



# Automatic Target Recognition Vessels In Passive Sonar Using Emerging Technologies Of Artificial Intelligence And Deep Learning

Hassan Akbarian<sup>1</sup> | Mohammad Hosein Sedaaghi<sup>2✉</sup>

1. Department of Electrical Engineering, Sahand University of Technology, Tabriz, Iran. E-mail: [h\\_akbarian98@sut.ac.ir](mailto:h_akbarian98@sut.ac.ir)

2. Department of Electrical Engineering, Sahand University of Technology, Tabriz, Iran. E-mail: [sedaaghi@sut.ac.ir](mailto:sedaaghi@sut.ac.ir)

## Article Info

### Article type:

Research Article

### Article history:

Received

2023-08-11

Received in revised form

2023-11-16

Accepted

2024-01-16

Published online

2024-01-16

### Keywords:

*Emerging technology, Deep learning, Passive sonar, Artificial intelligence, Automatic target detection*

## ABSTRACT

**Objective:** Artificial intelligence is a part of computer science that emphasizes the creation of intelligent machines in defense equipment and military equipment. Intelligent systems for automatic underwater target recognition are increasingly used in passive sonar to reduce human intervention and related challenges in accurately identifying vessels. Today, highly advanced methods of machine learning and deep learning are being used by the world's navies to identify acoustic targets.

**Methodology:** In this article, recent works in the field of automatic underwater acoustic target recognition are reviewed, and a new method based on deep learning algorithms is presented. In this method, first, the raw audio signals are received from the hydrophones, and after performing the necessary pre-processing, using the Short-time Fourier transform, the spectrogram images related to the passive sonar acoustic data are generated and fed to the model layers for model validation and classification.

**Results:** The obtained results show that the multi-layer structures of the proposed model can automatically extract several features are required for the classification of different ship classes. In this article, common deep learning algorithms are used to identify targets, which can increase identification accuracy and reduce evaluation errors by searching for the most informative features of sonar data.

**Conclusion:** The obtained results show that the recognition accuracy of the proposed model is more than 97%, and its validation loss is less than 3%. In this method, with the relative improvement of classification accuracy, the speed of target recognition has increased significantly.

**Cite this article:** akbarian, H., & Sedaaghi, M. H. (2023). Automatic target recognition of surface vessels in passive sonar using emerging technologies of artificial intelligence and deep learning. *Defensive Future Studies*, 8(30), 149-182.

DOI: 10.22034/dfs.2024.2007357.1719



© The Author(s)  
DOI:

Publisher: AJA Command and Staff University



## شناسایی خودکار شناورهای سطحی در سونار غیرفعال با استفاده از

### فناوری‌های نوظهور هوش مصنوعی و یادگیری عمیق

حسن اکبریان<sup>۱</sup> | محمدحسین صدیقی<sup>۲</sup> ✉

۱. گروه مخابرات، دانشکده مهندسی برق، دانشگاه صنعتی سهند، نام شهر تبریز، نام کشور ایران، رایانامه: h\_akbarian98@sut.ac.ir

۲. استاد، گروه مخابرات، دانشکده مهندسی برق، دانشگاه صنعتی سهند، نام شهر تبریز، نام کشور ایران، رایانامه: sedaaghi@sut.ac.ir

#### چکیده

#### اطلاعات مقاله

**هدف:** سیستم‌های هوشمند شناسایی خودکار اهداف زیرآبی، به‌طور فزاینده‌ای در سونار غیرفعال استفاده می‌شوند تا دخالت انسانی و چالش‌های مربوط به آن را در شناسایی دقیق شناورها کاهش دهند. امروزه روش‌های بسیار پیشرفته یادگیری عمیق به‌منظور شناسایی خودکار اهداف صوتی، توسط نیروهای دریایی جهان در حال بهره‌برداری می‌باشند.

**روش‌شناسی:** در این مقاله روشی جدید در زمینه شناسایی خودکار اهداف صوتی زیر آب مبتنی بر الگوریتم‌های یادگیری عمیق ارائه شده است. در این روش ابتدا سیگنال‌های صوتی خام از هایدروفون‌ها دریافت شده و پس از انجام پیش‌پردازش‌های لازم، با استفاده از تبدیل فرکانسی زمان\_کوتاه، تصاویر طیف‌نگار مربوط به داده‌های صوتی سونار غیرفعال تولیدشده و به لایه‌های پنهان مدل برای اعتبارسنجی و طبقه‌بندی، تغذیه می‌شود.

**یافته‌ها:** نتایج به‌دست‌آمده نشان می‌دهد این مدل می‌تواند به‌طور خودکار چندین ویژگی را که برای دسته‌بندی کلاس‌های مختلف کشتی مورد نیاز است را استخراج نمایند و با جستجوی آموزنده‌ترین ویژگی از داده‌های سوناری، موجب افزایش دقت شناسایی و کاهش خطای ارزیابی گردند.

**نتیجه:** دقت شناسایی مدل پیشنهادی بیش از ۹۷٪ و خطای ارزیابی آن کمتر از ۳٪ است. در این روش با بهبود نسبی دقت طبقه‌بندی، سرعت شناسایی اهداف به‌طور قابل ملاحظه‌ای افزایش یافته است.

نوع مقاله:

مقاله پژوهشی

تاریخ دریافت:

۱۴۰۲/۰۵/۲۰

تاریخ بازنگری:

۱۴۰۲/۰۸/۲۵

تاریخ پذیرش:

۱۴۰۲/۱۰/۲۶

تاریخ انتشار:

۱۴۰۲/۱۰/۲۶

کلیدواژه‌ها:

فناوری نوظهور، یادگیری عمیق، سونار غیرفعال، هوش مصنوعی، شناسایی خودکار اهداف.

**استناد:** اکبریان، حسن & صدیقی، محمدحسین. (۱۴۰۲). شناسایی خودکار شناورهای سطحی در سونار غیرفعال با استفاده

از فناوری‌های نوظهور هوش مصنوعی و یادگیری عمیق. آینده‌پژوهی دفاعی، ۸(۳۰)، ۱۴۹-۱۸۲.

DOI: 10.22034/dfs.2024.2007357.1719



ناشر: دانشگاه فرماندهی و ستاد ارتش جمهوری اسلامی ایران

© نویسندگان.

## مقدمه

تشخیص دقیق و صحیح اهداف نفوذی زیر آبی برای عملیات تروریستی و خرابکارانه، در جنگ‌های نوپدید کنونی بسیار حائز اهمیت است. استفاده از شبکه‌های هیدروفونی در زیر آب برای پایش و مراقبت از اسکله‌های نظامی، بنادر حساس و شناورهای مهم در لنگرگاه‌ها، عاملی مهم در جلوگیری از شکست‌های احتمالی در آینده می‌شود. در شناسایی و تشخیص اهداف زیر آبی با استفاده از اطلاعات به‌دست‌آمده از سونوبوی‌ها و هیدروفون‌ها از طریق روش‌های مدرن هوش مصنوعی، یادگیری ماشین و یادگیری عمیق از میزان دقت و صحت بالایی نسبت به روش‌های شناسایی موجود برخوردار بوده که موجب پایین آمدن درصد خطا و کاهش هشدارهای اشتباه می‌شود. سیستم‌های سونار در هر دو زمینه غیرنظامی و نظامی کاربرد دارند. در حوزه عملیات نظامی، استفاده از سیستم‌های سونار غیرفعال<sup>۱</sup> (PSS) ضروری می‌باشند، به‌عنوان مثال، برای زیردریایی‌ها، علاوه بر کاربرد ناوبری، به‌عنوان ابزاری اصلی برای شناسایی و طبقه‌بندی اهداف سطحی و زیر سطحی، مورد استفاده قرار می‌گیرند (Fernandes & Moura & Seixas, 2022). سونار غیرفعال یکی از روش‌های اصلی تشخیص هدف برای کشتی‌های جنگی مدرن، زیردریایی‌ها، وسایل نقلیه زیردریایی بدون سرنشین<sup>۲</sup> (UUV) و غیره است. در مقایسه با سونار فعال، سونار غیرفعال دارای اختفای خوب و امنیت بالایی است (Wang & et al., 2022).

شناسایی خودکار اهداف صوتی زیر آب<sup>۳</sup> با استفاده از سیگنال‌های منتشرشده از شناورهای سطحی و زیرسطحی، به یکی از مهم‌ترین چالش‌ها در علوم آکوستیک دریایی تبدیل شده است. صدهای منتشرشده از کشتی به سه قسمت نویز ماشین‌آلات، نویز صدای پروانه و جریان آب، و کاویتاسیون تقسیم می‌شود. مطالعات نشان داده است که نویز غالب منتشرشده از کشتی‌ها در سرعت پایین تحت تأثیر نویز ماشین‌آلات قرار می‌گیرد و در سرعت‌های بالاتر، کاویتاسیون به منبع نویزی غالب تبدیل می‌شود. برای کشتی‌های تجاری با سرعت ۸ گره دریایی، نویز کاویتاسیون شروع به انتشار می‌کند. صوت ناشی از ماشین‌آلات داخل کشتی باعث ایجاد ارتعاش در سازه می‌شود که از طریق پوشش پوسته منتقل می‌شود. موتورهای دیزلی با سرعت پایین، نویز تونال با فرکانس پایین

<sup>1</sup> Passive Sonar System (PSS)

<sup>2</sup> Autonomous Underwater Vehicles (UUVs)

<sup>3</sup> Underwater Acoustic Target Recognition (UATR)

مرتبط با نرخ احتراق موتور تولید می‌کنند. سایر ماشین‌آلات از جمله ژنراتورها، پمپ‌ها، و سیستم‌های گرمایش، تهویه مطبوع همگی به صدای کلی منتشرشده به آب اطراف کمک می‌کنند. سطح سروصدای تابشی ناشی از ماشین‌آلات داخل کشتی به نوع و اندازه کشتی و همچنین لرزش پایه‌های مورد استفاده برای موتورها، ژنراتورها و غیره بستگی دارد (Smith & Rigby, 2022). فناوری سنتی در تشخیص هدف صوتی زیر آب به‌وسیله استخراج دستی ویژگی‌ها با استفاده از دانش مهارتی و طبقه‌بندی کننده‌های کم‌عمق انجام می‌شود. این ویژگی‌ها شامل ویژگی‌های حوزه زمان، ویژگی‌های حوزه فرکانس، ویژگی‌های زمان-فرکانس و ویژگی‌های مدولاسیون زمانی است. با این حال، به دلیل ماهیت قوی متغیر زمانی و غیر گاوسی سیگنال‌های صوتی زیر آب، عملکرد طبقه‌بندی اشیا بر اساس ویژگی‌های سنتی ضعیف بوده و تعمیر آن محدود است و دارای راندمان پردازش پایینی است (Jin & Zeng, 2023).

سیگنال‌های صوتی سیگنال‌های توالی زمانی هستند، اما دامنه فرکانس آن‌ها معمولاً حاوی اطلاعات بیشتری است. بنابراین، لازم است پیش‌پردازش و استخراج ویژگی داده‌های خام برای کاهش ابعاد داده‌ها و حذف نویز قبل از ورود به مدل‌های شناسایی انجام شود. سیگنال‌های صوتی منتشرشده از کشتی‌ها یکی از اصلی‌ترین داده‌های تحقیقاتی در زمینه شناسایی اهداف صوتی زیر آب هستند (Luo & et al., 2023). وقتی کشتی‌ها در آب حرکت می‌کنند، صداهایی تولید می‌کنند که اصطلاحاً به آن‌ها صوت منتشرشده از کشتی گفته می‌شود. با توجه به ویژگی‌های منحصر به فرد سیگنال‌های منتشرشده از کلاس‌های مختلف شناورهای سطحی و زیرسطحی، می‌توان با تجزیه و تحلیل این سیگنال‌های صوتی، کلاس خاصی از کشتی‌ها یا زیردریایی‌ها و یا حتی یک شناور ویژه را شناسایی کرد. در روش‌های سنتی، شناسایی اهداف شناور زیر آب از طریق کاربران آموزش‌دیده سونار انجام می‌گیرد که به دلیل نیاز به پایش مداوم صوت‌های دریافتی توسط سونار غیرفعال و اثرات ناشی از شرایط جوی می‌تواند نادرست باشد. از این رو، ایجاد سیستم شناسایی خودکار و قوی، به‌منظور جایگزینی برای کاربر انسانی ضروری به نظر می‌رسد.

