار.

است که در آن زمان مورد نیاز برای شناسایی بسیار مهم است. برای بررسی اینکه آیا طول بخش‌های کوچک‌تر در بخش‌بندی می‌تواند تأثیری در عملکرد کلی داشته باشد، طول‌های مختلف مانند ۵۰ میلی‌ثانیه، ۱۰۰ میلی‌ثانیه، ۳۰۰ میلی‌ثانیه و یک ثانیه آزمایش شدند. براساس نتایج اولیه، مشخص شد که طول ۱۰۰ میلی‌ثانیه در تقسیم‌بندی نسبت به مقادیر دیگر بهتر است. در شکل ۲ نمونه‌های مختلفی از سگمنت‌های نوفه‌ای و دارای ریزپرنده در حوزه زمان (ستون اول) به همراه اسپکتروگرام‌های برخی از سگمنت‌های دیگر که به صورت تصادفی از بین داده‌ها انتخاب شده‌اند، نمایش داده شده است.

۳-۲. استخراج ویژگی

در این مطالعه سه روش (دو روش خودکار و مبتنی بر یادگیری عمیق و یک روش دستی مبتنی بر ضرایب کپسترال) برای استخراج ویژگی‌های ارزشمند ارائه می‌شوند. در روش استخراج دستی ویژگی‌ها از ضرایب کپسترال و طبقه‌بند اِس‌وِی‌ام استفاده شده است. در این

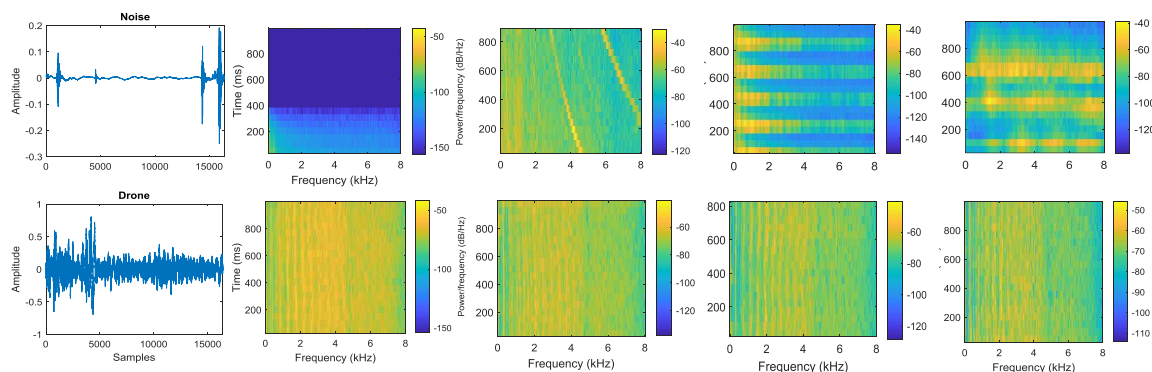
شناسایی صدای ریزپرنده در یک محیط با شرایط نوفه‌ای مشابه است. لازم به ذکر است که داده‌های کوادکوپترها در یک محیط سر بسته ثبت شده‌اند. داده‌ها به صورت تک کاناله بوده و با نرخ ۴۴/۱ کیلوهرتز نمونه‌برداری شده‌اند. اطلاعات بیش‌تر راجع به پایگاه داده‌های استفاده شده در این مطالعه را می‌توان در مراجع [۱، ۱۹، ۲۰] پیدا کرد. داده‌ها از دو کوادکوپتر تجاری شامل ریزپرنده‌های بُبپ^۱ و مَمبو^۲ ثبت شده‌اند.

۲-۲. پیش‌پردازش

در مرحله اول ابتدا همه داده‌های موجود، در بسامد ۱۶ کیلوهرتز دوباره نمونه‌برداری شدند. سپس، در مرحله دوم فایل‌های صوتی ثبت‌شده به بخش‌های کوچک‌تر با طول‌های مختلف تقسیم‌بندی شدند. این کار باعث می‌شود الگوریتم مورد نظر ویژگی‌ها را با دقت بیشتری در مقایسه با استفاده از کل داده به طور هم‌زمان استخراج کند. هدف دیگر برای تقسیم‌بندی، بهینه‌سازی فرآیند آموزش و آزمون الگو برای استفاده به صورت زمان واقعی

^۱ Bebop

^۲ Mambo



شکل ۲ نمونه‌های مختلفی از علامت‌های صوتی در حوزه زمان (ستون اول از سمت چپ) و اسپکتروگرام بخش‌های مختلف که به صورت تصادفی رسم شده‌اند؛ اسپکتروگرام‌های ردیف اول برای سگمت‌های نوفه‌ای هستند که به صورت تصادفی از بین داده‌های نوفه‌ای انتخاب شده‌اند؛ هم‌چنین اسپکتروگرام‌های ردیف دوم برای بخش‌های دارای داده ریزپرنده بوده که به صورت تصادفی از بین داده‌های کلاس ریزپرنده انتخاب شده‌اند.

روش ضرایب کپسترال وجود ندارد، بلوک‌دیگرام کلی این روش در شکل ۳ نمایش داده شده است. در این روش، تنظیم شبه‌سنگ‌هایی مانند تعداد بانک صافی و محدوده‌ی آن‌ها تاثیر زیادی در عملکرد سامانه نهایی دارد. به‌عنوان مثال، تصویر بانک صافی استفاده‌شده به‌منظور استخراج ویژگی در شکل ۴ نمایش داده شده است. مزیت این روش نسبت به روش‌های یادگیری عمیق، قابل تفسیر بودن آن‌ها می‌باشد.

