



دانشکده مهندسی برق و رباتیک

پایان نامه کارشناسی ارشد مهندسی مخابرات سیستم

شناسایی، تشخیص و طبقه‌بندی اصوات ضربه‌ای از روی سیگنال صوتی آن‌ها به  
کمک شبکه عصبی

نگارنده: مرتضی سعادت‌ی

استاد راهنما

دکتر حسین مروی

تیر ۱۳۹۶

شماره: ۱۸۴۹۳  
تاریخ: ۹۷، ۴، ۲۷

باسمه تعالی



مدیریت تحصیلات تکمیلی

فرم شماره (۳) صورتجلسه نهایی دفاع از پایان نامه دوره کارشناسی ارشد

با نام و یاد خداوند متعال، ارزیابی جلسه دفاع از پایان نامه کارشناسی ارشد خانم / آقای مرتضی سعادت با شماره دانشجویی ۹۳۰۹۴۳۴ رشته مخابرات گرایش سیستم تحت عنوان: شناسایی، تشخیص و طبقه بندی اصوات ضربه ای از روی سیگنال صوتی آنها به کمک شبکه عصبی که در تاریخ ۱۳۹۶/۰۴/۲۷ با حضور هیأت محترم داوران در دانشگاه صنعتی شاهرود برگزار گردید به شرح ذیل اعلام می گردد:

قبول (با امتیاز ۱۶/۲ درجه خوب)	<input checked="" type="checkbox"/>	مردود	<input type="checkbox"/>
نوع تحقیق: نظری	<input checked="" type="checkbox"/>	عملی	<input type="checkbox"/>

عضو هیأت داوران	نام و نام خانوادگی	مرتبه علمی	امضاء
۱- استانداره‌نمای اول	عمین مراد	دانشیار	
۲- استانداره‌نمای دوم	—	—	—
۳- استاد مشاور	—	—	—
۴- نماینده تحصیلات تکمیلی	علی فصیح	استادیار	
۵- استاد ممتحن اول	علیرضا احمدزاد	دانشیار	
۶- استاد ممتحن دوم	هادی کرمانلو	استادیار	

نام و نام خانوادگی رئیس دانشکده:  
تاریخ و امضاء و مهر دانشکده:

تبصره: در صورتی که کسی مردود شود حداکثر یکبار دیگر (در مدت مجاز تحصیل) می تواند از پایان نامه خود دفاع نماید (دفاع مجدد نباید زودتر از ۴ ماه بعد از تاریخ ...)

خدای را بسی شاکرم که از روی کرم، پدر و مادری فداکار نسیم ساخته تا در سایه درخت پربار  
وجودشان بیایم و از ریشه آنها شاخ و برگ گیرم و از سایه وجودشان در راه کسب علم و دانش  
تلاش نمایم. والدینی که بودنشان تاج افتخاری است بر سرم و نامشان دلیلی است بر بودنم،  
چرا که این دو وجود، پس از پروردگار، مایه هستی ام بوده اند دستم را گرفتند و راه رفتن را در این  
وادی زندگی پر از فراز و نشیب آموختند. آموزگارانانی که برایم زندگی، بودن و انسان بودن را

معنا کردند.

تقدیم به وجود باارزشان...

## تشکر و قدردانی:

با سپاس فراوان از زحمات استاد محترم و گرانقدر جناب آقای دکتر مروی که از ابتدای راه و در طی انجام این پژوهش، با راهنمایی‌های خود مرا در نگارش این پایان‌نامه یاری نمودند.

همچنین از جناب آقای مهندس مهدی مقیمی و همچنین جناب آقای مهندس فرشاد رستمی به دلیل یاری‌ها و راهنمایی‌های بی چشم داشت ایشان که بسیاری از سختی‌ها را برایم آسانتر نمودند، بسیار سپاسگزارم.

## تعهد نامه

اینجانب مرتضی سعادت‌تی دانشجوی دوره کارشناسی ارشد رشته برق - مخابرات گرایش سیستم‌های دانشکده برق و رباتیک دانشگاه صنعتی شاهرود نویسنده پایان‌نامه شناسایی، تشخیص و طبقه‌بندی اصوات ضربه‌ای از روی سیگنال صوتی آن‌ها به کمک شبکه عصبی تحت راهنمایی دکتر حسین مروی متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام « دانشگاه صنعتی شاهرود » و یا « **Shahrood University of Technology** » به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا یافته‌های آن‌ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

## تاریخ

### امضای دانشجو

#### مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم‌افزارها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می‌باشد. این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

## چکیده:

یکی از مسائل مهم در زندگی روزمره تأمین امنیت جامعه است. در این راستا در سال‌های اخیر سیستم‌های شناسایی حرکات خطرناک و غیر معمول در حال گسترش و استفاده هستند. یکی از مسائل مرتبط با تشخیص حرکات غیرمعمول و خطرناک و پیشگیری از اتفاقات نامطلوب، تشخیص و تفکیک خودکار انواع صداهایی همچون انفجار، رعدوبرق، شلیک گلوله و... است. همچنین شناسایی نوع اسلحه شلیک‌کننده نیز برای تحقیقات بعدی برای شناسایی عوامل اغتشاش بسیار کاربردی خواهد بود.

در این راستا در این پایان‌نامه به تشخیص نوع اسلحه از روی سیگنال صوتی ناشی از شلیک آن خواهیم پرداخت. بدین منظور با استفاده از یک شبکه عصبی پرسپترون چندلایه و ویژگی‌هایی نظیر کپستروم حقیقی، پیشگویی ضرایب خطی و ضرایب کپسترال در مقیاس مل به شناسایی و تفکیک اسلحه‌ها خواهیم پرداخت. اسلحه‌های استفاده شده در این پایان‌نامه شامل ۶ نوع اسلحه سبک کمری می‌باشد که سیگنال‌های صوتی استفاده شده در آن مورد تایید صنعت می‌باشد.

نتایج حاصل از روش پیشنهادی نشان می‌دهد رویکرد استفاده از شبکه عصبی پرسپترون چندلایه و استخراج ویژگی‌های مناسب، رویکردی مناسب در تشخیص نوع اسلحه دارد.

**کلمات کلیدی:** تشخیص نوع اسلحه، اصوات ضربه‌ای، شبکه عصبی پرسپترون چندلایه،

کپستروم حقیقی، ضرایب پیشگویی خطی، ضرایب کپسترال در مقیاس مل

## فهرست

فصل اول.....	۱
مقدمه.....	۱
۱-۱ پیشگفتار.....	۲
۲-۱ ضرورت انجام تحقیق.....	۲
۳-۱ کاربرد شناسایی اصوات.....	۳
۴-۱ سیستم شناسایی و تشخیص اصوات ضربه‌ای.....	۴
۵-۱ مزایای سیستم تشخیص اصوات ضربه‌ای.....	۶
فصل دوم: مروری بر کارهای پیشین و مبانی نظری.....	۷
۱-۲ مقدمه.....	۸
۲-۲ تشخیص اصوات ضربه‌ای.....	۸
۳-۲ استخراج ویژگی.....	۹
۱-۳-۲ ضرایب کپسترال در مقیاس مل.....	۹
۲-۳-۲ ضرایب کپسترال در مقیاس خطی.....	۱۱
۳-۳-۲ ضرایب کپسترال در مقیاس گاماتن.....	۱۱
۴-۳-۲ ضرایب ادراکی پیشگویی خطی.....	۱۱
۵-۳-۲ ضرایب پیش‌بینی خطی.....	۱۲
۶-۳-۲ تبدیل موجک.....	۱۴
۷-۳-۲ ضرایب کپسترال حقیقی.....	۱۵
۸-۳-۲ ضرایب کپسترال فرکانس بارک.....	۱۶
۴-۲ طبقه بند.....	۱۷
۱-۴-۲ آموزش با نظارت.....	۱۷
۱-۴-۲-۱ ماشین بردار پشتیبان.....	۱۷
۲-۴-۲-۱ درخت تصمیم‌گیری.....	۱۸
۲-۴-۲ آموزش بدون نظارت.....	۱۸
۲-۴-۲-۱ شبکه‌های کوهون.....	۱۸
۲-۴-۲-۲ شبکه‌ها پفیلد.....	۱۹
۳-۴-۲-۲ شبکه عصبی پرسپترون.....	۱۹
۱-۳-۲-۴-۲ شبکه عصبی پرسپترون تک لایه.....	۲۰
۲-۳-۲-۴-۲ شبکه عصبی پرسپترون چندلایه (MLP).....	۲۱



۲۳	۵-۲ مروری بر کارهای پیشین .....
۲۹	فصل سوم: روش پیشنهادی.....
۳۰	۱-۳ مقدمه .....
۳۱	۲-۳ پنجره گذاری.....
۳۲	۳-۳ ویژگیهای مورد استفاده .....
۳۲	۱-۳-۳ ضرایب کپسترال در مقیاس مل.....
۳۳	۲-۳-۳ ضرایب کپسترال حقیقی .....
۳۴	۳-۳-۳ ضرایب پیش‌بینی خطی.....
۳۴	۴-۳ طبقه‌بند مورد استفاده.....
۳۶	۵-۳ جمع‌بندی .....
۳۷	فصل چهارم: نتایج روش پیشنهادی.....
۳۸	۱-۴ مقدمه.....
۳۸	۲-۴ پایگاه داده.....
۳۸	۱-۲-۴ رولور رو جر ۳۵۷ مگنوم.....
۴۰	۲-۲-۴ کلت ام ۱۹۱۱.....
۴۱	۳-۲-۴ رولور مانورهمین.....
۴۴	۴-۲-۴ کلت برتا ام ۹.....
۴۶	۵-۲-۴ کلت دزرت ایگل .....
۴۸	۶-۲-۴ رولور.....
۵۰	۳-۴ استخراج ویژگی از پایگاه داده.....
۵۱	۱-۳-۴ ضرایب کپسترال در مقیاس مل.....
۵۲	۲-۳-۴ ضرایب کپسترال حقیقی.....
۵۵	۳-۳-۴ ضرایب پیش‌بینی خطی.....
۵۶	۱-۳-۳-۴ ضرایب پیش‌بینی خطی.....
۵۷	۲-۳-۳-۴ ضرایب پیش‌بینی خطی.....
۶۰	۳-۳-۳-۴ ضرایب پیش‌بینی خطی.....
۶۳	۴-۴ تحلیل و نتیجه‌گیری.....
۶۸	۵-۴ جمع‌بندی .....
۶۹	فصل پنجم: جمع‌بندی.....
۷۴	منابع و مراجع .....

## فهرست نمودارها

- نمودار ۱-۴ مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش MFCC..... ۵۲
- نمودار ۲-۴ مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش Real Cepstrum..... ۵۴
- نمودار ۳-۴ مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش LPC 10..... ۵۷
- نمودار ۴-۴ مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش LPC 20..... ۵۹
- نمودار ۵-۴ مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش LPC50..... ۶۱
- نمودار ۶-۴ مقایسه دقت روش‌های پیشنهادی برای شناسایی اسلحه ۱..... ۶۵
- نمودار ۷-۴ مقایسه دقت روش‌های پیشنهادی برای شناسایی اسلحه ۲..... ۶۵
- نمودار ۸-۴ مقایسه دقت روش‌های پیشنهادی برای شناسایی اسلحه ۳..... ۶۶
- نمودار ۹-۴ مقایسه دقت روش‌های پیشنهادی برای شناسایی اسلحه ۴..... ۶۶
- نمودار ۱۰-۴ مقایسه دقت روش‌های پیشنهادی برای شناسایی اسلحه ۵..... ۶۷
- نمودار ۱۱-۴ مقایسه دقت روش‌های پیشنهادی برای شناسایی اسلحه ۶..... ۶۷

## فهرست شکل‌ها

- شکل ۱-۱ مقایسه سیگنال اصوات ضربه‌ای با سیگنال غیر ضربه‌ای ..... ۵
- شکل ۱-۲ ساختار اصلی سیستم‌های تشخیص اصوات ..... ۸
- شکل ۲-۲ شبکه عصبی پرسپترون ..... ۲۰
- شکل ۳-۲ توابع خطی جدانشدنی و جدانشدنی ..... ۲۱
- شکل ۴-۲ شبکه عصبی پرسپترون چندلایه ..... ۲۲
- شکل ۱-۳ بلوک دیاگرام شناسایی، تشخیص و طبقه‌بندی اصوات ضربه‌ای ..... ۳۰
- شکل ۲-۳ پنجره گذاری سیگنال صوتی ..... ۳۲
- شکل ۳-۳ بلوک دیاگرام ضرایب کیسترال در مقیاس مل ..... ۳۳
- شکل ۴-۳ بلوک دیاگرام ضرایب کیسترال حقیقی ..... ۳۳
- شکل ۵-۳ شبکه عصبی مورد استفاده برای طبقه‌بندی ..... ۳۵
- شکل ۱-۴ روجر ۳۵۷ مگنوم ..... ۳۹
- شکل ۲-۴ سیگنال صوتی شلیک اسلحه ۱ در حوزه زمان ..... ۳۹
- شکل ۳-۴ سیگنال صوتی شلیک اسلحه ۱ در حوزه فرکانس ..... ۴۰
- شکل ۴-۴ کلت ..... ۴۱
- شکل ۵-۴ سیگنال صوتی شلیک کلت در حوزه زمان ..... ۴۱
- شکل ۶-۴ سیگنال صوتی شلیک کلت در حوزه فرکانس ..... ۴۲
- شکل ۷-۴ رولور مانورهمین ..... ۴۳
- شکل ۸-۴ سیگنال صوتی رولور مانورهمین در حوزه زمان ..... ۴۳
- شکل ۹-۴ سیگنال صوتی رولور مانورهمین در حوزه فرکانس ..... ۴۴
- شکل ۱۰-۴ کلت برتا ام ۹ ..... ۴۴
- شکل ۱۱-۴ سیگنال صوتی کلت برتا ام ۹ در حوزه زمان ..... ۴۵
- شکل ۱۲-۴ سیگنال صوتی کلت برتا ام ۹ در حوزه فرکانس ..... ۴۶
- شکل ۱۳-۴ کلت دزرت ایگل ..... ۴۷
- شکل ۱۴-۴ سیگنال صوتی کلت دزرت ایگل در حوزه زمان ..... ۴۸
- شکل ۱۵-۴ سیگنال صوتی کلت دزرت ایگل در حوزه فرکانس ..... ۴۸
- شکل ۱۶-۴ رولور ..... ۴۹
- شکل ۱۷-۴ سیگنال صوتی رولور در حوزه زمان ..... ۵۰

## فهرست جدول‌ها

جدول ۴-۱: نتایج حاصل از ارزیابی طبقه با استفاده از ویژگی MFCC	۵۲
جدول ۴-۲: نتایج حاصل از ارزیابی طبقه با استفاده از ویژگی Real Cepstrum	۵۴
جدول ۴-۳: نتایج حاصل از ارزیابی طبقه با استفاده از ویژگی LPC10	۵۶
جدول ۴-۴: نتایج حاصل از ارزیابی طبقه با استفاده از ویژگی LPC20	۵۸
جدول ۴-۵: نتایج حاصل از ارزیابی طبقه با استفاده از ویژگی LPC50	۶۱
جدول ۴-۶: مقایسه دقت روشهای پیشنهادی	۶۳

# فصل اول:

## مقدمہ

## ۱-۱ پیشگفتار:

تشخیص، شناسایی و طبقه‌بندی اصوات ناگهانی و ضربه‌ای یکی از مسائل مهم و حائز اهمیت می‌باشد. تشخیص و تفکیک خودکار انواع صداهایی همچون انفجار، رعدوبرق، شلیک گلوله و... از یکدیگر توسط یک سیستم یکی از نیازهای جامعه‌ی امروزی می‌باشد. بی‌شک تاکنون شاهد این مسئله بوده‌ایم که در عملیات پلیس با مجرمین مسلح و یا تروریست‌ها با نیروهای نظامی تیراندازی اتفاق می‌افتد؛ امروزه آگاهی از این اتفاقات اهمیت ویژه‌ای برای پلیس، سربازها و امنیت خصوصی و عمومی دارد. در این راستا نکته حائز اهمیت این است که سیستمی طراحی شود تا نوع اسلحه‌ی شلیک‌شده را شناسایی کند و این اطلاعات را در اختیار پلیس قرار دهد، این کار با استفاده از ویژگی‌های آکوستیکی عمل شلیک گلوله انجام می‌شود. با توجه به ۲ دیدگاه امنیت خصوصی که شامل حفاظت اشخاص و پرسنل مهم است و نیز دیدگاه توسعه‌ی امنیت عمومی، تشخیص رخداد شلیک گلوله در فضای شهری و پیش‌بینی و تخمین نوع اسلحه کمک بسیار ارزشمندی به رسیدگی وضعیت موجود و مدیریت آن می‌کند.

در این فصل ابتدا ضرورت انجام این تحقیق را بیان کرده و سپس به کاربرد شناسایی اصوات می‌پردازیم. آنگاه به بیان سیستم شناسایی و تشخیص اصوات و نحوه‌ی عملکرد آن را شرح می‌دهیم. نتیجه‌گیری پایان‌بخش این فصل خواهد بود.

## ۱-۲ ضرورت انجام تحقیق:

کاربرد نظامی سیستم‌های شناسایی نوع اسلحه یکی از رایج‌ترین و مهم‌ترین روش‌ها برای شناسایی توان نیروهای نظامی‌باشد، چراکه نیروهای نظامی خودی در شرایط عملیات زمینی همواره از سوی تیراندازهای کمین‌گرفته‌ی دشمن که ممکن است از هر مکانی به سمت آن‌ها شلیک کنند، درخطر هستند؛ به همین دلیل شناسایی نوع اسلحه‌ی شلیک‌کننده‌ی گلوله بلافاصله بعد از شلیک اولین گلوله، به نیروها کمک می‌کند که در سریع‌ترین زمان ممکن خود را برای محافظت و مقابله با

تیراندازهای مخفی شده مهیا کنند. به عبارت دیگر این سیستم به سربازها می‌گوید که دشمن با استفاده از چه اسلحه‌ای به سوی آن‌ها شلیک می‌کند.

هدف از انجام این پایان‌نامه شناسایی، تشخیص و طبقه‌بندی اصوات ضربه‌ای با تأکید بر شناسایی نوع اسلحه‌ی سبک با استفاده از صدای شلیک گلوله می‌باشد.

اساس و عملکرد سیستم بدین گونه است که شلیک گلوله شامل دو نوع رخداد آکوستیکی است. اول صدای انفجار دهانه است که ناشی از عمل کردن چاشنی گلوله است. رخداد دوم برای گلوله‌های مافوق صوت اتفاق می‌افتد که صدای موج تکانه‌ای است. با دریافت صدای یکی یا هر دو این دو رخداد آکوستیکی که توسط آرایه‌ای از سنسورها، می‌توان تخمینی از شناسایی نوع اسلحه‌ی شلیک‌شده بدست آورد.

### ۱-۳ کاربرد شناسایی اصوات:

سیستم شناسایی اصوات محیط دارای دامنه گسترده‌ای می‌باشد که با توجه به زمینه‌های مختلف کاربردهای متفاوتی دارد.

از کاربردهای سیستم شناسایی می‌توان به تشخیص گوینده [۱]، [۲] تبدیل گفتار به نوشتار [۳]، [۴] و تشخیص گفتار [۵]، [۶] اشاره نموده و همچنین دیگر کاربردهای آن نظیر تعاملات ربات و انسان [۷]، [۸] دسته‌بندی ادوات موسیقی [۹]، تخمین و شناسایی محل ایجاد صدا [۱۰]، [۱۱] تشخیص و شناسایی نوع خودرو و هواپیما [۱۲]، [۱۳]، تشخیص گونه‌های مختلف پرندگان و حیوانات اشاره کرد [۱۴]، [۱۵].

امروزه این سیستم در هوشمند سازی خانه‌ها هم نقش مؤثری ایفا می‌کند و مورد توجه پژوهشگران این حوزه می‌باشد [۱۶].

## ۴-۱ سیستم شناسایی و تشخیص اصوات ضربه‌ای

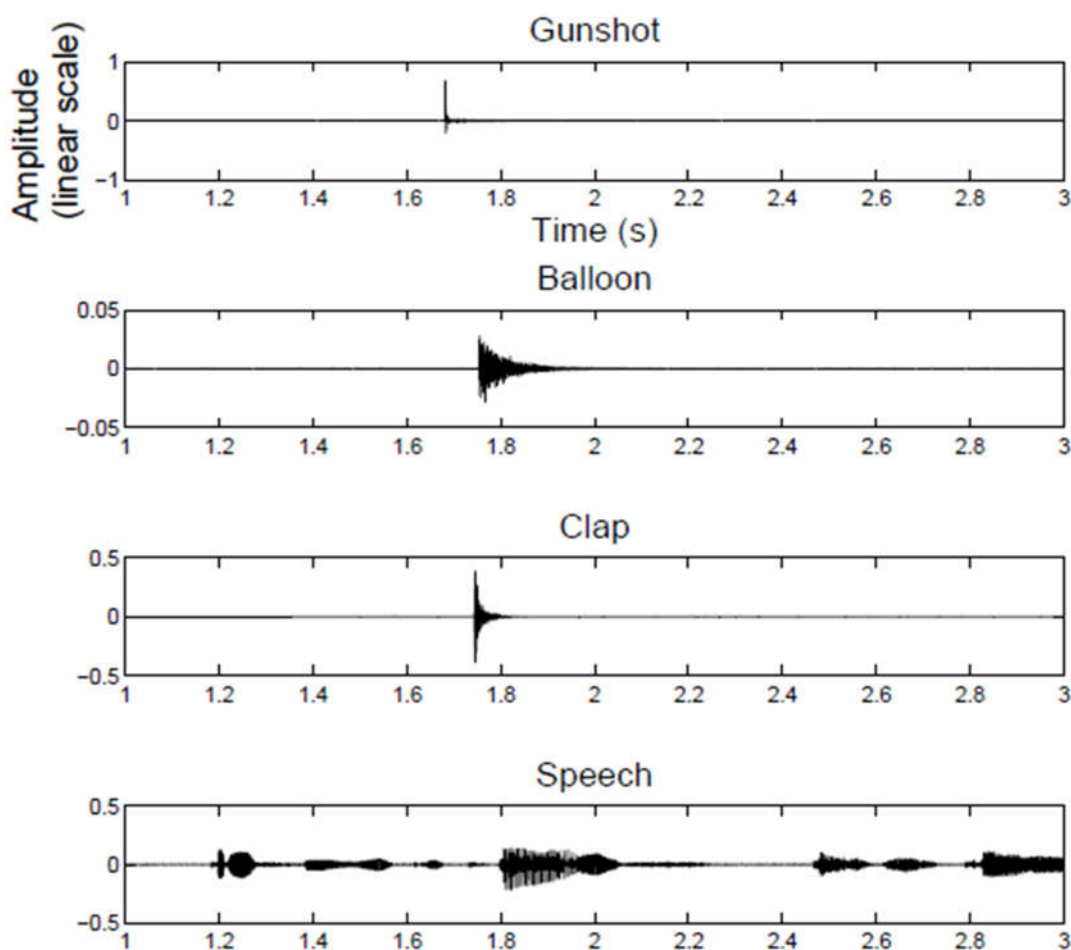
به‌طور کلی صداهای ضربه‌ای به دلیل متراکم شدن ناگهانی هوا ایجاد شده و همزمان سرعت انرژی سیگنال آن افزایش می‌یابد. زمان به وجود آمدن این نوع از اصوات کوتاه بوده و در کل نشانگر این نکته می‌باشد که رویدادی در پیرامون رخ داده است. از این ویژگی افزایش ناگهانی و سریع می‌توان برای شناسایی صدای ضربه‌ای استفاده کرد. اصواتی نظیر فریاد زدن، شلیک کردن، انفجار، دست زدن، رعدوبرق، بسته شدن درب و شکستن شیشه در دسته‌ی صداهای ضربه‌ای قرار می‌گیرد.

در این پایان‌نامه پس از تشخیص ضربه‌ای بودن صدای ورودی به شناسایی نوع اسلحه می‌پردازیم. همان‌طور که در قبل اشاره شد صدای شلیک گلوله در دسته‌ی اصوات ضربه‌ای قرار دارد و در همین راستا در ادامه به تشریح جزئیاتی در این‌باره پرداخته و در مورد نوع صدا و چگونگی کلاس‌بندی این اصوات توضیح خواهیم داد.

در شکل (۱-۱) نمونه‌ای از سیگنال‌های به‌کاررفته در هر دسته از سیگنال‌های ضربه‌ای و غیر ضربه‌ای نشان داده شده است. با مقایسه سیگنال‌ها می‌توان دریافت که سیگنال‌های ضربه‌ای دسته‌های مختلف علاوه بر اینکه از نظر ساختار و فرکانس موجود در آن‌ها با سیگنال‌های غیر ضربه‌ای مثل گفتار متفاوت هستند بلکه با یکدیگر نیز متفاوت‌اند.

با توجه به مشاهدات در شکل (۱-۱) دریافتیم که از نظر ظاهری و فرکانس‌های موجود سیگنال‌های ضربه‌ای متفاوتند و حال اینکه این تفاوت نیز در سیگنال‌های ضربه‌ای نیز وجود دارد. به‌طور مثال در مقایسه اصوات ضربه‌ای و غیر ضربه‌ای همانند شلیک گلوله، انفجار بادکنک و کف زدن که سیگنال ضربه‌ای بوده و گفتار که یک سیگنال غیر ضربه‌ای می‌باشد این تفاوت به‌راحتی قابل تشخیص می‌باشد.





