

Presenting an intelligent extraction method in audio watermarking systems based on lifting wavelet transform and support vector machine

Seyed Mostafa Pourhashemi¹, Mohammad Mosleh^{1*}, Yousof Erfani^{1, 2}

¹Department of Computer Engineering, Dezful Branch, Islamic Azad University, Dezful, Iran

²Department of Electrical & Computer Engineering, AEL Group, McMaster University, Hamilton, ON, Canada
smpourhashemi@iaud.ac.ir; mosleh@iaud.ac.ir; erfaniy@mcmaster.ca

Corresponding author address: Mohammad Mosleh, Department of Computer Engineering, Faculty of Engineering, Dezful Branch, Islamic Azad University, Dezful, Iran.

Abstract- Nowadays, the rapid growth of the use of information technology and computer networks has increased the transfer of information in a digital form. For this reason, the protection of data has become one of the most important challenges in this field. Watermarking is introduced as one of the newest and most important techniques for data protection. Audio watermarking is considered to be the most challenging type of watermarking due to the nature of audio files. The most extraction methods used in audio watermarking algorithms, which mainly use non-intelligent techniques based on the reverse of embedding rules for extraction phase of audio watermarking, often they are not able to extract watermarking exactly and have a lot of errors in extracting. Our proposed solution to solve this problem is to use an intelligent algorithm to extract the watermark. The purpose of this article is to provide a method that covered the weakness of non-intelligent extraction methods using trained machine learning classifier and helped to improve system performance. For the embedding operation, the Lifting Wavelet Transform (LWT) has been used in the proposed method; in the extraction operation, the Support Vector Machine (SVM) classifier is also used. The trained classifier is able to detect the effects of various attacks on the audio files and consequently, intelligent and precise extraction of watermark. The results of various experiments under different conditions indicate that this intelligent method has achieved appropriate imperceptibility and high capacity along with high robustness.

Keywords- Audio watermarking, Lifting wavelet transform (LWT), Classifier, Machine learning, Support vector machine (SVM)

ارائه یک روش استخراج هوشمند در سیستم‌های نهان‌نگاری صوتی بر اساس تبدیل موجک ارتقاء یافته و ماشین بردار پشتیبان

سید مصطفی پورهایمی^۱، محمد مصلح^{*۱}، یوسف عرفانی^{۱،۲}

۱- گروه مهندسی کامپیوتر، واحد دزفول، دانشگاه آزاد اسلامی، دزفول، ایران

۲- گروه مهندسی برق و کامپیوتر، دانشگاه McMaster، همیلتون، انتاریو، کانادا

smpourhashemi@iaud.ac.ir, mosleh@iaud.ac.ir, erfaniy@mcmaster.ca

* نشانی نویسنده مسئول: محمد مصلح، دزفول، کوی آزادگان، دانشگاه آزاد اسلامی، واحد دزفول، دانشکده فنی و مهندسی، گروه مهندسی کامپیوتر.

چکیده- امروزه رشد سریع استفاده از فناوری اطلاعات و شبکه‌های کامپیوتری، باعث افزایش انتقال اطلاعات در فرم دیجیتال شده است. به همین دلیل حفاظت از داده‌های ارسالی به یکی از مهم‌ترین چالش‌ها در این زمینه مبدل گشته است. نهان‌نگاری بعنوان یکی از جدیدترین و مهمترین تکنیک‌ها برای حفاظت داده‌ها معرفی شده و مورد استفاده قرار می‌گیرد. نهان‌نگاری صوتی بدلیل ماهیت فایل-های صوتی بعنوان چالش برانگیزترین نوع نهان‌نگاری مطرح شده است. اکثر روش‌های استخراج در الگوریتم‌های نهان‌نگاری که عمدتاً از تکنیک‌های غیر هوشمند و بر اساس معکوس قوانین درج، برای فاز استخراج در نهان‌نگاری صوتی استفاده می‌کنند، اغلب قادر به استخراج دقیق نهان‌نگاره نیستند و خطای زیادی در تشخیص دارند. راهکار پیشنهادی ما برای حل این مشکل، استفاده از الگوریتمی هوشمند برای استخراج نهان‌نگاره است. هدف این مقاله، ارائه روشی است که با استفاده از دسته‌بند آموزش دیده مبتنی بر یادگیری ماشین نقاط ضعف روش‌های استخراج غیرهوشمند را پوشش داده و به بهبود عملکرد سیستم کمک کند. برای عملیات درج در روش پیشنهادی از تبدیل موجک ارتقاء یافته^۱ استفاده نموده و در بخش استخراج نیز از دسته‌بند ماشین بردار پشتیبان^۲ استفاده می‌شود. دسته‌بند آموزش دیده، قادر به تشخیص تأثیر حملات مختلف بر روی فایل صوتی و در نتیجه استخراج هوشمندانه و دقیق نهان‌نگاره می‌باشند. نتایج آزمایشات مختلف تحت شرایط متفاوت، نشان می‌دهد که این روش هوشمند، به شفافیت مناسب و ظرفیت بالا، در کنار مقاومت بالا دست یافته است.

واژه‌های کلیدی: نهان‌نگاری صوتی، تبدیل موجک ارتقاء یافته، دسته‌بند، یادگیری ماشین، ماشین بردار پشتیبان

۱- مقدمه

تصویری است، بنابراین نهان‌نگاری صوتی یکی از پرچالش‌ترین زمینه‌های نهان‌نگاری می‌باشد. از آنجایی که ثابت شده است که سیستم شنوایی انسان بسیار حساس‌تر از سیستم بینایی انسان می‌باشد [۱]، (آزمایشات متفاوتی در زمینه عصب شنوایی انجام پذیرفته است [۲])، بنابراین حفظ شفافیت در نهان‌نگاری صوتی بمراتب سخت‌تر از نهان‌نگاری تصویر می‌باشد. به دلیل توانایی

هدف اصلی نهان‌نگاری صوتی دیجیتال، درج یک سری از اطلاعات مخفی در فایل‌های صوتی (بعنوان یکی از مهمترین عناصر مولتی-مدیا) می‌باشد، بگونه‌ای که قابل شنیدن نبوده و در برابر حملات پردازش سیگنال نیز مقاومت خوبی داشته باشد. بدلیل آنکه تشخیص نهان‌نگاره در فایل‌های صوتی بمراتب ساده‌تر از فایل‌های

۱۲ DS-CDMA، ۱۳ FHMA و ۱۴ FDMA. در نهایت بهترین تکنیک دسترسی بر اساس آزمایشات مختلف در شرایط گوناگون، انتخاب می‌گردد. رویکرد استفاده شده برای استخراج نیز بر اساس قوانین معکوس درج و بصورت غیرهوشمند می‌باشد [۶]. هو و هسو یک طرح نهان‌نگاری صوتی کور ارائه داده‌اند. عملیات درج نهان‌نگاره، بوسیله‌ی ماژوله کردن بردارها در حوزه‌ی تبدیل کسینوسی گسسته^{۱۵} انجام می‌پذیرد. برای این منظور از مادولاسیون شاخص کوانتیزاسیون^{۱۶} بهره گرفته شده است. در انتهای درج نیز عملیات انتقال هموار^{۱۷} برای هر فریم صورت می‌پذیرد. الگوریتمی نیز برای حفظ توازن انرژی در باند فرکانسی، با توانایی بازیابی اطلاعات نهان‌شده بطور مستقیم از باند معین ارائه شده است. رویکرد استفاده شده برای استخراج نیز بر اساس قوانین معکوس درج و بصورت غیرهوشمند می‌باشد [۷]. محسن فر و همکاران بر مبنای همکاری تکنیک تجزیه QR^{۱۸} و الگوریتم ژنتیک، یک طرح نهان-نگاری صوتی هوشمند ارائه داده‌اند. تجزیه‌ی QR نوعی تجزیه‌ی ماتریس به دو ماتریس متعامد و مثلثی می‌باشد. در فاز درج، یافتن بهترین مکان برای درج، بر اساس الگوریتم ژنتیک انجام می‌شود. بیت‌های نهان‌نگاره در ماتریس R درج می‌گردند. تابع برازندگی الگوریتم ژنتیک نیز بر اساس پارامتر شفافیت طرح تعیین می‌گردد. روش استخراج، روشی قاعده‌مند و بر اساس معکوس قوانین درج و بصورت غیر هوشمند می‌باشد [۸]. چن و همکاران بر مبنای قضیه کاروش-کاهن-توکر^{۱۹} یک طرح نهان‌نگاری صوتی دیجیتال را ارائه داده‌اند. در این طرح اطلاعات در ضرایب فرکانسی پایین سیگنال صوتی DWT شده، درج می‌گردد. در حقیقت برای تغییر دامنه‌ی فرکانس پایین از تئوری KKT برای حداقل نمودن اختلاف بین ضرایب سیگنال اصلی و سیگنال نهان‌نگاری شده استفاده می‌شود. روش پیشنهادی در این مقاله کور بوده و رویکرد استفاده شده برای استخراج نیز بر اساس قوانین معکوس درج و بصورت غیر هوشمند می‌باشد [۹]. هو و هسو دو طرح برای بهبود عملکرد مادولاسیون منطقی دیتر^{۲۰} جهت استفاده در نهان‌نگاری صوتی کور ارائه نموده‌اند. این الگوریتم بهبود یافته در زیر باند تخمینی سطح چهارم که از طریق تجزیه DWT بدست آمده است، عمل می‌نماید. طرح پیشنهادی اول بر اساس یک فیلتر پایین گذر پاسخ ضربه محدود^{۲۱} دو ضربه برای پوشاندن نویز حاصل از نهان‌نگاری می‌باشد. طرح پیشنهادی دوم نیز بمنظور جبران اعوجاج ارائه شده توسط مادولاسیون بردار است. رویکرد استفاده شده برای استخراج نیز بر اساس قوانین معکوس درج و بصورت غیرهوشمند می‌باشد [۱۰]. جیحون و همکاران یک طرح نهان‌نگاری صوتی کور برای درج داده بوسیله‌ی تغییر میانگین ضرایب DCT ارائه داده‌اند. ایده اصلی، تقسیم باند فرکانسی DCT به فریم‌های کوتاه و تغییر

بالایی که صوت در مخفی کردن داده‌ها دارد، امروزه توجه بسیاری از محققین به این زمینه معطوف شده است. از کاربردهای مهم نهان‌نگاری صوتی می‌توان به حفاظت از حق کپی (ادعای مالکیت)، تشخیص دستکاری، حفاظت در برابر کپی کردن، ثبت اثر انگشت، مانیتورینگ انتشار و حمل اطلاعات اشاره نمود.

