



پردیس علوم  
دانشکده ریاضی، آمار و علوم کامپیوتر

# شناسایی سازهای موسیقی سنتی به کمک روش‌های یادگیری ماشین

نگارنده

مریم همتی

استاد راهنما :

دکتر باقر باباعلی

پایان‌نامه برای دریافت درجه کارشناسی  
در رشته علوم کامپیوتر

بهمن ۱۴۰۰

## چکیده

تحلیل و بازیابی اطلاعات موسیقی امروزه مورد توجه بسیاری از محققان جهان قرار گرفته است و شناسایی سازها، متن آهنگ، ژانر موسیقی، خواننده اثر و استخراج هرگونه اطلاعات دیگر از شاخه‌های این حوزه محسوب می‌شوند. در موسیقی سنتی ایرانی هم مسائل تشخیص سازها، دستگاه‌ها و گوشه‌ها مطرح هستند اما متأسفانه مورد توجه مناسبی قرار نگرفته و محدود فعالیت‌هایی صرفاً بر اساس اشل صوتی پنج دستگاه اصلی و با سعی در تفکیک و شناسایی خودکار این دستگاه‌ها از یکدیگر داشته‌اند که با روش‌ها و روش‌های قدیمی و با دادگان‌های محدود و بدون جامعیت صورت گرفته است. [۱] از این رو، ما تلاش کردیم که با مطالعه استفاده از روش‌های پیشرفته‌ی یادگیری ماشینی که امروزه مطرح‌ترین روش‌ها بوده و استفاده‌ی گسترده‌ی آن‌ها در تحلیل و بازیابی اطلاعات انواع مختلف موسیقی در جهان، به تعمیم نتایج نمونه‌های تحقیقات صورت گرفته بر روی دادگان‌های مورد استفاده در پژوهش‌های معتبر این روش‌ها بر دادگانی متشکل از قطعات متنوع، بدون هیچ محدودیتی در تعداد سازها، داشتن یا نداشتن آواز و بدون ایجاد تغییری در عناوین یا ساختار کلی آن‌ها که دارای جامعیت از موسیقی سنتی ایرانی که جمع‌آوری کرده‌ایم، پرداخته و مسئله‌ی تشخیص تعداد و نوع سازهای قطعه به صورت خودکار و یا حتی چگونگی استخراج هر گونه ویژگی مشابهی مبتنی بر محتوای قطعات موسیقی سنتی ایرانی را به کمک این سامانه بررسی کردیم.

در ادامه به تمام جزئیات روش‌های مورد استفاده در سامانه‌ی پیشنهادی، نوع عملکرد سامانه، دادگان مورد استفاده و هرگونه اطلاعات و تعاریف پایه‌ای مرتبط با انواع روش‌های یادگیری، نظریه موسیقی و فعالیت‌های صورت گرفته در این زمینه‌ها به تفصیل پرداخته شده است.

# سپاسگزاری

از جناب آقای دکتر باقر باباعلی که راهنمایی این پروژه را به عهده داشته‌اند و در طول تحصیل نیز همواره بنده را راهنمایی نمودند کمال تشکر را دارم. همچنین از شرکت بیب‌تونز و شرکت نواک که برای همکاری صمیمانه و تسهیل انجام این پروژه نیز بسیار سپاس گزارم.

## پیشگفتار

تشخیص صداها گاهی بسیار ساده و گاهی بسیار پیچیده و دشوار است؛ به طور مثال تشخیص صدای یک مرد و یک زن هر چند که کوتاه باشد، تمایز صدای پیانو و گیتار در بین سازهای غربی و یا تنبک و سنتور، کار ساده ای است اما تشخیص صدای تار و سه تار، انواع سازهای کوبه ای و یا ویولن و کمانچه گاهی برای موسیقیدان ها بسیار دشوار بوده و با تغییر پارامترهایی مانند کوک سازها، سبک نواختن و ساختار فیزیکی سازها و ... میتوان این امر را دشوارتر هم کرد.

در مورد موسیقی و به ویژه موسیقی سنتی ایرانی و ویژگی آن در فصل دوم به تفصیل پرداخته می شود اما به طور کلی، صوت نتیجه ی ارتعاش است و در محیط مادی مانند هوا یا آب به صورت موج انتشار می یابد و ما در دستگاه شنوایی مان آن را با فعل و انفعالات فیزیولوژیکی درک می کنیم. پس صدا از جنس موج بوده و مفاهیمی چون شدت، بلندی، طول موج، بسامد یا فرکانس برای صدا قابل تعریف هستند و چگونگی شنیدن صدایی که ساز تولید میکند هم میتوان اینطور بیان کرد که هنگامی که سیمی از ساز مرتعش می شود و یک نت نواخته شده یا از پرده ای به پرده ای دیگر میرویم و به اصطلاح صدا را زیر و یا بم می کنیم، در واقع فرکانس ارتعاش سیم را تغییر می دهیم و این تغییر در محیط هم منتقل شده و فرکانس به ارتعاش درآوردن استخوان داخل گوش نیز، دقیقا معادل فرکانس ارتعاش سیم ساز می باشد. به همین دلیل ما دقیقا همان صدایی را می شنویم که ساز تولید می کند.

فرض کنید در سازهای مختلف یک نت در این اکتاو را بنوازیم این به این معنی است که فرکانس امواج یا به عبارتی تعداد نوسان های سیم مرتعش، یا ارتعاش صدا، در تمامی اصوات یکسان است، در این صورت باز هم صدا متفاوت خواهد بود و این نقش عامل شکل موج در صدا را مشخص میکند.

شکل امواج با وجود نواختن نت یکسان در یک گام، بسیار متفاوت است بنابراین آنچه موجب تشخیص اصوات از یکدیگر می گردد، شکل موج صداها شنیده شده است. عوامل موثر در شکل موج بسیار پیچیده و زیاد هستند که بحث های فنی عمیق تری را طلب می کند. مثلا آنچه در موسیقی به رنگ صدا، معروف است، دقیقا به عاملی به نام جمع هارمونی ها و شکل موج صوت مربوط می شود و همین تفاوت ها میتواند عامل تشخیص صدا باشد. [۲]

حال با توجه به تنوع بالای سازهای مورد استفاده در موسیقی سنتی ایرانی که بعضا در نواحی

مختلف کشور هم با کوک و سبک ها و ساختار کمی متفاوت نواخته میشوند ، گاهی بسیار مشابه بوده و خصوصا در آثاری که به صورت گروهی نواخته میشوند ، تشخیص تعداد و انواع سازها کاری بسیار دشوار و برای افرادی که با موسیقی و تفاوت سازها آشنا نباشند ، تشخیص آنها می توان گفت که غیر ممکن بوده و دارای درصد خطایی بسیار بالاست .

اگر بخواهیم که از پردازش سیگنال موسیقی به منظور بازیابی محتوای موسیقی بهره ببریم و بدون ساده سازی مسئله و محدود کردن دادگان به قطعات تکنوازی ، برای حل مسئله به روش یادگیری نظارت شده ، به دادگانی با حجم بالا که با جزییات و دقت زیادی برچسب دار شده باشد و از نظر تنوع سازهای مختلف یکدست بوده و تحت تاثیر بایاس خاصی نباشد ، نیاز داریم که تهیه آن بسیار زمانبر ، پرهزینه و دشوار بوده و فرآیند برچسب زنی قطعات هم باید توسط متخصص این زمینه صورت گیرد که میزان خطای انسانی را هم کاهش دهیم که باعث دشوارتر شدن تهیه ی دادگانی مناسب می شود .

ما در این پروژه از SimCLR استفاده کرده و زنجیره بزرگی از داده های صوتی را برای ایجاد یک چارچوب ساده برای یادگیری متضاد و خود نظارتی نمایش های موسیقایی به نام CLMR استفاده می کنیم که یک رویکرد اجرایی بر روی داده های موسیقی خام با نمایش دامنه زمانی است و برای یادگیری بازنمایی های مفید نیازی به برچسب ندارد. ما CLMR را در وظیفه پایین دستی طبقه بندی موسیقی بر روی مجموعه داده های جمع آوری شده ارزیابی می کنیم و مطالعه ای را برای آزمایش میزان موثر بودن نوآوری های مرتبط با موسیقی مورد استفاده ما نسبت به SimCLR ارائه می دهیم . علاوه بر این، نشان می دهیم که نمایش های CLMR با استفاده از مجموعه داده های خارج از دامنه قابل انتقال هستند، که نشان می دهد روش ما تعمیم پذیری قوی در طبقه بندی موسیقی دارد. [۳] در نهایت، ما نشان می دهیم که روش پیشنهادی یادگیری کارآمد داده را در مجموعه داده های برچسب گذاری شده کوچک تر کاهش می دهد و بدین منظور از دادگان نوا که شامل تعداد زیادی قطعات تکنوازی از پنج ساز موسیقی سنتی ایرانی و برچسب دار هست [۴] ، برای آموزش و تعمیم به دادگان جدید استفاده کرده ایم.

## فهرست مطالب

۱	یادگیری ماشین در هوش مصنوعی	۱
۱	۱.۱ شبکه‌ی عصبی و یادگیری عمیق	۱.۱
۲	۱.۱.۱ یادگیری انتها به انتها	۱.۱.۱
۳	۲.۱.۱ داده افزایی	۲.۱.۱
۳	۲.۱ یادگیری با نظارت	۲.۱
۴	۳.۱ یادگیری بی نظارت	۳.۱
۴	۱.۳.۱ یادگیری خود نظارتی	۱.۳.۱
۵	۲.۳.۱ یادگیری متضاد	۲.۳.۱
۵	۴.۱ یادگیری نیمه نظارتی	۴.۱
۶	۵.۱ یادگیری کارآمد از نظر داده	۵.۱
۷	۶.۱ یادگیری بازنمایی	۶.۱
۱۰	۱.۶.۱ یادگیری بازنمایی خود نظارتی	۱.۶.۱
۱۳	۷.۱ شرح چند مفهوم	۷.۱
۱۳	۱.۷.۱ رمزگذار و رمزگشا	۱.۷.۱
۱۵	۲.۷.۱ طرح ریز	۲.۷.۱
۱۶	۳.۷.۱ تابع زیان	۳.۷.۱
۱۷	۲ موسیقی	۲
۱۷	۱.۲ ردیف و دستگاه	۱.۲
۱۸	۲.۲ سازهای موسیقی سنتی ایرانی	۲.۲
۲۲	۳.۲ ارکان موسیقی سنتی	۳.۲
۲۳	۴.۲ بازیابی و استخراج اطلاعات موسیقایی	۴.۲
۲۳	۱.۴.۲ بازیابی اطلاعات در موسیقی سنتی ایرانی	۱.۴.۲
۲۵	۳ مرور کارهای انجام شده دیگران	۳
۲۶	۱.۳ در موسیقی ایرانی	۱.۳

۲۹	سامانه پیشنهادی	۴
۳۱	کلیات روش سامانه	۱.۴
۳۲	داده افزایی	۱.۱.۴
۳۳	ترکیب دسته‌ای	۲.۱.۴
۳۴	رمزگذار	۳.۱.۴
۳۴	تابع زیان متضاد	۴.۱.۴
۳۵	کدگذاری پیشگویانه متضاد	۵.۱.۴
۳۵	ارزیابی خطی	۶.۱.۴
۳۵	بهینه‌سازها	۷.۱.۴
۳۷	شرح دادگان و جمع‌بندی	۵

## فصل ۱

# یادگیری ماشین در هوش مصنوعی

این روزها بسیاری از اتفاقات و کارهای صورت گرفته در اطراف ما بدون دخالت انسانها و توسط ماشینها صورت گرفته و بعضا بازدهی بیش تر و عملکردی دقیق تر از انسانها دارند و مدام در حال بهبود و پیشرفت هم هستند. «آیا ماشین فکر میکند؟» این سوالی بود که آلن تورینگ با مطرح کردن آن در یکی از مقالات خود در سال ۱۹۵۰ نقطه ی آغازی برای پژوهش و فعالیتها در زمینه ی هوش مصنوعی شد و سالها بعد در سال ۱۹۵۹ آرتور ساموئل واژه ماشین لرنینگ را ابداع کرد و پایه ای برای امکان پذیر شدن تمام فعالیت های امروزه یادگیری ماشین شد.

به طور کلی در علم یادگیری ماشین، به موضوع طراحی ماشینهایی پرداخته می شود که با استفاده از مثالهای داده شده به آنها و تجربیات خودشان، بیاموزند. در واقع، در این علم تلاش می شود تا با بهره گیری از الگوریتمها، یک ماشین به شکلی طراحی شود که بدون آنکه صراحتا برنامه ریزی و تک تک اقدامات به آن گفته شود بتواند بیاموزد و عمل کند. در یادگیری ماشین، به جای برنامه نویسی همه چیز، دادهها به یک الگوریتم عمومی داده می شوند و این الگوریتم است که براساس دادههایی که به آن داده شده منطق خود را می سازد. با توجه به تنوع و تفاوتی که داده ها و اهداف ما از یادگیری ماشین دارند، روش های یادگیری متنوع و زیادی با جزئیات، ساختار و اهدافی مختلف وجود دارند که در ادامه به طور خلاصه به توضیح آنها پرداخته میشود.

### ۱.۱ شبکه ی عصبی و یادگیری عمیق

یادگیری عمیق بخشی از روش های یادگیری ماشین است که بر روش هایی تمرکز دارد که مبتنی بر شبکه های عصبی مصنوعی هستند. یادگیری عمیق به رایانه ها می آموزد آنچه را که به طور طبیعی برای انسان انجام می شود، انجام دهند. به عبارتی دیگر، یادگیری عمیق زیرمجموعه ای از یادگیری ماشین است که در آن از الگوریتم هایی استفاده می شود که شبکه ای بسازد که مغز انسان را شبیه



سازی می‌کند؛ در این نوع شبکه گره‌های منفردی وجود دارند که لایه‌ها را تشکیل می‌دهند، دقیقاً مانند نورون‌های مغز ما مناطق مختلف را به هم متصل می‌کنند. به گره‌های یک لایه و یال‌های بین آنها، وزنی به آن‌ها اختصاص داده می‌شود که تأثیری را که پارامتر بر نتیجه پیش‌بینی کلی دارد تغییر می‌دهد. در ساده‌ترین شبکه‌های عصبی تمام داده‌های آموزش در لایه ورودی گرفته شده، داده‌ها را از میان لایه‌های پنهان عبور داده و مقادیر را بر اساس وزن هر گره تغییر شکل می‌دهد. در آخر مقدار آن را در لایه خروجی برمی‌گرداند.

این روش عموماً، به تعداد زیادی از داده‌ها نیازمند بوده و همچنین به قدرت پردازشی بالایی نیاز دارد و واحدهای پردازنده گرافیکی قدرتمندی که در یادگیری عمیق استفاده می‌شوند، هزینه زیادی دارند و به علاوه با تنظیم مناسب یک شبکه عصبی ممکن است مدتی طول بکشد تا نتایج سازگار و موثقی به دست آورید. آزمایش و آموزش شبکه عصبی فرایندی متعادل‌کننده بین تصمیم‌گیری در مورد مهم‌ترین ویژگی‌های مدل است.

واژه‌ی عمیق به عمق یا تعداد لایه‌های پنهان شبکه‌ی عصبی اشاره دارد که شبکه‌های عمیق می‌توانند تا بیش از صد لایه مخفی هم داشته باشند، باید توجه داشت که عمدتاً لایه‌های شبکه عصبی یادگیری عمیق توسط افراد طراحی و ساخته نمی‌شود، بلکه توسط مجموعه داده‌ها و با استفاده از فرایند یادگیری چند منظوره ایجاد شده و توپولوژی آن نیز ممکن است ساده یا بسیار پیچیده باشد؛ هرچه لایه‌های بیشتری در شبکه وجود داشته باشد، یادگیری عمیق ویژگی‌های بیشتری را می‌تواند تشخیص دهد. اما متأسفانه، هرچه لایه‌های بیشتری در شبکه وجود داشته باشد، زمان بیشتری صرف محاسبه آن خواهد شد و تمرین سخت‌تر خواهد بود.

در یادگیری عمیق مراحل مدل‌سازی، طبقه‌بندی داده‌ها و همچنین استخراج ویژگی‌ها به شکل خودکار انجام شده و یادگیری ماشینی همگرا است؛ این نوع یادگیری به روش‌هایی اشاره می‌کند که هنگام افزودن اطلاعات، سطح معینی از عملکرد را به وجود می‌آورد. یکی از ویژگی‌های مثبت این الگوریتم این است که با افزایش اندازه، داده‌ها به پیشرفت خود ادامه می‌دهند.

## ۱.۱.۱ یادگیری انتها به انتها

یادگیری انتها به انتها در زمینه هوش مصنوعی و یادگیری ماشینی تکنیکی است که در آن مدل تمام مراحل بین فاز ورودی اولیه و نتیجه خروجی نهایی را یاد می‌گیرد. این یک فرایند یادگیری عمیق است که در آن همه بخش‌های مختلف به‌جای آموزش متوالی به‌طور همزمان آموزش داده می‌شوند. یک مثال خوب از این تکنیک، ایجاد یک رونوشت (خروجی) نوشته شده از یک کلیپ صوتی (ورودی) ضبط شده است. در اینجا مدل آموزش تمام مراحل را که در این بین اتفاق می‌افتد دور

می‌زند و بر این نکته تأکید می‌شود که بتواند توالی کامل مراحل و وظایف را انجام دهد. [۵] نمونه دیگر، خودروهای خودران هستند. سیستم‌های آنها برای یادگیری و پردازش خودکار اطلاعات با استفاده از یک شبکه عصبی پیچشی (CNN) آموزش دیده‌اند و در این مورد، سیستم‌ها از ورودی‌های انسانی که قبلاً ارائه شده بود به عنوان راهنمایی برای تکمیل وظایف استفاده می‌کنند. [۶]

## ۲.۱.۱ داده افزایی

به طور کلی، افزایش داده‌ها در تجزیه و تحلیل داده‌ها تکنیک‌هایی هستند که برای افزایش حجم داده‌ها با افزودن کپی‌های کمی تغییر یافته از داده‌های موجود یا داده‌های مصنوعی جدید ایجاد شده از داده‌های موجود استفاده می‌شوند. این به عنوان یک منظم‌کننده عمل می‌کند و به کاهش برآزش بیش از حد در هنگام آموزش یک مدل یادگیری ماشینی کمک می‌کند.

