



Diagnosing COPD Disease through Lung Sounds Analysis with Focusing on Temporal Features and using a Network based on Temporal Attention and Bidirectional Recurrent Gates

Saadati, Davoud¹ / Mirzakuchaki, Sattar^{2*}

¹ - M.Sc. Student, Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran

² - Professor, Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran

ARTICLE INFO

DOI: 10.22041/ijbme.2023.562783.1805

Received: 29 September 2022

Revised: 24/12/2022 – 7/3/2023 – 14/6/2023

Accepted: 6 August 2023

KEY WORDS

Diagnosis
Lung Disease
Sound Processing
Recurrence Plot
Temporal Attention
Recurrent Network
Transfer Learning
Machine Learning

ABSTRACT

Examining the sound of body organs is one of the methods of diagnosing various diseases, which is used by specialists to analyze abnormal sounds. Since most of the deaths caused by the disease occur in poor countries that have a shortage of equipment and specialists, the development of diagnostic methods based on machine learning and audio processing, which are available, non-invasive and inexpensive, can lead to early diagnosis and save the lives of millions of people. In previous studies, inputs that reflect the frequency characteristics of the sound have been used, in this article, we also use a recurrent representation that reflects the temporal characteristics of the sound and is given as an input to convolutional networks in order to benefit from its transfer learning advantages. By adding the temporal attention mechanism and the bidirectional recurrent gates, the audio data sequence which is a time series is investigated and each data is weighted according to its value. The data used in this article is from the ICBHI lung sound database, which has been used in many other articles. The presented method was able to classify lung sounds into three categories: healthy, chronic obstructive pulmonary disease (COPD) and other diseases with an accuracy of 97%, which is a better result than other methods that used this database.

*Corresponding Author

Address	Department of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran		
Postal Code	16846-13114	Tel	+98-21-73225737
E-Mail	m_kuchaki@iust.ac.ir	Fax	+98-21-73225777





تشخیص بیماری COPD از طریق تحلیل صدای ریه با تمرکز بر ویژگی‌های زمانی و به کارگیری شبکه‌ی مبتنی بر توجه زمانی و گیت‌های بازگشتی دوطرفه

سعادت‌ی، داود^۱ / میرزا کوچکی، ستار^{۲*}

^۱ - دانشجوی کارشناسی ارشد، گروه الکترونیک، دانشکده‌ی مهندسی برق، دانشگاه علم و صنعت ایران، تهران، ایران

^۲ - استاد، گروه الکترونیک، دانشکده‌ی مهندسی برق، دانشگاه علم و صنعت ایران، تهران، ایران

مشخصات مقاله

شناسه‌ی دیجیتال: 10.22041/ijbme.2023.562783.1805

پذیرش: ۱۵ مرداد ۱۴۰۲

بازنگری: ۱۴۰۱/۱۰/۳ - ۱۴۰۱/۱۲/۱۶ - ۱۴۰۲/۳/۲۴

ثبت در سامانه: ۷ مهر ۱۴۰۱

چکیده

واژه‌های کلیدی

بررسی صدای اندام‌های بدن یکی از روش‌های تشخیص بیماری‌های مختلف است که توسط پزشکان برای تحلیل صداهای عارضه‌ای به کار گرفته می‌شود. از آن‌جا که بیش‌تر مرگ و میرهای ناشی از بیماری، در کشورهای فقیری رخ می‌دهد که کمبود تجهیزات و متخصص دارند، ایجاد روش‌های تشخیصی مبتنی بر یادگیری ماشین و پردازش صوت که علاوه بر دردسترس بودن، غیرتهاجمی و ارزان نیز باشند با تشخیص زودهنگام می‌تواند باعث نجات جان میلیون‌ها انسان شود. در مطالعات پیشین غالباً ورودی‌های بازتاب‌کننده‌ی ویژگی‌های فرکانسی صوت مورد استفاده قرار گرفته است که در این مقاله علاوه بر آن از یک نمایش بازگشتی نیز استفاده شده که ویژگی‌های زمانی صوت را بازتاب کرده و به عنوان ورودی به شبکه‌های کانولوشنی داده می‌شود تا از مزیت‌های یادگیری انتقالی آن بهره گرفته شود. هم‌چنین با اضافه کردن مکانیسم توجه زمانی و شبکه‌ی گیت‌های بازگشتی دوطرفه، توالی داده‌های صوتی که یک سری زمانی بوده مورد بررسی قرار گرفته و هر یک از داده‌ها با توجه به ارزش آن‌ها وزن‌دهی شده است. داده‌های مورد استفاده در این مقاله از پایگاه داده‌ی صدای ریه ICBHI بوده که از آن در مقاله‌های بسیاری استفاده شده است. روش ارائه شده در این مقاله توانسته در طبقه‌بندی صدای ریه به سه دسته‌ی سالم، بیماری انسداد مزمن ریوی (COPD) و سایر بیماری‌ها به دقت ۹۷٪ برسد که نسبت به سایر روش‌هایی که از این پایگاه داده استفاده کرده‌اند نتیجه‌ی بهتری است.

تشخیص بیماری
بیماری ریوی
پردازش صوت
نمایش بازگشتی
توجه زمانی
شبکه‌ی بازگشتی
یادگیری انتقالی
یادگیری ماشین

*نویسنده‌ی مسئول

نشانی گروه الکترونیک، دانشکده‌ی مهندسی برق، دانشگاه علم و صنعت ایران، تهران، ایران

تلفن +۹۸-۲۱-۷۳۲۲۵۷۳۷

کد پستی ۱۶۸۴۶-۱۳۱۱۴

دورنگار +۹۸-۲۱-۷۳۲۲۵۷۷۷

پست الکترونیک m_kuchaki@iust.ac.ir



۱- مقدمه

طبق آمار سازمان بهداشت جهانی^۱ بیماری انسداد مزمن ریوی^۲ (COPD) سومین عامل مرگ در جهان است و حدود ۶٪ از کل مرگ و میر در جهان را در بر گرفته که بیش از ۹۰٪ آن‌ها در کشورهای با اقتصاد ضعیف و متوسط رخ می‌دهد. این بیماری با نرخ مرگ بیش از ۳ میلیون نفر در سال، مرگ‌آورترین بیماری در بین بیماری‌های ریوی است. ذات‌الریه^۳ سالانه میلیون‌ها نفر را می‌کشد و پیش‌بینی می‌شود تا سال ۲۰۳۰ نزدیک به ۹ میلیون کودک زیر ۵ سال با این بیماری کشته شوند و این تنها بخشی از آمار مرگ ناشی از بیماری‌های ریوی است. همچنین پاندمیک ویروس کووید ۱۹ که از اواخر سال ۲۰۱۹ تا زمان نگارش این مقاله موجب مرگ بیش از ۶/۴ میلیون نفر در جهان و ۱۴۲ هزار نفر در ایران شده است، نیاز به یک ابزار تشخیصی ارزان و در دسترس برای تشخیص اولیه را نشان می‌دهد تا با شناسایی افراد ناقل از انتشار بیش‌تر آن جلوگیری شود.

روش‌های متداولی که معمولاً برای تشخیص بیماری‌های ریوی توسط پزشکان استفاده می‌شود شامل معاینه‌ی بدنی، اسپرومتری، رادیوگرافی، بافت‌برداری، سی‌تی اسکن و غیره است. اکثر این روش‌ها تهاجمی^۴ و پرهزینه بوده و نیازمند تجهیزات و متخصص می‌باشند. از آن‌جا که این روش‌ها به تجهیزات و متخصص نیاز دارند می‌توان نرخ بالای مرگ و میر را ناشی از کمبود متخصص و امکانات پزشکی در مقایسه با تعداد بیماران به ویژه در مناطق محروم و فقیر دانست. بنابراین یافتن یک روش در دسترس، کم‌هزینه و غیرتهاجمی^۵ برای تشخیص عارضه‌های ریوی ضروری است. یکی از رویکردهای غیرتهاجمی رایج، گوش دادن است. تشخیص بیماری با آنالیز صدای ریه به کمک گوشی پزشکی^۶ روشی رایج و قابل اتکا بین متخصصان ریه در تشخیص اولیه بوده که به تجربه، مهارت و قدرت شنوایی متکی است. در بسیاری از مطالعات انجام شده مبتنی بر سی‌تی اسکن، رادیولوژی و اشعه‌ی ایکس تمایز تغییرات پاتومورفولوژیکی ناشی از عفونت COPD در مقایسه با سایر بیماری‌ها و نیز هر بیماری نسبت به دیگری اثبات شده است [۱]. بنابراین صدای ناشی از هر عفونت به دلیل تمایز بودن دارای ویژگی‌های نهفته‌ی منحصر به فردی بوده که در جدول (۱) و شکل‌های (۱) تا (۳) قابل مشاهده است. پزشکان و متخصصان نیز از گذشته با استفاده از گوشی‌های پزشکی صدای ریه را تحلیل کرده و بیماری آن را تشخیص می‌دادند.