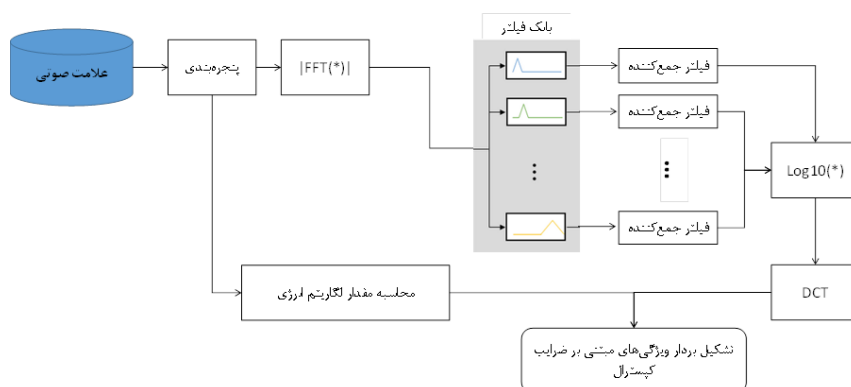
۲-۳-۲. شبکه‌های عصبی کانولوشن

شبکه‌های عصبی کانولوشن (سی‌ان‌ان‌اس) یکی از مهم‌ترین روش‌های یادگیری عمیق است که در آن چندین لایه به روش قدرتمند آموزش داده می‌شوند. این روش در کاربردهای مختلف بسیار مؤثر و متداول است. در سال‌های اخیر، این روش مورد توجه بسیاری در زمینه‌های مختلف پردازش علامت‌های صوتی، مانند شناسایی

روش ضرایب کپسترال از بخش‌های ۱۰۰ میلی‌ثانیه‌ای استخراج شده و با استفاده از روش اس‌وی‌ام طبقه‌بندی می‌شوند. در روش خودکار، همه فرآیند استخراج ویژگی‌ها و طبقه‌بندی با استفاده از الگوریتم یادگیری عمیق انجام می‌شود. در ادامه ساختارهای مختلفی از الگوریتم‌های یادگیری عمیق استفاده شده در این مطالعه با جزئیات بیش‌تر توضیح داده می‌شود.

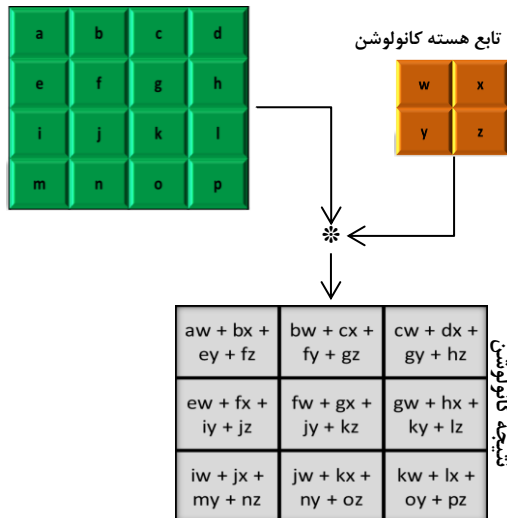
۲-۳-۱. استخراج ویژگی‌های مبتنی بر ضرایب کپسترال

ضرایب کپسترال ویژگی‌های بسیار رایجی هستند که در اکثر زمینه‌ها از علامت صوتی به‌منظور شناسایی و طبقه‌بندی آن، استخراج می‌شوند. در یک الگو صافی (فیلتر) منبع صوتی، ضرایب کپسترال توصیفی از صافی ارائه می‌کنند. اساس این روش، بیان طیف هموارشده علامت ورودی با استفاده از تعداد محدودی از ضرایب است. با وجود اینکه استاندارد واحدی برای محاسبه



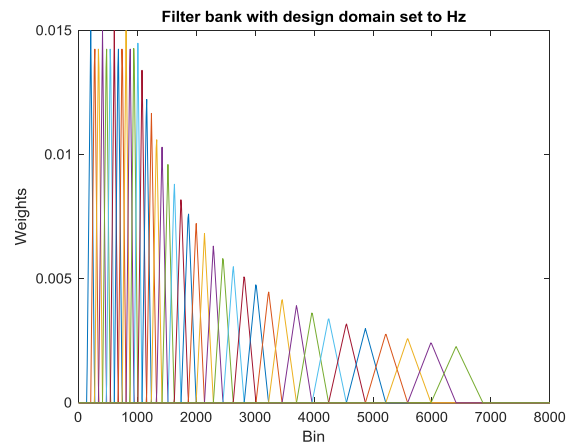
شکل ۳ بلوک‌دیگرام روش استخراج ویژگی مبتنی بر ضرایب کپسترال.

نمونه‌ای از عملگر ماکس-پولینگ^۲ در شکل ۶ نمایش داده شده است. لایه متراکم اغلب به‌عنوان آخرین قسمت از شبکه‌های سی‌ان‌ان استفاده می‌شود.