پارامترهای تعیین‌کننده در طراحی یک سیستم نهان‌نگاری صوتی، برای کاربردهای مختلف، متفاوت است. برخی از این پارامترها برای همه‌ی سیستم‌های نهان‌نگاری دارای اهمیت‌اند و بعضی فقط در کاربردهای خاص تعیین‌کننده هستند. یک الگوریتم نهان‌نگاری صوتی قابل اعتماد با کارایی قابل قبول، باید خصوصیات زیر را در نظر داشته باشد [۳]:

شفافیت^۳، که حداقل مقدار قابل قبول برای شفافیت، 20 dB است؛ مقاومت^۴، که حالت بهینه زمانی است که میزان خطای سیستم در برابر حملات، 0 باشد؛ ظرفیت^۵، که حداقل میزان ظرفیت در یک سیستم نهان‌نگاری صوتی باید 20 بیت بر ثانیه^۶ باشد.

هیچ روش برجسته‌ای وجود ندارد که هر سه پارامتر فوق را بصورت توأماً تأمین نماید. ولی همواره مصالحه‌ای بین سه ویژگی مذکور وجود دارد که افزایش یکی از آن‌ها موجب کاهش بقیه می‌شود [۴] و این مقاله بدنبال بدست آوردن یک حالت بهینه برای این مصالحه می‌باشد.

در ادامه تعدادی از جدیدترین و معتبرترین مقالات در این حوزه را مورد بررسی اجمالی قرار گرفته است. لی و همکاران یک طرح نهان‌نگاری صوتی مقاوم را بر مبنای تجزیه مقادیر منفرد^۷ و الگوریتم تکامل تفاضلی^۸ و با استفاده از روش کوانتیزاسیون مدولاسیون دیتر^۹ ارائه داده‌اند. ابتدا از دو تبدیل موجک گسسته^{۱۰} و LWT بطور مجزا برای تجزیه‌ی سیگنال میزبان و محاسبه‌ی ضرایب تخمینی استفاده می‌گردد. در ادامه از تجزیه‌ی SVD برای بدست آوردن مقادیر منفرد^{۱۱} بدلیل مقاومت بالای آن‌ها در برابر حملات استفاده می‌شود. پس از آن، تکنیک کوانتیزاسیون DM برای کوانتیزه نمودن SVs و درج نهان‌نگاره مورد استفاده قرار می‌گیرد. همچنین حل مساله‌ی مصالحه‌ی بین معیارهای مقاومت و شفافیت، بطور مؤثر توسط الگوریتم بهینه‌سازی DE صورت می‌پذیرد. رویکرد استفاده شده برای استخراج نیز بر اساس قوانین معکوس درج و بصورت غیرهوشمند می‌باشد [۵]. خلیل و ادیب یک طرح نهان‌نگاری صوتی دیجیتال بر مبنای تکنیک‌های دسترسی چندگانه را ارائه داده‌اند. طرح پیشنهادی اجازه‌ی درج چندین زیر-نهان‌نگاره در یک سیگنال صوتی را می‌دهد. سه تکنیک متفاوت دسترسی برای این منظور بکار گرفته شده است:

بر کیفیت سیگنال صوتی نخواهد داشت. بعلاوه اینکه ضرایب تخمین، کمترین تأثیر را از حملات پردازش صوت خواهند پذیرفت.

نکته مهم دیگر این است که بدلیل آنکه در روش پیشنهادی، عملیات آموزش دسته‌بند در فاز استخراج انجام پذیرفته است، لذا دسته‌بند بصورت مؤثر قادر به تشخیص تأثیر حملات مختلف بر روی فایل صوتی بوده و مقاومت طرح در برابر حملات، بسیار مناسب خواهد بود. دلیل اینکار این است که حملات مخرب در فاز استخراج مشخص شده و در فاز درج هیچگونه حمله‌ای بر روی فایل صوتی صورت نمی‌پذیرد. بعلاوه اینکه روش پیشنهادی این مقاله، روشی کور می‌باشد که بدون نیاز به سیگنال اولیه قادر به استخراج نهان‌نگاره است.

در بخش دوم این مقاله، توضیحاتی در خصوص تبدیل LWT که تبدیل اصلی در روش پیشنهادی می‌باشد، آورده شده است. بخش سوم، روش پیشنهادی را با جزئیات کامل معرفی نموده و نتایج آزمایشات نیز بطور کامل و گام به گام در بخش چهارم آمده است. در نهایت، نتیجه گیری مقاله در بخش پنجم آمده است.

۲- تبدیل موجک ارتقاء یافته

در این بخش، به معرفی تبدیل LWT بعنوان تبدیل اصلی در طرح پیشنهادی پرداخته شده است.

تبدیل LWT بر اساس تجزیه‌ی چندسطحی سیگنال عمل کرده و یک طرح ارتقاء یافته از تبدیل موجک گسسته می‌باشد. در این تبدیل، یک تابع درون‌یابی چند جمله‌ای برای یافتن مقادیر فرکانس بالا استفاده شده و همچنین این تبدیل برای ساختن توابع مقیاس‌دهی بمنظور یافتن مقادیر فرکانس پایین بکار گرفته می‌شود. عملکرد این تبدیل پس از دریافت سیگنال صوتی، بر اساس سه مرحله‌ی تقسیم کردن^{۲۴}، پیش‌بینی^{۲۵} و بروز رسانی^{۲۶} می‌باشد.

در مرحله‌ی تقسیم کردن، سیگنال به نمونه‌های زوج و فرد تقسیم می‌گردد؛ زیرا از این طریق می‌توان حداکثر همبستگی بین نمونه‌های مجاور را برای مرحله‌ی پیش‌بینی استفاده کرد. در مرحله‌ی پیش‌بینی، نمونه‌های زوج در ضریب پیش‌بینی^{۲۷} ضرب شده و نتایج به نمونه‌های فرد اضافه می‌گردد تا ضرایب تفصیلی^{۲۸} (ضرایب بالا گذر^{۲۹}) را تولید نماید. در مرحله‌ی بروز رسانی، ضرایب تفصیلی محاسبه شده توسط مرحله‌ی پیش‌بینی، در ضریب بروز رسانی^{۳۰} ضرب شده و نتایج به نمونه‌های زوج اضافه می‌شود تا ضرایب بزرگ^{۳۱} (ضرایب پایین گذر^{۳۲}) را بدست آورند.

نمونه‌های هر فریم بر مبنای بیت‌های نهان‌نگاره‌ای که در آن‌ها درج است، می‌باشد. در این طرح از رگرسیون خطی و انحراف معیار برای استخراج بیت‌های نهان‌نگاره استفاده شده است [۱۱].

مصلح و همکاران یک طرح نهان‌نگاری صوتی هوشمند را با ترکیب نمودن تکنیک SVD و SVM ارائه نموده‌اند. نهان‌نگاره شامل دو بخش می‌باشد: یک بخش همان رشته‌ی دودویی تصادفی است که برای آموزش دسته‌بند استفاده می‌گردد و بخش دوم نیز معادل دودویی عکس نهان‌نگاره می‌باشد. عملیات درج نهان‌نگاره در ضرایب SV حاصل از SVD انجام می‌پذیرد. استخراج نهان‌نگاره نیز بر مبنای دسته‌بند آموزش یافته‌ی SVM انجام می‌شود. رویکرد روش مذکور، نیمه-کور می‌باشد [۱۲]. لی و همکاران با استفاده از مادولاسیون طیف گسترده^{۲۲} یک طرح نهان‌نگاری صوتی ارائه داده‌اند. این طرح پیش از محاسبه‌ی پارامتر همبستگی، مشخصه‌های ادراکی سیگنال صوتی نهان‌نگاره شده را استخراج می‌نماید. اینکار برای بهینه کردن عملیات استخراج و خارج کردن آن از حالت استخراج مرسوم، انجام شده است [۱۳]. عرفانی و همکاران یک طرح نهان‌نگاری صوتی بر مبنای نمایش هسته‌ی ادراکی سیگنال صوتی که به Spikegram مشهور است، ارائه داده‌اند. Spikegram، یکی از جدیدترین روش‌ها برای نمایش سیگنال صوتی می‌باشد. برای افزایش مقاومت طرح، Spikegram با مجموعه‌ای از گام‌تُن‌ها ترکیب شده است. در عملیات درج، فاز مربوط به ضرایب سیگنال داده شده در حوزه‌ی فوریه تغییر می‌یابد. نهان‌نگاره فقط در هسته‌هایی با دامنه تغییرات بالا، درج می‌شود. در فاز استخراج نیز علامت و فاز گام‌تُن‌ها در Spikegram برای افزایش قدرت تشخیص استخراجگر مورد استفاده قرار می‌گیرد [۱۴].