در راستای حفظ امنیت دریایی و جلوگیری از نفوذ دشمن، به‌منظور تشخیص صحیح و به‌موقع اهداف شناور سطحی، استفاده از روش‌های هوش مصنوعی، یادگیری ماشین و بالأخص روش نوین یادگیری عمیق، در شناسایی دقیق موقعیت و نوع اهداف، بسیار حائز اهمیت و ضروری است. به دلیل مزیت‌های ذاتی تشخیص اهداف صوتی به‌صورت غیرفعال،

در مقایسه با روش‌های دیگر، این روش توسط نیروهای دریایی جهان مورد استفاده قرار می‌گیرد. امروزه با پیشرفت علم و فناوری، یادگیری عمیق، با قابلیت پردازش مستقل داده‌ها و یادگیری ویژگی، توجه گسترده‌ای را به خود جلب کرده است. مدل‌های یادگیری عمیق توانایی یادگیری قوی دارند و می‌توانند ویژگی‌های بالقوه سیگنال‌ها را استخراج کنند. استفاده از یادگیری ماشین و روش‌های یادگیری عمیق مدرن در تشخیص هدف سونار غیرفعال، موفقیت بسیار خوبی را به همراه داشته است (Huynh, Doan & Kim & 2019). به‌منظور استفاده از شبکه‌های عصبی عمیق برای شناسایی خودکار اهداف، به مقادیر زیادی از داده‌های آموزشی نیاز است تا سطح قابل توجهی از انواع ویژگی‌ها را به دست آورند. با این حال، زمانی که تعداد نمونه‌های موجود کم است، ویژگی‌های یادگیری شده توسط مدل، بیش‌ازحد با داده‌های آموزشی سازگار شده. قابل تعمیم به داده‌های جدید نمی‌باشند (Chandran & Suraj & Mujeeb & Supriya, 2021). به دلیل مشکلات و هزینه‌های بالا در جمع‌آوری مجموعه داده‌های سوناری، نمونه‌های مؤثر مجموعه داده‌های سیگنال‌های صوتی منتشرشده از کشتی کافی نیست. کمبود داده منجر به بیش‌برازش مدل شبکه عمیق در مقیاس بزرگ می‌شود که همگرایی آن دشوار بوده و در نهایت بر دقت طبقه‌بندی تأثیر می‌گذارد (Jin & Liu & Wu, 2020). لی و همکاران<sup>۱</sup> (۲۰۲۲) ثابت کردند که برای شناسایی هدف صوتی زیر آب، جلوگیری از هزینه‌های محاسباتی بالا در پردازش اطلاعات صوتی زیر آب به‌منظور کاهش پیچیدگی‌های محاسباتی، یک رویکرد بسیار مهم است.

در این مقاله سعی شده است به این سؤال اساسی پاسخ داده شود که آیا با استفاده از روش‌های نوین و نوظهور هوش مصنوعی و یادگیری عمیق می‌توان به مدلی از شناسایی اهداف صوتی زیر آب دست یافت که علاوه بر دارا بودن دقت شناسایی اهداف مناسب و مقبول، دارای سرعت عملکرد بالا و حجم کمی از محاسبت باشد تا در سامانه‌های پرتابل و موبایلی بتوان استفاده کرد؟ بدین جهت در این تحقیق، ابتدا با استفاده از فرآیند بخش‌بندی سیگنال‌های آکوستیکی، تعداد داده‌های لازم برای آموزش و ارزیابی شبکه به‌دست آمده است. با توجه به اینکه مؤلفه‌های رزونانس بالا سیگنال آکوستیکی دارای نویزهای کمتری بوده و شباهت آن به سیگنال اصلی بیشتر است (Xiaoquan & Yuan & Cheng, 2018)، در مرحله پیش‌پردازش مؤلفه رزونانس بالا که اغلب در یک باند باریک

<sup>1</sup> Lei & et al.

در فرکانس پایین متمرکز می‌شوند، استخراج گردیده و پس از انجام حذف نویز و کاهش نرخ نمونه‌برداری، تصاویر طیف‌نگار داده‌ها با استفاده از تبدیل فرکانسی STFT به‌دست خواهد آمد. در مرحله استخراج ویژگی، ویژگی‌های بالارزش از داده‌های سوناری برای افزایش نرخ شناسایی مثبت استخراج می‌شوند. مرحله بعدی، استفاده از ویژگی‌های استخراج‌شده به‌عنوان ورودی طبقه‌بندی‌ها است که در آن بخشی از داده‌ها برای آموزش و بقیه برای آزمایش و ارزیابی عملکرد مدل مورد استفاده قرار می‌گیرد. با توجه به عملکرد برجسته شبکه‌های یادگیری عمیق در تشخیص صدا، در این تحقیق از شبکه‌های یادگیری عمیق برای استخراج و تشخیص ویژگی‌های سیگنال صوتی کشتی در زیر آب استفاده شده است. در این مقاله سعی شده است علاوه بر دستیابی به نرخ شناسایی معقول و مناسب، با کاهش تعداد پارامترهای شبکه، سرعت شناسایی مدل نیز تا حد کافی کاهش یافته و حد وسط خوبی بین دقت و سرعت شناسایی برقرار گردد. با استفاده از این مدل پیشنهادی، می‌توان به دقت شناسایی بیش از ۹۷٪ و زیان ارزیابی کمتر از ۰.۳٪ دست یافت که نسبت به روش‌های دیگر شناسایی اهداف صوتی زیرآب، از توانایی مناسب و سرعت شناسایی بهتری برخوردار خواهد بود.

### مبانی نظری و پیشینه‌های پژوهش

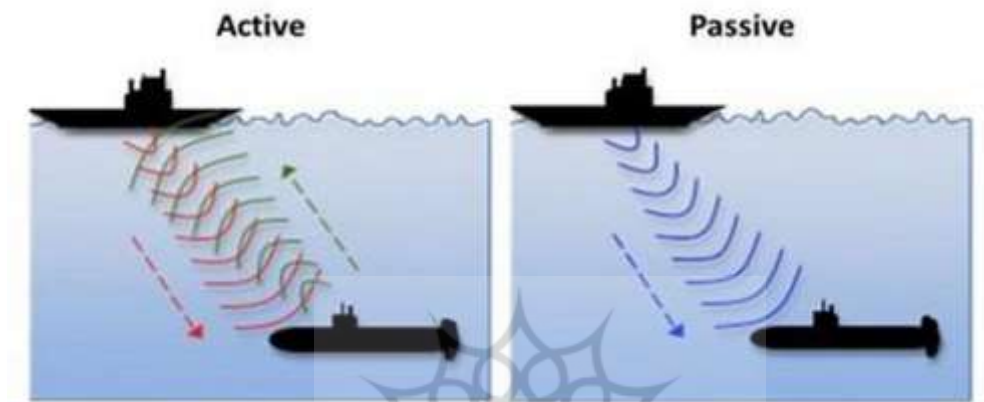
#### سونار<sup>۱</sup>

ناوبری و تشخیص فاصله توسط صوت (سونار)، تکنولوژی است که با استفاده از انتشار صدا در زیر آب قادر به شناسایی دیگر ناوها یا کشتی‌ها است. سیستم‌های سونار شباهت زیادی با سیستم‌های رادار و الکترواپتیکال<sup>۲</sup> دارند با این تفاوت که محیط انتشار امواج در رادار، هوا بوده ولی در سونار محیط انتشار آب است. تفاوت دیگر در نوع امواج است که در رادارها از نوع امواج الکترومغناطیس با فرکانس بالا بوده و در سونار از امواج صوتی با فرکانس پایین استفاده می‌شود. این تفاوت بسیار زیاد در فرکانس باعث برتری سونارها در استفاده از الگوریتم‌های پیچیده‌تر نسبت به رادارها شده است. از طرفی، در سونار محیط انتشار بسیار مؤثر بوده، درحالی‌که در رادار محیط انتشار تأثیر زیادی ندارد. عملکرد سونار بر اساس انتشار موج بین یک هدف و یک گیرنده است. سیستم سونار به‌طور متداول به

<sup>1</sup> Sound Navigation and Ranging

<sup>2</sup> Electro-Optical Systems

دو نوع سونار غیرفعال<sup>۱</sup> و سونار فعال<sup>۲</sup> تقسیم می‌شود. در یک سیستم سونار غیرفعال، امواج صوتی منتشرشده به وسیله‌ی هدف توسط گیرنده دریافت می‌شود که مشابه با آشکارسازی مادون قرمز غیرفعال است. در یک سیستم سونار فعال، امواج از یک فرستنده ارسال و به یک هدف برخورد می‌کنند و انعکاس توسط گیرنده دریافت می‌شود که مشابه با رفتار رادار است (باقری، ۱۳۹۳: ۷).



شکل (۱) نحوه عملکرد سونارهای فعال و غیرفعال (باقری، ۱۳۹۳: ۷)

### هایدروفون<sup>۳</sup>

هایدروفون یا همان میکروفون زیر آب دستگاهی است که برای دریافت صوت در زیر آب طراحی شده است، واژه هایدروفون مشتق شده از واژگان یونانی هایدرو به معنای آب و فون به معنای صوت است<sup>۴</sup>. در حقیقت هایدروفون‌ها با دریافت فشار حاصل از موج صوتی آن را به سیگنال قابل پردازش در خروجی تبدیل می‌نماید. امروزه از هایدروفون‌های پیزوالکتریک به صورت گسترده در کاربردهای زیر آبی نظامی مانند سیستم‌های سونار تجهیزات جنگی (از قبیل زیردریایی‌ها و کشتی‌های جنگی) و همچنین کاربردهای غیرنظامی مورد استفاده قرار می‌گیرند. عملکرد هایدروفون‌ها بر اساس تبدیل فشار صوتی

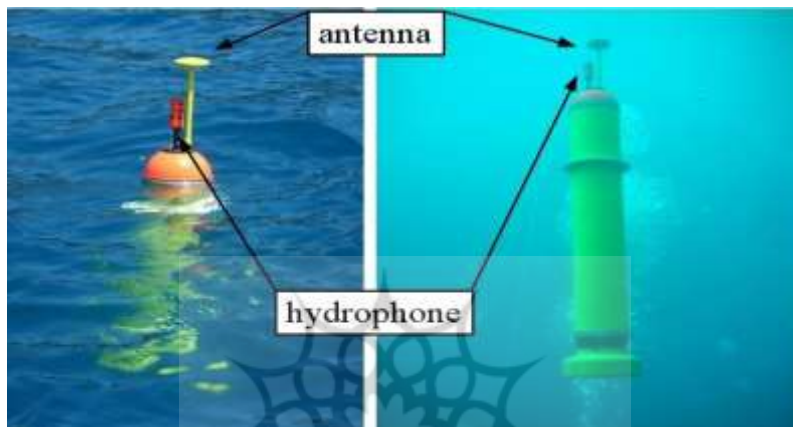
<sup>1</sup> Passive Sonar

<sup>2</sup> Active Sonar

<sup>3</sup> Hydrophone

<sup>4</sup> "Hydro" = "water" and "phone" = "sound"

به سیگنال الکتریکی قابل پردازش است. حسگر پیزوالکتریک<sup>۱</sup> مورد استفاده در هایدروفون، برای دریافت صوت و تبدیل آن به سیگنال الکتریکی، فشار صوت تولیدشده از اهداف را دریافت نموده و کرنش و تنش<sup>۲</sup> تولیدشده از این فشار را با استفاده از مواد پیزوالکتریک مستقیماً به ولتاژ تبدیل می‌نماید. معمولاً برای دریافت صوت، چندین حسگر را در جهت‌های مختلف در کنار هم قرار می‌دهند تا بتوانند جهت صوت را تشخیص دهند (شمس ناتری، ۱۳۹۷: ۴۶).



شکل (۲) هایدروفون استقرار یافته روی سطح و زیر سطح آب دریا (شمس ناتری، ۱۳۹۷: ۴۶).