با پیشرفت فناوری و تکنیک‌های پردازش سیگنال دیجیتال و یادگیری ماشین، بسیاری از روش‌های محاسباتی برای تشخیص خودکار بیماری‌های ریوی با دقت بالا از طریق تجزیه و تحلیل صدای ریه ایجاد شده که در ادامه به آن‌ها اشاره شده است.

در مقاله‌ی [۲] مدلی متشکل از CNN استاندارد VGGish و شبکه‌ی بازگشتی دوطرفه معرفی شده که سیگنال صدا را با استفاده از لگاریتم مل به آرایه‌ی دوبعدی (طیف نگاره) تبدیل کرده و به دقت طبقه‌بندی ۸۷/۴۱ درصد دست یافته است. در مقاله‌ی [۳] از تبدیل فوری‌ی زمان کوتاه (STFT) برای تبدیل سیگنال‌های صوتی ریه به تصاویر و به دنبال آن از شبکه‌ی عصبی پیچشی (CNN) برای طبقه‌بندی استفاده شده است. در این مطالعه روی دو مدل طبقه‌بندی کننده‌ی CNN_SVM و CNN استاندارد بررسی‌هایی انجام شده که برای طبقه‌بند CNN_SVM دقت ۶۵/۵٪ و برای طبقه‌بند CNN استاندارد دقت ۶۳/۰۹٪ به دست آمده است. در مقاله‌ی [۴] از ویژگی‌های آماری تبدیل هیلبرت-هنگ برای طبقه‌بند یادگیری عمیق مبتنی بر CNN به منظور طبقه‌بندی سیگنال‌های صوتی ریه استفاده شده است. دقت طبقه‌بندی با روش ارائه شده در این مطالعه ۹۳/۶۷٪ است. در مقاله‌ی [۵] از MFCC برای تبدیل سیگنال‌های صدای تک‌بعدی ریه به تصاویر دوبعدی استفاده شده است. در این مطالعه از CNN با بهینه‌ساز ADAM برای طبقه‌بندی استفاده شده و دقت طبقه‌بندی به دست آمده با استفاده از این روش ۷۴٪ گزارش شده است. در مقاله‌ی [۶] یک رویکرد مبتنی بر شبکه‌ی عصبی بازگشتی (RNN) برای طبقه‌بندی سیگنال‌های صوتی ریه در دسته‌های مختلف ارائه شده است. در این مطالعه از تکنیک‌های پنجره‌سازی برای تقسیم‌بندی سیگنال‌های صوتی ریه با قاب‌های مختلف استفاده شده، ۱۳ سطح از ویژگی MFCCs استخراج شده و حداکثر دقت طبقه‌بندی برابر با ۸۱٪ به دست آمده است. در مقاله‌ی [۷] از لگاریتم مل برای تبدیل سیگنال صدای تک‌بعدی ریه به تصاویر دوبعدی و از یک مدل مبتنی بر CNN برای طبقه‌بندی استفاده شده و دقت طبقه‌بندی برابر با ۸۱/۶۲٪ گزارش شده است. در مقاله‌ی [۸] روش EMD برای استخراج IMF-ها و به دنبال آن انتخاب IMF و استخراج ویژگی‌های آماری پیشنهاد شده است. در این مطالعه پارامترهای آماری (میانگین، میانه، انحراف استاندارد، چولگی و کشیدگی) ترکیب‌های فضای ویژگی را تشکیل داده و با طبقه‌بندی کننده‌ی درخت، دقت

^۱ Invasive^۵ Non Invasive^۶ Stethoscope^۱ World Health Organization (WHO)^۲ Chronic Obstructive Pulmonary Disease^۳ Pneumonia

ذاتی و تبدیل موجک پیوسته روی صوت یک نمایش جدید در حوزه‌ی زمان-فرکانس ارائه شده است. در مقاله‌ی [۱۱] از ویژگی‌های MFCC همراه با شبکه‌ی عصبی مصنوعی^۶ (ANN) برای طبقه‌بندی صداهای عادی، خس خس و کراکل استفاده شده که در آن دقت طبقه‌بندی برای کراکل ۷۵٪، برای صدای خس خس ۱۰۰٪ و برای صدای نرمال ۸۰٪ به دست آمده است. علاوه بر ویژگی‌های فرکانسی صوت باید توجه داشت که صوت یک سری زمانی است و ترتیب و توالی داده می‌تواند معنادار باشد. برای مثال در بیماری‌های ریوی صداهای عارضه‌ای می‌تواند در بخش‌های به خصوصی از یک دوره‌ی تنفسی خود را نشان دهد و یا زمان هر دوره‌ی تنفس کم یا زیاد شود. در جدول (۱) تعدادی از صداهای عارضه‌ای و دلیل رخ دادن و ارتباط آن‌ها با بیماری‌ها ارائه شده است [۱۴]. همچنین در شکل‌های (۱)، (۲) و (۳) می‌توان تفاوت‌های صدای ریه در هر کلاس را مشاهده کرد. در تحقیق حاضر علاوه بر ویژگی‌های فرکانسی صوت، ویژگی‌های زمانی صوت نیز با به کارگیری ورودی نمایش بازگشتی و شبکه‌ی بازگشتی و توجه زمانی مورد ارزیابی قرار گرفته است.

۹۶/۲۹٪ به دست آمده است. در مقاله‌ی [۹] از نسبت فرکانسی توان چگالی طیفی^۱ (PSD) و ویژگی‌های تبدیل هیلبرت-هوانگ^۲ (HHT) برای تشخیص صداهای معمولی ریه و کراکل استفاده شده و با طبقه‌بند ماشین بردار پشتیبان^۳ (SVM) دقت بالای ۹۰٪ به دست آمده است. در مقاله‌ی [۱۰] رویکرد دیگری برای طبقه‌بندی سیگنال‌های صوتی عارضه‌ای معمولی و پیوسته پیشنهاد شده که فقط سیگنال‌های خس خس در دسته‌ی سیگنال‌های عارضه‌ای پیوسته قرار گرفته است. در این مطالعه از ضرایب فرکانس مل کیسترال^۴ (MFCCs) برای استخراج ویژگی و از مدل مخلوط گاوسی^۵ (GMM) برای طبقه‌بندی سیگنال‌ها استفاده شده و دقت ۹۴/۲٪ به دست آمده است. در اکثر مطالعات انجام شده به ویژگی‌های فرکانسی صوت توجه بیشتری شده و اغلب با بهینه‌سازی تبدیل‌های حوزه‌ی زمان-فرکانس، ورودی‌هایی تولید شده که ویژگی‌های فرکانسی صوت را بهتر بیان کرده است. در مقاله‌ی [۱۲] از تبدیل S بهینه شده برای تبدیل صوت استفاده شده که با افزایش فرکانس تابع پنجره باریک‌تر و در فرکانس‌های پایین گسترده‌تر شده است. در مقاله‌ی [۱۳] با اعمال تجزیه‌ی مد

جدول (۱) - صداهای عارضه‌ای ریه، ویژگی‌های فرکانسی و زمانی آن‌ها، دلیل رخداد و بیماری مرتبط با آن‌ها

نوع	زمان‌بندی	فرکانس صدا	علت	بیماری مرتبط
Wheeze	هر دو	بیش‌تر از ۴۰۰ هرتز	باریک شدن راه جریان هوایی	آسم، COPD
Rhonchi	دوفازی، بیش‌تر با زدم	کم‌تر از ۲۰۰ هرتز	ترشح در برونش، ضخیم شدن مخاط	برونشیت، COPD
Fine Crackle	دوفازی، بیش‌تر دم	بیش‌تر از ۶۵۰ هرتز	باز شدن انفجاری مجاری تنفسی کوچک	ذات‌الرئیه، فیبروز ریه
Coarse Crackle	دم	کم‌تر از ۳۵۰ هرتز	حباب هوا در نایژه‌های بزرگ	برونشکتازی، COPD

فایل به دو بخش ۱۰ ثانیه‌ای تقسیم شده که هر بخش حداقل دارای یک دوره‌ی تنفسی است.

