مقاله پژوهشی

تشخیص دستگاه قطعه‌های موسیقی سنتی ایرانی بر مبنای استخراج توالی نت‌ها و استفاده از شبکه‌های LSTM

سینا غضنفری‌پور، مرتضی خادمی و عباس ابراهیمی مقدم

چکیده: دسته‌بندی دستگاه قطعات موسیقی سنتی ایرانی توسط کامپیوتر برای علاقه‌مندان موسیقی دستگاهی ایرانی، موضوعی بسیار جالب ولی پیچیده و چالش‌برانگیز است. این مسئله اولاً به دلیل کاربردهای فراوان آن در زمینه‌هایی مانند آهنگ‌سازی و آموزش موسیقی و ثانیاً به خاطر نیاز افراد عادی به کامپیوتر برای تشخیص دستگاه از اهمیت بالایی برخوردار است. در این مقاله روشی برای تشخیص دستگاه و زیردستگاه یک قطعه موسیقی ایرانی بر پایه استخراج نت‌های متوالی، دسته‌بندی سلسله‌مراتبی و استفاده از شبکه‌های LSTM ارائه شده است. در این روش، قطعه موسیقی در مرحله اول به یکی از سه دسته کلی، دسته‌بندی می‌شود. دسته اول صرفاً شامل دستگاه ماهور، دسته دوم شامل دستگاه‌های شور و نوا و دسته سوم شامل دستگاه‌های همایون، سه‌گاه و چهارگاه است. سپس برای هر دسته بسته به نوع آن، تعداد متفاوت دسته‌بندهای دیگر اعمال می‌شود تا این که یکی از ٦ دستگاه و یکی از ١١ زیردستگاه موسیقی سنتی ایرانی مشخص گردد. این تحقیق به هیچ سبک نوازندگی و ساز خاصی محدود نشده و تحت تأثیر سرعت و تکنیک‌های نوازندگی قرار نمی‌گیرد. قطعات برچسب‌گذاری شده در پایگاه داده "آرگ" که برای این تحقیق به وجود آمده است، به صورت تکنوازی هستند؛ اگرچه تعداد اندکی از آنها از هم‌نوایی سازهای کوبه‌ای (مانند تنبک) نیز در کنار سازهای ملودی بهره‌مند می‌باشند. نتایج نشان می‌دهند که تشخیص ٦ دستگاه اصلی و ١١ زیردستگاه به ترتیب با دقت میانگین ٧٤٫٥٪ و ٦٦٫٣٥٪ انجام گرفته که نسبت به تحقیقات کم‌شمار مشابه، نتایج بهتری دارد.

***کلیدواژه:* تشخیص دستگاه موسیقی، توالی نت، دسته‌بندی سلسله‌مراتبی، یادگیری عمیق، LSTM.**

۱- مقدمه

امروزه الگوریتم‌های یادگیری به طور گسترده‌ای در بسیاری از زمینه‌های مختلف مانند بازاریابی، امنیت و بازیابی اطلاعات، کاربرد دارند. این الگوریتم‌ها برای دسته‌بندی صوت و به ویژه موسیقی به اندازه دسته‌بندی تصویر متداول نیستند، اما تشخیص و جداسازی قطعات موسیقی بر اساس گونه[1]، محتوا (مانند سازبندی یا تکنوازی/ چندنوازی بودن قطعات) و حتی نوع نواخته‌شدن آن (مانند تکنیک‌های نوازندگی)، به یکی از کاربردهای مهم آنها تبدیل گردیده است. در زمینه موسیقی سنتی و دستگاهی ایرانی به دلیل پیچیدگی ذاتی و قواعد خاصی که درون خود دارد، تحقیق چندانی صورت نپذیرفته است، اما در مورد موسیقی غربی، به دلیل ساختار نسبتاً ساده‌تر و رواج بیشتر آن در دنیا، تحقیقات بسیار وسیع‌تری صورت گرفته است. در موسیقی غربی، محققین با استفاده از ویژگی‌های گوناگون توانسته‌اند قطعات موسیقی را بر اساس گونه آنها مانند پاپ[2]، کلاسیک[3]، جاز[4]، هیپ‌هاپ[5]، متال[6] و ... از یکدیگر جدا کنند. به طور مثال استفاده از ویژگی سرعت ضرب‌آهنگ[7] و ریتم[8] در تشخیص گونه موسیقی غربی بسیار مؤثر بوده است [۱]. در موسیقی ایرانی به دلیل دارابودن گام‌های گوناگون و اجرای موسیقی با سرعت ضرب‌آهنگ‌های متفاوت در یک دستگاه موسیقی، به کارگیری چنین ویژگی‌هایی معمولاً نتایج مناسبی را در پی نخواهد داشت.

در این مقاله پس از استخراج نت[9]های متوالی نواخته‌شده در یک قطعه موسیقی ایرانی، از LSTM[10] که از خانواده شبکه‌های عصبی بازگشتی[11] است و در پردازش وظایف مبتنی بر توالی، مناسب و کارامد می‌باشد، برای دسته‌بندی دستگاه‌ها و زیردستگاه‌های موسیقی ایرانی استفاده شده است. همچنین تعداد زیادی LSTM در یک ساختار سلسله‌مراتبی به کار گرفته شده‌اند تا دقت دسته‌بندی بهبود پیدا کند. سامانه طراحی‌شده قادر به انتساب یک قطعه موسیقی به یکی از ۶ دستگاه ماهور، شور، نوا، سه‌گاه، همایون و چهارگاه و همچنین ۱۱ زیردستگاه مختلف است. علت فقدان دستگاه راست‌پنج‌گاه در دسته‌بندی این است که در پایگاه داده مورد استفاده این تحقیق، قطعه موسیقی با برچسب دستگاه راست‌پنج‌گاه و تفکیک‌شده از ماهور نداریم و توالی نت‌ها (گام) که در این تحقیق اساس تفکیک بین دو دستگاه می‌باشد، در آن یکسان است. از این رو این دو دستگاه ادغام شده‌اند و با یک برچسب واحد در پایگاه داده موجود هستند. در سامانه پیشنهادی، ابتدا قطعات موسیقی به ۳ دسته تقسیم می‌شوند. قطعات دستگاه ماهور در دسته یک، قطعات در دستگاه‌های شور و نوا در دسته دو و قطعات در دستگاه‌های همایون، سه‌گاه و چهارگاه نیز به

این مقاله در تاریخ ۲۰ خرداد ماه ۱۴۰۰ دریافت و در تاریخ ۲۹ بهمن ماه ۱۴۰۰ بازنگری شد.

سینا غضنفری‌پور، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: ghazanfaripour@mail.um.ac.ir).

مرتضی خادمی (نویسنده مسئول)، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: khademi@um.ac.ir).

عباس ابراهیمی مقدم، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران، (email: a.ebrahimi@um.ac.ir).

1. Genre
2. Pop
3. Classical
4. Jazz
5. Hip Hop
6. Metal
7. Tempo
8. Rhythm
9. Note
10. Long Short Term Memory
11. Recurrent Neural Network

صورت جداگانه در دسته سه قرار گرفته‌اند. یک LSTM برای این دسته‌بندی آموزش دیده که در مرحله اول منجر به تشخیص دستگاه ماهور از سایر دستگاه‌ها می‌شود. در مرحله دوم می‌توان آهنگ‌های مربوط به دستگاه نوا را از شور و همایون را از سه‌گاه و چهارگاه تمایز داد. همچنین در همین مرحله زیردستگاه‌های ماهور دو و ماهور سُل نیز دسته‌بندی می‌شوند. در مرحله سوم دستگاه‌های سه‌گاه و چهارگاه نیز از یکدیگر تشخیص داده شده و در نتیجه تمامی دستگاه‌های موسیقی ایرانی مورد نظر و تعدادی از زیردستگاه‌ها دسته‌بندی می‌شوند. در مرحله چهارم تمامی زیردستگاه‌های مد نظر تفکیک می‌شوند.

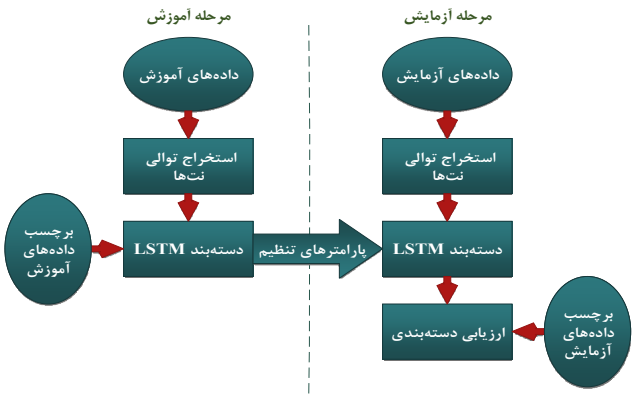
در ادامه مقاله، در بخش ۲ نگاهی اجمالی به مقوله موسیقی در تحقیقات روز دنیا و ارائه سابقه‌ای از تحقیق در زمینه تشخیص گونه موسیقی خواهیم داشت. در بخش ۳ روش پیشنهادی تشخیص دستگاه موسیقی ایرانی شرح داده شده و در بخش ۴، پس از معرفی پایگاه داده مورد استفاده، نتایج شبیه‌سازی و ارزیابی سامانه پیشنهادی ارائه گردیده است. این مقاله با نتیجه‌گیری در بخش ۵ پایان می‌یابد.