روش‌های مرسوم همواره بدنال یافتن حالت بهینه‌ی مصالحه‌ی بین پارامترهای سیستم نهان‌نگاری صوتی از طریق ارائه‌ی تکنیکی برای درج در حوزه‌ی خاص، یافتن نقاط و نواحی خاص و یا ترکیب این دو بوده و با دارا بودن سادگی و سربار محاسباتی پایین، ولی بدلیل وابستگی زیادشان به آستانه‌ی تصمیم و بکارگیری یک سری قوانین از پیش تعریف شده، کارایی تشخیص و دیکدر بودن آن‌ها تضعیف شده است. بدین منظور این مقاله با استفاده از دسته‌بند آموزش دیده‌ی یادگیری ماشین سعی در رفع این نقاط ضعف داشته و به بهبود عملکرد سیستم چشم دوخته است.

برای درج نهان‌نگاره نیز از ضرایب تخمین^{۲۳} حاصل از LWT استفاده می‌شود. دلیل انتخاب این تبدیل قدرتمند و نیز این ضرایب، این است که تغییرات این ضرایب اندک بوده و تأثیر زیادی

۳-۱- درج نهان‌نگاره

بدلیل توانایی بالای تبدیل LWT و ضرایب تخمینی حاصل از آن، از این تبدیل برای درج نهان‌نگاره استفاده شده است. نهان‌نگاره‌ی اصلی رشته‌ی دودویی بوده و با W نمایش داده شده است. همچنین یک رشته‌ی دودویی شبه تصادفی قاعده‌مند ایجاد نموده و به ابتدای نهان‌نگاره چسبانده می‌شود. این رشته بعداً در فاز استخراج برای آموزش دسته‌بندهای مبتنی بر یادگیری ماشین مورد استفاده قرار می‌گیرد. این رشته با T نمایش داده شده است. در طرح پیشنهادی، ابتدا سیگنال اصلی به فریم‌هایی تقسیم بندی شده و با استفاده از تکنیک نمونه‌برداری پایین^۳ هر فریم به دو زیرفریم با سایز برابر تقسیم می‌گردد. پس از آن تبدیل LWT سطح دو^۴ بر هر کدام از زیرفریم‌ها اعمال شده و ضرایب تبدیل بدست می‌آیند. در نهایت بر اساس اختلاف انرژی محلی زیرفریم‌ها و تغییر آن در حوزه‌ی موجک، اقدام به درج نهان‌نگاره می‌شود. جزئیات روش پیشنهادی برای عملیات درج در ادامه آمده است:

(۱) ابتدا سیگنال صوتی میزبان (S) به K تا فریم با سایز 2L تقسیم شده و هر فریم با S_k نامگذاری می‌گردد.

$$(k=1, 2, \dots, K)$$

(۲) هر کدام از فریم‌های S_k با استفاده از تکنیک نمونه‌برداری پایین به دو زیرفریم مشابه زیر تقسیم بندی گشته و زیرفریم‌ها بر اساس معادلات (۱) و (۲) محاسبه می‌شود.

$$S_k^1 = \{S_k(1), S_k(2), \dots, S_k(L)\} \quad (1)$$

$$S_k^2 = \{S_k(L+1), S_k(L+2), \dots, S_k(2L)\} \quad (2)$$

L طول هر زیرفریم بوده و k شماره هر فریم می‌باشد.

(۳) در ادامه از تمامی زیرفریم‌ها تبدیل LWT سطح دو گرفته شده و ضرایب تخمینی این تبدیل بر اساس معادله (۳) محاسبه گشته و با A_k^i نمایش داده می‌شود.

$$A_k^i = \left\{ a_k^i(j) \mid j=1, 2, \dots, \frac{L}{4} \right\}, i=1, 2 \text{ and } k=1, 2, \dots, K \quad (3)$$

L طول زیرفریم، k شماره فریم، i شماره زیرفریم و j شماره هر ضریب می‌باشد.

(۴) در ادامه با استفاده از معادله (۴)، سطح انرژی هر کدام از زیرفریم‌ها محاسبه می‌گردد.

$$E_k^i = \sum_{j=1}^{\frac{L}{4}} |a_k^i(j)|^2, \quad (4)$$

$$i=1, 2 \text{ and } k=1, 2, \dots, T+W$$

W طول نهان‌نگاره‌ی اصلی و T طول رشته‌ی دودویی شبه تصادفی می‌باشد.

از جمله مزایای LWT نسبت به DWT می‌توان به موارد زیر اشاره نمود:

- پیاده‌سازی سریعتری نسبت به DWT دارد؛ بدلیل آنکه حجم محاسبات آن، تقریباً نصف محاسبات DWT می‌باشد. بنابراین برای پیاده‌سازی نرم‌افزارهای کاربردی بلادرنگ که خواهان مصرف انرژی پایینتری می‌باشند، بسیار مناسب است.
- این تبدیل اجازه می‌دهد تا محاسبات بطور کامل در محل تبدیل موجک انجام پذیرد. عبارت دیگر، هیچ حافظه کمکی مورد نیاز نیست و سیگنال اصلی را می‌توان با تبدیل موجک آن جایگزین کرد.
- این طرح اجازه‌ی پیاده‌سازی تبدیل‌های موجک عددی صحیح برگشت پذیر را خواهد داد. تبدیل موجک ساده، شامل عملیات ممیز شناور می‌باشد که با توجه به قانون گرد کردن اعداد ممیز شناور، بعضاً خطاهایی را به دنبال دارد. در صورتی که LWT قادر به فشرده‌سازی سیگنال بدون از دست دادن اطلاعات و بازسازی کامل سیگنال می‌باشد.

- تبدیل موجک خطی، غیر خطی و انطباقی، در این طرح امکان پذیر بوده و تبدیل نتیجه شده، معکوس پذیر و برگشت پذیر خواهد بود.

- همچنین در نمونه برداری‌های بی‌قاعده می‌تواند مورد استفاده قرار گیرد. [۱۵]

۳- روش پیشنهادی

همانطور که پیشتر گفته شد اکثر تکنیک‌های پیشین در زمینه‌ی نهان‌نگاره‌ی صوتی با استفاده از روشی مرسوم بر پایه‌ی یک سری قوانین از پیش تعیین شده و بر اساس معکوس قوانین درج اقدام به استخراج نهان‌نگاره می‌نمایند. در این مقاله روشی مبتنی بر دسته‌بند یادگیری ماشین SVM ارائه خواهد شد که با استفاده از الگوریتمی هوشمند و بصورت یادگیر، به استخراج دقیق و بهینه‌ی نهان‌نگاره منجر شود. این دسته‌بند هوشمند بدلیل آنکه در فاز استخراج مورد آموزش و یادگیری قرار می‌گیرد، لذا تأثیرات حملات مخرب را درک نموده و قادر به تشخیص تأثیر حملات مختلف بر روی فایل صوتی و استخراج هوشمندانه نهان‌نگاره می‌باشند. در فاز درج نهان‌نگاره نیز از ضرایب تخمینی حاصل از تبدیل قدرتمند LWT بهره گرفته شده است. جزئیات فاز درج و فاز استخراج بطور کامل در ادامه آمده است.

می‌گیرند. بنابراین با روش‌های غیرهوشمندی سر و کار خواهیم داشت که در اکثر مواقع قادر به تشخیص اثرات مخرب حملات بر روی فایل‌های صوتی نخواهد بود. از طرفی با استفاده مناسب از دسته‌بندهای یادگیری ماشین می‌توان این اثرات مخرب را بهتر تشخیص داده و نتایج مناسب‌تری گرفت. جزییات روش پیشنهادی در عملیات استخراج نهان‌نگاره بصورت کامل در ادامه آمده است:

(۱) ابتدا سیگنال صوتی دریافت شده (\bar{S}) به k تا فریم با سایز $2L$ تقسیم‌بندی شده و آن‌ها با (\bar{S}_k) نمایش داده می‌شود. $k=1,2,\dots,K$
 (۲) بر اساس معادلات (۱) و (۲)، هر فریم با استفاده از تکنیک نمونه‌برداری پایین به دو زیرفریم $(\bar{S}_k^i, i=1,2)$ تقسیم می‌گردد.
 (۳) تبدیل LWT سطح دو بر روی هر کدام از زیرفریم‌ها اعمال شده و ضرایب تخمین $(B_k^i = \{b_k^i\}, i=1,2)$ تولید می‌گردد.
 (۴) با استفاده از دنباله شبه تصادفی پیشنهادی، رشته‌ی دودویی با طول T بازتولید می‌شود. از آن‌ها بعنوان کلاس مربوط به ضرایب هر فریم و برای آموزش دسته‌بندها استفاده خواهد شد.
 (۵) در اینجا محاسبه‌ی مجموعه‌ی آموزشی $(DS_{Training})$ بعنوان برچسب (کلاس) با استفاده از ضرایب حاصل از تبدیل LWT سطح دو و بر اساس معادله‌ی (۱۱) انجام می‌پذیرد.

$$DS_{Training} = \left\{ \left(b_k^1(1), b_k^1(2), \dots, b_k^1\left(\frac{L}{4}\right), b_k^2(1), b_k^2(2), \dots, b_k^2\left(\frac{L}{4}\right), w_k \right) \right\} \quad (11)$$

جاییکه $k=1,2,\dots,T$
 k شماره فریم، L طول زیرفریم، w کلاس مربوط به ضرایب هر فریم می‌باشد.