۳-۳-۲. شبکه‌های عصبی بازگشتی

شبکه‌های عصبی بازگشتی (آران‌ان) معمولاً روی علامت‌هایی اعمال می‌شوند که بین مقادیر آن‌ها در طول زمان همبستگی وجود دارد. در حالی که شبکه‌های عصبی مرسوم، تمام مقادیر علامت ورودی را به‌صورت مستقل از هم در نظر می‌گیرند. در بسیاری از شرایط عملیاتی مانند پردازش علامت صوتی، ارتباط معنی‌داری بین توالی‌های مختلف وجود دارد. شبکه‌های آران‌ان دارای یک واحد حافظه هستند که می‌توانند توالی‌های علامت ورودی را پردازش کنند. مهم‌ترین مزیت آران‌ان در نظر گرفتن وابستگی زمانی و استخراج ویژگی‌های زمانی است.



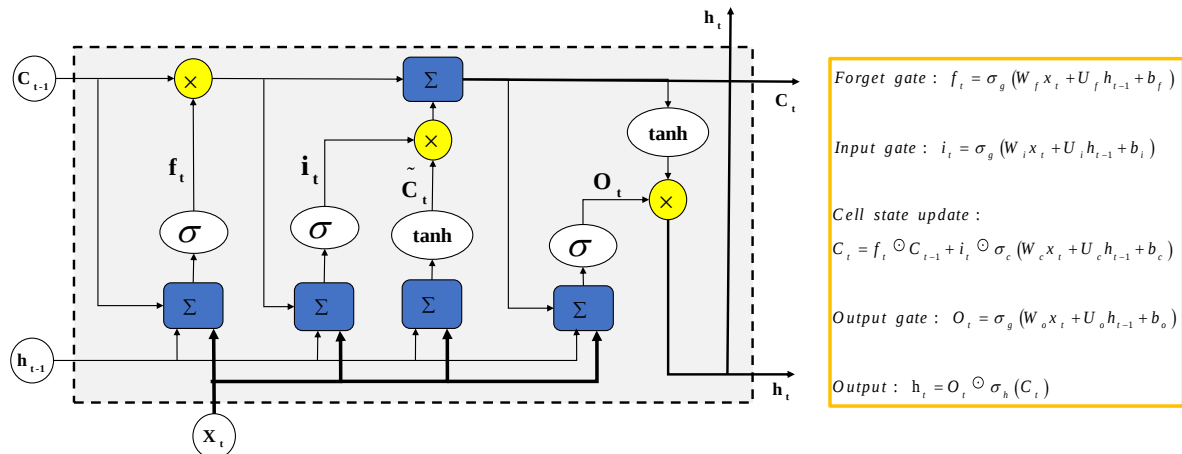
شکل ۴ بانک فیلتر استفاده‌شده در روش مبتنی بر ضرایب کپسترال.

ریزپرنده‌ها [۱]، طبقه‌بندی صداها و اطراف [۲۱] و غیره قرار گرفته است. شبکه‌های سی‌ان‌ان از نرون‌ها، وزن‌ها و بایاس‌های قابل تنظیم تشکیل شده‌اند. هر نرون بخشی از علامت صوتی را به‌عنوان ورودی دریافت می‌کند و سپس حاصل ضرب وزن‌ها در مقادیر ورودی را محاسبه می‌کند. خروجی این فرآیند با استفاده از یک تابع فعال‌ساز غیرخطی به برجسب کلاس تبدیل می‌شود.

به‌طور کلی، شبکه‌های سی‌ان‌ان از سه بخش تشکیل می‌شوند: لایه‌های کانولوشن، لایه‌های پولینگ^۱ و لایه‌های متراکم (لایه‌هایی که در آن‌ها همه نرون‌ها کاملاً به هم متصل هستند). لایه کانولوشن از مجموعه‌ای از صافی‌ها به نام هسته تشکیل شده است که در علامت ورودی کانالو می‌شوند. خروجی فرآیند کانولوشن بین هسته‌ها و علامت ورودی، ویژگی‌های استخراج‌شده اولیه را فراهم می‌کند. فرآیند کانولوشن به‌صورت تصویری در شکل ۵ نشان داده شده است. فرآیند پولینگ، که به‌عنوان مرحله کاهش نرخ نمونه‌برداری نیز شناخته می‌شود، روی بردار ویژگی اعمال می‌شود. در این مرحله، مقادیر هر گروه از نرون‌های همسایه با یک عملیات ریاضی (مانند میانگین‌گیری یا بیشینه گرفتن) ترکیب می‌شود. این مرحله شبیه یک صافی پایین‌گذر عمل کرده و انتخاب‌های رایج برای آن بیشینه‌گیری یا متوسط‌گیری است. این لایه مانع از بیش‌برازش شده و بار محاسبات را کاهش می‌دهد.

^۲ Max-Pooling

^۱ Pooling



شکل ۷ معماری مرسوم یک یاخته از ساختار ال‌اس‌تی‌ام (سمت چپ) و روابط حاکم بر ساختار ال‌اس‌تی‌ام (سمت راست)؛ در این ساختار i_t واحد ورودی، O_t واحد خروجی و f_t واحد فراموشی است که به واحد پنهان (h_t) در ساختارهای آران‌ان اضافه شده است.