۲- سابقه دسته‌بندی گونه موسیقی

از روش‌های پرکاربرد و دارای سابقه در حل مسایل دسته‌بندی، روش‌های یادگیری ماشین است. این روش‌ها می‌توانند که راه حل‌های مؤثری را در پردازش سیگنال موسیقی مانند تشخیص ضرب و احساسات موسیقی و تشخیص آکورد[1] ارائه دهند [۲] تا [۱۱]. نخستین بار در سال ۲۰۰۲ دسته‌بندی گونه موسیقی غربی به عنوان یک مسئله بازشناسی مطرح شد و مجموعه‌ای از ویژگی‌های مناسب برای توصیف محتوای موسیقی از قبیل رنگ صدا[2]، ریتم و زیر و بمی[3] گردید [۱۲]. در [۱۳] و [۱۴] از ویژگی‌های دست‌ساز[4] که برای استخراج ویژگی‌های خاص توسط توسط انسان طراحی و انتخاب می‌شوند، همراه با دسته‌بندهای مختلف برای تشخیص گونه موسیقی استفاده شده است. در [۱۵] برای دسته‌بندی چهار نوع موسیقی فولک، راک، پاپ و زیتر چینی[5] از دسته‌بند ماشین بردار بردار پشتیبان[6] و ویژگی‌های ضرایب طیفی مل[7] استفاده گردیده و با الگوریتم جستجوی فاخته[8] سعی در بهینه‌سازی دسته‌بند شده است. علاوه علاوه بر ماشین بردار پشتیبان از RVM[9] و درخت تصمیم[10] [۱۶] و از الگوریتم نزدیک‌ترین همسایه Kام[11] و بیز[12] نیز استفاده گردیده که در آن ویژگی‌های طیفی چهار گونه کلاسیک، فولک، غزل و صوفی، استخراج شده و از طریق نگاشت در صفحه احساس برای دسته‌بندی گونه موسیقی مورد استفاده قرار گرفته است [۱۷]. در موردی دیگر برای مدل‌سازی موسیقی فولک چینی از میدان تصادفی شرطی[13] و برای خوشه‌بندی از GMM[14] استفاده شده تا گونه قطعه موسیقی را تخمین بزند [۱۸]. از ویژگی‌هایی همچون SR[15] و ZC[16] نیز می‌توان در تشخیص گونه موسیقی استفاده کرد [۱۹]. در [۲۰] از رنگ صدا و ضرایب طیفی مل به عنوان ویژگی و از الگوریتم نزدیک‌ترین همسایه Kام به عنوان دسته‌بند استفاده گردیده و از ویژگی‌هایی نظیر زیر و بمی و تداوم زمانی[17] برای این منظور بهره برده شده است [۲۱]. همچنین به منظور بهبود دسته‌بندی موسیقی با هدف کاهش پیچیدگی محاسباتی، از روش انتخاب ویژگی[18] استفاده شده است [۲۲].

از دیگر روش‌های یادگیری، یادگیری عمیق است که اخیراً به لطف دسترسی به واحدهای پردازش گرافیکی[19] که هزینه‌های محاسباتی مربوط به آموزش یادگیری عمیق را کاهش می‌دهد، مورد توجه قرار گرفته است. در مقایسه با ویژگی‌های دست‌ساز، اکثر روش‌های یادگیری عمیق به طور خودکار اطلاعات متمایزکننده را از نمونه‌های موجود در مجموعه داده‌ها کشف و استخراج می‌کنند. شبکه عصبی پیچشی یک روش یادگیری عمیق بسیار قدرتمند است [۲۳]. نخستین استفاده از آن در بازیابی اطلاعات موسیقی به سال ۲۰۱۲ برمی‌گردد [۲۴]. پس از آن، استفاده از شبکه عصبی پیچشی برای تشخیص و شناسایی خودکار آکوردهای موسیقی [۲۵]، تشخیص شروع هر نت موسیقی[20] [۲۶]، دسته‌بندی موسیقی به وسیله ویژگی‌های به دست آمده حاصل از اسپکتروگرام[21] [۲۷] و [۲۸] و حل سایر چالش‌های بازیابی اطلاعات موسیقی [۲۴] مورد توجه قرار گرفته است. در سال ۲۰۱۴ از یک مدل شبکه عصبی پیچشی آموزش داده شده و یک فرایند جداسازی صداهای هارمونیک[22] و کوبه‌ای[23] قبل از تولید اسپکتروگرام استفاده شده است [۲۹]. آموزش شبکه عصبی پیچشی برای دادگان بزرگ زمان‌بر می‌باشد و در بعضی تحقیقات برای مقابله مؤثر با این مشکل بر روی تنظیم صحیح پارامترهای شبکه تمرکز شده است [۳۰]. در [۳۱] اثرات ترکیب ویژگی‌های دست‌ساز و ویژگی‌های شبکه عصبی پیچشی روی تشخیص گونه موسیقی مورد بررسی قرار گرفته است. ترکیب ویژگی‌های دست‌ساز در حوزه تصویر (اسپکتروگرام) با سایر ویژگی‌های به دست آمده مستقیم از سیگنال صوتی نیز برای دسته‌بندی گونه موسیقی، مورد بررسی قرار گرفته است [۳۲] و [۳۳]. از روش‌های تعمیم‌یافته یادگیری عمیق (مثلاً CLNN[24] و MCLNN[25]) نیز در موسیقی می‌توان بهره برد [۳۴] و [۳۵]. همچنین روشی با استفاده از اسپکتروگرام و شبکه عصبی پیچشی با هدف جلوگیری از بیش‌برازش[26] داده‌ها ارائه شده است [۳۶]. از شبکه عصبی پیچشی و شبکه عصبی بازگشتی نیز در کنار یکدیگر می‌توان استفاده کرد [۳۷] و [۳۸]. در حوزه تشخیص شروع نت موسیقی نشان داده شده که شبکه عصبی پیچشی با پیش‌پردازش دستی کمتر، اما با هزینه محاسباتی

1. Chord
2. Timbre
3. Pitch
4. Handcraft
5. Guzheng
6. Support Vector Machine
7. MFCC
8. Cuckoo
9. Relevance Vector Machine
10. Decision Tree
11. K-Nearest Neighbors
12. Bayes
13. Conditional Random Field
14. Gaussian Mixture Models
15. Spectral Rolloff
16. Zero Crossings
17. Duration
18. Feature Selection
19. Graphics Processing Unit
20. Onset
21. Spectrogram
22. Harmonic
23. Percussion
24. ConditionaL Neural Networks
25. Masked Conditional Neural Networks
26. Overfitting



شکل ۱: ساختار کلی روش پیشنهادی.

بالاتر، قابل مقایسه با شبکه عصبی بازگشتی است [۳۹]. به علاوه، از متن آهنگ نیز در جهت بهبود نتایج جستجو به قیمت افزایش محاسبات می‌توان بهره جست [۴۰].

از آنجا که یک قطعه موسیقی چیزی جز توالی‌ای از نت‌ها که با قاعده خاص نواخته می‌شوند نیست، می‌توان از LSTM که برای حل مسایل پردازش سیگنال‌های مبتنی بر توالی زمانی مناسب است، استفاده کرد. برای مثال در [۴۱] از LSTM و GRU[1] که دو نوع از شبکه‌های عصبی بازگشتی هستند برای دسته‌بندی گونه موسیقی استفاده گردیده است. همچنین ادغام شبکه عصبی پیچشی و LSTM مورد بررسی قرار گرفته است [۴۲]. این تحقیق، ویژگی‌های مورد نظر را استخراج و اسپکتروگرام را به عنوان ورودی شبکه عصبی پیچشی استفاده کرده است. سپس خروجی‌های شبکه عصبی پیچشی را به لایه کاملاً همبند[2] وارد می‌کند. همچنین از LSTM و شبکه عصبی پیچشی برای تشخیص گونه و سبک[3] نواختن استفاده شده که ویژگی‌های ریتم و طیفی[4] نیز برای دسته‌بندی مورد استفاده قرار گرفته‌اند [۴۳]. در [۴۴] از ترکیب LSTM و ماشین بردار پشتیبان برای دسته‌بندی گونه موسیقی استفاده شده است. تمرکز بر روی جزئیات شبکه‌های LSTM نیز مورد بررسی قرار گرفته است. مثلاً [۴۵]، اثرات تعداد متفاوت لایه‌های LSTM را در دسته‌بندی زیرگونه‌های موسیقی جاز بررسی نموده و نتایج حاصل را با نتایج به دست آمده از روش‌های ماشین بردار پشتیبان و الگوریتم نزدیک‌ترین همسایه Kام همراه با ویژگی‌های ضرایب طیفی مل مقایسه کرده است.