### یادگیری عمیق<sup>۳</sup>

یادگیری عمیق، زیرمجموعه‌ای از یادگیری ماشین<sup>۴</sup>، مبتنی بر شبکه‌های عصبی مصنوعی<sup>۵</sup> بوده که از عملکرد سلول‌های مغز انسان به نام ساختار عصبی الهام می‌گیرد. در یادگیری عمیق، ضروری نیست همه‌چیز آموزش داده شوند. الگوریتم‌های یادگیری عمیق را می‌توان به روش‌های یادگیری نظارت‌شده<sup>۶</sup>، نیمه نظارت‌شده<sup>۷</sup> و بدون نظارت<sup>۸</sup> تقسیم‌بندی نمود. روش یادگیری نظارت‌شده بر مبنای آموزش یک مدل با استفاده از داده‌های دارای برچسب‌های صحیح که از قبل طبقه‌بندی شده‌اند، به منظور دستیابی به یک مدل، بنا

<sup>1</sup> Piezoelectric

<sup>2</sup> Tension and Strain

<sup>3</sup> Deep Learning

<sup>4</sup> Machine Learning

<sup>5</sup> Artificial Neural Networks (ANN)

<sup>6</sup> Supervised

<sup>7</sup> Semi-Supervised

<sup>8</sup> Unsupervised



شده است، به گونه‌ای که وقتی داده‌های جدید به این ساختار وارد می‌شوند، این مدل می‌تواند نوع خروجی‌ها را به راحتی تخمین بزند. توابع یادگیری بدون نظارت برای یک مجموعه داده بدون ابرچسب، می‌تواند مدلی را آموزش دهد که بدون اطلاع از سیگنال‌های در نظر گرفته شده برای ارزیابی مدل، اطلاعات موجود در داده را بررسی و طبقه‌بندی صحیحی را انجام دهد. استفاده از الگوریتم‌های یادگیری عمیق این روزها به طور گسترده افزایش یافته است. این روش، با تحلیل حجم زیادی از داده‌ها، بسیار عالی کار می‌کند. تکنیک‌های یادگیری عمیق عملکرد بهتر و دقت بیشتری نسبت به سایر تکنیک‌های یادگیری ماشین دارند. شبکه‌های عصبی سنتی فقط شامل ۲ یا ۳ لایه پنهان هستند، در صورتی که شبکه‌های عمیق می‌توانند بیش از ۱۵۰ لایه داشته باشند (Li & Huang, 2019). یادگیری عمیق بر مبنای مجموعه‌ای از الگوریتم‌ها است که ویژگی‌های سطح بالا در داده‌ها را با استفاده از یک شبکه عمیق که شامل چندین لایه پردازشی است، مدل‌سازی می‌کند. بدین صورت که داده‌های سطح پایین را در هر لایه دریافت می‌کند، آن‌ها را پردازش کرده و در نتیجه داده‌های سطح بالاتر را به دست می‌آورد. بدین ترتیب، برای به دست آوردن اطلاعات مورد نیاز ماشین، لزومی به نظارت کامل انسان در هر لحظه نیست. در این حوزه، شبکه‌ها و معماری‌های مختلفی وجود دارد که از جمله این شبکه‌ها می‌توان به شبکه عصبی عمیق، شبکه باور عمیق<sup>۲</sup>، شبکه عصبی بازگشتی<sup>۳</sup>، شبکه عصبی کانولوشنی<sup>۴</sup> CNN، شبکه باقی‌مانده<sup>۵</sup> ResNet، GAN<sup>۶</sup> و غیره اشاره کرد.

### پیشینه پژوهش

در سال‌های اخیر، هوش مصنوعی به عنوان روشی مبتنی بر یادگیری ذهنی به کانون تحقیقات دفاعی و نظامی تبدیل شده است. محمدی فاتح و ابراهیمی (۱۳۹۹) روشی برای شناسایی و رتبه‌بندی فناوری‌های اطلاعاتی نوظهور در عرصه دفاعی نظامی ارائه نمودند. بنابراین، یادگیری عمیق نه تنها می‌تواند به طور خودکار ویژگی‌های آکوستیکی را از سیگنال خام بیاموزد، بلکه می‌تواند استخراج ویژگی‌های عمیق تر و حتی ترکیب ویژگی‌ها

<sup>1</sup> Deep Neural Network (DNN)

<sup>2</sup> Deep Belief Network

<sup>3</sup> Recurrent Neural Network

<sup>4</sup> Convolution Neural Network

<sup>5</sup> Residual Network

<sup>6</sup> Generative Adversarial Network

را بر اساس برخی از پارامترهای ویژگی سطح بالا انجام دهد. جیانگ، ژائو و وانگ<sup>۱</sup> (۲۰۲۲) مدلی با ترکیب شبکه خصمانه مولد کانولوشنال عمیق<sup>۲</sup> اصلاح شده و مدل S-ResNet را برای به دست آوردن دقت طبقه‌بندی خوب پیشنهاد کردند، که توانسته است به دقت طبقه‌بندی ۹۳/۰۴ درصد دست یابد. تیان، چن، وانگ و لیو<sup>۳</sup> (۲۰۲۱) یک واحد باقیمانده چند مقیاسی<sup>۴</sup> پیشنهاد کرده است که قادر به ساخت شبکه پشته کانولوشن عمیق<sup>۵</sup> است. این ساختار انعطاف‌پذیر و متعادل که شبکه عصبی عمیق باقیمانده چند مقیاسی<sup>۶</sup> نامیده می‌شود، برای طبقه‌بندی هدف صوتی زیر آب بکار رفته و توانسته است به دقت شناسایی ۸۳/۱۵ درصد دست یابد. هو، وانگ و لیو<sup>۷</sup> (۲۰۲۱) یک شبکه عصبی کانولوشنال قابل تفکیک<sup>۸</sup> جدید برای شناسایی سیگنال‌های منتشرشده از کشتی‌ها در شکل موج‌های حوزه زمان پیشنهاد کرده است. ویژگی‌های عمیق حاوی اطلاعات داخلی هدف می‌باشند که توسط یک شبکه کانولوشن DWS استخراج می‌شوند. میانگین نرخ شناسایی طبقه‌بندی هنگام آزمایش بر روی یک مجموعه سیگنال آکوستیکی به ۹۰/۹ درصد می‌رسد. چن و همکاران<sup>۹</sup> (۲۰۲۱) روشی برای شناسایی اهداف صوتی زیرآبی پیشنهاد کردند که طیف حاصل از ضبط آنالیز فرکانس پایین<sup>۱۰</sup> را به‌عنوان ورودی شبکه عصبی کانولوشنال (CNN) در نظر گرفته و یک CNN مبتنی بر LOFAR برای شناسایی آنلاین ایجاد کردند. روش LOFAR-CNN پیشنهادی توانسته است به دقت شناسایی ۹۵/۲۲ درصد دست یابد. صفری، ظهیری و خوزین‌قند<sup>۱۱</sup> (۲۰۲۳) با استفاده از مدولاسیون پروانه بر روی سیگنال ارسالی (موسوم به سونار میکرو داپلر<sup>۱۲</sup>) و هسته‌های مختلف، یک ماشین بردار پشتیبان<sup>۱۳</sup> (SVM) برای تشخیص خودکار اهداف متحرک سوناری را پیشنهاد کرده‌اند. دقت

<sup>1</sup> Jiang & Zhao & Wang

<sup>2</sup> Deep Convolutional GAN

<sup>3</sup> Tian, Chen, Wang & Liu

<sup>4</sup> multiscale residual unit (MSRU)

<sup>5</sup> Deep convolution stack

<sup>6</sup> multiscale residual deep neural network (MSRDN)

<sup>7</sup> Hu, Wang & Liu

<sup>8</sup> depthwise separable convolutional neural network

<sup>9</sup> Chen & et al.

<sup>10</sup> Low frequency analysis and recording (LOFAR)

<sup>11</sup> Saffari, Zahiri & Khozein Ghanad

<sup>12</sup> sonar micro- Doppler

<sup>13</sup> support vector machine (SVM)

شناسایی صحیح اهداف برای نسبت‌های مختلف سیگنال به نویز (SNR) و زوایای دید متفاوت بود. هانگ و دیگران<sup>۱</sup> (۲۰۲۱) یک روش طبقه‌بندی با استفاده از ویژگی‌های همجوشی<sup>۲</sup> و شبکه باقیمانده ۱۶ لایه (ResNet18) پیشنهاد کرده است. آزمایش‌های شناسایی بر روی مجموعه نویز منتشر شده از کشتی به نام ShipsEar از یک محیط واقعی انجام شده که دقت شناسایی ۹۴/۳ درصد را نشان می‌دهد. لو و همکاران<sup>۳</sup> یک روش تشخیص هدف صوتی زیر آب (UATR) را بر اساس ResNet پیشنهاد کردند. در روش پیشنهادی، یک روش تحلیل طیفی چند پنجره‌ای<sup>۴</sup> (MWSA) برای حل مشکلی که روش سنتی تحلیل زمان-فرکانس<sup>۵</sup> در استخراج ویژگی‌های سیگنال چندگانه به‌طور هم‌زمان دارد، استفاده کرده است. دقت شناسایی به‌دست‌آمده در این روش ۹۶/۳۲ درصد است. جین و زنگ<sup>۶</sup> (۲۰۲۳) با ترکیب شبکه‌های عصبی ResNet و DensNet به مدلی دست‌یافته‌اند که دارای دقت شناسایی ۹۷/۶۹ درصد است. لی، سونگ و فنگ<sup>۷</sup> (۲۰۲۲) با ترکیب روش‌های تشخیص پوش مدولاسیون بر روی نویز<sup>۸</sup> (DEMON) و ضبط آنالیز فرکانس پایین (LOFAR) به‌منظور استخراج ویژگی‌های اهداف صوتی زیر آب و استفاده از شبکه عصبی کانولوشنی CNN توانسته‌اند به‌دقت شناسایی ۹۴/۰۰ درصد دست یابند. چن و همکاران<sup>۹</sup> (۲۰۲۲) یک روش مبتنی بر یک حافظه کوتاه‌مدت دو جهته<sup>۱۰</sup> (Bi-LSTM) برای کشف ویژگی‌های یک ماسک زمان-فرکانس برای استخراج ویژگی‌های متمایز سیگنال صوتی زیر آب پیشنهاد داده‌اند. شنگ و ژو<sup>۱۱</sup> (۲۰۲۲)، یک روش شناسایی اهداف صوتی زیر آب مبتنی بر ترانسفورماتور<sup>۱۲</sup> UATR را برای شناسایی دو کلاس از اهداف پیشنهاد کرده است که می‌تواند اطلاعات سراسری و محلی روی طیف‌نگارها را ضبط کند و در نتیجه عملکرد UATR را بهبود بخشد. حداکثر دقت شناسایی در این

<sup>1</sup> Hong & et al.

<sup>2</sup> fusion feature

<sup>3</sup> Luo & et al.

<sup>4</sup> Multi-Window Spectral Analysis (MWSA)

<sup>5</sup> time-frequency analysis

<sup>6</sup> Jin & Zeng

<sup>7</sup> Li, Song & Feng

<sup>8</sup> Detection Of Envelope Modulation On Noise (DEMON)

<sup>9</sup> Chen & et al.

<sup>10</sup> bidirectional long short-term memory (Bi-LSTM)

<sup>11</sup> Sheng & Zhu

<sup>12</sup> UATR transformer

روش ۹۶/۹ درصد بوده است. کمال، چاندران و سوپریا<sup>۱</sup> (۲۰۲۱) با استفاده از روش حافظه طولانی کوتاه مدت توانستند به دقت طبقه‌بندی ۹۵/۲ درصد در شناسایی اهداف سونار غیرفعال دست یابند. لی، گو، سان و وانگ<sup>۲</sup> (۲۰۲۲) با استفاده از طیف‌نگار مل ویژگی‌های فرکانسی صوت را استخراج نموده و توسط الگوریتم ResNet طبقه‌بندی را با دقت ۹۷/۲۵ درصد انجام داده است. جوشی، پاریک و امباتکار<sup>۳</sup> (۲۰۲۳) مدلی را پیشنهاد دادند که طیف‌نگار داده‌های صوتی را با استفاده از ضریب فرکانس مل محاسبه نموده و با دقت ۹۶/۷۸ درصد داده‌ها را طبقه‌بندی نمودند.