شکل ۱-۱ مقایسه سیگنال اصوات ضربه‌ای با سیگنال غیر ضربه‌ای [۱۰]

بنابراین بسته به اینکه سیگنال مربوط به کدام دسته از اصوات ضربه‌ای می‌باشد طول مدتی که انرژی سیگنال افزایش یافته متغیر می‌باشد. به‌طور مثال در شلیک گلوله مدت‌زمان افزایش انرژی (رخداد ضربه) بسیار کوتاه بوده ولی انرژی اولیه‌ی آن از بقیه بیشتر می‌باشد درحالی‌که دودسته دیگر در مدت‌زمان بیشتری به وقوع می‌پیوندند. شایان‌ذکر است که این تغییرات در فرکانس‌های سیگنال موجود می‌باشد و تفاوت‌هایی به چشم می‌خورد. برای مثال محدوده‌ی فرکانس در دسته انفجار بادکنک شبیه به دسته‌ی انفجار بود؛ اما مقدار بیشینه انرژی ما کمتر از دسته انفجار می‌باشد که ما در دودسته دیگر محدوده‌ی فرکانس بسیار گسترده از سایر دسته‌ها داریم. همان‌طور که بیان شد، سیگنال‌های دسته‌های مختلف دارای ساختار و رفتار متفاوت در حوزه‌ی زمان و فرکانس می‌باشد.

## ۱-۵ مزایای سیستم تشخیص اصوات ضربه‌ای

امروزه این سیستم در حوزه‌های مختلف نقش مؤثری ایفا می‌کند چراکه در حال حاضر پژوهشگران در حوزه‌ی پردازش صوت فعالیت بسیاری برای صنعتی کردن این سیستم کرده و علاوه بر پردازش گفتار به موضوعاتی نظیر عیب‌یابی قطعات با توجه به نوع صدای آن، شناسایی هواپیما و خودرو و... پرداخته‌اند که در زیر به توصیف چند نمونه از آن‌ها می‌پردازیم.

یکی از پرکاربردترین مزایای استفاده از این سیستم به‌عنوان سیستم نظارتی و امنیتی در مکان‌هایی نظیر مغازه‌ها، خانه‌ها، بانک‌ها و پارکینگ‌های عمومی می‌باشد که از این سیستم می‌توان برای تشخیص نفوذ استفاده کرد. در این حالت می‌توان یک فعالیت غیرعادی را تشخیص داد و شناسایی اصواتی مانند شلیک گلوله، فریاد زدن، انفجار و شکسته شدن شیشه نیز در راستای همین امر می‌باشد [۱۷] ، [۱۸]

همچنین می‌توان در تعامل انسان ربات و نیز جهت‌یابی در هدایت ربات از آن بدین گونه استفاده نمود که ربات صداهای محیط را تشخیص می‌دهد و بر مبنای آن جهت حرکت خود را تعیین نماید و نیز در صورت بروز مشکل از آن محیط دور شود [۷].

## ۱-۶ ساختار پایان‌نامه:

این پایان‌نامه شامل پنج فصل می‌باشد که در فصل اول تعریف سیستم شناسایی و تشخیص اصوات ضربه‌ای و نحوه عملکرد آن بررسی شد. در فصل دوم به تشریح پژوهش و تحقیقاتی که در این باره انجام شده و مبانی نظری استفاده‌شده در روش پیشنهادی می‌پردازیم.

در فصل سوم به معرفی سیستم پیشنهادی پرداخته‌ایم.

در فصل چهارم نتایج روش پیشنهادی و تحلیل نتایج موردبررسی قرار گرفته‌است.

در فصل پنجم که پایان‌بخش این پایان‌نامه است، نتیجه‌گیری و جمع‌بندی کلی بیان شده است.

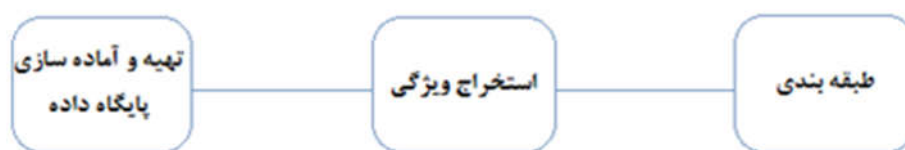
فصل دوم:  
مروری بر کارهای پیشین و  
مبانی نظری

## ۲-۱ مقدمه:

در این فصل با توجه به شناختی که از سیستم شناسایی اصوات ، کاربردها و همچنین الزام وجود چنین مکانیزمی برای شناسایی آن داریم؛ به شرح مختصری از تحقیقات انجام شده در ارتباط با این مهم می پردازیم و طبق روال هر فصل نتیجه گیری پایان بخش این فصل می باشد.

## ۲-۲ تشخیص اصوات ضربه ای:

در راستای تشخیص صدای گلوله تحقیقات مختلف و زیادی در سالهای اخیر انجام شده و سیستم هایی پیشنهاد شده است که در تمامی این سیستم های پیشنهادی سه بخش تهیه و آماده سازی پایگاه داده، استخراج ویژگی و طبقه بندی نقش اصلی و سازنده را دارند.



شکل ۲-۱ ساختار اصلی سیستم های تشخیص اصوات

در راستای تشخیص صدا این نکته حائز اهمیت است که پایگاه استاندارد وجود ندارد و در هر پژوهش محقق با توجه به کاربرد آن، پایگاه مورد نظر را تهیه و یا می سازد. لازم به ذکر است تعداد طبقه های موجود در هر پایگاه داده و نمونه های هر یک از این طبقه ها با توجه به بازخورد پژوهش متفاوت می باشد. در ارتباط با بخش استخراج ویژگی تحقیقات گسترده ای انجام شده که در هر پژوهش محقق با توجه به پایگاه داده تهیه شده و کاربردهای آن ویژگی جدید را پیشنهاد می دهد. به طور کلی رویداد جدید این پژوهش ها در نشان دادن این ویژگی نوین می باشد که یا راندمان تشخیص صحیح سیستم را افزایش داده و یا باعث کاهش ابعاد بردار ویژگی شده است. در بخش طبقه بندی با توجه به نوع ویژگی های استخراج شده از طبقه بندی های متفاوتی استفاده شده است [۱۹]. (غالب طبقه بندی های

استفاده شده در زمینه‌های تشخیص صداهای محیط شامل مدل مخفی مارکوف<sup>۱</sup>، مدل مخلوط گوسی<sup>۲</sup> و ماشین پشتیبان<sup>۳</sup> می‌باشد.

## ۲-۳ استخراج ویژگی:

در سیستم شناسایی اصوات بخش استخراج ویژگی نقش بسیار سازنده‌ای و تعیین‌کننده‌ای دارد. در تجزیه و تحلیل سیگنال صدا ویژگی‌های مختلفی استفاده می‌شود که انتخاب ویژگی موردنظر بسته به کاربرد آن صورت می‌گیرد، چراکه شرایط مناسب کاربرد هر یک با دیگری متفاوت است. برای مثال، ممکن است برخی ویژگی‌ها در محیط‌های نویزی جواب بهتری دهند و یا اینکه حجم محاسبات کمتری را طلب کنند. این ویژگی‌ها گاه در کل سیگنال صدا و گاه در یک پنجره از آن محاسبه می‌شوند، ولی در تحلیل سیگنال صدا به‌طور معمول ویژگی‌های سیگنال پنجره‌بندی شده تحلیل می‌شود. این ویژگی‌ها گاه یک عددند، نظیر انرژی یا میانگین پنجره سیگنال صدا و گاه یک بردارند. برای مثال، خود اندازه سیگنال صدا در حوزه زمان را می‌توان به‌صورت برداری یک بعدی با طول تعداد نقاط پنجره صدا در نظر گرفت. از این‌رو، گاه از این ویژگی‌ها بانام ضرایب ویژگی یا بردارهای ویژگی نیز یاد می‌شود. روش‌های متعددی برای استخراج ویژگی ارائه شده است از جمله مهم‌ترین این روش‌ها، می‌توان به الگوریتم‌های ضرایب کپسترال در مقیاس مل (MFCC)<sup>۴</sup> و ضرایب ادراکی پیش بینی خطی (PLP)<sup>۵</sup> و ضرایب پیش‌بینی خطی (LPC)<sup>۶</sup> اشاره کرد.

## ۲-۳-۱ ضرایب کپسترال در مقیاس مل:

شاخص‌ترین روش که در بسیاری از مقالات به‌عنوان روش استاندارد استخراج ویژگی شناخته می‌شود روش ضرایب کپسترال در مقیاس مل (MFCC) می‌باشد. این ویژگی یکی از عمومی‌ترین

<sup>1</sup> Hidden Markov Model (HMM)

<sup>2</sup> Gaussian Mixture Model (GMM)

<sup>3</sup> Support Vector Machine (SVM)

<sup>4</sup> Mel Frequency Cepstral Coefficients (MFCC)

<sup>5</sup> Perceptual Linear Prediction (PLP)

<sup>6</sup> Linear Prediction Cepstral (LPC)

تکنیک‌های استخراج ویژگی در زمینه تشخیص صوت است که به‌عنوان یک ویژگی مبتنی بر فرکانس مورد توجه بوده و بسیار دقیق‌تر از ویژگی حوزه‌ی زمان می‌باشد. در روش MFCC از تبدیل فوریه گسسته استفاده می‌شود و فیلتر بانک مورد استفاده، فیلتر بانک مل می‌باشد. مراحل استخراج ویژگی ضرایب کپسترال فرکانس مل بدین صورت می‌باشد:

۱- ابتدا با استفاده از یک فیلتر، سیگنال پیش تأکید می‌شود. تابع انتقال این فیلتر طبق رابطه (۲) است.

$$H(z) = 1 - az^{-1} \quad (۱-۲)$$

۲- سیگنال پیش تأکید شده، به یک سری فریم‌های زمانی که تعداد دلخواهی از نمونه‌ها می‌باشد تقسیم‌بندی می‌شود. در بیشتر سیستم‌ها برای انتقال دادن از یک فریم به فریم دیگر از همپوشانی استفاده می‌شود. سپس هر فریم زمانی برای حذف ناپیوستگی در لبه‌ها از یک پنجره همینگ عبور داده می‌شود.

۳- پس از فریم بندی و پنجره کردن، مجذور اندازه تبدیل فوریه هر فریم محاسبه می‌شود تا مؤلفه‌های فرکانسی سیگنال حوزه‌ی زمان استخراج شوند. برای سرعت بخشیدن به امر پردازش، الگوریتم تبدیل فوریه سریع بکار می‌رود.

۴- سپس فریم‌های تبدیل‌شده به حوزه فوریه از مل فیلترها عبور داده می‌شوند. فیلترهای مل فیلترهای مثلثی شکلی هستند که بافاصله‌های مساوی در مقیاس مل از هم قرار دارند.

۵- مرحله بعد از خروجی مل فیلترها لگاریتم گرفته می‌شود

۶- در مرحله آخر از تبدیل کسینوسی استفاده می‌شود [۲۰].

## ۲-۳-۲ ضرایب کپسترال در مقیاس خطی:

روش ضرایب کپسترال در مقیاس خطی<sup>۱</sup> همانند روش MFCC می‌باشد با این تفاوت که فیلتر بانکی که استفاده می‌شود دیگر بر مبنای مقیاس مل نمی‌باشد بلکه در مقیاس خطی تقسیم‌بندی شده و طول تمام فیلترهای مثلثی با یکدیگر یکسان می‌باشد [۲۱].

## ۲-۳-۳ ضرایب کپسترال در مقیاس گاماتن:

روش ضرایب کپسترال در مقیاس گاماتن<sup>۲</sup> ترکیبی از روش‌های MFCC و PLP می‌باشد. به این ترتیب که برای قسمتی که بجای فیلتر بانک در مقیاس مل، از فیلتر بانک در مقیاس گاماتن استفاده می‌کنند. بعد از اعمال فیلتر بانک، مانند روش PLP از تکنیک بلندی هم‌تراز استفاده می‌کنند.

فیلتر گاماتن عبارت است از یک سری فیلتر میان‌گذر که سیستم شنیداری انسان را مدل می‌کنند. تابع ضربه هرکدام از این فیلترها به صورت معادله (۲-۲) می‌باشد. در این معادله  $a$  یک ثابت است که معمولاً برابر با ۱ در نظر گرفته می‌شود.  $\varphi$  عبارت است از جابه‌جایی فاز  $n$  مرتبه فیلتر می‌باشد. در این معادله  $f_c$  فرکانس مرکزی در حوزه هرتز می‌باشد [۲۲].

$$g(t) = a^{n-1} e^{-2\pi t} \cos(2\pi f_c t + \varphi) \quad (2-2)$$

## ۲-۳-۴ ضرایب ادراکی پیش بینی خطی:

الگوریتم ادراکی پیش‌گویی خطی (PLP) به همراه ضرایب کپسترال فرکانس مل (MFCC) از پرکاربردترین الگوریتم‌های استخراج ویژگی می‌باشد و اینکه کدام یک دارای عملکرد بهتری می‌باشد بستگی به نوع کاربرد دارد. در مواردی که عدم تطابق بین داده، آزمایش و آموزش وجود دارد روش ادراکی پیش‌گویی خطی بهتر عمل می‌کند به عبارت دیگر در مقابل نویز مقاوم‌تر است.

<sup>1</sup> Linear Frequency Cepstral Coefficients

<sup>2</sup> Gammatone Frequency Cepstral Coefficients

مراحل بدست آوردن ضرایب ادراکی پیشگویی خطی بدین ترتیب می‌باشد.

۱- فریم بندی و پنجره کردن؛ سیگنال به یک سری فریم‌های زمانی که تعداد دلخواهی از نمونه‌ها می‌باشد تقسیم‌بندی می‌شود.

۲- مجذور اندازه تبدیل فوریه با استفاده از الگوریتم تبدیل فوریه سری انجام می‌پذیرد.

۳- در این مرحله فرکانس در مقیاس هرتز طبق رابطه (۳-۲) به فرکانس در مقیاس باراک تبدیل می‌شود. در واقع سیگنال صوت از سری فیلترهای دوزنقه‌ای شکل یا فاصله‌های مساوی در مقیاس بارک عبور می‌کند.

$$f_{bark} = 6 \ln \left[ \frac{f}{600} + \sqrt{\left( \frac{f}{600} \right)^2 + 1} \right] \quad (۳-۲)$$

۴- در این مرحله، اسپکتروم توسط یک مدل تمام قطب رابطه (۴-۲) و با استفاده از روش خودهمبستگی تقریب زده می‌شود. قبل از آن باید تبدیل فوریه معکوس روی نمونه‌های اسپکتروم اعمال شود. برای مدل‌های تمام قطب مرتبه M فقط (M+1) مقادیر اولیه خودهمبستگی لازم است.

$$H(Z) = \frac{1}{1 - \sum_{i=1}^m a_i z^{-i}} \quad (۴-۲)$$

۵- آخرین مرحله استخراج ضرایب پیش‌بینی ادراکی خطی، مرحله تحلیل کپستروم است. در این مرحله ضرایب  $a_i$  به دست آمده از مدل تمام قطب توسط رابطه بازگشتی به ضرایب کپسترال تبدیل می‌شوند [۲۳].

## ۲-۳-۵ ضرایب پیش‌بینی خطی:

پیش‌بینی خطی یکی از روش‌های قدرتمند برای محاسبه و تخمین پارامترهای سیگنال گفتار، مانند فرکانس‌های پیچ، فرمنت و توابع مقطع کانال صوتی می‌باشد. اهمیت این روش در دقت بالای



تخمین و سرعت بالای محاسبات می‌باشد. ایده اصلی در این روش این است که یک نمونه از سیگنال گفتار را می‌توان از نمونه‌های قبلی آن بدست آورد. ضریب پیش‌بینی خطی روشی است متداول در حوزه پردازش سیگنال و پردازش گفتار که به منظور فشرده‌سازی با کد کردن گفتار با کیفیت بالا در نرخ بیت پایین استفاده می‌شود. ضریب پیش‌بینی خطی روشی کارآمد بوده که تخمین‌های بسیار دقیقی از پارامترهای گفتار را فراهم می‌سازد [۲۴].

این مدل از دو بخش تشکیل شده است که یک بخش فقط شامل صفر و دیگری تمام قطب است. در عمل چون همه واک‌ها تنها دارای قطب‌اند؛ تنها از بخش تمام قطب استفاده می‌شود که با فرمول

(۵-۲) بیان شده است. روش LPC یک روش دقیق، مقاوم و قابل اعتماد برای تخمین پارامترهای یک سیستم متغیر با زمان خطی می‌باشد.

$$H(Z) = \frac{y(Z)}{x(Z)} = \frac{1}{1 - \sum_{j=1}^p a_j Z^{-j}} \quad (۵-۲)$$

ضرایب  $a_j z$  را ضرایب پیش‌بینی خطی یا ضرایب LPC می‌نامند. در این رابطه مرتبه پیشگویی برابر  $p$  است. دلیل استفاده از این نام در این فیلتر این است که اگر بخواهیم نمونه زمانی لحظه  $n$  سیگنال  $y$  یعنی  $y(n)$  را بدست آوریم، آن را با استفاده از ترکیب خطی  $p$  نمونه قبلی  $y$  پیش‌بینی می‌کنیم و آن را مطابق (۶-۲)، با  $y(n)$  نشان می‌دهیم.

$$y(n) = \sum_{i=1}^p a_i y(n-i) \quad (۶-۲)$$

اگر مطابق رابطه (۷-۲) به سیگنال گفتار یک پنجره اعمال کنیم داریم:

$$y(n) = w(n)s(n) \quad (۷-۲)$$

که تفاوت بین نمونه‌های تخمین در نمونه‌های واقعی را خطای پیشگویی می‌گویند و رابطه (۸-۲) بیان شده است.

$$e(n) = y(n) - \sum_{i=1}^p a_i y(n-i) \quad (۸-۲)$$

با داشتن خطا، میانگین مجذور خطای پیشگویی می‌تواند به صورت (۹-۲) تعریف شود.

$$E = \sum_{n=0}^{N-1} e^2(n) = \sum_{n=0}^{N-1} \left( y(n) - \sum_{j=1}^p a_j y(n-j) \right)^2 \quad (۹-۲)$$

حال ضرایب را نیز می‌توان با مینیمم کردن میانگین مجذور خطا به دست آورد. که در این صورت می‌توان از آن مشتق گرفت. با این عمل برای هر ضریب  $a_i$  به تعداد  $p$  معادله خواهیم داشت که در رابطه (۱۰-۲) نشان داده شده است.

$$\sum_{j=1}^p a_j \phi(i,j) = \phi(i,0) \text{ for } i \in \{1,2,3, \dots, P\} \quad (۱۰-۲)$$

از دیگر کاربردهای مطرح LPC استفاده در کد کردن سیگنال است

## ۲-۳-۶ تبدیل موجک:

تبدیل موجک<sup>۱</sup> یک تبدیل انتگرالی است که از توابع موجک پایه استفاده می‌کند. از موجک به‌عنوان ابزار مناسبی در ریاضیات، فیزیک در پردازش سیگنال، محاسبات عددی و فشرده‌سازی اطلاعات استفاده می‌کنند. در تبدیل موجک تابع را می‌توان با استفاده از ترکیب نسخه‌های اصلاح‌شده توابع هسته نمایش داد. توابع هسته را مادر و نسخه‌های اصلاح‌شده آن را نیز توابع فرزند می‌نامند. یک تابع رازمانی تابع موجک مادر می‌نامیم که بتواند در شرط پذیرفتگی صدق کند. یک تابع مادر می‌بایست نوسانی بوده و همچنین دامنه آن نیز به‌سرعت به سمت صفر میل کند. اگر این دو شرط را همراه با شرط صفر بودن مؤلفه جریان مستقیم یا مؤلفه فرکانس صفر را برای یک تابع در نظر بگیریم، گفته می‌شود شرط پذیرفتگی در تابع مربوطه برقرار می‌باشد. به عبارت دیگر برای یک تابع  $h \in L(R)$  شرط پذیرفتگی برقرار می‌باشد [۲۵].

<sup>1</sup> Wavelet transform

اگر:

$$C_n = \int_{-\infty}^{+\infty} \frac{|H(w)|^2}{|w|} dw < \infty \quad (11-2)$$

که  $H(W)$  تبدیل فوریه تابع  $h(t)$  می‌باشد. ثابت  $C_n$  ثابت پذیرفتگی تابع  $h(t)$  و لازم است که برای وجود عکس تبدیل موجک شرط فوق برقرار باشد. هر تابعی که در شرط پذیرفتگی فوق صدق کند. می‌تواند یک تابع موجک مادر باشد. برای تابع داده شده  $h(t)$  شرط  $C_n < \infty$  برقرار است.

تنها اگر  $H(0)=0$  باشد. موجک‌ها ذاتا همانند باند عبور یک فیلتر در حوزه‌ی فوریه و یا به‌طور مشابه  $\int_{-\infty}^{+\infty} h(t) dt = 0$  که نشانگر مقدار متوسط صفر در حوزه‌ی زمان می‌باشد عمل می‌کند. تبدیل موجک از یک تابع  $f \in L^2(\mathbb{R})$  با توجه به شرط پذیرفتگی تابع موجک مادر  $h(t)$  تعریف می‌شود.

$$W_f(a, b) = \int_{-\infty}^{+\infty} f(t) h_{a,b}^*(t) dt \quad (12-2)$$

که نماد \* نشانگر مزدوج مختلط می‌باشد. اگرچه اکثر توابع موجک، توابع حقیقی می‌باشند، توابع فرزند از یک تابع موجک مادر  $h(t)$  با تغییر در پارامترهای انتقال و مقیاس آن تولید می‌شود.

$$h_{a,b}(t) = \frac{1}{\sqrt{a}} h\left(\frac{t-b}{a}\right) \quad (13-2)$$

که  $a > 0$  پارامتر مقیاس انتقال می‌باشد. ثابت  $a^{-\frac{1}{2}}$  برای نرمال‌سازی انرژی تابع موجک می‌باشد که از این طریق انرژی تابع موجک فرزند با انرژی تابع موجک مادر برابر می‌شود. اخیرا توابع موجک جدیدی که به‌طور کلی جالب‌تر از توابع موجک متعامد می‌باشند، پدید آمده‌اند؛ که می‌توان از آنها به‌عنوان توابع مادر استفاده نمود. [۲۶].

### ۲-۳-۷ ضرایب کپسترال حقیقی:

یکی از مشخصه‌هایی که از سیگنال گفتار استخراج می‌شود در بسیاری از کاربردها مورد استفاده قرار می‌گیرد ضرایب کپسترال می‌باشد. آنالیز کپسترال روشی متداول در حوزه‌ی

پردازش سیگنال و گفتار صوتی می‌باشد که به‌منظور فشرده‌سازی یا کد کردن گفتار باکیفیت بالا در نرخ پایین مورد استفاده قرار می‌گیرد. این روش به‌طور گسترده برای کد نمودن گفتار، سنتز گفتار، شناخت گفتار، شناخت گوینده، احراز هویت گوینده و ذخیره‌سازی استفاده می‌شود. این ضرایب نه‌تنها فیلتر مجرای گفتار را در خوددارند، بلکه حاوی اطلاعات سیگنال تحریک نیز هستند؛ بنابراین مشخصه مناسبی برای تشخیص گفتار، دسته‌بندی اصوات و... می‌باشند. ضرایب کپسترال حقیقی<sup>۱</sup> با استفاده از توابع لگاریتم با مقدار حقیقی تعریف شده‌اند. برای محاسبه ضرایب کپسترال حقیقی اندازه یا قسمت حقیقی موردنظر است و از آن DFT گرفته می‌شود [۲۷].

## ۲-۳-۸ ضرایب کپسترال فرکانس بارک:

روش ضرایب کپسترال فرکانس بارک<sup>۲</sup> از ترکیب دو روش ضرایب کپسترال فرکانس مل (MFCC) و ضرایب پیش‌بینی ادراکی خطی ایجاد شده است. مراحل استخراج این ویژگی بدین ترتیب می‌باشد [۲۸] و [۲۹].

۱- فریم بندی و پنجره کردن

۲- مجذور اندازه تبدیل فوریه با استفاده از الگوریتم تبدیل فوریه سریع

۳- سه مرحله آنالیز بیان شده در ضرایب پیش‌بینی ادراکی خطی

۴- اعمال لگاریتم

۵- اعمال تبدیل کسینوسی گسسته

---

<sup>1</sup> Real Cepstrum

<sup>2</sup> Bark Frequency Cepstral Coefficients

## ۲-۴ طبقه بند:

یکی از پرکاربردترین طبقه بندهای مورداستفاده شبکه عصبی مصنوعی می‌باشد. شبکه عصبی مصنوعی الگویی برای پردازش اطلاعات می‌باشد شبکه عصبی مصنوعی با پردازش روی داده‌های تجربی، دانش یا قانون نهفته در ورای داده‌ها را به ساختار شبکه منتقل می‌کند که به این عمل یادگیری می‌گویند. اصولاً توانایی یادگیری مهم‌ترین ویژگی یک سیستم هوشمند است. سیستمی که بتواند یاد بگیرد منعطف‌تر است و ساده‌تر برنامه‌ریزی می‌شود، بنابراین بهتر می‌تواند در مورد مسائل و معادلات جدید پاسخگو باشد. شبکه عصبی مصنوعی<sup>۱</sup> از نظر آموزش به دودسته آموزش با نظارت و آموزش بدون نظارت تقسیم می‌گردد [۳۰].

### ۲-۴-۱ آموزش با نظارت:

آموزش با استفاده از زوج بردارهای نمونه صورت می‌گیرد به طوری که به هر بردار ورودی یک بردار خروجی مشخص نسبت داده می‌شود. با ارائه این مجموعه بردارها به شبکه، وزن‌ها بر اساس الگوریتم یادگیری شبکه اصلاح می‌گردند. این‌گونه آموزش را آموزش با معلم می‌نامند. در این شیوه علاوه بر نمونه‌های آموزشی، پاسخ مطلوب هم وجود دارد. در این ارتباط می‌توان به ماشین بردار پشتیبانی، درخت تصمیم و پرسپترون اشاره کرد.