موسیقی دیگر ملل، همچون موسیقی عربی [۴۶]، موسیقی هندی [۴۷] تا [۵۰] و یا اندونزیایی [۵۱] در تحقیقات مورد توجه بوده‌اند، اما در زمینه موسیقی ایرانی تحقیقات زیادی انجام نشده است. از تحقیقات انجام‌شده در زمینه موسیقی ایرانی می‌توان به سامانه طراحی‌شده در [۵۲] اشاره کرد که پس از دریافت یک سیگنال صوتی، با استفاده از شبکه عصبی با توابع پایه شعاعی[5]، ماهور یا غیر ماهور بودن قطعه ورودی را مشخص می‌کند. همچنین پایگاه داده‌ای شامل ۱۳۵ قطعه موسیقی حاصل از نواخته‌شدن ساز سه‌تار مورد استفاده قرار گرفته است. عدم شناسایی سایر دستگاه‌های موسیقی ایرانی، تعداد داده بسیار کم و عدم استفاده از برخی تکنیک‌های نوازندگی نظیر ویبراسیون و سرش‌نغمه‌ای[6] که برای زیباترکردن قطعات

1. Gated Recurrent Unit
2. Fully Connected
3. Style
4. Spectral
5. Radial Basis Functions
6. Glissando

شکل ۲: روند استخراج بردار توالی نت.

استفاده می‌شوند، از جمله مسایل حل‌نشده این تحقیق هستند. در یکی دیگر از پژوهش‌ها [۵۳] با استفاده از شبکه عصبی و گام‌های موجود در قطعات و استخراج گام‌های هر یک از دستگاه‌ها، دستگاه قطعاتی را که گام‌ها و نت‌های آن از قبل مشخص است، شناسایی می‌کند. اما این روش خودکار نیست و نیازمند یک ناظر جهت استخراج فواصل موجود در هر قطعه موسیقی می‌باشد. در [۵۴] از ویژگی ضرایب طیفی مل و دسته‌بند ماشین بردار پشتیبان برای دسته‌بندی دستگاه‌ها و سازهای موسیقی ایرانی با استفاده از پایگاه داده "نوا" که شامل ۱۷۸۶ قطعه موسیقی با طول‌های متفاوت می‌باشد، استفاده گردیده است. دقت این روش برای دسته‌بندی دستگاه در بهترین حالت ۳۴ درصد می‌باشد.

۳- روش پیشنهادی

شکل ۱ ساختار کلی روش پیشنهادی را نشان می‌دهد. ابتدا داده‌های آموزشی برای استخراج توالی نت‌های نواخته‌شده، به سامانه وارد گردیده و سپس فرکانس‌های هر پیک انرژی سیگنال (ویژگی زیر و بمی صدا) استخراج شده و با نگاشت این فرکانس‌ها به نت‌ها، بردار توالی نت ساخته می‌شود. سپس این بردار برای آموزش به مدل شبکه LSTM وارد می‌گردد. پس از مرحله آموزش و تنظیم پارامترهای شبکه LSTM، داده‌های آزمایش نیز پس از استخراج بردار توالی نت، توسط شبکه دسته‌بندی شده و نهایتاً جهت ارزیابی سامانه پیشنهادی، با برچسب‌های از پیش تعیین شده مقایسه می‌گردد.

۳-۱ استخراج توالی نت‌ها

شکل ۲ فرایند تشخیص و استخراج نت‌های نواخته‌شده در یک قطعه موسیقی را نشان می‌دهد. این مرحله، اصلی‌ترین قدم در تشخیص دستگاه

موسیقی می‌باشد. برای انجام این مرحله، ابتدا باید فرکانس اساسی نت‌ها استخراج شود. هرچه این فرکانس‌ها در یک قطعه موسیقی با دقت بیشتری استخراج شوند، دستگاه آن قطعه نیز درست‌تر تعیین خواهد شد. برای این منظور در ابتدا محدوده تحلیل فرکانسی برای استخراج نت‌ها، بین ۲۶ تا ۲۰۰۰ هرتز تعیین می‌شود که حدوداً معادل نت‌های "لا صفر" (A۰) تا "سی ششم" (B۶) است. این محدوده، ۷۵ نت اصلی را شامل می‌شود و بنابراین برای کشف نت‌های حاصل از سیگنال موسیقی، پوشش کامل فرکانسی را ایجاد می‌کند. روش‌های متفاوتی را برای استخراج فرکانس اساسی نت‌ها می‌توان استفاده کرد که در اینجا از پنجره‌گذاری، محاسبه تابع خودهمبستگی و فیلترکردن استفاده شده است.

سیگنال موسیقی یک سیگنال ایستا نیست و مشخصات فرکانسی آن با زمان تغییر می‌کند. تحت این شرایط، ابزارهای مختلف مانند تبدیل فوریه یا تابع خودهمبستگی، کارایی لازم را ندارند. بنابراین هنگام بررسی سیگنال صوتی آن را به قسمت‌های کوتاه زمانی تقسیم می‌کنیم، به طوری که در آن بازه زمانی بتوان سیگنال را با تقریب خوبی ایستا فرض کرده و ابزارهای فوق، کارامد گردند. سپس به آنالیز هر قسمت و استخراج ویژگی‌های مد نظر آن می‌پردازیم که این عمل پنجره‌گذاری نام دارد. پنجره‌ها انواع گوناگون با نقاط قوت و ضعف متفاوت دارند. به طور مثال پنجره مستطیلی در عین سادگی در لبه‌ها ایجاد ناپیوستگی و اعوجاج می‌کند. برای کاهش این اعوجاج از پنجره‌هایی استفاده می‌گردد که در ابتدا و انتها مقادیر نزدیک صفر داشته و به مرور در مرکز پنجره به یک نزدیک می‌گردند. در روش پیشنهادی این مقاله از پنجره هنینگ[1] استفاده می‌شود. انتخاب طول پنجره نیز به مانند انتخاب نوع آن حایز اهمیت است. طول پنجره را باید طوری انتخاب کنیم که در آن فقط یک نت امکان حضور داشته باشد. همچنین برای جلوگیری از هدررفتن اطلاعات سیگنال صوتی، پنجره‌ها مقداری با یکدیگر هم‌پوشانی دارند. این میزان معمولاً بین ۲۵ تا ۷۵ درصد طول پنجره است.

در این تحقیق، پس از پنجره‌گذاری سیگنال و استخراج قطعات با طول و میزان هم‌پوشانی مشخص، برای شناسایی فرکانس‌های اساسی در سیگنال هارمونیک موسیقی از تابع خودهمبستگی تعدیل‌شده بهره جسته‌ایم. تابع خودهمبستگی برای پیداکردن الگوهای تکراری یک سیگنال (مانند ریتم یا گام در سیگنال موسیقی) بسیار مفید است [۵۵] تا [۵۷]. از سوی دیگر ویژگی محدودیت دامنه فرکانسی سیگنال موسیقی نیز با خصوصیت کارامدی این روش در فرکانس‌های متوسط و پایین بسیار سازگار می‌باشد. به جای تابع خودهمبستگی از روش‌های دیگری همچون فیلتر تخمین زیر و بمی[2] [۵۸]، تعیین زیر و بمی کپستروم[3] [۵۹] و مجموع باقیمانده هارمونیک‌ها[4] [۶۰] نیز می‌توان برای استخراج نت‌ها استفاده کرد اما در این تحقیق به علت پیچیدگی محاسباتی بیشتر آنان نسبت به تابع خودهمبستگی از آنها استفاده نشده است.

پس از اعمال پنجره و محاسبه تابع خودهمبستگی، اقدام به اعمال فیلتر به عنوان یک عمل پس‌پردازش[5] می‌شود. از آنجا که فرکانس‌های اساسی سیگنال موسیقی در فرکانس‌های پایین و متوسط پدیدار می‌شوند، از فیلتر پایین‌گذر استفاده می‌گردد. در روش پیشنهادی از فیلتر میانه[6] که یک فیلتر پایین‌گذر است به دلیل عملکرد مناسب، سادگی و حفظ جزئیات سیگنال استفاده شده است. فیلتر میانه فیلتری غیر خطی می‌باشد و در بین فیلترهای FIR[7] معمولاً عملکرد مناسب‌تری از خود در نرم‌کردن سیگنال به نمایش می‌گذارد. اعمال فیلتر میانه با طول مشخص (معمولاً عددی فرد) باعث می‌شود تا فرکانس‌های استخراج‌شده دور از میانه حذف گردند و بدین وسیله فرکانس‌های اساسی سیگنال موسیقی، دقیق‌تر و با خطای کمتری تشخیص داده شوند. اعمال این فیلتر، فرکانس‌های حاصل از پنجره‌های متوالی با فاصله زمانی کم را به یک فرکانس نمونه که همان مقدار عنصر میانی است، تبدیل می‌کند و بنابراین کاهش طول بردار توالی نت را نیز در پی دارد.