در یادگیری ماشین مفهومی به نام خط یا خطوط لوله وجود دارد، منظور از لوله جریان یک طرفه داده است که به صورت چرخه‌ای و تکراری است. اختصاصاً در یادگیری ماشین خطوط لوله این دید را به ما می‌دهند که داده‌ها پیش از جای‌گیری چگونه تنظیم می‌شوند و چگونه استقرار پیدا می‌کنند.

برای انجام داده‌افزایی موثر و مفید باید دانست که در خط لوله یادگیری ماشین خود در چه مرحله‌ای داده‌افزایی را انجام داد، پاسخ بدیهی به نظر می‌رسد؛ این داده‌افزایی قبل از ورود داده‌ها به مدل انجام می‌شود. اما در اینجا دو گزینه پیش‌روست؛ نخست آن است که از قبل، تمام تبدیلات لازم را انجام داده که با این کار، اساساً اندازه پایگاه داده یا دادگان خود را افزایش می‌دهید، دیگر آن است که این تبدیلات را روی یک دسته کوچکی از داده‌ها انجام داده و سپس آن را به مدل یادگیری ماشین خود بدهید. گزینه نخست، تحت عنوان افزایش برون خط خوانده می‌شود. این روش برای پایگاه‌های داده نسبتاً کوچک‌تر ترجیح داده می‌شود، چرا که شما در این روش می‌توانید افزایش اندازه پایگاه داده را با ضربی برابر با تعداد تبدیلاتی که انجام داده‌اید، به پایان برسانید (برای مثال، با قرینه‌سازی تمام تصاویر، اندازه پایگاه داده با ضرب دو افزایش می‌یابد). گزینه دوم، تحت عنوان افزایش برخط یا افزایش در حین اجرا خوانده می‌شود. این روش برای پایگاه‌های داده بزرگتر ترجیح داده می‌شود، چرا که شما نمی‌توانید از عهده افزایش زیاد در اندازه پایگاه داده برآید. بجای آن، شما تبدیلات را روی دسته‌های کوچکی از داده‌ها که می‌خواهید به مدل خود بدهید، اجرا می‌کنید. برخی از چارچوب‌های یادگیری ماشین، افزایش آنلاین را پشتیبانی می‌کنند، که می‌تواند روی GPU تسریع شود. [۷]

## ۲.۱ یادگیری با نظارت

یادگیری تحت نظارت با نظارت یا نظارت شده، زیرمجموعه‌ای از یادگیری ماشین است که در این روش حجم زیادی از داده‌های برجسته زده شده و مدلی که آن را با توجه به شناختی که از داده‌ها و ویژگی آن‌ها وجود داشته است، ساخته شده است در دست داریم و مدل به کمک الگوریتم‌ها و مدل‌های متنوع ریاضیاتی و آماری آموزش داده می‌شود. این روش به طور گسترده در کارهای موسیقایی مانند تشخیص آکورد، تشخیص کلید شمارش ضرب برجسته‌گذاری موسیقی و پیشنهاد موسیقی و حتی در تشخیص دستگاه و ساز در موسیقی سنتی ایرانی [۴] مورد استفاده قرار گرفته است. اما همانطور که گفته شد، در این روش به حجم زیادی از موسیقی‌های برجسته‌دار نیاز داریم که تهیه

آن دشوار، پر هزینه و زمان‌بر است [۳] در حالی که موسیقی خام و بدون برچسب در مقادیر زیادی وجود دارد. علاوه بر آن، هر دادگان یا مسئله مدل مخصوص خود را نیاز دارد و با تغییر جزئی داده‌ها یا پرسش و هدف آموزش مدل را باید تغییر داد و همین امر هم می‌تواند دشوار و زمان‌بر باشد.

## ۳.۱ یادگیری بی نظارت

یادگیری بی نظارت، نظارت نشده یا بدون نظارت هم زیر مجموعه‌ای از یادگیری ماشین است که در این روش دیگر نیاز به داده‌های برچسب‌دار نداشته و با حداقل نظارت انسان به انجام کار خود می‌پردازد. در این روش، مدل در ابتدا با گروه بندی یا خوشه بندی اولیه داده‌ها شروع به کار کرده، با بررسی و برآورد تراکم خوشه‌ها آنها را تغییر داده و همین روند را تا جایی ادامه می‌دهد تا با افزایش و کاهش ابعاد خوشه‌ها به بهترین خوشه بندی برسد و این امر در اکثر اوقات نیازمند به کارگیری الگوریتم‌های پیچیده تری از روش یادگیری با نظارت است. این روش با ممکن ساختن یادگیری از دادگان‌های بزرگ و در دسترس بدون برچسب و با یادگیری نمایش‌هایی با هدف کلی که می‌توانند برای کارهای مختلف پایین‌دستی مورد استفاده مجدد قرار گیرند، تلاش بر از بین بردن محدودیت‌های روش یادگیری با نظارت دارد اما علی‌رغم اهمیت یادگیری بی نظارت برای سیگنال‌های صوتی خام، اما هنوز در یادگیری بی نظارت برای کارهای موسیقایی پیشرفت‌های قابل مقایسه با یادگیری با نظارت مشاهده نشده است. اما نقطه عطف این روش یادگیری در موسیقی، رویکرد خود نظارتی است که اخیراً در حوزه فرکانس زمانی برای وظایف طبقه‌بندی کلی صدا به موفقیت‌های چشمگیری رسیده است [۹-۱۲] اما از یادگیری بازنمایی‌های مؤثر صدای خام به شیوه‌ای بدون نظارت برای کارهای موسیقی‌گریزان باقی مانده است.

### ۱.۳.۱ یادگیری خود نظارتی

یک زیرشاخه و نوعی نسخه بسط یافته یا پیشرفته تر یادگیری بی نظارت است که کارایی داده‌ای بسیار بالایی داشته و بر خلاف یادگیری بی نظارت بر خوشه بندی یا گروه بندی داده‌ها تمرکز نمی‌کند که اخیراً هم پیشرفت چشمگیری را در تحقیقات مختلف نشان داده است. [۱۳-۱۷] این روش از داده‌های ورودی که برچسب‌دار نیستند، برای ارائه‌ی سیگنال‌های نظارت بهره‌بردار کرده و به نوعی سیستم یاد می‌گیرد بخشی از ورودی را از سایر بخش‌های آن ورودی پیش‌بینی کند. پس می‌توان آن را را نسخه مستقلی از یادگیری با نظارت نیز در نظر گرفت زیرا دارای توانایی استفاده از داده‌های برچسب‌دار کم‌تر در هنگام تنظیم دقیق در یک کار خاص نیز هست [۱۶، ۱۸، ۱۹]. از مزایای یادگیری خود نظارتی می‌توان علاوه بر عدم نیاز به مداخله انسان، به قابلیت‌های بهبود یافته آن و مقیاس پذیر بودنش هم اشاره کرد؛ قبل تر دیدیم که بزرگترین مشکل استفاده از روش با نظارت در تهیه داده‌های برچسب‌دار مناسب است، روش یادگیری خود نظارتی حتی با وجود حجم عظیمی از داده می‌تواند این فرایند را به صورت خودکار انجام دهد و پاسخی مناسب به این

مشکل دارد. برای یادگیری خود نظارتی، جایی که حفظ سازگاری داده های زمانی یا فضایی برای نمایش یادگیری های معنادار فرض شده است، سیگنال های خام یا اولیه تبدیل شده و مدل ها با بازسازی و یا زیان (تلفات) متضاد در برابر سیگنال های اصلی بهینه می شوند. به عبارتی دیگر، بدون درستی مبنایی میتواند هیچ تابع زیانی برای آموزش وجود نداشته باشد؛ یادگیری خود نظارتی در عوض از طریق تابع زیان جایگزین آموزش میدهد. یکی از راه های حفظ اطلاعات مفید در طول یادگیری خود نظارتی، تعریف تابع زیان جایگزین با توجه به یک مسئله ابتدایی نسبتاً ساده صورت میگیرد، با این ایده که نمایشی که برای این مسئله خوب باشد، برای کارهای پایین دستی نیز مفید خواهد بود و رویکردهای آگاهانه زیادی برای تولید این کارها و مسائل ابتدایی یا ساده شده وجود دارند [۲۰، ۲۱] و در روش های دیگر، مانند یادگیری بازنمایی متضاد، زیان جایگزین را مستقیماً بر روی بازنمایی های آموخته شده فرمول بندی می کند و بر متضاد نسخه های چندگانه، کمی متفاوت از هر نمونه تکیه می کند که اغلب با استفاده از استراتژی های نمونه گیری منفی [۱۶، ۲۲، ۲۳] یا نهایی کردن چارجوب ها برای نمایش ها انجام میشود. [۱۷]

### ۲.۳.۱ یادگیری متضاد

یادگیری متضاد یک تکنیک یادگیری ماشینی است که برای یادگیری ویژگی های کلی مجموعه داده بدون برچسب با آموزش مدل که نقاط داده مشابه یا متفاوت هستند استفاده می شود. در واقع سعی بر این است که زمان آموزش مدل در یک مسئله، ابتدا به صورت خود نظارتی یک بازنمایی مناسب از داده ها را یاد گرفته و بعد، این مدل از قبل آموزش دیده شده را بر روی داده های برچسب دار به طور دقیق تنظیم کنید؛ کمی قبل تر تا حدی به ایده ی یادگیری بازنمایی به صورت خود نظارتی را بیان کردیم، اما نکته ای که اهمیت دارد این است که اگر چند تابع (عموماً توابع انتقال) داشته باشیم که به کمک آن ها هر داده تا اندکی تغییر دهیم، به طوری که از داده ی اصلی خیلی فاصله نگیرد، و تنها از یک دید دیگر به همان داده بنگرد، می توان یک تابع زیان متضاد تعریف کرد، با تلاش به کمینه سازی فاصله بازنمایی داده اصلی و بازنمایی دید آن. به طور مثال، در حوزه تصویر، توابعی مانند چرخش تصویر، معکوس جانی و را در نظر می گیرند. پس تعریف درست این توابع بسیار به دانش پایه ای از آن حوزه کاری نیاز دارد. [۲۴]

### ۴.۱ یادگیری نیمه نظارتی

یادگیری نیمه نظارتی یا نیمه نظارت شده هم نوعی دیگر از روش های یادگیری ماشینی است که میتوان گفت ترکیبی از یادگیری بی نظارت و با نظارت می باشد؛ این روش از مقدار کمی داده ی برچسب دار و مقدار زیادی داده ی بدون برچسب استفاده کرده و در روش یادگیری نیمه نظارتی در ابتدا، مدل تا جایی با مقدار اندکی داده ی برچسب دار آموزش داده می شود که نتایج مطلوبی به دست آید و بعد از مدل حاوی مجموعه داده آموزش بدون برچسب یا برچسب ساختگی برای پیش بینی خروجی استفاده

شده و برچسب‌های حاصل از داده‌های آموزش برچسب‌دار با برچسب‌های ساختگی و همچنین، ورودی داده‌های حاصل از داده‌های آموزش برچسب‌دار با ورودی‌های موجود در داده‌های بدون برچسب جمع و ترکیب شده و نهایتاً، به همان ترتیبی که در مجموعه داده‌ی کاملاً برچسب‌دار عمل شد، مدل آموزش داده می‌شود. یکی از رویکردهای محبوب در یادگیری نیمه نظارت شده این است که الگوریتم‌های خوشه‌بندی و طبقه‌بندی را ادغام کنیم. این الگوریتم‌ها نقش موثری در شناسایی مرتبط‌ترین نمونه‌ها در مجموعه داده دارند. می‌توان به این نمونه‌ها برچسب زد و برای آموزش مدل یادگیری نبا نظارت در کار طبقه‌بندی استفاده کرد.

## ۵.۱ یادگیری کارآمد از نظر داده

یادگیری کارآمد داده نوعی یادگیری ماشینی است که می‌تواند دامنه‌های پیچیده را بدون نیاز به حجم زیاد داده درک کند. الگوریتم‌های یادگیری ماشینی سنتی معمولاً برای رسیدن به نتایج منطقی به داده‌های بزرگ متکی هستند. با نیاز به حجم زیادی از داده‌ها، آنها می‌توانند روندها و اشتراکاتی را که وجود دارد، تشخیص دهند. در مقایسه، یادگیری کارآمد داده از مجموعه داده‌های کوچک‌تری با هدف یادگیری سریع‌تر و در عین حال حفظ همان سطح عملکرد در مقایسه با سایر روش‌های یادگیری ماشینی استفاده می‌کند. این امر به ویژه برای دامنه‌هایی با مقادیر محدود داده، مانند مراقبت‌های بهداشتی شخصی و تجزیه و تحلیل احساسات، مهم می‌شود.

یادگیری ماشینی کارآمد داده اغلب از روش یادگیری تقویتی استفاده می‌کند. این بدان معناست که این الگوریتم‌ها از سیستم پاداش و تنبیه به عنوان بخشی از برنامه نویسی خود استفاده می‌کنند. به جای برنامه ریزی برای تکمیل یک کار، الگوریتم‌های تقویت‌کننده به تنهایی مشکل را حل می‌کنند. برنامه می‌داند که می‌خواهد به حالت «پاداش» برسد، مانند شناسایی صحیح یک احساس، نه به حالت «مجازات»، و ناتوانی در شناسایی درست یک احساس. همانطور که از طریق مشکلات موجود کار می‌کند، یاد می‌گیرد که چگونه تجزیه و تحلیل را در طول انجام آن بهبود بخشد. یادگیری تقویتی شبیه روش‌هایی است که کودکان یاد می‌گیرند چگونه اشیاء را شناسایی کنند، مانند تشخیص اینکه یک ماشین آتش‌نشانی یک ماشین آتش‌نشانی است زیرا یک وسیله نقلیه بزرگ و قرمز است.

یادگیری کارآمد داده می‌تواند از راه‌های زیر خصوصاً در کارهای تجاری، بسیار مفید باشد از جمله:

پردازش سریع داده‌ها: به جای صرف زمان زیادی برای تجزیه و تحلیل مقادیر زیادی از داده‌ها، فرآیندهای کارآمد داده می‌توانند به سرعت مقادیر کمی از داده‌ها را ارزیابی کرده و نتایج و پیش‌بینی‌های مرتبط را برگردانند.

بدون نیاز به کلان داده: در روش‌های یادگیری ماشینی سنتی، قبل از انجام پیش‌بینی‌های دقیق، به مقدار زیادی داده نیاز است. این می‌تواند به این معنی باشد که داده‌ها فقط تا زمانی که به اندازه کافی برای استفاده موثر جمع‌آوری نشده‌اند در فضای ذخیره‌سازی قرار می‌گیرند. در مقایسه،

یادگیری کارآمد می تواند تجزیه و تحلیل داده ها و بازگرداندن پیش بینی ها را بسیار سریعتر آغاز کند.

حفاظت از حریم خصوصی داده ها : از آنجایی که برای این فرآیندها فقط به مقدار کمی داده نیاز است، نیازی به ذخیره مقادیر زیادی از داده ها نیست و بی جهت خطر نقض داده ها را تهدید می کند. علاوه بر این، نیاز کمتری به داده های ناشناس وجود دارد، زیرا داده های کمتری لازم است. این بیشتر خطر به خطر افتادن اطلاعات خصوصی را به عنوان بخشی از فرآیند تجزیه و تحلیل کاهش می دهد.

## ۶.۱ یادگیری بازنمایی

این روش ، رویکردی از یادگیری ماشین است که امکان کشف بازنمایی های مورد نیاز برای تشخیص یا طبقه بندی ویژگی را از داده های خام به دستگاه یا سامانه داده و به این صورت نیاز به مهندسی ویژگی های دستی و دخالت مستقیم انسان را با اجازه دادن به ماشین برای یادگیری ویژگی ها و اعمال آن ها در یک فعالیت معین کاهش می دهد.

در یادگیری بازنمایی، داده ها به سامانه فرستاده می شوند و آن به تنهایی نمایش یا بازنمایی داده ها را یاد می گیرد. این روشی برای تعیین نمایشی داده ای از ویژگی ها، توابع فاصله و توابع شباهت است که نحوه عملکرد مدل پیش بینی را مشخص می کنند.

روش های مختلفی برای یادگیری بازنمایی های مختلف وجود دارد. برای مثال،

- در مورد مدل های احتمالی، هدف یادگیری نمایشی است که توزیع احتمال ویژگی های بیانگر اطلاعات اساسی را برای ورودی مشاهده شده به تصویر کشیده ، سپس می توان از چنین نمایش آموخته ای برای پیش بینی استفاده کرد.
- در یادگیری عمیق، نمایش ها با ترکیب تبدیل های غیرخطی متعدد داده های ورودی با هدف ارائه نمایش های انتزاعی و مفید برای کارهایی مانند طبقه بندی، پیش بینی و غیره شکل می گیرند.

داده های موجود دارای ساختاری به نام بعد هستند که داده های ورودی را با توجه به معیارها و ویژگی های مورد نظر ما دسته بندی می کنند ؛ فرض کنید به ازای هر داده ، برداری برای ذخیره سازی آن در نظر گرفته شده است که افزایش هر بعد به معنای افزایش مولفه ای به بردار بوده و به این ترتیب هم باعث افزایش فضای مورد نیاز برای ذخیره سازی آن شده و هم عددی نامنفی به مقادیر توابع فاصله اضافه شده ، که باعث افزایش حجم نمونه و محاسبات و کاهش سرعت در فعالیت سامانه می شود .

یادگیری بازنمایی با کاهش داده هایی که ابعاد بزرگی دارند به داده های کم بعد کار کرده و به این ترتیب کشف الگوها و ناهنجاری ها را آسان تر می کند و در عین حال درک بهتری از رفتار کلی

داده‌ها ارائه می‌دهد.

اساساً، وظایف یادگیری ماشینی مانند طبقه‌بندی، اغلب ورودی‌هایی را می‌طلبند که از نظر ریاضی و محاسباتی برای پردازش راحت باشد، که باعث ایجاد انگیزه در یادگیری بازنمایی می‌شود. برای بررسی و پردازش بر روی داده‌های دنیای واقعی، مانند عکس‌ها، ویدیوها و داده‌های حسگر، تعریف مشخصی از کیفیت مناسب وجود نداشته و همواره تلاش در افزایش کیفیت‌های سابق بوده و رویکردی که مناسب به نظر می‌رسد، بررسی داده‌ها برای چنین ویژگی‌ها یا بازنمایی‌هایی به جای وابستگی به تکنیک‌های صریح است.