### روش‌شناسی پژوهش (روش اجرای پژوهش)

#### دیتاست

تمام داده‌های آکوستیکی منتشر شده از کشتی‌ها، از پایگاه داده‌ای به نام ShipsEar حاصل می‌شود که توسط سانتوز دومینگوئز و همکاران<sup>۴</sup> (۲۰۱۶) جمع‌آوری گردیده است. این رکوردها با ضبط کننده‌های صوتی دیجیتال مستقل، Hyd SR-1 ساخته شده است. این ضبط کننده فشرده، شامل هیدروفونی با حساسیت اسمی 193.5 dB re 1V/1Pa- و پاسخ یکنواخت در محدوده فرکانس ۱ هرتز الی ۲۸ کیلوهرتز است. چهار کلاس مربوط به سیگنال‌های منتشر شده از کشتی و یک کلاس مربوط به نویزهای پس‌زمینه است. ۸۰٪ از داده‌های دارای برچسب برای تنظیم دقیق نظارت شده و ۲۰٪ باقیمانده از داده‌های دارای برچسب برای آزمایش دقت شناسایی، به کار می‌روند. ۱۱ نوع کشتی عبوری در چهار کلاس آزمایشی (بر اساس اندازه کشتی) و یک کلاس صدای پس‌زمینه، همان‌طور که در جدول ۱ نشان داده شده است، ادغام می‌شوند.

جدول (۱) تبدیل کلاس‌های مختلف سیگنال‌های صوتی به ۵ کلاس مختلف

جزئیات	نوع کلاس سیگنال‌های صوتی
Fishing Boats, Trawlers, Mussel Boats, Tugboats And Dredgers	Class A
Motorboats, Pilot Boats, And Sailboats	Class B

<sup>1</sup> Kamal, Chandran, Supriya

<sup>2</sup> Li, Guo, Sun, Li, Wang.

<sup>3</sup> Joshi, Pareek, Ambatkar

<sup>4</sup> Santos-Domínguez & et al.

Class C	Passenger Ferries
Class D	Ocean Liners And Ro-Ro Vessels
Class E	Background Noise Recordings

همچنین کلاس‌های مختلف کشتی‌ها در شکل ۳ نشان داده شده است که شامل کشتی ماهیگیری<sup>۱</sup>، قایق پایلوت<sup>۲</sup>، کشتی مسافربری<sup>۳</sup> و اقیانوس‌پیما<sup>۴</sup> است.



(الف)



(ب)



(ج)



(د)

شکل (۳) چهار کلاس کشتی شامل (الف) کشتی ترال ماهیگیری (ب) قایق پایلوت (ج) کشتی مسافربری و (د) اقیانوس‌پیما

### پیش‌پردازش سیگنال‌های صوتی

اولین مرحله در پیش‌پردازش داده‌های صوتی، حذف نویزهای محیطی و انتشاری است. شناورهای سطحی و زیر سطحی از جمله منابع مولد نویز انتشاری هستند که به دلیل حرکت پروانه شناور و عمل کاویتاسیون در پشت شناور ایجاد می‌شود. اغلب این نویزها

<sup>1</sup> fishing trawler

<sup>2</sup> pilot boat

<sup>3</sup> passenger ship

<sup>4</sup> ocean liner

در محدوده فرکانسی ۳ کیلوهرتز می‌باشند، هرچند برخی از نویزها تا محدوده ۱۰ کیلوهرتز هم دریافت شده‌اند. در اینجا برای از بین بردن این نویزها از فیلتر میانه<sup>۱</sup> استفاده شده است.

در مرحله بعد برای آشکار کردن بخش‌های زمانی معنادار در یک سیگنال صوتی که به آن صحنه‌های شنیداری<sup>۲</sup> نیز می‌گویند، آن‌ها را قطعه‌بندی می‌کنند. در این فرآیند ابتدا بین اجزای پراهمیت و کم‌اهمیت در سیگنال صوتی تفکیک‌پذیری انجام می‌گیرد. اجزای بااهمیت، که معمولاً آن‌ها را با نام قطعه<sup>۳</sup> می‌شناسیم دارای اطلاعات مهمی هستند که باید آنالیز شوند. در بین قطعات صوتی، سکوت وجود دارد که در استخراج ویژگی مورد نظر، بخش‌های سکوت حاوی اطلاعات مهمی نیستند. برای این کار از محاسبه انرژی زمان کوتاه و آستانه گذاری برای آن استفاده می‌گردد. تابع انرژی زمان کوتاه یک سیگنال صوتی به صورت معادله زیر تعریف می‌شود (Ozydin,2019).

(۱)

$$E_n = \frac{1}{N} \sum_m [x(m) \cdot w(n-m)]^2$$

که  $x(m)$  سیگنال صوتی زمان گسسته به طول  $N$  و  $n$  اندیس زمانی برای انرژی زمان کوتاه است.  $w(n)$  معرف پنجره مستطیلی بوده که به صورت رابطه زیر نشان داده می‌شود.

(۲)

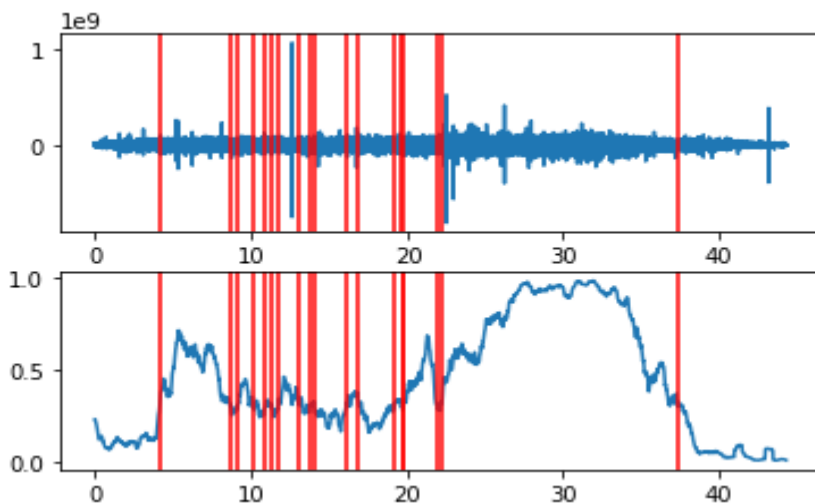
$$w(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \text{otherwise} \end{cases}$$

پژوهشگاه علوم انسانی و مطالعات فرهنگی  
پرتال جامع علوم انسانی

<sup>1</sup> Median Filter

<sup>2</sup> Auditory Scenes

<sup>3</sup> Segment

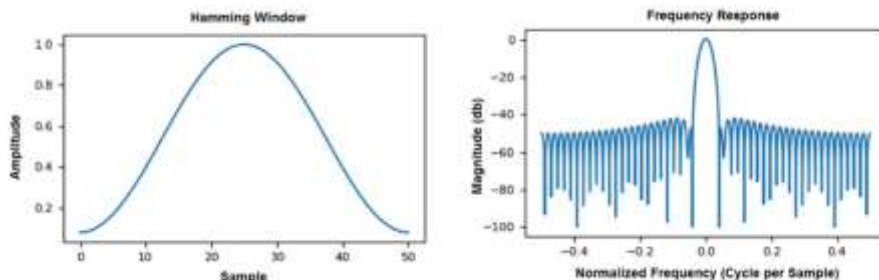


شکل (۴) حذف سکوت و نواحی کم‌اهمیت در سیگنال‌های آکوستیکی

به‌منظور تحلیل زمان-فرکانس سیگنال با استفاده از ابزارهای در دسترس مانند تبدیل فوریه و ...، نیاز است که سیگنال اصلی به‌صورت سیگنال ایستا در نظر گرفته شود. به‌دلیل تجهیزات و ماشین‌آلات بکار رفته در کشتی‌ها سیگنال صوتی زیر آب تولید شده، یک سیگنال پویا و غیر ایستا است. بنابراین بسیاری از خواص تحلیل سیگنال نظیر تبدیل فوریه را نمی‌توان برای این دسته از سیگنال‌ها به کار برد. در واقع مشخصات آماری این سیگنال‌ها در طول زمان تغییر می‌کند و ثابت نیست. برای حل این مشکل، با استفاده از پنجره‌سازی سیگنال، این سیگنال‌ها را به فریم‌های کوچک تبدیل می‌کنیم. این کار با این استدلال انجام می‌شود که مشخصات آماری این سیگنال‌های صوتی در بازه زمانی کوچک ثابت است و در این بازه زمانی کوچک می‌توان سیگنال را شبه ایستا در نظر گرفت و از مزایای سیگنال ایستا بهره برد. معمولاً از پنجره تحلیلی دارای همپوشانی برای در نظر گرفتن تعداد بیشتری از نمونه‌های سیگنال استفاده می‌شود. یکی از پرکاربردترین پنجره‌ها، پنجره همینگ<sup>۱</sup> است که با معادله زیر نشان داده شده است (Tan, Jiang, 2019).

$$W(n) = 0.54 - 0.46 \cos\left(2\pi / (N - 1)\right) \quad (۳)$$

<sup>۱</sup> Hamming Window



شکل (۵) پنجره همینگ و پاسخ فرکانسی آن به منظور کاهش پراکندگی طیفی در سیگنال آکوستیکی

داده‌های آکوستیکی به دست آمده از مجموعه داده، دارای نرخ نمونه برداری  $۵۲/۷۳۴$  کیلوهرتز می‌باشند. به منظور افزایش سرعت پردازش و کاهش حجم محاسباتی، این داده‌ها مجدداً با نرخ پایین تر ( $۲۶/۳۶۷$  کیلوهرتز) نمونه برداری می‌شوند. کوچک کردن یک دنباله  $c[n]$  توسط ضرب  $D$ ، عملیاتی است که در آن یکی از هر  $D$  عنصر مربوط به  $c[n]$  حفظ شده و بقیه مؤلفه‌ها صفر می‌گردد. بنابراین، خروجی  $d[m]$  یک نمونه گیر با فاکتور  $D$  است که با معادله زیر به دست می‌آید:

$$\begin{aligned} d[m] &= c[mD] \\ d[m] &= \sum_{n \in \mathbb{Z}} c[n] h[mD - n] \end{aligned} \quad (۴)$$

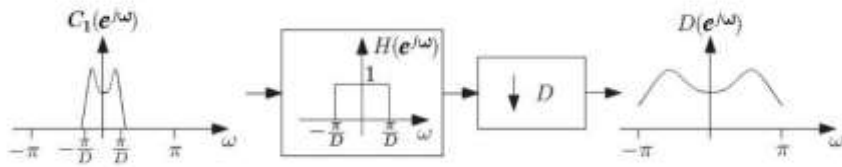
قبل از کاهش نمونه برداری توسط  $D$ ، ابتدا دنباله  $c[n]$  از یک فیلتر  $h[n]$  عبور داده می‌شود.  $h[n]$  به همراه نمونه گیر با فاکتور  $D$ ، توسط معادله زیر نشان داده می‌شود:

$$h[n] = \frac{1}{D} \sin c \left( \frac{n}{D} \right) \quad (۵)$$

یک روش برای به دست آوردن خروجی تبدیل شده از ورودی، گسترش پاسخ فرکانسی سیگنال ورودی در محدوده  $[-\pi, \pi]$  با ضرب  $D$  و سپس ایجاد همپوشانی با فاصله گذاری  $\pi$  است. همچنین این سیگنال را می‌توان با ایجاد همپوشانی توسط فیلتر ضد آلیاسینگ<sup>۱</sup> بر روی ورودی با فاصله  $\pi/D$  و سپس گسترش آن توسط  $D$  تشکیل داد. عمل کاهش نمونه برداری در شکل زیر نشان داده شده است (Yonina, 2018).