۴-۲. ساختارهای پیشنهادی برای شناسایی ریزپرنده‌ها

بیش‌تر مطالعات موجود در ادبیات متمرکز بر مهندسی ویژگی است که عملکرد آن‌ها به دانش قبلی متخصصان و شناخت آن‌ها از ماهیت و ذات علامت وابسته است. در این مطالعه، سه روش با استفاده از مهندسی ویژگی‌ها (با استخراج ضرایب کپسترال) و یادگیری عمیق برای استخراج خودکار ویژگی‌ها و همچنین افزایش دقت طبقه‌بندی ارائه شده است. همان‌طور که در شکل ۴ نشان داده شده، ساختارهای پیشنهادی از دو بخش اصلی تشکیل شده‌اند. بخش اول، شامل مهندسی ویژگی‌ها بوده و از ضرایب کپسترال و طبقه‌بند اِس‌وی‌ام برای شناسایی ریزپرنده‌ها استفاده می‌کند. بخش دوم شامل یک الگوی ترکیبی از شبکه‌های سی‌ان‌ان و ال‌اس‌تی‌ام است. همان‌طور که مشاهده می‌شود، ساختار پیشنهادی شامل سه لایه سی‌ان‌ان دو-بعدی و دو لایه ال‌اس‌تی‌ام است. این لایه‌ها برای استخراج مهم‌ترین ویژگی‌های مکانی و زمانی علامت‌های صوتی استفاده می‌شوند. در این ساختار بعد از استخراج ویژگی، طبقه‌بندی انجام می‌شود. به عبارت دیگر، ویژگی‌های استخراج‌شده به ورودی لایه متراکم (لایه‌ای که در آن نرون‌ها کاملاً به هم متصل شده‌اند) جهت طبقه‌بندی سگمنت‌های نوفه‌ای و دارای ریزپرنده اعمال می‌شوند. یک مشکل اساسی در بیش‌تر ساختارهای یادگیری عمیق، مشکل بیش‌برازش است. در

۴-۳-۲. شبکه‌های ال‌اس‌تی‌ام

به صورت نظری، شبکه‌های آران‌ان عملکرد خوبی در حفظ ارتباط علامت‌های ورودی نشان می‌دهند. در حالی که در کاربردهای عملی، برای اندازه‌گیری وابستگی‌های زمانی با محدودیت‌های زیادی روبرو هستند. برای حل چالش‌ها، ساختار ال‌اس‌تی‌ام (ساختار دارای حافظه بلند مدت و کوتاه مدت) که نسخه توسعه‌یافته از شبکه آران‌ان می‌باشد، ارائه شده است. شبکه‌های ال‌اس‌تی‌ام^۱ از حافظه بلند مدت و کوتاه مدت برای اندازه‌گیری تغییرات بلند مدت و کوتاه مدت علامت‌ها استفاده می‌کنند. همان‌طور که در شکل ۷ نشان داده شده است، هر یاخته اصلی ال‌اس‌تی‌ام به سه واحد وصل است: یک واحد ورودی، یک واحد خروجی و یک واحد فراموشی.

در معادلات موجود در شکل ۷ متغیرهای h_{t-1} و C_{t-1} به ترتیب خروجی و حالت یاخته ال‌اس‌تی‌ام قبلی هستند. x_t بردار ورودی واحد ال‌اس‌تی‌ام است. متغیرهای W ، U و b ماتریس‌های وزن و شبه‌سنج‌های بردار بایاس هستند که نیاز دارند در طول فرآیند آموزش تنظیم شوند. شبه‌سنج σ_g نشان‌دهنده تابع سیگموئید است. همچنین متغیرهای σ_h و σ_c توابع تانژانت هایپربولیک هستند. در معادلات فوق عملکرد $\odot$ نشان‌دهنده ضرب درایه به درایه است. یک یاخته ال‌اس‌تی‌ام می‌تواند بردار وزن‌ها را براساس حالت قبلی (C_{t-1}) و ورودی فعلی (i_t) بروزسانی کند.

¹ LSTM

جدول ۱ شبه‌سنگ‌های راه‌اندازی شبکه یادگیری عمیق به منظور استخراج ویژگی.

لایه‌ها	شبه‌سنگ‌های یادگیری
سی‌ان‌ان	تعداد صافی‌ها = ۱۲۸، ابعاد هسته‌ها = ۲*۲، تابع فعال‌سازی = ReLU
ماکس - پولینگ	Pool size = ۲*۲
ال‌اس‌تی‌ام	تعداد یاخته‌ها = ۲۵۶

جدول ۲ شبه‌سنگ‌های راه‌اندازی شبکه یادگیری عمیق به منظور استخراج ویژگی.

شبه‌سنگ	توصیف
بهینه‌ساز	بهینه‌ساز = آدام، نرخ یادگیری = ۰٫۰۰۰۵
تابع هزینه	باینری کراس-اینترپی*
تابع فعال‌سازی	سیگموئید (آخرین لایه)
لایه متراکم	تعداد واحدها = ۶۴، تابع فعال‌سازی = ReLU

* Binary cross-entropy

جدول ۳ جزئیات شبه‌سنگ‌های لایه‌های شبکه ترکیبی سی‌ان‌ان-ال‌اس‌تی‌ام پیشنهادی.