در مرحله بعد و پس از استخراج فرکانس‌های اساسی قطعه موسیقی، این فرکانس‌های متوالی را توسط بلوک "استخراج نت" به نت تبدیل کرده و برداری از توالی نت‌های نواخته‌شده به دست می‌آوریم. برای تحقق این هدف صرفاً لازم است که فرکانس مرکزی و بازه فرکانسی هر نت مشخص باشد. به طور مثال نت "می‌کرن" دارای فرکانس مرکزی ۳۲۱ هرتز و بازه ۳۱۵ الی ۳۲۸ هرتز می‌باشد. حال هر فرکانس اساسی خروجی از مرحله قبل که در این بازه است به این نت نگاشت می‌شود. فرکانس مرکزی و طول بازه حول آن برای نت‌های مختلف متفاوت است. در این تحقیق، استخراج نت‌های یک قطعه موسیقی به نت‌های اصلی محدود نشده و نت‌های دیگر حاصل از نیم‌پرده‌ها و ربع‌پرده‌ها نیز که در موسیقی ایرانی وجود دارند، در نظر گرفته شده‌اند. این کار باعث فواصل فرکانسی کمتر بین نت‌های آن (در مقایسه با موسیقی غربی) و دشواری بیشتر در تشخیص نت می‌شود، اما برای تشخیص دستگاه موسیقی ایرانی لازم است. در انتهای این مرحله، برداری به طول مشخص و حاصل از توالی نت‌های موجود در هر قطعه موسیقی به وجود می‌آید که به دسته‌بندهای LSTM اعمال می‌شود.

۳-۲ دسته‌بند LSTM

شبکه‌های عصبی بازگشتی از چارچوب‌های محبوب روش‌های یادگیری هستند. یک شبکه عصبی بازگشتی با شبکه‌های عصبی سنتی تفاوت دارد و می‌تواند اطلاعات حالت‌های گذشته را به خاطر سپرده و با کمک آن، حالت فعلی را پیش‌بینی کند. اما علی‌رغم این توانایی، به شکاف اطلاعاتی گذشته و حال حساس است و اگر این شکاف بزرگ باشد، کارایی آن به شدت کاهش می‌یابد [۶۱]. شبکه LSTM زیردسته‌ای از شبکه‌های عصبی بازگشتی است. اگر LSTM ویژگی مهمی را در دنباله ورودی و گام‌های ابتدایی تشخیص دهد، می‌تواند این اطلاعات را طی مسیری طولانی منتقل کند. از این رو LSTM برای پردازش، دسته‌بندی و پیش‌بینی سری‌های زمانی با تأخیر زیاد مناسب است و مشکل وابستگی طولانی‌مدت شبکه‌های عصبی بازگشتی را حل کرده است و به طور ویژه برای سیگنال‌های با وابستگی زمانی مانند سیگنال موسیقی، به خوبی عمل می‌کند. در LSTM با مفاهیم جدیدی مواجه می‌شویم که در شبکه عصبی بازگشتی وجود نداشتند. در این شبکه سه دروازه[8] (فراموشی، ورودی و خروجی) و یک سلول حافظه که از طریق آن شبکه نسبت به کنترل جریان داده درون خود اقدام می‌کند، وجود دارد. شبکه‌های LSTM ساختار زنجیره‌ای داشته و ماژول تکرارشونده‌ای با ۴ لایه دارند که با یکدیگر در تعامل هستند. ساختار LSTM در تحقیقات گوناگونی از

1. Hanning
2. Pitch Estimation Filter
3. Cepstrum Pitch Determination
4. Summation of Residual Harmonics
5. Postprocessing
6. Median
7. Finite Impulse Response
8. Gate

شکل ۳: لایه‌های LSTM طراحی‌شده.

جدول ۱: توزیع قطعات پایگاه داده "آرگ" بر حسب نوع دستگاه.

نوع دستگاه	سه‌گاه	نوا	همایون	ماهور	شور	چهارگاه
تعداد قطعه	۹۰	۱۴۹	۹۱	۷۷	۷۶	۱۲۳

جمله [۶۲] آورده شده که پارامترهای وزن و بایاس شبکه LSTM در طول آموزش شبکه محاسبه می‌شوند.

پیکربندی شبکه LSTM پیشنهادی برای این تحقیق دارای ۵ لایه می‌باشد که در شکل ۳ آمده است. تمامی دسته‌بندهای سلسله‌مراتبی شکل ۴ از این ساختار استفاده می‌کنند و فقط در لایه خروجی تفاوت دارند. همچنین در آموزش این شبکه‌ها می‌توان از انواع بهینه‌سازها استفاده کرد که رایج‌ترین آنها بهینه‌ساز آدام[1] است. این بهینه‌ساز قابلیت تنظیم کاهش نرخ یادگیری برای رسیدن به دقت حداکثری را دارد [۶۳]. در این تحقیق نیز از همین نوع بهینه‌ساز استفاده شده است.

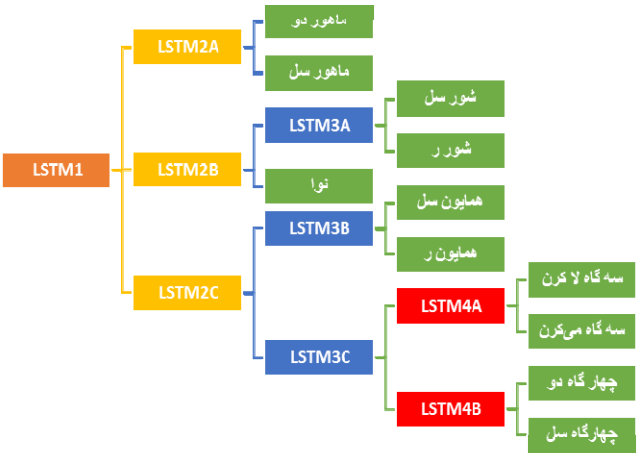
شکل ۴ ساختار سلسله‌مراتبی طراحی‌شده در این تحقیق برای تشخیص دستگاه و زیردستگاه را به صورت نمودار درختی به تصویر کشیده است. دسته‌بندی چندمرحله‌ای پیشنهادی شامل ۹ عدد LSTM فوق است. در مرحله اول آزمایش، هر قطعه موسیقی ورودی پس از تبدیل به بردار توالی نت، ابتدا توسط LSTM۱ دسته‌بندی می‌شود تا دریابیم که کدام یک از سه دسته ابتدایی ماهور، شور- نوا یا همایون- سه‌گاه- چهارگاه است. چنانچه قطعه موسیقی، ماهور تشخیص داده شود، برای تعیین یکی از دو زیردستگاهش وارد اولین دسته‌بند مرحله دوم یعنی LSTM۲A می‌گردد. اما چنانچه جزو دسته دوم یا سوم تشخیص داده شود، برای تفکیک دقیق‌تر به ترتیب وارد یکی از دو دسته‌بند دیگر مرحله دوم یعنی LSTM۲B یا LSTM۲C خواهد شد. در این مرحله شور از نوا و همایون از زوج سه‌گاه- چهارگاه متمایز می‌گردد. دسته‌بندهای مرحله سوم یعنی LSTM۳A، LSTM۳B و LSTM۳C به ترتیب زیردستگاه‌های شور و همایون و دو دستگاه سه‌گاه و چهارگاه را از یکدیگر جدا می‌کنند. دسته‌بندهای مرحله چهارم یعنی LSTM۴A و LSTM۴B نیز به ترتیب وظیفه تشخیص زیردستگاه‌های سه‌گاه و چهارگاه را بر عهده دارند.

٤- نتایج شبیه‌سازی و ارزیابی

در این بخش پس از معرفی پایگاه داده تولید و استفاده‌شده در تحقیق با نام "آرگ"[2]، نتایج شبیه‌سازی سامانه پیشنهادی ارائه خواهد گردید.

٤-١ پایگاه داده "آرگ"

در زمینه موسیقی غربی پایگاه‌های داده متفاوتی مانند LMD[3]، Ballroom و ISMIR۲۰۰۴ وجود دارند، اما پرکاربردترین آنها در زمینه تشخیص گونه موسیقی GTZAN است. در زمینه موسیقی ایرانی پایگاه



شکل ۴: LSTM سلسله‌مراتبی طراحی‌شده.

داده مشخص، شامل قطعات با طول یکسان و برچسب‌دار در دسترس نیست. در [۶۴] از پایگاه داده‌ای استفاده شده که شامل ۱۱۴۳ قطعه ۳۰ ثانیه است اما آن پایگاه داده صرفاً شامل دستگاه ماهور می‌شود. بنابراین برای تحقیق در زمینه تشخیص دستگاه‌های موسیقی ایرانی به پایگاه داده جدیدی نیاز بود. در این تحقیق اقدام به تولید پایگاه داده‌ای به نام "آرگ" مبتنی بر دستگاه‌های موسیقی ایرانی شده است. با توجه به نظرسنجی‌ای که از اهالی موسیقی، نوازندگان و آهنگسازان صورت پذیرفت، این نتیجه حاصل گردید که شنیدن ۲۰ ثانیه برای تشخیص دستگاه موسیقی ایرانی برای فردی آشنا به موسیقی، می‌تواند تا حد زیادی کفایت کند. از همین رو طول قطعات ۲۰ ثانیه و ترجیحاً به صورت تکنوازی در نظر گرفته شد که به تفکیک منابع صوت نیاز نباشد. این قطعات توسط سازهایی مانند پیانو، تار، سه‌تار، عود، کمانچه، نی، سنتور و ویولن نواخته شده‌اند و همچنین تعدادی از این قطعات با همراهی سازهای کوبه‌ای مانند تنبک نواخته شده‌اند.