<sup>۱</sup> Anti Aliasing





شکل (۶) سیگنال حاصل بعد از اعمال فیلتر ضد آلباسینگ و کاهش نمونه برداری با فاکتور

D(Yonina, 2018)

### محاسبه طیف نگار<sup>۱</sup>

با استفاده از تبدیل زمان فرکانس، مانند تبدیل فوریه، می توان داده های آکوستیکی منتشر شده از کشتی ها را به داده هایی در حوزه فرکانس تبدیل نمود و این امکان فراهم می گردد تا اطلاعات فرکانسی سیگنال صوتی ساطع شده از کشتی ها قابل استخراج باشد. زمانی که تبدیل فوریه بر روی سیگنال صوتی اعمال می گردد، فقط مقادیر فرکانس را استخراج نموده و اطلاعات زمانی سیگنال را از دست می دهد. اگر از این فرکانس ها به عنوان ویژگی استفاده گردد، سیستم ممکن است اطلاعات ابتدایی و انتهایی را از دست دهد. روشی برای محاسبه ویژگی های سیستم وجود دارد که مقادیر فرکانس را همراه با زمانی که در آن مشاهده شده اند، استخراج می کند. به این نمایش بصری فرکانس های یک سیگنال صوتی نسبت به زمان، طیف نگار گفته می شود. در نمایش طیف گرام، یک محور نشان دهنده زمان و محور دوم نشان دهنده فرکانس ها و رنگ ها نشان دهنده بزرگی (دامنه) فرکانس مشاهده شده در یک زمان خاص است (Wang, Ying and Tian:2020). تبدیل فوریه کوتاه مدت<sup>۲</sup> یا به اختصار STFT یکی از روش های اولیه مقایسه سیگنال با این توابع مقدماتی بوده و به صورت معادله زیر تعریف می شود.

$$F_x(t, f; h) = \int_{-\infty}^{+\infty} x(u)h(ut)e^{-j2\pi fu} du \quad (۶)$$

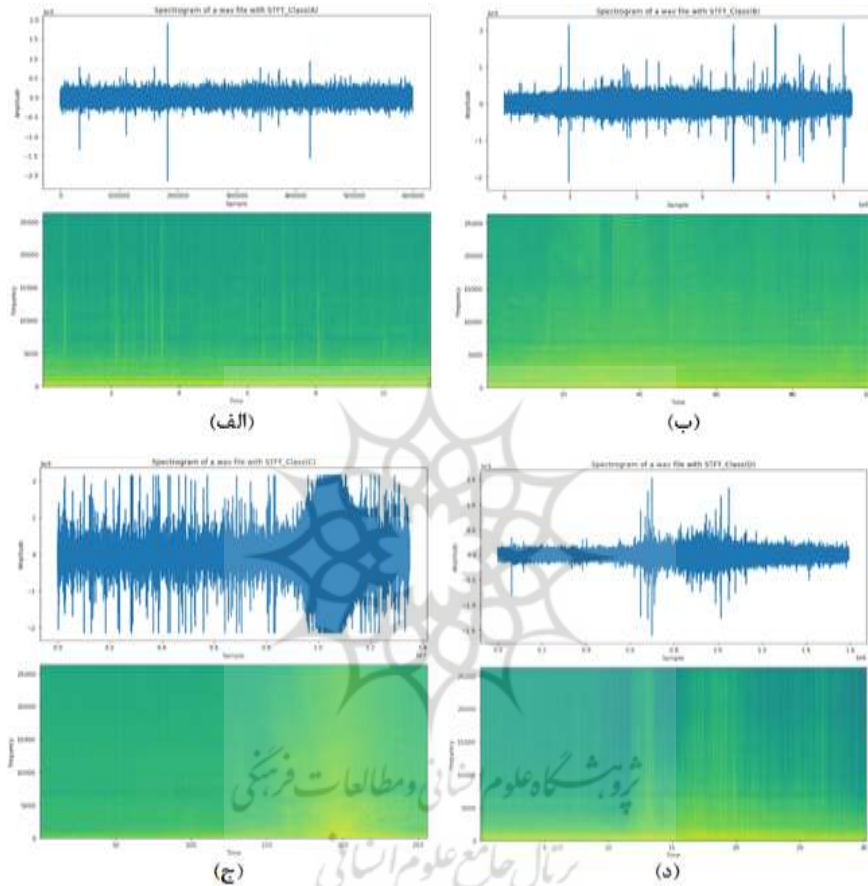
که در آن  $h(t)$  یک پنجره تحلیلی زمان کوتاه است که در اطراف  $h=0$  و  $f=0$  محلی سازی شده است. پنجره  $h(t)$  در امتداد سیگنال  $x(u)$  در نظر گرفته شده و برای هر جابجایی  $h(u-t)$  تبدیل فوریه تابع  $x(u)h(u-t)$  محاسبه می گردد. طیف نگار با معادله زیر به دست می آید:

<sup>1</sup> Spectrogram

<sup>2</sup> Short-Time Fourier Transform

$$S_x(t, f) = \left| \int_{-\infty}^{+\infty} x(u)h(u-t)e^{-j2\pi fu} du \right| \quad (7)$$

شکل زیر تصاویر طیف نگار مربوط به سیگنال‌های صوتی منتشرشده از ۴ کلاس کشتی را نشان می‌دهد.

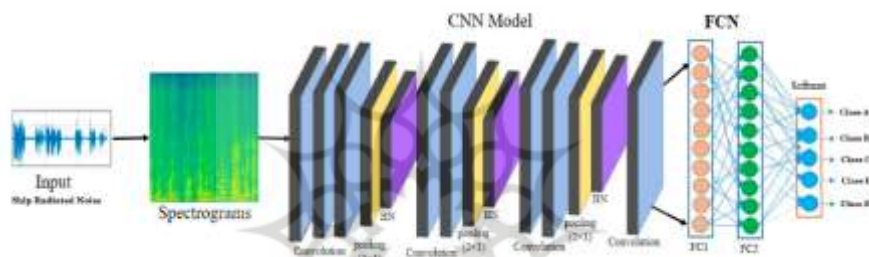


شکل (۷) تبدیل فوریه زمان کوتاه و طیف‌نگار مربوط به سیگنال‌های سوناری. الف) کلاس A ، ب) کلاس B، ج) کلاس C و د) کلاس D.

### شبکه‌های عصبی مبتنی بر کانولوشن

شبکه‌های عصبی کانولوشن یکی از مهم‌ترین روش‌های یادگیری عمیق هستند که داده‌ها در آن‌ها به وسیله چندین مرحله از لایه‌ها آموزش می‌بینند. این روش‌ها قدرتمند بسیار کاربردی بوده و از بهترین روش‌ها در شناسایی خودکار اهداف می‌باشند.

در شکل زیر معماری کلی از یک شبکه عصبی کانولوشن چندلایه برای طبقه‌بندی تصاویر طیف‌نگار نشان داده شده است. در هر شبکه عصبی کانولوشن دو مرحله برای آموزش وجود دارد. مرحله تغذیه روبه‌جلو<sup>۱</sup> و مرحله پس‌انتشار<sup>۲</sup>. در مرحله اول تصویر ورودی طیف‌نگار به شبکه تغذیه می‌شود و عملیات کانولوشن در هر لایه بر روی ورودی هر مرحله اعمال می‌شود. در این مرحله به‌منظور محاسبه میزان خطای شبکه از یک تابع خطا<sup>۳</sup> استفاده شده که با پاسخ صحیح مقایسه می‌گردد. در مرحله بعد بر اساس میزان خطای محاسبه‌شده مرحله پس‌انتشار، با محاسبه گرادیانت هر پارامتر، پارامترهایی که بر خطای ایجادشده در شبکه تأثیرگذار هستند، تغییر پیدا می‌کنند. این به‌روزرسانی پارامترها در مراحل بعدی لایه‌های کانولوشنی تکرار می‌شوند تا ویژگی‌های سطح بالا را استخراج نمایند.



شکل (۸) مراحل مختلف لایه‌های کانولوشن جهت استخراج ویژگی‌های سطح بالای داده‌های آکوستیکی.

در هر لایه کانولوشنال، اندازه فیلتر  $3 \times 3$  است. تابع فعال‌سازی استفاده شده ReLU و آخرین تابع softmax است.

در این مقاله، به‌منظور کاهش محاسبات برای پردازش داده‌ها از طریق کاهش ابعاد ویژگی، به‌جای استفاده از لایه‌های کاملاً متصل برای طبقه‌بندی، در الگوریتم‌های شبکه‌های کانولوشنی از لایه ادغام میانگین سراسری<sup>۴</sup> استفاده می‌شود، زیرا تعداد پارامترها را کاهش می‌دهد و خطر بیش‌برازش را کاهش می‌دهد. در لایه آخر، بردار حاصل از عبور نقشه‌های

<sup>1</sup> feed forward

<sup>2</sup> backpropagation

<sup>3</sup> loss function

<sup>4</sup> Global average pooling

ویژگی از طریق یک لایه ادغام میانگین جهانی به softmax وارد می‌شود. تفاوت اصلی بین ادغام میانگین جهانی و یک لایه کاملاً متصل این است که ادغام میانگین جهانی میانگین تمام مقادیر یک نقشه ویژگی را می‌گیرد، در حالی که یک لایه کاملاً متصل دارای مجموعه‌ای از وزن‌ها است که به همه مقادیر در یک نقشه ویژگی اعمال می‌شود (Zhu, Liu, Yang, and Wang: 2022). این منجر به کاهش تعداد پارامترها در مدل می‌شود که می‌تواند برای آموزش در مجموعه داده‌های کوچک‌تر و برای جلوگیری از برازش بیش‌ازحد مفید باشد.

مشابه لایه‌های ادغام بیشینه<sup>۱</sup>، از لایه‌های ادغام میانگین سراسری (GAP) برای کاهش ابعاد فضایی یک تانسور سه‌بعدی استفاده می‌شود. باین‌حال، لایه‌های GAP کاهش ابعاد بیشتری را انجام می‌دهند که در آن یک تانسور با ابعاد  $h \times w \times d$  به ابعاد  $d \times 1 \times 1$  کاهش می‌یابد. لایه‌های GAP هر نقشه ویژگی  $h \times w$  را به یک عدد واحد کاهش می‌دهند و صرفاً میانگین تمام مقادیر  $hw$  را در نظر می‌گیرند. GAP به صورت معادله زیر تعریف می‌شود:

$$GAP(F) = \frac{1}{W \cdot H} \sum_{x=1}^W \sum_{y=1}^H F_{x,y}, \quad F \in \mathbb{R}^{W \times H \times C} \quad (8)$$

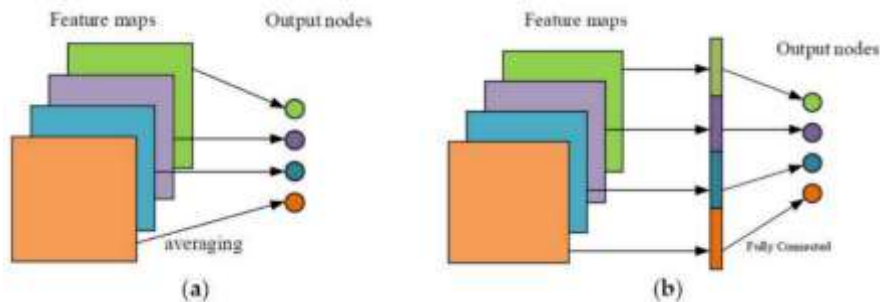
$F$  نشان‌دهنده نقشه ویژگی ورودی، و  $x$  و  $y$  مختصات مکان‌های مختلف فضایی<sup>۲</sup> می‌باشند.  $W$ ،  $H$  و  $C$  به ترتیب عرض، ارتفاع و عمق  $F$  (یعنی تعداد کانال‌ها) را نشان می‌دهند. اگر لایه طبقه‌بندی نهایی حاوی بردار وزن  $W \in \mathbb{R}^C$  برای یک کلاس هدف باشد، آنگاه نمره کلاس  $s$  را می‌توان به صورت محاسبه کرد.

$$s = W^T f = \sum w_c \cdot f_c, \quad f = GAP(f) \text{ and } f \in \mathbb{R}^C \quad (9)$$

اثر اتصال کامل و ادغام میانگین سراسری بر روی نقشه‌های ویژگی در شکل ۸ نشان داده شده است.

<sup>1</sup> max pooling layers

<sup>2</sup> different spatial locations



شکل (۸) نمایش ادغام سراسری در مقابل اتصال کامل. (الف) ادغام میانگین سراسری (ب) لایه‌های کاملاً متصل.