لایه	نوع	تابع فعال‌سازی	ابعاد هسته	تعداد صافی، یاخته، واحد
۱	سی‌ان‌ان	ReLU	۲*۲	۱۲۸
۲	ماکس - پولینگ	-	۲*۲	-
۳	درپوت	-	-	-
۴	سی‌ان‌ان	ReLU	۲*۲	۱۲۸
۵	ماکس - پولینگ	-	۲*۲	-
۶	درپوت	-	-	-
۷	سی‌ان‌ان	ReLU	۲*۲	۱۲۸
۸	ماکس - پولینگ	-	-	-
۹	درپوت	-	-	-
۱۰	فلتن*	-	-	-
۱۱	ال‌اس‌تی‌ام	-	-	۲۵۶
۱۲	ال‌اس‌تی‌ام	-	-	۲۵۶
۱۳	درپوت	-	-	-
۱۴	دنس**	ReLU	-	۶۴
۱۵	دنس	سیگموئید	-	۱

* Flatten

** Dense

این مطالعه برای جلوگیری از این مشکل از روش درپوت^۱ استفاده شده است. مطابق این فن، برخی از وزن‌های شبکه در مرحله آموزش به‌طور تصادفی صفر می‌شوند. این فرآیند مانع از همگرایی وزن‌های شبکه عصبی عمیق به یک کمینه محلی می‌شود.

۲-۵. معیارهای ارزیابی عملکرد

به‌منظور مقایسه عملکرد روش‌های مختلف، از معیارهای حساسیت، دقت و اختصاصیت در این مطالعه استفاده شده است. این معیارها با استفاده از روابط زیر محاسبه می‌شوند:

$$Sens = \frac{TP}{TP + FN} \quad (1)$$

$$Spec = \frac{TN}{TN + FP} \quad (2)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$F1 - score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (4)$$

در روابط فوق، مثبت صحیح (TP) به تعداد سگمنت‌هایی که حاوی علامت ریزپرنده بوده (به‌عنوان کلاس مثبت) و به درستی طبقه‌بندی شده‌اند، گفته می‌شود. منفی صحیح (TN) تعداد سگمنت‌های نوفه‌ای (به‌عنوان کلاس منفی) هستند که به درستی طبقه‌بندی شده‌اند. مثبت نادرست و منفی نادرست نیز به ترتیب با FP و FN نشان داده می‌شوند.

۳. نتایج

عملکرد روش‌های پیشنهادی با استفاده از معیارهای دقت، اختصاصیت و حساسیت ارزیابی شده است. به‌منظور دستیابی به عملکرد بهینه، شبه‌سنگ‌های شبکه استخراج ویژگی به‌صورت تجربی و مطابق با مقادیر موجود در جدول ۱ تنظیم شده‌اند. همچنین، شبه‌سنگ‌هایی که در لایه طبقه‌بندی‌کننده استفاده شده‌اند، در جدول ۲ خلاصه گردیده‌اند. در نهایت، جزئیات الگوی مبتنی بر ساختار پیشنهادی سی‌ان‌ان-ال‌اس‌تی‌ام در جدول ۳ ارائه شده است.

¹ Dropout

آزمون رسم شده که نشان‌دهنده عملکرد مطلوب ویژگی‌های اِم‌اِف‌سی‌سی در طبقه‌بندی صداها ریزپرنده است.

۴. بحث

همان‌طور که قبلاً ذکر شد، در این مطالعه یک چهارچوب جدید برای شناسایی خودکار ریزپرنده با استفاده از شبکه‌های سی‌ان‌ان و اِل‌اِس‌تی‌ام ارائه شده است. به‌منظور استخراج ویژگی‌های مکانی، از یک الگوی سی‌ان‌ان دو-بعدی استفاده شد. بنابراین، علامت‌های صوتی یک‌بعدی به بخش‌های دو-بعدی تبدیل شدند. پس از آن، از شبکه‌های اِل‌اِس‌تی‌ام برای استخراج ویژگی‌های زمانی استفاده شد. برای مقابله با مشکل بیش‌برازش، از روش درپوت با احتمال ۰/۳ بعد از بعضی از لایه‌ها استفاده شد. این روش در فرآیند آموزش ۳۰ درصد از وزن‌ها را به‌طور تصادفی صفر می‌کند. شبه‌سنج‌های شبکه استخراج ویژگی به‌صورت تجربی و مطابق با مقادیر موجود در جدول ۱ تنظیم شده‌اند. هم‌چنین، شبه‌سنج‌های لایه طبقه‌بندی‌کننده در جدول ۲ خلاصه شده‌اند.