دلیل استفاده از یادگیری بازنمایی، کسب اطمینان از ارائه‌ی نتایج ثابت از مدل است تا دقت آن را افزایش داده و عملکرد آن را بهبود بخشیم. در این بخش، ما به چگونگی یادگیری بازنمایی می‌توانیم عملکرد مدل را در چند یادگیری مختلف بهبود بخشیم:

### یادگیری عمیق

در یادگیری عمیق، یادگیری بازنمایی نتیجه عملکردی است که یک مدل یاد می‌گیرد که در آن یادگیری در مولفه یا متغیرهای مدل ثبت می‌شود، زیرا تابع در طول آموزش ورودی را به خروجی تبدیل می‌کند. یادگیری بازنمایی در اینجا به ماهیت یا ویژگی‌های ورودی تبدیل شده اشاره دارد و نه به متغیرهای مدل یا تابعی که باعث آن است. نقش غیررسمی هم توسط معماری و ساختار مدل و هم متغیرهای آموخته شده (مثلاً این که آیا یک متغیر در نمایش بخشی یا تمام ورودی نقشی ایفا می‌کند) در نگاشت ورودی به خروجی ایفا می‌شود.

### یادگیری با نظارت

بیان شد که هنگامی که مدل ورودی  $X$  را به خروجی  $Y$  ترسیم می‌کند، به این یادگیری با نظارت گفته می‌شود و در این روش از داده‌های برچسب‌گذاری شده استفاده می‌شود و کامپیوتر سعی می‌کند با مقایسه خروجی مدل با دانش پایه‌ای که به آن داده شده، خود را اصلاح کند و فرآیند یادگیری، نگاشت را از ورودی به خروجی بهینه می‌کند. این فرآیند تا زمانی که تابع بهینه‌سازی به حداقل‌های جهانی برسد تکرار می‌شود. اما حتی زمانی که تابع بهینه‌سازی به حداقل‌های جهانی می‌رسد، داده‌های جدید همیشه خوب عمل نمی‌کنند و در نتیجه بیش از حد برازش می‌شوند. در حالی که یادگیری با نظارت برای یادگیری نگاشت از ورودی به خروجی به مقدار قابل توجهی داده نیاز ندارد، باعث اهمیت نوع ویژگی‌های آموخته شده می‌شود. هنگامی که ویژگی‌های آموخته شده در الگوریتم یادگیری با نظارت گنجانده شوند، دقت پیش‌بینی به مقدار قابل توجهی بهبود می‌یابد.

به طور مثال در شبکه‌های عصبی که از «شبکه» گره‌های به هم پیوسته در لایه‌های مختلف استفاده می‌کنند، شبکه قوانین محاسباتی را برای انتقال داده‌های ورودی از لایه ورودی شبکه به لایه خروجی شبکه ایجاد می‌کند و هر یال دارای یک وزن است که به نوعی نشان‌دهنده ارزش آن است. رابطه بین لایه‌های ورودی و خروجی، که توسط وزن‌ها مشخص و مقداردهی می‌شود، توسط یک تابع

شبکه مرتبط با یک شبکه عصبی توصیف می‌شود. وظایف یادگیری مختلف را می‌توان با به حداقل رساندن یک تابع هزینه بر روی تابع شبکه با توابع شبکه‌ای که به درستی تعریف شده‌اند، به دست آورد.

## یادگیری بی نظارت

یادگیری بی نظارت نوعی یادگیری ماشینی است که در آن برجسب‌ها به منظور توجه به مشاهدات نادیده گرفته می‌شود. این روش برای طبقه بندی یا رگرسیون مورد استفاده نبوده و هدف اصلی استفاده از آن، مواردی مانند کشف الگوهای زیربنایی، خوشه‌بندی داده‌ها، حذف نویز، تشخیص نقاط پرت، و تجزیه داده‌ها و موارد دیگر است.

هنگام کار با داده‌ها، باید در مورد ویژگی‌هایی که استفاده می‌کنیم بسیار مراقب باشیم تا مطمئن شویم الگوهای تولید شده دقیق هستند. مشاهده شده است که داشتن داده‌های بیش‌تر لزوماً به معنای داشتن نمایشی بهتر از آن‌ها نیست. ما باید مدلی را توسعه دهیم که هم انعطاف پذیر و هم گویا باشد تا ویژگی‌های استخراج شده بتواند اطلاعات حیاتی را منتقل کند.

یادگیری ویژگی‌های بی نظارت ویژگی‌ها را از داده‌های ورودی بدون برجسب با پیروی از روش‌هایی مانند یادگیری فرهنگ لغت، تجزیه و تحلیل مؤلفه‌های مستقل، رمزگذارهای خودکار، فاکتورسازی ماتریس و اشکال مختلف خوشه بندی، یاد می‌گیرد.

یادگیری بازنمایی بدون نظارت اغلب به دنبال کشف ویژگی‌های کم‌بعدی است که ساختاری را در زیر داده‌های ورودی با ابعاد بالا محصور می‌کند.

به طور مثال، خوشه‌بندی K-means یک رویکرد کوانتیزاسیون برداری است. یک مجموعه  $n$  بردار از طریق خوشه‌بندی K-means به  $k$  خوشه یا همان زیرمجموعه، تقسیم می‌شود که هر بردار متعلق به خوشه‌ای با نزدیک‌ترین میانگین است. با وجود استفاده از روش‌های حریصانه، مسئله‌ای از نظر محاسباتی NP-hard است.

خوشه‌بندی K-means مجموعه‌ای از ورودی‌ها را قبل از به‌دست آوردن ویژگی‌های مبتنی بر مرکز، به گروه‌های  $k$  تقسیم می‌کند. این ویژگی‌ها را می‌توان به روش‌های مختلفی مشخص کرد. ساده‌ترین روش افزودن  $k$  ویژگی دوگانه به هر نمونه است، با هر ویژگی  $z$  دارای مقدار یکی از k-means آموخته‌شده  $z$  امین مرکز نزدیک‌ترین نقطه به نمونه مورد بررسی است. پس از پردازش با تابع پایه شعاعی می‌توان از فواصل خوشه‌ای به عنوان ویژگی استفاده کرد.

پس تا اینجا متوجه شدیم که برخلاف وظایف یادگیری معمولی مانند طبقه‌بندی، که هدف نهایی کاهش طبقه‌بندی‌های اشتباه است، یادگیری بازنمایی یک هدف میانی یادگیری ماشینی است که برای آن بیان یک هدف آموزشی مستقیم و واضح را دشوار است. اما اهمیت این روش در چیست؟

عملکرد هر مدل یادگیری ماشینی به شدت به نمایش‌هایی که می‌آموزد به خروجی بستگی دارد. نمایشی که یاد می‌گیرد به نوبه خود به طور مستقیم به مدل و آنچه که به عنوان ورودی تغذیه می‌



شود بستگی دارد که می تواند داده های خام یا خروجی یک مدل بالادستی (یادگیری انتقال) باشد که داده های ورودی خام را تغییر می دهد. درک انواع مختلف بازنمایی ها، از جمله این که چه چیزی یک نمایش خاص را برای یک کار خاص بهتر می کند، به متخصصان کمک می کند از یک دیدگاه منسجم تر و گسترده از ساختارهای مختلف مدل یادگیری خصوصا از نوع مدل یادگیری عمیق بهره مند شوند.

## ۱.۶.۱ یادگیری بازنمایی خود نظارتی

پیش تر در مورد معایب یادگیری نظارت شده ، که اکثرا مربوط به دشواری ها و چالش های تهیه و استفاده ی دادگان های مناسب ، در آن هاست ، اشاره کردیم که انگیزه را برای استفاده از روش های نوینی مانند یادگیری خود نظارتی که تا حدی این مشکلات را برطرف می کنند ، بیش تر کردند و مختصرا هم در مورد این روش توضیح دادیم . در این بخش با جزئیاتی بیش تر به شرح آن به عنوان یک روش یادگیری بازنمایی و بیان ویژگی هایش می پردازیم.

خود نظارتی ، به وظایف یادگیری اشاره دارد که از یک شبکه عصبی عمیق ( شبکه ای با تعداد لایه های بالا ) می خواهد بخشی از داده های ورودی را پیش بینی کرده و یا برچسبی که از نظر پیاده سازی قابل استخراج از آن است را با توجه و با استفاده از بخشی دیگر از ورودی به آن نظیر کند . این برخلاف یادگیری با نظارت است که از شبکه پیش بینی خروجی هدف ارائه شده دستی را می خواهد. الگوریتم های خود نظارتی در درجه اول در استراتژی خود برای تعریف برچسب های مشتق شده برای پیش بینی متفاوت هستند ؛ عمدتا به این صورت عمل می کنند که با طرح و حل مسئله یا مسائلی ساده تر به عنوان بهانه ای برای شروع فرآیند آموزش استفاده کرده و انتخاب این مسئله ، تفاوت ها را در نتایج بازنمایی آموخته شده حاصل را تعیین می کند. بنابراین همین انتخاب بسیار در در انجام کارهای پایین دستی ، یعنی دقیق همان هدف اصلی ما از انجام یادگیری ، مختلف موثر است. این استراتژی ها با موفقیت برای بهبود کارایی نمونه یادگیری در روش های مختلف از تصویر [۱۷، ۱۹] ، ویدیو [۲۵، ۲۶] ، گفتار [۱۴] ، متن گراف ها به کار گرفته شده اند. این روش ها را می توان برای تقویت وظایف مختلف پایین دستی از جمله نه تنها تشخیص ساده ، بلکه همچنین تشخیص و محلی سازی پیش بینی متراکم (تبدیل سیگنال) تشخیص ناهنجاری و غیره به کار برد .

علاوه بر این، برخی از نتایج نشان می دهند که کیفیت بازنمایی خود نظارتی حتی تابعی لگاریتمی از مقدار داده های پیش آموزشی بدون برچسب است و اگر این روند ادامه داشته باشد ، بهبود عملکرد آن در طول زمان به صورت رایگان قابل دستیابی و ممکن است و این به معنی حل مشکلی است که آن را به عنوان مشکل اصلی روش یادگیری با نظارت مطرح کرده بودیم ، زیرا جمع آوری داده ها و قدرت محاسباتی بهبود یافته و امکان استفاده از مجموعه های پیش آموزش بزرگ را بدون نیاز به حاشیه نویسی دستی داده های جدید و برچسب گذاری آن ها را فراهم می کند. [۲۷]

در ادامه مختصرا نمادهای لازم برای تعریف مسئله یادگیری بازنمایی خود نظارتی را معرفی می

کنیم و آن را با سایر نمونه‌های یادگیری رایج مقایسه می‌کنیم. یادگیری با نظارت به یک مجموعه داده برچسب‌گذاری شده برای مسئله‌ی هدف که می‌خواهیم حل کنیم، نیاز دارد و یک مدل پیش‌بینی می‌سازد که تخمین می‌زند؛ در زمینه یادگیری عمیق، مدل پیش‌بینی معمولاً از یک تابع استخراج‌کننده بازنمایی مانند  $h\theta$  و یک تابع طبقه‌بندی تشکیل شده است، ما این مدل پیش‌بینی را با به حداقل رساندن یک تابع زیان  $L$  به سیستم منفی آموزش می‌دهیم.

با این حال  $h\theta$  ممکن است صدها میلیون متغیر داشته باشد، که به میلیون‌ها نقطه داده برچسب‌گذاری شده در دادگان نیاز دارد تا به درستی برازش شود. این میلیون‌ها نقطه داده مشروح در اکثر شرایط اصلاً در دسترس نیستند، اما بسیاری از آنها اساساً منبع رایگانی از نقاط داده بدون برچسب دارند. به عنوان مثال، تعداد زیادی از داده‌های سیگنال صوتی خام  $x$  در مقابل مقدار محدود داده‌های گفتار رونویسی شده  $y$  در تشخیص گفتار را در نظر بگیرید.

روش‌های یادگیری بی نظارت اغلب از چنین داده‌های بدون برچسب با ساختن مدل‌های متفاوت یاد می‌گیرند. سایر رویکردهای رایج بدون نظارت مانند رمزگذارهای خودکار و خوشه‌بندی بازنمایی‌های نهفته‌ی فشرده را یاد می‌گیرند.

برای مثال رمزگذارهای خودکار اغلب یک هدف بازسازی را بهینه می‌کنند؛ به صورتی که تابع استخراج‌کننده بازنمایی یک ویژگی فشرده را از ورودی استخراج می‌کند و تابع طبقه‌بندی از آن برای بازسازی ورودی اصلی استفاده می‌کند.

آموزش بازنمایی خود نظارتی را می‌توان به عنوان یک مورد خاص از یادگیری بدون نظارت دید، زیرا هر دو روش بدون حاشیه‌نویسی یاد می‌گیرند، در حالی که روش‌های مرسوم بی نظارت بر اهداف بازسازی یا برآورد چگالی تکیه دارند، روش‌های بازنمایی خود نظارتی بر وظایف بهانه‌ای تکیه می‌کنند که از دانش در مورد روش داده‌های مورد استفاده برای آموزش بهره‌برداری می‌کنند. اگرچه روش‌های یادگیری با نظارت تمایل به یادگیری ویژگی‌های قوی‌تری نسبت به روش‌های یادگیری بی نظارت دارند، اما برای ایجاد برچسب‌های مورد نیاز، به کار پرهزینه و زمان‌بر از حاشیه‌نویس‌های انسانی نیاز دارند. هدف تکنیک‌های یادگیری بازنمایی خود نظارتی، بهترین در هر دو فضا است:

- آموزش استخراج‌کننده ویژگی قدرتمند با استفاده از یادگیری متمایز

- بدون نیاز به حاشیه‌نویسی دستی نمونه‌های آموزشی

شناسه و معرف یک روش خود نظارتی، وظیفه یا مسئله‌ی بهانه‌ای آن است که از یک فرآیند  $P$  برای تولید شبه‌برچسب‌ها و هدفی برای هدایت یادگیری تشکیل شده است. با توجه به مجموعه داده‌های خام، فرآیند بهانه‌ای، شبه‌برچسب‌های  $z$  و نقاطی از داده‌هایی که با احتمال مشخصی تغییر یافته‌اند، ایجاد می‌کند.

به عنوان مثال، بخشی از یک سیگنال گفتاری  $x$  را می‌توان با پوشاندن بخشی از سیگنال تغییر داد، و شبه‌برچسب  $z$  به عنوان بخش پوشانده شده از ورودی تعریف می‌شود. سپس می‌توان یک

شبکه عصبی را با هدف پیش‌بینی بخش گمشده  $z$  با توجه به  $x$  تا حدی پوشانده شده، آموزش داد. بیش‌تر فعالیت‌های تحقیقاتی خود نظارتی به استخراج وظایف بهانه‌ای  $P$  می‌پردازد، که یادگیری بازنمایی‌های هدف کلی  $h\theta$  را که کارایی بالا و یادگیری کارآمد از نظر داده را از وظایف پایین‌دستی امکان‌پذیر می‌سازد. نحوه‌ی کار خود نظارتی مختصراً به شرح زیر است:

۱. داده‌های حاشیه‌نویسی شده یا برچسب‌دار برای وظیفه هدف، دادگان  $D_i$  را تشکیل داده و مجموعه‌ی تمام داده‌های بدون برچسب موجود،  $D_s$  که از نظر اندازه هم بزرگ‌تر است را، تشکیل می‌دهند.

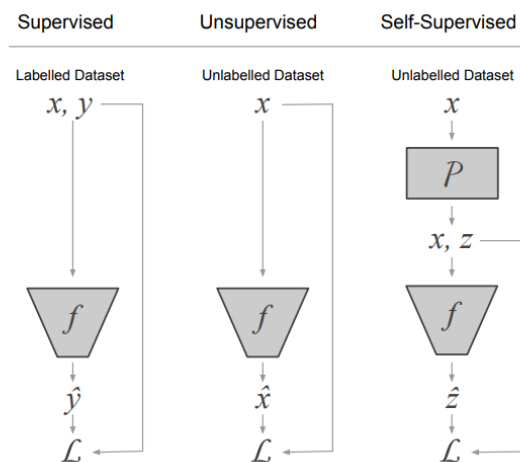
۲. وظیفه‌ی بهانه‌ای دادگان جدیدی از داده‌های شبه‌برچسب‌دار را، با توجه به اطلاعاتی که پیش‌تر قرارداد کردیم، به این صورت تولید می‌کند  $\bar{D}_s = \{x_i, z_i\} = P(D_s)$ . باید توجه داشت از آن جایی که فرآیند  $P$  اغلب به متغیرهای تبدیل نمونه یا پوشش آن‌ها بستگی دارد، معمولاً در شروع هر دوره از آموزش، تکرار می‌شود.

۳. مدل بهانه‌ای، برای بهینه‌سازی در هدف خود نظارتی در  $D_s$  آموزش داده شده است، مهم‌تر این‌که، این تخمین خوب  $\theta^*$  را از میان تعداد فراوانی متغیر در  $h\theta$  ارائه می‌دهد، اما بدون نیاز به برچسب حاشیه‌نویسی. در بسیاری از موارد ورودی  $x_i$  یک داده نقطه‌ای است و شبه‌برچسب  $z_i$  یک برچسب کلاس با مقدار اسکالر است. با این حال، ورودی  $x_i$  بالا می‌تواند شامل چندین نقطه داده با شبه‌برچسب  $z_i$  باشد که توضیح دهد شبکه چگونه باید این نقاط داده را به هم مرتبط کند. به طور مشابه، در پیش‌بینی تبدیل، ورودی  $x_i$  می‌تواند از چند تکه به هم ریخته تشکیل شده باشد در حالی که شبه‌برچسب  $z_i$  ترتیب به هم ریخته را به ترتیب اصلی مرتبط می‌کند.