(۶) در این مرحله اقدام به آموزش دسته‌بند SVM بر اساس مجموعه آموزشی $DS_{Training}$ می‌گردد. مزیت استفاده از عملیات آموزش دسته‌بند در فاز استخراج این است که دسته‌بند یادگیر بصورت مؤثر توانایی تشخیص تأثیر حملات مختلف بر روی فایل صوتی را پیدا کرده و استخراج هوشمندانه‌تری خواهیم داشت. همانطور که پیشتر نیز بیان شد، مقالات پیشینی (مانند مرجع [۱۲]) که اقدام به استفاده از دسته‌بندهای یادگیر داشته‌اند، عملیات آموزش دسته‌بندها را در فاز درج انجام می‌دادند. بنابراین دسته‌بندها قادر به تشخیص تأثیرات مخرب حملات بر روی فایل صوتی نبوده و استخراج گاهاً با مشکل مواجه می‌شد. ولی روش پیشنهادی این مقاله، بدلیل استفاده از یک رویکرد مناسب، قادر به انجام عملیات آموزش دسته‌بندها در فاز استخراج می‌باشد.

(۵) شرطی که برای درج بیت "1" مدنظر قرار می‌گیرد این است که $E_k^2 \geq E_k^1$ باشد و اگر این شرط ارضاء نشود باید انرژی‌ها تغییر داده شود. بررسی تغییرات انرژی بر اساس معادلات (۵) تا (۷) انجام می‌پذیرد:

$$\Delta E_k = |E_k^1 - E_k^2| + \lambda \quad (5)$$

$$E_k^1 = \begin{cases} E_k^1 & \lambda \geq |E_k^1 - E_k^2| \\ E_k^1 - \frac{\Delta E_k}{2} & \lambda < |E_k^1 - E_k^2| \end{cases} \quad (6)$$

$$E_k^2 = \begin{cases} E_k^2 & \lambda \geq |E_k^1 - E_k^2| \\ E_k^2 + \frac{\Delta E_k}{2} & \lambda < |E_k^1 - E_k^2| \end{cases} \quad (7)$$

λ یک عدد ثابت است.

(۶) شرطی که برای درج بیت "0" مدنظر قرار می‌گیرد این است که $E_k^2 < E_k^1$ باشد و اگر این شرط ارضاء نشود باید انرژی‌ها بر اساس معادلات (۸) و (۹) تغییر داده شود:

$$E_k^1 = \begin{cases} E_k^1 & \lambda \geq |E_k^1 - E_k^2| \\ E_k^1 + \frac{\Delta E_k}{2} & \lambda < |E_k^1 - E_k^2| \end{cases} \quad (8)$$

$$E_k^2 = \begin{cases} E_k^2 & \lambda \geq |E_k^1 - E_k^2| \\ E_k^2 - \frac{\Delta E_k}{2} & \lambda < |E_k^1 - E_k^2| \end{cases} \quad (9)$$

(۷) در ادامه بر اساس معادله‌ی (۱۰) و با توجه به انرژی محلی مادولاسیون شده، ضرایب زیرفریم‌ها اصلاح می‌شود.

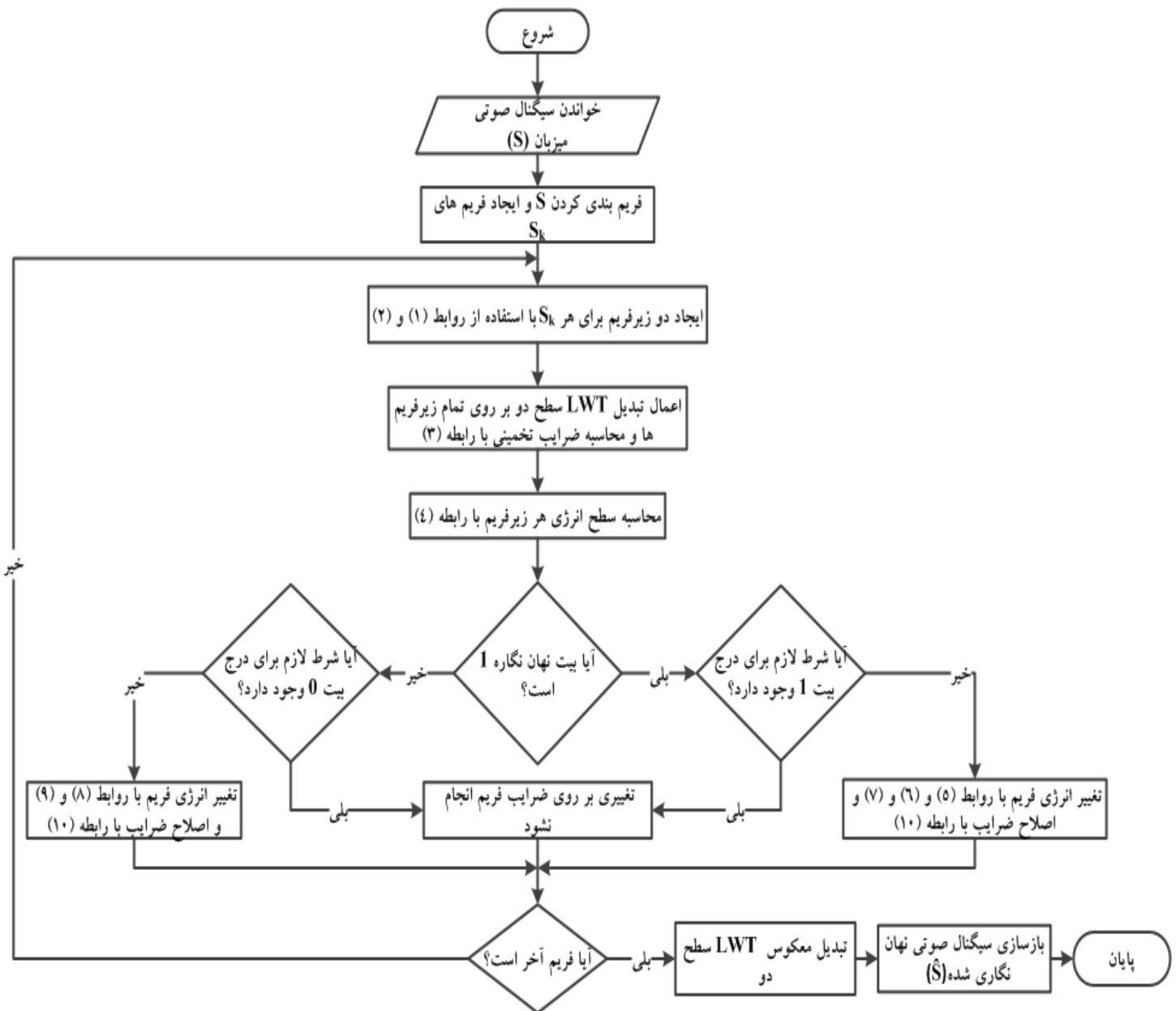
$$a_k^i(j) = \text{sign}(a_k^i(j)) \cdot \sqrt{\frac{((a_k^i(j)))^2 \cdot E_k^{ii}}{E_k^i}} \quad (10)$$

جاییکه $i=1,2$ و $j=1,2,\dots,\frac{L}{4}$ و $k=1,2,\dots,T+W$

(۸) در نهایت تبدیل معکوس LWT سطح دو^{۳۵} گرفته شده و سیگنال صوتی نهان‌نگاری شده (\bar{S}) بازسازی می‌گردد. فلوچارت مربوط به عملیات درج نهان‌نگاره در روش پیشنهادی، در شکل (۱) ارائه شده است.

۳-۲- استخراج نهان‌نگاره

همانطور که پیشتر گفته شد اکثر تکنیک‌هایی که در زمینه‌ی نهان‌نگاری صوتی مطرح شده‌اند، از روش استخراجی غیر هوشمند و مبتنی بر یکسری قوانین از پیش تعیین شده (معکوس قوانین درج) استفاده می‌کنند. روش‌های مرسوم هیچگونه هوشمندی نداشته و صرفاً بدلیل سادگی و پیچیدگی کم، مورد استفاده قرار



شکل ۱: فلوچارت الگوریتم پیشنهادی درج

W طول نهان نگاره‌ی اصلی و T طول رشته‌ی دودویی شبه تصادفی است.

این عملیات برای تمامی W تا فریم انجام پذیرفته و نهان‌نگاره بصورت هوشمند استخراج می‌گردد. فلوچارت مربوط به عملیات استخراج نهان نگاره در روش پیشنهادی، در شکل (۲) ارائه شده است.

۴- نتایج پیاده سازی

روش پیشنهادی بطور کامل در نرم افزار Matlab-2012 پیاده سازی شده است. برای ارزیابی روش نیز از سه نوع فایل صوتی

(۷) پس از پایان آموزش دسته‌بند، نوبت به استخراج نهان‌نگاره می‌باشد. جزییات مربوط به این بخش در ادامه آمده است: (لازم بذکر است که عملیات استخراج از یک فریم بعد از آخرین فریم آموزش، یعنی فریم $(T+1)$ شروع می‌گردد. با استفاده از دسته‌بند SVM آموزش یافته، اقدام به استخراج بیت‌های نهان‌نگاره از طریق برچسب برگشتی با استفاده از معادله‌ی (۱۲) می‌شود.

$$W'_k = \begin{cases} 0 & \text{if } return_label_k = -1 \\ 1 & \text{if } return_label_k = +1 \end{cases}, \quad (12)$$

$$T + 1 \leq k \leq T + W$$

EAQUAL اندازه‌گیری می‌شود [۱۷]. همانگونه که در جدول (۱) آمده است، ODG می‌تواند اعدادی در بازه‌ی $[-4, 0]$ را اختیار کند.