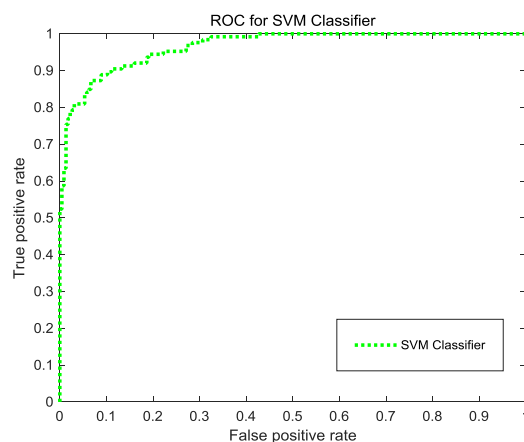
به‌منظور دستیابی به عملکرد بهینه، باید مقدار شبه‌سنج‌های مختلف در شبکه پیشنهادی تنظیم شوند. به‌عنوان مثال، در جدول ۴ تاثیر تعداد واحدهای موجود در لایه‌های پنهان بررسی شده است. همان‌طور که مشاهده می‌شود، تنظیم تعداد واحدهای استفاده شده در لایه‌های پنهان، عملکرد واحد طبقه‌بندی‌کننده در شبکه مورد نظر را بهبود می‌بخشد. همان‌طور که در جدول ۴ گزارش شده است، بهترین نتیجه (دقت برابر با ۹۵/۳۱ درصد) با استفاده از ساختار دوم به‌دست آمده است. در این ساختار تعداد ۶۴ واحد در لایه پنهان اولیه برای طبقه‌بندی استفاده شده است. هم‌چنین، در این ساختار لایه پنهان دوم حذف شده است. مزیت ساختار دوم پیشنهادی این است که در عین سادگی (نسبت به سایر ساختارهای پیشنهادشده در این مطالعه) بهترین عملکرد را در مقایسه با سایر ساختارهای پیشنهادشده فراهم کرده است. در شکل ۸ نمودار آراسی برای روش مبتنی بر ضرایب اِم‌اِف‌سی‌سی رسم شده است. نمودار آراسی نشان می‌دهد که روش مبتنی بر ضرایب اِم‌اِف‌سی‌سی دارای

تاثیر تعداد واحدهای موجود در لایه پنهان مورد بررسی قرار گرفته و دقت به‌دست آمده با استفاده از تعداد واحدهای مختلف در جدول ۴ گزارش شده است.

جدول ۴ شبه‌سنج‌های یادگیری لایه‌های مختلف طبقه‌بندی؛ زمان‌های آموزش برای کل شبکه پیشنهادی به همراه لایه‌های طبقه‌بندی به‌صورت میانگین (بعد از چهار بار تکرار اجرا).

زمان آموزش	دقت (درصد)	لایه پنهان دوم	لایه پنهان اول	دی‌ان‌ان*
۶۰ دقیقه	۹۳/۵۱	-	۳۲ واحد	دی‌ان‌ان ۱*
۶۳ دقیقه	۹۵/۳۱	-	۶۴ واحد	دی‌ان‌ان ۲
۷۱ دقیقه	۹۴/۳۵	۳۲ واحد	۳۲ واحد	دی‌ان‌ان ۳
۷۸ دقیقه	۹۴/۷۳	۶۴ واحد	۶۴ واحد	دی‌ان‌ان ۴

* DNN



شکل ۸ نمودار آراسی برای طبقه‌بند اِس‌وی‌ام و ویژگی‌های مبتنی بر ضرایب کپسترال.

هم‌چنین، به‌منظور ارزیابی تصویری عملکرد روش پیشنهادی، از معیار آراسی^۱ استفاده شده است. در شکل ۸ نمودار آراسی برای روش مبتنی بر ضرایب اِم‌اِف‌سی‌سی^۲ و طبقه‌بند ماشین بردار پشتیبان رسم شده است. لازم به ذکر است که نمودار آراسی با استفاده از نتایج داده

^۱ ROC

^۲ MFCC

پایینی هستند. علاوه بر روش‌های مبتنی بر یادگیری عمیق، در این مطالعه یک چهارچوب مبتنی بر مهندسی ویژگی‌ها نیز برای شناسایی ریزپرنده ارائه شد. همان‌طور که در بخش‌های قبلی توضیح داده شد، در این چهارچوب از ضرایب کپسترال و طبقه‌بند IS-OI برای جداسازی علامت‌های نوفه‌ای از علامت‌های ریزپرنده استفاده شد. نتایج به‌دست‌آمده در جدول ۵ نشان می‌دهند که ساختار مبتنی بر مهندسی ویژگی‌ها در عین سادگی می‌تواند عملکرد قابل مقایسه‌ای با ساختارهای پیچیده مبتنی بر یادگیری عمیق فراهم کند. استفاده از ضرایب کپسترال به‌عنوان ویژگی در کنار طبقه‌بند IS-OI می‌تواند دقت ۹۱/۹۰ درصد، حساسیت ۹۲/۳۶ درصد و اختصاصیت ۸۹/۸۸ درصد را در طبقه‌بندی علامت‌های نوفه‌ای و ریزپرنده به‌دست آورد.

در جدول ۶ نتایج روش پیشنهادی با عملکرد سایر مطالعات موجود از نظر دقت طبقه‌بندی مقایسه شده است. همان‌طور که مشاهده می‌شود، همه روش‌های پیشنهادی دارای عملکرد بهتری نسبت به روش‌های ارائه‌شده در جدول ۶ هستند. مزیت روش پیشنهادی مبتنی بر یادگیری عمیق این است که در آن‌ها همه کارهای استخراج و انتخاب ویژگی به‌صورت خودکار انجام می‌شود. هم‌چنین، روش‌های مبتنی بر یادگیری عمیق عملکرد مناسبی را فراهم می‌کنند. با وجود اینکه روش‌های مبتنی بر مهندسی ویژگی‌ها (مانند روش پیشنهادی مبتنی بر ضرایب بسامد مل) دارای عملکرد پایینی نسبت به روش‌های یادگیری عمیق هستند، ولی دارای قابلیت تفسیرپذیری بالایی هستند.