۴. تابع خروجی بهانه‌ای کنار گذاشته می‌شود و تابع بازنمایی  $h\theta^*$  به عنوان یک راه حل جزئی برای حل مشکل هدف مورد علاقه با استفاده از مدل منتقل می‌شود. بسیار مهم است، زمانی که متغیرهای بازنمایی  $\theta^*$  از قبل به خوبی از مرحله خود نظارتی در معادله ۳ برازش شده باشند، ممکن است فقط تعداد کمی از متغیرها نیاز به یادگیری یا اصلاح برای حل مشکل هدف داشته باشند، بنابراین می‌توان آن را با یک برچسب کوچک حل کرد. مجموعه داده هدف  $D_i$  دو راه متداول برای حل مسئله هدف با استفاده از  $\theta^*$  وجود دارد - تنظیم دقیق و بازخوانی خطی

نمایش بالا فرض می‌کند که وظیفه هدف برچسب‌گذاری شده و با یادگیری با نظارت آموزش داده شده است، زیرا این رایج‌ترین مورد استفاده است. با این حال، وظایف هدف بدون برچسب مانند خوشه‌بندی یا بازیابی نیز در صورت جایگزین شدن در مرحله ۴ به وضوح می‌توانند از پیش‌آموزش خود نظارتی سود ببرند. [۲۷]

این روش دارای جزئیات زیادی است و صرفاً به کلیات آن در این قسمت پرداختیم. در ادامه هم کمی مفصل‌تر به روش استفاده از آن در سامانه‌ی پیشنهادی پرداخته خواهد شد.



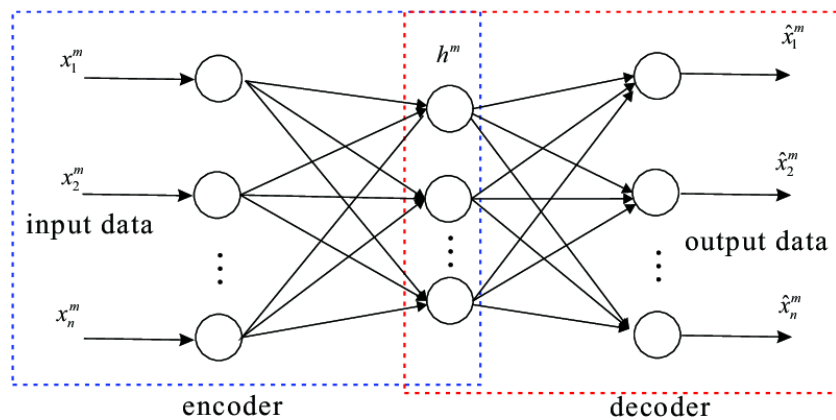
شکل ۱.۱: مقایسه نمونه یادگیری‌های با نظارت، بی نظارت و خود نظارتی برای آموزش مدل  $f$  با استفاده از داده‌های خام  $x$ ، برچسب‌های  $y$  و تابع زیان  $L$ . روش‌های خود نظارتی وظیفه بهانه‌ای  $P$  را معرفی می‌کنند که شبه‌برچسب‌های  $z$  را برای آموزش  $f$  ایجاد می‌کنند. [۲۷]

## ۷.۱ شرح چند مفهوم

روش مورد استفاده در سامانه‌ی پیشنهادی، همان‌طور که در فصل مربوطه خواهید دید، دارای ۴ مرحله اصلی است که در مورد مفهوم مورد استفاده در مرحله اول آن یعنی داده افزایی، قبل‌تر صحبت کردیم و در ادامه این بخش، به توضیح چند مفهوم مورد استفاده در مراحل دیگر آن به ترتیب می‌پردازیم:

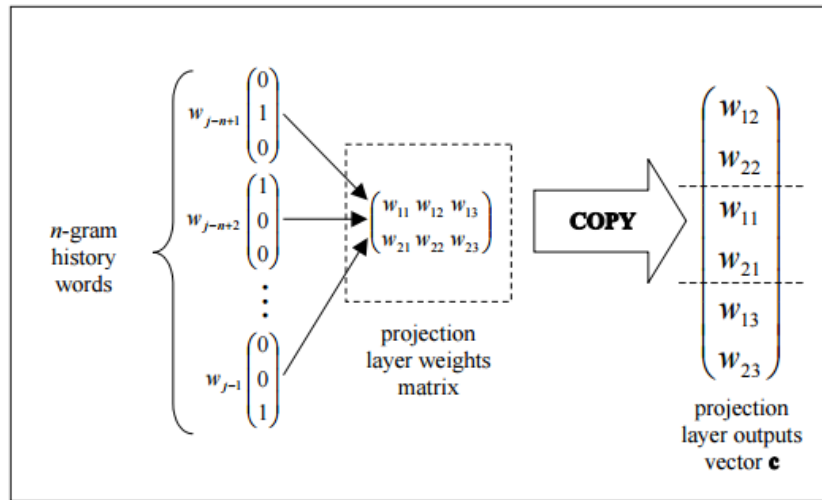
### ۱.۷.۱ رمزگذار و رمزگشا

رمزگذار شبکه (عصبی) ای است که ورودی را به هر شکل و نمایی که دریافت می‌کند و آن را به صورت نقشه ویژگی، بردار یا هر نمایی که سامانه به آن نیاز داشته باشد و در انجام پردازش‌ها راحت‌تر و مناسب‌تر باشد، خروجی می‌دهد. این بردارهای ویژگی، اطلاعات و یا ویژگی‌هایی که معرف ورودی هستند را، دربرمی‌گیرد. رمزگشا هم یک شبکه است (معمولاً همان ساختار شبکه با رمزگذار اما در جهت مخالف) که بردار ویژگی را از لایه‌های بعد از رمزگذار و پس از انجام پردازش‌ها می‌گیرد و بهترین تطابق را با ورودی واقعی یا خروجی مورد نظر می‌دهد. رمزگذارها با رمزگشاها بی‌هیچ برچسبی آموزش داده می‌شوند (به صورت بی نظارت). تابع زیان بر اساس محاسبه تغییرات بین ورودی واقعی و بازسازی شده است. بهینه‌ساز سعی خواهد کرد هم



شکل ۲.۱: لایه‌های رمزگذار و رمزگشا و لایه‌های پنهان میان آن‌ها در شبکه عصبی عمیق

رمزگذار و هم رمزگشا را آموزش دهد تا این زیان و تلفات بازسازی را کاهش دهد. پس از آموزش، رمزگذار بردار ویژگی را برای ورودی می‌دهد که می‌تواند توسط رمزگشا برای ساخت ورودی با ویژگی‌هایی که بیش‌ترین اهمیت را دارند تا ورودی بازسازی شده به عنوان ورودی واقعی قابل تشخیص باشد، استفاده کند. همین تکنیک در کاربردهای مختلف مانند ترجمه، مدل‌های تولیدی و غیره استفاده می‌شود. دانستن این نکته مهم است که در کاربرد واقعی، تلاش در بازسازی ورودی واقعی نیست، بلکه می‌خواهند ورودی‌ها را به خروجی‌های خاصی نگاشت دهند، ترجمه کنند و یا ارتباط دهند.



شکل ۳.۱: این شکل نشان می دهد که چگونه خروجی لایه طرح ریزی می تواند به طور موثر با رونوشت کردن ستون‌ها از ماتریس وزن‌های لایه طرح ریزی جمع آوری شود.

### ۲.۷.۱ طرح‌ریز

لایه طرح‌ریزی، زیرنمایه‌های گسسته یک دنباله به طول  $n$  را به یک فضای برداری پیوسته نگاشت می‌کند. لایه طرح‌ریزی به گونه‌ای است که برای متن‌هایی که چندین بار حاوی یک کلمه هستند، مجموعه وزن‌های یکسانی اعمال می‌شود تا هر قسمت از بردار طرح ریزی را تشکیل دهد. این سازمان‌دهی به طور موثر مقدار داده‌های موجود را برای آموزش وزن‌های لایه طرح‌ریزی افزایش می‌دهد زیرا هر کلمه از هر الگوی آموزشی متن به طور جداگانه تغییراتی در مقادیر وزن ایجاد می‌کند.

لایه پنهان خروجی لایه طرح ریزی را پردازش می‌کند و با تعدادی نورون مشخص شده در پیکربندی توپولوژی، ایجاد می‌شود. هر نورون با تعدادی وزن برابر با اندازه واژگان نشان داده می‌شود. لایه طرح‌ریزی با استفاده نکردن از یک تابع فعال سازی غیرخطی با لایه‌های مخفی و خروجی متفاوت بوده و هدف آن صرفاً ارائه یک ابزار کارآمد برای نمایش دنباله داده شده بر روی یک فضای برداری پیوسته کاهش یافته برای پردازش بعدی توسط لایه های مخفی و خروجی است که برای طبقه بندی چنین بردارهایی آموزش دیده اند.

با توجه به ماهیت یک و صفری عناصر بردار ورودی، خروجی یک کلمه خاص با زیرنمایه  $i$  صرفاً ستون  $i$  ام ماتریس آموزش دیده وزن‌های لایه طرح‌ریز، که در آن هر ردیف از ماتریس وزن نورون را نشان می‌دهد، است.

### ۳.۷.۱ تابع زیان

تابع ضرر تابعی است که فاصله بین خروجی جریان الگوریتم و خروجی مورد انتظار را محاسبه می کند و چگونگی عملکرد الگوریتم ما برای مدل کردن داده ها را ارزیابی می کند. یک تابع زیان تصمیمات را به هزینه های مرتبط آنها ترسیم می کند. در واقع اندازه گیری به عنوان یک سیگنال بازخورد برای تنظیم روش کار الگوریتم استفاده می شود و این مرحله تنظیم ، همان چیزی است که ما آن را یادگیری می نامیم.

## فصل ۲

### موسیقی

موسیقی دانش و هنری گسترده است که پیدایش، رشد و توسعه آن قدمتی به درازای عمر انسان داشته و در بسیاری از فرهنگ ها بخش مهمی از شیوه زندگی مردم بوده و هست. اجزای تشکیل دهنده ی موسیقی، صدا و سکوت هستند که تنظیم صداها در زمان با توجه به عناصر نواک یا زیر و بمی (تعیین کننده ی ملودی و هارمونی) و ضرب آهنگ یا ریتم به طوری که احساسی را منتقل کند، تعریف اولیه و عمومی موسیقی است. هنرهای زیبا، نمایانگر احساسات و روحیات هر قومی است و موسیقی اقوام و کشورهای مختلف نیز که تحت تأثیر احساسات و نحوه زندگی مردم است، تفاوت‌های زیادی با یکدیگر دارند پس یکی از دیدگاه‌های تقسیم بندی موسیقی، ناحیه پیدایش است. موسیقی سنتی ایرانی یک دسته مطرح در تقسیم بندی از این دیدگاه محسوب میشود. موسیقی سنتی ایرانی همان موسیقی دستگاهی می باشد. این نوع موسیقی شامل گوشه ها و نغمه هایی است که طبق ویژگی های ملودیک و ساختاریشان در یک گروه طبقه بندی شده اند. این نوع موسیقی طبق سنت ها، فرهنگ ها و آداب و رسوم های هر کشور و قوم می باشد و برای شناخت و تسلط داشتن به موسیقی سنتی باید ردیف موسیقی سنتی را آموخت که در دوران قاجار موزیسین ها و نوازندگان آن را به عنوان یک مجموعه منظم شکل دادند.

#### ۱.۲ ردیف و دستگاه

ردیف در نظام دستگاهی موسیقی سنتی ایرانی، هفت دستگاه به نام های شور، سه گاه، همایون، چهارگاه، نوا، راست پنجگاه، ماهور دارد. همچنین پنج آواز با نام های ابوعطا، بیات ترک، افشاری و دشتی که کوک آن شبیه به دستگاه شور است و جزو دستگاه شور محسوب می شود و آواز اصفهانی که کوکی مشابه با دستگاه همایون دارد و جزو دستگاه همایون می باشد. [۲۸]

دستگاه از دو کلمه مجزا «دست» و «گاه» به وجود آمده است. گاه در لغت به معنای پرده ساز



می باشد و ترکیب دو کلمه دستگاه به معنای چگونگی قرار گرفتن دست بر روی پرده های ساز است. چگونگی قرار گرفتن دست ها در هر قطعه، آهنگ و نغمه متفاوت است. تفاوت آواز با دستگاه فقط در تعداد گوشه ها می باشد. گوشه ها در دستگاه بیشتر از آواز می باشد. هفت دستگاه (به نامهای شور، سه گاه، چهارگاه، ماهور، همایون، نوا و راست پنجگاه) و پنج آواز با یکدیگر حدود ۲۰۰ تا ۴۰۰ گوشه دارد. این گوشه ها ممکن است در آواز و دستگاه از لحاظ ملودیک و ساختاری شبیه هم باشند. به مجموعه ای از آوازاها و دستگاه ها که حالت ها، ویژگی های ساختاری و ملودیک آنها شبیه باشد ردیف می گویند. از جمله ردیف موسیقی ایرانی می توان به ردیف آوازی و ردیف سازی اشاره داشت.

## ۲.۲ سازهای موسیقی سنتی ایرانی

در این نوع موسیقی از انواع سازهای زهی، کوبه ای و بادی با تنوع بالایی استفاده می شود. سازهای زهی شامل سازهایی است که از طریق سیم صدا تولید می کند؛ مانند چنگ، تار، سه تار، کمانچه، قیچک و ...، سازهای کوبه ای با ضربه زدن و خراشیدن صدا تولید می کنند و از آنها برای برجسته کردن ضرباهنگ قطعات موسیقی استفاده می شود مثلاً دایره، دف، تنبک، طبل و ... و در آخر، سازهای بادی در اثر دمیدن به صدا در می آیند. نی و نی انبان از سازهای بادی معروف می باشند. حال به معرفی کوتاهی از چند ساز و ویژگی های آنها می پردازیم:

- ساز کمانچه، یکی از پرکاربردترین و تنها ساز کششی ارکسترهای موسیقی ایرانی است. اصولاً ترکیب نوای یک ساز کششی و زخمه، ترکیب دل نشینی می شود و از همین رو سال هاست موسیقی ایرانی را با تار و کمانچه می شناسند. ساختار منحصر به فرد، امکان بالایی اجرای انواع پرده ها از سویی نوازنده و توان بیان حالات مختلف موسیقی ردیف دستگاهی سبب شده است که این ساز از آغاز دهه پنجاه و به ویژه در دهه های شصت و هفتاد و هشتاد از محجور بودن در آمده و به جایگاه پیشین خود، یعنی پیش از ورود ساز ویلون به ایران برسد. کاسه طنینی این ساز تقریباً کره ای شکل و تو خالی است. سطح بالایی آن باز است که روی آن پوست می کشند و خرک را روی آن نصب می کنند. آرشه یا کمانه یا کمان میله ای است چوبی که کمی هم منحنی و دسته ای از موی دم اسب به دو سر آن وصل شده است. موی های آرشه کشیده و محکم نیست و هنگام نوازندگی با انگشت های دست راست نوازنده به میزان کشیدگی مطلوب می رسد. البته چنانچه در شماره دیگری از معرفی سازهای ایرانی خواهیم گفت، این ساز اشکال مقامی نیز در لرستان، آذربایجان و خراسان شمالی دارد و نوع لرستانی آن به کمانچه پشت باز معروف است.
- فرهنگ دهخدا سنتور را این گونه معرفی کرده است «از سازهای ایرانی به شکل ذوزنقه که دارای سیم های بسیاری است و با دو زخمه چوبی نواخته می شود. رایج ترین نوع سنتور (۹

خرکي) داراي ۷۲ سیم است که به دسته های ۴ تایی و در ۱۸ دسته تقسیم می شود. «. این ساز را می توان سازی دانست که در کنار سه تار، بیشترین نوازنده را دارد. سنتور را بسیاری از علاقه مندان به موسیقی دوست دارند. صدای شفاف و آسانی نسبی در آغاز صدادهی آن برای نوآموزان، از دلایل اصلی توجه به ساز سنتور است. معمولاً دو شیوه اصلی برای نواختن این ساز وجود دارد که از نظر صدادهی، تفاوت های زیادی دارند. شیوه نوازندگی بدون نمد و با نمد. نمد سنتور تکه ای که روی مضراب سنتور چسبانده می شود و صدای آن را از نظر سونوریت به صدای پیانوشیه تر می کند، اما شیوه نواختن بدون نمد آن به صدای سنتور ایرانی و سنتی تر نزدیک تر است. در سه دهه گذشته و با حضور آهنگسازانی چون پرویز مشکاتیان و فرامرز پایور که خود از نوازندگان چیره دست و صاحب سبک این ساز بودند، سنتور در ارکسترهای ایرانی جلوه بیشتری پیدا کرده است. از نوازندگان مطرح این ساز می توان به حبیب سماعی، فرامرز پایور، پرویز مشکاتیان، رضا ورزنده، فضل الله توکل، پشنگ کامکار، اردوان کامکار و مجید کیانی اشاره کرد.

- بنا بر روایتی، تار را شاه سازهای ایرانی می دانند. البته بیره هم نگفته اند چون ساز تار به دلیل پرده بندی های زیادی که دارد، توانایی اش در اجرای ظرایف موسیقی ردیف دستگاهی بسیار بالاست. صدای خوش و خاص آن دل های زیادی را تسخیر کرده است. کمتر اثری در موسیقی ایرانی است که صدای تار از آن به گوش نرسد و کمتر ارکستر ایرانی را می توان نام برد که تار در آن نقشی نداشته باشد. ساز تار، شش سیم دارد که نوازنده با مضرابی از جنس برنج با ضربه های منظم، آن را می نوازد. در ساخت کاسه طنینی این ساز، معمولاً از چوب درخت توت استفاده می شود. روی کاسه و نقاره پوستی کشیده می شود که شفافیت، زیبایی و عمق صدای تار مدیون این پوست است. همه خاندان آقا علی اکبرخان فراهانی، شامل میرزا عبدالله، میرزا حسینقلی، علی اکبرخان شهنازی، عبدالحسین شهنازی از نوازندگان چیره دست تار بوده اند، همچنین از نوازندگان نامدار و صاحب سبک دیگر این ساز می توان به درویش خان، مرتضی نی داوود، جلیل شهناز، فرهنگ شریف، هوشنگ ظریف، حسین علیزاده، محمدرضا لطفی، داریوش طلایی و کیوان ساکت اشاره کرد.