### ۲-۴-۱-۱ ماشین بردار پشتیبانی:

ماشین بردار پشتیبانی یکی از روش‌های یادگیری با نظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. این روش از جمله‌ی روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی از جمله شبکه‌های عصبی پرسپترون نشان داده است. یک دسته‌بند یا مرزی است که با معیار قرار دادن بردارهای پشتیبان، بهترین طبقه‌بندی و تفکیک بین داده‌ها را برای ما مشخص می‌کند. مبنای کاری

---

<sup>1</sup> Artificial Neural Network

طبقه بند کننده ماشین بردار پشتیبانی دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کنیم خطی را انتخاب کنیم که حاشیه اطمینان بیشتری داشته باشد [۳۱] و [۳۲].

#### ۲-۴-۱-۲ درخت تصمیم‌گیری:

درخت تصمیم یکی از ابزارهای قوی و متداول برای دسته‌بندی و پیش‌بینی می‌باشد و همچنین یک ابزار برای پشتیبانی از تصمیم است که از درختان برای مدل کردن استفاده می‌کند. درخت تصمیم یک ساختار درختی شبیه فلوچارت است. در این ساختار هر گره داخلی آزمونی را بر روی یک ویژگی مشخص می‌کند. درخت تصمیم به‌طور معمول در تحقیق در عملیات استفاده می‌شود، به‌طور خاص در آنالیز تصمیم، برای مشخص کردن استراتژی که با بیشترین احتمال به هدف برسد بکار، می‌رود [۳۳].

#### ۲-۴-۲ آموزش بدون نظارت:

در این نوع آموزش، بردارهای ورودی به شبکه ارائه گردیده بدون اینکه بردارهای خروجی مربوط به شبکه داده شوند، وزن‌های شبکه به صورتی اصلاح می‌شوند که بردارهای ورودی مشابه در یک گروه‌بندی قرار گیرند. پاسخ شبکه بر اساس نزدیک‌ترین بردار به بردار ورودی خواهد بود. در این حالت تنها داده‌های ورودی وجود دارند، بدون این‌که ناظر مقدار مطلوب را مشخص کند. یادگیری بدون ناظر را خود یادگیری هم می‌گویند. از شبکه‌های به‌کارگیرنده الگوریتم بدون ناظر می‌توان به شبکه‌های کوهونن، هاپفیلد و پرسپترون اشاره کرد.

#### ۲-۴-۱-۲ شبکه‌های کوهونن:

هدف از یک شبکه کوهونن<sup>۱</sup>، ایجاد یک طبقه بند الگوهاست که خودسازمانده بوده و از الگوریتم یادگیری کوهونن برای اصلاح اوزان استفاده نماید. معمولاً شبکه کوهونن شامل یک آرایی دوبعدی از

---

<sup>1</sup> Kohonen Neural Network

عصب‌ها است که در آن تمامی ورودی‌ها به تمامی عصب‌ها وارد می‌شود. هر عصب دارای مجموعه‌ای از اوزان است که می‌توان آن‌ها را به صورت الگوهای نمونه در نظر گرفت. زمانی که یک الگوی ورودی به شبکه وارد می‌شود، عصب حاوی الگوی نمونه‌ای که بیشترین شباهت را با الگوی ورودی دارد، بزرگ‌ترین پاسخ را تولید می‌کند. یکی از تفاوت‌های این شبکه با دیگر شبکه‌های خودسازمانده این است که در این شبکه، الگوهای نمونه به گونه‌ی خاصی ذخیره می‌شوند؛ یعنی نمونه‌های مشابه در عصب‌هایی قرار می‌گیرند که از نظر فیزیکی نزدیک به هم قرار داشته باشند و نمونه‌های متفاوت در عصب‌های دورتر از هم قرار می‌گیرند.

#### ۲-۲-۴-۲ شبکه‌های پفیلد:

نوعی از شبکه‌های عصبی مصنوعی بازگشتی است که می‌تواند در نقش سیستمی دارای حافظه وابسته شامل اجزایی دو حالتی ظاهر شود. همگرایی شبکه‌های پفیلد<sup>۱</sup> به یکی از کمینه‌های موضعیش قطعی است اما همگرایی به جواب از قبل ضبط‌شده هرچند بسیار محتمل است، اما قطعی نیست [۳۴].

#### ۲-۲-۴-۳ شبکه عصبی پرسپترون:

یکی از انواع پرکاربرد شبکه‌های عصبی، شبکه عصبی پرسپترون<sup>۲</sup> می‌باشد. در این شبکه وزن‌ها و بایاس‌ها می‌توانند برای تولید یک هدف مشخص، آموزش داده شوند. قوانین یادگیری مورداستفاده در این راستا را قوانین آموزش پرسپترون می‌نامند. شبکه‌های پرسپترون از آن جهت که توانایی مناسبی برای تکامل به‌وسیله بردارهای ورودی دارند، بسیار شایان توجه می‌باشند. این شبکه‌ها مخصوصاً در حل مسائل ساده طبقه‌بندی بسیار مناسب می‌باشند. این نوع شبکه عصبی در حل مسائل، بسیار سریع و قابل اطمینان است. این شبکه عصبی بر مبنای یک واحد محاسباتی به نام پرسپترون ساخته می‌شود. یک پرسپترون برداری از ورودی‌ها با مقادیر حقیقی را گرفته و یک ترکیب خطی از ورودی‌ها

<sup>1</sup> Hopfield Neural Network

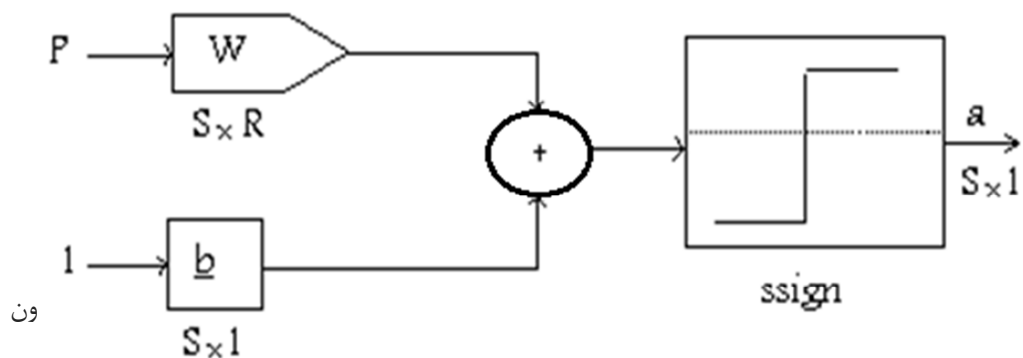
<sup>2</sup> Perceptron Neural Network

را محاسبه می‌کند. اگر حاصل از یک مقدار آستانه بیشتر بود خروجی پرسپترون برابر با 1 و در غیر این صورت معادل 1- خواهد بود. شبکه‌های عصبی پرسپترون، به‌ویژه پرسپترون چندلایه در زمره کاربردی‌ترین شبکه‌های عصبی می‌باشند، این شبکه‌ها قادرند با انتخاب مناسب تعداد لایه‌ها و سلول‌های عصبی، که اغلب هم زیاد نیستند، یک نگاشت غیرخطی را با دقت دلخواه انجام دهند [۳۵].

۲-۴-۲-۳-۱ شبکه عصبی پرسپترون تک لایه:

اولین شبکه‌ای که مورد بررسی قرار می‌گیرد شبکه پرسپترون تک لایه می‌باشد که ساختار آن در

شکل زیر تشریح شده است:



از خصوصیات اصلی پرسپترون ها قابلیت یادگیری یا train شدن می‌باشد. این یادگیری در پرسپترون ها supervised می‌باشد.

به این مفهوم که ما باید تعدادی ورودی به همراه خروجی صحیح داشته باشیم تا پرسپترون بتواند آن را تقلید نماید. یادگیری پرسپترون ها به شکل زیر می‌باشد:

۱- یک خروجی تولید می‌کنند.

۲- خروجی را با خروجی اصلی مقایسه می‌کنند.

۳- مقادیر را کمی تنظیم می‌کنند تا به خروجی نزدیک تر شوند.

۴- بعد از تکرار شدن این مراحل به تعداد کافی پرسپترون اصطلاحاً به رفتار صحیح همگرا یا

converge می‌شود. به این روش یادگیری  $\delta$  rule یا قانون دلتا می‌گویند.



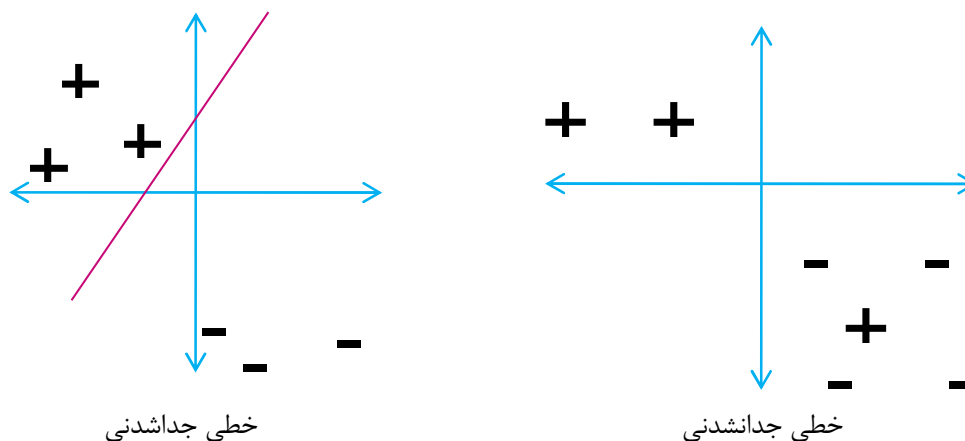
در این روش یادگیری تغییر وزن در هر مرحله به صورت زیر محاسبه می‌شود:

$$w_i = x_i \quad (۱۴-۲)$$

که دلتا در آن تفاوت خروجی موردنظر و خروجی نورون بوده و  $x_i$  مقدار ورودی است. در حالت کلی یادگیری پرسپترون عبارت است از پیدا کردن مقادیر درستی برای  $W$  بنابراین فضای فرضیه  $H$  در یادگیری پرسپترون عبارت است از مجموعه تمام مقادیر حقیقی ممکن برای بردارهای وزن خروجی پرسپترون توسط رابطه زیر مشخص می‌شود:

$$(X_1, X_2, \dots, X_n) = \begin{cases} 1, & \text{if } w_0 + w_1X_1 + \dots + W_nX_n < 0 \\ -1, & x \geq 0 \end{cases} \quad (۱۵-۲)$$

یک پرسپترون قادر است مثال‌هایی را یاد بگیرد که به صورت خطی جدا پذیر باشند. این گونه مثال‌ها مواردی هستند که به طور کامل توسط یک hyperplane قابل جداسازی می‌باشند.

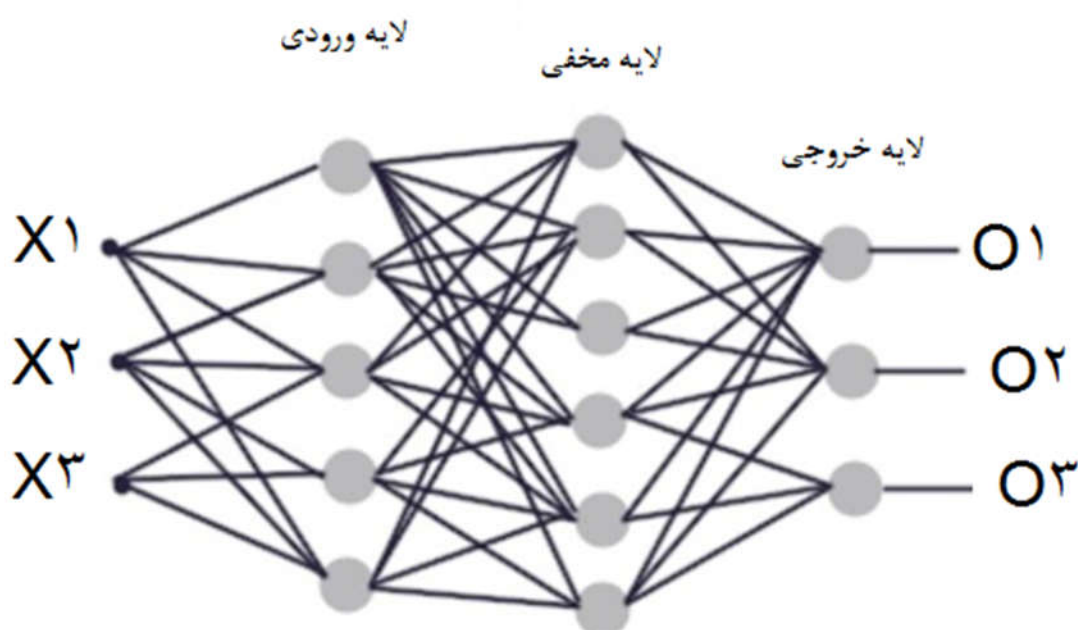


شکل ۲-۳ توابع خطی جداسدنی و جدانشدنی

۲-۴-۲ شبکه عصبی پرسپترون چندلایه (MLP):

برخلاف پرسپترونها شبکه‌های چندلایه می‌توانند برای یادگیری مسائل غیرخطی و همچنین مسائلی با تصمیم‌گیری‌های متعدد بکار روند.

شبکه عصبی پرسپترون چندلایه<sup>۱</sup> مجموعه‌ای از نورون‌ها است که در لایه مختلفی پشت سر هم قرار گرفته‌اند. مقادیر ورودی پس از ضرب در وزن‌های موجود در گذر گاه‌های بین لایه‌ها به نورون بعدی رسیده و در آنجا باهم جمع می‌شوند و پس از عبور از تابع شبکه مربوطه خروجی نورون‌ها را تشکیل می‌دهند. در پایان خروجی به دست آمده با خروجی مورد نظر مقایسه شده و خطای به دست آمده جهت اصلاح وزن‌های شبکه به کار می‌رود، این امر اصطلاحاً آموزش شبکه عصبی نامیده می‌شود.



شکل ۲-۴ شبکه عصبی پرسپترون چندلایه [۳۶]

مدل پرسپترون چندلایه، سه لایه دارد یک لایه ورودی، یک لایه خروجی و یک لایه بین آن‌ها که مستقیماً به داده‌های ورودی و نتایج خروجی متصل نیست. در واقع این لایه را لایه پنهان می‌نامند. هر واحد در لایه پنهان و لایه خروجی مانند یک پرسپترون عمل می‌کند، با این تفاوت که تابع استفاده شده به جای تابع پلکانی تابع سیگموئید است. واحدهای لایه ورودی صرفاً وظیفه توزیع مقادیر ورودی را به لایه بعدی بر عهده دارند و بنابراین هیچ محاسبه‌ای را انجام نمی‌دهند. نحوه عمل پرسپترون چندلایه‌ای مشابه پرسپترون تک لایه‌ای است. بدین صورت که الگویی به شبکه عرضه

<sup>1</sup> Multi-Layer Perceptron (MLP)

می‌شود و خروجی آن محاسبه می‌گردد، مقایسه خروجی واقعی و خروجی مطلوب باعث می‌گردد که ضرایب وزنی شبکه تغییر یابد به طوری که در دفعات بعد خروجی درست‌تری حاصل شود. قاعده فراگیری روش تنظیم ضرایب وزنی شبکه را بیان می‌کند. قاعده ساده فراگیری پرسپترون تک لایه‌ای در مورد پرسپترون چندلایه‌ای کاربرد ندارد.

بنابراین استفاده از تابع سیگموئید بدان معنی است که واحدهای میانی تا اندازه‌ای نسبت به خروجی‌های مدل آگاهی دارند به طوری که می‌توان ضرایب وزنی آن‌ها را برای کاهش میزان خطا تنظیم کرد. وقتی به شبکه آموزش ندیده‌ای الگویی عرضه شود، خروجی‌های تصادفی تولید می‌کند. لذا ابتدا تابع خطایی تعریف می‌شود تا تفاوت خروجی واقعی و خروجی‌های تصادفی را تولید کند. برای اینکه میزان تابع خطا به طور دائم کاهش داده شود ضرایب وزنی خطوط ارتباطی واحدها میزان می‌شود. قاعده دلتا مقدار تابع خطا را محاسبه کرده و آن به این علت است. نهایتاً ضرایب وزنی هر لایه پس از انتشار را به عقب از یک لایه به لایه پیشین آن انتشار می‌دهد. عبارت واحد جداگانه تنظیم می‌شود و بدین صورت میزان خطا کاهش می‌یابد [۳۶].

## ۲-۵ مروری بر کارهای پیشین:

در این بخش از فصل دوم به بررسی مقالاتی که حول محور تشخیص صدای ضربه‌ای منتشر شده می‌پردازیم.

در [۳۷] از پایگاه داده‌ای با ۷ کلاس از صداهای محیطی استفاده شده است. در بخش شناسایی نیز از سه روش همبستگی متقابل، روش اکتشافی مبتنی بر پیش‌بینی انرژی<sup>۱</sup> و فیلتر موجک<sup>۲</sup> استفاده می‌کند و سپس با استفاده از مقایسه نتایج بهترین عملکرد را به فیلتر موجک و ضعیف‌ترین را به پیش‌بینی انرژی نسبت داده است. نکته حائز اهمیت این است که در این پروژه بخش شناسایی و تشخیص به طور مستقل از هم کار می‌کنند. در بخش تشخیص با ویژگی‌هایی همچون MFCC، مشتق اول،

<sup>1</sup> Energy prediction based detection

<sup>2</sup> Wavelet filtering based detection

مشتمل بر دوام آن، نرخ عبور از صفر، نقطه قطع طیف، LPC، انرژی و ترکیب‌های متفاوت آن‌ها بررسی شده و به طبقه بند اعمال می‌گردد؛

پس از انجام آزمایش‌ها با توجه به نتایج به دست آمده در این پروژه به این مهم دست‌یافته‌اند که وقتی از ویژگی‌های نرخ عبور از صفر، نقطه طیف و نقطه تعادل طیفی MFCC با ضریب ۱۶ و انرژی در کنار هم استفاده می‌شود کمترین نرخ خطای زمانی حاصل می‌شود و مقدار خطای ۱۵٪ می‌باشد.

در [۳۸] برای شناسایی سیگنال صوتی از طبقه بندهای مختلفی برای مقایسه‌ی بین روش‌های ایستا و روش‌های غیرایستا مبتنی بر موجک بهره برده می‌شود؛ در این مقاله از روش‌های زمان کوتاه، موجک گسسته، موجک پیوسته و توزیع وینر-ویل استفاده می‌کند و برای استخراج ویژگی روش‌های MFCC، LPC و PLP را مورد بررسی قرار می‌دهد. ویژگی‌های استخراج شده را به طبقه بندهای ماشین بردار پشتیبان، تخمین بیشترین شباهت<sup>۱</sup> و مدل مخفی مارکوف برای ارزیابی می‌دهد و سپس نتایج بدست آمده را مقایسه کرده و بهترین نتیجه برای شناسایی سیگنال صوتی را به طبقه‌بند تبدیل موجک گسسته با ویژگی تبدیل فوریه پیوسته نسبت داده که در این حالت نرخ تشخیص سیگنال صوتی برابر با ۷۰٪ اعلام شده است.

هدف در مقاله [۳۹] افزایش تعداد کلاس پایگاه داده می‌باشد که تأثیر این افزایش را بررسی نرخ تشخیص مشاهده نماید. برای این منظور از ۶۱ کلاس به‌عنوان پایگاه داده استفاده می‌نماید. برای استخراج این ویژگی روش‌های MFCC، مشتق اول و دوم را مورد بررسی قرار داد و ویژگی‌های استخراج‌شده را به طبقه بند مدل مخفی مارکوف اعمال می‌کند. سپس از نتایج بدست آمده بدین‌صورت نتیجه‌گیری می‌کند که دقت تشخیص به ازای ۶۱ کلاس ۲۴٪ بوده است. پس نتیجه گرفت می‌شود که با افزایش تعداد کلاس نرخ تشخیص کاهش می‌یابد.

نویسندگان در مقاله [۴۰] به تشخیص تیراندازی در محیط‌های نویزی پرداخته است. بدین صورت که ابتدا صدا وارد سیستم صوت می‌شود با استفاده از فیلتر ضربه‌ای بودن نوع سیگنال صوتی تأیید

---

<sup>1</sup> Correlatlation

می‌گردد و سپس برای استخراج ویژگی به مرحله بعد ارسال می‌گردد. برای استخراج ویژگی از روش‌های LPC و MFCC با ضریب ۱۳ استفاده می‌کند. ویژگی استخراج شده را به طبقه بند مدل مخفی مارکوف می‌دهد. پس از بررسی و مقایسه نتایج این گونه استنباط می‌گردد که در محیط‌های بدون نویز مدل مخفی مارکوف با استفاده از ویژگی‌های LPC و MFCC به صورت پایدار به خوبی در شناسایی نوع صدای ضربه‌ای عمل می‌کند، اما در محیط‌هایی با حضور نویز شدید استفاده از طبقه بند با همبستگی متقابل عملکرد بهتری دارد. همچنین از مزایای استفاده از همبستگی متقابل می‌توان به محاسبات آسان و عملکرد بهتر نسبت به طبقه‌بندی‌های دیگر در محیط‌هایی با حضور نویز اشاره کرد. در این مقاله [۴۱] برای شناسایی سیگنال صوتی، پژوهشگر در ابتدا برای استخراج ویژگی از تبدیل موجک با استفاده از موجک گاماتن استفاده کرده و نتایج حاصل از این ویژگی را با موجک رابیشنر مقایسه می‌کند. سپس ویژگی‌های بدست آمده را به طبقه بند ماشین برداری پشتیبان می‌دهد. با مقایسه نتایج بدست آمده موجک گاماتن با توجه به نوع عملکرد دارای بازدهی بهتری نسبت به سایر موجک‌ها در محیط‌های بدون نویز و با نویز را دارد.

محققان برای ارتباط و تعامل بین انسان و ربات در مقاله [۸] سیستم شنوایی ربات را برای تشخیص صداهای محیط مورد بررسی قرار داده‌اند. برای استخراج ویژگی از ردگیری انطباق گابور<sup>۱</sup> و MFCC استفاده کرده و عملکرد این دو ویژگی را مورد بررسی قرار گرفته است؛ سپس ویژگی‌های بدست آمده را برای طبقه‌بندی به مدل مخفی مارکوف و مدل مخلوط گوسی می‌دهد. با استفاده از نتایج به دست آمده بهترین نتایج زمانی برای تشخیص صدا حاصل می‌گردد که از ویژگی‌های ردگیری انطباق گابور و طبقه بند مارکوف مخفی استفاده می‌کنیم؛ همچنین شایان ذکر است استفاده از ویژگی ردگیری انطباق و طبقه بند مخفی مارکوف نتایج بهتری در شناسایی سیگنال صوتی به استفاده از ویژگی MFCC را دارا می‌باشد.

---

<sup>1</sup> MP-Gabor

در [۴۲] به طراحی یک سیستم برای شناسایی سیگنال صوتی در یک محیط با حضور نویز می‌پردازد. برای استخراج ویژگی از ویژگی‌های ادراکی MFCC و PLP استفاده می‌نمایند و ویژگی‌های استخراج‌شده را برای طبقه‌بندی به ۳ سطح طبقه‌بندی می‌کنند:

۱- طبقه‌بندی در محیط با حضور نویز و یا عدم حضور نویز.

۲- طبقه‌بندی در محیط‌های بدون حضور نویز که به تفکیک سیگنال گفتار و یا غیر گفتار می‌پردازد.

۳- طبقه‌بندی سیگنال گفتار که فریاد باشد یا غیر فریاد.

به همین منظور از دو مدل مخلوط گوسی GMM و بردار پشتیبان SVM استفاده می‌کند. نتایج آزمایش‌ها نشان می‌دهد که طبقه بند مدل ترکیبی دقت بالایی در تشخیص سیگنال صوتی را دارد.

همچنین در مقاله [۴۳] برای شناسایی سیگنال صوتی از ۹ کلاس سیگنال به‌عنوان پایگاه داده استفاده می‌نماید. در ابتدا برای استخراج ویژگی از سیگنال صوتی ویژگی‌های مبتنی بر تبدیل موجک در کنار MFCC را مورد بررسی قرار می‌دهد. سپس ویژگی‌های استخراج‌شده را به طبقه بند ماشین بردار پشتیبان می‌دهد. پس از انجام آزمایش و با استفاده از نتایج بدست آمده بدین‌صورت نتیجه می‌گیرد که با استفاده از دودسته ویژگی کنار هم نرخ تشخیص را به ۹۳٪ می‌رساند که این نتیجه در مقایسه با استفاده از MFCC و یا تبدیل موجک به‌تنهایی بسیار بهتر می‌باشد و نرخ تشخیص را بهبود می‌بخشد.

در [۴۴] از ۴ کلاس به‌عنوان پایگاه داده استفاده می‌کند. برای استخراج ویژگی روش‌های MFCC، LPC، با ضریب ۱۶ و همبستگی متقابل را مورد بررسی قرار می‌دهد. سپس ویژگی‌های استخراج‌شده را به طبقه بند مدل مخفی مارکوف می‌دهد. در این مقاله با توجه به شباهت‌هایی که در سیگنال‌ها با سیگنال موردنظر دارد عمل شناسایی را انجام می‌دهد. پس از مقایسه نتایج به دست آمده بدین‌صورت

نتیجه می‌گیرد که ویژگی MFCC و LPC در محیط بدون نویز عملکرد بهتری دارد اما در محیط‌هایی با حضور نویز ویژگی‌های همبستگی متقابل نتایج بهتری را ارائه می‌دهند.