اگرچه موسیقی کلاسیک ایرانی شامل ۱۲ دسته (۷ دستگاه و ۵ آواز) می‌باشد، ولی برای شروع تحقیق در این زمینه با توجه به پیچیدگی ذاتی موسیقی ایرانی و هم‌پوشانی بسیاری از این دسته‌ها، قطعاتی از شش دستگاه سه‌گاه، چهارگاه، همایون، شور، نوا و ماهور انتخاب شده‌اند. از آنجا که تشخیص دستگاه ماهور و راست‌پنج‌گاه حتی برای موسیقیدان‌ها نیز به دلیل شباهت زیاد گام آنها چالشی جدی است، در پایگاه داده "آرگ" برای هر دو نوع دستگاه مذکور از برچسب ماهور استفاده شده است. نهایتاً ۶۰۶ قطعه موسیقی ۲۰ثانیه‌ای در پایگاه داده "آرگ" قرار گرفته که هر کدام به طور دقیق، با نظر متخصصین برچسب‌گذاری شده‌اند. در جدول ۱ توزیع قطعات موسیقی این پایگاه داده را روی شش دستگاه مشاهده می‌کنید. در این پایگاه داده، ۴۰۵ قطعه موسیقی برای آموزش، ۱۳۰ قطعه برای اعتبارسنجی[4] و ۷۱ قطعه برای آزمایش استفاده شده است.

٤-٢ نتایج شبیه‌سازی

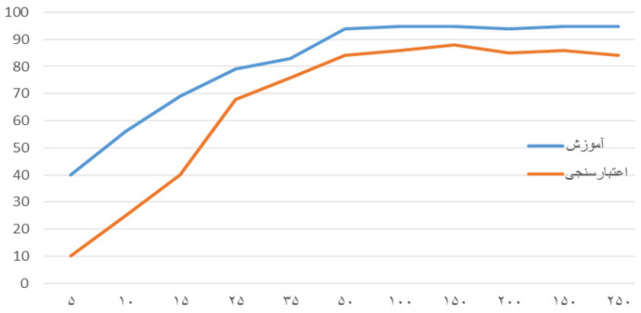
در روش پیشنهادی ارائه‌شده، پارامترهای متفاوتی برای بهبود عملکرد سامانه مورد نظر انتخاب گردیده است. همان طور که در توصیف پایگاه داده "آرگ" گفته شد، طول قطعات موسیقی به عنوان اولین پارامتر، برای کلیه قطعات پایگاه داده ۲۰ ثانیه انتخاب گردیده است. در مرحله پنجره‌گذاری روی این قطعات نیز از پنجره هنینگ به طول ۶۰ میلی‌ثانیه و میزان هم‌پوشانی ۲۵ درصد استفاده شده است. نتیجه این مرحله و

1. Adam

۲. دسترسی به پایگاه داده آرگ از طریق آدرس زیر امکان‌پذیر است:
http://dx.doi.org/10.13140/RG.2.2.20688.99842

3. Latin Music Database

4. Validation



شکل ۵: دقت داده‌های آموزش و اعتبارسنجی بر حسب تعداد دوره با اندازه پشته ۴۵.

اعمال تابع خودهمبستگی روی پنجره‌ها، استخراج برداری با حدود ۴۰۰ فرکانس اساسی می‌باشد. در مرحله فیلترینگ، طول فیلتر میانه (اعمال‌شده روی بردار اخیر) نیز برابر ۹ انتخاب شده است؛ بدین معنا که از میان ۹ فرکانس اساسی استخراج‌شده، فرکانس میانه را برای تعیین نت انتخاب می‌کند. بدین وسیله بردار توالی نت با طول بین ۴۰ تا ۴۵ نت بر حسب میزان حذف سکوت‌های درون قطعه به دست می‌آید. برای این که طول بردارها مساوی باشد، عدد ۴۵ یعنی حداکثر تجربی را برای طول بردارهای توالی نت انتخاب می‌کنیم و در شرایطی که تعداد کمتر از این بود، آخرین نت به دست آمده در انتهای بردار تکرار می‌گردد. این طول برای بردار توالی نت را می‌توان با تغییر طول پنجره هنینگ و میزان هم‌پوشانی بین آنها به هر مقدار دلخواه رساند، اما با توجه به این که طبق تجارب شنیداری، معمولاً در هر ۰٫۵ ثانیه یک نت نواخته می‌شود، عدد منتخب فوق برای تشخیص دستگاه در قطعات ۲۰ثانیه‌ای مناسب به نظر می‌رسد. همچنین دسته‌بند به کار گرفته شده دارای پارامترهای زیادی برای اطمینان از عملکرد مناسب می‌باشد که باید به صورت دقیق تنظیم شوند. پارامتر اول اندازه پشته[1] است. برای تنظیم این پارامتر به تعداد کل داده‌های آموزش، سرعت آموزش کل شبکه و دقت بهتر دسته‌بندی توجه می‌شود. بر این اساس اندازه پشته ۴۵ در نظر گرفته شده و بنابراین تمامی ۴۰۵ داده آموزش طی ۹ پشته به شبکه وارد می‌شوند. پارامتر مهم دیگر دوره‌های[2] مختلف آموزش است. همان گونه که در شکل ۵ نمایش داده شده است، با توجه به اندازه پشته و تعداد کل داده‌های پایگاه داده "آرگ"، مقادیر مختلفی بین ۵ تا ۲۵۰ دوره، مورد آزمایش قرار گرفته است. طبق شکل مذکور، بعد از طی حدود ۵۰ دوره، شبکه به نقطه حداقل خطا نزدیک شده و بعد از آن، مقدار دقت افزایش خاصی پیدا نمی‌کند. در نتیجه مقدار ۵۰ برای تعداد دوره آموزش با هدف حداکثر دقت آزمایش و سرعت مناسب آموزش در نظر گرفته می‌شود.

دلیل اصلی استفاده از طرح سلسله‌مراتبی، رضایت‌بخش نبودن نتایج LSTM به فرم دسته‌بندی واحد است، زیرا یک سیستم غیر سلسله‌مراتبی با تعداد خروجی زیاد در لایه آخر به خوبی عمل نمی‌کند. به طور مثال دقت تشخیص برای دسته‌بندی ۶ دستگاه و ۱۱ زیردستگاه به ترتیب برابر با ۴۹٫۶٪ و ۴۳٫۱٪ است که با افزایش یک لایه اضافه با ۲۵۶ نورون، این نتایج به ۵۱٫۴٪ و ۴۶٫۹٪ ارتقا پیدا کرده که این میزان دقت سامانه همچنان که نشان داده خواهد شد راضی‌کننده نیست. به همین جهت سامانه پیشنهادشده با طرحی سلسله‌مراتبی ارائه گردیده که تعداد خروجی لایه آخر هر دسته‌بند کاهش یابد.

در این ساختار سلسله‌مراتبی به جز در LSTM۱، انتشار خطا در هر

1. Batch
2. Epoch

جدول ۲: نتایج LSTMهای ساختار سلسله‌مراتبی.

دسته‌بندهای LSTM	دقت
LSTM۱	۸۷٫۳٪
LSTM۲A	۶۹٫۲٪
LSTM۲B	۶۴٪
LSTM۲C	۸۱٫۸٪
LSTM۳A	۵۶٫۳٪
LSTM۳B	۶۶٫۷٪
LSTM۳C	۶۶٫۷٪
LSTM۴A	۷۷٫۸٪
LSTM۴B	۵۸٫۳٪

جدول ۳: میزان دقت سامانه پیشنهادی در تشخیص هر دستگاه بر حسب درصد.

	نوا	چهارگاه	سه‌گاه	همایون	شور	ماهور
ماهور	۰٫۰	۸٫۳	۰٫۰	۰٫۰	۰٫۰	**۹۱٫۷**
شور	۱۶٫۷	۰٫۰	۰٫۰	۸٫۳	**۷۵٫۰**	۰٫۰
همایون	۰٫۰	۰٫۰	۰٫۰	**۶۶٫۷**	۳۳٫۳	۰٫۰
سه‌گاه	۰٫۰	۸٫۳	**۷۵٫۰**	۰٫۰	۰٫۰	۱۶٫۷
چهارگاه	۰٫۰	**۷۵٫۰**	۰٫۰	۲۵٫۰	۰٫۰	۰٫۰
نوا	**۶۳٫۶**	۹٫۱	۰٫۰	۰٫۰	۲۷٫۳	۰٫۰

طبقه از سامانه طراحی‌شده باعث کاهش دقت در دسته‌بندهای بعدی می‌شود که نتایج هر کدام به صورت جداگانه در جدول ۲ آمده است.