- خیلی از مردم شناسان و محققان موسیقی معتقدند که ساز نی نه تنها قدیمی ترین ساز ایران، بلکه قدیمی ترین ساز جهان است البته بعد از سازهای کوبه ای. نی ها در جهان و ایران انواع بسیار زیادی دارند که اگر فقط برخی از نی های ایرانی را معرفی کنیم، چندین صفحه مجله را به خود اختصاص می دهد. اگر به موسیقی ایرانی هم علاقه مند نباشید، بعید است که صدای نی شما را جذب نکرده باشد. این ویژگی ذاتی ساز نی است. نی ایرانی با اجرای تنها دو نت می تواند شنونده را جذب خود کند و گوش نوازی اش را به رخ همگان بکشد. نی هفت بندی که در موسیقی ردیف دستگاهی می شود، از ساز های بادی کهن و از جنس گیاه نی است که طول آن از شش گره و هفت بند تشکیل شده؛ به همین دلیل به آن نی هفت

بند می گویند. هنگام نواختن، هوا از دهانه نی وارد می شود و با انگشت گذاری بر سوراخ ها و با تغییر فشار هوا صدا های مختلف به وجود می آید. وسعت معمول صدای نی حدود دو اکتاو و نیم است. از نوازندگان مطرح نی می توان به حسن کسایی، عبدالنقی افشارنیا، محمدعلی کیانی نژاد، جلیل عندلیبی، حسن ناهید، محمد موسوی و حسین یآوری اشاره کرد.

- ساز تنبک، به نوعی تنها ساز کوبه ای و غیر ملودیک موسیقی ایرانی است. هر چند سازهایی چون دایره و دف و انواع دمام و سازهایی چون تاس و نقاره و کوزه، امروزه در ارکسترهای موسیقی ایرانی وجود دارد و استفاده می شود، اما ساز کوبه ای موسیقی ردیف دستگاہی تنبک است که گاهی بدان تمبک نیز گفته می شود. سازهایی که در کنار تنبک از آنها نام برده شد، عموماً به عنوان سازهایی کوبه ای همراهی کننده با سازهایی ملودیک موسیقی ایرانی در ۵۰ سال گذشته و بر گرفته از موسیقی مقامی نواحی ایران در موسیقی دستگاہی جایگاهی پیدا کردند که در بخش دوم از معرفی سازهایی ایرانی به آن خواهیم پرداخت. ساز تنبک در تحول و تطورهایی یک صد سال اخیر به دلیل امکانات بالا و تکنیک های زیاد و متنوع نوازندگی آن توانسته است با هر شرایطی خودش را وفق دهد و پای ثابت همه نوع ارکسترهای موسیقی یارانی باقی بماند. امروزه نوازندگان چیره دست، ساز، تنبک دستی، دف و اکثر سازهایی کوبه ای استفاده شده در موسیقی ردیف دستگاہی ایران را دارند. از نوازندگان برجسته این ساز می توان به حسین تهرانی، بهمن رجبی، کامبیز گنجه ای، محمد قوی حلم، مرتضی اعیان، جهانگیر ملک، همایون شجریان، ناصر فرهنگ فر، محمود فرهمند و نوید افقه اشاره کرد.

- ساز سه تار همیشه ساز خلوت اهل دل بوده است. از آنجا که ساختار و پرده بندی ساز سه تار شباهت با سازتار دارد، معمولاً نوازندگان تار دستی هم بر سه تار دارند، اما این دو ساز از نظر شخصیت صوتی و رنگ آمیزی صوتی دو ساز کاملاً جدایی از یکدیگر هستند. موج نخست توجه به سه تار با درویش خان در دوره پهلوی اول و اواخر قاجار آغاز می شود و آغاز دور دوم شنیدن صدای سه تار نزد علاقه مندان به موسیقی را می توان با نوازندگی مرحوم احمد عبادی در رادیو دانست. محمدرضا لطفی و حسین علیزاده نیز سهم بسزایی در شناساندن صدای سه تار به مخاطبان موسیقی ایرانی داشتند البته نمی توان از دهه ۶۰ و ساز مرحوم جلال ذوالفنون چشم پوشی کرد. این نوازنده سه تار با رویکرد خاص خود در نوازندگی و آهنگ سازی و استفاده گروهی از سه تار که از نمادهای بارز آن آلبوم گل صدبرگ با صدای شهرام ناظری و آلبوم پرند با تکنوازی خود ذوالفنون اشاره کرد. از نوازندگان برجسته این ساز می توان به، درویش خان، یوسف فروتن، سعید هرمزی، حسین علیزاده، محمدرضا لطفی و جلال ذوالفنون اشاره کرد.

• ساز بریط را می‌توان از اصیل‌ترین سازهای موسیقی ایران باستان به همراه چنگ به شمار آورد. آثار باستانی عیلامی متعلق به هزاره سوم-چهارم پیش از میلاد، نگاره ای را نشان می‌دهد چندین زن به شکل منظم در حال نواختن بریط هستند. چون سطح ساز بریط از چوب پوشیده شده است، در زبان عربی آن را عود نامیده‌اند (العود در زبان عربی به معنای چوب است). به باور برخی پژوهشگران موسیقی مانند اکارد نوبوئر این واژه در اصل عربی شده واژه «رود» فارسی می‌باشد که هم نام دیگر بریط است و هم به معنای سیم در سازهای زهی می‌باشد. این ساز پس از اینکه به اروپا برده شد، نام لوت بر آن نهادند. واژه لوت از نگارش کلمه العود به وجود آمده و به تدریج به لوت تبدیل شده است. اصولاً بریط یا همان عود در اندازه‌های مختلف ساخته می‌شود که اندازه متداول و معمول همان عودهای ساخت ایران است. نمونه‌های ساخت کشورهای عربی دارای کاسه‌ای بزرگ و عودهای ترکیه کوچک و عودهای ایرانی متوسط است. شکم این ساز بسیار بزرگ و گلابی شکل و دسته آن بسیار کوتاه است. به طوری که قسمت اعظم طول سیم‌ها در امتداد شکم قرار گرفته‌است. سطح رویی شکم از جنس چوب است که بر آن پنجره‌هایی مشبک ایجاد شده‌است. بریط فاقد «دستان» است و خرک ساز کوتاه و تا اندازه‌ای کشیده‌است. بریط دارای ده سیم یا ۵ سیم جفتی است البته در برخی مواقع استادان قالب شکنی کرده و دو یا یک سیم در قسمت پایین قبل از سیم دو به ساز اضافه می‌کنند که این سیم‌ها فا زیر کوک می‌شود. سیم‌های جفت با هم همصدا (کوک) می‌شوند و هر یک از سیم‌های دهگانه، یک گوشی مخصوص به خود دارد؛ گوشی‌ها در دو طرف جعبه گوشی (سر ساز) قرار گرفته‌اند. بریط (عود) بم‌ترین ساز بین سازهای زهی ایرانی است؛ نت نویسی آن با کلید سل است (در واقع نت نویسی آن با توجه به وسعت و بمی صدای ساز بر اساس کلید فا می‌باشد که برای سهولت نت خوانی و نوازندگی یک اکتاو بالاتر نوشته می‌شود) که جمعاً دو اکتاو است. «اکتاو» بم‌تر از نت نوشته شده حاصل می‌شود. سیم بم (سل پائین) معمولاً نقش «واخوان» دارد و گاه این سیم جفت نیست. صدای بریط (عود) به نحوی است که صدای اکتاو چهارم پیانو از راست به چپ برابری دارد و در اصل باید بریط (عود) را با کلید «فا» نواخت یعنی صدای اصلی بریط یک اکتاو پایین‌تر از آن است که امروز متداول شده‌است. صدای بریط بم، نرم و در عین حال گرم و جذاب و نسبتاً قوی است. این ساز نقش تک‌نواز و هم‌نواز هر دو را به خوبی می‌تواند ایفا کند. همانگونه که مستحضرید بریط صدایی بم و تا حدودی تو دماغی دارد که دلیل آن نوع زه (سیم) ساز و عدم وجود پرده بندی (مانند ویلن) روی دسته ساز است. پرده بندی موسیقی ایرانی بخوبی روی این ساز قابل بیان است. در ایران نیز پس از احیای ساز عود تلاش‌های مثر ثمری صورت گرفت. از جمله عود نوازان معروف می‌توان به منصور نریمان، حسن منوچهری، محمود رحمانی پور، اکبر محسنی و عبدالوهاب شهیدی اشاره کرد. در سال‌های اخیر هم که استادان دیگری به این جمع اضافه شدند و اکثر از کسانی هستند که در نواختن ساز عود نوآوری‌هایی داشتند، از جمله این افراد ارسالان کامکار، حسین بهروزی‌نیا، محمد فیروزی و جمال جهانشاد هستند.

## ۳.۲ ارکان موسیقی سنتی

تفکیک مرز بین موسیقی سنتی ایرانی با دیگر انواع موسیقی به خصوص موسیقی کلاسیک غربی که بیشتر پژوهش‌ها در این حوزه انجام شده، ضروری به نظر می‌رسد. با این وجود، به طور کامل و قطعی نمی‌توان مبانی نظری موسیقی شرقی (از جمله موسیقی سنتی ایرانی) را از مبانی نظری موسیقی غربی تفکیک نمود و برخی مباحث و ارکان، در هر دو نوع موسیقی، مشترک است.

اصوات موسیقی، به واسطه چهار رکن اصلی و بنیادی از دیگر صداها متمایز می‌شوند: نواک، کشش (ارزش زمانی نغمه، دیرند، فاصله زمانی، مدت تداوم صوت، ریتم یا وزن)، دینامیک (شدت صوت یا دامنه) و طنین (رنگ صوتی یا شیوش) در تفکیک و تمیز دستگاه‌های موسیقی سنتی ایرانی از یکدیگر، ویژگی نواک از اهمیت بیشتری برخوردار است. نواک که در متون مختلف، با عناوین دیگر چون زیربومی، پیچ فرکانسی، ارتفاع صوت یا نام نغمه نیز بکار رفته است، صفتی است ادراکی از صدا که از طریق تنظیم فرکانس (بسامد) یک موج سینوسی، با دامنه دلخواه برای انطباق با صدای مورد نظر به دست می‌آید. از این رو، نواک هر نغمه را می‌توان در سه مفهوم بسامد، نام نغمه و طول موج نیز توصیف کرد.

بسامد، تعداد نوسان در مدت زمان معین (معمولاً یک ثانیه) است که واحد آن هرتر است. طول موج، تجسم دیگری از بسامد است و واحد اندازه‌گیری آن سانتی‌متر (یا متر) است. هرچه طول موج کوتاه‌تر باشد، صدای حاصل زیرتر خواهد بود و برعکس.

الفبای موسیقی، نغمه (نُت) نام دارد. برای همه صداها موسیقایی، از بم‌ترین تا زیرترین آن‌ها، فقط هفت نام وجود دارد که حداکثر نُه بار تکرار می‌شوند. وسعت اصواتی که در موسیقی بکار گرفته می‌شوند، حدود نه اکتاو است. یعنی دامنه صوتی برای اجرای کلیه قطعات موسیقی، در هر سبک و با هر سازی، بین بم‌ترین نغمه، یعنی  $C0$  (نغمه دو در اکتاو صفر) با فرکانس 16.351 هرتر (در کوک دیاپازون) و زیرترین نغمه، یعنی  $B8$  قرار دارد. نام‌گذاری این هفت نغمه به دو شکل هجایی یا الفبایی رایج است. نام‌گذاری نغمه‌ها در برخی کشورها (از جمله فرانسه، ایتالیا و ایران) از نظام هجایی و در کشورهای انگلیسی‌زبان (انگلستان و آمریکا) و آلمانی‌زبان (اتریش و آلمان)، از نظام الفبایی تبعیت می‌کنند. (شکل ۱.۲)

مفهوم فواصل در موسیقی ایرانی و اثرات آن بر ساختار و طبقه‌بندی ردیف موسیقی ایرانی، یکی از مهمترین مسائلی است که مورد توجه موسیقیدانها بوده است. اختلاف در نواک دو نغمه، فاصله موسیقایی را ایجاد میکند. نیم پرده در موسیقی غربی، کوچکترین فاصله میان دو نغمه پیاپی است. این فاصله در موسیقی ایرانی ربع پرده است؛ با این تفاوت که این فاصله در موسیقی غربی تعدیل شده است ولی این امر در موسیقی ایرانی (و به طور گسترده‌تر در موسیقی مشرق زمین) به دلیل ماهیت این نوع موسیقی و ترجیح اساتید اهل فن انجام نگرفته است تا از این طریق، دست نوازنده را در اعمال سلاقی خود در کوک فواصل ربع پرده و بداهه نوازی باز بگذارد.



شکل ۱.۲: دو سیستم نام‌گذاری جهانی برای هفت نغمه موسیقی به همراه نمایش آن‌ها در نوشتار موسیقی

## ۴.۲ بازیابی و استخراج اطلاعات موسیقایی

با افزایش روزافزون حجم داده‌های موسیقی ذخیره شده در منابع قابل دسترس و گوناگون، نیاز به سیستم‌های خودکار بازیابی و استخراج اطلاعات از موسیقی اهمیت ویژه‌ای یافته است. یک دسته از روش‌های مورد توجه، بازیابی اطلاعات موسیقی مبتنی بر محتوا است که اخیراً یکی از موضوعات پرطرفدار پژوهشی در حوزه پردازش سیگنال موسیقی و بازیابی اطلاعات بوده و تکنیک‌ها و ابزارهای مختلفی نیز در این زمینه توسعه یافته است. بطور کلی، سیگنال موسیقی به عنوان یک پدیده پیچیده حاوی حجم زیاد و متنوعی از اطلاعات در خصوص ژانر، احساس، هنرمند، ساز و غیره است. تنوع بالای اطلاعات موجود در سیگنال موسیقی، باعث مطرح شدن حیطه گسترده‌ای از مسائل در «بازیابی اطلاعات موسیقی مبتنی بر محتوا» مانند قطعه بندی یک قطعه موسیقی به بخش‌های آواز و غیر آواز شناسایی خواننده دسته بندی ژانر جستجو با زمزمه تشخیص بار احساسی موسیقی تشخیص ساز موسیقی حاشیه نویسی خودکار موسیقی و غیره، جهت مطالعه و پژوهش شده است که نمونه‌ی پیشرفته و پیاده سازی شده خیلی از این مسائل را در برنامه‌های پرکاربرد هم می‌توانید پیدا کنید. [۴]

### ۱.۴.۲ بازیابی اطلاعات در موسیقی سنتی ایرانی

علی‌رغم پژوهش‌های زیادی که در حوزه پردازش و بازیابی اطلاعات موسیقایی بر مبنای محتوا برای موسیقی سایر نواحی، به ویژه موسیقی غربی، صورت گرفته‌اند، پژوهش‌های انجام شده در حوزه پردازش و بازیابی موسیقی سنتی ایرانی بسیار ناچیز است دلایل متعددی برای کمبود پژوهش در این حوزه قابل طرح است که یکی از مهمترین آن، نبود دادگان معتبر و قابل دسترس برای مسائل مختلف مطرح در این حوزه است. بنابراین یک گام مهم در راستای رونق دادن به پژوهش در حوزه پردازش موسیقی سنتی ایرانی، گردآوری دادگان معتبر و قابل دسترس همگان است. در حوزه موسیقی

سایر نواحی و بویژه موسیقی غربی، دادگان های متعددی برای هر یک از مسائل پژوهشی وجود دارد [۳۰]. گردآوری یک دادگان همه منظوره به نحوی که برای همه مسائل پژوهشی حوزه قابل استفاده باشد، سخت و یا تقریباً غیر ممکن است که نبود این چنین دادگانی برای موسیقی غربی دلیلی بر این ادعاست. دو تا از مسائل پایه و اساسی در راستای بازیابی اطلاعات موسیقایی سنتی ایرانی مبتنی بر محتوا، تشخیص دستگاه و تشخیص ساز است. تشخیص ساز موسیقی نیز از مباحث اساسی در حوزه بازیابی اطلاعات موسیقی مبتنی بر محتوا است که در حوزه موسیقی سنتی ایرانی به نحوه جامع و کاملی به آن پرداخته نشده است موسیقی می تواند شامل صدای یک یا چند ساز باشد. در فرایند تشخیص ساز به طور کلاسیک که یک مساله برجسب زنی توالی محسوب می شود، هر قطعه موسیقی به قطعاتی ریز تقسیم شده و به قطعه برجسب یک یا چند ساز زده می شود.