۵. نتیجه‌گیری

در این مطالعه یک چهارچوب جدید مبتنی بر شبکه‌های یادگیری عمیق برای شناسایی علائم صوتی کوادکوپترهای تجاری از علائم نوفه‌ای ارائه شد. عملکرد ساختارهای یادگیری عمیق مختلف شامل SI-IN ، I-I و ترکیب آن‌ها باهم مقایسه شدند. علاوه بر این یک ساختار مبتنی بر مهندسی ویژگی‌ها با استفاده از ضرایب کپسترال و طبقه‌بند IS-OI ارائه شد. نتایج به‌دست‌آمده نشان دادند که استفاده از ترکیب ساختارهای SI-IN و

عملکرد مطلوبی در طبقه‌بندی سگمنت‌های نوفه‌ای و ریزپرنده است.

در این مطالعه چهار ساختار مختلف، برای شناسایی علامت‌های ریزپرنده از نوفه ارائه شد. هم‌چنین مقایسه‌ای بین عملکرد هر چهار ساختار پیاده‌سازی‌شده انجام شد. نتایج به‌دست‌آمده از مقایسه عملکرد روش‌های مختلف در جدول ۵ گزارش شده است. همان‌طور که مشاهده می‌شود، ساختارهای SI-IN و I-I نسبت به ترتیب دقت‌های ۹۳/۹۲ و ۹۱/۵۶ درصد را در جداسازی علامت‌های ریزپرنده و نوفه فراهم کردند. با این وجود از نظر هر چهار معیار استفاده‌شده (دقت، اختصاصیت و حساسیت) ساختار SI-IN عملکرد بهتری نسبت به ساختار I-I دارد.

جدول ۵ نتایج طبقه‌بندی حاصل از روش‌های پیاده‌سازی‌شده، مشاهده می‌شود که روش یادگیری عمیق SI-IN - I-I نسبت به دارای بهترین عملکرد در بین سایر روش‌ها است.

روش	معیار		
	دقت (درصد)	اختصاصیت (درصد)	حساسیت (درصد)
SI-IN	۹۳/۹۲	۹۵/۹۴	۹۵/۳۵
I-I	۹۱/۵۶	۹۵/۵۸	۹۰/۲۰
SI-IN - I-I	۹۵/۳۱	۹۶/۲۴	۹۵/۶۱
I-I - IS-OI	۹۱/۹۰	۸۹/۸۸	۹۲/۳۶

همان‌طور که در جدول ۵ گزارش شده است، در بین ساختارهای یادگیری عمیق استفاده‌شده، بهترین نتیجه به‌وسیله ساختار ترکیبی SI-IN - I-I نسبت به‌دست آمده است. ساختار ترکیبی SI-IN - I-I نسبت به دقت ۹۵/۳۱ درصد، حساسیت ۹۵/۶۱ درصد، اختصاصیت ۹۶/۲۴ درصد و مقدار F1 - Score ^۱ برابر با ۹۵/۱۲ درصد را در طبقه‌بندی سگمنت‌های نوفه‌ای و سگمنت‌های دارای علامت ریزپرنده به‌دست آورده است. مزیت همه ساختارهای یادگیری عمیق خودکار بودن فرآیند استخراج ویژگی و طبقه‌بندی در آن‌ها می‌باشد. در مقابل ساختارهای مبتنی بر یادگیری عمیق دارای قابلیت تفسیر

^۱ F1-Score

جدول ۶ مقایسه عملکرد روش‌های پیشنهادی با سایر روش‌های موجود، متغیر FC بیانگر لایه تماماً متصل است.

منبع	سال	ویژگی	طبقه‌بند	دقت (درصد)
[۲۲]	۲۰۱۸	إف‌إف‌تی - بیسد فیچر*	ای‌ان‌ان**	۸۶٫۷
[۱۴]	۲۰۱۷	مل اسپکتروگرام	آران‌ان	۸۱٫۰
[۲۳]	۲۰۱۹	إس‌تی‌إف‌تی اسپکتروگرام	سی‌ان‌ان	۹۰٫۱
روش پیشنهادی		إل‌إس‌تی‌ام	فلتن + إف‌سی + سافت‌ماکس***	۹۱٫۵۶
		سی‌ان‌ان	فلتن + إف‌سی + سافت‌ماکس	۹۳٫۹۲
		سی‌ان‌ان - إل‌إس‌تی‌ام	فلتن + إف‌سی + سافت‌ماکس	۹۵٫۳۱
		إم‌إف‌سی‌سی	إس‌وی‌ام	۹۱٫۹۰

* FFT-based features

** ANN

*** Flatten + FC + SoftMax

- [2] X. Yue, Y. Liu, J. Wang, H. Song, H. Cao, "Machine learning-based drone detection and classification: state-of-the-art in research," IEEE Access, vol. 7, no. 4, pp. 138668-138682, 2019.
- [3] P. Nguyen, M. Ravindranatha, A. Nguyen, R. Han, T. Vu, "Investigating cost-effective rf-based detection of drones," Proceedings of the 2nd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use, pp. 17-22, 2016.
- [4] P. Nguyen, H. Truong, M. Ravindranathan, A. Nguyen, R. Han, T. Vu, "Matthan: Drone presence detection by identifying physical signatures in the drone's RF communication," Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, pp. 211-224. 2017.
- [5] B. Knoedler, R. Zemmari, W. Koch, "On the detection of small UAV using a GSM passive coherent location system," 17th international radar symposium (IRS), IEEE, pp. 1-4, 2016.
- [6] Y. Liu, X. Wan, H. Tang, J. Yi, Y. Cheng, X. Zhang, "Digital television based passive bistatic radar system for drone detection," IEEE Radar Conference (RadarConf), IEEE, pp. 1493-1497, 2017.
- [7] A. Nagy, B. Kovari, "Programming language neutral design pattern detection," 16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI), IEEE, pp. 215-219, 2015.
- [8] A. Bernardini, F. Mangiatordi, E. Pallotti, L. Capodiferro, "Drone detection by acoustic signature identification," Electronic Imaging, vol. 2017, no. 10, pp. 60-64, 2017.
- [9] N. Takahashi, M. Gygli, L. Van Gool, "Aenet: Learning deep audio features for video analysis," IEEE Transactions on