جدول (۲) - بیماری و تعداد نمونه‌های پایگاه داده‌ی ICBHI

بیماری ریوی	تعداد بیماران
سالم	۲۶
انسداد مزمن ریوی (COPD)	۶۴
آسم	۱
برونشکتازی	۷
ذات‌الرئیه	۶
برونشولیت	۶
عفونت دستگاه تنفسی تحتانی	۲
عفونت دستگاه تنفسی فوقانی	۱۴
تعداد کل موارد	۱۲۶

۲- پایگاه داده

برای آزمایش روش پیشنهادی از داده‌های پایگاه داده‌ی ICBHI [۱۵] استفاده شده است. این پایگاه داده شامل ۵/۵ ساعت صدای ضبط شده از ریه‌ی ۱۲۶ نفر است. در جدول (۲) تعداد و انواع بیماری‌های موجود در این پایگاه داده ارائه شده است. با توجه به این که COPD بیش‌ترین سرانه‌ی مرگ و میر را نسبت به سایر بیماری‌های ریوی دارد در این مقاله از داده‌های مربوط به بیماری COPD استفاده شده و این داده‌ها از افراد سالم و سایر بیماری‌ها متمایز شده است. بنابراین در این تحقیق صداهای ریه در سه دسته‌ی سالم، COPD و سایر بیماری‌ها طبقه‌بندی شده است. مدت زمان هر فایل صوتی این پایگاه داده ۲۰ ثانیه بوده که در این مقاله برای بیش‌تر شدن تعداد نمونه‌ها با توجه به طول هر دوره‌ی تنفسی (۳ تا ۶ ثانیه)، هر

^۱ Mel-Frequency Cepstral Coefficients

^۵ Gaussian Mixture Model

^۶ Artificial Neural Network

^۱ Power Spectral Density

^۲ Hilbert-Huang Transform

^۳ Support Vector Machine

۳- روش پیشنهادی

تمام شبیه‌سازی‌های این پژوهش با زبان برنامه‌نویسی پایتون در محیط کولب نسخه‌ی Pro (غیررایگان) با Gpu و حافظه‌ی RAM ۲۵ گیگابایت صورت گرفته است. در تولید ورودی نمایش بازگشتی از فایل‌های ۱۰ ثانیه‌ای، حافظه‌ی RAM بسیار زیادی مورد نیاز بوده که این مشکل با موازی‌سازی شبکه‌ی پیشنهادی برطرف شده و ورودی‌های بازگشتی با به کارگیری RAM کم‌تر از ۲۰ گیگابایت تولید شده است.

۳-۱- پیش‌پردازش

سیگنال‌های صدای اندام تنفسی ریه با اطلاعات غیرضروری و مخربی مانند نویز، صداهای محیطی، صدای قلب و غیره همراه است. برای حذف نویز و صداهای غیرضروری می‌توان از فیلترهای خطی حوزه‌ی فرکانس استفاده کرد اما به دلیل ماهیت غیرثابت بودن^۱ این سیگنال‌ها بهتر است برای رسیدن به عمل‌کرد بهینه از روش‌های تجزیه‌ی سیگنال استفاده شود.

۳-۱-۱- حذف نویز و صداهای هم‌فرکانس با صدای ریه

با توجه به مطالعات انجام شده [۱۴] صدای ریه در محدوده‌ی فرکانسی ۵۰ تا ۲۵۰۰ هرتز قرار دارد بنابراین برای روش پیشنهادی در مرحله‌ی پیش‌پردازش با رابطه‌ی (۱) دامنه‌ی صدا بین ۱- تا ۱ نرمالیزه شده و سپس برای حذف نویز، سیگنال صدا با فیلتر میان‌گذر باترورث مرتبه‌ی ۶ از محدوده‌ی ۵۰ تا ۲۵۰۰ هرتز عبور داده شده است. در شکل‌های (۱) تا (۳) سیگنال صدای اصلی (الف) و فیلتر شده (ب) برای صدای بیماری در کلاس‌های مختلف نشان داده شده است.

$$x_{norm}(t) = \frac{x(t)}{\max(|x|)} \quad (1)$$

از آن‌جا که صداهای محیطی، نویز و صدای سایر اندام‌ها با صدای ریه هم‌پوشانی داشته و برای مثال اجزای اصلی فرکانس صدای قلب (s1 و s2) در محدوده‌ی ۱۰ تا ۲۵۰ هرتز بوده [۱۶] که با اجزای فرکانس پایین صدای ریه هم‌پوشانی دارد، پس از اعمال فیلتر باترورث برای حذف نویز، جهت تفکیک و حذف یا کم اثر کردن این صداهای اضافی از صدای ریه، سیگنال‌های صوتی با استفاده از روش تجزیه‌ی مد ذاتی^۲ (EMD) به ۵ تابع مد ذاتی^۳ (IMFs) تجزیه شده است [۱۷]. سپس IMF-های دارای کم‌ترین همبستگی با سیگنال اصلی حذف شده و بدین

ترتیب در سیگنال باقی‌مانده هم‌پوشانی فرکانسی صدای ریه با صداهای مزاحم بسیار کاهش یافته است. در شکل‌های (۱) تا (۳) در بخش (ج) صداهای دارای هم‌پوشانی با صدای ریه که با کمک روش تجزیه‌ی مد ذاتی حذف شده یا کاهش یافته در نمایش طیف‌نگاره‌ها به وضوح قابل مشاهده است.

۳-۱-۲- ورودی‌های شبکه‌ی پیشنهادی

در شبکه‌ی مورد استفاده برای استخراج ویژگی و طبقه‌بندی از ۳ کانال به عنوان ورودی استفاده شده است. بنابراین از تبدیل لگاریتم مل (Log_mel)، ضرایب کپسترال فرکانسی مل^۴ (MFCCs) و نمایش بازگشتی^۵ (RP) به عنوان ورودی شبکه‌ی پیشنهادی استفاده شده که Log_mel و MFCCs ویژگی‌های فرکانسی صدا و RP ویژگی‌های زمانی صدا را بازتاب داده که در مطالعات کم‌تر مورد توجه قرار گرفته است.

۳-۱-۲-۱- لگاریتم مل

برای تبدیل Log_mel، صدا به مقیاس مل برده شده و از آن لگاریتم گرفته شده است. مقیاس مل واحدی از گام بوده به طوری که فواصل مساوی در صداهای فرکانس بالا و فرکانس پایین شبیه همان فاصله‌ای است که گوش انسان می‌شنود. مقیاس مل با اعمال یک سری روابط ریاضی روی فرکانس‌های صوتی به دست می‌آید. این مقیاس رفتار سیستم شنیداری انسان را تقلید کرده و طیف‌نگاره‌ی حاصل از آن ویژگی‌های فرکانسی را در راستای محور زمان بازتاب می‌کند.

در شکل‌های (۱) تا (۳) برای سیگنال صوتی اصلی (الف)، سیگنال صوتی فیلتر شده با فیلتر باترورث (ب) و سیگنال بهینه شده با EMD (ج) طیف‌نگاره‌ی لگاریتم مل نشان داده شده که تاثیر فیلتر باترورث و EMD بر فرکانس‌های صدا قابل مشاهده است که باترورث نویزهای فرکانس بالا و EMD سیگنال‌هایی با فرکانس‌های مزاحم موجود در محدوده‌ی فرکانسی صدای ریه را کم یا حذف می‌کند [۷].