## فصل ۳

# مرور کارهای انجام شده دیگران

هدف از یادگیری بازنمایی، شناسایی ویژگی‌هایی است که وظایف پیش‌بینی را برای تغییرات پیچیده داده‌های طبیعی آسان‌تر و قوی‌تر می‌کند [۳۱]. در یادگیری بازنمایی بی‌نظارت، مدل‌سازی مولد یا تولیدی و مدل‌های مبتنی شباهت معمولاً با تلاش برای بازسازی مشاهدات بر اساس بازنمایی‌های آموخته‌شده، بازنمایی‌های مفیدی از داده‌ها پیدا می‌کنند. [۳۲، ۳۳] هدف یادگیری بازنمایی خود نظارتی شناسایی عوامل توضیحی داده‌ها با استفاده از هدفی است که با توجه به بازنمایی‌های آموخته شده به طور مستقیم فرموله شده است. [۱۴، ۱۷، ۲۰، ۲۱]

در مقایسه با vision، کار بر روی یادگیری خود نظارتی در صدا هنوز بسیار محدود است، اما تعدادی از آثار وجود دارد که اخیراً ظاهر شده‌اند. کدگذاری پیش‌بینی متضاد یک رویکرد جهانی برای یادگیری متضاد است و برای طبقه‌بندی گوینده و واج با استفاده از صدای خام، در میان کارهای دیگر، موفق بوده است. [۱۴، ۳۴] چندین فعال تحت خود نظارتی را معرفی می‌کند که وظایف رگرسیون یا تبعیض باینری را حل کرده که به طور مشترک یک رمزگذار را برای تشخیص گفتار بهینه می‌کنند. برای بهبود نمایش‌های شرایط آکوستیک ناسازگار و قابلیت انتقال آن‌ها، افزاینده‌ها را به سیگنال گفتار ورودی اعمال می‌کنند. [۳۵]

در بازیابی اطلاعات موسیقی، پیشرفت‌های اخیر در تخمین گام با نظارت خود انجام شده است، با وجود اینکه بدون برجسب‌های حقیقت زمینی آموزش دیده‌اند. L<sub>3</sub>-Net تعبیه‌های عمیق را از مکاتبات سمعی و بصری در ویدئوها از طریق یادگیری خود نظارت می‌آموزد. کار آنها از طیف‌نگارهای مل برای صدا استفاده می‌کند و به بیش از ۴۰ میلیون نمونه آموزشی صوتی و تصویری برای یادگیری تعبیه‌های بهینه نیاز دارد. [۹]

Audio۲Vec همچنین در حوزه فرکانس زمانی عمل می‌کند و با بازسازی برش‌های طیف‌گرا از برش‌های گذشته و آینده یاد می‌گیرد. با داده‌های محدود، Audio۲Vec از مدل‌های با نظارت در طبقه‌بندی صدا و ابزار بهتر عمل می‌کند. [۱۰]

CLAR از یک هدف یادگیری متضاد استفاده می‌کند و ضرر را در مجموعه‌ای از بازنمایی‌های آموخته شده از هر دو طیف صوتی خام و mel-spectrograms محاسبه می‌کند. [۱۱]



COLA از روش مشابهی فقط با طیف‌نگارهای مل استفاده می‌کند و از مقایسه‌های دوخطی به جای شباهت کسینوس استفاده می‌کند. [۱۲] هر دو کار بر روی دستور گفتار، طبقه‌بندی صدای محیطی، و طبقه‌بندی زیر و بمی و ابزار در مجموعه داده NSynth ارزیابی می‌شوند. [۳۶]

## ۱.۳ در موسیقی ایرانی

تاکنون پژوهشی در خصوص شناسایی خودکار موسیقی سنتی ایرانی توسط پژوهشگران غیر ایرانی و متخصص در حوزه بازیابی اطلاعات موسیقی، صورت نگرفته است. شاید دلیل این امر، گنگ و مبهم بودن مسئله و یا ناآشنایی آنها با موسیقی سنتی ایرانی باشد. در این صورت، این امر را می‌توان ناشی از شناخته نشدن موسیقی ایرانی به صورت گسترده و خارج از ایران دانست؛ به نحوی که نیاز به انجام چنین پژوهش‌هایی، برای پژوهشگران غیر ایرانی ایجاد نشده است. اما جدا از پژوهشگران خارجی، طبق اطلاعاتی که تا به امروز در دسترس ما هست، پژوهش‌های داخل کشور هم اندک بوده و اکثراً به طور خاص به دسته بندی دستگاه‌ها و گوشه‌های موسیقی سنتی ایرانی پرداخته و یا با موضوع دسته بندی و تشخیص خودکار دستگاه‌های موسیقی ایرانی انجام شده‌اند و با نگاهی به فهرست منابع و تاریخ انتشار این پژوهش‌ها، می‌توان دریافت که بیشتر این پژوهش‌ها به موازات یکدیگر و به طور مجزا انجام شده و بیشتر آنها احتمالاً به دلیل بی‌اطلاع بودن و یا دسترسی نداشتن به منابع مرتبط و مشابه قبلی، از این آثار بهره‌نجمسته‌اند. این پژوهش‌ها در قالب پایان‌نامه مقاطع مختلف تحصیلی و مقاله توسط متخصصان و پژوهشگران زمینه‌های مختلف مانند مهندسی کامپیوتر و مهندسی برق و الکترونیک، مهندسی صدا، مهندسی مکانیک، علوم ریاضی و زیست‌پزشکی منتشر شده‌اند که متأسفانه بیشتر این پژوهشگران در زمینه مبانی نظری موسیقی سنتی ایرانی اشراف و تسلط لازم و کاملی را نداشته، یا با علم و نوازندگی سازهای موسیقی سنتی ایرانی بیگانه بوده و در بیشتر موارد اطلاعات آنها ناکافی و کم بوده است و به همین دلیل این پژوهش‌ها را محدودتر و دقت صحت آنها را تا حدی تحت تاثیر قرار داده است [۱]. در اکثر این پژوهش‌ها با توجه هدف آن پژوهش یک دادگان مستقل و جدید ساخته شده که عموماً تک منظوره بوده (تشخیص ساز یا دستگاه)، از جامعیت برخوردار نیستند یعنی تعداد محدودی قطعه تک‌نوازی بدون آواز بوده در حالی که در حقیقت در موسیقی سنتی حجم زیادی را تشکیل نمی‌دهند، از تنوع و حجم کافی برای هر ساز یا هر دستگاه برخوردار نیستند و داده‌های موجود برای سازها و دستگاه‌های مختلف، متعادل نیست.

همان‌طور که گفته شد، تعداد کل پژوهش‌های انجام شده که به طور خاص به دسته‌بندی دستگاه‌ها و گوشه‌های موسیقی سنتی ایرانی پرداخته‌اند، محدود است که یکی از آنها تنها به موضوع تعیین فرکانس پایه نغمه در یک ساز ایرانی پرداخته

و وارد مقوله تشخیص دستگاه و گوشه نشده است. [۱]  
در [۳۷] از یک شبکه عصبی با توابع شعاعی پایه برای تشخیص دستگاه‌ها و سایر دستگاه‌ها

برای ساز سه‌تار بهره برده است. دادگان استفاده شده شامل ۱۳۵ قطعه موسیقی است که ۶۰ تای آن در دستگاه ماهور و بقیه در پنج دستگاه دیگر بوده است. بعد از آموزش شبکه RBF برای ۷۰ درصد داده‌های موجود در دادگان، نهایتاً به دقت حدود ۷۳ درصد در تشخیص دستگاه ماهور رسیده است.

در پژوهشی دیگر بر این اساس که نت‌های نواخته شده توسط ساز نقش کلیدی در تشخیص دستگاه‌های موسیقی ایفا میکند، سعی بر آن داشته که نت‌های قطعه موسیقی را با دقت بالایی استخراج نماید و در ادامه با مشخص کردن فواصل بین این نت‌ها و با توجه به منحصر به فرد بودن الگوهای این فواصل برای دستگاه‌های مختلف، دستگاه قطعه موسیقی مورد نظر را تشخیص دهد. در این پژوهش برای ارزیابی روش پیشنهادی، از ۴۶ قطعه موسیقی (۴۲ قطعه توسط ساز تار و ۴ قطعه توسط سنتور نواخته شده است) در پنج دستگاه مختلف استفاده شده و به دقت ۹۳ درصد رسیده است. [۳۸]

در پژوهش [۳۹] برای دسته بندی ردیف میرزا عبدالله از ویژگی‌های مختلفی نظیر ناهمگونی، ضرایب کپسترال بر مبنای مقیاس مل، فرکانس گام، میانگین و انحراف معیار سنتروید طیفی بهره برده شده است. برای دسته‌بندی روش‌های مختلفی نظیر ماشین بردار پشتیبان شبکه عصبی پرسپترون و K تا نزدیکترین همسایه، آزموده شده است که ماشین بردار پشتیبان به دقت بالاتری دست یافته است. دادگان استفاده شده شامل ۱۲۵۰ قطعه موسیقی از سازهای زهی زخمه‌ای تار و سه‌تار توسط چهار استاد معروف ایرانی است و دربرگیرنده هفت دستگاه و شش آواز است.

در پژوهشی [۴۰] دیگر، دادگانی با عنوان دادگان موسیقی سنتی ایرانی مریم به منظور تشخیص دستگاه در موسیقی سنتی ایرانی به صورت مستقل از نوع ساز معرفی شده است. این دادگان شامل ۱۱۳۷ قطعه موسیقی است که ۶۳۱ تای آن صدای نی با صدای برخی سازهای دیگر در پس‌زمینه است و در بقیه قطعات، صدای ویولن به عنوان صدای پیش‌زمینه است. قطعات این دادگان در هفت دستگاه است و تعداد قطعات انتخاب شده در هر دستگاه متفاوت است.

دستگاه شور با ۴۴۵ قطعه بیشترین و دستگاه سه‌گاه با ۷۴ قطعه کمترین تعداد قطعه در این دادگان را دارد. محققین این پژوهش با انتخاب دو ساز نی و ویولن، سعی بر ارائه دادگانی برای مسئله تشخیص دستگاه به صورت مستقل از ساز داشته‌اند که انتخاب تنها دو ساز برای این ادعا مقبول به نظر نمی‌رسد. در ادامه، این محققین از ۸۰ درصد داده‌های هر دستگاه در دادگان مریم برای آموزش

و از ۲۰ درصد باقیمانده برای آزمون آن استفاده کرده‌اند. ایشان با این ادعا که دستگاه بر مبنای ۱۶ ثانیه موسیقی قابل تشخیص است، هر قطعه موسیقی به قطعات ۱۶ ثانیه‌ای قطعه‌بندی شده است. بر روی هر قطعه ۱۶ ثانیه‌ای تبدیل فوریه زمان کوتاه ۱۳ اعمال شده و ویژگی‌های حاصل به یک شبکه عصبی ژرف (با معماری: پنج لایه پیچشی + دو لایه GRU + دو لایه اتصال کامل) خورانیده شده است و متوسط امتیاز F1 حدود ۸۶ درصد بر روی هفت دستگاه گزارش شده است.

در پژوهش بعدی [۴۱] برای تشخیص دستگاه، با بهره‌گیری از منطق و نظریه مجموعه‌های فازی و با این فرض که هر نت نواخته شده، یک مجموعه فازی است؛ هر قطعه موسیقی را مجموعه‌ای از مجموعه‌های فازی در نظر می‌گیرد و بر این اساس، به محاسبه شباهت بین دستگاه قطعه موسیقی ورودی و دستگاه‌های مرجع می‌پردازد دادگان استفاده شده شامل جمعاً ۲۱۰ قطعه موسیقی سنتی

ایرانی با ۸۹ قطعه در دستگاه شور و نوا، ۳۰ قطعه در دستگاه سه‌گانه، ۴۱ قطعه در دستگاه ماهور و راست پنج‌گانه، ۲۶ قطعه در دستگاه همایون، و ۲۴ قطعه در دستگاه چهارگانه است. قطعات عمدتاً شامل آواز سه استاد آواز و تکنوازی چهار ساز تار، سه تار، سنتور و کمانچه است در این پژوهش نشان داده شده است که در روش پیشنهادی، یک دقیقه موسیقی از هر قطعه برای تشخیص دستگاه آن قطعه لازم و کافی است.

در این پژوهش [۴۲] برای دسته‌بندی هفت دستگاه موسیقی ایرانی، از شبکه عصبی پرسپترون چند لایه استفاده می‌شود. ورودی‌های شبکه عصبی، بیست‌ونهمین بلندی از طیف فرکانس هر قطعه موسیقی است. نتایج نشان می‌دهد شبکه می‌تواند دستگاه قطعات آزمون را با دقت حدود ۶۵ درصد برای نی، ۷۲ درصد برای ویولن و ۵۶ درصد برای آواز تشخیص دهد. در این پروژه [۴۳] هم از یک شبکه عصبی پرسپترون با یک لایه مخفی برای دسته‌بندی پنج دستگاه شور، ماهور، همایون، سه‌گانه و چهارگانه استفاده شده است. ورودی شبکه، بردارهای باینری ۲۴ مولفه‌ای است که هر مولفه آن گویای یک نت در یک اکتاو است. شبکه به کمک ۱۲۰ الگوی آموزشی تولید شده، آموزش داده شده است و دقت ۱۰۰ درصد بر روی این ۱۲۰ الگوی ممکن گزارش شده است. همان‌طور که بیان شد اکثر پژوهش‌ها بر موسیقی ایرانی تنها بر روی تشخیص دستگاه صورت گرفته است.

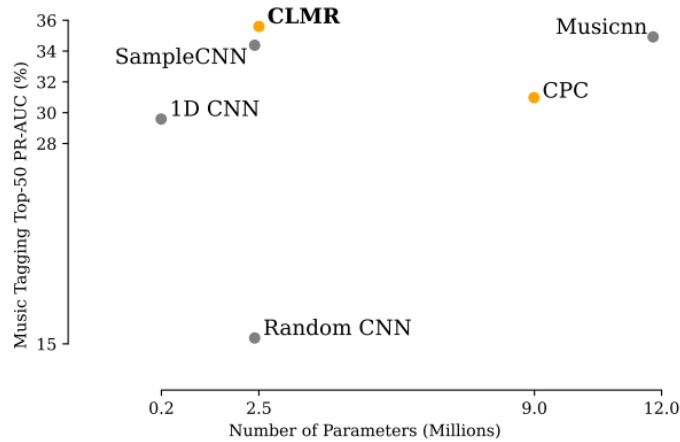
## فصل ۴

### سامانه پیشنهادی

روش های یادگیری با نظارت به طور گسترده در کارهای موسیقی استفاده شده است. این روش ها به داده های برجسب دار نیاز دارند، که ایجاد آنها برای موسیقی به ویژه دشوار، پرهزینه و زمان بر است ، در حالی که داده های موسیقی بدون برجسب خام در مقادیر زیادی موجود است . جایگزین های بی نظارت برای یادگیری عمیق انتها به انتها برای موسیقی قانع کننده هستند، به خصوص اگر بتوانند به مجموعه داده های کوچک تر تعمیم دهند.

در این پروژه ، ما اطلاعات یک چارچوب یادگیری متضاد ساده برای تصاویر، SimCLR را با پیشرفت های اخیر در یادگیری بازنمایی برای صدا در حوزه زمان ترکیب می کنیم. [۱۶] به عنوان تعریف های اولیه ، SimCLR چارچوبی برای یادگیری متضاد نمایش های بصری است که نمایش ها را با به حداکثر رساندن توافق بین نماها یا دیدهای تقویت شده متفاوت از نمونه داده های مشابه از طریق از زیان متضاد در فضای پنهان می آموزد. فضای پنهان هم ، که به عنوان فضای ویژگی پنهان یا فضای نهفته نیز شناخته می شود، تعبیه مجموعه ای از اقلام در یک منیفولد است که در آن اقلامی که شباهت بیشتری به یکدیگر دارند در فضای نهفته نزدیک به یکدیگر قرار می گیرند.

ما همچنین خط لوله ای از افزایش داده ها را در صدای موسیقی ارائه می کنیم تا چارچوبی ساده برای یادگیری متضاد و تحت روش خود نظارتی از بازنمایی شکل های موج خام موسیقی تشکیل دهیم. برای مقایسه اثربخشی این چارچوب ساده در مقایسه با یک هدف یادگیری با نظارت خود پیچیده تر، ما همچنین بازنمایی های آموخته شده توسط کدگذاری پیش بینی متضاد (CPC) را ارزیابی می کنیم . [۱۴] مدل های تحت نظارت خود بر روی وظیفه برجسب گذاری موسیقی پایین دستی، یعنی کار هدف ما ، ارزیابی می شوند و ما را قادر می سازند تطبیق پذیری آنها را ارزیابی کنیم؛ برجسب های موسیقی بسیاری از ویژگی های موسیقی را توصیف می کنند، به عنوان مثال در دادگان نوا ، نوع ساز، نوازنده و دستگاه مشخص است. [۴] عناوین و نتایج مطالعات ما به شرح زیر است :



شکل ۱.۴: مقایسه عملکرد و پیچیدگی مدل‌های با نظارت (به رنگ خاکستری) و مدل‌های خود نظارتی (به رنگ نارنجی) در طبقه‌بندی موسیقی شکل‌های موج صوتی خام روی دادگان برای ارزیابی بازنمایی‌های موسیقی. مدل‌های با نظارت به صورت انتها به انتها آموزش داده شده‌اند، در حالی که CLMR و CPC بدون دانش پایه‌ای از پیش آموزش داده شده‌اند. با این حال امتیازات آن‌ها با آموزش یک طبقه‌بندی کننده خطی بر روی بازنمایی‌های آموخته شده‌شان به دست آمده و عملکرد قابل رقابتی با مدل‌های نظارت شده دارند [۳]

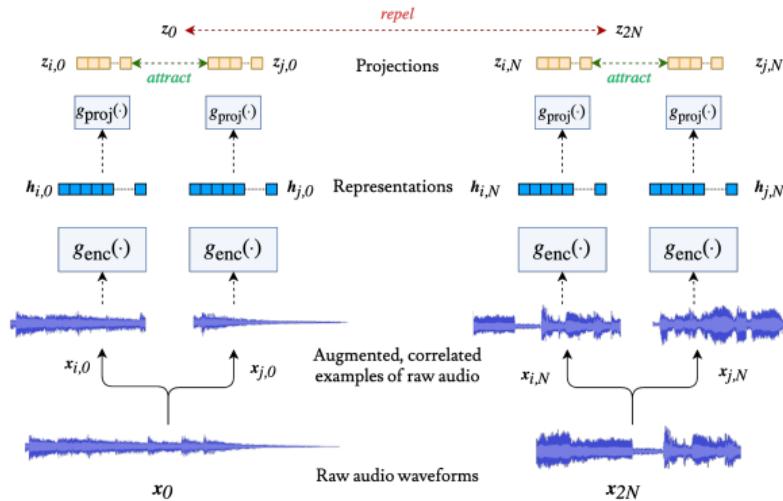
- CLMR علیرغم پیش آموزش خود نظارتی و آموزش طبقه‌بند خطی در انجام کار مورد هدف با سیگنال‌های خام صدای موسیقی، در مقایسه با مدل‌های با نظارت، عملکرد قوی‌تری در کار طبقه‌بندی موسیقی به دست می‌آورد. (شکل ۱.۵)
- CLMR طبقه بندی کارآمد را امکان پذیر می کند: ما عملکرد قابل مقایسه ای را با استفاده از کمتر از یک صدم از داده های برجسب دار بدست می آوریم.
- قابلیت انتقال خارج از دامنه بازنمایی های آموخته شده از قبل از آموزش CLMR را بر روی مجموعه های کاملاً متفاوت صوتی موسیقی نشان می دهیم.
- CLMR می تواند از هر مجموعه داده ای از صدای موسیقی خام بیاموزد، بدون نیاز به تغییر و تنظیم دقیق در داده های ورودی همچنین مدل ها برای پیش آموزش نیازی به برجسب های ثبت شده دستی ندارند.
- به علاوه ما یک مطالعه فرسایشی در مورد اثربخشی افزایش داده های صوتی فردی هم ارائه می دهیم.