جدول ۱: مقادیر معرف ODG

Explanation	Quality	ODG value
Imperceptible	Excellent	0.0
Perceptible	Good	-1.0
Slightly annoying	Fair	-2.0
Annoying	Poor	-3.0
Very annoying	Bad	-4.0

برای محاسبه‌ی میزان مقاومت طرح در برابر حملات از پارامترهای BER و NC استفاده شده است. پارامتر BER بر اساس معادله‌ی (۱۴) و پارامتر NC بر اساس معادله‌ی (۱۵) تعریف می‌گردند [۸]:

$$BER(W, \bar{W}) = \frac{\sum_{i=1}^M \sum_{j=1}^M W(i, j) \oplus \bar{W}(i, j)}{M \times M} \quad (14)$$

$$NC(W, \bar{W}) = \frac{\sum_{i=1}^M \sum_{j=1}^M W(i, j) \bar{W}(i, j)}{\sqrt{\sum_{i=1}^M \sum_{j=1}^M W^2(i, j)} \cdot \sqrt{\sum_{i=1}^M \sum_{j=1}^M \bar{W}^2(i, j)}} \quad (15)$$

در روابط (۱۴) و (۱۵)، نماد \oplus همان عملگر یای انحصاری^۳، و نیز W و \bar{W} به ترتیب نهان‌نگاره‌ی اصلی و استخراج شده می‌باشند. همچنین M به طول نهان‌نگاره اشاره دارد. مقدار پارامتر BER هرچه به ۰ نزدیک‌تر باشد و مقدار پارامتر NC هرچه به ۱ نزدیک‌تر باشد، نشان‌دهنده‌ی مقاومت بالاتر طرح پیشنهادی خواهد بود. برای بررسی مقاومت طرح پیشنهادی برخی از حملات که در نرم افزار Stirmark [۱۸] و نیز حملاتی که در نرم افزار Matlab شبیه سازی شده اند، استفاده شده است.

متفاوت شامل Rock, Blues, Electronic استفاده شده که تمامی آن‌ها سیگنال‌های صوتی 16-bit mono در فرمت wave می‌باشند که بصورت 44,100 Hz نمونه برداری شده‌اند. تمامی این فایل‌های صوتی استاندارد با طول ۱۰ ثانیه بوده [۸] و محاسبات ما در این مقاله بر مبنای بُرش ۱ ثانیه‌ای از ابتدای این فایل صوتی استاندارد می‌باشد. شکل موج این فایل‌های صوتی در شکل (۳) نمایش داده شده است. همچنین دنباله‌ی نهان‌نگاره نیز در اندازه‌ی 30×30 ارائه شده که در شکل (۴) قابل رؤیت می‌باشد [۱۶].

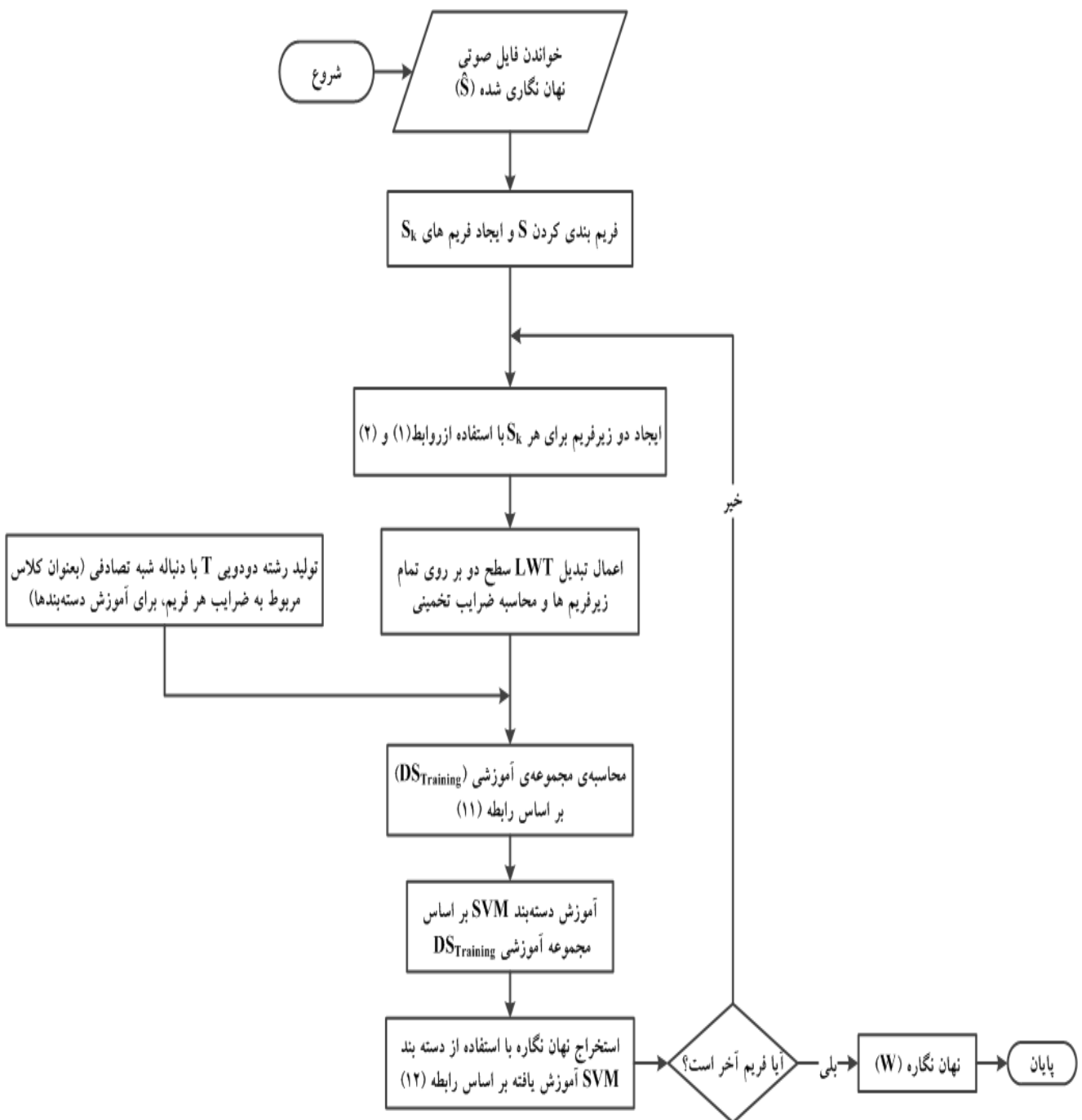
نهان‌نگاره‌هایی که در سیستم‌های نهان‌نگاری صوتی برای درج استفاده می‌گردند، بصورت تصاویر سیاه و سفید دیده می‌شوند. بدلیل اینکه نهان‌نگاره‌ها ترکیبی از ۰ و ۱‌ها هستند که ۰‌ها نشان‌دهنده‌ی نقاط تیره (سیاه) در طرح نهان‌نگاره و ۱‌ها نمایانگر نقاط روشن (سفید) در طرح می‌باشند. از آنجایی که برای درج، نهان‌نگاره (در هر ابعادی که باشد) ابتدائاً بصورت یک رشته‌ی بیتی (بردار تک بعدی^۴) متشکل از ۰ و ۱‌های پشت سر هم در نظر گرفته می‌شود، بنابراین تعداد کل بیت‌های نهان‌نگاره دارای اهمیت می‌باشد.

برای ارزیابی کارایی طرح پیشنهادی از پنج پارامتر نرخ سیگنال به نویز^{۳۷}، درجه تفاوت هدف^{۳۸}، نرخ خطای بیتی^{۳۹}، همبستگی طبیعی^{۴۰} و ظرفیت جاده‌ی^{۴۱} بهره گرفته شده است. پارامتر SNR که برای نشان دادن شفافیت طرح نهان‌نگاری صوتی پیشنهادی مورد استفاده قرار می‌گیرد، برای محاسبه شباهت بین فایل صوتی اصلی و فایل صوتی نهان‌نگاری شده، توسط مؤسسه بین‌المللی صنعت فونوگرافی^{۴۲} ارائه شده و بصورت معادله‌ی (۱۳) تعریف می‌گردد [۸]:

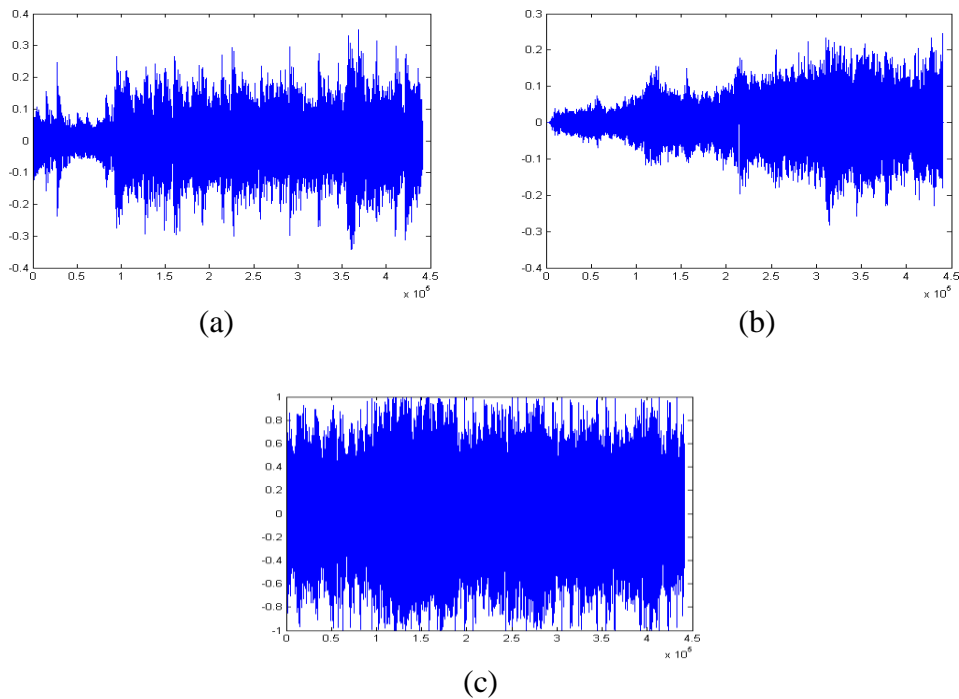
$$SNR = 10 \log_{10} \frac{\sum_n S^2[n]}{\sum_n [S[n] - \bar{S}[n]]^2} \quad (13)$$

در رابطه‌ی بالا، $S[n]$ و $\bar{S}[n]$ به ترتیب، سیگنال اصلی و سیگنال نهان‌نگاری شده می‌باشند.

دیگر پارامتر تاثیرگذار برای نشان دادن میزان کیفیت شنوایی طرح، پارامتر ODG می‌باشد که توسط نرم افزار متن باز به نام



شکل ۲: فلوچارت الگوریتم پیشنهادی استخراج



شکل ۳: شکل موج سیگنال‌های صوتی (a-Rock, b-Blues, c-Electronic)

IR

شکل ۴: نهان نگاره (امضای صاحب اثر)

همانگونه که از نتایج جدول (۲) قابل ملاحظه است، ضرایب صوتی حاصل از تبدیل LWT، در فایل صوتی از نوع Electronic کمترین تغییرات را در برابر عملیات درج در طرح پیشنهادی داشته و به همین دلیل بیشترین میزان شفافیت در زمان استفاده از این نوع فایل صوتی بدست آمده است. با استدلال مشابه، دومین میزان شفافیت بالا در سیستم نهان‌نگاری صوتی پیشنهادی در زمان استفاده از فایل صوتی از نوع Rock بدست آمده است. همچنین در زمان استفاده از فایل صوتی از نوع Blues ضرایب حاصل از تبدیل، بیشترین تغییرات را به نسبت دو فایل صوتی دیگر پذیرفته و به همین خاطر میزان شفافیت نسبت به آنها کاهش پیدا کرده است. ولی نکته حائز اهمیت این است که دامنه‌ی تغییرات مقادیر معیارهای شفافیت پایین بوده و برای فایل‌های صوتی متفاوت، مقادیر نزدیک به هم بدست آمده است. این مسأله نشان‌دهنده‌ی یکپارچه بودن کل سیستم پیشنهادی بر روی فایل‌های صوتی متفاوت و موفقیت در دسترسی به شفافیت بالا می‌باشد.

همانطور که پیشتر اشاره شد، برای آموزش دسته‌بند یادگیری ماشین، یک رشته‌ی ۶۰۰ بیتی توسط دنباله‌ی شبه تصادفی

برای ارزیابی ظرفیت نهان‌نگاری الگوریتم پیشنهادی نیز پارامتر Payload بکار گرفته شده است. این پارامتر بر اساس معادله‌ی (۱۶) تعریف می‌گردد:

$$Payload = \frac{M}{N} bps \quad (16)$$

که M تعداد بیت‌های نهان نگاره و N طول بخشی از سیگنال صوتی میزبان که در آن درج انجام شده است، می‌باشد. در جدول (۲) میزان شفافیت بدست آمده بر اساس پارامترهای SNR و ODG برای هر کدام از فایل‌های صوتی نشان داده شده است.

جدول ۲: نتایج شفافیت طرح پیشنهادی با توجه به پارامترهای SNR و

ODG

فایل صوتی	SNR (dB)	ODG
Rock-and-roll	42.61	-0.37
Blues	41.92	-0.41
Electronic	43.43	-0.31
میانگین	42.65	-0.36

بطور مشابه عملاتی نیز توسط نرم افزار Matlab بر روی سیگنال صوتی نهان نگاری شده اعمال می شود تا باعث تخریب سیگنال و تغییر ضرایب گردد. عملاتی که توسط این دو نرم افزار اعمال می گردد، عملاتی استاندارد بوده و در اکثر مقالات مشابه برای ارزیابی مقاومت یک طرح نهان نگاری صوتی مورد استفاده قرار می گیرد. پس از اعمال عملیات و تغییر ضرایب سیگنال صوتی نهان نگاری شده، عملیات استخراج نهان نگاره صورت گرفته و بر اساس دسته بندی SVM که آموزش دیده است و قادر به تشخیص تأثیر عملیات مختلف بر روی فایل صوتی می باشد، استخراج هوشمندانه ی نهان نگاره انجام می گردد.

از جمله این عملیات می توان به موارد زیر اشاره نمود:

- Noise addition: اضافه نمودن نویز گوسین سفید^{۴۹} به سیگنال نهان نگاری شده.
- Low-pass filtering: یک فیلتر پایین گذر با فرکانس قطع 4kHz بر روی سیگنال نهان نگاری شده اعمال می گردد.
- MP3 compression: فشرده کردن سیگنال نهان نگاری شده در نرخ بیت 64kbps و بازسازی مجدد آن.
- Re-quantization: کوانتیزه کردن رو به پایین سیگنال نهان نگاری شده به 8-bit و در ادامه کوانتیزاسیون مجدد به 16-bit.
- Re-sampling: نمونه برداری رو به پایین سیگنال صوتی نهان نگاری شده از 44.1 به 22.05kHz و در ادامه بازگردانی نمونه برداری به نرخ 44.1kHz.
- Cropping: ده نمونه از هر ۴۴۱ سگمنت، برداشته شده و پس از اضافه نمودن نویز سفید، مجدداً جایگزین می گردد. [۱۶]

بر اساس اطلاعات جداول (۳) و (۴)، سیستم پیشنهادی در برابر اکثر عملیات به مقدار ایده آل و کاملی برای BER و NC دست یافته است. در مابقی موارد نیز میزان خطا بسیار اندک بوده و نتایج نسبت به روش هایی که از استخراج مبتنی بر معکوس قوانین درج استفاده نموده اند، مناسب تر می باشد و این مهم بدلیل استفاده از روش استخراج هوشمند مبتنی بر ماشین یادگیر بدست آمده است.

نتایج جداول (۳) و (۴) حاکی از مقاومت بالای طرح پیشنهادی در زمان استفاده از دسته بند SVM می باشد. همانگونه که از نتایج جداول (۳) و (۴) قابل ملاحظه است، در زمان بکارگیری فایل صوتی از نوع Electronic و در زمان اعمال عملیات مختلف، کمترین تخریب بر روی ضرایب تغییر یافته، انجام شده و در برابر اکثر عملیات، استخراج نهان نگاره توسط دسته بند SVM با موفقیت کامل انجام پذیرفته و نرخ خطای طرح پیشنهادی نزدیک به صفر و یا مقادیر بسیار پایین بدست آمده است. با استدلال مشابه،

ساخته شده و به ابتدای نهان نگاره اضافه شده است. از آنجایی که طول نهان نگاره ی اصلی ۹۰۰ بیت می باشد، بنابراین در مجموع ۱۵۰۰ بیت در فایل صوتی، درج شده است. البته با توجه به قانون درج در حوزه نهان نگاری صوتی، به دلیل اینکه ۶۰۰ بیت اول درج شده، یک دنباله دودویی شبه تصادفی برای آموزش دسته بند SVM بوده و جزو نهان نگاره ی اصلی (امضای صاحب اثر) نمی باشد، بنابراین همان ۹۰۰ بیت نهان نگاره ی اصلی، مبنای محاسبات ظرفیت قرار می گیرد. علیرغم این حجم از اطلاعات نهان نگاری شده، میزان SNR و ODG بدست آمده قابل توجه بوده و نشانگر شفافیت بالای طرح پیشنهادی می باشد.

در مرحله ی استخراج از دسته بند SVM آموزش یافته، برای تشخیص بیت های نهان نگاره بهره گرفته شده است. این دسته بند بصورت با ناظر^{۴۴} عمل نموده و از نوع دسته بندهای تابعی^{۴۵} می باشد. مهمترین پارامتر این دسته بند برای مدل کردن و آنالیز داده ها و نیز تشخیص الگو، هسته ی SVM می باشد. بدین منظور از هسته ی متداول و قدرتمند به نام تابع پایه شعاعی^{۴۶} استفاده شده است. هسته ی RBF بر اساس توزیع گوسین عمل می کند. بعلاوه، روشی که برای یافتن ابر صفحه^{۴۷} مورد استفاده قرار گرفته است، الگوریتم کمترین مربعات^{۴۸} می باشد [۱۹]. همچنین دو کلاس 0 و 1 که مشخص کننده ی بیت های نهان نگاره ی درج شده نیز می باشند، برای دسته بندی مورد استفاده قرار گرفته اند. ویژگی های مورد استفاده در این دسته بندی، ضرایب صوتی حاصل از تبدیل LWT می باشند که بر اساس روابط درج، تغییر یافته و عملاتی نیز بر روی آنها اعمال شده است. جداول (۳) و (۴) نتایج بررسی مقدار پارامترهای BER و NC را در برابر عملیات اعمال شده توسط نرم افزارهای Matlab و Stirmark نشان می دهد.