۳-۱-۲-۲- ضرایب کپسترال فرکانسی مل

برای تبدیل MFCCs ابتدا از سیگنال پنجره‌گذاری شده، تبدیل فوریه‌ی گسسته گرفته شده، سپس لگاریتم دامنه‌ی طیفی به دست آمده محاسبه شده و بانک فیلتر مل که مجموعه‌ای از فیلترهای مثلثی بوده روی آن اعمال شده و به دنبال آن تبدیل کسینوس گسسته‌ی ضرایب کپسترال فرکانس مل به دست

^۴ Mel Frequency Cepstral Coefficients

^۵ Recurrence Plot Pre-Trained Network

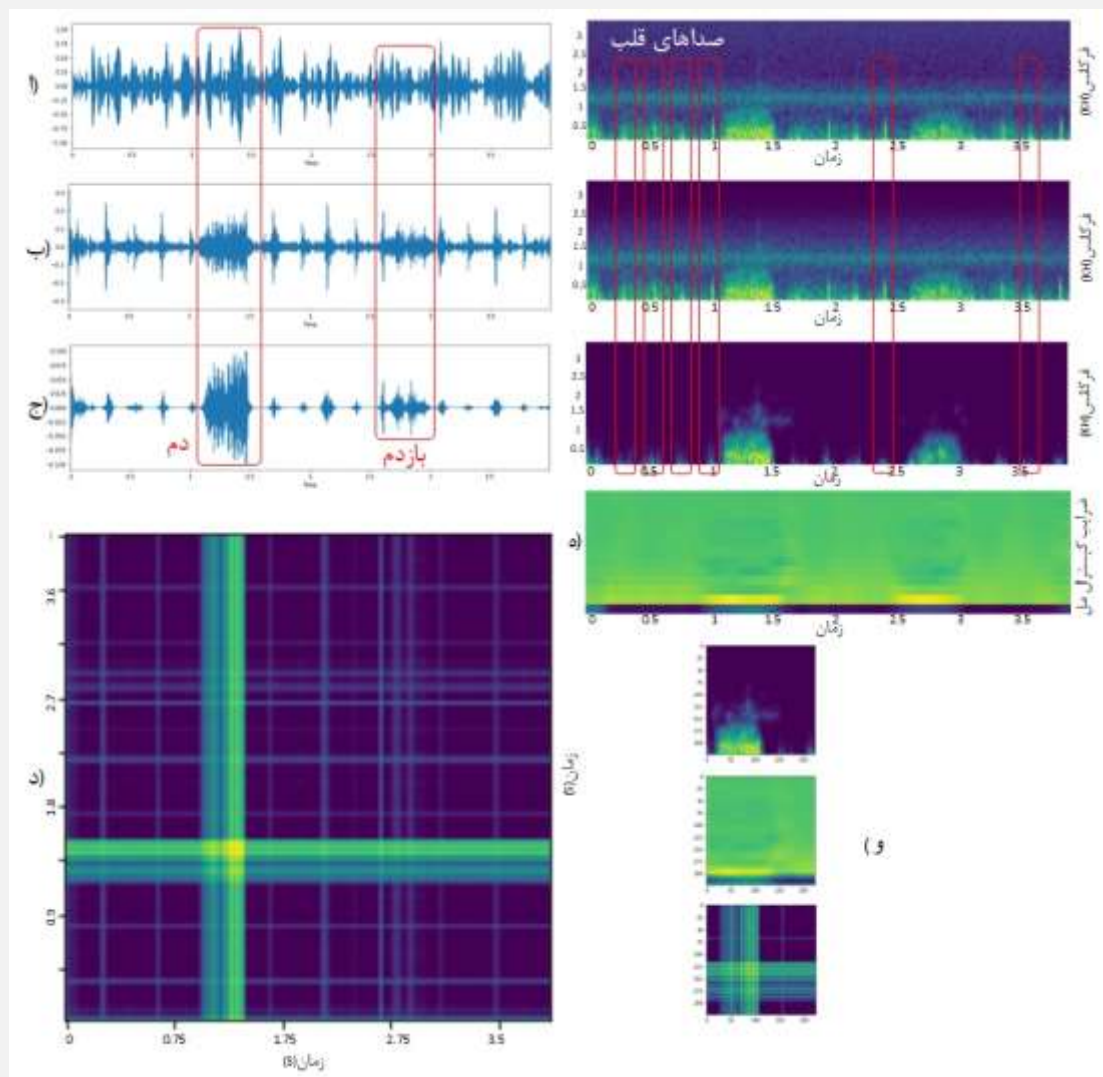
^۱ Non Stationary

^۲ Empirical Mode Decomposition

^۳ Intrinsic Mode Functions

در شکل‌های (۱) تا (۳) در بخش (ه) برای صدای بیماری در کلاس‌های مختلف طیف‌نگاره‌ی حاصل از ضرایب استخراج شده‌ی این تبدیل در راستای محور زمان نشان داده شده است.

آمده است. در بانک فیلتر از ۳۳ فیلتر مثلثی برای استخراج باندهای فرکانسی تا محدوده‌ی ۲۵۰۰ هرتز استفاده شده و ۱۳ مورد از ضرایب کپسترال فرکانس مل استخراج شده است [۶].



شکل (۱) - صدای ریه مربوط به کلاس سالم که در این ۴ ثانیه یک دوره‌ی تنفسی عادی را در خود جای داده است، در این فایل صدای قلب نیز ضبط شده که در بخش پیش‌پردازش اثر آن و سایر صداهای دارای هم‌پوشانی با صدای ریه کاهش داده شده است، الف) صدای ریه ضبط شده و طیف‌نگاره‌ی لگاریتم مل آن، ب) صدای ریه بعد از اعمال فیلتر میان‌گذر باترورت مرتبه‌ی ۶ در محدوده‌ی ۵۰ تا ۲۵۰۰ هرتز و طیف‌نگاره‌ی لگاریتم مل آن، ج) صدای ریه پس از حذف صداهای دارای هم‌پوشانی با صدای ریه و طیف‌نگاره‌ی لگاریتم مل آن که به خوبی حذف نویزها و کاهش اثر سیگنال‌های صدای قلب را نشان داده است، د) نمایش بازگشتی که هر دو محور آن زمان است، ه) طیف‌نگاره‌ی MFCCs با استخراج ۱۳ ضریب، و) برای ورودی شبکه‌ی پیشنهادی هر فایل ۱۰ ثانیه‌ای به ده بخش ۱ ثانیه‌ای تقسیم شده و تبدیل‌های معرفی شده بر آن‌ها اعمال شده که در این بخش ورودی مربوط به ثانیه‌ی ۲ صدای ریه در کلاس سالم نشان داده شده است

بین ۱ تا n دارد. محتوای این بردارها می‌تواند پوزیشن، سرعت، دما، فشار هوا، رطوبت و غیره باشد. سیستم به وسیله‌ی زنجیره‌ای از این بردارها که مسیر حرکت سیستم را در یک فضای ریاضیاتی نمایش می‌دهد توصیف می‌شود. در واقع با RP می‌توان تکراری بودن تحول حالت را در فضای فازی یک

۳-۲-۱-۳- نمایش بازگشتی

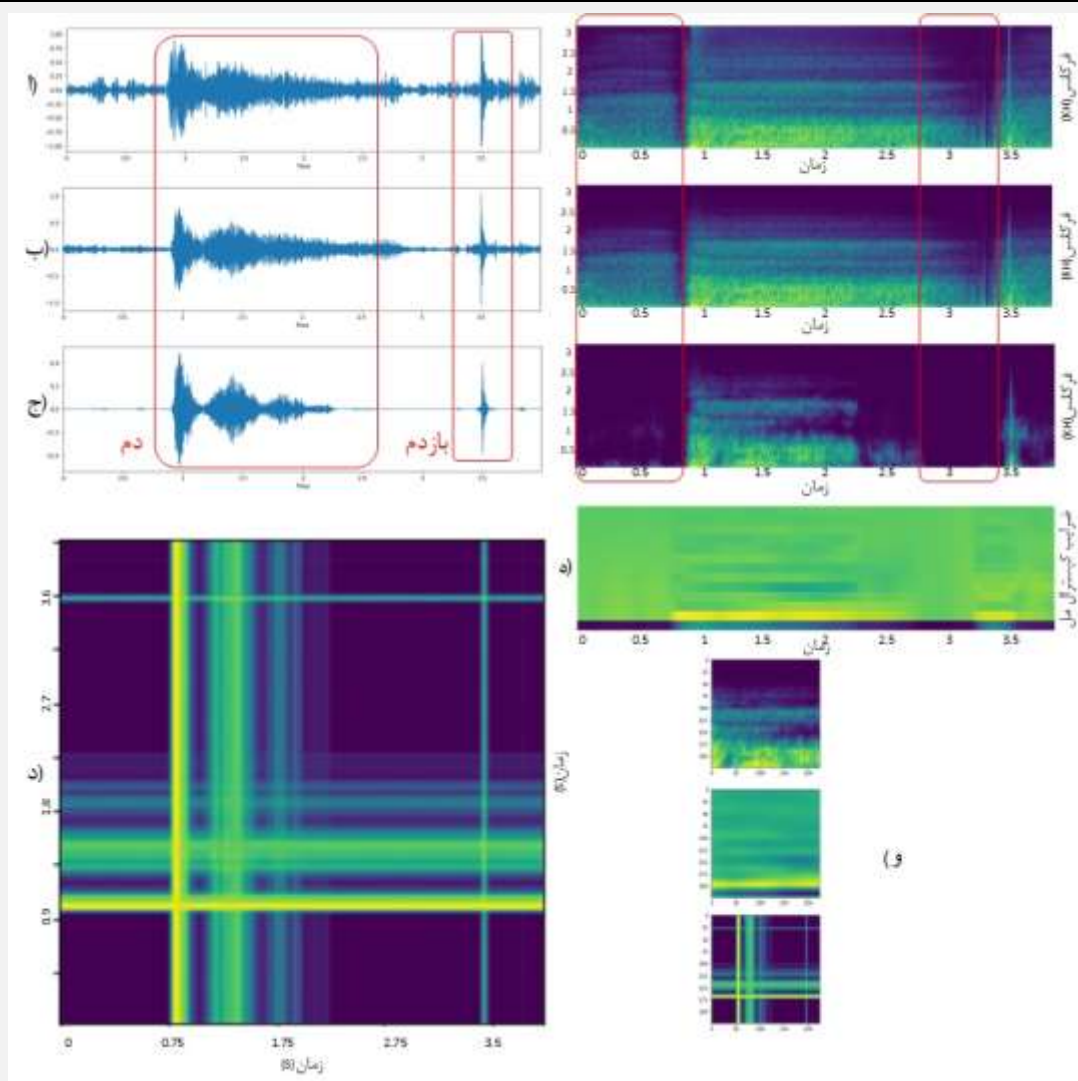
اکمن در سال ۱۹۸۷ [۱۸] یک متد برای نمایش بازگشتی^۱ بودن سیستم‌های دینامیک معرفی کرده است. فرض کنید مسیر حرکت یک سیستم دینامیک در فضای فازی به صورت برداری از x_i ها بوده که i بیان‌گر تعداد حالت‌ها است و مقداری