## ۱.۴ کلیات روش سامانه

این پروژه به طور کلی بر اساس SimCLR، که همانطور که پیش‌تر اشاره شد، یک چارچوب یادگیری متضاد ساده از بازنمایی‌های بصری است، انجام گرفته است. علی‌رغم یک رویکرد پیش‌آموزشی تشخیص‌دهنده و بدون برجسب، یک طبقه‌بند خطی، عملکردی قابل مقایسه با مدل‌های کاملاً با نظارت در بسیاری از معیارهای طبقه‌بندی تصویر به دست آورد. هدف یادگیری آن به حداکثر رساندن توافق بازنمایی‌های نهفته نماهای تقویت‌شده یک تصویر با استفاده از زیان متضاد است. در CLMR، ما این چارچوب را با حوزه صدای موسیقی خام تطبیق می‌دهیم. در حالی که اکثر اجزای اصلی CLMR در کارهای دیگر ظاهر شده‌اند، توانایی آن در مدل‌سازی فرم‌های موجی موسیقی را نمی‌توان تنها با یک انتخاب طراحی و باید با ترکیب آنها توضیح داد. ابتدا چهار مولفه اصلی را در بخش‌های فرعی زیر توضیح خواهیم داد:

- یک ترکیب تصادفی از افزایش داده‌ها که دو نمونه همبسته و فزوده‌شده از یک قطعه صوتی مشابه، «دوتایی مثبت» را تولید می‌کند که با  $x_i$  و  $x_j$  مشخص می‌شود.
- یک شبکه عصبی رمزگذار  $g_{enc}(\cdot)$  که نمونه‌های افزوده‌شده را به بازنمایی یا نمایش‌های نهفته آنها نگاشت می‌کند.
- شبکه عصبی پروژکتور  $g_{proj}(\cdot)$  که بازنمایی‌های کدگذاری شده را به فضای پنهانی که در آن زیان متضاد فرموله می‌شود، نگاشت می‌کند.
- یک تابع زیان متضاد، که هدف آن شناسایی  $x_j$  از مثال‌های منفی در دسته  $\{x_{k \neq i}\}$  برای یک  $x_i$  معین است.

چارچوب کامل در شکل ۲.۵ رسم شده است.



شکل ۲.۴: چارچوب کاملی که بر روی صدای خام کار می کند و در آن یادگیری متضاد در فضای پنهان همبسته و افزوده شده‌ی دوتایی‌های نمونه از موج های صوتی یا خام موسیقی ، فرموله می شود [۳] .

#### ۱.۱.۴ داده افزایی

ما یک زنجیره جامع از افزوده‌های صوتی برای شکل‌های موج صوتی خام موسیقی طراحی کردیم تا شناسایی دوتایی‌های صحیح نمونه‌ای را برای مدل سخت‌تر کنیم. هر افزوده متوالی به طور تصادفی روی  $x_i$  و  $x_j$  به طور مستقل اعمال می‌شود. هر افزوده احتمال مستقل  $p_{transform}$  را برای اعمال بر روی فایل صوتی دارد. ترتیب افزایش‌های اعمال شده در صدا به دقت در نظر گرفته می‌شود، به عنوان مثال، اعمال یک اثر تاخیری پس از طنین به طور تجربی نتیجه کاملاً متفاوتی در موسیقی می‌دهد.

۱. یک قطعه تصادفی به اندازه‌ای مشخص از یک قطعه موسیقی انتخاب می‌شود، بدون اینکه سکوت حذف شود (سکوتی که گاهی در ابتدا یا انتهای قطعه موسیقی وجود دارد). دو مثال  $x_i$  و  $x_j$  از یک قطعه صوتی می‌توانند همپوشانی داشته باشند یا بسیار جدا و متفاوت باشند و به مدل اجازه می‌دهد همه‌ی ساختارها را استنتاج کند.

۲. قطبیت سیگنال صوتی معکوس شده است، یعنی دامنه در  $-1$  ضرب شده است.

۳. افزوده نویز سفید گاوسی با نسبت سیگنال به نویز  $80$  دسیبل به سیگنال اصلی اضافه شده است. (نویز سفید گاوسی افزودنی (AWGN) یک مدل نویز پایه است که در تئوری اطلاعات برای تقلید از تأثیر بسیاری از فرآیندهای تصادفی که در طبیعت رخ می‌دهند

استفاده می‌شود. سفید به این ایده اشاره دارد که دارای قدرت یکنواخت در سراسر باند فرکانسی برای سیستم اطلاعاتی است.

۴. بهره در بازه  $[۰, -۶]$  کاهش می‌یابد.

۵. یک فیلتر فرکانس روی سیگنال اعمال می‌شود. سکه انداختن اتفاقی تعیین می‌کند که فیلتر پایین گذراست یا بالا گذر. فرکانس های قطع از توزیع های یکنواخت در  $[۰, ۴۰۰۰]$  و  $[۲۰۰, ۱۲۰۰]$  هرگز به دست می‌آیند.

۶. سیگنال با تاخیر ۰.۵ به سیگنال اصلی اضافه می‌شود. تأخیر به طور تصادفی بین ۲۰۰ تا ۵۰۰ میلی‌ثانیه و با افزایش ۵۰ میلی‌ثانیه نمونه برداری می‌شود.

۷. سیگنال گام جابجا شده است. فاصله انتقال گام از توزیع یکنواخت نیم‌تون‌ها بین  $[۰, ۵]$ ، یعنی یک چهارم کامل در مقایسه با مقیاس سیگنال اصلی، گرفته می‌شود.

۸. طنین برای تغییر آکوستیک سیگنال اضافه شده است. اندازه فضای پاسخ ضربه، طنین و ضریب میرایی از توزیع یکنواخت به  $[۰, ۱۰۰]$  کشیده شده است.

فضای افزایش‌ها به این عملیات محدود نمی‌شود و می‌تواند به‌عنوان مثال، به‌کارگیری تصادفی کر، اعوجاج و سایر مدل‌سازی‌ها گسترش یابد. برخی از اینها عملکرد را در یادگیری خود نظارتی برای تشخیص خودکار گفتار در حوزه زمان نیز بهبود می‌بخشد.

#### ۲.۱.۴ ترکیب دسته‌ای

اندازه دسته بزرگ‌تر از حد خاصی هدف یادگیری متضاد را سخت‌تر می‌کند، اما می‌تواند عملکرد مدل را به‌طور قابل ملاحظه‌ای بهبود بخشد. اندازه دسته بزرگ‌تر از حد خاصی مثلاً  $N$  هدف یادگیری متضاد را سخت‌تر می‌کند، اما می‌تواند عملکرد مدل را به‌طور قابل ملاحظه‌ای بهبود بخشد. [۱۴]

یک آهنگ از دسته را به‌عنوان نمونه انتخاب می‌کنیم، آن را به دو نمونه اضافه کرده و با آنها به عنوان دوتایی مثبت رفتار می‌کنیم. ما  $2(N-1)$  نمونه باقیمانده در دسته را به‌عنوان نمونه‌های منفی در نظر گرفته و نمونه‌های منفی را مشخصاً نمونه برداری نکردیم. اندازه‌های دسته‌ای بزرگ‌تر یک مشکل عملی برای صدای خام در هنگام آموزش بر روی یک GPU ایجاد می‌کند، زیرا ابعاد ورودی آنها برای نرخ نمونه بالاتر افزایش می‌یابد. هنگام آموزش بر روی چندین پردازنده گرافیکی، ما از نرمال‌سازی دسته‌ای جامع استفاده می‌کنیم، به‌عنوان مثال، ما آمار دسته‌ای را در طول آموزش موازی روی همه دستگاه‌ها جمع می‌کنیم تا از نشت احتمالی آمار دسته‌ای جلوگیری کنیم، زیرا نمونه‌های مثبت در یک دستگاه نمونه برداری می‌شوند. که از دست دادن تمرین را بهبود می‌بخشد، اما با یادگیری نمایش‌های مفید مقابله می‌کند.



### ۳.۱.۴ رمزگذار

برای مقایسه مستقیم یک مدل با نظارت state-of-the-art آنها به انتها که در طبقه‌بندی موسیقی روی شکل‌های موج خام مورد استفاده قرار می‌گیرد در مقابل یک مدل تحت خود نظارتی، از معماری SampleCNN (یک شبکه‌ی عصبی کانولوشنال عمیق با استفاده از فیلترهای بسیار کوچک برای طبقه‌بندی موسیقی) به عنوان رمزگذار خود استفاده می‌کنیم. [؟] (یک مدل state-of-the-art که گاهی اوقات پیشرفته یا پیشرو هم شناخته می‌شود، به بالاترین سطح پیشرفت عمومی، به عنوان یک دستگاه، تکنیک یا زمینه علمی که در یک زمان خاص به دست آمده است، اشاره دارد.)

در این پیکربندی، gene رمزگذار SampleCNN از ۹ بلوک کانولوشن تک بعدی تشکیل شده است که هر کدام دارای اندازه فیلتر ۳، حالت عادی دسته‌ای، فعال سازی ReLU (مخفف واحد خطی اصلاح شده است و نوعی تابع فعال سازی است. از نظر ریاضی به صورت  $y = \max(x, 0)$  تعریف می‌شود. ReLU متداول ترین تابع فعال سازی در شبکه‌های عصبی به خصوص در CNN ها است. اگر مطمئن نیستید که از چه عملکرد فعال سازی در شبکه خود استفاده کنید، ReLU معمولاً انتخاب خوبی است.)

لایه خروجی نهایی حذف می‌شود، که یک بردار ویژگی ۵۱۲ بعدی  $h_i$  را برای هر ورودی صدا ایجاد می‌کند. بردارهای ویژگی از رمزگذار را می‌توان مستقیماً در هدف یادگیری استفاده کرد، اما فرمول‌بندی هدف روی رمزگذاری‌هایی که با یک تابع پارامتری به فضای پنهان متفاوت نگاشت شده‌اند، به اثربخشی نمایش‌ها کمک می‌کند از یک لایه غیرخطی  $z_i = W^2 \text{ReLU}(W^1)h_i$  با ابعاد خروجی ۱۲۸ به عنوان شروع برآورد gproj استفاده می‌کنیم. ما با دسته و پیکربندی رمزگذار فوق تا مستقیماً عملکرد خود نظارتی خود را با روش کاملاً با نظارت به همان اندازه مقایسه کنیم و به دلیل بزرگی اندازه‌ی دادگان زمان بسیار زیادی صرف اجرای آن شد.

### ۴.۱.۴ تابع زیان متضاد

با توجه به یافته‌های اخیر در مورد چندین تابع هدف در یادگیری متضاد، تابع زیان متضاد مورد استفاده در این مدل، تابع زیان آنروپی مقیاس‌پذیر دما نرمال شده است که معمولاً به عنوان زیان NT-Xent شناخته می‌شود:

$$l_{i,j} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_j)/\tau)}$$

شباهت زوجی با استفاده از شباهت کسینوس اندازه‌گیری می‌شود و پارامتر دما  $\tau$  به مدل کمک می‌کند تا از منفی‌های بزرگ یاد بگیرد. تابع نشانگر  $\mathbb{1}_{[k \neq i]}$  برابر ۱ خواهد بود اگر و تنها اگر  $k \neq i$  و در غیر این صورت مقدار آن ۰ خواهد بود. این زیان برای تمام زوج مرتب یا دوتایی‌های  $(z_i, z_j)$  و  $(z_j, z_i)$  در حالتی که  $i \neq j$  برقرار باشد، محاسبه می‌شود.

## ۵.۱.۴ کدگذاری پیشگویانه متضاد

ما برای مقایسه مستقیم تر،  $genc$  رمزگذار CPC اصلی را به یک معماری عمیق تر تنظیم کرده .  $genc$  رمزگذار شامل ۷ لایه با هر کدام ۵۱۲ فیلتر و اندازه های فیلتر [۲، ۲، ۴، ۴، ۴، ۶، ۱۰] است. به جای تکیه بر حداکثر ادغام، اندازه ها و گام های فیلتر برای پارامترسازی و تسهیل نمونه گیری پایین تنظیم می شوند. ( در پردازش سیگنال، نمونه برداری پایین، فرآیند کاهش نرخ نمونه برداری از یک سیگنال است. این معمولاً برای کاهش سرعت داده یا اندازه داده ها انجام می شود. ضریب نمونه برداری معمولاً یک عدد صحیح یا یک کسر گویا بزرگتر از واحد است. )  
ما همچنین تعداد مراحل پیش بینی را به ۲۰ افزایش داده و از شبکه خواستیم تا ۱۰۰ میلی ثانیه صدا را در آینده پیش بینی کند. اندازه دسته روی ۶۴ تنظیم شده است که از آن ۱۵ مثال منفی در زبان متضاد گرفته شده است.

## ۶.۱.۴ ارزیابی خطی

ارزیابی بازنمایی های آموخته شده توسط مدل های خود نظارتی معمولاً با ارزیابی خطی انجام می شود، که اندازه گیری می کند که کلاس های مربوطه تا چه حد تحت بازنمایی های آموخته شده به صورت خطی قابل تفکیک هستند. ما بازنمایی هایی را برای تمام نقاط داده از یک شبکه ثابت CLMR پس از همگرا شدن پیش آموزش به دست می آوریم و یک طبقه بند خطی را با استفاده از این نمایش های خود نظارت در وظیفه پایین دستی یعنی برای حل مسئله نهایی، طبقه بندی موسیقی آموزش می دهیم. برای CPC، نمایش ها از خود رگرسیون (خودرگرسیون یک مدل سری زمانی است که از مشاهدات مراحل زمانی قبلی به عنوان ورودی معادله رگرسیون برای پیش بینی مقدار در مرحله زمانی بعدی استفاده می کند.) استخراج می شوند و یک بردار زمینه به دست می آید که به طور میانگین سرتاسری برای به دست آوردن یک بردار تک با ابعاد ۵۱۲ ادغام می شود. برای CLMR، آخرین بردار ۵۱۲ بعدی  $h$  از رمزگذار به جای  $z$  از ابتدای برآورد استفاده می شود، زیرا برای همه آزمایش های ما نتایج بهتری به همراه داشت. ما معیارهای ارزیابی را بر روی یک مجموعه آزمون نگه داشته شده محاسبه می کنیم، که میانگین آن در سه مرحله در مجموعه آموزشی با استفاده از seed یا دانه های تصادفی مختلف محاسبه می شود.

## ۷.۱.۴ بهینه سازها

ما از بهینه ساز Adam با نرخ یادگیری 0.0003 و  $\beta_1 = 0.9, \beta_2 = 0.999$  در طول آموزش استفاده می کنیم و از مقدار اولیه He برای همه لایه های کانولوشن استفاده می کنیم. ( الگوریتم Adam یک الگوریتم بهینه سازی جایگزین برای کاهش گرادیان تصادفی برای آموزش مدل های

یادگیری عمیق بوده و یک روش نرخ یادگیری تطبیقی است، به این معنی که نرخ یادگیری فردی را برای پارامترهای مختلف محاسبه می کند. بهینه سازی تطبیقی هم تکنیکی است که به صورت پویا کامپایل مجدد بخش هایی از یک برنامه را بر اساس نمایه اجرای فعلی انجام می دهد. با یک پیاده سازی ساده، یک بهینه ساز تطبیقی ممکن است به سادگی بین کامپایل سازی درست در لحظه و تفسیر دستورالعمل ها معاوضه ایجاد کند. Adam در بیشتر موارد در بین بهینه سازهای تطبیقی بهترین و نرخ یادگیری تطبیقی برای مجموعه داده هایی که داده های پراکنده ای دارند، بهترین است. و از آن جایی که در شبکه عصبی ReLU فعال می شود، مقداردهی اولیه He یکی از روش هایی است که می توان انتخاب کرد تا واریانس آن خروجی ها را تقریباً به یک رساند. ( پارامتر  $\tau$  روی 0.5 تنظیم شده است، چرا که ما نتایج ثابتی را بدون در نظر گرفتن اندازه های مختلف دسته و دمای  $\tau \in \{0.1, 0.5, 1.0\}$  مشاهده کردیم. برای ارزیابی خطی، ما از بهینه ساز Adam با نرخ یادگیری 0.0003 کاهش وزن  $10^{-6}$  استفاده می کنیم.

انتشار پس انتشار تنها در سر نهایی (خطی) برای همه آزمایش ها در اینجا انجام می شود. همچنین زمانی که امتیازهای اعتبارسنجی برای 5 دوره بهبود نمی یابد، از یک مکانیسم توقف اولیه استفاده می کنیم. (انتشار پس انتشار، مخفف انتشار خطاها به عقب، الگوریتمی برای یادگیری با نظارت شبکه های عصبی مصنوعی با استفاده از کاهش گرادیان است. با توجه به یک شبکه عصبی مصنوعی و یک تابع خطا، این روش گرادیان تابع خطا را با توجه به وزن های شبکه عصبی محاسبه می کند.)

## فصل ۵

### شرح دادگان و جمع‌بندی

برای انجام این پروژه مجموعاً بیش از ۹۶۰ ساعت موسیقی سنتی ایرانی در قالب ۱۶۸۹۵ قطعه از ۱۸۱ هنرمند جمع‌آوری شده که برای استفاده از آنها به عنوان ورودی در سامانه پیشنهادی که در بخش قبل به صورت مفصل توضیح داده شد، به برچسب‌گذاری و تغییر اسم آنها نیازی نبوده و سیستم از آنها به عنوان داده‌های خام، استفاده می‌کند.

برای تطابق این داده‌ها با مدل مورد استفاده، تنها تغییری که در قطعات داده شده، یکسان‌سازی تمام قطعات با تنظیم فرکانس نمونه برداری بر روی ۲۲۰۵۰ هرتز، فرمت mp۳ و تنظیم کانال آنها به صورت مونو است.

با توجه به هدف پژوهش ما که تشخیص ساز در قطعه موسیقی هست نیاز به بررسی فراوانی استفاده از سازهای مختلف در موسیقی سنتی بودیم تا سازهای پرتکرار را شناسایی کنیم، به این منظور ۱۰۰ آلبوم را به صورت تصادفی و نمونه در نظر گرفته و اعداد با توجه به اطلاعات منتشر شده رسمی در اطلاعات آلبوم‌ها استخراج شده‌اند که در ۱.۵ جدول قابل مشاهده هستند.

که همانطور از نتایج جدول مشخص است، مرز بین سازهای پرتکرار کامل مشخص بوده و می‌توانیم بدون کاهش کلیات برچسب سازها را یازده ساز اول یعنی تنبک، سنتور، کمانچه، نی، تار، سه تار، عود، دف، قیچک، دایره و رباب در نظر گرفته و سایر سازهای دیگر مانند سازهای موسیقی کلاسیک غربی هم که در موسیقی تلفیقی به کار رفته‌اند مانند پیانو، ویولن، ویولن سل را هم می‌توانیم صرف نظر کنیم اما قابلیت در نظر گرفتن آنها را هم داریم.