دقت سامانه پیشنهادی برای تشخیص هر یک از شش دستگاه اصلی در جدول ۳ نشان داده شده است. همان طور که جدول نشان می‌دهد، دقت تشخیص دستگاه ماهور ۹۱٫۷٪ است که کارامدی روش را در تشخیص این دستگاه نشان می‌دهد. به طور مثال برای دستگاه ماهور، قطعه mahour(۶) که نت‌ها در آن به خوبی قابل تفکیک هستند، دستگاه متناظر به خوبی تشخیص داده شده است. اگرچه عملکرد سامانه پیشنهادی برای برخی از دستگاه‌ها مانند ماهور، برجسته بوده است اما دسته‌بندی برخی دستگاه‌های شبیه به هم، سخت‌تر است. به طور مثال میزان دقت سامانه در تشخیص دستگاه نوا، کمترین مقدار را داراست و اعداد نسبتاً بزرگ در سلول‌های تلاقی شور و نوا این دشواری دسته‌بندی را به دلیل شباهت دستگاه‌ها به خوبی نشان می‌دهند. این مسئله روی دقت میانگین سامانه اثر منفی می‌گذارد. برای اندازه‌گیری این اثر منفی، دو دستگاه را ادغام کرده و یک دسته‌بندی ۵کلاسه را توسط سامانه، آزمایش می‌کنیم. تحت این شرایط، میانگین دقت حدود ۵٪ بهبود یافته و از ۷۴٫۵٪ به ۷۹٫۹۴٪ می‌رسد.

طبق جدول ۴، میانگین دقت سامانه برای تشخیص ۱۱ زیردستگاه ۶۶٫۳۵٪ است و زیردستگاه‌های سه‌گاه و چهارگاه دارای کمترین دقت تشخیص می‌باشند. عواملی چون شباهت سه‌گاه با چهارگاه و همچنین اشتراک نت شاهد در چهارگاه "سل" و همایون "سل"، باعث ایجاد خطا در تشخیص دو زیردستگاه مذکور شده است. البته دلایل گوناگونی باعث جلوگیری از افزایش دقت میانگین در تشخیص زیردستگاه‌ها می‌شود. وجود نت شاهد که نت پرتکرار نواخته‌شده در هر زیردستگاه است، صرفاً یکی از عوامل محدودکننده‌ای است که نتایج آن در این جدول به چشم می‌خورد.

در هنگام نواخته‌شدن قطعات موسیقی، نت‌ها با سرعت بالایی اجرا می‌شوند و ممکن است گاهی چندین نت با یکدیگر ترکیب گردند. بر این اساس ممکن است هنگام انجام این محاسبات بر روی یک قطعه کامل موسیقی، نت‌ها به صورت کامل و منطبق با آنچه که نواخته می‌شوند،

جدول ۴: میزان دقت سامانه پیشنهادی در تشخیص هر زیردستگاه بر حسب درصد.

	نوا	چهارگاه سل	چهارگاه دو	سه‌گاه می‌کرن	سه‌گاه لاکرن	همایون ر	همایون سل	شور ر	شور سل	ماهور سل	ماهور دو
ماهور دو	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۱۶/۷	**۸۳/۳**
ماهور سل	۰/۰	۱۶/۷	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	**۶۶/۶**	۱۶/۷
شور سل	۱۶/۷	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۱۶/۷	۰/۰	**۶۶/۶**	۰/۰	۰/۰
شور ر	۱۶/۷	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	**۸۳/۳**	۰/۰	۰/۰	۰/۰
همایون سل	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	**۶۶/۶**	۰/۰	۳۳/۴	۰/۰	۰/۰
همایون ر	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	**۶۶/۶**	۰/۰	۳۳/۴	۰/۰	۰/۰	۰/۰
سه‌گاه لاکرن	۰/۰	۰/۰	۱۶/۶	۰/۰	**۵۰/۰**	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰	۱۶/۷
سه‌گاه می‌کرن	۰/۰	۰/۰	۰/۰	**۶۶/۶**	۱۶/۷	۰/۰	۰/۰	۰/۰	۰/۰	۱۶/۷	۰/۰
چهارگاه دو	۰/۰	۱۶/۷	**۶۶/۶**	۰/۰	۰/۰	۱۶/۷	۰/۰	۰/۰	۰/۰	۰/۰	۰/۰
چهارگاه سل	۰/۰	**۵۰/۰**	۱۶/۷	۰/۰	۰/۰	۰/۰	۳۳/۳	۰/۰	۰/۰	۰/۰	۰/۰
نوا	**۶۳/۶**	۰/۰	۹/۱	۰/۰	۰/۰	۰/۰	۰/۰	۹/۱	۱۸/۲	۰/۰	۰/۰

جدول ۵: مقایسه نتایج.

	میانگین دقت تشخیص ۶ دستگاه	میانگین دقت تشخیص ۱۱ زیردستگاه
روش پیشنهادی	۷۴/۵٪	۶۶/۳۵٪
LSTM غیر سلسله‌مراتبی	۴۹/۶٪	۴۳/۱٪
روش [۵۴]	۳۳/۱۴٪	۲۵/۹۶٪

استخراج نگردند. اما از آنجا که تعداد نت‌های اجراشده در یک قطعه موسیقی بسیار زیاد است، چنین خطاهایی نمی‌تواند تأثیر چندانی بر توانایی تشخیص سامانه پیشنهادی بگذارد.

به طور کلی، خطاهایی را که طی دسته‌بندی دستگاه‌های موسیقی به آنها برخورد کرده‌ایم، می‌توان به سه دسته خطاهای ناشی از پایگاه داده، استخراج ویژگی و دسته‌بند تقسیم‌بندی کرد که همراه با مثال‌هایی از پایگاه داده تشریح شده‌اند:

- تعداد زیاد گونه‌ها و زیرگونه‌ها با هم‌پوشانی زیاد خصوصیات موسیقیایی در پایگاه داده از عوامل ایجاد خطاست. برای مثال سامانه در دسته‌بندی قطعه shour(۶) اشتباه کرده که ناشی از شباهت دو دستگاه شور و نواست.
- تعدادی از قطعات از هم‌نوازی ساز کوبه‌ای همراه با ساز ملودی به وجود آمده‌اند تا اثر چندنوازی بودن قطعات نیز در سامانه بررسی شود. این عامل باعث ایجاد خطا در استخراج ویژگی زیر و بمی می‌شود، زیرا ماهیت تکنوازی‌بودن را از بین برده و اکتشاف نت را از ملودی نواخته‌شده با مشکل مواجه می‌کند. به عنوان نمونه، هم‌نوازی ساز کوبه‌ای همراه با ساز ملودی باعث خطا در تشخیص دستگاه قطعه nava(۳) شده است.
- از عمده‌ترین محدودیت‌ها، حساسیت سامانه به تعداد داده‌های آموزشی کم در پایگاه داده است که منجر به کاهش دقت عملکرد دسته‌بند سامانه و همچنین بیش‌برازش می‌گردد که با اعمال داده‌های اعتبارسنجی به شبکه از بروز بیش‌برازش جلوگیری شده است. همچنین نت شاهد یکسان زیردستگاه‌ها از عوامل تشخیص اشتباه سامانه است که تشابه آن در زیردستگاه‌های گوناگون، باعث ایجاد خطا در دسته‌بندی می‌شود. علت اصلی اشتباه سامانه در دسته‌بندی قطعه homayoun(۴) (دارای نت شاهد سل) شباهت نت شاهد آن با دستگاه شور می‌باشد که سامانه را دچار خطا کرده است.

همچنین برای مقایسه با تحقیقات دیگر، روش پیشنهادشده در [۵۴] را روی پایگاه داده "آرگ" اعمال کرده‌ایم. این روش به ترتیب از استخراج ویژگی‌های ضرایب طیفی مل، ایجاد بردار هویت، اعمال کاهش بعد و دسته‌بند ماشین بردار پشتیبان برای تشخیص ۷ دستگاه استفاده کرده است. جدول ۵ نشان می‌دهد که روش پیشنهادی در این مقاله با اختلاف زیادی دارای دقت بالاتری در تشخیص دستگاه موسیقی ایرانی است.

۵- نتیجه‌گیری

در این مقاله با هدف دسته‌بندی دستگاه موسیقی سنتی ایرانی، سامانه‌ای مبتنی بر ۹ دسته‌بند LSTM و با رویکردی سلسله‌مراتبی پیشنهاد شده است. نتایج آزمایش‌ها نشان داد که دقت سامانه پیشنهادی برای تشخیص ۶ دستگاه اصلی موسیقی ایرانی بین ۶۳/۶٪ تا ۹۱/۷٪ متغیر بوده و دارای میانگین ۷۴/۵٪ می‌باشد. این دقت برای تشخیص ۱۱ زیردستگاه بین ۵۰٪ تا ۸۳/۳٪ با میانگین ۶۶/۳۵٪ متغیر بوده است. همچنین این نتایج نشان داد که سامانه پیشنهادی در مقایسه با تحقیقات اندک دیگری که روی موسیقی سنتی ایران انجام شده است، دارای قابلیت خوبی در دسته‌بندی دستگاه‌ها و زیردستگاه‌ها است. در فرایند این تحقیق، پایگاه داده‌ای برای آموزش و آزمایش سامانه پیشنهادی در ۶ دستگاه کلی و ۱۱ زیردستگاه تولید و برچسب‌گذاری شده است. قطعات برچسب‌گذاری شده در پایگاه داده به صورت تکنوازی هستند، اگرچه تعداد اندکی از قطعات از هم‌نوایی‌سازی کوبه‌ای در کنار سازهای ملودی به وجود آمده‌اند. بهینه‌سازی الگوریتم و مهندسی بهتر ویژگی‌های موسیقی ایرانی باعث افزایش دقت و کارایی سامانه پیشنهادی خواهد شد.