در این تحقیق ویژگی‌های استخراج شده از داده‌های سوناری در حوزه زمان-فرکانس در قالب تصاویر طیف‌نگار ذخیره می‌شوند. این تصاویر به‌عنوان ورودی مدل طبقه‌بندی کننده پیشنهادی مورد استفاده قرار می‌گیرند. به‌منظور اجرای آزمایش‌های در نظر گرفته‌شده، داده‌ها با فرکانس نمونه‌برداری ۲۶/۳۶۷ کیلوهرتز دوباره نمونه‌برداری می‌شوند. هر یک از فایل‌های صوتی به بخش‌های<sup>۱</sup> متعددی تقسیم می‌شوند تا برای ورودی الگوریتم‌های یادگیری عمیق پردازش شوند. طول خاص بخش‌بندی بر اساس ماهیت سیگنال به‌دست‌آمده و ابعاد ورودی الگوریتم مورد استفاده، تعیین می‌شود. با در نظر گرفتن ویژگی‌های سیگنال‌های صوتی سونار غیرفعال، منابع محاسباتی و دقت طبقه‌بندی، هر سیگنال به بخش‌های ۴ ثانیه‌ای تقسیم کردیم. هر بخش توسط مدل به‌صورت مستقل طبقه‌بندی می‌شود. با انجام عمل بخش‌بندی و محاسبه طیف فرکانسی، ۵۶۷۱ تصویر طیف‌نگار در ابعاد  $3 \times 224 \times 224$  به‌دست می‌آید که متعلق به ۵ کلاس تعریف شده از انواع شناورها است. ۷۰٪ از این داده‌ها برای آموزش<sup>۲</sup>، ۲۰٪ برای اعتبارسنجی<sup>۳</sup> و ۱۰٪ برای آزمایش<sup>۴</sup> استفاده شده است. عملکرد الگوریتم‌ها با

<sup>1</sup> Segment

<sup>2</sup> Train

<sup>3</sup> Validation

<sup>4</sup> Test

پارامترهایی مانند صحت<sup>۱</sup>، دقت<sup>۲</sup>، بازیابی<sup>۳</sup> و امتیاز F1<sup>۴</sup> ارزیابی می‌شود. معادلات ارزیابی ارزیابی از طریق روابط زیر قابل محاسبه می‌باشند (Irfan, Jiangbin, Ali, and Iqbal :2021).

$$\text{False Positive Rate} = \frac{FP}{TN + FP} \quad (10)$$

$$\text{True Positive Rate} = \frac{TP}{TP + FN} \quad (11)$$

$$\text{precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positive}} \quad (12)$$

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{TP + TN + FP + FN} \quad (13)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (15)$$

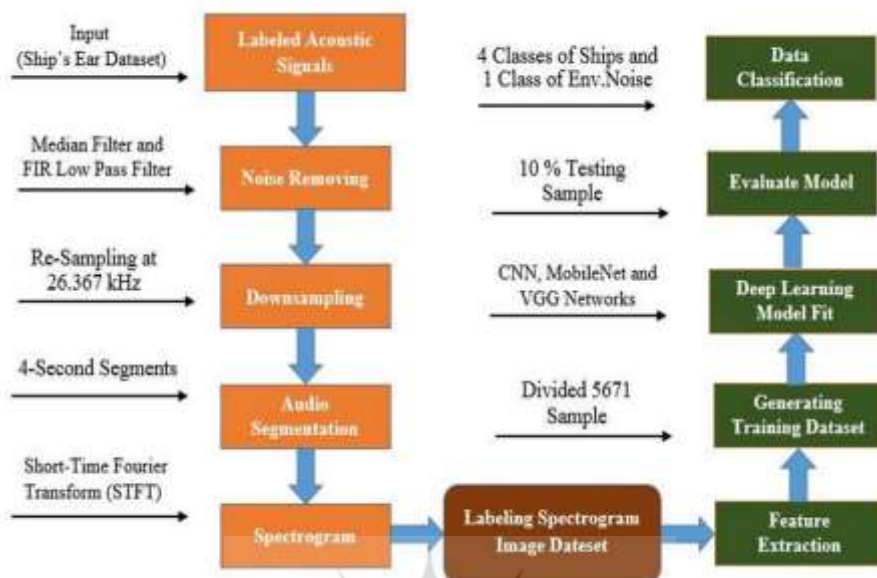
ارزیابی بدین صورت انجام می‌گیرد که پس از اعمال داده‌های آزمایشی به مدل پیشنهادی، تعداد داده‌هایی که به درستی تشخیص داده شده‌اند، به عنوان معیار مثبت صحیح (TP)، داده‌هایی که به اشتباه شناسایی شده‌اند، مثبت کاذب (TN)، داده‌هایی که به درستی رد شده‌اند، منفی صحیح (TN) و داده‌هایی که به اشتباه رد شده‌اند، منفی کاذب (FN) در نظر گرفته شده و شمارش می‌شوند. با استفاده از این دسته‌بندی، معیارهای معرفی شده برای ارزیابی مدل‌های پیشنهادی مورد استفاده قرار می‌گیرند و در پایان نتایج حاصل از به کارگیری الگوریتم‌های مختلف یادگیری عمیق، با هم مقایسه شده‌اند. ساختار کلی مدل پیشنهادی در شکل ۹ نشان داده شده است.

<sup>1</sup> Accuracy

<sup>2</sup> Precision

<sup>3</sup> Recall

<sup>4</sup> F1-Score



شکل (۹) بلوک دیاگرام مدل پیشنهادی برای شناسایی و طبقه‌بندی داده‌های آکوستیکی سونار غیر فعال.

### شبیه‌سازی و نتایج

در این پژوهش برای شبیه‌سازی مدل از برنامه پایتون با استفاده از کتابخانه Keras با پس‌زمینه Tensorflow استفاده شده و در ۵۰ تکرار<sup>۱</sup> با استفاده از بهینه‌ساز نزولی گرادیان تصادفی<sup>۲</sup> با اندازه دسته‌های ۶۴ تایی از داده‌های طیف‌نگار و نرخ یادگیری اولیه ۰/۰۰۱ آموزش داده شد. دقت و تابع هزینه مدل برای داده‌های آموزشی و داده‌های اعتبارسنجی به دست آورده شده است. مدل‌های مبتنی بر شبکه‌های کانولوشنی در دو مرحله حرکت روبه‌جلو<sup>۳</sup> و پس انتشار<sup>۴</sup> آموزش داده می‌شوند. ویژگی‌های سطح پایین و سطح بالا که در داده‌های تزریق شده به مدل وجود دارد، توسط فیلترهای بهینه‌شده در مرحله پس انتشار استخراج می‌گردند و برای هر کلاس، فیلترها مهم‌ترین نواحی را برای تشخیص

<sup>۱</sup> Epoch

<sup>۲</sup> Stochastic Gradient Descent Optimizer

<sup>۳</sup> Feed-Forward

<sup>۴</sup> Back-Propagation

تصاویر اسپکتروگرام مربوط به آن کلاس از کشتی‌ها را نشان می‌دهد. در نتیجه، دقت نهایی مدل اصلاح‌شده پیشنهادی توسط فیلترها یا ویژگی‌های استخراج‌شده از لایه‌های مختلف مدل تعیین می‌گردد و تصویرسازی فیلترها به‌منظور ارزیابی عملکرد مورد استفاده قرار می‌گیرد. معمولاً فیلترهای لایه اول ویژگی‌های رنگی و لبه‌های جهت‌دار موجود در تصاویر طیفی را جداسازی و تفکیک می‌کنند. فیلترهای میانی بافت‌های به‌وجود آمده از همپوشانی رنگ‌های مختلف و لبه‌ها را شناسایی و نشان‌گذاری می‌کنند. بدین ترتیب الگوهای خاصی از جهت‌گیری و رنگ‌بندی مربوط به طیف‌نگارهای کلاس‌های مختلف کشتی توسط فیلترهای لایه‌های پایانی کانولوشنی مدل پیشنهادی استخراج می‌شوند. نمودار تابع هزینه و دقت مدل در هر دوره در شکل ۱۰ نشان داده شده است. نتایج به‌دست‌آمده نشان می‌دهد که استفاده از روش‌های یادگیری عمیق منتج به نتایج مطلوبی در شناسایی صحیح اهداف آکوستیکی و کاهش تشخیص اشتباه می‌گردد. همچنین مشاهده می‌گردد در الگوریتم‌های مورد آزمایش واقع شده، شبکه‌های دوبعدی موبایل‌نت و شبکه‌های VGG از دقت طبقه‌بندی بهتری نسبت به شبکه‌های CNN، ResNet، LeNet، DensNet و ... برخوردار هستند. این امر به این دلیل است که در شبکه‌های موبایل‌نت از نوع جدیدی از کانولوشن‌های دوبعدی با عنوان کانولوشن قابل تفکیک عمقی<sup>۱</sup> به‌عنوان واحد اصلی استفاده می‌شود که شامل کانولوشن عمقی<sup>۲</sup> و کانولوشن نقطه‌ای<sup>۳</sup> است. همچنین در شبکه VGG از لایه‌های مختلف کانولوشنی است که دارای عمق بیشتر و پارامترهای بیشتری نسبت به شبکه‌های CNN است. علاوه بر این، در این تحقیق به دلیل استفاده از روش ادغام میانگین در انتهای لایه‌های الگوریتم‌های کانولوشنی مدل پیشنهادی به‌جای لایه کاملاً متصل<sup>۴</sup> سعی شده است از پیچیدگی و افزایش محاسبات کاسته شود. به‌ویژه در شبکه موبایل‌نت، به دلیل حذف کانولوشن‌های اضافی نقطه‌ای و

<sup>1</sup> Depthwise Separable Convolution

<sup>2</sup> Depthwise Convolution

<sup>3</sup> Pointwise Convolution

<sup>4</sup> Fully Connected



عمقی در انتهای الگوریتم، مدت زمان انجام محاسبات آموزش و ارزیابی، صرف شده توسط روش پیشنهادی، کمتر از مدل‌های طبقه‌بندی کننده مبتنی بر روش‌های کانولوشنی مورد اشاره است. نتایج ارزیابی مدل توسعه‌یافته نشان می‌دهد که مدل ارائه‌شده کارایی لازم جهت تشخیص و طبقه‌بندی کلاس‌های مختلف کشتی را دارد. در برخی از روش‌ها نیز مدل‌های یادگیری ماشین مانند ماشین بردار پشتیبان<sup>۱</sup> (مدل ارائه‌شده توسط صفری، ظهیری و خوزین‌قند، ۲۰۲۳) و نزدیک‌ترین همسایه<sup>۲</sup> k برای طبقه‌بندی اهداف صوتی زیر آب استفاده کردند (مدل ارائه‌شده توسط عظیمی سجادی، دوبک و یائو، ۲۰۰۰). روش‌های شناسایی اهداف توسط الگوریتم‌های یادگیری ماشین طولانی و فرسایشی می‌باشند. در روش‌های مبتنی بر تکنیک‌های یادگیری ماشین، نیاز است که ابتدا مسئله را به چندین قسمت تقسیم کرده و سپس نتیجه را ترکیب نمود. این روش‌ها نمی‌توانند مقدار قابل توجهی از داده‌ها را به‌طور خودکار پردازش نموده و ویژگی‌های عمیق را استخراج نمایند که در طبقه‌بندی اهداف بسیار حائز اهمیت است. علاوه بر این، روش‌های متداول یادگیری ماشین قادر به یادگیری مناطق کلاتر نیستند و بیشتر به ویژگی‌های اهدافی که به‌صورت دستی برای مدل تعریف می‌گردد، نسبت به اطلاعات کامل پس‌زمینه داده‌های آموزشی، متکی می‌باشند. این مشکلات با استفاده از معماری یادگیری عمیق قابل حل است.