پس از درج نهان نگاره در فایل صوتی میزبان و بازسازی سیگنال صوتی نهان نگاری شده، عملیات مختلفی توسط نرم افزار Stirmark که بطور خاص جهت آزمایش مقاومت فایل های صوتی ایجاد شده است، انجام گرفته که سعی در تخریب فایل صوتی و تغییر ضرایب آن خواهد داشت. دریافت سیگنال ورودی و اعمال عملیات بر روی آن، در نرم افزار Stirmark بصورت خط فرمان زیر انجام می شود:

```
./stirmarkaudio -n -i input.wav -w output
```

در خط فرمان فوق، input.wav فایل صوتی ورودی می باشد. مجموعه عملیات استاندارد بر روی این فایل اعمال شده و خروجی بصورت یک فایل صوتی با فرمت wav و با نام حمله ی اعمالی در مسیر output ذخیره می گردد. بطور مثال، اگر حمله ی Amplify بر روی فایل صوتی ورودی اعمال شود، خروجی با نام Amplify.wav در مسیر output ذخیره شده و معیارهای مقاومت بر روی آن مورد آزمایش و ارزیابی قرار خواهد گرفت.

در اختیار گیرنده قرار نمی‌گیرد، ممکن است در تشخیص خود دچار خطای بیشتری شوند. روش پیشنهادی این مقاله، علاوه بر کور بودن و داشتن امنیت بالا، از نظر تشخیص نهان‌نگاره نیز بسیار موفق عمل نموده و قابلیت عملیاتی شدن دارد.

در جداول (۶) و (۷)، توجه به این نکته حائز اهمیت است که مقاومت هر طرح نهان‌نگاری صوتی در ارتباط با شفافیت و ظرفیت طرح باید ارزیابی گردد. لذا روش‌هایی که نسبت به طرح پیشنهادی این مقاله، مقاومتی با مقدار اندکی بالاتر در برابر برخی حملات داشته‌اند، ظرفیت نهان‌نگاری پایینتری داشته و شفافیت طرحشان از شفافیت طرح پیشنهادی این مقاله پایینتر بوده است. بنابراین باید هر سه معیار مقاومت، ظرفیت و شفافیت بصورت یکپارچه مورد بررسی قرار گیرد، که از این نظر نیز طرح پیشنهادی این مقاله، موفق‌تر عمل نموده است. در مجموع همانطور که از جداول (۵)، (۶) و (۷) نتیجه می‌شود، کل سیستم پیشنهادی در توازن بین سه معیار شفافیت، مقاومت و ظرفیت، موفق عمل نموده است. این توازن و موفقیت مرهون درج دقیق و استخراج هوشمند نهان‌نگاره می‌باشد.

۵- نتیجه‌گیری

در این مقاله از یک روش هوشمند مبتنی بر یادگیری ماشین در فاز استخراج استفاده شد. بدین منظور، دسته‌بند SVM مورد استفاده قرار گرفت. با استفاده از روش هوشمند فوق‌الذکر، نقاط ضعف روش استخراج مرسوم نیز پوشش داده شد. دسته‌بندهای یادگیر بصورت مؤثر قادر به تشخیص تأثیر حملات مختلف بر روی فایل صوتی و استخراج هوشمندانه و بهینه نهان‌نگاره می‌باشند. برای درج نهان‌نگاره نیز اختلاف انرژی محلی ضرایب حاصل از تبدیل LWT مورد استفاده قرار گرفت. علیرغم اینکه دسته‌بندهای یادگیر ذاتاً پیچیدگی بیشتری نسبت به روش مرسوم مبتنی بر همبستگی دارند، ولی طرح پیشنهادی ما از ویژگی‌های خاصی بهره می‌برد که باعث بهینه‌سازی مصالحه بین سه پارامتر مثلث طلایی در نهان‌نگاری (شفافیت، مقاومت و ظرفیت) می‌گردد.

نتایج پیشنهادی از دستیابی به ظرفیت نهان‌نگاری بالا (900bps) با Average-SNR برابر با (42.65 dB) و Average-ODG برابر با (-0.36) حکایت دارد. همچنین مقاومت طرح در برابر حملات رایج Stirmark و حملات Matlab مورد آزمایش قرار گرفته و نتایج مربوط به پارامترهای BER و NC بسیار مناسب است. یکی از راهکارهای پیشنهادی برای ادامه مسیر این تحقیق، ارائه یک روش استخراج هوشمند مبتنی بر الگوریتم‌های بهینه‌سازی در کنار دسته‌بندهای مبتنی بر یادگیری ماشین، برای افزایش دقت در فاز استخراج می‌باشد.

دومین میزان مقاومت بالا و دستیابی به نرخ خطای پایین در زمان استفاده از فایل صوتی از نوع Blues بدست آمده است. همچنین در زمان استفاده از فایل صوتی از نوع Rock، تخریب بیشتری توسط حملات بر روی ضرایب تغییر یافته، انجام شده که نتیجتاً تشخیص درست نهان‌نگاره در مرحله‌ی استخراج را نسبت به دو نوع فایل صوتی دیگر مختل کرده و میزان خطای بیتی را نسبت به آن دو اندکی بالا برده است. البته شایان ذکر است که در تمامی موارد، میزان خطا بسیار پایین بوده و نتایج بدست آمده در خصوص فایل‌های صوتی متفاوت، نزدیک به هم می‌باشد که نشان‌دهنده موفقیت کل سیستم پیشنهادی در دستیابی به مقاومت بالا در برابر حملات مخرب و تشخیص درست نهان‌نگاره با استفاده از دسته‌بند آموزش دیده‌ی SVM می‌باشد.

در ادامه نتایج روش پیشنهادی با برخی دیگر از مقالات، از نظر پارامترهای Payload, SNR, ODG و نیز روش آشکارسازی در جدول (۵) مورد مقایسه قرار گرفته است. همچنین مقایسه‌ای بین مقاومت طرح پیشنهادی در برابر حملات رایج Matlab، با سایر مقالات، از نظر معیارهای BER و NC به ترتیب در جداول (۶) و (۷) انجام شده است. در جدول (۷) فقط مقالاتی که از معیار NC نیز برای آزمایش مقاومت طرح پیشنهادی‌شان استفاده نموده‌اند، آورده شده است.

در خصوص جدول (۵)، ذکر این نکته حائز اهمیت است که اگرچه با افزایش ظرفیت، تغییرات بر روی ضرایب فایل صوتی بیشتر شده و نتیجتاً شفافیت طرح کاهش پیدا خواهد کرد، ولی طرح پیشنهادی به دلیل روش استفاده شده در عملیات درج، در عین دستیابی به ظرفیت بالا، به شفافیت بالایی نیز دست یافته است. بعلاوه اینکه روش پیشنهادی کور بوده و بدون نیاز به سیگنال اولیه قادر به تشخیص هوشمند نهان‌نگاره می‌باشد.

روش آشکارسازی نیمه‌کور برای تکمیل مرحله‌ی استخراج به بخشی از سیگنال صوتی اولیه نیازمند بوده و بدون داشتن آن قادر به استخراج نهان‌نگاره نمی‌باشد؛ بنابراین علاوه بر مشکلاتی که ممکن است به لحاظ امنیتی برای سیستم‌های نهان‌نگاری صوتی مبتنی بر روش آشکارسازی نیمه‌کور ایجاد شود، این سیستم‌ها از نظر پیاده‌سازی عملیاتی و صنعتی نیز با مشکلاتی مواجه خواهند شد. البته بدلیل اینکه بخشی از اطلاعات از قبل در اختیار گیرنده قرار می‌گیرد، لذا درصد خطای تشخیص پایین می‌آید. در مقابل، روش آشکارسازی کور قرار دارد که طبق توضیحات پیشین، بدون نیاز به سیگنال اولیه قادر به تشخیص نهان‌نگاره می‌باشد. بنابراین علاوه بر افزایش ضریب ایمنی سیستم نهان‌نگاری، به لحاظ پیاده‌سازی عملیاتی و صنعتی نیز قابلیت بالاتری دارد. البته این سیستم‌ها طبیعتاً بدلیل آنکه هیچ اطلاعاتی از سیگنال صوتی اولیه

جدول ۳: میزان BER, NC روش پیشنهادی در برابر حملات Stirmark بر روی فایل‌های صوتی متفاوت

حملات	Rock		Blues		Electronic	
	BER	NC	BER	NC	BER	NC
No Attack	0	1	0	1	0	1
Add Brumm	0	1	0	1	0	1
Add noise	0	1	0	1	0	1
Amplify	0	1	0	1	0	1
Compressor	0.22	0.9983	0.11	0.9995	0.11	0.9997
Extra Stereo	0	1	0	1	0	1
FFT-Invert	0.22	0.9984	0.11	0.9996	0	1
Invert	0.22	0.9984	0.22	0.9980	0	1
Lsbzero	0	1	0	1	0	1
Normalize	0	1	0	1	0	1
Nothing	0	1	0	1	0	1
Original	0	1	0	1	0	1
FFT-Real_Reverse	1.00	0.9922	0.88	0.9935	0.55	0.9946
RC-Lowpass	1.77	0.9810	1.33	0.9882	1.00	0.9917
RC-Highpass	2.00	0.9803	1.66	0.9851	1.22	0.9897
میانگین	0.362	0.9965	0.287	0.9976	0.192	0.9983