إل‌إس‌تی‌ام بهترین عملکرد (میانگین دقت برابر با ۹۵٫۳۱ درصد، میانگین اختصاصیت ۹۶٫۲۴ درصد و میانگین حساسیت ۹۵٫۶۱ درصد) را در بین سایر ساختارهای پیاده‌سازی شده فراهم می‌کنند. نتایج نشان دادند که یادگیری وابستگی‌های مکانی با استفاده از ساختار سی‌ان‌ان و اندازه‌گیری وابستگی‌های زمانی با استفاده از ساختار إل‌إس‌تی‌ام دارای عملکرد مناسب است. مزیت ساختار ترکیبی ارائه‌شده نسبت به روش‌های مبتنی بر مهندسی ویژگی، خودکار بودن تمام مراحل شناسایی در آن‌ها است. با این حال، عیب اصلی در تمام روش‌های مبتنی بر یادگیری عمیق قابلیت تفسیرپذیری پایین در آن‌ها می‌باشد.

۶. تقدیر و تشکر

نویسندگان از حمایت انجمن علوم صوتی ایران و پژوهشکده صوتیات تشکر می‌کنند.

۷. تضاد منافع

این مطالعه تضاد منافع ندارد.

فهرست منابع

- [1] S. Al-Emadi, A. Al-Ali, A. Mohammad, A. Al-Ali, "Audio based drone detection and identification using deep learning," 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), IEEE, pp. 459-464, 2019.

- [21] J. Salamon, J.P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279-283, 2017.
- [22] K. Juhyun, K. Dongho, "Neural network based real-time UAV detection and analysis by sound," *Journal of Advanced Information Technology and Convergence*, vol. 8, no. 1, pp. 43-52, 2018.
- [23] C. Ponce, A. Aldrich, J. Martinez-Carranza, C. Rascon, "Detection of nearby UAVs using CNN and Spectrograms," *Proceedings of the International Micro Air Vehicle Conference and Competition (IMAV)*, vol. 29, 2019.
- Multimedia, vol. 20, no. 3, pp. 513-524, 2017.
- [10] Z.-H. Ling, S.-Y. Kang, H. Zen, A. Senior, M. Schuster, X.-J. Qian, H.M. Meng, L. Deng, "Deep learning for acoustic modeling in parametric speech generation: A systematic review of existing techniques and future trends," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 35-52, 2015.
- [11] H.C. Vemula, "Multiple Drone Detection and Acoustic Scene Classification with Deep Learning," *Doctoral dissertation*, Wright State University, 2018.
- [12] E. Sprengel, M. Jaggi, Y. Kilcher, T. Hofmann, "Audio based bird species identification using deep learning techniques," *Conference and Labs of the Evaluation Forum (CLEF)*, pp. 547- 559, 2016.
- [13] M. Nijim, N. Mantrawadi, "Drone classification and identification system by phenome analysis using data mining techniques," *Proceeding IEEE Symposium Technologies Homeland Security (HST)*, pp. 1-5, 2016.
- [14] S. Jeon, J.-W. Shin, Y.-J. Lee, W.-H. Kim, Y. Kwon, H.-Y. Yang, "Empirical study of drone sound detection in real-life environment with deep neural networks," *arXiv:1701.05779*, 2017.
- [15] A. Bernardini, F. Mangiatordi, E. Pallotti, L. Capodiferro, "Drone detection by acoustic signature identification," *Electronic Imaging*, vol. 2017, no. 10, pp. 60_64, 2017.
- [16] J. Kim, C. Park, J. Ahn, Y. Ko, J. Park, J.C. Gallagher, "Real-time UAV sound detection and analysis system," *Proceeding IEEE Sensors Applied Symposium (SAS)*, pp. 1-5, 2017.
- [17] Y. Seo, B. Jang, S. Im, "Drone detection using convolutional neural networks with acoustic STFT features," *Proceeding 15th IEEE International Conference Advanced Video Signal Based Surveill (AVSS)*, pp. 1-6, 2018.
- [18] E. Matson, B. Yang, A. Smith, E. Dietz, J. Gallagher, "UAV detection system with multiple acoustic nodes using machine learning models," *Proceeding 3rd IEEE International Conference Robotic Computing (IRC)*, pp. 493-498, 2019.
- [19] K.J. Piczak, "ESC: Dataset for environmental sound classification," *Proceedings of the 23rd ACM International Conference on Multimedia*, pp. 1015-1018, 2015.
- [20] P. Warden, "Speech commands: A dataset for limited-vocabulary speech recognition," *arXiv preprint arXiv:1804.03209*, 2018.