مراجع

[1] R. Mayer, R. Neumayer, and A. Rauber, "Combination of audio and lyrics features for genre classification in digital audio collections," in *Proc. of the 16th ACM Int. Conf. on Multimedia*, pp. 159-168, Vancouver, Canada, 26-31 Oct. 2008.

[2] R. Rajan and H. A. Murthy, "Music genre classification by fusion of modified group delay and melodic features," in *23rd National Conf. on Communications, NCC'17*, 6 pp. Chennai, India, 2-4 Mar 2017.

[3] Y. Wang, "Research on music recognition algorithm based on RBF neural network," *Revista de la Facultad de Ingenieria*, vol. 32, no. 8, pp. 707-712, Jan. 2017.

[4] G. K. Birajdar and M. D. Patil, "Speech/music classification using visual and spectral chromagram features," *J. of Ambient Intelligence and Humanized Computing*, vol. 11, no. 1, pp. 329-347, Jan. 2020.

[27] T. Nakashika, C. Garcia, and T. Takiguchi, "Local-feature-map integration using convolutional neural networks for music genre classification," in *Proc. 13th Annual Conf. of the Int. Speech Communication Association, INTERSPEECH'12*, pp. 1752-1755, Portland, ON, USA, Sept. 2012.

[28] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 6, no. 3, pp. 610-621, Jan. 1973.

[29] G. Gwardys and D. M. Grzywczak, "Deep image features in music information retrieval," *International J. of Electronics and Telecommunications*, vol. 60, no. 4, pp. 321-326, Dec. 2014.

[30] S. Sigtia and S. Dixon, "Improved music feature learning with deep neural networks," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP'14*, pp. 6959-6963, Florence, Italy, 4-9 May 2014.

[31] Y. M. G. Costa, L. S. Oliveira, and C. Silla, "An evaluation of convolutional neural networks for music classification using spectrograms," *Applied Soft Computing*, vol. 52, no. C, pp. 28-38, Mar. 2017.

[32] L. Nanni, *et al.*, "Combining visual and acoustic features for music genre classification," *Expert Systems with Applications*, vol. 45, no. C, pp. 108-117, Mar. 2016.

[33] L. Nanni, Y. M. G. Costa, D. R. Lucio, C. N. Silla, and S. Brahnam, "Combining visual and acoustic features for bird species classification," in *Proc. IEEE 28th Int. Conf. on Tools with Artificial Intelligence, ICTAI'16*, pp. 396-401, San Jose, CA, USA, 6-8 Nov. 2016.

[34] F. Medhat, D. Chesmore, and J. Robinson, "Masked conditional neural networks for audio classification," in *Proc. Int.Conf. on Artificial Neural Networks.* pp. 349-358, Alghero, Italy, 11-14 Sept. 2017.

[35] F. Medhat, D. Chesmore, and J. Robinson, "Automatic classification of music genre using masked conditional neural networks," in *Proc. IEEE Int. Conf. on Data Mining, ICDM'17*, pp. 979-984, New Orleans, LA, USA, 18-21 Nov. 2017.

[36] L. R. Aguiar, M. G. Y. Costa, and C. Silla, "Exploring data augmentation to improve music genre classification with convnets," in *Proc. Int. Joint Conf. on Neural Networks, IJCNN'18*, 8 pp., Rio de Janeiro, Brazil, 8-13 Jul 2018.

[37] L. Feng, S. Liu, and J. Yao, Music Genre Classification with Paralleling Recurrent Convolutional Neural Network, arXiv preprint arXiv:1712.08370, Dec. 2017.

[38] S. Panwar, A. Das, M. Roopaei, and P. Rad, "A deep learning approach for mapping music genres," in *12th System of Systems Engineering Conf., SoSE'17*, 5 pp., Waikoloa, HI, USA, 18-21 Jun. 2017.

[39] J. Schlüter and S. Böck, "Musical onset detection with convolutional neural networks," in *Proc. 6th Int. Workshop on Machine Learning and Music, MML'13,* 4 pp. Prague, Czech Republic, 23-23 Sept. 2013.

[40] S. Oramas, *et al.*, "Multimodal deep learning for music genre classification," *Trans. of the International Society for Music Information Retrieval*, vol. 1, no. 1, pp. 4-21, Sept. 2018.

[41] J. Jakubik, "Evaluation of gated recurrent neural networks in music classification tasks," in *Proc. of 38th Int. Conf. on Information Systems Architecture and Technology, ISAT'17,* pp. 27-37, Szklarska Poręba, Poland, 17-19 Sept. 2018.

[42] N. Chen and S. Wang, "High-level music descriptor extraction algorithm based on combination of multi-channel CNNs and LSTM," in *Proc. 18th Int. Society for Music Information Retrieval Conf.*, *ISMIR'17*, pp. 509-514, Suzhou, China, 23-27 Oct. 2017.

[43] D. Ghosal and M. H. Kolekar, "Musical genre and style recognition using deep neural networks and transfer learning," in *Proc. APSIPA Annual Summit and Conf.,* pp. 978-988, Hawaii, HI, USA, 12-15 Nov. 2018.

[44] P. Fulzele, R. Singh, N. Kaushik, and K. Pandey, "A hybrid model for music genre classification using LSTM and SVM," in *Proc. 11th Int. Conf. on Contemporary Computing, IC3'18*, 3 pp., Noida, India, 2-4 Aug. 2018.

[45] R. J. M. Quinto, R. O. Atienza, and N. M. C. Tiglao, "Jazz music sub-genre classification using deep learning," in *Proc. IEEE Region 10 Conf.*, *TENCON'17*, pp. 3111-3116, Penang, Malaysia, 5-8 Nov. 2017.

[46] L. Soboh, I. Elkabani, and Z. Osman, "Arabic cultural style based music classification," in *Proc. Int. Conf. on New Trends in Computing Sciences ICTCS'17,* pp. 6-11, Amman, Jordan, 11-13 Oct. 2017.

[47] S. Kanchana, K. Meenakshi, and V. Ganapathy, "Comparison of genre based tamil songs classification using term frequency and

[5] J. H. Foleiss and T. F. Tavares, "Texture selection for automatic music genre classification," *Applied Soft Computing*, vol. 89, no. C, Article ID: 106127, Apr. 2020.

[6] W. H. Chang, J. L. Li, Y. S. Lin, and C. C. Lee, "A genre-affect relationship network with task-specific uncertainty weighting for recognizing induced emotion in music," in *Proc. IEEE Int. Conf. on Multimedia and Expo, ICME'18*, 6 pp., San Diego, CA, USA, 23-27 Jul. 2018.

[7] A. Elbir, H. O. İlhan, G. Serbes, and N. Aydın, "Short time fourier transform based music genre classification," in *Proc. Electric Electronics, Computer Science, Biomedical Engineerings' Meeting, EBBT'18*, 4 pp., Istanbul, Turkey, 18-19 Apr.. 2018.

[8] E. Simas Filho, E. Borges Jr., and A. Fernandes Jr., "Genre classification for brazilian music using independent and discriminant features," *Journal of Communication and Information Systems*, vol. 33, no. 1, pp. 104-112, May 2018.

[9] Y. M. G. Costa, L. S. Oliveira, A. L. Koerich, F. Gouyon, and J. G. Martins, "Music genre classification using LBP textural features," *Signal Processing*, vol. 92, no. 11, pp. 2723-2737, Nov. 2012.

[10] A. K. Singh, R. Singh, and A. Dwivedi, "Mel frequency cepstral coefficients based text independent Automatic Speaker Recognition using matlab," in *Proc. Int. Conf. on Reliability Optimization and Information Technology, ICROIT'14*, pp. 524-527, Faridabad, India, 6-8 Feb. 2014.

[11] C. Silla, C. A. A. Kaestner, and A. L. Koerich, "Automatic music genre classification using ensemble of classifiers," in *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics*, pp. 1687-1692, Montreal, Canada, 7-10 Oct. 2007.

[12] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, Jul. 2002.

[13] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798-1828, Aug. 2013.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, May 2015.

[15] W. Shi and X. Fan, "Speech classification based on cuckoo algorithm and support vector machines," in *Proc. 2nd IEEE Int. Conf. on Computational Intelligence and Applications, ICCIA'17*, pp. 98-102, Beijing, China, 8-11 Sept. 2017.

[16] S. Sharma, P. Fulzele, and I. Sreedevi, "Novel hybrid model for music genre classification based on support vector machine," *IEEE Symp. on Computer Applications & Industrial Electronics, ISCAIE'18*, pp. 395-400, , Penang, Malaysia, 28-29 Apr. 2018.