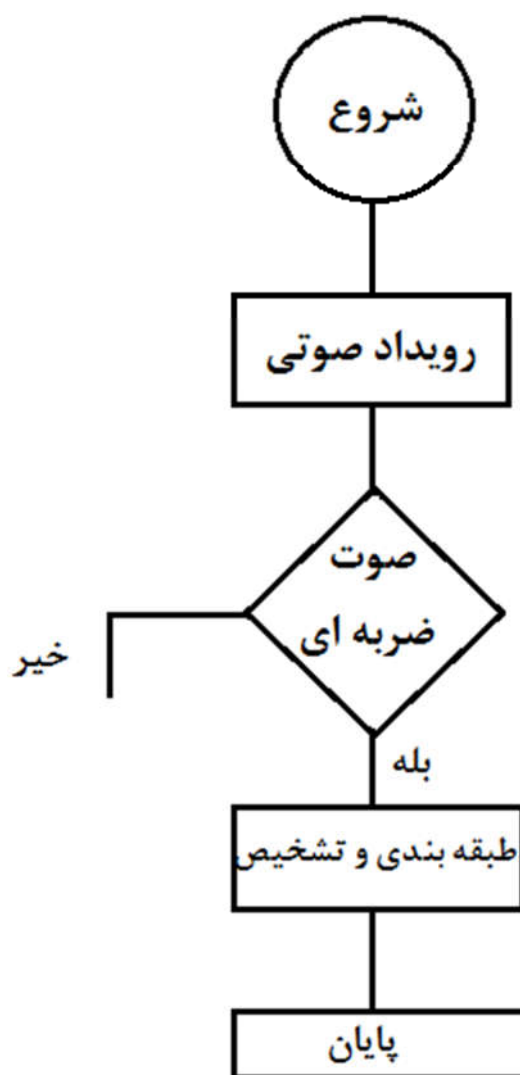


فصل سوم:

روش پیمایشی

### ۱-۳ مقدمه

هدف از انجام این پایان نامه تشخیص خودکار اصوات ضربه‌ای در قالب صدای اسلحه است. از اساسی‌ترین مراحل سیستم‌های تشخیصی، مرحله استخراج ویژگی و مرحله طبقه‌بندی است. ما در این پایان نامه از روشی که در شکل (۱-۳) نشان داده شده استفاده می‌کنیم.



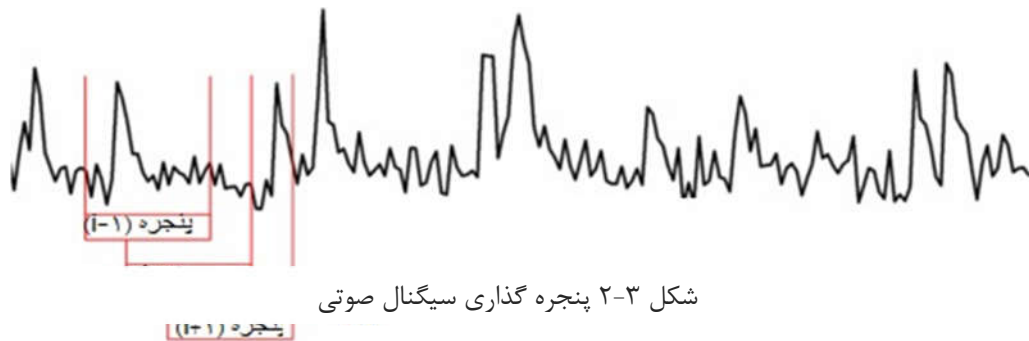
شکل ۱-۳ بلوک دیاگرام شناسایی، تشخیص و طبقه‌بندی اصوات ضربه‌ای

این سیستم از سه بخش اصلی ورودی سیگنال صوتی، تشخیص اصوات ضربه، طبقه‌بندی و تشخیص نوع اسلحه شلیک‌شده تشکیل‌شده است. کارایی این سیستم بدین‌صورت است که، اگر سیگنال ورودی ضربه‌ای باشد برنامه آن را برای تشخیص و دسته‌بندی به بخش طبقه‌بندی و تشخیص ارسال می‌کند ولی اگر سیگنال صوتی ضربه‌ای نباشد برنامه تمام می‌شود. در بخش طبقه‌بندی و تشخیص نوع اسلحه شلیک‌شده مشخص می‌گردد که در ادامه به تشریح تمام مراحل می‌پردازیم.

در کار بر روی داده‌های صوتی باید توجه داشت این سیگنال‌ها، سیگنال‌هایی غیر ایستا هستند، بنابراین برای تحلیل آن‌ها نیاز است این سیگنال‌ها را به پنجره‌های کوچک تقسیم کنیم بطوریکه بتوان فرض کرد در این پنجره‌ها سیگنالی ایستا هستند. سپس هر یک از سیگنال‌های ورودی با استفاده از روش‌های MFCC، LPC و Real Cepstrum استخراج ویژگی می‌کنیم و نهایتاً با استفاده از یک شبکه عصبی پرسپترون چندلایه طبقه‌بندی اصوات را انجام می‌دهیم. در ادامه این فصل با تکیه بر مبانی نظری بیان‌شده در فصل قبل، روش پیشنهادی برای طبقه‌بندی اصوات ضربه‌ای و الگوریتم‌های استفاده‌شده در روش پیشنهادی را بیان می‌کنیم.

### ۳-۲ پنجره گذاری:

بدین منظور ما اصوات مورد نظرمان را که در مراحل بعدی مورد استفاده قرار می‌دهیم با پنجره‌های مستطیلی به طول ۲۵۶ نمونه و با روی هم افتادگی ۲۵ درصد یعنی ۶۴ نمونه، بخش‌بندی می‌کنیم؛ بنابراین پس از این مرحله مانند شکل (۳-۲) هر سیگنال صوتی ما به تعدادی قاب به طول ۲۵۶ نمونه بخش‌بندی می‌شود که در هر یک از این قاب‌ها با توجه به فرکانس نمونه‌برداری صوت‌ها در پایگاه داده مورد استفاده، می‌توانیم سیگنال صوتی را سیگنالی ایستا در نظر بگیریم.



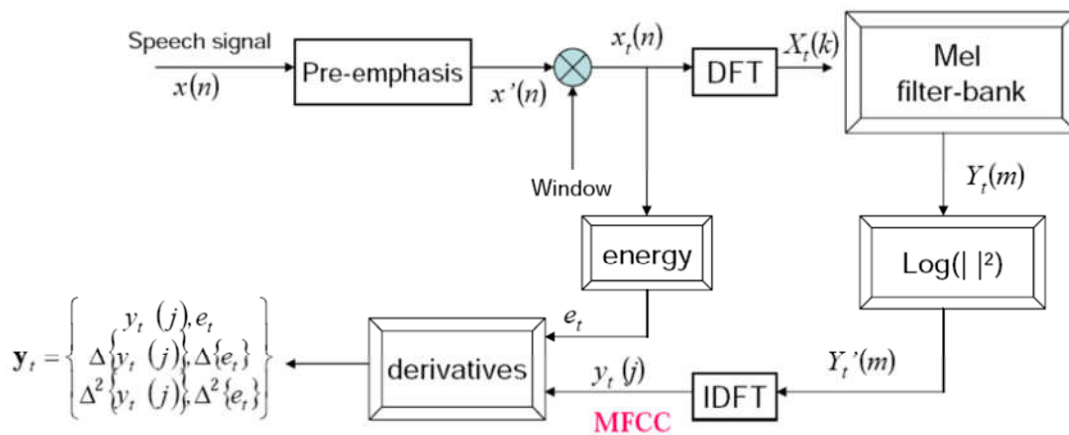
شکل ۳-۲ پنجره گذاری سیگنال صوتی

### ۳-۳ ویژگی‌های مورد استفاده :

پیش از طبقه‌بندی اصوات، به‌عنوان ورودی طبقه بند، از اصوات موردنظرمان باید استخراج ویژگی نمائیم. بدین منظور هر صوت را با پنجره مستطیلی به طول ۲۵۶ نمونه و با همپوشانی ۶۴ نمونه به سیگنال‌های ایستا بخش‌بندی می‌کنیم و پس‌از آن از روش‌های MFCC، LPC و Real Cepstrum از هر کدام استخراج ویژگی می‌کنیم که به‌تفصیل در ادامه آن‌ها را بیان می‌کنیم.

### ۳-۳-۱ ضرایب کپسترال در مقیاس مل:

در روش MFCC همانند شکل (۳-۳) باید پس از پنجره گذاری سیگنال صوتی، انرژی سیگنال صوتی را محاسبه کرده و به‌عنوان یکی از ویژگی‌ها مدنظر قرارداد؛ همچنین از سیگنال پنجره گذاری شده با طی روندی شامل محاسبه تبدیل فوریه گسسته، استفاده از فیلتر بانک مل، محاسبه لگاریتم اندازه و نهایتاً محاسبه معکوس تبدیل فوریه گسسته ویژگی دیگری استخراج کرد. با توجه به اینکه ما از ۱۲ فیلتر مل استفاده کردیم خروجی این قسمت شامل ۱۲ ویژگی خواهد بود؛ بنابراین تاکنون ما ۱۳ ویژگی داریم. اکنون از این ۱۳ ویژگی مشتق مرتبه اول و مشتق مرتبه دوم آن‌ها را نیز محاسبه می‌کنیم؛ که نهایتاً در مجموع ۳۹ ویژگی خواهیم داشت. بنابراین در ابتدا هر سیگنال صوتی را قاب‌بندی کردیم و با استخراج ویژگی MFCC از این قاب‌ها، برای هر قاب ۳۹ ویژگی به دست می‌آوریم.



شکل ۳-۳ بلوک دیاگرام ضرایب کپسترال در مقیاس مل [۲۰]

### ۲-۳-۳ ضرایب کپسترال حقیقی :

برای استخراج ویژگی به روش کپستروم حقیقی مطابق شکل (۳-۴) باید پس از قاب‌بندی

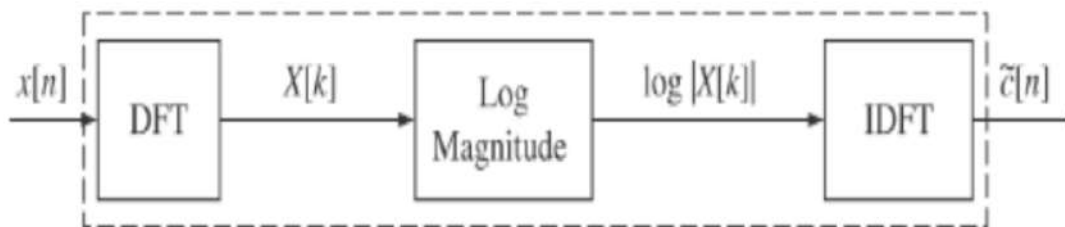
سیگنال صوتی، از نمونه‌های سیگنال صوتی تبدیل فوریه گسسته بگیریم، پس از آن لگاریتم

اندازه هر نمونه را محاسبه می‌کنیم و نهایتاً معکوس تبدیل فوریه گسسته خواهیم گرفت.

با طی مراحل بیان‌شده برای هر قاب صوتی، مجموعه‌ای از ویژگی‌ها به‌اندازه طول قاب صوتی داریم

درنهایت در این روش کافی است تعداد مناسب از مجموعه ویژگی را برداریم. ما در این استخراج

ویژگی، ۲۰ نمونه اول را به‌عنوان ویژگی برای هر قاب صوتی انتخاب می‌کنیم.



شکل ۳-۴ بلوک دیاگرام ضرایب کپسترال حقیقی [۲۷]

### ۳-۳-۳ ضرایب پیش‌بینی خطی:

برای تقریب یک سیگنال صوت با استفاده از روش ضرایب پیشگویی خطی، باید ابتدا سیگنال صوتی را پنجره گذاری کنیم و این سیگنال پنجره گذاری شده را که با تقریب خوبی یک سیگنال ایستا می‌توان در نظر گرفت که می‌توانیم با تعداد مناسب ضرایب پیشگویی تقریب بزنیم.

به‌منظور بررسی تاثیر تعداد ضرایب در تشخیص اصوات ضربه‌ای، در روش پیشنهادی تعداد این ضرایب را ۱۰، ۲۰ و ۵۰ در نظر می‌گیریم.

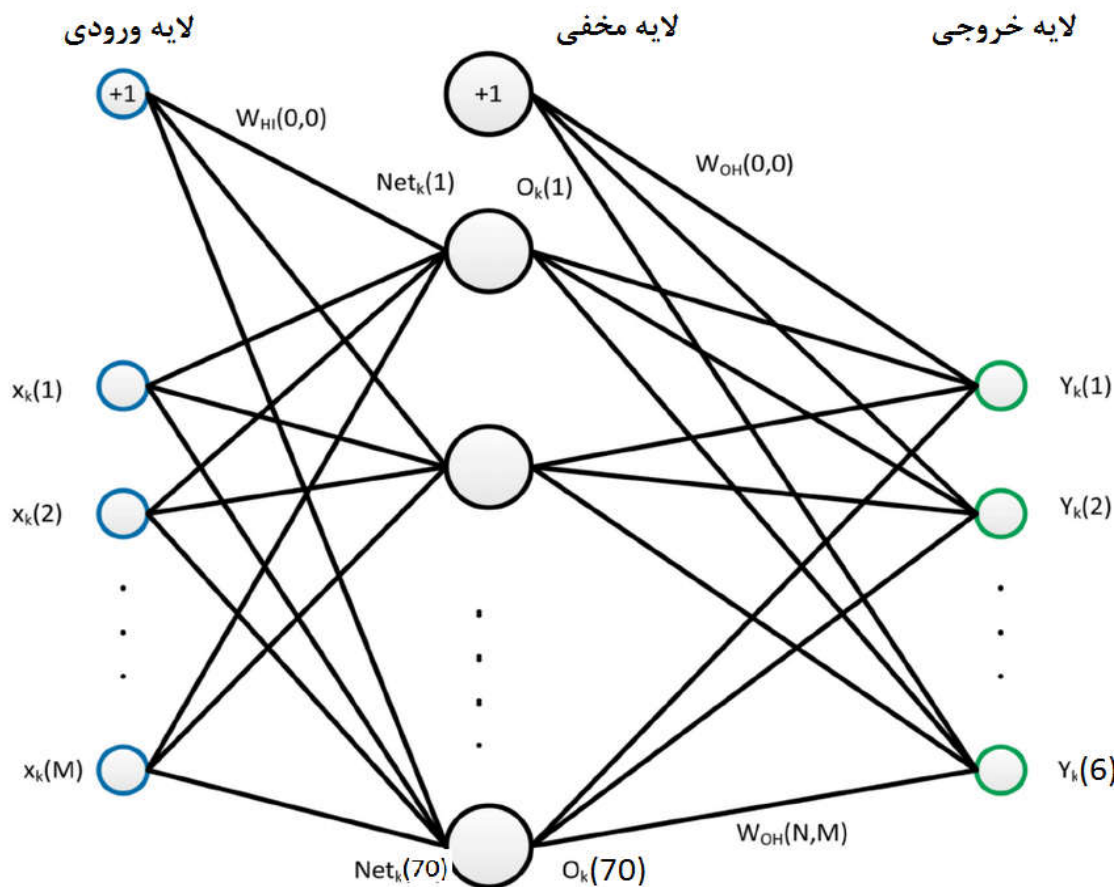
برای مرحله استخراج ویژگی با ضرایب پیشگویی خطی، پایگاه داده را با سه ضریب مختلف استخراج ویژگی می‌کنیم؛ بنابراین برای هر قاب از یک صوت ضربه‌ای در روش LPC با ۱۰ ضریب، ۱۰ ویژگی استخراج کردیم و در مرحله طبقه‌بندی از آن استفاده می‌کنیم؛ در روش LPC با ۲۰ ضریب، ۲۰ ویژگی استخراج کردیم و در مرحله طبقه‌بندی از آن استفاده می‌کنیم و در روش LPC با ۵۰ ضریب، ۵۰ ویژگی استخراج کردیم و در مرحله طبقه‌بندی از آن استفاده می‌کنیم.

### ۳-۴ طبقه‌بند مورد استفاده:

برای قسمت طبقه‌بندی روش پیشنهادی از یک شبکه‌ی عصبی پرسپترون چندلایه استفاده می‌کنیم. با توجه به تعداد ویژگی‌های استخراج‌شده در مرحله قبل برای هر روش تعداد نورون‌های لایه‌ی ورودی را متناسب با ویژگی‌های استخراج‌شده انتخاب می‌کنیم. با توجه به اینکه پایگاه داده ما که شامل اصوات ۶ کلاس از صداهاست برای لایه خروجی شبکه عصبی ۶ نورون در نظر می‌گیریم. تعداد نورون‌های لایه مخفی را برای تمام روش‌های استخراج ویژگی یکسان و برابر با ۷۰ در نظر می‌گیریم؛ چراکه بتوانیم تاثیر دقت طبقه‌بندی را فقط ناشی از استخراج ویژگی‌های متفاوت بررسی کنیم.

بنابراین شبکه عصبی که برای طبقه‌بندی استفاده می‌کنیم یک شبکه عصبی پرسپترون دولایه (بدون احتساب لایه ورودی) می‌باشد که تعداد نورون‌های لایه مخفی برابر با ۷۰ و تعداد نورون‌های

لایه خروجی برابر ۶ می‌باشد و با روش پس انتشار خطا در مرحله آموزش وزن‌های شبکه اصلاح می‌شود.



شکل ۳-۵ شبکه عصبی مورد استفاده برای طبقه‌بندی

به منظور بررسی تاثیر استخراج ویژگی‌های متفاوت برای تمام روش‌های مورد استفاده تعداد دوره‌های آموزش شبکه با آزمون و خطا مقادیر ۱۰۰۰، ۱۵۰۰ و ۲۰۰۰ دوره در نظر گرفته‌ایم. طبقه بند ما یک شبکه عصبی پرسپترون دو لایه مطابق شکل (۳-۵) می‌باشد که تعداد نرون‌های لایه ورودی آن بسته به روش استخراج ویژگی از پایگاه داده متفاوت خواهد بود.

### ۳-۵ جمع‌بندی

در این فصل روش پیشنهادی برای طبقه‌بندی اصوات ضربه‌ای را بیان کردیم. بدین منظور از سه روش ضرایب کپسترال در مقیاس مل، ضرایب کپسترال حقیقی و ضرایب پیش‌بینی خطی با سه ضریب ۱۰ و ۲۰ و ۵۰ برای استخراج ویژگی از اطلاعات پایگاه داده بهره بردیم و برای مرحله طبقه‌بندی از یک شبکه عصبی مصنوعی پرسپترون دولایه که با روش پس انتشار خطا در مرحله آموزش اصلاح وزن‌ها را انجام می‌دهد، استفاده کردیم.

در فصل بعد بایان نتایج حاصل از روش پیشنهادی و مقایسه نتایج ناشی از استفاده از ویژگی‌های متفاوت به تحلیل و بررسی نتایج خواهیم پرداخت.



فصل چهارم:

نتایج روش پیمایشی

#### ۴-۱ مقدمه:

در این فصل با توجه به روش پیشنهادی بیان شده و با معرفی پایگاه داده تهیه شده نتایج حاصل از انجام روش پیشنهادی را برای استخراج ویژگی‌های متفاوت تحلیل و بررسی می‌کنیم.

#### ۴-۲ پایگاه داده:

پایگاه داده از ۶ کلاس صدای ضربه‌ای ناشی از شلیک اسلحه سبک تشکیل شده است؛ شامل اسلحه رولور رو جر ۳۵۷ مگنوم، اسلحه کلت ام ۱۹۱۱ (M1911)، اسلحه رولور مانورهاین MR-73، اسلحه کلت برتا ام ۹ (Beretta M 9)، اسلحه کلت دزرت ایگل (Desert Eagle) و اسلحه رولور SAA 1873

#### ۴-۲-۱ رولور رو جر ۳۵۷ مگنوم:

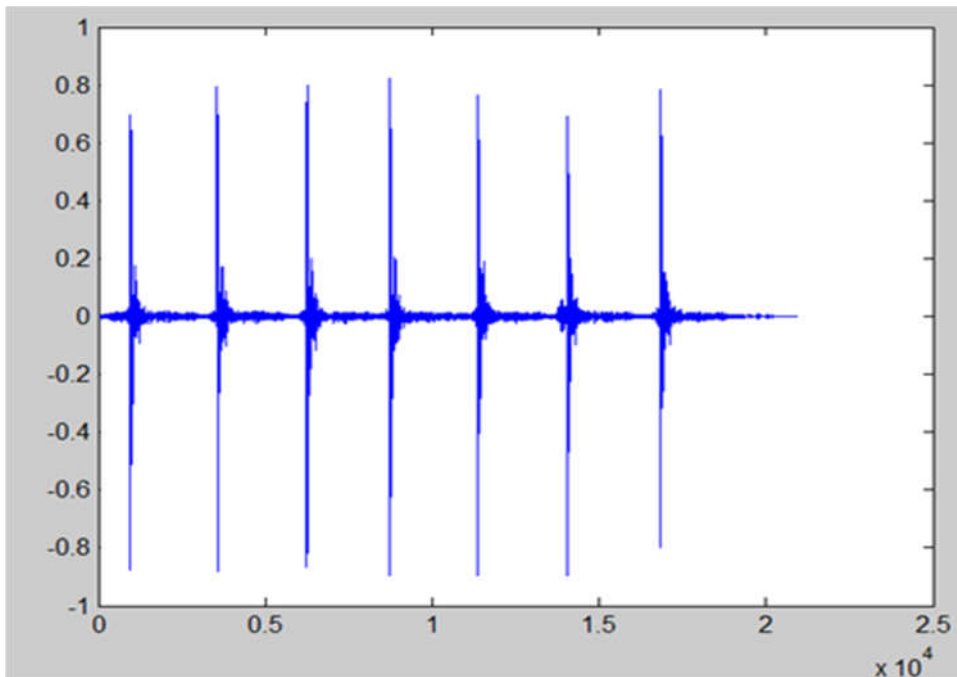
این رولورها از سال ۱۹۸۵، در کارخانه Ruger و Sturm واقع در امریکا ساخته می‌شوند. مدل GP-100 در واقع جانشین ارزنده مدل‌های معروف به ۶ می‌باشند که عمدتاً در عملیات پلیسی و امنیتی استفاده می‌شدند. تمامی مدل‌های GP-100، تمام از جنس فولاد ضدزنگ ساخته می‌شوند. توپی همگی از نوع ۶ تیر و ماشه آن‌ها دوزمانه می‌باشد. مدل‌های کالیبر ۰/۳۵۷ و بعضی مدل‌های کالیبر ۰/۳۸، دارای یک زائده اضافی در زیر لوله هستند که در بدنه سلاح جاسازی شده و مقاومت لوله را افزایش می‌دهد. سوزن آتش و ایمنی آن به گونه‌ای است که تنها زمانی عمل می‌کند که ماشه سلاح کاملاً کشیده شود. متعلقات ماشه، به گونه‌ای واحد طراحی شده‌اند که باز و بسته کردن سلاح را تسهیل می‌کند. ضامن سلاح علاوه بر قفل چخماق، مانع از چرخیدن توپی سلاح می‌شود. دستگاه نشانه‌روی سلاح در قسمت عقب، در دو نوع ثابت و قابل تنظیم در دسترس می‌باشد. تمامی رولورهای GP-100، دارای ضریب اطمینان و دوام فوق‌العاده‌ای هستند.



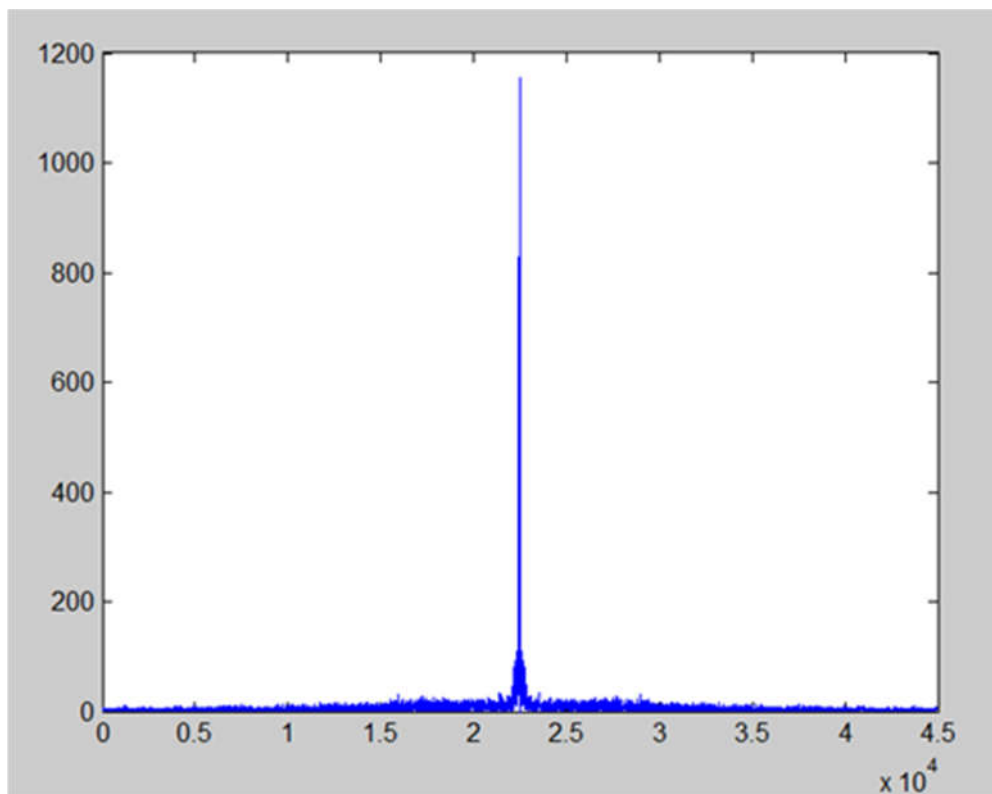
شکل ۴-۱ روجر ۳۵۷ مگنوم [۴۵]

این اسلحه دارای کالیبر ۳۵۷ مگنوم و ۳۸ اسپیشیال با طول لوله ۱۰۲،۷۶ و ۱۵۲ میلی‌متر در ۳ مدل تولید می‌شود. مکانیزم ماشه دوحالته و ظرفیت توپی ۶ گلوله را دارد. وزن آن ۱۰۰۰ تا ۱۳۰۰ گرم در مدل‌های مختلف می‌باشد [۴۵].