جدول ۱.۵: توزیع فراوانی سازهای استفاده شده در بخشی از دادگان

تنبک	سنتور	کمانچه	نی	تار	سه تار	عود	دف	قیچک	دایره
۶۴	۴۶	۴۵	۳۹	۳۷	۳۴	۳۲	۳۰	۱۴	۱۲

ریاب	کوزه	قانون	صراحی	دوتار	دهل	کمان	سبو	کاسه
۱۰	۴	۳	۳	۲	۲	۲	۱	۱

قوشمه	طبل	دمام	شمشال	کاخن	سرنا	تنبور	پرهیب	دیوان
۱	۱	۱	۱	۱	۱	۱	۱	۱

اصلی‌ترین عامل بازدارنده از انجام مطالعات پردازش سیگنال موسیقی به منظور بازیابی محتوای موسیقی در موسیقی سنتی ایرانی را نبود و محدودیت در تهیه‌ی دادگان مناسب بیان کردیم و ادعا کردیم که بدون نیاز به صرف زمان و هزینه برای برچسب‌زنی حجم زیادی از قطعات موسیقی و با استفاده از قطعات موسیقی به عنوان داده‌های خام، می‌توانیم به شناسایی خودکار سازهای به‌کار رفته در قطعات موسیقی سنتی و یا هرگونه استخراج اطلاعات پردازیم؛ بدین منظور ما از سامانه‌ی CLMR را پیشنهاد کردیم، که یک چارچوب یادگیری متضاد بازنمایی خود نظارتی است که بازنمایی‌های مفیدی از شکل‌های موج خام صدای موسیقی را آموخته در حالی که نیازی به پیش‌پردازش صدای ورودی نداشته و بدون دانش پایه آموزش داده می‌شود، که امکان پیش‌آموزش ساده و مستقیم روی مجموعه داده‌های موسیقی در مقیاس بی‌سابقه را فراهم می‌کند.

همچنین، نشان دادیم که با استفاده از این سامانه و با آموزش یک طبقه‌بند خطی در طبقه‌بندی موسیقی در دادگانی برچسب‌دار مانند، نوا [۴]، بازنمایی‌های طبق وظیفه‌ی خاصی آموخته‌شده را، می‌توان آزمایش کرد و در مقایسه با مدل کاملاً با نظارت به عملکرد رقابتی دست یافت. به علاوه، CLMR می‌تواند با استفاده از مجموعه دادگانی با موسیقی‌های دارای برچسب چندین برابر کم‌تر نیز، به عملکرد قابل مقایسه‌ای دست یابد و قابلیت انتقال خارج از دامنه بازنمایی‌های آموخته‌شده از پیش‌آموزش روی دادگانی کاملاً متفاوت از موسیقی را نشان داد.

این سامانه علاوه بر حل مشکل تهیه‌ی دادگان مناسب، امکان پیش‌برد مسئله به صورت‌های پیچیده‌تر را هم امکان‌پذیر می‌کند؛ به طور مثال مسئله‌ی تشخیص تعداد سازهای یک قطعه، طبق اطلاعاتی که در دست داریم، در پژوهش‌های صورت گرفته در موسیقی سنتی ایرانی مورد بررسی نبوده، زیرا تمام قطعات به صورت تکنوازی بوده و دارای برچسب تک در مورد ساز بوده‌اند که در واقعیت موسیقی سنتی ما، حجم این قطعات به طور کلی در برابر کل قطعات بسیار کم‌تر بوده و یا مسئله‌ی تشخیص دستگاه و گوشه بر روی قطعات تکنوازی، به نوعی دارای ساده‌سازی است. تلاش ما

در جمع‌آوری دادگانی با حجم بالا و دارای جامعیت بود به نوعی که نه در تعداد سازها و نه در نوع آنها و یا بودن یا نبودن آواز، هیچ‌گونه محدودیتی اعمال نشده است. در مجموع سادگی آموزش مدل بدون برجسب و بدون پیش پردازش صدا، همراه با نتایج دلگرم‌کننده به دست آمده با یک لایه خطی منفرد بهینه‌سازی شده برای یک کار موسیقی چالش‌برانگیز، پیشرفت‌های هیجان‌انگیزی به سمت یادگیری بدون نظارت بر روی صدای موسیقی خام است. و با استناد به این نتایج، امیدوارم هستیم که پژوهش‌های پیش‌تری با روش‌های این‌چنینی صورت گیرد و مسائلی که در موسیقی غربی به خوبی قادر به یافتن جواب برای آنها بوده‌اند، مانند؛ تشخیص خواننده، تشخیص متن موسیقی، تشخیص گام و یا حتی تکمیل قطعات قدیمی و تا حدی از بین رفته با استفاده از قطعاتی که در دست داریم و خیلی مسائل دیگر هم در موسیقی ما مورد بررسی قرار گیرند.

جدول ۲.۵: لیست هنرمندان

<p>حشمت رجب زاده احسان ذبیحی فر حسین علیشاپور علی کاظمی هومان رومی مسعود شعاری پرویز مشکاتیان پرواز همای علیرضا افتخاری روح الله خالقی همایون نصیری شهرام میرجلالی سعید نایب محمدی حسین رضا اسدی ابوالحسن صبا علی رستیمان حسن خدایی نیا بهرام حصیری پویا سرابی سامر حبیبی بهرام با جلان محمد رضا درویشی هوشنگ کامکار سلیمان سالک رسول اکبری علیرضا حاجی طالب سجاد پورقناد داوود آزاد ارسلان کامکار علی زند وکیلی ابوالفضل صادقی نژاد شهاب اکبری ایرج صهبایی مهیار طریحی سیامک شجریان میرزا علی چهارزی</p>	<p>ایرج بسطامی سعید نایب محمدی صهبا مطلبی امیر شریفی علیرضا فلسفی سام امیرحسین هوشنگ کامکار شهرام ناظری جلال تاج اصفهانی غلامحسین بنان پیمان یزدانیان پدارم بلوچی رضا پرویززاده فاضل جمشیدی اسدالله ملک جلیل شهناز سیاوش ایمانی تهمورس پورناظری بابک شریفی مجد بهنام معصومی صدیق تعریف داریوش طلایی بیژن کامکار انوشیروان روحانی حمید متبسم علیرضا گلبانگ رامین بحیرائی جاوید افسری راد مسعود جاهد گروه چارتار بهرام دهقانپار ابولحسن اقبال آذر بهر روز رونده سهیل مخبری کیوان علی محمدی مرتضی فلاحتی</p>	<p>کیوان ساکت حمید قنبری پیام جهانمانی نگین زمردی نگار خارکن سالار عقیلی حسن خان حسین علیزاده سید عبدالحسین مختاباد محمد رضا لطفی حسام اینانلو سیامک ایقانی بهر روز همتی حسین پرنیا مرتضی محجوبی حسن کسائی امیرعباس ستایشگر فرشاد جمالی پژمان حدادی پریچهر خواجه فرهاد فخرالدینی ناصر فرهنگ فر گروه کامکارها علی قمصری زکریا یوسفی سیامک بنایی محمد امین اکبرپور اشکان کمانگری اردوان کامکار آرش قاسمی فرشاد عباسی گروه بال و شال گروه بوم جمشید پورمهر رامین بحیرائی فرید یداللهی</p>	<p>همایون شجریان هومن مهدویان مجتبی عسگری فرزاد فضلی محمد رضا ابراهیمی گروه دستان عیسی غفاری فرامرز پایور محمد موسوی اردشیر کامکار فرید خردمند سحاب علم علی پژوهشگر علی اکبر شکارچی مهرداد پیکرزاده علی اصغر بهاری مجید درخشانی فردین خلعتبری رضا قلی میرزا ظلی پویان بیگلر حسین عمومی بهرام سارنگ داریوش پیرنیاکان علی رهبری علیرضا خشتی سامان احتشامی محمد باقر زینالی حسین بهروزی نیا صائب کاکاوند فردین کریم خاوری آرین رحمانیان علیرضا برزگر سپهر سراجی علی اصغر شاه زیدی محمد شمس علیرضا گلشن</p>	<p>محمد رضا شجریان وحید تاج محمد معتمدی سیامک جهانگیری علی اکبر مرادی علی مومنی نور علی برومند علی جهاندار (گروه عارف) جلال ذوالفقون علیرضا قربانی گروه موسیقی هفت خوان سینا علم لطف الله مجد ایرج رحمان پور حسین تهرانی رضا ورزنده سعید خلج سهیل حکمت آرا حسام الدین سراج مصباح قمصری مجید خلج عطا جنگوک حمید رضا نوربخش سید علی اصغر کردستانی مجید علیزاده همایون خرم محسن کرامتی رضا صبری علی انصاری سیاوش کامکار حسین خواجه امیری (ایرج) بامداد ملکی حسین مهران محمود حشمت محمد اسماعیلی ابوسعید مرضایی حسین حمیدی</p>
---	---	--	---	--

# واژه‌نامه فارسی به انگلیسی

Music information retrieval	بازیابی و استخراج اطلاعات موسیقیای
Overfitting	برازش بیش از حد
Online	برخط
Offline	برون خط
Pre training	پیش آموزش
Fine tuning	تنظیم دقیق
Data Augmentations	داده افزایی
Encoder	رمزگذار
Decoder	رمزگشا
Index	زیرنما
Encoder neural network	شبکه عصبی رمزگذار
Projector neural network	شبکه عصبی طرح ریز
Waveform	شکل موج
Pretext task	وظیفه بهانه‌ای
End to end learning	یادگیری انتها به انتها
Representation learning	یادگیری بازنمایی
Supervised learning	یادگیری با نظارت
Unsupervised learning	یادگیری بی نظارت
Self supervised learning	یادگیری خود نظارت
Semi supervised learning	یادگیری نیمه نظارت شده
Constractive learning	یادگیری متضاد



## کتاب نامه

- [۱] Vafaeian, Amir, et al. "Automatic Identification and Classification of the Iranian Traditional Music Scales (Dastgāh) and Melody Models (Gusheh): Analytical and Comparative Review on Conducted Research." *Human Information Interaction* 5.2 (2018): 46-72.
- [۲] Kinsler, L. E., and A. R. Frey. "E., 1962," *Fundamentals of Acoustics*."
- [۳] Spijkervet, Janne, and John Ashley Burgoyne. "Contrastive learning of musical representations." *arXiv preprint arXiv:2103.09410* (2021).
- [۴] Baba Ali, B., A. Gorgan Mohammadi, and A. Faraji Dizaji. "Nava: A Persian Traditional Music Database for the Dastgah and Instrument Recognition Tasks." *Advanced Signal Processing* 3.2 (2019): 125-134.
- [۵] Bokhove, Christian, and Christopher Downey. "Automated generation of 'good enough'transcripts as a first step to transcription of audio-recorded data." *Methodological innovations* 11.2 (2018): 2059799118790743.
- [۶] Sonata, I., et al. "Autonomous car using CNN deep learning algorithm." *Journal of Physics: Conference Series*. Vol. 1869. No. 1. IOP Publishing, 2021.
- [۷] Gandhi, Arun." Data Augmentation | How to use Deep Learning when you have Limited Data—Part 2." *Nanonets*, June 2021 , <https://nanonets.com/blog/data-augmentation-how-to-use-deep-learning-when-you-have-limited-data-part-2/>.

- [⌘] Koops, Hendrik Vincent, et al. "Annotator subjectivity in harmony annotations of popular music." *Journal of New Music Research* 48.3 (2019): 232-252.
- [Ⓐ] Cramer, Jason, et al. "Look, listen, and learn more: Design choices for deep audio embeddings." *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019.
- [Ⓜ] Tagliasacchi, Marco, et al. "Pre-training audio representations with self-supervision." *IEEE Signal Processing Letters* 27 (2020): 600-604.
- [Ⓝ] Al-Tahan, Haider, and Yalda Mohsenzadeh. "CLAR: contrastive learning of auditory representations." *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021.
- [Ⓞ] Saeed, Aaqib, David Grangier, and Neil Zeghidour. "Contrastive learning of general-purpose audio representations." *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021.
- [Ⓟ] Dosovitskiy, Alexey, et al. "Discriminative unsupervised feature learning with convolutional neural networks." *Advances in neural information processing systems* 27 (2014).
- [Ⓠ] Oord, Aaron van den, Yazhe Li, and Oriol Vinyals. "Representation learning with contrastive predictive coding." *arXiv preprint arXiv:1807.03748* (2018).
- [Ⓡ] Hjelm, R. Devon, et al. "Learning deep representations by mutual information estimation and maximization." *arXiv preprint arXiv:1808.06670* (2018).
- [Ⓢ] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *International conference on machine learning*. PMLR, 2020.
- [Ⓣ] Grill, Jean-Bastien, et al. "Bootstrap your own latent-a new approach to self-supervised learning." *Advances in Neural Information Processing Systems* 33 (2020): 21271-21284.

- [۱۸] Henaff, Olivier. "Data-efficient image recognition with contrastive predictive coding." International Conference on Machine Learning. PMLR, 2020.
- [۱۹] Chen, Ting, et al. "Big self-supervised models are strong semi-supervised learners." Advances in neural information processing systems 33 (2020): 22243-22255.
- [۲۰] Doersch, Carl, Abhinav Gupta, and Alexei A. Efros. "Unsupervised visual representation learning by context prediction." Proceedings of the IEEE international conference on computer vision. 2015.
- [۲۱] Zhang, Richard, Phillip Isola, and Alexei A. Efros. "Colorful image colorization." European conference on computer vision. Springer, Cham, 2016.
- [۲۲] Tian, Yonglong, Dilip Krishnan, and Phillip Isola. "Contrastive multiview coding." European conference on computer vision. Springer, Cham, 2020.
- [۲۳] He, Kaiming, et al. "Momentum contrast for unsupervised visual representation learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
- [۲۴] Tiu, Ekin. "Understanding Contrastive Learning". Towards Data Science , 7 Jan 2021, <https://towardsdatascience.com/>
- [۲۵] Xu, Dejing, et al. "Self-supervised spatiotemporal learning via video clip order prediction." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.
- [۲۶] Alwassel, Humam, et al. "Self-supervised learning by cross-modal audio-video clustering." Advances in Neural Information Processing Systems 33 (2020): 9758-9770.
- [۲۷] Ericsson, Linus, et al. "Self-Supervised Representation Learning: Introduction, Advances and Challenges." arXiv preprint arXiv:2110.09327 (2021).

[۲۸] کیانی ، مجید . هفت دستگاہ موسیقی ایران ، تهران ، انتشارات سوره مهر، ۱۳۹۳ .

- [۲۹] Kamien, Roger. Music: An Appreciation/brief. Mcgraw-Hill, 2000.
- [۳۰] Miller, Lloyd. Music and Song in Persia (RLE Iran B): The Art of Avaz. Routledge, 2012.
- [۳۱] Bengio, Yoshua, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives." IEEE transactions on pattern analysis and machine intelligence 35.8 (2013): 1798-1828.
- [۳۲] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).
- [۳۳] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).
- [۳۴] Pascual, Santiago, et al. "Learning problem-agnostic speech representations from multiple self-supervised tasks." arXiv preprint arXiv:1904.03416 (2019).
- [۳۵] Ravanelli, Mirco, et al. "Multi-task self-supervised learning for robust speech recognition." ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020.
- [۳۶] Engel, Jesse, et al. "Neural audio synthesis of musical notes with wavenet autoencoders." International Conference on Machine Learning. PMLR, 2017.
- [۳۷] محمودیان، سار. بنوشی، ایوب. "دسته‌بندی خودکار گام ماهور موسیقی ایرانی توسط یک شبکه عصبی مصنوعی." دومین کنفرانس بین المللی آکوستیک و ارتعاشات، دانشگاه صنعتی شریف، دی ۱۳۹۱
- [۳۸] عنداله زادگان، صابر. جعفری، شهرام. مرتضی، دیروند. "تشخیص خودکار دستگاه و گام موسیقی سنتی ایرانی مبتنی بر تکنوازی سازهای تار و سنتور به وسیله استخراج هوشمند." بیستمین کنفرانس ملی سالانه انجمن کامپیوتر ایران، دانشگاه فردوسی مشهد، اسفند ۱۳۹۳.
- [۳۹] LAYEGH, Mahmood ABBASI, Siamak HAGHIPOUR, and Yazdan NAJAFI SAREM. "Classification of the Radif of Mirza Abdollah a

canonic repertoire of Persian music using SVM method." Gazi University Journal of Science Part A: Engineering and Innovation 1.4 (2013): 57-66.

- [۴۰] RezezadehAzar, Shahla, et al. "Instrument-Independent Dastgah Recognition of Iranian Classical Music Using AzarNet." arXiv preprint arXiv:1812.07017 (2018).
- [۴۱] Abdoli, Sajjad. "Iranian Traditional Music Dastgah Classification." ISMIR. 2011.
- [۴۲] Beigzadeh, Borhan, and Mojtaba Belali Koochesfahani. "Classification of Iranian traditional musical modes (DASTGÄH) with artificial neural network." Journal of Theoretical and Applied Vibration and Acoustics 2.2 (2016): 107-118.
- [۴۳] Hajimolahoseini, Habib, Rassoul Amirfattahi, and Maryam Zekri. "Real-time classification of Persian musical Dastgahs using artificial neural network." The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012). IEEE, 2012.

## **Abstract**

Analysis and retrieval of music information has been considered by many researchers around the world today and the identification of instruments, lyrics, music genre, singer of the work and the extraction of any other information are among the branches of this field. In Iranian traditional music, there are also issues of instruments, Dastgahs and Gushehs, but unfortunately, they have not been given proper attention and few activities have been done with old methods and methods and with limited and Comprehensive datas.

Therefore, we tried to generalize the results of research samples on the data used in research by studying the use of advanced machine learning methods that are the most popular and trendy methods these days which are widespread use in analyzing and retrieving information of different types of music in the world. The validity of these methods is based on a database consisting of various tracks, without any limitation on the number of instruments, having or not having a song, and without changing the general titles or structure of them. We automatically or even how to extract any similar feature based on the content of Iranian traditional music songs with the help of this system.

In the following, all the details of the methods used in the proposed system, the type of system operation, the data used and any basic information and definitions related to the types of learning methods, music theory and activities in these fields are discussed in detail.



College of Science  
School of Mathematics, Statistics, and Computer Science

# Identify Traditional Musical Instruments with the Help of Machine Learning Methods

**Author name**  
**Maryam Hemati**

Supervisor:  
Dr. Bagher BabaAli

A thesis submitted in partial fulfillment of the requirements for  
the degree of B.Sc. in Computer Science

February 2022