^۱ Recurrence Plot

و فاصله‌ها و نحوه‌ی قرارگیری خطوط می‌تواند به خوبی عمل کرد سیستم را نشان دهد و پیش‌بینی و تشخیص بیماری و درک رفتار کلی سیستم را بهبود بخشد [۱۹]. ضریب RP یک ماتریس مانند رابطه‌ی (۲) است که حالت سیستم را در i و j با هم مقایسه می‌کند. اگر این دو حالت تقریباً برابر باشند مقدار RP برابر با ۱ و اگر نابرابر بوده و یا از مقدار مشخصی فاصله یا خطای بیش‌تری داشتند مقدار آن برابر با صفر است.

$$R_{i,j} = \begin{cases} 1 & : X_i \approx X_j \\ 0 & : X_i \neq X_j \end{cases} \quad i, j = 1, 2, \dots, N \quad (2)$$

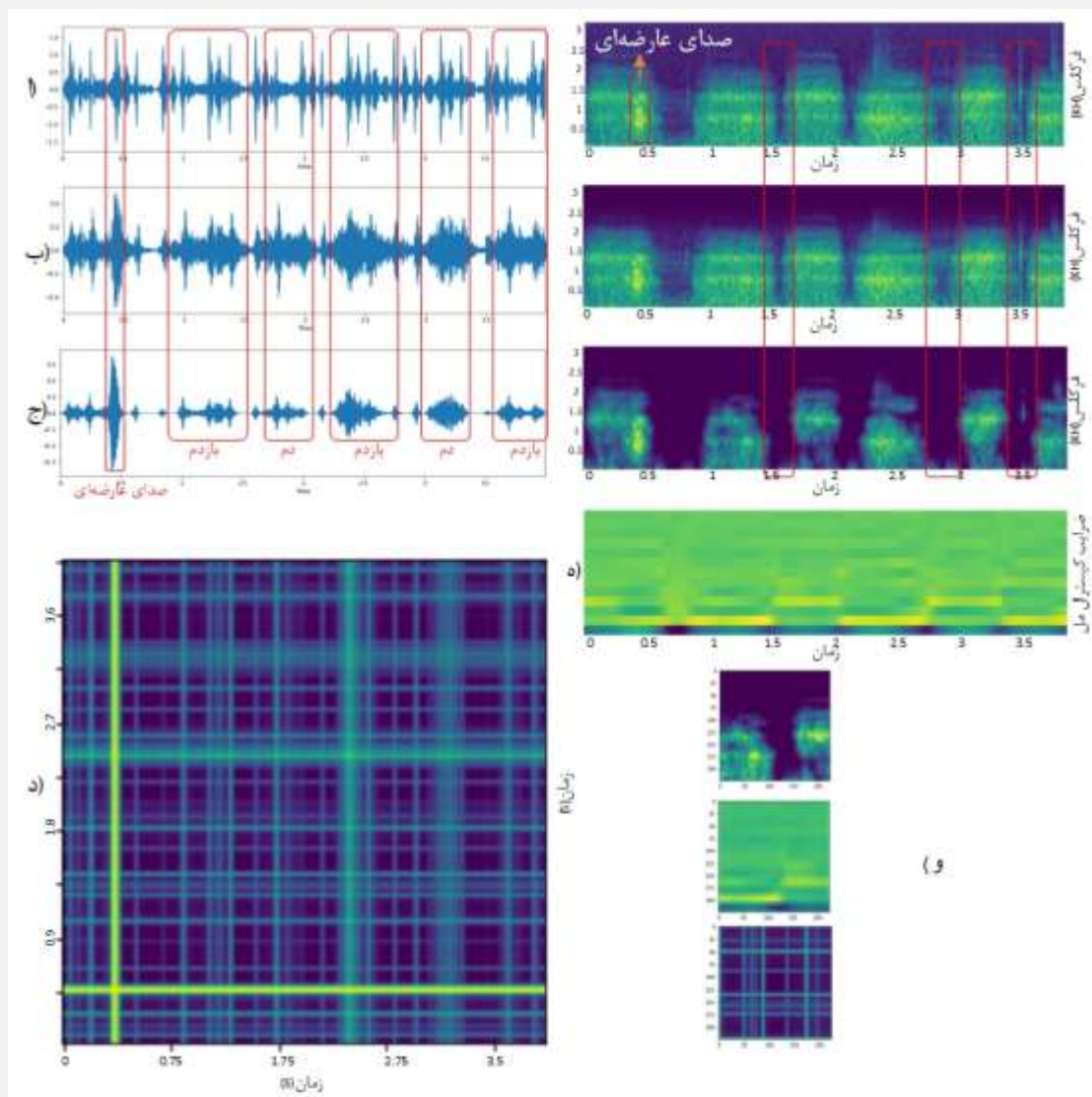
سیستم پویا نشان داد، یعنی اگر فاصله‌ی بین دو حالت در فضای فاز بسیار نزدیک باشد، در نظر گرفته می‌شود که بین دو حالت در فضای توپولوژی در طول زمان بازگشت رخ می‌دهد. با بررسی این نمودار می‌توان وقوع تکراری توالی‌های مشابهی از حالت‌ها را در دینامیک سیستم پیدا کرد. یک خط مورب در RP نشان می‌دهد که تکامل حالات در فضای فاز در زمان‌های مختلف مشابه است. یک خط عمودی (یا افقی) در RP نشان می‌دهد که یک حالت در طول یک زمان مشخص تغییر نکرده یا خیلی آهسته تغییر می‌کند. با بررسی خطوط افقی، عمودی و مورب می‌توان ویژگی‌های بسیار مفید دیگری را استخراج کرد



شکل (۲) - صدای ریه مربوط به کلاس COPD که در این ۴ ثانیه یک دوره‌ی تنفسی غیرعادی را در خود جای داده است، عمل دم در این بیمار به سختی و در زمان بیش‌تری انجام شده که با کمی صدای عارضه‌ای Wheeze همراه است، الف) صدای ریه‌ی ضبط شده و طیف‌نگاره‌ی لگاریتم مل آن، ب) صدای ریه بعد از اعمال فیلتر میان‌گذر باترورت مرتبه‌ی ۶ در محدوده‌ی ۵۰ تا ۲۵۰۰ هرتز و طیف‌نگاره‌ی لگاریتم مل آن، ج) صدای ریه پس از حذف صداهای دارای هم‌پوشانی با صدای ریه و طیف‌نگاره‌ی لگاریتم مل آن که به خوبی کاهش اثر سیگنال‌های مزاحم را نشان داده است، د) نمایش بازگشتی که هر دو محور آن زمان است، ه) طیف‌نگاره‌ی MFCCs با استخراج ۱۳ ضریب، و) در این بخش برای نمونه، ورودی مربوط به ثانیه‌ی ۲ صدای ریه در کلاس COPD نشان داده شده که هر ثانیه به تصویر با ابعاد (۲۲۴،۲۲۴) تبدیل شده و در مجموع ۱۰ تصویر (۲۲۴،۲۲۴) سه کاناله به عنوان ورودی به شبکه داده شده است

مطالعات غالباً ویژگی‌ها و تبدیل‌های فرکانسی مورد توجه قرار گرفته است. در شکل‌های (۱) تا (۳) در بخش (د) نمایش بازگشتی هر یک از صداهاى مربوط به کلاس‌های مختلف نشان داده شده و هر دو محور این نمایش از جنس زمان است.