در شکل (۴-۲) نمونه‌ای از سیگنال صوتی شلیک شده توسط اسلحه روجر ۳۵۷ مگنوم رسم شده است.



شکل ۴-۲ سیگنال صوتی شلیک اسلحه ۱ در حوزه زمان



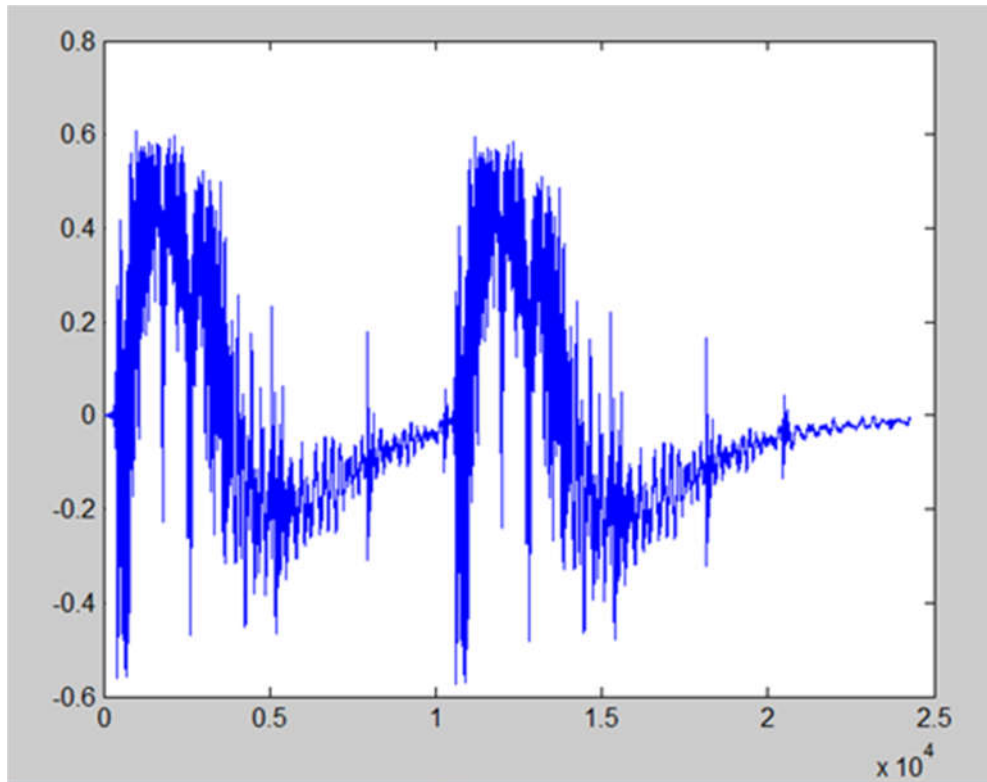
شکل ۳-۴ سیگنال صوتی شلیک اسلحه ۱ در حوزه فرکانس

#### ۴-۲-۲ کلت ام ۱۹۱۱:

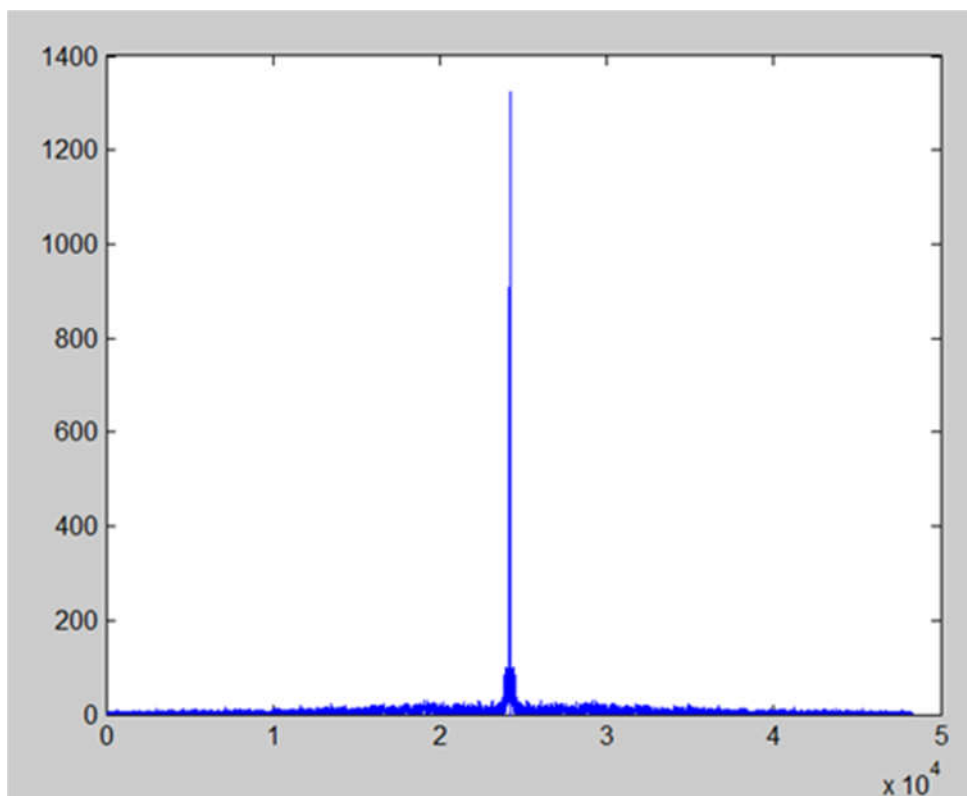
کلت ۱۹۱۱ سلاح کمربندی نیمه خودکاری است که از سال ۱۹۱۱ تا ۱۹۸۵ سلاح کمربندی سازمانی نیروهای مسلح آمریکا بود. اغلب آن را «ام ۱۹۱۱» یا «کلت ۴۵» می‌نامیدند. این سلاح در جنگ اول، جنگ دوم، جنگ کره و جنگ ویتنام بکار رفت. در ایران بیشتر به «کلت» مشهور است و دیگر سلاح‌هایی (مانند استار، براونینگ و غیره) را که شبیه به آن ساخته شده نیز کلت می‌گویند. کلت ۱۹۱۱ را جان براونینگ طرح کرد. در شکل (۴-۲) نمونه‌ای از سیگنال صوتی شلیک شده توسط اسلحه کلت ام ۱۹۱۱ نشان داده شده است. است کالیبر اصلی این سلاح ۰/۴۵ است ولی در کالیبرهای دیگر مانند ۰/۳۸ و ۷/۶۲ نیز ساخته شده است [۴۶].



شکل ۴-۴ کلت ام ۱۹۱۱ [۴۶]



شکل ۴-۵ سیگنال صوتی شلیک کلت ام ۱۹۱۱ در حوزه زمان



شکل ۴-۶ سیگنال صوتی شلیک کلت ام ۱۹۱۱ در حوزه فرکانس

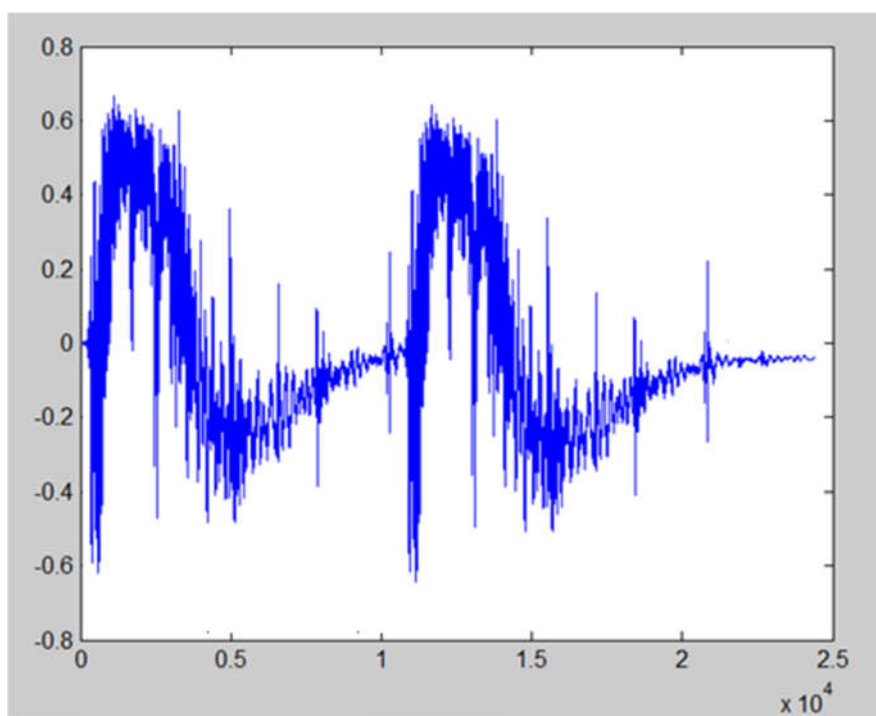
### ۴-۲-۳ رولور مانور هین MR-73 :

رولور MR-73، یکی از خوش‌دست‌ترین رولورهای ساخت فرانسه است. این رولور در سال ۱۹۷۳ و توسط کارخانه مانور هین طراحی و ساخته شد. این رولور، سلاح سازمانی ژاندارمری و برخی از واحدهای نیروی پلیس کشور فرانسه از جمله G.I.G.N (نیروهای ویژه رهایی گروگان) و R.A.I.D محسوب می‌شود. مدل‌های ورزشی این رولور در کالیبرهای ۰/۳۲ بلند اسمیت و سون ویا ۰/۳۸ اسپشیل، از جمله پرطرفدارترین رولورهای اروپایی به شمار می‌آیند. تمامی مدل‌های رولور MR-73، ساختاری تمام فولادی دارند. لوله این رولورها، با روش چکش‌کاری سخت یا خانکوبی ساخته می‌شوند که این امر خود باعث افزایش دقت و طول عمر مفید این سلاح شده است. برای رولور، یک نوع تویی سفارشی کالیبر ۹ × ۱۹ لوگر در نظر گرفته شده است که عمداً به درد کارآموزان پلیس می‌خورد. عمل مونتاژ قطعات این سلاح تماماً با دست انجام می‌شود که نتیجه آن، افزایش کیفیت مونتاژ و دقت عمل فوق‌العاده این سلاح کمری بوده است. شکل (۴-۵) نشان دهنده‌ی سیگنال صوتی شلیک شده توسط

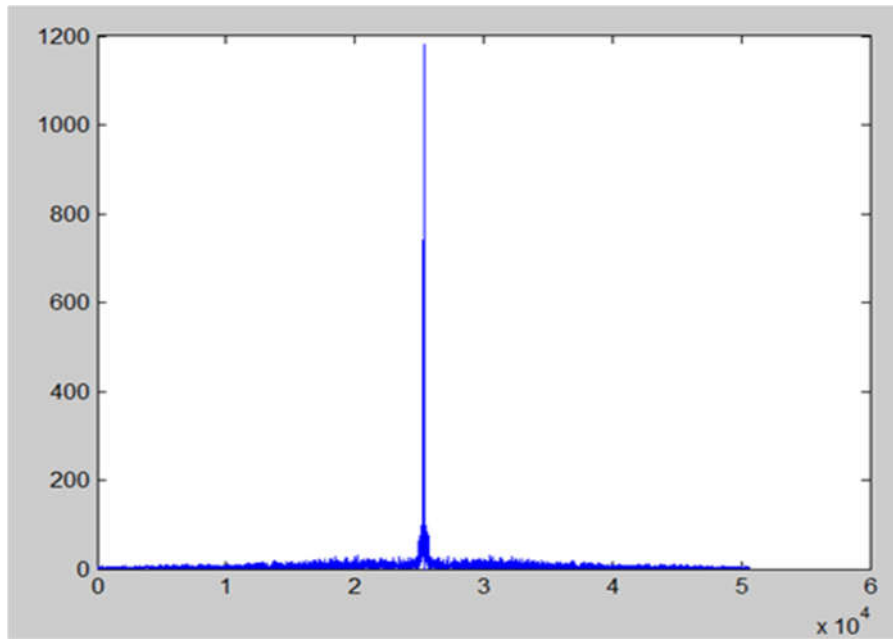
اسلحه رولور مانورھین MR-73 می باشد. از نقطه نظر کیفیت، می توان این رولور را چیزی در حد و حدود تیپانچه SIG P-210 دانست. از نظر مشخصات فنی مکانیزم ماشه این اسلحه دو زمانه، کالیبر آن ۳۷۵ مگنوم (۹ × ۱۹)، وزن ۹۱۰ گرم با لوله ۳ اینچی، ۱۰۵۰ گرم با لوله ۵ اینچی و طول: ۲۰۴ و ۲۶۴ میلی متر می باشد. طول لوله این اسلحه ۷۶ و ۱۳۳ میلی متر است که دارای ظرفیت توپی ۶ تیر می باشد [۴۷].



شکل ۴-۷ رولور مانورھین MR-73 [۴۷]



شکل ۴-۸ سیگنال صوتی رولور مانورھین MR-73 در حوزه زمان



شکل ۴-۹ سیگنال صوتی رولور مانور همین MR-73 در حوزه فرکانس

#### ۴-۲-۴ کلت برتا ام ۹ :

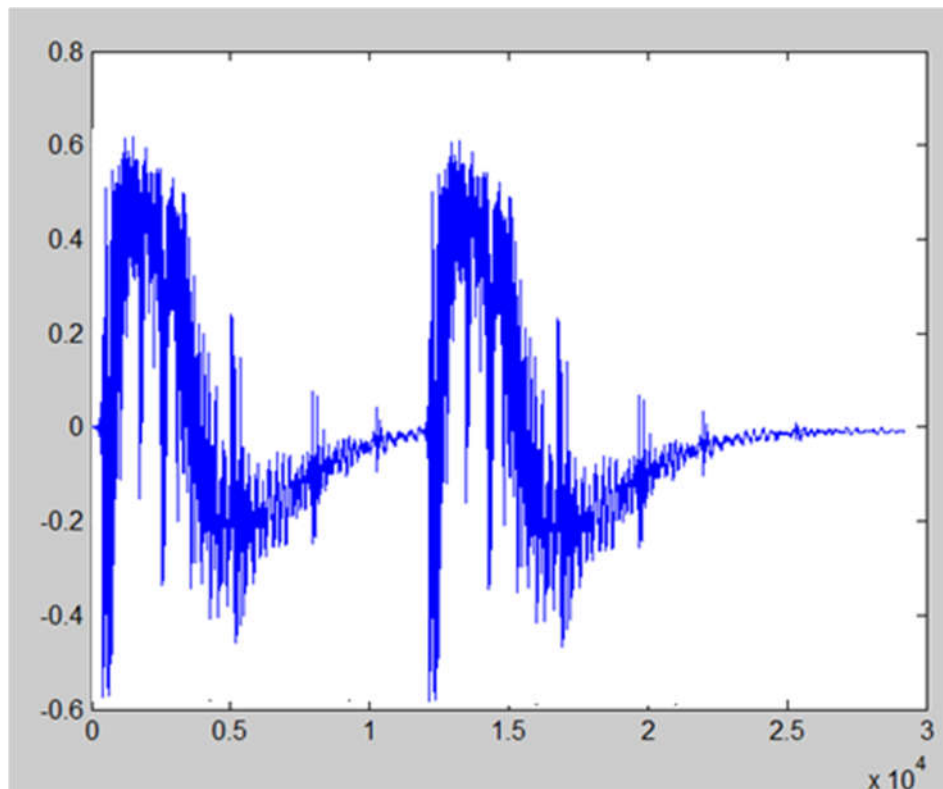
در این نوع سلاح برخلاف هفت تیر (رولور) که دارای چندین جان لوله است، تنها یک جان لوله وجود دارد و اسلحه پس از هر شلیک گلوله‌ای تازه را درون جان لوله جای می‌دهد. این سلاح از آن رو نیمه خودکار نامیده می‌شود که برای اولین شلیک، لازم است لوله کشویی به سمت عقب کشیده شود تا اولین گلوله مستقر شود؛ اما پس از آن با شلیک هر گلوله، پوکه به صورت خودکار خارج و بر اثر فشار گاز باروت گلوله‌ای تازه از خشاب درون جان لوله جای می‌گیرد.



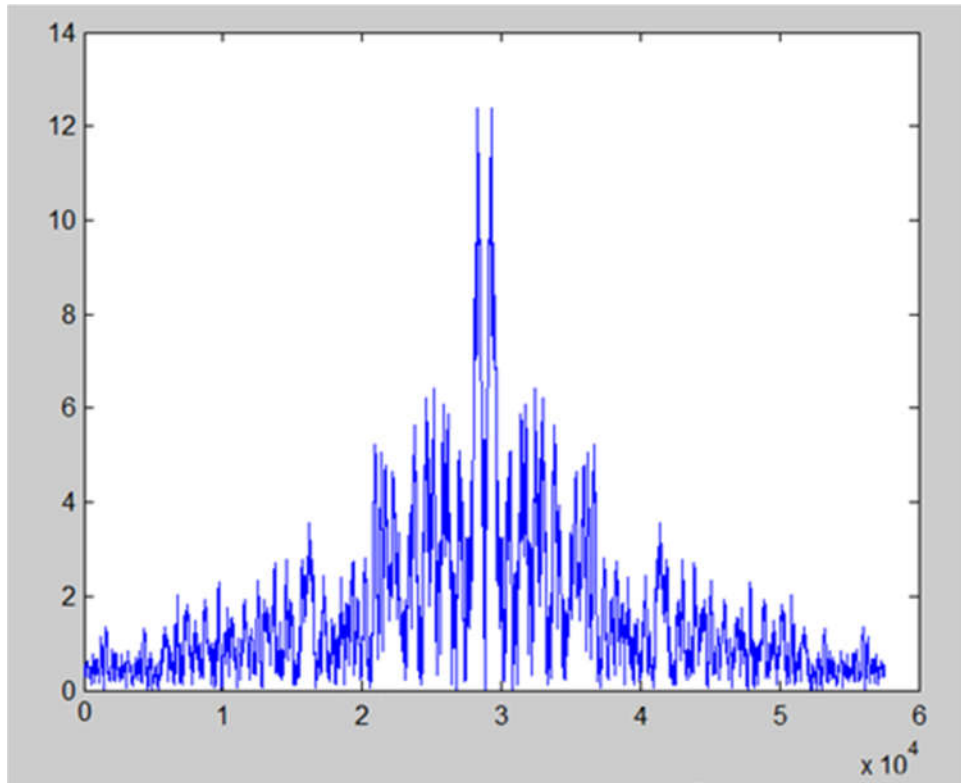
شکل ۴-۱۰ کلت برتا ام ۹ [۴۹]



برتا ام ۹ سلاح کمبری با کالیبر ۹ میلی متری از شرکت ایتالیایی برتا است که به سفارش نیروهای نظامی ایالات متحده آمریکا طراحی شده و از سال ۱۹۸۵ در خدمت ارتش آمریکا قرار گرفته است. در شکل (۴-۸) نمونه‌ای از سیگنال صوتی شلیک شده توسط کلت برتا ام ۹ را مشاهده می‌کنیم. برتا ام ۹ به‌عنوان سلاح کمبری سازمانی ارتش آمریکا جایگزین کلت ام ۱۹۱۱ شده است. تپانچه‌ام ۹ یکی از انواع برتا ۹۲ است. این اسلحه فشنگ‌های ۹×۱۹ میلی متری پارابلوم (فشنگ استاندارد تپانچه و مسلسل دستی ناتو) را شلیک می‌کند. خشاب استاندارد آن ظرفیت ۱۵ فشنگ را دارد و وزن خالی آن ۹۵۲ گرم و وزن پر ۱۱۶۲ گرم است. طول تفنگ هم ۲۱.۷ سانتی متر است که ۱۲.۵ سانتی متر طول لوله است. سرعت گلوله در هنگام خروج از دهانه این سلاح ۳۸۱ متر بر ثانیه است [۴۸]، [۴۹].



شکل ۴-۱۱ سیگنال صوتی کلت برتا ام ۹ در حوزه زمان



شکل ۴-۱۲ سیگنال صوتی کلت برتا ام ۹ در حوزه فرکانس

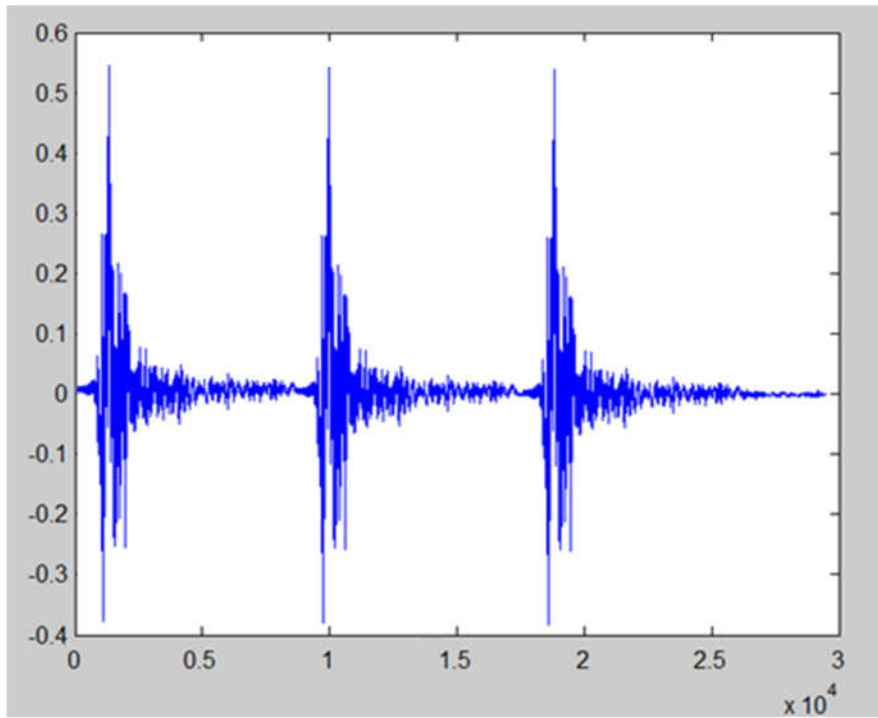
#### ۴-۲-۵ کلت دزرت ایگل:

این سلاح کمربندی علاوه بر داشتن قدرتی فوق‌العاده، از مکانیزم منحصر به فردی بهره می‌گیرد. عمده سلاح‌های کمربندی، از گلوله‌های مورت کوتاه استفاده می‌کنند. در حالی که عملکرد گلوله‌های این سلاح، توسط فشار گاز باروت صورت می‌گیرد. در واقع مکانیزم گلوله‌های این سلاح، از تفنگ M-16 اقتباس شده است. مزیت این نوع گلوله‌ها در این است که می‌توان از فشنگ‌های بسیار قدرتمندی استفاده کرد که قبلاً تنها توسط رولورها قابل شلیک بودند.

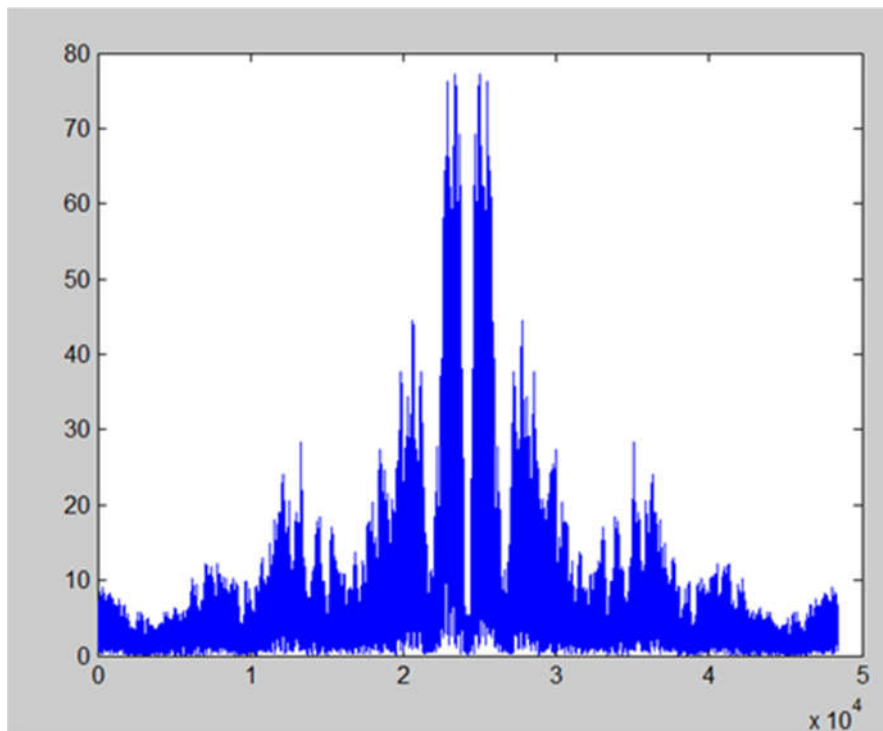


شکل ۴-۱۳ کلت دزرت ایگل (Desert Eagle) [۵۰]

به علت استفاده از گلوله‌های حجیم و پر قدرت در این سلاح، ظرفیت خشاب آن کمی کاهش پیدا کرده است. ظرفیت خشاب این سلاح در کالیبر ۰/۳۷۵ مگنوم ۹ تیر، در کالیبر ۰/۴۴ مگنوم ۸ تیر و در کالیبر AE ۰/۵ معادل ۷ تیر می‌باشد. برای مقایسه، بد نیست بدانید ظرفیت خشاب سلاح کمری SIG-Sauer P226، معادل ۱۶ تیر است [۵۰]. این اسلحه دارای سیستم مکانیکی فشار گاز CO2 با طول سلاح ۲۸۰ میلی‌متر و همچنین سرعت اولیه گلوله ۱۲۰ متر در ثانیه است. وزن سلاح ۱۱۰۰ گرم، طول لوله ۱۴۴ میلی‌متر، وزن کپسول گاز ۱۲ گرم و تعداد خان ۱۲ عدد می‌باشد.