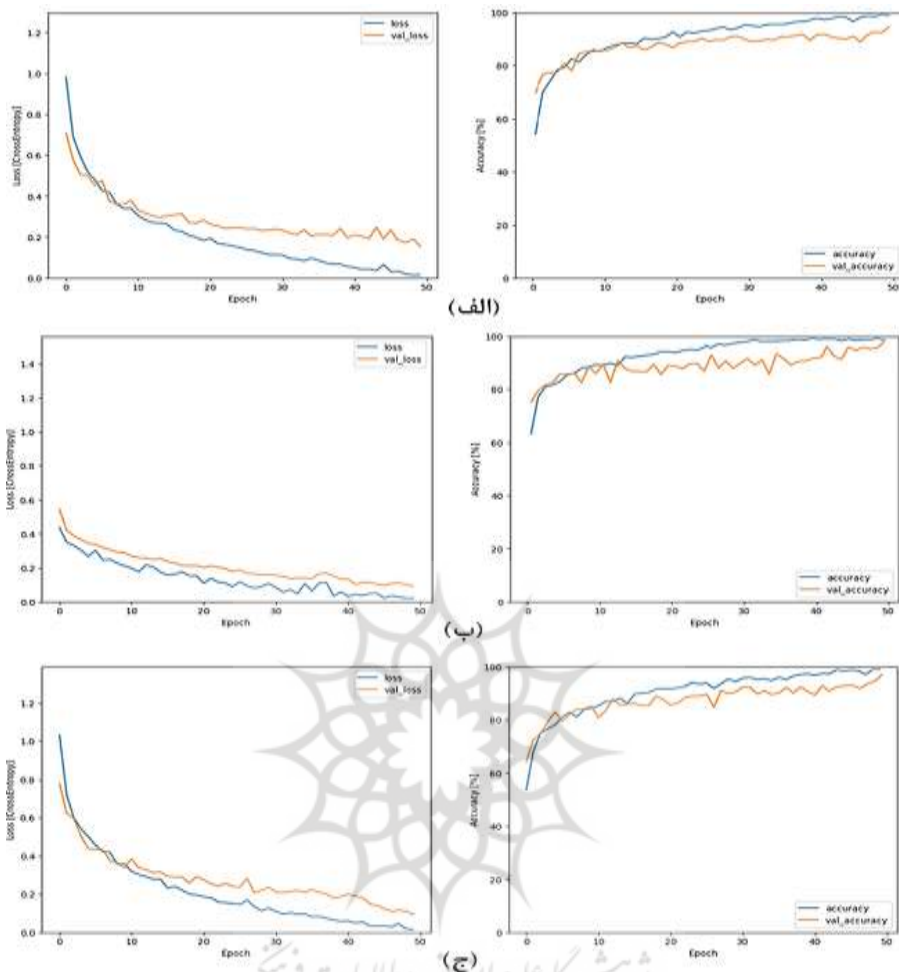
باخ و همکاران روشی را مبتنی بر طیف‌نگار و صرفاً با استفاده از شبکه‌های کم‌عمق CNN ارائه نمودند که نسبت به روش ما از دقت بسیار پایین‌تری برخوردار است (تقریباً ۸۷ درصد). روش جیانگ، ژائو و وانگ که مدلی با ترکیب شبکه خصمانه مولد کانولوشنال عمیق DCGAN اصلاح‌شده و مدل S-ResNet است، تعداد داده‌های ورودی را با استفاده از شبکه خصمانه مولد به‌صورت مصنوعی افزایش داده که دقت و سرعت شبکه پیچیده آن‌ها از روش ما کمتر است. چن و همکاران، بر خلاف روش ما از روش LOFAR برای

<sup>1</sup> Support Vector Machine (SVM)

<sup>2</sup> k-nearest neighbors algorithm

تولید طیف‌نگار استفاده کرده و در شبکه استاندارد آن را طبقه‌بندی نمودند که به دقت ۹۵/۲۲ دست یافتند. هرچند تعداد بسیار کمی از روش‌های شناسایی انجام‌شده در این حوزه، مانند روش جین و زنگ (ترکیبی از شبکه‌های ResNet و DensNet)، با ترکیب الگوریتم‌های مختلف شبکه عصبی، افزایش لایه‌های پنهان و عمق دادن به شبکه که منجر به افزایش پیچیدگی‌های محاسباتی می‌گردد، موفق شدند نسبت به روش پیشنهادی ما، به مقدار ناچیزی دقت شناسایی اهداف را ارتقا دهند (در حدود ۰/۳ درصد)، اما در عوض روش پیشنهادی در این پژوهش علاوه بر داشتن دقت طبقه‌بندی و شناسایی بسیار مناسب و معقول، از سرعت پردازش و طبقه‌بندی بسیار خوبی نیز برخوردار است که حتی توسط رایانه‌های معمولی و وسایل ارتباطی پورتابل قابل اجرا بوده و عملکرد خوبی را ارائه می‌دهد.





شکل (۱۰) نمایش دقت آموزش و زیان ارزیابی در الگوریتم‌های مختلف روش پیشنهادی (الف)

.CNN (ب) MobileNet و (ج) VGG19.

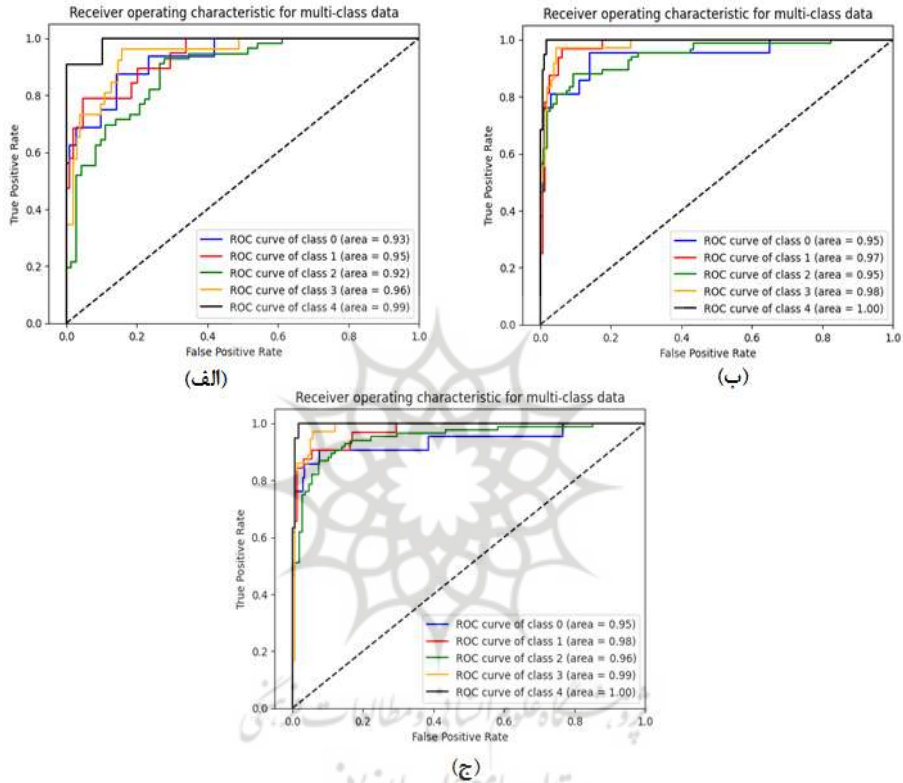
در روش‌های شناسایی خودکار اهداف با استفاده از تکنیک‌های یادگیری عمیق، به‌منظور ارزیابی قابلیت مدل‌ها در تشخیص کلاس‌های مختلف اهداف شناور سطحی، از منحنی ROC<sup>۱</sup> استفاده می‌گردد. این منحنی به‌وسیله ترسیم نرخ صحیح مثبت<sup>۲</sup> (TPR) نسبت به نرخ مثبت کاذب<sup>۳</sup> (FPR) ایجاد می‌شود. با توجه به نتایج به‌دست آمده و مقایسه

<sup>۱</sup> Receiver Operator Characteristic

<sup>۲</sup> True Positive Rate

<sup>۳</sup> False Positive Rate

نسبت شناسایی صحیح داده‌های آزمایشی مربوط به کلاس‌های مختلف کشتی‌ها، مشاهده می‌شود مدل VGG نسبت مدل‌های دیگر (ResNet، LSTM، LeNet، CNN و MobileNet) از نرخ صحیح مثبت بیشتر و بهتری برخوردار است. مساحت سطح زیر منحنی مربوط به کلاس‌های مختلف گویای این امر است.



شکل (۱۱) مقایسه شناسایی صحیح کلاس‌های مختلف با منحنی ROC در الگوریتم‌های

مختلف روش پیشنهادی (الف) CNN، (ب) MobileNet و (ج) VGG19.

نتایج دقت شناسایی سیگنال‌های صوتی زیرآب برای روش‌های مختلف یادگیری عمیق در جدول زیر آورده شده است. نتایج نشان می‌دهد که مدل طبقه‌بندی کننده با استفاده از الگوریتم VGG19 به دقت ۹۷/۳۴ درصد و با استفاده از الگوریتم MobileNet به دقت ۹۶/۸۶ درصد دست یافته است که این روش‌ها از روش‌های استاندارد LeNet و CNN عملکرد بهتری دارند. همچنین صحت شناسایی ۹۸/۴۷ درصد، فراخوانی ۹۹/۲۸ درصد و

امتیاز F1، ۹۹/۱۴ درصد، از دیگر نتایج به دست آمده هستند. با بررسی نتایج می توان دریافت که روش پیشنهادی در تمامی معیارهای ارزیابی برای تشخیص دقیق و صحیح اهداف شناورهای سطحی بر اساس سیگنال های آکوستیکی دریافتی از آنها، نسبت به سایر روش های رایج، عملکرد مناسب و قابل اعتمادتری در تشخیص خودکار اهداف دارد. جدول (۱) مقادیر صحت طبقه بندی، دقت شناسایی، فراخوانی، و امتیاز F1 روش های مختلف یادگیری عمیق بر حسب درصد برای شناسایی خودکار اهداف زیر آب. بهترین معیار در هر ستون به صورت پررنگ تر نشان داده شده است.

Method	Accuracy	Precision	Recall	F1-Score
<b>LeNet</b>	<b>٪۷۰/۰۰</b>	٪۷۸/۰۰	٪۸۵/۰۰	٪۷۷/۰۰
<b>CNN</b>	٪۹۳/۴۴	٪۹۴/۳۹	٪۹۶/۰۰	٪۹۲/۸۱
<b>VGG 19</b>	٪۹۷/۳۴	٪۹۸/۴۷	٪۹۹/۲۸	٪۹۹/۱۴
<b>MobileNet</b>	٪۹۶/۸۶	٪۹۸/۳۳	٪۹۹/۰۴	٪۹۸/۸۴

با توجه به اهمیت سرعت شناسایی خودکار اهداف به عنوان عامل برتری در میدان های نبرد و تصمیم گیری ها، سعی شده است از الگوریتم هایی استفاده گردد که علاوه بر داشتن دقت بالای شناسایی اهداف، از پیچیدگی محاسباتی کمتری نسبت به مدل های دیگر یادگیری عمیق برخوردار باشند. بازده محاسباتی و سرعت عمل مدل های مختلف در جدول (۲) نشان داده شده است. با نگرش به نتایج به دست آمده، مشاهده می شود که تعداد پارامترهای محاسباتی الگوریتم موبایل نت از سایر الگوریتم ها کمتر بوده و از سرعت بیشتری در پردازش داده های آکوستیکی و آموزش مدل برخوردار است. از آنجایی که فرآیندهایی مانند مدت زمان بارگذاری مدل و مدت زمان وارد کردن داده ها برای مدل های مختلف عملکرد یکسانی دارند، در مقایسه الگوریتم ها در نظر گرفته نمی شوند.

جدول (۲) مقادیر صحت طبقه بندی، دقت شناسایی، فراخوانی، و امتیاز F1 روش های مختلف یادگیری عمیق بر حسب درصد برای شناسایی خودکار اهداف زیر آب. بهترین معیار در هر ستون به صورت پررنگ تر نشان داده شده است.

Method	Number of Parameters (million)	Computation Time of each Step (second)	Computation Time of all epochs (minute)
<b>LeNet</b>	۳۶ mil	۶۰/۵ sec	۷۵ min
<b>CNN</b>	۵/۸ mil	۱۸/۵ sec	۲۱ min

<b>VGG 19</b>	۲۰/۲ mil	۳۹/۵ sec	۳۴ min
<b>MobileNet</b>	۲/۲ mil	۱۲/۶ sec	۱۸ min

### نتیجه‌گیری و پیشنهادها

تشخیص دقیق و صحیح اهداف نفوذی زیرآبی برای مقابله با عملیات تروریستی و خرابکارانه، در جنگ‌های نوپدید کنونی بسیار حائز اهمیت است. استفاده از شبکه‌های هیدروفونی در زیر آب برای پایش و مراقبت از اسکله‌های نظامی، بنادر حساس و شناورهای مهم در لنگرگاه‌ها، عاملی مهم در جلوگیری از شکست‌های احتمالی در آینده می‌شود. شناسایی و تشخیص اهداف زیر آبی با استفاده از اطلاعات به‌دست‌آمده از سونوبوی‌ها و هیدروفون‌ها با استفاده از روش‌های مدرن یادگیری عمیق از میزان دقت و صحت بالایی نسبت به روش‌های شناسایی موجود برخوردار بوده که موجب پایین آمدن درصد خطا و کاهش هشدارهای اشتباه می‌شود. هدف اصلی این کار توسعه طبقه‌بندی کننده هدف سونار غیر فعال است که می‌تواند با دقت کافی در محیط صوتی چالش‌برانگیزی از دریا، با بهره‌برداری از ساختارهای طیفی زمانی که اغلب ظریف و در سیگنال‌های دریافت شده نهفته هستند، انجام می‌دهد. در این مقاله، یک مدل جدید برای تشخیص هدف صوتی زیر آب پیشنهاد گردید که مبتنی بر روش‌های یادگیری عمیق با مکانیسم‌های اصلاح‌شده در انتهای شبکه در مجموعه داده‌های آکوستیکی است. برای استفاده از داده‌های آکوستیکی سونار در مدل پیشنهادی، پس از انجام پیش‌پردازش‌های لازم بر روی آن‌ها، تصاویر طیف‌نگار داده‌ها به‌دست‌آمده و مجموعه داده جدیدی به‌صورت تصاویر رنگی تشکیل خواهد شد. در مراحل شناسایی و طبقه‌بندی اهداف توسط مدل پیشنهادی، از شبکه‌های کانولوشنی نوین با کمترین تعداد پارامتر استفاده شده است. در این تحقیق نشان داده شده است که با استفاده از این شبکه می‌توان به‌طور هم‌زمان ویژگی‌های زمان و فرکانس داده‌ها را با تولید طیف‌نگارهای داده‌های آکوستیکی منتشرشده از کشتی‌ها استخراج کرد. مدل طبقه‌بندی کننده پیشنهادی، عملکرد خوبی در طبقه‌بندی و شناسایی سیگنال‌های هدف با دقت شناسایی بالا و خطای کم، از خود نشان داده است. در مقایسه عملکرد، نتایج آزمون‌های طبقه‌بندی نشان داد که دقت روش پیشنهادی به طرز قابل توجهی از تکنیک‌های یادگیری عمیق رایج بهتر است. در این پژوهش سعی شده تا تعادل مناسبی بین افزایش دقت شناسایی اهداف و کاهش پیچیدگی