شبکه‌ی پیشنهادی با درک خطوط رسم شده در RP ویژگی‌های زمانی صوت و اتفاق‌هایی که در هر دوره‌ی تنفسی شخص بیمار یا سالم رخ می‌دهد را با بهره‌گیری از مزیت‌های یادگیری انتقالی شبکه‌های کانولوشنی استخراج می‌کند در حالی که در سایر



شکل (۳) - صدای ریه مربوط به کلاس سایر بیماری‌ها که در این ۴ ثانیه سه دوره‌ی تنفسی غیرعادی را در خود جای داده است و صدای دم اول صدای عارضه‌ای Crackle را به همراه دارد، الف) صدای ریه‌ی ضبط شده و طیف‌نگاره‌ی لگاریتم مل آن، ب) صدای ریه بعد از اعمال فیلتر میان‌گذر باترورت مرتبه‌ی ۶ در محدوده‌ی ۵۰ تا ۲۵۰۰ هرتز و طیف‌نگاره‌ی لگاریتم مل آن، ج) صدای ریه پس از حذف صداهاى دارای هم‌پوشانی با صدای ریه و طیف‌نگاره‌ی لگاریتم مل آن که به خوبی حذف نویزها و کاهش اثر سیگنال‌های مزاحم را نشان داده است، د) نمایش بازگشتی که هر دو محور آن زمان است، ه) طیف‌نگاره‌ی MFCCs با استخراج ۱۳ ضریب، و) نمونه‌ای از ورودی شبکه‌ی پیشنهادی که پس از اعمال تبدیل‌های معرفی شده روی هر ثانیه از فایل ۱۰ ثانیه‌ای، ۳ تبدیل به دست آمده که به ابعاد (۲۲۴، ۲۲۴) تغییر سایز داده شده و برای هر فایل ۱۰ ثانیه‌ای داده‌ی ورودی معادل با (۱۰، ۳، ۲۲۴، ۲۲۴) دارد

شده است. برای طبقه‌بندی نیز از شبکه‌های عصبی استفاده شده و صداها در ۳ کلاس سالم، COPD و سایر بیماری‌ها طبقه‌بندی شده است.

۳-۲- استخراج ویژگی و طبقه‌بندی

استخراج ویژگی در مدل پیشنهادی به صورت ویژگی‌های عمیق و مدلی مبتنی بر رزنت^۱، توجه زمانی^۲ و شبکه‌ی بازگشتی انجام

^۱ Temporal Attention

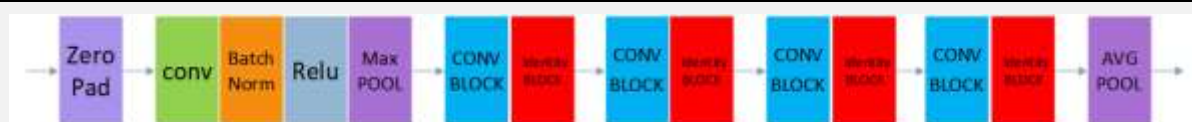
^۲ Resnet

۳-۲-۱- شبکه‌ی رزنت

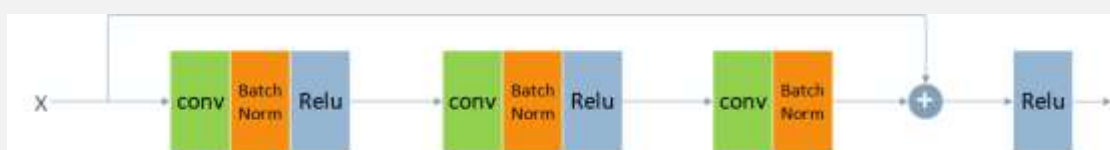
در حوزه‌ی تشخیص بیماری از طریق صدا به دلیل کم بودن داده‌ها ممکن است شبکه به خوبی آموزش داده نشود و یا نتایج برای داده‌های جدید عمومیت نداشته باشد، بنابراین برای رفع این مشکل می‌توان از شبکه‌های استاندارد دارای وزن‌های از پیش آموزش داده شده استفاده کرد. یکی از این شبکه‌ها ResNet بوده که با استفاده از پایگاه داده‌ی ImageNet آموزش داده شده و برای مدل پیشنهادی از این شبکه و یادگیری انتقالی وزن‌های آن استفاده شده است.

در شکل (۴) معماری شبکه‌ی رزنت ۵۰ نشان داده شده که با در کنار هم قرار گرفتن بلاک‌های کانولوشنی و همانی تشکیل شده است. تا پیش از معرفی ResNet مشخص بود که با افزایش عمق شبکه می‌توان به کارایی بهتری دست پیدا کرد. این موضوع به طور خاص با معماری‌هایی مثل AlexNet، VGGNet و GoogleNet نیز به طور مکرر تایید شده است. اما مشکل

بزرگ این است که تمام این معماری‌ها با این که عنوان می‌کنند عمق بیش‌تر بهتر است اما در عمل نمی‌توانند از یک حد خاصی بیش‌تر پیش بروند. با افزایش تعداد لایه‌ها مشاهده شده که دقت کاهش می‌یابد. اگر به یک مدل کم‌عمق که دقت خوبی دارد لایه‌های جدید اضافه شود یک مدل عمیق‌تر به دست می‌آید. انتظار می‌رود در صورتی که دقت مدل با افزایش لایه‌ها کاهش یابد، مدل این توانایی را داشته باشد که لایه‌های اضافه شده نقش نگاشت همانی^۱ را انجام دهند اما این اتفاق رخ نمی‌دهد. در رزنت با اتصال ورودی لایه به خروجی آن $(H(x)=F(x)+x)$ ، خروجی لایه‌ی قبل به لایه‌ی بعد منتقل می‌شود. در واقع برای شبکه ساده‌تر است که مقدار صفر را تولید کند $(F(x)=0)$ به طوری که در خروجی فقط ورودی باشد به جای آن که خودش ورودی را تولید کند. در شکل (۵) معماری بلاک همانی نشان داده شده است. رزنت ۵۰ تقریباً ۲۲/۵ میلیون پارامتر قابل یادگیری دارد [۲۰].



شکل (۴) - معماری شبکه‌ی رزنت ۵۰



شکل (۵) - معماری بلاک همانی شبکه‌ی رزنت

۳-۲-۲- شبکه‌ی توجه زمانی

توجه زمانی می‌تواند به عنوان یک مکانیسم انتخاب زمان پویا در نظر گرفته شود که تعیین کننده‌ی زمان توجه بوده و بخشی از سیگنال که به لحاظ محتوا دارای ارزش اطلاعاتی بیش‌تر است را حفظ می‌کند. در واقع این مکانیسم می‌تواند سیکل‌های دم و بازدم را مورد توجه قرار دهد و به بخش‌هایی که عمل تنفس انجام نمی‌شود توجهی نداشته باشد. مکانیسم توجه باعث ایجاد ساختار یادگیری بازگشتی شده که ارتباط زمانی بین رشته‌ی ورودی و خروجی را ثبت کرده و به شبکه‌ی استخراج کننده‌ی ویژگی اجازه می‌دهد تا بر کاراکترهایی که ارتباط معنایی دارند متمرکز شود. در مطالعه‌ی [۲۱] بیان شده است که اضافه شدن مکانیسم توجه نسبت به حالت مدل پایه همیشه عمل کرد بهتری دارد. همچنین اگر از توجه زمانی در لایه‌های پایینی استفاده شود نتیجه‌ی بهتری به دست خواهد آمد.