شکل ۴-۱۴ سیگنال صوتی کلت دزرت ایگل در حوزه زمان



شکل ۴-۱۵ سیگنال صوتی کلت دزرت ایگل در حوزه فرکانس

## ۴-۲-۶ رولور اس اس آ ۱۸۷۳:

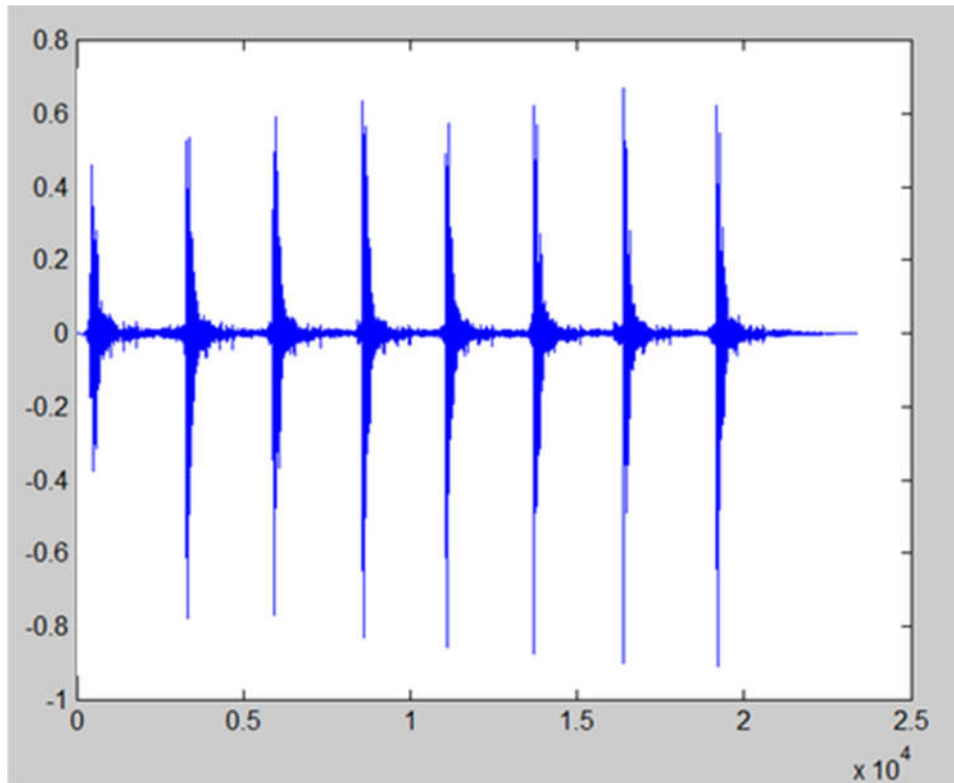
کلت تکنواخت ارتش، معروف به ( کلت مصلح (peacemaker) با نام اصلی (Colt single-action) Army رولوری بود که توسط کارخانه کلت برای سواره نظام ارتش آمریکا طراحی و در سال ۱۸۷۳ میلادی عرضه شد. این سلاح پرتعدادترین و رایج‌ترین رولور در غرب وحشی شد. نوع آن با لوله ۷/۵ اینچی برای یگان سواره نظام ساخته شده بود ولی در بین گاوچران‌ها با لوله ۵/۵ و ۴ اینچی رایج بود. این سلاح توسط یگان توپخانه ارتش آن کشور نیز اتخاذ شد.



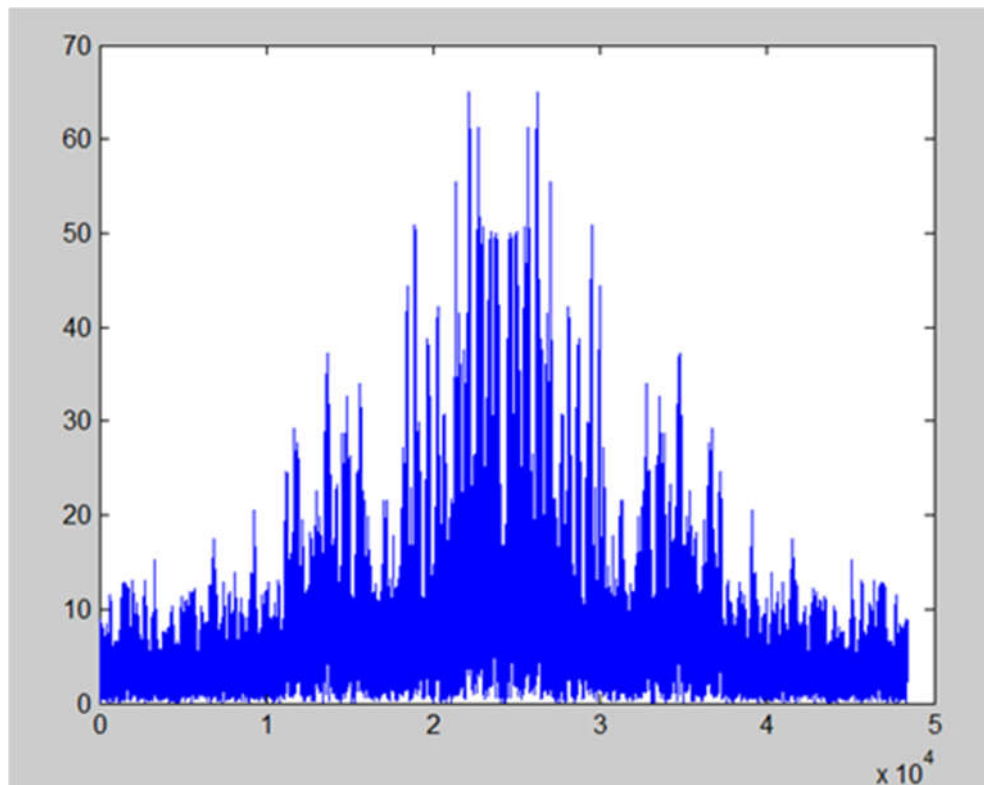
شکل ۴-۱۶ رولور اس اس آ ۱۸۷۳ [۵۱]

این سلاح هنوز توسط کارخانه‌های مشهوری ساخته می‌شود و مخصوصاً به دلیل دقت بیشتر نسبت به رولورهای خودکار برای اردو زدن در مناطق جنگلی و شکار و تیراندازی هدفی مفیدتر است. نحوه‌ی عملکرد این اسلحه تکنواخت است. به این صورت که برای شلیک باید چکش آن را مسلح کرد و سپس ماشه را چکاند و برای شلیک بعدی نیز چنین کرد. در شکل (۴-۱۲) نمونه‌ای از سیگنال شلیک شده توسط اسلحه رولور SAA 1873 نشان داده شده است. این سلاح ۶ تیر ظرفیت دارد ولی چون چکش آن در اثر ضربه شدید به احتمال زیاد شلیک می‌کند یکی از مخزن‌ها را خالی می‌گذارند. ظرفیت آن ۶ تیر و کالیبر آن ۰/۴۵ لانگ کلت است که نباید با ۰/۴۵ خودکار اشتباه شود. سیلندر آن برخلاف رولورهای دیگر ثابت است که به استحکام آن کمک می‌کند ولی باید تک‌تک در آن فشنگ گذاشت و

هنگام خارج کردن پوکه هم باید یکی یکی آن‌ها را درآورد. نهایتاً ۵۰ درصد از داده‌های موجود در پایگاه داده را به‌عنوان داده‌های آموزش و ۵۰ درصد از داده‌ها را به‌عنوان، داده‌های آزمایش استفاده می‌کنیم. در ادامه نتایج حاصل با استفاده از سیستم پیشنهادی را بر روی داده‌های آزمایش بیان خواهیم کرد [۵۱].



شکل ۴-۱۷ سیگنال صوتی رولور اس اس آ ۱۸۷۳ در حوزه زمان



شکل ۱۸-۴ سیگنال صوتی رولور اس اس ۱۸۷۳ در حوزه فرکانس

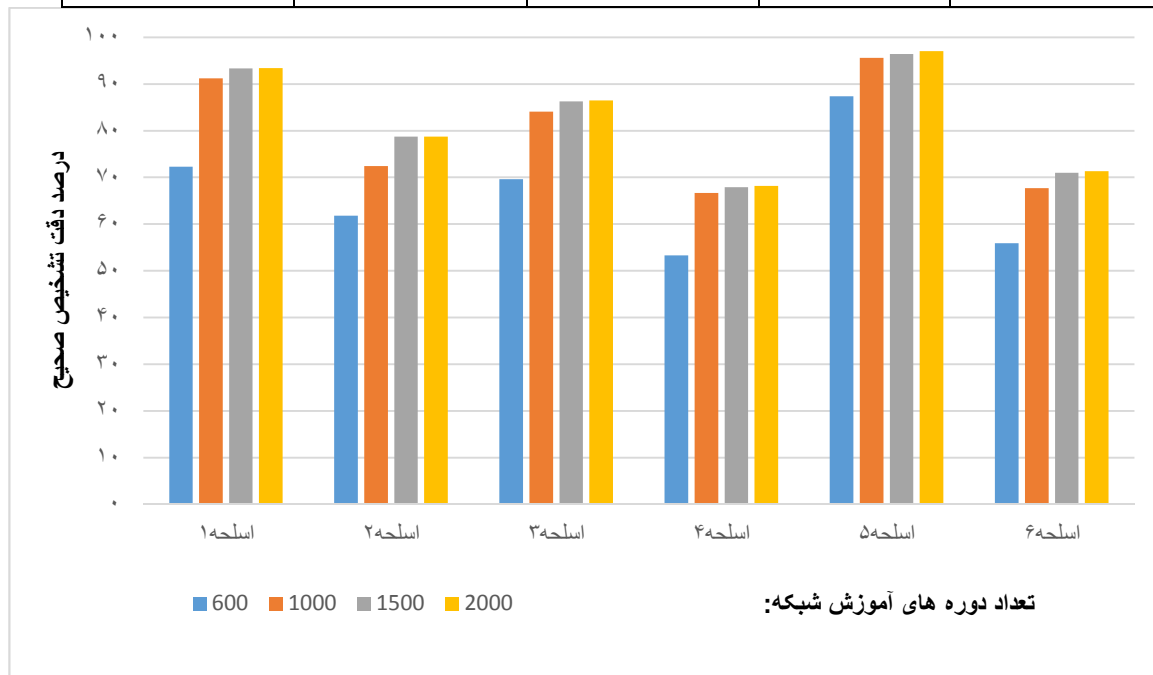
### ۳-۴ استخراج ویژگی از پایگاه داده:

#### ۱-۳-۴ ضرایب کپسترال در مقیاس مل :

با استفاده از استخراج ویژگی به وسیله MFCC از داده‌های آموزش با ۳۹ ضریب، شبکه عصبی را با تعداد مختلف از دوره‌های آموزش، آموزش دادیم و پس از آن با شبکه آموزش دیده با ارزیابی داده‌های آزمایش نتایج به شرح جدول ۱-۴ حاصل شده است. نمودار (۱-۴) مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش MFCC را نشان می‌دهد.

جدول ۴-۱: نتایج حاصل از ارزیابی طبقه با استفاده از ویژگی MFCC

درصد دقت تشخیص صحیح				تعداد دوره‌های آموزش شبکه
۲۰۰۰	۱۵۰۰	۱۰۰۰	۶۰۰	
۹۳/۴	۹۳/۳	۹۱/۲	۷۲/۳	اسلحه ۱
۷۸/۷	۷۸/۷	۷۲/۴	۶۱/۸	اسلحه ۲
۸۶/۵	۸۶/۳	۸۴/۱	۶۹/۶	اسلحه ۳
۶۸/۲	۶۷/۹	۶۶/۷	۵۳/۳	اسلحه ۴
۹۷	۹۶/۴	۹۵/۶	۸۷/۴	اسلحه ۵
۷۱/۳	۷۱/۲	۶۷/۷	۵۵/۹	اسلحه ۶



نمودار ۴-۱ مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش MFCC

با توجه به تعداد دوره‌های آموزش شبکه و دقت تشخیص صحیح اسلحه‌ها، نتایج استفاده از ویژگی

MFCC در جدول (۴-۱) نشان داده شده است. همان‌طور که در جدول (۴-۱) مشاهده می‌شود



- با استفاده از ۶۰۰ دوره آموزش شبکه اسلحه ۵ بالاترین دقت تشخیص صحیح را دارا می‌باشد و پس از آن اسلحه ۳ بیشترین دقت تشخیص صحیح را نسبت به بقیه دارد.
- با استفاده از ۱۰۰۰ دوره آموزش شبکه اسلحه ۵ بالاترین دقت تشخیص صحیح را دارا می‌باشد و پس از آن اسلحه ۱ بیشترین دقت تشخیص صحیح را نسبت به بقیه دارد.
- همچنین نتایج استفاده از ۱۵۰۰ و ۲۰۰۰ دوره آموزش شبکه نیز همانند ۱۰۰۰ دوره آموزش شبکه می‌باشد.
- با توجه به جدول (۴-۱) می‌توان نتیجه گرفت در تعداد دوره‌های آموزش از ۶۰۰ تا ۱۰۰۰ دقت تشخیص صحیح افزایش یافته است ولی از ۱۰۰۰ تا ۲۰۰۰ دوره آموزش شبکه تفاوت چندانی در نتایج دیده نمی‌شود.

با مشاهده نمودار (۴-۱) اسلحه ۵ از بالاترین دقت تشخیص صحیح در تمامی دوره‌های آموزش شبکه برخوردار است و پایین‌ترین دقت تشخیص صحیح متعلق به اسلحه ۴ می‌باشد.

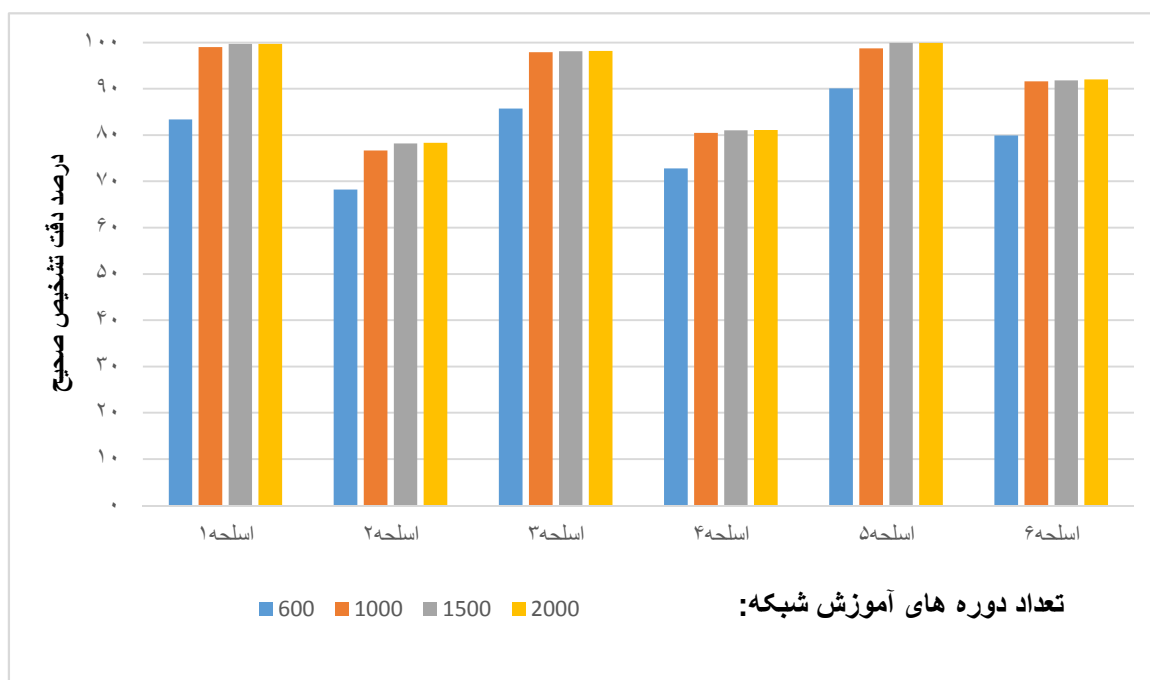
با توجه به مقایسه دقت طبقه بند براساس تعداد دوره‌های آموزش شبکه با استفاده از روش MFCC به ترتیب از بیشترین تا کمترین دقت تشخیص صحیح متعلق به اسلحه‌های ۵ و ۱ و ۳ و ۲ و ۴ و ۶ می‌باشد.

#### ۴-۳-۲ ضرایب کپسترال حقیقی :

با استفاده از استخراج ویژگی به وسیله کپستروم حقیقی از داده‌های آموزش با ۲۰ ضریب، شبکه عصبی را با تعداد مختلف از دوره‌های آموزش، آموزش دادیم و پس از آن با شبکه آموزش دیده با ارزیابی داده‌های آزمایش نتایج به شرح جدول ۴-۲ حاصل شده است. نمودار (۴-۲) مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش Real Cepstrum را نشان می‌دهد.

جدول ۲-۴: نتایج حاصل از ارزیابی طبقه با استفاده از ویژگی Real Cepstrum

درصد دقت تشخیص صحیح				تعداد دوره‌های آموزش شبکه
۶۰۰	۱۰۰۰	۱۵۰۰	۲۰۰۰	
۸۳/۴	۹۹/۰۱	۹۹/۷	۹۹/۷	اسلحه ۱
۶۸/۲	۷۶/۷	۷۸/۲	۷۸/۳	اسلحه ۲
۸۵/۷	۹۷/۹	۹۸/۱	۹۸/۲	اسلحه ۳
۷۲/۸	۸۰/۵	۸۱	۸۱/۱	اسلحه ۴
۹۰/۱	۹۸/۷	۹۹/۹	۹۹/۹	اسلحه ۵
۷۹/۹	۹۱/۶	۹۱/۸	۹۲	اسلحه ۶



نمودار ۲-۴ مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش Real Cepstrum

با توجه به تعداد دوره‌های آموزش شبکه و دقت تشخیص صحیح اسلحه‌ها، نتایج استفاده از ویژگی Real Cepstrum در جدول (۲-۴) نشان داده شده است. همانطور که در جدول (۲-۴) مشاهده می‌شود

- با استفاده از ۶۰۰ دوره آموزش شبکه اسلحه ۵ بالاترین دقت تشخیص صحیح را دارا می‌باشد و پس از آن اسلحه ۳ بیشترین دقت تشخیص صحیح را نسبت به بقیه دارد.
- با استفاده از ۱۰۰۰ دوره آموزش شبکه اسلحه ۱ بالاترین دقت تشخیص صحیح را دارا می‌باشد و پس از آن اسلحه ۵ بیشترین دقت تشخیص صحیح را نسبت به بقیه دارد.
- با استفاده از ۱۵۰۰ دوره آموزش شبکه اسلحه ۵ بالاترین دقت تشخیص صحیح را دارا می‌باشد و پس از آن اسلحه ۱ بیشترین دقت تشخیص صحیح را نسبت به بقیه دارد.
- همچنین نتایج استفاده از ۲۰۰۰ دوره آموزش شبکه نیز همانند ۱۵۰۰ دوره آموزش شبکه می‌باشد.
- با توجه به جدول (۲-۴) می‌توان نتیجه گرفت در تعداد دوره‌های آموزش از ۶۰۰ تا ۱۰۰۰ دقت تشخیص صحیح افزایش یافته است ولی از ۱۰۰۰ تا ۲۰۰۰ دوره آموزش شبکه تفاوت چندانی در نتایج دیده نمی‌شود.

با مشاهده نمودار (۲-۴) اسلحه ۵ از بالاترین دقت تشخیص صحیح در دوره‌های ۶۰۰، ۱۵۰۰ و ۲۰۰۰ آموزش شبکه برخوردار است و اسلحه ۱ بالاترین دقت تشخیص صحیح را با ۱۰۰۰ دوره آموزش شبکه دارد. پایین‌ترین دقت تشخیص صحیح متعلق به اسلحه ۳ می‌باشد.

- با توجه به مقایسه دقت طبقه بند براساس تعداد دوره‌های آموزش شبکه با استفاده از روش

Real Cepstrum به ترتیب از بیشترین تا کمترین دقت تشخیص صحیح متعلق به اسلحه‌های ۵ و ۱ و ۳ و ۶ و ۴ و ۳ می‌باشد.

#### ۴-۳-۳ ضرایب پیش‌بینی خطی :

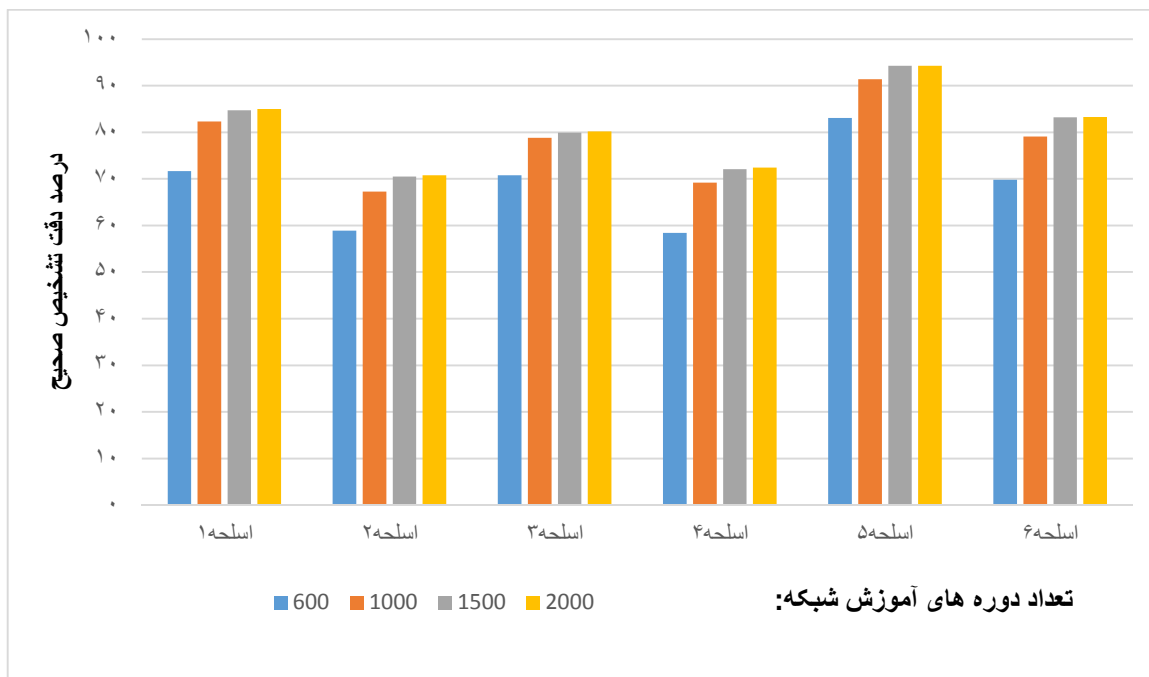
برای استخراج ویژگی از روش ضرایب پیش‌بینی خطی با ۳ ضریب ۱۰ و ۲۰ و ۵۰ استفاده کرده تا بالاترین دقت تشخیص صحیح را بدست آوریم.

#### ۴-۳-۳-۱ ضرایب پیش‌بینی خطی (LPC 10):

با استفاده از استخراج ویژگی به‌وسیله روش ضرایب پیش‌بینی خطی و استفاده از ۱۰ ضریب برای پیشگویی شبکه عصبی را با تعداد مختلف از دوره‌های آموزش، آموزش دادیم و پس‌از آن با شبکه آموزش‌دیده با ارزیابی داده‌های آزمایش نتایج به شرح جدول ۳-۴ حاصل شده است. نمودار (۳-۴) مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش LPC با ضریب ۱۰ را نشان می‌دهد.

جدول ۳-۴: نتایج حاصل از ارزیابی طبقه با استفاده از ویژگی LPC10

درصد دقت تشخیص صحیح				
۲۰۰۰	۱۵۰۰	۱۰۰۰	۶۰۰	تعداد دوره‌های آموزش شبکه
۸۵	۸۴/۷	۸۲/۳	۷۱/۷	اسلحه ۱
۷۰/۸	۷۰/۵	۶۷/۳	۵۸/۹	اسلحه ۲
۸۰/۲	۷۹/۹	۷۸/۸	۷۰/۸	اسلحه ۳
۷۲/۴	۷۲/۱	۶۹/۲	۵۸/۴	اسلحه ۴
۹۴/۳	۹۴/۳	۹۱/۴	۸۳/۱	اسلحه ۵
۸۳/۱	۸۳/۲	۷۹/۱	۶۹/۸	اسلحه ۶



نمودار ۳-۴ مقایسه دقت طبقه بند بر اساس تعداد دوره های آموزش با روش LPC 10

با توجه به تعداد دوره های آموزش شبکه و دقت تشخیص صحیح اسلحه ها، نتایج استفاده از ویژگی LPC با ضریب ۱۰ در جدول (۳-۴) نشان داده شده است. همانطور که در جدول (۳-۴) مشاهده می شود

- با استفاده از ۶۰۰ دوره آموزش شبکه اسلحه ۵ بالاترین دقت تشخیص صحیح را دارا می باشد و پس از آن اسلحه ۱ بیشترین دقت تشخیص صحیح را نسبت به بقیه دارد.
- همچنین نتایج استفاده از دوره های آموزش شبکه ۱۰۰۰، ۱۵۰۰ و ۲۰۰۰ نیز همانند ۶۰۰ دوره آموزش شبکه می باشد.
- با توجه به جدول (۳-۴) می توان نتیجه گرفت در تعداد دوره های آموزش از ۶۰۰ تا ۱۰۰۰ دقت تشخیص صحیح افزایش یافته است ولی از ۱۰۰۰ تا ۲۰۰۰ دوره آموزش شبکه تفاوت چندانی در نتایج دیده نمی شود.
- دقت تشخیص صحیح اسلحه های ۳ و ۶ با ۱۰۰۰ دوره آموزش شبکه بسیار نزدیک است.