جدول ۴: میزان BER, NC روش پیشنهادی در برابر حملات Matlab بر روی فایل‌های صوتی متفاوت

حملات	Rock		Blues		Electronic		میانگین	
	BER	NC	BER	NC	BER	NC	BER	NC
No Attack	0	1	0	1	0	1	0	1
Re-sampling (22.05 kHz)	0	1	0	1	0	1	0	1
Re-quantization	0.5	0.9941	0.2	0.9977	0	1	0.23	0.9972
Noise Addition	1.1	0.9881	0.8	0.9907	0.2	0.9966	0.70	0.9918
Low-pass Filtering (4 kHz)	1.0	0.9885	0.6	0.9919	0.3	0.9953	0.63	0.9919
MP3 (64 kbps)	0.4	0.9947	0.5	0.9932	0.2	0.9975	0.36	0.9951
Cropping	0	1	0	1	0	1	0	1
میانگین	0.42	0.9950	0.30	0.9962	0.10	0.9984		

جدول ۵: مقایسه روش پیشنهادی با برخی از جدیدترین مقالات از منظر پارامترهای Payload, SNR, ODG و نیز روش آشکارسازی

مراجع	Payload (bps)	SNR (dB)	ODG	روش آشکارسازی
[5]	136	35 to 39	Not Reported	کور
[6]	1 k	38	-0.36 to -2.7	کور
[7]	84 to 848	17.5 to 26.4	-0.04 to -0.3	کور
[8]	159	27.33	-0.36 to -0.81	نیمه کور
[9]	1 to 2 k	19.8 to 24.1	Not Reported	نیمه کور

مراجع	Payload (bps)	SNR (dB)	ODG	روش آشکارسازی
[10]	102 to 689	19.8 to 30.3	-0.015 to -1.667	کور
[11]	164	35	-0.36	کور
[12]	172	42.07	-0.33	نیمه کور
[13]	43	Not Reported	-0.62 to -3.64	کور
[14]	56	Not Reported	-0.1 to -0.85	کور
روش پیشنهادی	900	42.65	-0.36	کور

جدول ۶: مقایسه روش پیشنهادی با برخی از جدیدترین مقالات از منظر پارامتر BER

مراجع	Noise Addition	Low-pass Filtering (4 kHz)	MP3 (64 kbps)	Re-quantization	Re-sampling (22.05 kHz)
[5]	0	0	0	0	0
[6]	1	11	18	9	11
[7]	0.73	0	2.72	0.03	0
[8]	0	0 (9 kHz)	7	Not Reported	0
[9]	0	2.085 (5 kHz)	2.27 (80 kbps)	Not Reported	0.04
[10]	Not Reported	0	2.253	0	0
[11]	0	Not Reported	19	0	0
[12]	0	0	0	0.8	0
[13]	0.53	0 (10 kHz)	0.29	0.01	0.03
[14]	0	0	0 (32 kbps)	0	0
روش پیشنهادی	0.70	0.63	0.36	0.23	0

جدول ۷: مقایسه روش پیشنهادی با برخی از جدیدترین مقالات از منظر پارامتر NC

مراجع	Noise Addition	Low-pass Filtering (4 kHz)	MP3 (64 kbps)	Re-quantization	Re-sampling (22.05 kHz)
[5]	1	1	1	1	1
[11]	1	Not Reported	0.707	1	1
[12]	1	1	1	0.9881	0.9987
روش پیشنهادی	0.9918	0.9919	0.9951	0.9972	1

transactions on audio, speech, and language processing, 21(11), 2368-2378.

- [6] Khalil, M., & Adib, A. (2014). Audio watermarking with high embedding capacity based on multiple access techniques. *Digital Signal Processing*, 34, 116-125.
- [7] Hu, H. T., & Hsu, L. Y. (2015). Robust, transparent and high-capacity audio watermarking in DCT domain. *Signal Processing*, 109, 226-235.
- [8] Mohsenfar, S. M., Mosleh, M., & Barati, A. (2015). Audio watermarking method using QR decomposition and genetic algorithm. *Multimedia Tools and Applications*, 74(3), 759-779.
- [9] Chen, S. T., Hsu, C. Y., & Huang, H. N. (2015). Wavelet-domain audio watermarking using optimal modification on low-frequency amplitude. *IET Signal Processing*, 9(2), 166-176.
- [10] Hu, H. T., & Hsu, L. Y. (2017). Supplementary schemes to enhance the performance of DWT-RDM-based blind audio watermarking on audio, speech, and language processing, 21(11), 2368-2378.
- [1] Liu, W., & Hu, A. Q. (2017). A subband excitation substitute based scheme for narrowband speech watermarking. *Frontiers of Information Technology & Electronic Engineering*, 18(5), 627-643.
- [2] Bruce, I. C., Erfani, Y., & Zilany, M. S. (2018). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. *Hearing research*, 360, 40-54.
- [3] Bender, W., Gruhl, D., Morimoto, N., & Lu, A. (1996). Techniques for data hiding. *IBM systems journal*, 35(3.4), 313-336.
- [4] Mat Kiah, M. L., Zaidan, B. B., Zaidan, A. A., Mohammed Ahmed, A., & Al-Bakri, S. H. (2011). A review of audio based steganography and digital watermarking. *International Journal of Physical Sciences*, 6(16), 3837-3850.
- [5] Lei, B., Soon, Y., & Tan, E. L. (2013). Robust SVD-based audio watermarking scheme with differential evolution optimization. *IEEE*

مراجع

- ²⁷ predict factor
- ²⁸ detailed coefficients
- ²⁹ High Pass Coefficients
- ³⁰ update factor
- ³¹ coarse coefficients
- ³² Low Pass Coefficients
- ³³ Down-Sampling
- ³⁴ Two-Level LWT
- ³⁵ Inverse two-level LWT
- ³⁶ 1-D vector
- ³⁷ Signal to Noise Ratio (SNR)
- ³⁸ Objective Difference Grade (ODG)
- ³⁹ Bit Error Rate (BER)
- ⁴⁰ Normalized Correlation
- ⁴¹ Payload
- ⁴² International Federation Phonographic Industry (IFPI)
- ⁴³ Exclusive OR (XOR)
- ⁴⁴ Supervised
- ⁴⁵ Functional
- ⁴⁶ Radial Basis Function (RBF)
- ⁴⁷ Hyper-plane
- ⁴⁸ Least squares (LS)
- ⁴⁹ White Gaussian Noise

- watermarking. *Circuits, Systems, and Signal Processing*, 36(5), 1890-1911.
- [11] Jeyhoon, M., Asgari, M., Ehsan, L., & Jalilzadeh, S. Z. (2017). Blind audio watermarking algorithm based on DCT, linear regression and standard deviation. *Multimedia Tools and Applications*, 76(3), 3343-3359.
 - [12] Mosleh, M., Latifpour, H., Kheyrandish, M., Mosleh, M., & Hosseinpour, N. (2016). A robust intelligent audio watermarking scheme using support vector machine. *Frontiers of Information Technology & Electronic Engineering*, 17(12), 1320-1330.
 - [13] Li, R., Xu, S., & Yang, H. (2016). Spread spectrum audio watermarking based on perceptual characteristic aware extraction. *IET Signal Processing*, 10(3), 266-273.
 - [14] Erfani, Y., Pichevar, R., & Rouat, J. (2017). Audio Watermarking Using Spikegram and a Two-Dictionary Approach. *IEEE Transactions on Information Forensics and Security*, 12(4), 840-852.
 - [15] Sweldens, W. (1996). The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and computational harmonic analysis*, 3(2), 186-200.
 - [16] Latifpour, H., Mosleh, M., & Kheyrandish, M. (2015). An intelligent audio watermarking based on KNN learning algorithm. *International Journal of Speech Technology*, 18(4), 697-706.
 - [17] Lerch, A. (2002). Zplane development, EAQUAL-Evaluate Audio Quality, version: 0.1.3alpha. <http://www.mp3-tech.org/programmer/misc.html>.
 - [18] Lang, A. (2005). Stirmark benchmark for audio (smba): Evaluation of watermarking schemes for audio. Version 1.3.1.
 - [19] Alpaydin, E. (2009). *Introduction to machine learning*. MIT press.

پاورقی‌ها:

- ¹ Lifting Wavelet Transform (LWT)
- ² Support Vector Machine (SVM)
- ³ Imperceptibility
- ⁴ Robustness
- ⁵ Capacity
- ⁶ bit per second (bps)
- ⁷ Singular Value Decomposition (SVD)
- ⁸ Differential Evolution (DE)
- ⁹ Dither Modulation (DM)
- ¹⁰ Discrete Wavelet Transform (DWT)
- ¹¹ Singular Values (SVs)
- ¹² Direct-Sequence Code Division Multiple Access
- ¹³ Frequency Hopped Multiple Access
- ¹⁴ Frequency Division Multiple Access
- ¹⁵ Discrete Cosine Transform (DCT)
- ¹⁶ Quantization Index Modulation (QIM)
- ¹⁷ smooth transition
- ¹⁸ QR Decomposition
- ¹⁹ Karush-Kuhn-Tucker (KKT) theorem
- ²⁰ Rational Dither Modulation (RDM)
- ²¹ Finite Impulse Response (FIR)
- ²² Spread spectrum modulation
- ²³ Approximation Coefficients
- ²⁴ Split
- ²⁵ Predict
- ²⁶ Update