در بخش توجه زمانی شبکه‌ی پیشنهادی از یک ورودی به ابعاد $(1, 224, 224)$ استفاده شده که می‌تواند به صورت هر یک از سه تبدیل معرفی شده باشد و در این‌جا برای بهره‌گیری از خصوصیات زمانی از ورودی نمایش بازگشتی (RP) استفاده شده است. این ورودی پس از استخراج ویژگی‌های زمانی، ماتریسی از ضرایب که بسته به اهمیت زمانی مقادیری بین ۰ تا ۱ دارد را در خروجی فعال‌ساز Relu به ابعاد $(64, 112, 112)$ که در لایه‌های ابتدایی رزنت قرار دارد ضرب می‌کند. برای بخش توجه زمانی به جای استفاده از پولینگ که نقش کاهش ابعاد ویژگی‌های استخراج شده را دارد از لایه‌های کانولوشنی با گام‌های متفاوت استفاده شده است تا شبکه با یادگیری، بهترین تابع معادل با پولینگ را ایجاد کرده و در نهایت به ابعادی برابر با خروجی فعال‌ساز Relu در رزنت برسد. در شکل (۶) توجه زمانی اضافه شده به لایه‌های ابتدایی رزنت نشان داده شده و در

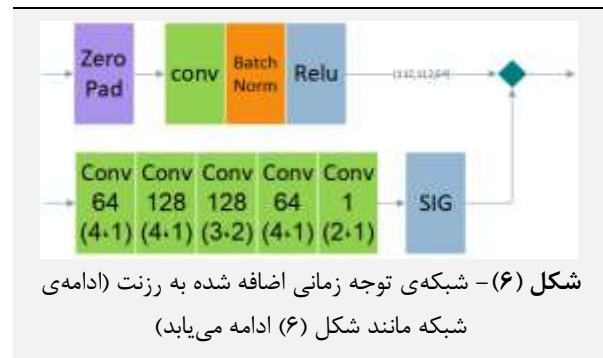
^۱ Identity Mapping

استفاده از تبدیل‌های معرفی شده نمایش‌هایی ایجاد شده است که بازتاب‌کننده‌ی ویژگی‌های زمانی و فرکانسی صوت باشند. در سایر مطالعات، عمده‌ی توجه به نمایش ویژگی‌های فرکانسی معطوف بوده اما در این پژوهش با به کارگیری RP ویژگی‌های زمانی صدا نیز مورد استفاده قرار گرفته است. از آن‌جا که تولید RP برای یک فایل ۱۰ ثانیه‌ای به حافظه‌ی RAM بسیار زیادی نیاز دارد، یک راه حل عمومی کاهش نرخ نمونه‌برداری از فایل صوتی است که امکان از دست رفتن اطلاعات را دارد. اما در روش پیشنهادی برای این که هیچ اطلاعاتی از دست نرود شبکه‌ای برای پردازش موازی جداگانه‌ی هر ثانیه از صدا ارائه شده که در شکل (۷) قابل مشاهده است. بنابراین هر فایل ۱۰ ثانیه‌ای به ۱۰ بخش تقسیم شده و سپس ورودی‌هایی معادل ۱ ثانیه به شبکه‌ی پیشنهادی داده شده است. در شبکه‌ی رزنت توجه زمانی، RP به عنوان ورودی به شبکه‌ی توجه زمانی داده شده است تا با استخراج ویژگی‌های زمانی صوت و ضرب آن‌ها با ویژگی‌های استخراج شده توسط رزنت در لایه‌های اولیه، ویژگی‌ها بر اساس ارزش اطلاعاتی آن‌ها وزن‌دهی شده و در لایه‌های بعدی رزنت ویژگی‌های عمیق‌تری استخراج شود. به طور کلی از لایه‌های دراپ‌اوت در شبکه‌ها برای جلوگیری از وقوع بیش‌برازش^۲ استفاده می‌شود. خروجی شبکه‌های موازی رزنت توجه زمانی با استفاده از گیت‌های بازگشتی دوطرفه در شبکه‌ی بازگشتی، ویژگی‌هایی را استخراج می‌کند که به خروجی ثانیه‌ی قبل و بعد ارتباط دارد. در واقع شبکه‌ی بازگشتی ویژگی‌ها را با توجه به توالی زمانی سیگنال صوتی استخراج می‌کند که ارتباط معنایی با داده‌های قبلی و بعدی دارند. در نهایت شبکه صداها را در سه دسته‌ی سالم، COPD و سایر بیماری‌ها طبقه‌بندی می‌کند.

۴- آزمایش‌های انجام شده و یافته‌ها

برای بررسی عمل‌کرد شبکه‌ی پیشنهادی پنج آزمایش طراحی شده است که شبکه را در حالت‌های مختلف ارزیابی می‌کند. سه کلاس سالم، COPD و سایر بیماری‌ها هر یک شامل ۷۰٪ فایل صوتی ۱۰ ثانیه‌ای بوده که به طور برابر از هر کلاس ۷۰٪ داده‌ها برای آموزش شبکه، ۱۵٪ برای ارزیابی آموزش و ۱۵٪ نیز برای تست و انجام آزمایش‌ها مورد استفاده قرار گرفته است. در آموزش داده‌ها با به کارگیری الگوریتم بهینه‌سازی Adam، شبکه با نرخ یادگیری ۰/۰۰۰۱ تا بیش از ۳۰۰ دوره آموزش داده شده که در هر آزمایش وزن‌های دارای کم‌ترین خطای ارزیابی به عنوان بهترین حالت شبکه انتخاب شده تا برای

آن ورودی (۱،۲۲۴،۲۲۴) از ۵ لایه‌ی کانولوشنی با گام‌های (۱،۴)، (۱،۴)، (۲،۳)، (۱،۴) و (۱،۲) عبور کرده که برای مثال گام (۲،۳) به این معنی است که کرنل در راستای افقی ۳ گام (درایه) به سمت راست و در راستای عمودی ۲ گام به سمت پایین پرش کرده و تأثیری که در بردار ویژگی لایه‌ی قبل دارد برابر با یک لایه‌ی پولینگ (۲،۳) می‌باشد. در نهایت پس از عبور از تابع فعال‌ساز سیگموئید که مقادیر را بین ۰ تا ۱ نگاشت می‌کند، یک بردار با ابعاد (۱،۱۱۲،۱) به دست آمده که یعنی به ازای هر ستون از محور زمان یک ضریب به عمق ۱ وجود داشته که با اعمال ضرایب استخراج شده بر ویژگی‌های خروجی در لایه‌ی Relu از رزنت ارزش اطلاعات قرار گرفته شده در محور زمان مشخص می‌شود. تعداد پارامترهایی که توجه زمانی به شبکه اضافه می‌کند تقریباً برابر با ۳۰۰ هزار پارامتر است.



۳-۲-۳- شبکه‌ی بازگشتی

گیت‌های بازگشتی دوطرفه^۱ (BGRU) اطلاعات را با گیت‌های بعدی و قبلی در یک فایل ۱۰ ثانیه‌ای مبادله کرده تا رفتار زمانی و ارتباط بین ثانیه‌های قبل و بعد در سیگنال صوت مورد توجه قرار گیرد. این بخش می‌تواند با استفاده از تمام اطلاعات ورودی موجود در گذشته و آینده، یک چارچوب زمانی خاص را آموزش دهد. در مطالعه‌ی [۲۲] یک نمونه از لایه‌ی بازگشتی دوطرفه معرفی شده که در ثانیه‌ی t ویژگی‌های استخراج شده با استفاده از وزن‌هایی به دست می‌آید که از طریق حالت مستقیم ویژگی ثانیه‌ی قبل (t-1) و از طریق حالت برگشتی ویژگی‌های ثانیه‌ی بعد (t+1) را مورد استفاده قرار داده تا وزن‌ها را آموزش دهد. ویژگی‌های استخراج شده، وابسته به اطلاعات ورودی ثانیه‌ی قبلی و بعدی این سری زمانی هستند [۲].