با مشاهده نمودار (۳-۴) اسلحه ۵ از بالاترین دقت تشخیص صحیح در تمامی دوره‌های آموزش شبکه برخوردار است. پایین‌ترین دقت تشخیص صحیح متعلق به اسلحه ۲ می‌باشد.

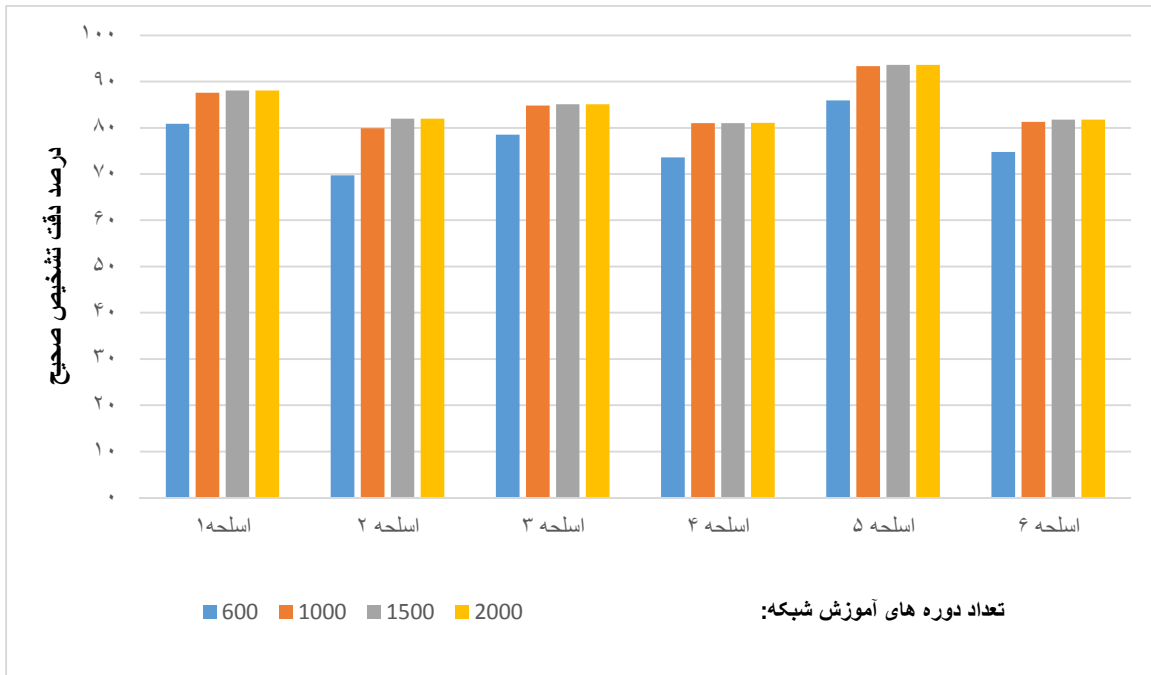
- با توجه به مقایسه دقت طبقه بند براساس تعداد دوره‌های آموزش شبکه با استفاده از روش LPC با ضریب ۱۰ به ترتیب از بیشترین تا کمترین دقت تشخیص صحیح متعلق به اسلحه‌های ۵ و ۱ و ۶ و ۳ و ۴ و ۲ می‌باشد.

#### ۴-۳-۳-۲ ضرایب پیش‌بینی خطی (LPC 20):

با استفاده از استخراج ویژگی به‌وسیله روش ضرایب پیشگویی خطی و استفاده از ۲۰ ضریب برای پیشگویی شبکه عصبی را با تعداد مختلف از دوره‌های آموزش، آموزش دادیم و پس‌از آن با شبکه آموزش‌دیده با ارزیابی داده‌های آزمایش نتایج به شرح جدول ۴-۴ حاصل شده است. نمودار (۴-۴) مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش LPC با ضریب ۲۰ را نشان می‌دهد.

جدول ۴-۴: نتایج حاصل از ارزیابی طبقه با استفاده از ویژگی LPC20

درصد دقت تشخیص صحیح				
۲۰۰۰	۱۵۰۰	۱۰۰۰	۶۰۰	تعداد دوره‌های آموزش شبکه
۸۸/۱	۸۸/۱	۸۷/۶	۸۰/۹	اسلحه ۱
۸۲	۸۲	۷۹/۹	۶۹/۷	اسلحه ۲
۸۵/۲	۸۵/۱	۸۴/۸	۷۸/۵	اسلحه ۳
۸۱/۱	۸۱	۸۱	۷۳/۶	اسلحه ۴
۹۳/۶	۹۳/۶	۹۳/۳	۸۵/۹	اسلحه ۵
۸۱/۸	۸۱/۸	۸۱/۳	۷۴/۸	اسلحه ۶



نمودار ۴-۴ مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش LPC 20

با توجه به تعداد دوره‌های آموزش شبکه و دقت تشخیص صحیح اسلحه‌ها، نتایج استفاده از ویژگی

LPC با ضریب ۲۰ در جدول (۴-۴) نشان داده شده است. همانطور که در جدول (۴-۴)

مشاهده می‌شود.

- با استفاده از ۶۰۰ دوره آموزش شبکه اسلحه ۵ بالاترین دقت تشخیص صحیح را دارا می‌باشد و پس از آن اسلحه ۱ بیشترین دقت تشخیص صحیح را نسبت به بقیه دارد.
- دقت تشخیص صحیح اسلحه‌های ۳ و ۶ با ۶۰۰ دوره آموزش شبکه بسیار نزدیک به هم است.
- با استفاده از ۱۰۰۰ دوره آموزش شبکه اسلحه ۱ بالاترین دقت تشخیص صحیح را دارا می‌باشد و پس از آن اسلحه ۵ بیشترین دقت تشخیص صحیح را نسبت به بقیه دارد.
- دقت تشخیص صحیح اسلحه‌های ۲ و ۴ با ۱۰۰۰ دوره آموزش شبکه بسیار نزدیک به هم است.

- همچنین نتایج استفاده از دوره‌های آموزش شبکه ۱۵۰۰ و ۲۰۰۰ نیز همانند ۱۰۰۰ دوره آموزش شبکه می‌باشد.

- با توجه به جدول (۴-۴) می‌توان نتیجه گرفت در تعداد دوره‌های آموزش از ۶۰۰ تا ۱۰۰۰ دقت تشخیص صحیح افزایش یافته است ولی از ۱۰۰۰ تا ۲۰۰۰ دوره آموزش شبکه تفاوت چندانی در نتایج دیده نمی‌شود.

با مشاهده نمودار (۴-۴) اسلحه ۵ از بالاترین دقت تشخیص صحیح در تمامی دوره‌های آموزش

شبکه برخوردار است. پایین‌ترین دقت تشخیص صحیح متعلق به اسلحه ۲ می‌باشد.

- با توجه به مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش شبکه با استفاده از روش

LPC با ضریب ۲۰ به ترتیب از بیشترین تا کمترین دقت تشخیص صحیح متعلق به اسلحه‌ی ۵

و ۱ و ۳ و ۶ و ۴ و ۲ می‌باشد.

#### ۴-۳-۳-۳ ضرایب پیش‌بینی خطی (LPC 50):

با استفاده از استخراج ویژگی به‌وسیله روش ضرایب پیش‌بینی خطی و استفاده از ۱۰ ضریب برای

پیش‌بینی شبکه عصبی را با تعداد مختلف از دوره‌های آموزش، آموزش دادیم و پس‌از آن با شبکه

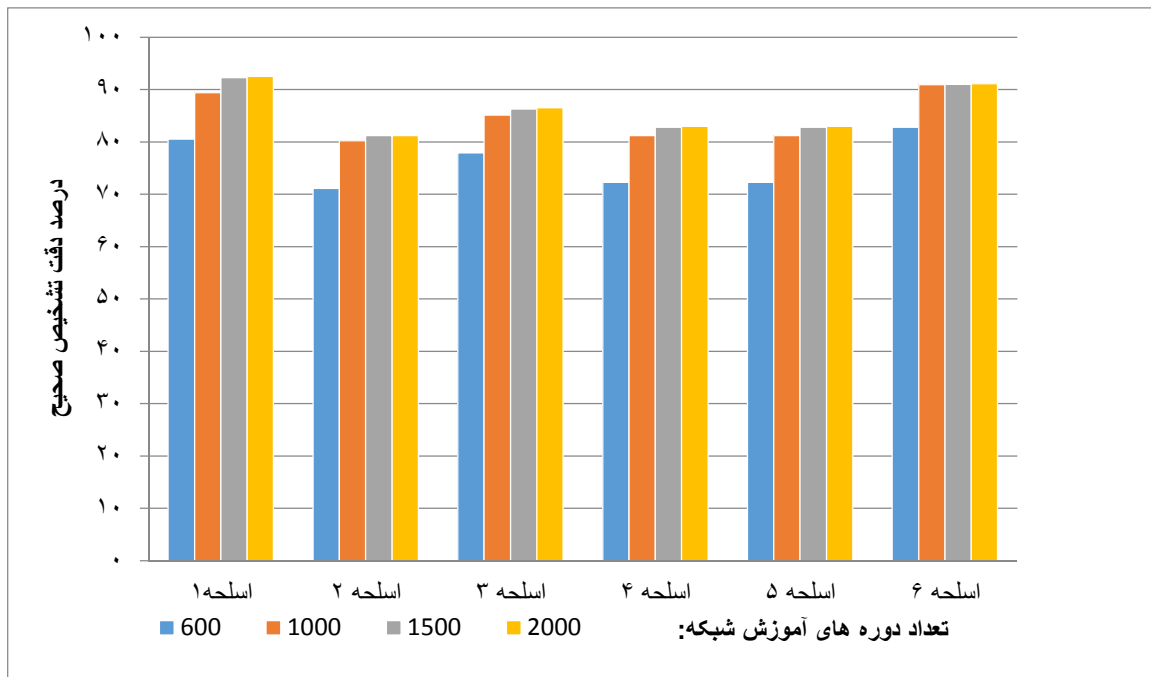
آموزش‌دیده با ارزیابی داده‌های آزمایش نتایج به شرح جدول ۴-۵ حاصل شده است. نمودار (۴-۵)

مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش LPC با ضریب ۵۰ را نشان می‌دهد.



جدول ۴-۵: نتایج حاصل از ارزیابی طبقه با استفاده از ویژگی LPC50

درصد دقت تشخیص صحیح				
۲۰۰۰	۱۵۰۰	۱۰۰۰	۶۰۰	تعداد دوره‌های آموزش شبکه
۹۲/۵	۹۲/۳	۸۹/۴	۸۰/۵	اسلحه ۱
۸۱/۲	۸۱/۲	۸۰/۲	۷۱/۱	اسلحه ۲
۸۶/۵	۸۶/۳	۸۵/۱	۷۷/۹	اسلحه ۳
۸۳	۸۲/۸	۸۱/۲	۷۲/۳	اسلحه ۴
۹۷/۵	۹۷/۴	۹۶/۷	۸۲/۶	اسلحه ۵
۹۱/۱	۹۱	۹۰/۹	۸۲/۸	اسلحه ۶



نمودار ۴-۵ مقایسه دقت طبقه بند بر اساس تعداد دوره‌های آموزش با روش LPC50

با توجه به تعداد دوره‌های آموزش شبکه و دقت تشخیص صحیح اسلحه‌ها، نتایج استفاده از ویژگی LPC با ضریب ۵۰ در جدول (۴-۵) نشان داده شده است. همانطور که در جدول (۴-۵) مشاهده می‌شود

- با استفاده از ۶۰۰ دوره آموزش شبکه اسلحه ۶ بالاترین دقت تشخیص صحیح را دارا می‌باشد و پس از آن اسلحه ۵ بیشترین دقت تشخیص صحیح را نسبت به بقیه دارد.
- دقت تشخیص صحیح اسلحه‌های ۱، ۵ و ۶ با ۶۰۰ دوره آموزش شبکه بسیار نزدیک به هم است.
- با استفاده از ۱۰۰۰ دوره آموزش شبکه اسلحه ۵ بالاترین دقت تشخیص صحیح را دارا می‌باشد و پس از آن اسلحه ۶ بیشترین دقت تشخیص صحیح را نسبت به بقیه دارد.
- دقت تشخیص صحیح اسلحه‌های ۱ و ۶ با ۱۰۰۰ دوره آموزش شبکه بسیار نزدیک به هم است.
- با استفاده از ۱۵۰۰ دوره آموزش شبکه اسلحه ۵ بالاترین دقت تشخیص صحیح را دارا می‌باشد و پس از آن اسلحه ۱ بیشترین دقت تشخیص صحیح را نسبت به بقیه دارد.
- همچنین نتایج استفاده از ۲۰۰۰ دوره آموزش شبکه نیز همانند ۱۵۰۰ دوره آموزش شبکه می‌باشد.
- با توجه به جدول (۴-۵) می‌توان نتیجه گرفت در تعداد دوره‌های آموزش از ۶۰۰ تا ۱۰۰۰ دقت تشخیص صحیح افزایش یافته است ولی از ۱۰۰۰ تا ۲۰۰۰ دوره آموزش شبکه تفاوت چندانی در نتایج دیده نمی‌شود.

با مشاهده نمودار (۴-۵) اسلحه ۵ از بالاترین دقت تشخیص صحیح در تمامی دوره‌های آموزش

شبکه برخوردار است. پایین‌ترین دقت تشخیص صحیح متعلق به اسلحه ۲ می‌باشد.

- با توجه به مقایسه دقت طبقه بند براساس تعداد دوره‌های آموزش شبکه با استفاده از روش LPC با ضریب ۵۰ به ترتیب از بیشترین تا کمترین دقت تشخیص صحیح متعلق به اسلحه-های ۵ و ۱ و ۶ و ۳ و ۴ و ۲ می‌باشد.

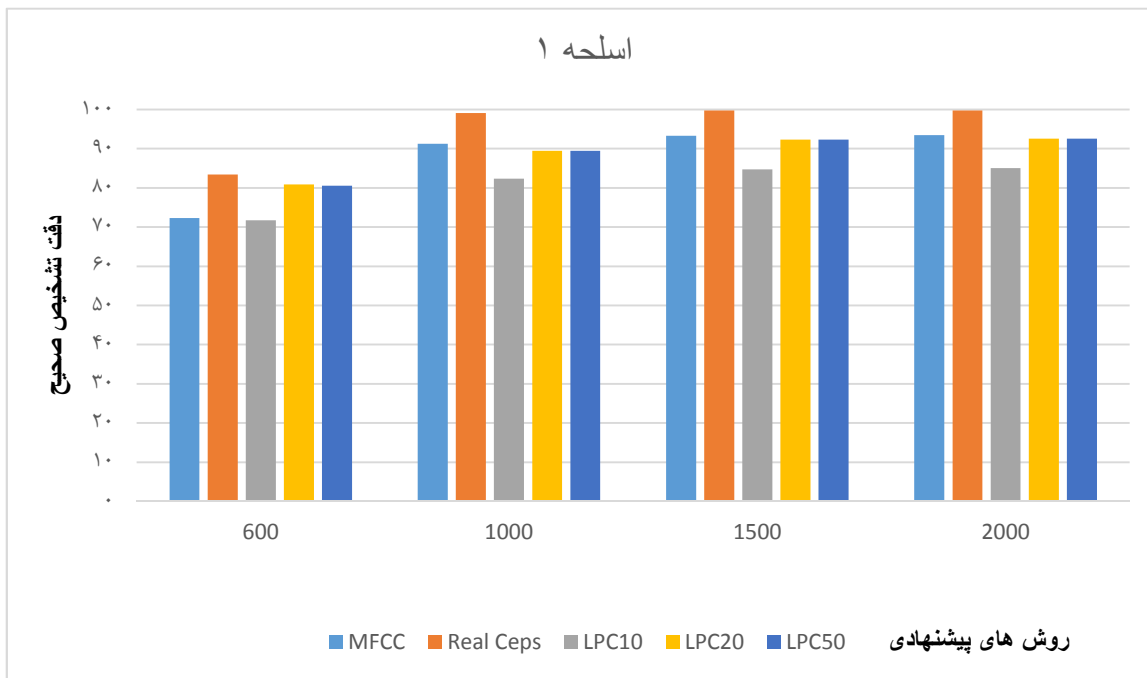
#### ۴-۴ تحلیل و نتیجه‌گیری:

برای تحلیل و نتیجه‌گیری از نتایج آزمایش‌ها، تمام نتایج را در قالب جدول ۵-۶ و بر اساس دقت هر روش برای هر اسلحه نشان داده‌ایم.

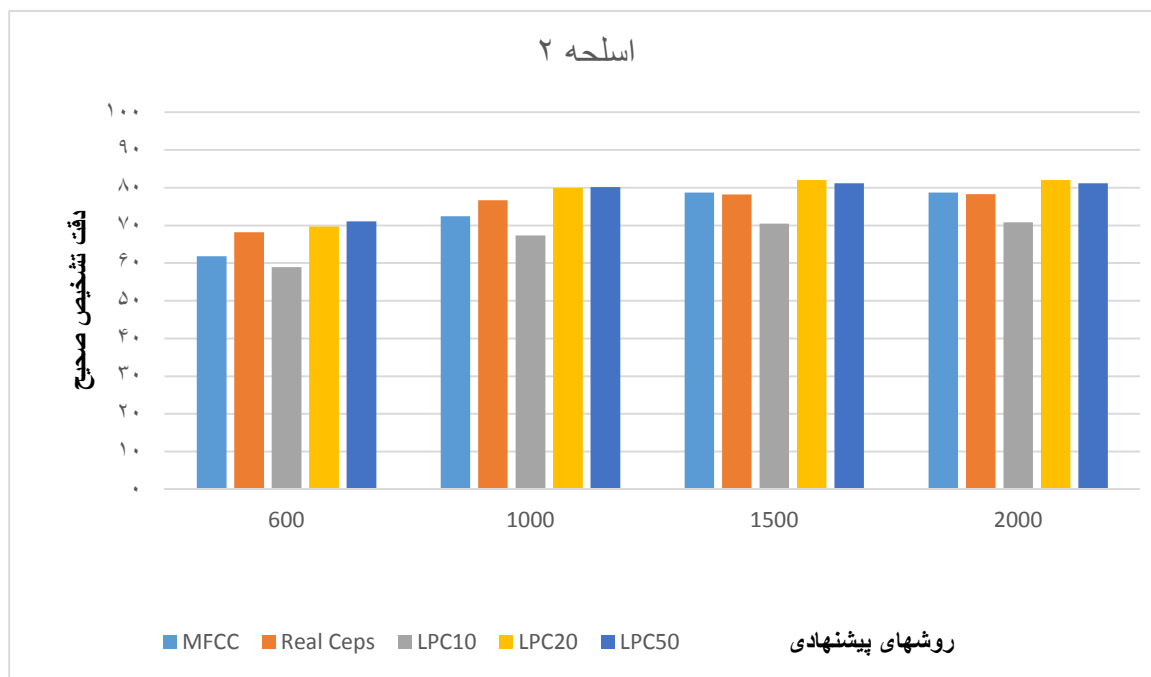
جدول ۴-۶: مقایسه دقت روش‌های پیشنهادی

تعداد دوره‌های آموزش				روش استخراج ویژگی	اسلحه
۲۰۰۰	۱۵۰۰	۱۰۰۰	۶۰۰		
درصد دقت تشخیص صحیح					
۹۳/۴	۹۳/۳	۹۱/۲	۷۲/۳	MFCC	اسلحه ۱
۹۹/۷	۹۹/۷	۹۹/۱	۸۳/۴	Real Ceps	اسلحه ۱
۸۵	۸۴/۷	۸۲/۳	۷۱/۷	LPC10	اسلحه ۱
۸۸/۱	۸۸/۱	۸۷/۶	۸۰/۹	LPC20	اسلحه ۱
۹۲/۵	۹۲/۳	۸۹/۴	۸۰/۵	LPC50	اسلحه ۱
۷۸/۷	۷۸/۷	۷۲/۴	۶۱/۸	MFCC	اسلحه ۲
۷۸/۳	۷۸/۲	۷۶/۷	۶۸/۲	Real Ceps	اسلحه ۲
۷۰/۸	۷۰/۵	۶۷/۳	۵۸/۹	LPC10	اسلحه ۲
۸۲	۸۲	۷۹/۹	۶۹/۷	LPC20	اسلحه ۲
۸۱/۲	۸۱/۲	۸۰/۲	۷۱/۱	LPC50	اسلحه ۲
۸۶/۵	۸۶/۳	۸۴/۱	۶۹/۶	MFCC	اسلحه ۳

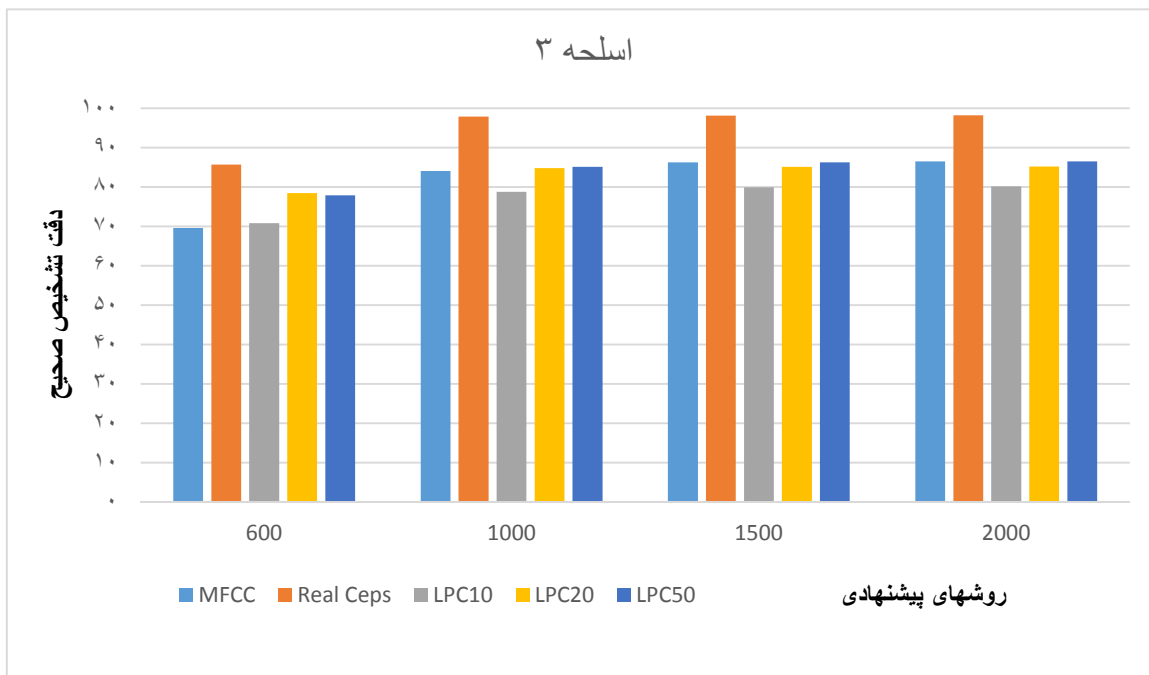
٩٨/٢	٩٨/١	٩٧/٩	٨٥/٧	Real Ceps	اسلحه ٣
٨٠/٢	٧٩/٩	٧٨/٨	٧٠/٨	LPC10	اسلحه ٣
٨٥/٢	٨٥/١	٨٤/٨	٧٨/٥	LPC20	اسلحه ٣
٨٤/٥	٨٤/٣	٨٥/١	٧٧/٩	LPC50	اسلحه ٣
٤٨/٢	٤٧/٩	٤٤/٧	٥٣/٣	MFCC	اسلحه ٤
٨١/١	٨١	٨٠/٥	٧٢/٨	Real Ceps	اسلحه ٤
٧٢/٤	٧٢/١	٤٩/٢	٥٨/٤	LPC10	اسلحه ٤
٨١/١	٨١	٨١/٠	٧٣/٤	LPC20	اسلحه ٤
٨٣	٨٢/٨	٨١/٢	٧٢/٣	LPC50	اسلحه ٤
٩٧	٩٤/٤	٩٥/٤	٨٧/٤	MFCC	اسلحه ٥
٩٩/٩	٩٩/٩	٩٨/٧	٩٠/١	Real Ceps	اسلحه ٥
٩٤/٣	٩٤/٣	٩١/٤	٨٣/١	LPC10	اسلحه ٥
٩٣/٤	٩٣/٤	٩٣/٣	٨٥/٩	LPC20	اسلحه ٥
٩٧/٥	٩٧/٤	٩٤/٧	٨٢/٤	LPC50	اسلحه ٥
٧١/٣	٧١/٢	٤٧/٧	٥٥/٩	MFCC	اسلحه ٦
٩٢	٩١/٨	٩١/٤	٧٩/٩	Real Ceps	اسلحه ٦
٨٣/٣	٨٣/٢	٧٩/١	٤٩/٨	LPC10	اسلحه ٦
٨١/٨	٨١/٨	٨١/٣	٧٤/٨	LPC20	اسلحه ٦
٩١/١	٩١	٩٠/٩	٨٢/٨	LPC50	اسلحه ٦