[17] D. Chaudhary, N. P. Singh, and S. Singh, "Genre based classification of hindi music," in *Proc. Int. Conf. on Innovations in Bio-Inspired Computing and Applications,* pp. 73-82, Kochi, India, 23-24 Nov. 2019.

[18] J. Li, J. Ding, and X. Yang, "The regional style classification of chinese folk songs based on GMM-CRF model," in *Proc. of the 9th Int. Conf. on Computer and Automation Engineering, ICCAE'17,* pp. 66-72, Sydney, Australia, 18-21 Feb. 2017.

[19] C. Kaur and R. Kumar, "Study and analysis of feature based automatic music genre classification using Gaussian mixture model," in *Proc. Int. Conf. on Inventive Computing and Informatics, ICICI'17*, pp. 465-468, , Coimbatore, India, 23-24 Nov. 2017.

[20] D. G. Bhalke, B. Rajesh, and D. S. Bormane, "Automatic genre classification using fractional fourier transform based mel frequency cepstral coefficient and timbral features," *Archives of Acoustics*, vol. 42, no. 2, pp. 213-222, Jan. 2017.

[21] A. Sridharan, *Music Similarity Estimation*, Master's Projects, 607, 2018, DOI: https://doi.org/10.31979/etd.8nz2-b9yavol

[22] A. Acharya, Detecting the Trend in Musical Taste Over the Decade: A Novel Feature Extraction Algorithm to Classify Musical Content with Simple features, arXiv preprint arXiv:1901.02053, 2018.

[23] Y. LeCun, *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, pp. 541-551, 1989.

[24] E. J. Humphrey and J. P. Bello, "Rethinking automatic chord recognition with convolutional neural networks," in *Proc. 11th Int. Conf. on Machine Learning and Applications*, vol. 2, pp. 357-362, Boca Raton, FL, USA, 12-15 Dec. 2012.

[25] E. J. Humphrey, J. P. Bello, and Y. LeCun, "Moving beyond feature design: deep architectures and automatic feature learning in music informatics," in *Proc. 13th Int. Society for Music Information Retrieval Conf. ,ISMIR'12*, pp. 403-408, Porto, Portugal, 8-12 Oct. 2012.

[26] J. Schlüter and S. Böck, "Improved musical onset detection with convolutional neural networks," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP'14,* pp. 6979-6983, Florence, Italy, 4-9 May 2014.

inverse document frequency," *Research J. Pharm. and Tech*, vol. 10, no. 5, pp. 1449-1454, Jul. 2017.
[48] A. Sridharan, M. Moh, and T. Moh, "Similarity estimation for classical indian music," in *Proc. 17th IEEE Int. Conf. on Machine Learning and Applications, ICMLA'18*, pp. 814-819, Orlando, FL, USA, 17-20 Dec. 2018.
[49] S. Chowdhuri, "PhonoNet: multi-stage deep neural networks for raga identification in hindustani classical music," in *Proc. of the 2019 on Int. Conf. on Multimedia Retrieval*, pp. 197-201, Ottawa, Canada, 10-13 Jun. 2019.
[50] M. Bhatt and T. Patalia, "Neural network based Indian folk dance song classification using MFCC and LPC," *Int. J. Intell. Eng. Syst.*, vol. 10, no. 3, pp. 173-183, Jun. 2017.
[51] F. Mahardhika, H. L. H. S. Warnars, Y. Heryadi, and Lukas, "Indonesian's dangdut music classification based on audio features," in *Proc. Indonesian Association for Pattern Recognition Int. Conf., INAPR'18*, pp. 99-103 Jakarta, Indonesia, 7-8 Sept.2018.
[۵۲] س. محمودان و ا. بنوشی، "دسته‌بندی خودکار گام ماهور موسیقی ایرانی توسط یک شبکه عصبی مصنوعی،" *دومین کنفرانس بین‌المللی آکوستیک و ارتعاشات دانشگاه صنعتی شریف*، صص. ۹-۱ ، تهران، دی ۱۳۹۱.
[53] H. Hajimolahoseini, R. Amirfattahi, and M. Zekri, "Real-time classification of Persian musical dastgahs using artificial neural network," in *Proc. 16th CSI Int. Symp. on Artificial Intelligence and Signal Processing, AISP'12,* pp. 157-160, Shiraz, Iran, 2-3 May 2012.
[۵۴] ب. باباعلی، آ. گرگان محمدی و ا. فرجی دیزجی، "نوا: دادگان موسیقی سنتی ایرانی برای تشخیص دستگاه و سازهای اصیل ایرانی،" *پردازش سیگنال پیشرفته*، جلد ۸، شماره ۲، صص. ۱۳۴-۱۲۵، پاییز و زمستان ۱۳۹۸.
[55] Md. Kamrul Hasan, S. Hussain, M. T. Hossain Setu, and Md. N. Ibne Nazrul, "Signal reshaping using dominant harmonic for pitch estimation of noisy speech," *Signal Process.* vol. 86, no. 5, pp. 1010-1018, May 2006.
[56] Q. Wang, X. Zhao, and J. Xu, "Pitch detection algorithm based on normalized correlation function and central bias function," in *Proc. 10th Int. Conf. on Communications and Networking in China, ChinaCom'15*, pp. 617-620, Shanghai, China, 15-17 Aug. 2015.
[57] B. S. Atal, "Automatic speaker recognition based on pitch contours," *the J. of the Acoustical Society of America*, vol. 52, no. 6B, pp. 1687-1697, Dec. 1972.
[58] S. Gonzalez and M. Brookes, "A pitch estimation filter robust to high levels of noise (PEFAC)," in *Proc. 19th European Signal Processing Conf.*, pp. 451-455, Barcelona, Spain, 29 Aug.-3 Sept. 2011.
[59] A. M. Noll, "Cepstrum pitch determination," *The J. of the Acoustical Society of America*, vol. 41, no. 2, pp. 293-309, Feb. 1967.
[60] T. Drugman and A. Alwan, *Joint Robust Voicing Detection and Pitch Estimation Based on Residual Harmonics*, arXiv preprint arXiv:2001.00459, Dec. 2019.
[61] http://colah.github.io/posts/2015-08-Understanding-LSTMs
[62] A. Graves, N. Jaitly, and A. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 273-278, Olomouc, Czech Republic, 8-12 Dec. 2013.
[63] E. Charniak, *Introduction to Deep Learning*, the MIT Press, 2019.
[۶۴] س. غضنفری‌پور، ح. نظام‌آبادی‌پور و ع. راشدی، "ترکیب ویژگی‌ها به کمک الگوریتم جستجوی گرانشی در بازیابی موسیقی ایرانی مبتنی بر محتوا در دستگاه ماهور،" *اولین کنفرانس محاسبات تکاملی و هوش جمعی*، صص. ۷۰-۶۵، *کرمان*، ۲۱-۱۹ اسفند ۱۳۹۴.

سینا غضنفری‌پور تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد مهندسی برق گرایش الکترونیک به‌ترتیب در سال‌های ۱۳۹۲ و ۱۳۹۴ از دانشگاه شهید باهنر کرمان به پایان رسانده است و هم‌اکنون در مقطع دکتری مهندسی برق گرایش مخابرات سیستم در دانشگاه فردوسی مشهد مشغول به تحصیل است. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: پردازش سیگنال موسیقی، پردازش صوت، بازشناسی الگو و پردازش زبان طبیعی.

مرتضی خادمی تحصیلات خود را در مقاطع کارشناسی و کارشناسی ارشد مهندسی برق به‌ترتیب در سال‌های ۱۳۶۴ و ۱۳۶۶ در دانشگاه صنعتی اصفهان به پایان رسانده است. نام‌برده از سال ۱۳۶۶ الی ۱۳۷۰ به عنوان عضو هیات علمی (مربی) در دانشگاه فردوسی مشهد به کار مشغول بود. پس از آن به دوره دکترای مهندسی برق در دانشگاه ولونگونگ (استرالیا) وارد گردیده و در سال ۱۳۷۴ موفق به اخذ درجه دکترا در مهندسی برق از دانشگاه مذکور گردید. دکتر خادمی از سال ۱۳۷۴ مجددا در دانشکده مهندسی دانشگاه فردوسی مشهد مشغول به فعالیت گردید و اینک نیز استاد این دانشکده است. زمینه‌های علمی مورد علاقه نام‌برده مخابرات ویدئویی، فشرده‌سازی ویدئو، پردازش تصویر، پردازش سیگنال‌های پزشکی و پنهان‌سازی اطلاعات در ویدئو می‌باشد.

عباس ابراهیمی مقدم مدرک کارشناسی و کارشناسی ارشد برق گرایش مخابرات خود را به ترتیب از دانشگاه‌های صنعتی شریف و صنعتی خواجه نصیر اخذ کرده است. ایشان مدرک دکتری خود را از دانشگاه مک‌مستر کانادا دریافت کرده و از سال ۱۳۹۰ به عنوان استادیار در دانشگاه فردوسی مشهد فعالیت علمی می‌نمایند. زمینه‌های تحقیقاتی مورد علاقه نام‌برده، پردازش گفتار، پردازش تصویر و ویدیو، بینایی ماشین و پردازش سیگنال‌های حیاتی می‌باشد.