محاسباتی (افزایش سرعت مدل) برقرار گردد، به طوری که با بهبود نسبی دقت طبقه‌بندی، تعداد پارامترها و میزان محاسبات به میزان محسوسی کاهش یابد. به طور کلی، هدف از اجرای این پژوهش، دستیابی به مدلی بود تا بتواند قابلیت اجرا بر روی رایانه‌ها و تلفن‌های همراه موجود را داشته باشد. روش پیشنهادی پشتیبانی فنی خوبی برای طبقه‌بندی هدف صوتی زیر آب و شناسایی سیستم سونار ارائه می‌کند. همچنین از این مدل علاوه بر استفاده در خدمات نظامی، می‌توان در امداد و نجات دریایی، سیستم‌های کنترل و نظارت زیر آب برای جلوگیری از نفوذهای زیر آب به وسیله غواصان و شناورهای بدون سرنشین زیر آبی و همچنین صنعت شیلات و ماهی‌گیری بهره‌برداری نمود.

### قدردانی

از خبرگان توانمندی که در مدت پژوهش، سخاوتمندانه دانش و وقت خود را در اختیار محققان این پژوهش قرار دادند، بسیار سپاسگزاریم.

### منابع

- باقری، وحید. (۱۳۹۳). استخراج ویژگی از داده‌های سونار غیرفعال مقاوم در برابر نویز، پایان‌نامه کارشناسی ارشد، دانشکده مهندسی برق، دانشگاه خلیج فارس بوشهر.
- شمس ناتری، مجتبی. (۱۳۹۷). طراحی، تحلیل، شبیه‌سازی و مدل‌سازی هیدروفون با قابلیت تشخیص جهت صوت با استفاده از تکنولوژی MEMS، رساله دکترا، دانشکده مهندسی برق، دانشگاه صنعتی نوشیروانی بابل.
- محمدی فاتح، اصغر و ابراهیمی، سیدعباس. (۱۳۹۹). شناسایی و رتبه‌بندی فناوری‌های اطلاعاتی نوظهور در بخش دفاعی - نظامی، فصلنامه آینده‌پژوهی دفاعی. ۱۷(۵): ۱۷۱-۱۴۳.

- Chandran, S. Kamal, S. Mujeeb, A. Supriya, M. H. (2021). Passive Sonar Target Classification Using Deep Generative  $\beta$ -VAE. *IEEE SIG. Process. Lett.*, VOL. 28. <https://doi.org/10.1109/LSP.2021.3071255>.
- Chen, J. Liu, C. Xie, J. An, J. Huang, N. (2022). Time-Frequency Mask-Aware Bidirectional LSTM: A Deep Learning Approach for Underwater Acoustic Signal Separation. *Sensors 2022*, 22, 5598. <https://doi.org/10.3390/s22155598>.

- Chen, J. Liu, J. Liu, C. Zhang, J. Han, B. (2021). Underwater Target Recognition based on Multi-Decision LOFAR Spectrum Enhancement: A Deep Learning Approach. *Future Internet* 2021, 13, 265. <https://doi.org/10.3390/fi13100265>.
- Doan, V.S. Huynh, T. Kim, D. (2020). Underwater acoustic target classification based on dense convolutional neural network. *IEEE Geosci. Remote. Sens. Lett., early access*. <https://doi.org/10.1109/LGRS.2020.3029584>.
- Fernandes, J. D. C. V., de Moura Junior, de Seixas, N. N. (2022). Deep Learning Models for Passive Sonar Signal Classification of Military Data. *Remote Sens.* 2022, 14, 2648. <https://doi.org/10.3390/rs14112648>.
- Hong, F. Liu, C. Guo, L. (2021). Underwater Acoustic Target Recognition with ResNet18 on ShipsEar Dataset. in *Proc. 4th IEEE Int. Conf. Electronic. Tech. (ICET)*, Chengdu, China, pp. 1240–1244. <http://dx.doi.org/10.1109/ICET51757.2021.9451099>.
- Hu, G. Wang, K. Liu, L. (2021). Underwater Acoustic Target Recognition Based on Depthwise Separable Convolution Neural Networks. *Sensors* 2021, 21, 1429. <https://doi.org/10.3390/s21041429>.
- Irfan, M. Jiangbin, Z. Ali, S. Iqbal, M. Masood, Z. Hamid, U. (2021). DeepShip: an Underwater Acoustic Benchmark Dataset and a Separable Convolution Based Autoencoder for Classification. *EXPERT. SYST. APPL.* vol. 183, pp. 1-12. <https://doi.org/10.1016/j.eswa.2021.115270>.
- Jiang, Z. Zhao, C. Wang, H. (2022). Classification of Underwater Target Based on S-ResNet and Modified DCGAN Models. *Sensors* 2022, 22, 2293. <https://doi.org/10.3390/s22062293>.
- Jin, A. Zeng, X. (2023). A Novel Deep Learning Method for Underwater Target Recognition Based on Res-Dense Convolutional Neural Network with Attention Mechanism. *J. Mar. Sci. Eng.* 2023, 11, 69. <https://doi.org/10.3390/jmse11010069>.
- Jin, A. Zeng, X. (2023). A Novel Deep Learning Method for Underwater Target Recognition Based on Res-Dense Convolutional Neural Network with Attention Mechanism. *J. Mar. Sci. Eng.* 2023, 11, 69, pp. 1-20. <https://doi.org/10.3390/jmse11010069>.
- Jin, G. Liu, F. Wu, H. (2019). Deep learning-based framework for expansion, recognition and classification of underwater acoustic signal. *J. Exp. Theor. Artif. Intell.* 2020, 32, 205–218. <http://dx.doi.org/10.1080/0952813X.2019.1647560>.
- Ke , X. Yuan, F. Cheng, E. (2018). Underwater Acoustic Target Recognition Based on Supervised Feature-Separation Algorithm. *Sensors* 2018, 18, 4318. <https://doi.org/10.3390/s18124318>.



- Lei, Z. Lei, X. Wang, N. (2022). Present status and challenges of underwater acoustic target recognition technology: A review. *Front. Phys.* 2022, 10, 1018. <https://doi.org/10.3389/fphy.2022.1044890>.
- Li, L. Song, S. Feng, X. (2022). Combined LOFAR and DEMON Spectrums for Simultaneous Underwater Acoustic Object Counting and F0 Estimation. *J. Mar. Sci. Eng.* 2022, 10(10), 1565. <https://doi.org/10.3390/jmse10101565>.
- Luo, X. Chen, L. Zhou, Cao, H. H. (2023). A Survey of Underwater Acoustic Target Recognition Methods Based on Machine Learning. *J. Mar. Sci. Eng.* 2023, 11, 384. <https://doi.org/10.3390/jmse11020384>.
- Luo, X. Zhang, M. Liu, T. Huang, M. Xu, X. (2021). An Underwater Acoustic Target Recognition Method Based on Spectrograms with Different Resolutions. *J. Mar. Sci. Eng.* 2021, 9(11), 1246. <https://doi.org/10.3390/jmse9111246>.
- Ozydin, S. (2019). Examination of Energy Based Voice Activity Detection Algorithms for Noisy Speech Signals. *Euro. Journal. Sci. Tech. (EJOSAT)*, Special Issue, pp. 157-163.
- Saffari, A. Zahiri, S. H. Khozein Ghanad, N. (2023). Using SVM Classifier and Micro-Doppler Signature for Automatic Recognition of Sonar Targets. *Archives of Acoustics*, Vol. 48, No. 1, pp. 49–61. <https://doi.org/10.1016/j.dt.2022.05.007>.
- Santos-Domínguez, D. Torres-Guijarro, S. Cardenal-López, A. Pena-Gimenez, A. (2016). ShipsEar: an Underwater Vessel Noise Database. *APPL ACOUST.* 2016, 113, pp.64-69.
- Sheng, F. Zhu, X. (2023). A Transformer-Based Deep Learning Network form Underwater Acoustic Target Recognition. *IEEE Geoscience And Remote Sensing Letters*, Vol. 19, 2022. <https://doi.org/10.1109/LGRS.2022.3201396>.
- Smith, T. A. Rigby, J. (2022). Underwater radiated noise from marine vessels: A review of noise reduction methods and technology. *Ocean Engineering* 266 (2022) 112863. <https://doi.org/10.1016/j.oceaneng.2022.112863>.
- Tian, S. Chen, D. Wang, H. Liu, J. (2021). Deep Convolution Stack For Waveform In Underwater Acoustic Target Recognition. *Sci. Rep.* 2021, 11, 9614. <https://doi.org/10.1016/j.knosys.2022.110119>.
- Wang, X. Ying, T. Tian, W. (2020). Spectrum Representation Based on STFT. in *Proc. 13'rd IEEE Int. Conf. Image. Signal. Process. BioMedic. Eng. Inf. (CISP-BMEI)*, pp. 435–438. <http://dx.doi.org/10.1109/CISP-BMEI51763.2020.9263516>.
- Wang, Y. Wang, H. Li, Q. Xiao, Y. Ban, X. (2022). Passive Sonar Target Tracking Based on Deep Learning. *J. Mar. Sci. Eng.* 2022, 10, 181, <https://doi.org/10.3390/jmse10020181>.

- Xu, J. Huang, Z. Li, C. et al. (2019). Advances in Underwater Target Passive Recognition Using Deep Learning", Journal of Signal Processing. 35(9): 1460-1475. <https://doi.org/10.1155/2018/1214301>.
- Zhu, F. Liu, C. Yang, J. Wang, S. (2022). An Improved MobileNet Network with Wavelet Energy and Global Average Pooling for Rotating Machinery Fault Diagnosis. Sensors 2022, 22, 4427. <https://doi.org/10.3390/s22124427>.
- L. Tan, L. Jiang, J. (2019). Digital Signal Processing Fundamentals and Applications. in Academic Press, Cambridge, CA., 3rd ed, Massachusetts, United States, ch.7, pp. 248-268
- Yonina, E. C. (2018). Sampling Theory Beyond Bandlimited Systems. in Cambridge University Press, Cambridge, CA., 3rd ed, TJ International, United Kingdom, ch.9, pp. 328-331
- Kamal, S. Chandran, C. S. Supriya, M.H. (2021). Passive sonar automated target classifier for shallow waters using end-to-end learnable deep convolutional LSTMs"; Engineering Science and Technology., pp.1-12. <https://doi.org/10.1016/j.jestch.2021.01.014>.
- Joshi, D., Pareek, J., Ambatkar, P. (2023). Comparative Study of Mfcc and Mel Spectrogram for Raga Classification Using CNN. Indian Journal of Science and Technology 16(11). 816-822, <https://doi.org/10.17485/IJST/v16i11.1809>.
- Li, C., Guo, J., Sun, Z., Li, J., Wang, P. (2022). A Deep Attention Model for Environmental Sound Classification from Multi-Feature Data. Appl. Sci. 2022, 12, 5988, <https://doi.org/10.3390/app12125988>