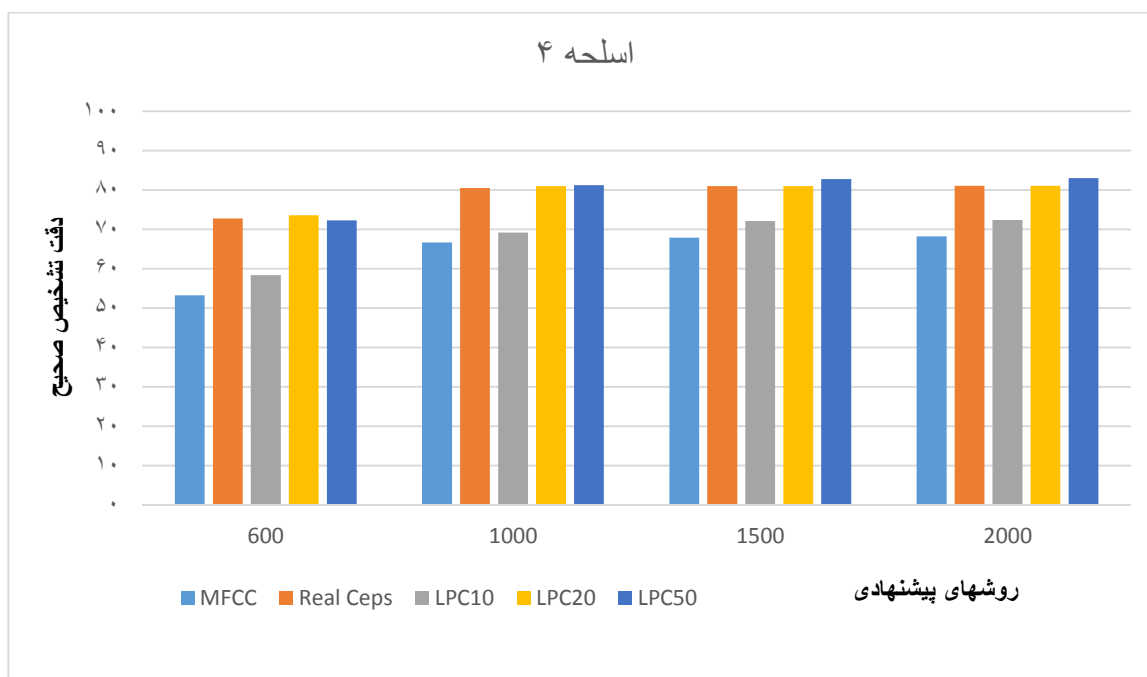
نمودار ۴-۶ مقایسه دقت روش های پیشنهادی برای شناسایی اسلحه ۱



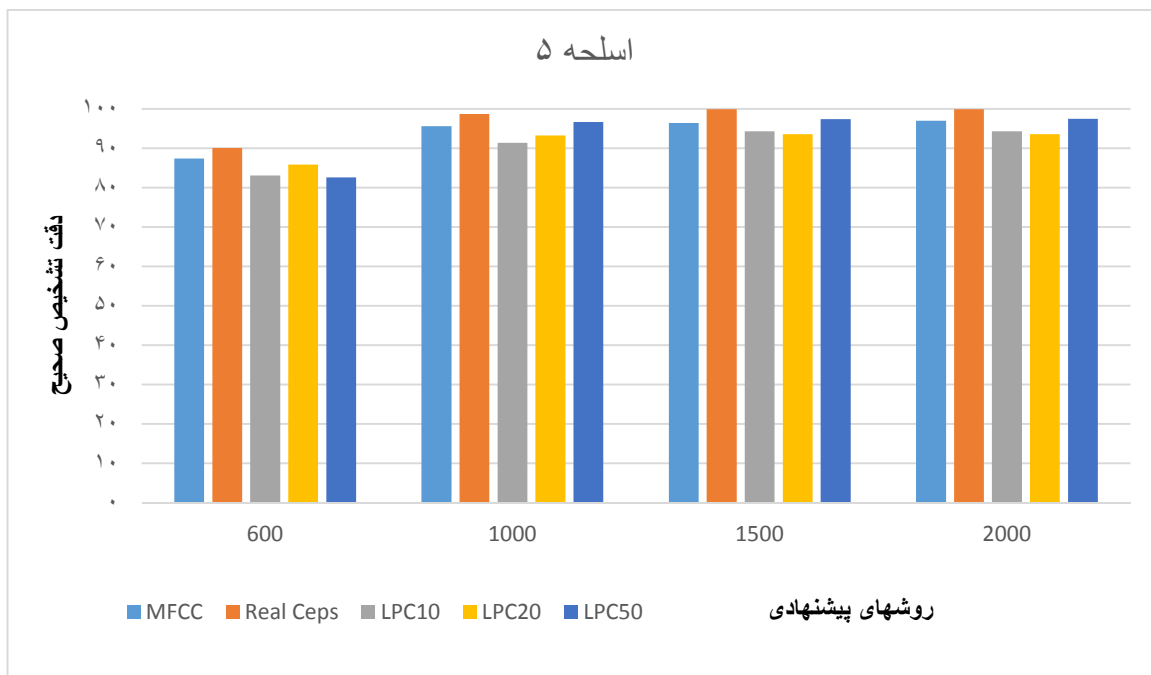
نمودار ۴-۷ مقایسه دقت روش های پیشنهادی برای شناسایی اسلحه ۲



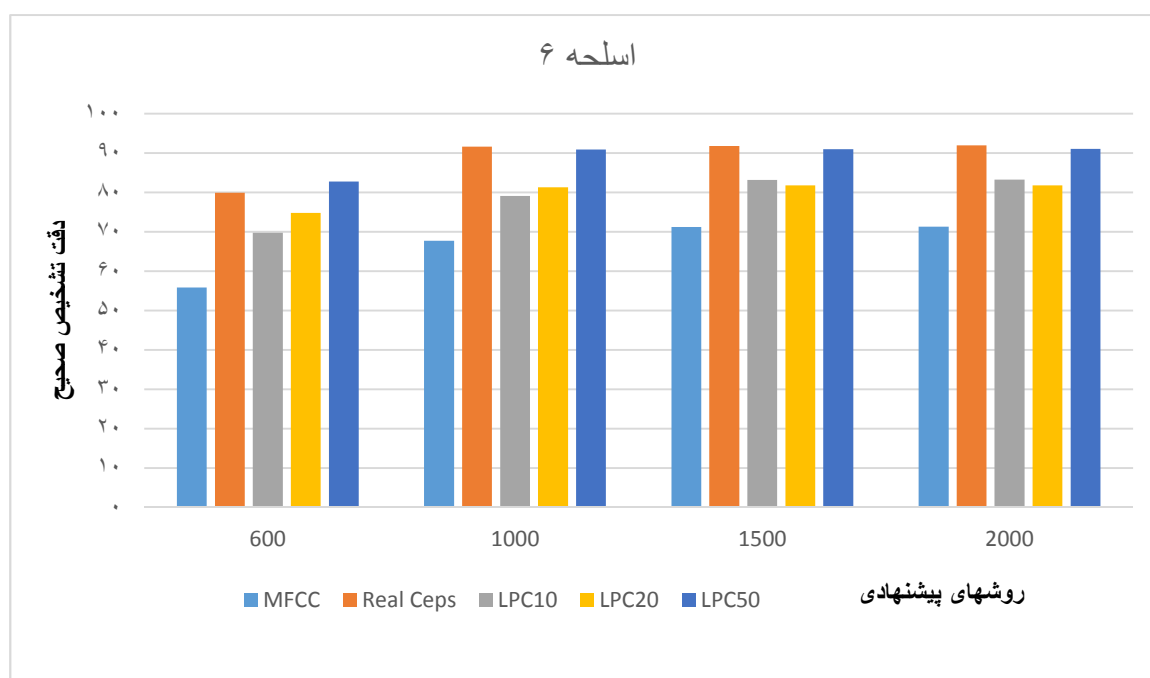
نمودار ۴-۸ مقایسه دقت روشهای پیشنهادی برای شناسایی اسلحه ۳



نمودار ۴-۹ مقایسه دقت روشهای پیشنهادی برای شناسایی اسلحه ۴



نمودار ۴-۱۰ مقایسه دقت روش های پیشنهادی برای شناسایی اسلحه ۵



نمودار ۴-۱۱ مقایسه دقت روش های پیشنهادی برای شناسایی اسلحه ۶

از بررسی نتایج مطرح شده می‌توان دریافت که برای اسلحه ۱ استفاده از ویژگی کپسترال حقیقی، برای اسلحه ۲ استفاده از ویژگی پیش‌بینی خطی با ضریب ۵۰، برای اسلحه ۳ استفاده از ویژگی کپسترال حقیقی، برای اسلحه ۴ استفاده از ویژگی پیش‌بینی خطی با ضریب ۵۰، برای اسلحه ۵ استفاده از ویژگی کپسترال حقیقی و برای اسلحه ۶ استفاده از ویژگی کپسترال حقیقی بهترین جواب را داده است؛ بنابراین نمی‌توان قاطعانه بیان کرد یک ویژگی خاص برای تشخیص اصوات ضربه‌ای کاربردی‌تر است.

همچنین برای طبقه بند نیز می‌توان دید افزایش تعداد دوره‌های آموزش از حدی به بعد تأثیر چندانی در بهبود نتایج ندارد و نسبت به بار محاسباتی تحمیلی افزایش دقت ارزش چندانی نخواهد داشت.

#### ۴-۵ جمع‌بندی:

در این فصل پایگاه داده مورد استفاده در این پایان‌نامه را معرفی کردیم و در ادامه بایان نتایج حاصل از اعمال روش پیشنهادی بر روی پایگاه داده مورد استفاده، به تحلیل و بررسی نتایج پرداختیم. در ادامه در فصل بعد جمع‌بندی و نتیجه حاصل از انجام این پایان‌نامه را بیان می‌کنیم و برای ادامه روند پژوهش در این حوزه ایده‌هایی را مطرح خواهیم کرد.



# فصل پنجم:

## جمع بندی

در این پایان‌نامه هدف تشخیص نوع اسلحه با استفاده از صدای شلیک آن می‌باشد که بدین منظور در فصل اول به معرفی اصوات ضربه‌ای و مقدمات ورود به این فصل پرداختیم. در فصل دوم پژوهش‌های انجام‌شده در این حوزه و مبانی نظری مورد استفاده در این پایان‌نامه را بررسی کردیم و پس از آن روش پیشنهادی برای تشخیص اسلحه و نتایج حاصل‌شده از اعمال روش پیشنهادی در قالب فصول سه و چهار بیان کردیم. اکنون در این فصل به جمع‌بندی نهایی، نتیجه‌گیری ارائه راهکاری برای ادامه پژوهش خواهیم پرداخت. در این پایان‌نامه با استفاده از روش‌های مختلف استخراج ویژگی از اصوات ضربه‌ای و استفاده از یک شبکه عصبی مصنوعی به‌عنوان طبقه بند به دسته‌بندی اصوات ضربه‌ای پرداختیم.

بدین منظور از ویژگی‌های MFCC، Real Cepstrum و LPC با تعداد ۱۰ و ۲۰ و ۵۰ ضریب برای استخراج ویژگی از اصوات استفاده کردیم و از یک شبکه عصبی پرسپترون دولایه با تعداد ۶ نورون در لایه خروجی و ۷۰ نورون در لایه مخفی و تعداد مناسب نورون در لایه‌ی ورودی با توجه به ویژگی‌های استخراج‌شده به‌عنوان طبقه بند استفاده کردیم.

داده‌های استفاده‌شده برای سیستم توضیح داده‌شده شامل شلیک ۶ اسلحه سبک کمری می‌باشد که ۵۰٪ از داده‌ها به‌عنوان داده‌های آموزش و ۵۰٪ از داده‌ها به‌عنوان داده‌های آزمایش مورد استفاده قرار گرفت.

- با مشاهده نتایج بدست آمده، استفاده از بهترین روش برای هر اسلحه به شرح ذیل می‌باشد:
- در اسلحه ۱ بیشترین دقت تشخیص صحیح با توجه به استفاده از روش کپستروم حقیقی می‌باشد.
  - در اسلحه ۲ بیشترین دقت تشخیص صحیح منوط به استفاده از روش پیش بینی خطی با ضریب ۲۰ می‌باشد.
  - در اسلحه ۳ بیشترین دقت تشخیص صحیح منوط به استفاده از روش کپستروم حقیقی می‌باشد.
  - در اسلحه ۴ بیشترین دقت تشخیص صحیح منوط به استفاده از روش پیش بینی خطی با ضریب ۵۰ می‌باشد.
  - در اسلحه ۵ بیشترین دقت تشخیص صحیح منوط به استفاده از روش کپستروم حقیقی می‌باشد.
  - در اسلحه ۶ بیشترین دقت تشخیص صحیح منوط به استفاده از روش‌های کپستروم حقیقی و پیش بینی خطی با ضریب ۵۰ می‌باشد.

نتایج حاصل از پیاده‌سازی در این پایان‌نامه نشان می‌دهد که بهترین درصد تشخیص نوع اسلحه برای اسلحه شماره ۵ می‌باشد که سیگنال صوتی ناشی از شلیک آن با ۵ اسلحه دیگر بیشترین تفاوت را در حوزه‌ی زمانی دارد؛ به‌طوری‌که با استفاده از ویژگی کپستروم حقیقی تا ۹۹/۹ درصد صحت دقت

تشخیص را برای این اسلحه داریم. همچنین بیشترین خطا در تشخیص نوع اسلحه مربوط به سلاح‌های ۲ و ۳ و ۴ هستند که سیگنال صوتی ناشی از شلیک آن‌ها در حوزه زمان بسیار شبیه یکدیگر است، بطوریکه برای آموزش شبکه حتی با بیش از ۱۵۰۰ دوره بعضاً خطایی به اندازه ۱/۲۸٪ نیز قابل مشاهده است. برای استخراج ویژگی از روش ضرایب پیش بینی خطی از سه ضریب ۱۰ و ۲۰ و ۵۰ استفاده کردیم و پس از مشاهده نتایج بدست آمده بیشترین دقت تشخیص صحیح در بین این ضرایب متعلق به ضرایب پیش بینی خطی با ضریب ۵۰ می‌باشد.

## پیشنهادات:

با توجه به آزمایشات انجام شده در این پایان‌نامه مشاهده شده است که ویژگی‌های متفاوتی برای تشخیص و دسته‌بندی هرکدام از اسلحه‌ها وجود دارد. به طوری که هریک از این اسلحه‌ها دارای ویژگی‌های منحصر به فردی می‌باشند که نسبت به سایرین برتری دارند. بنابراین بطور کلی نمی‌توان برای تشخیص نوع اسلحه سبک از روی صدای شلیک گلوله، تنها یک ویژگی خاص را مدنظر قرارداد. بنابراین با توجه به نتایج بدست آمده ناشی از استفاده ویژگی‌های مختلف برای پایگاه داده موردنظر پیشنهاد می‌شود:

۱- از ترکیب ویژگی‌ها در طبقه بند برای تشخیص نوع اسلحه استفاده شود.

۲- با توجه به استفاده از طبقه بند شبکه عصبی در این پایان نامه از طبقه بندهای دیگر نیز استفاده

گردد.

۳- برای داشتن راندمان بالاتر نیز می توان از ترکیب طبقه بندها نیز استفاده شود.

## منابع و مراجع :

- [1] A. E. Omer, "Joint MFCC-and-vector quantization based text-independent speaker recognition system," in *Communication, Control, Computing and Electronics Engineering (ICCCCEE), 2017 International Conference on*, pp. 1-6, 2017.
- [2] Z. N. Karam, W. M. Campbell, and N. Dehak, "Graph relational features for speaker recognition and mining," in *Statistical Signal Processing Workshop (SSP), 2011 IEEE*, pp. 525-528, 2011.
- [3] B. Wheatley, G. Doddington, C. Hemphill, J. Godfrey, E. Holliman, J. McDaniel, and D. Fisher, "Robust automatic time alignment of orthographic transcriptions with unconstrained speech," in *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, pp. 533-536, 1992.
- [4] R. Sprugnoli, G. Moretti, M. Fuoli, D. Giuliani, L. Bentivogli, E. Pianta, R. Gretter, and F. Brugnara, "Comparing two methods for crowdsourcing speech transcription," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 8116-8120
- [5] L. Burget, P. Schwarz, M. Agarwal, P. Akyazi, K. Feng, A. Ghoshal, O. Glembek, N. Goel, M. Karafiát, and D. Povey, "Multilingual acoustic modeling for speech recognition based on subspace Gaussian mixture models," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010, pp. 4334-4337.
- [6] H. Erdogan, J. R. Hershey, S. Watanabe, and J. Le Roux, "Phase-sensitive and recognition-boosted speech separation using deep recurrent neural networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, 2015, pp. 708-712.
- [7] S. Chu, S. Narayanan, C.-C. J. Kuo, and M. J. Mataric, "Where am I? Scene recognition for mobile robots using audio features," in *Multimedia and Expo, 2006 IEEE International Conference on*, 2006, pp. 885-888
- [8] X. Guo, Y. Toyoda, H. Li, J. Huang, S. Ding, and Y. Liu, "Environmental sound recognition using time-frequency intersection patterns," *Applied Computational Intelligence and Soft Computing*, Vol. 2012, pp. 2-20, 2012

- [9] J. G. A. Barbedo and G. Tzanetakis, "Musical instrument classification using individual partials," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, No. 1, pp. 111-122, 2011.
- [10] O. Dalby, D. Butler, and J. W. Birkett, "Analysis of gunshot residue and associated materials—a review," *Journal of forensic sciences*, Vol. 55, No. 4, pp. 924-943, 2010.
- [11] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and gunshot detection and localization for audio-surveillance systems," in *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, 2007, pp. 21-26.
- [12] S. Astapov, J. Preden, and E. Suurjaak, "A method of real-time mobile vehicle identification by means of acoustic noise analysis implemented on an embedded device," in *Electronics Conference (BEC), 2012 13th Biennial Baltic*, 2012, pp. 283-286.
- [13] A. Aljaafreh and L. Dong, "An evaluation of feature extraction methods for vehicle classification based on acoustic signals," in *Networking, Sensing and Control (ICNSC), 2010 International Conference on*, 2010, pp. 570-575.
- [14] F. Weninger and B. Schuller, "Audio recognition in the wild: Static and dynamic classification on a real-world database of animal vocalizations," in *acoustics, speech and signal processing (ICASSP), 2011 IEEE international conference on*, 2011, pp. 337-340
- [15] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt, "Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring," *Pattern Recognition Letters*, Vol. 31, No. 12, pp. 1524-1534, 2010
- [16] J.-C. Wang, H.-P. Lee, J.-F. Wang, and C.-B. Lin, "Robust environmental sound recognition for home automation," *IEEE Transactions on Automation Science and Engineering*, vol. 5, no. 1, pp. 25-31, 2008

[17] A. Rabaoui, M. Davy, S. Rossignol, and N. Ellouze, "Using one-class SVMs and wavelets for audio surveillance," *IEEE Transactions on information forensics and security*, vol. 3, no. 4, pp. 763-775, 2008.

[18] T. Ahmed, M. Uppal, and A. Muhammad, "Improving efficiency and reliability of gunshot detection systems," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 513-517

[۱۹] فیاضی فر، ن، (۱۳۹۳)، پایان نامه ارشد: " تشخیص صداهای ضربه‌ای با استفاده از تکنیک‌های پردازش سیگنال"، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شاهرود

[20] T. C. Nguyen, L. D. Pham, H. M. Nguyen, B. G. Bui, D. T. Ngo, and T. Hoang, "A High Performance Dynamic ASIC-Based Audio Signal Feature Extraction (MFCC)," in *Advanced Computing and Applications (ACOMP), 2016 International Conference on*, 2016, pp. 113-120

[21] I. Mporas, T. Ganchev, M. Sifarakis, and N. Fakotakis, "Comparison of speech features on the speech recognition task," *Journal of Computer Science*, vol. 3, no. 8, pp. 608-616, 2007.

[22] X. Zhao and D. Wang, "Analyzing noise robustness of MFCC and GFCC features in speaker identification," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 7204-7208.

[23] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *the Journal of the Acoustical Society of America*, vol. 87, no. 4, pp. 1738-1752, 1990.

[24] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*: Prentice Hall, 1978

[25] Y. Hu and P. C. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," *IEEE transactions on Speech and Audio processing*, vol. 12, no. 1, pp. 59-67, 2004.

[26] S.-H. Chen, H.-T. Wu, Y. Chang, and T.-K. Truong, "Robust voice activity detection using perceptual wavelet-packet transform and Teager energy operator," *Pattern Recognition Letters*, vol. 28, no. 11, pp. 1327-1332, 2007.



- [27] M. El Badaoui, F. Guillet, and J. Daniere, “New applications of the real cepstrum to gear signals, including definition of a robust fault indicator,” *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1032-1046, 2004.
- [28] M. Sumithra and A. Devika, “A study on feature extraction techniques for text independent speaker identification,” in *Computer Communication and Informatics (ICCCI), 2012 International Conference on*, 2012, pp. 1-5.
- [29] P. Kumar, A. Biswas, A. Mishra, and M. Chandra, “Spoken language identification using hybrid feature extraction methods,” *arXiv preprint arXiv:1003.5623*, 2010.
- [30] P. L. Rocha and W. L. S. Silva, “Artificial neural networks used for pattern recognition of speech signal based on DCT parametric models of low order,” in *Industrial Informatics (INDIN), 2016 IEEE 14th International Conference on*, 2016, pp. 46-51.
- [31] A. Rabaoui, H. Kadri, Z. Lachiri, and N. Ellouze, “Using robust features with multi-class SVMs to classify noisy sounds,” in *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on*, 2008, pp. 594-599.
- [32] S. Besbes and Z. Lachiri, “Multi-class SVM for stressed speech recognition,” in *Advanced Technologies for Signal and Image Processing (ATSIP), 2016 2nd International Conference on*, 2016, pp. 782-787.
- [33] E. Yüncü, H. Hacıhabiboglu, and C. Bozsahin, “Automatic speech emotion recognition using auditory models with binary decision tree and svm,” in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, 2014, pp. 773-778.
- [34] J. Apolinario, P. Mendonca, R. Chaves, and L. Calôba, “Cryptanalysis of speech signals ciphered by TSP using annealed Hopfield neural network and genetic algorithms,” in *Circuits and Systems, 1996., IEEE 39th Midwest symposium on*, 1996, pp. 821-824.
- [35] U. Ogbureke, J. Cabral, and J. Berndsen, “Using multilayer perceptron for voicing strength estimation in HMM-based speech synthesis,” in *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*, 2012, pp. 683-688.

- [36] A. Gupta, "Synthesis and performance analysis of a Recurrent Fuzzy Multilayer Perceptron for speech recognition," in *Methods and Models in Computer Science (ICM2CS), 2010 International Conference on*, 2010, pp. 22-26.
- [37] M. Vacher, D. Istrate, L. Besacier, J.-F. Serignat, and E. Castelli, "Life sounds extraction and classification in noisy environment," in *5th IASTED-SIP*, 2003.
- [38] Cowling, Michael, and Renate Sitte. "Comparison of techniques for environmental sound recognition." *Pattern recognition letters* 24.15 (2003): 2895-2907.
- [39] A. Mesaros, T. Heittola, A. Eronen, and T. Virtanen, "Acoustic event detection in real life recordings," in *Signal Processing Conference, 2010 18th European*, 2010, pp. 1267-1271.
- [40] I. L. Freire and J. A. Apolinário Jr, "Gunshot detection in noisy environments," in *Proceeding of the 7th International Telecommunications Symposium, Manaus, Brazil*, 2010, pp. 1-4.
- [41] X. Valero and F. Alías, "Gammatone wavelet features for sound classification in surveillance applications," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, 2012, pp. 1658-1662
- [42] J.-L. Rouas, J. Louradour, and S. Ambellouis, "Audio events detection in public transport vehicle," in *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*, 2006, pp. 733-738.
- [43] M. Hrabina and M. Sigmund, "Comparison of feature performance in gunshot detection depending on noise degradation," in *Radioelektronika (RADIOELEKTRONIKA), 2017 27th International Conference*, 2017, pp. 1-4.
- [44] A. W. Zimmermann, "Versatile M1911-style handgun and improved magazine for rifles and handguns," ed: Google Patents, 2006.
- [45] A. W. Zimmermann, "Versatile M1911-style handgun and improved magazine for rifles and handguns," ed: Google Patents, 2006
- [46] J. Walter, *Guns of the elite forces*: Frontline Books, 2005.
- [47] <http://www.beretta.com>, retrieved at 2017-01-14

[48] S. Wikipedia, *Beretta Firearms: Beretta Pistols, Beretta Submachine Guns, Beretta 92, Beretta M9, Beretta*: University-Press Org, [48] 2013 J. Walter, *Guns of the elite forces*: Frontline Books, 2005 .

[49] S. Wikipedia, *Beretta Firearms: Beretta Pistols, Beretta Submachine Guns, Beretta 92, Beretta M9, Beretta*: University-Press Org 2013,

[50] [https://en.wikipedia.org/wiki/IMI\\_Desert\\_Eagle](https://en.wikipedia.org/wiki/IMI_Desert_Eagle), retrieved at 2017-01-14

[51] D. Alder, *Colt Single Action: From Patersons to Peacemakers*: Skyhorse Publishing, Inc., 2015.



## Abstract:

One of the important issues in the daily life is the security of society. In recent years, dangerous and unusual behavior detection systems are being developed and used. One of the major issues with recognizing unusual and dangerous behaviors and preventing undesirable events is the automatic detection and differentiation of sounds such as explosions, thunderstorms, bullets and so on.

In this thesis, we determine the type of weapon from the sound signal generated by the firing. For this purpose, we use a multi-layer perceptron neural network and features such as real-time casts, linear prediction predictors, and collet-scale coefficients on the Mel scale to identify and distinguish guns.

The weapons used in this dissertation include 6 types of lumbar guns, which are the sound signals used by the industry.

The obtained results of the proposed method show that the approach of using the multilayer perceptron neural network and extracting appropriate features is an appropriate approach in detecting the type of weapon.

**keywords:** Gun Type Detection, Impulse Sound, Neural Network  
Multilayer Perceptron, Real cepstrum, LPC, MFCC



**Shahrood University of Technology**  
**Faculty of Electrical Engineering and Robotics**  
**MSc Thesis in Communication**

**Recognition and classification of explosive sounds from its  
signal by using Neural Network**

By: Morteza Saadati

Supervisor:

Dr. Hossein Marvi

July 2017