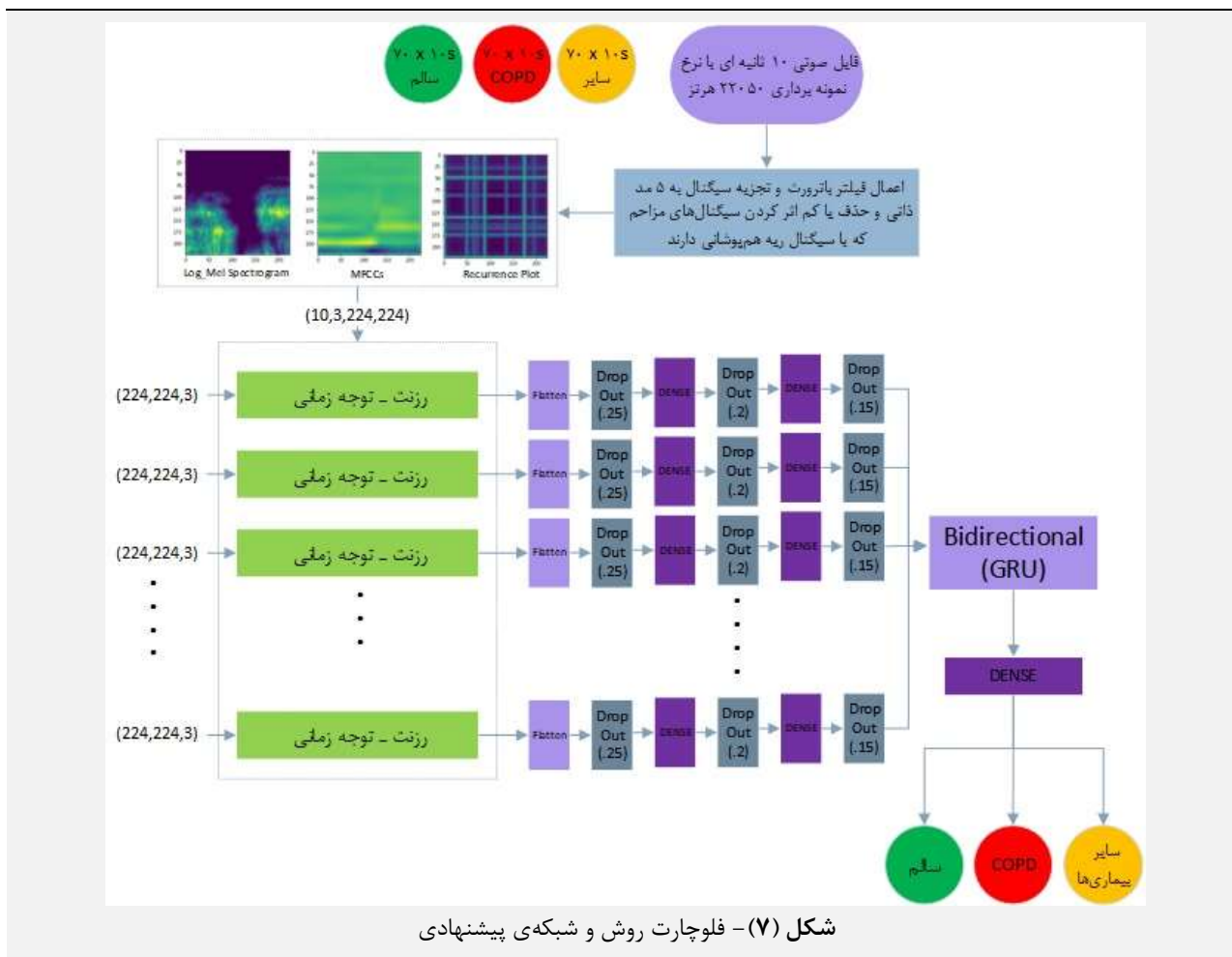
در این پژوهش پس از حذف نویز با فیلتر میان‌گذر باترورث مرتبه‌ی ۶ در محدوده‌ی ۵۰ تا ۲۵۰۰ هرتز، اجزای فرکانسی صدای قلب که با صدای ریه هم‌پوشانی دارد با استفاده از EMD از صدای اصلی تفکیک، حذف و یا کم‌اثر شده است. سپس با

^۲ Overfitting

^۱ Bidirectional Gated Recurrent Unit

هم‌چنین از تابع اختلاف آنتروپی موضوعی^۱ به عنوان تابع خطای ارزیابی استفاده شده است.

آزمایش‌ها مورد استفاده قرار گرفته و دقت به دست آمده نتیجه‌ی یک مرحله انجام آزمایش روی داده‌های تست است.



۳-۴- آزمایش ۳: RP

در این آزمایش نیز به شبکه توجه زمانی اضافه نشده و هر سه کانال آن به صورت نمایش بازگشتی (RP) است. در این حالت بیش‌ترین مقدار دقت تقسیم‌بندی داده‌های تست برابر با ۸۵/۷۱٪ و خطای ارزیابی آموزش شبکه برابر با ۰/۳۶۸۹ است.

۴-۴- آزمایش ۴: MFCCs، Log Mel و RP

در این آزمایش نیز توجه زمانی به شبکه اضافه نشده و برای آن از ۱۰ ورودی (۲۲۴،۲۲۴،۳) استفاده شده که کانال‌ها شامل آرایه‌های دویبعدی از نمایش بازگشتی، MFCCs و لگاریتم مل است. در این حالت نتیجه‌ی بهتری نسبت به حالت‌هایی که در آن‌ها از یک نوع ورودی استفاده شده به دست آمده است. بیش‌ترین دقت تقسیم‌بندی داده‌های تست در یک مرحله‌ی آزمایش برابر با ۹۶/۸۳٪ و خطای ارزیابی آموزش شبکه برابر با ۰/۰۹۰۳ بوده که به میزان قابل توجهی بهبود یافته است.

۱-۴- آزمایش ۱: MFCCs

در این آزمایش نتایج برای حالتی که توجه زمانی به شبکه‌ی رزنت اضافه نشده نشان داده شده است. در واقع برای هر فایل صوتی ۱۰ ثانیه‌ای، ۱۰ ورودی (۲۲۴،۲۲۴،۳) ایجاد شده که هر سه کانال دارای تصویر یکسانی از ویژگی‌های MFCCs است. در این حالت بیش‌ترین مقدار دقت تقسیم‌بندی داده‌های تست در یک مرحله‌ی آزمایش برابر با ۸۰/۹۵ درصد و خطای ارزیابی آموزش شبکه برابر با ۰/۴۲۸۲ بوده که نشان می‌دهد شبکه در آموزش دچار بیش‌برازش نشده است.

۲-۴- آزمایش ۲: Log Mel

در این آزمایش به شبکه توجه زمانی اضافه نشده و هر سه کانال آن طیف‌نگاره‌ی لگاریتم مل است. در این حالت بیش‌ترین مقدار دقت تقسیم‌بندی داده‌های تست در یک مرحله‌ی آزمایش برابر با ۸۸/۸۹٪ و خطای ارزیابی آموزش شبکه برابر با ۰/۳۶۶۱ است.

^۱ Categorical Cross_Entropy

نتیجه‌ی بهتری داشته است. در انتها با اضافه کردن توجه زمانی و تکمیل شدن شبکه‌ی پیشنهادی به ازای سه ورودی دقت ۹۷ درصد به دست آمده که بیان‌گر این است که شبکه یاد می‌گیرد که چه بخش‌هایی از صدا حاوی اطلاعات ارزشمندتری می‌باشد. اما ایرادی که می‌توان از این آزمایش نسبت به آزمایش‌های قبل گرفت زمان بیشتری است که صرف آموزش بخش توجه زمانی می‌شود. در جدول (۴) نتایج این مقاله با نتایج سایر مطالعات مقایسه شده است. روش پیشنهادی در آزمایش ۴ و ۵ از سایر مطالعات دقت بهتری داشته و در واقع در روش پیشنهادی بدون اضافه کردن شبکه‌ی توجه زمانی، تنها با اضافه شدن نمایش بازگشتی که ویژگی‌های زمانی را بازتاب می‌کند بالاترین دقت در بین تمام مقالات به دست آمده است.

جدول (۴) - مقایسه‌ی نتایج این پژوهش با سایر مطالعات

مقاله	مدل	دقت (ACCURACY)
[۱۱]	VGGish_Bigru (Log_mel)	٪ ۸۷/۴۱
[۲۵]	CNN_SVM (STFT)	٪ ۶۵/۵
[۲۶]	CNN	٪ ۹۳/۶۷
[۲۷]	CNN (MFCCs)	٪ ۷۴
[۲۸]	RNN (MFCCs)	٪ ۸۱
[۲۹]	CNN (Log_mel)	٪ ۸۱/۶۲
[۳۰]	CNN (EMD)	٪ ۹۶/۳۹
حاضر	آزمایش ۱	٪ ۸۰/۹۵
حاضر	آزمایش ۲	٪ ۸۸/۸۹
حاضر	آزمایش ۳	٪ ۸۵/۷۱
حاضر	آزمایش ۴	٪ ۹۶/۸۳
حاضر	آزمایش ۵	٪ ۹۷

۶- محدودیت‌ها و پیشنهادها

از محدودیت‌های این پژوهش و سایر پژوهش‌های انجام شده در زمینه‌ی تشخیص بیماری از طریق صدای ریه می‌توان به کم بودن نمونه‌های پایگاه داده اشاره کرد که موجب می‌شود نتایج به دست آمده عمومیت نداشته و صرفاً روی داده‌های استفاده شده نتیجه‌ی خوبی داشته باشد. انتظار می‌رود که با افزایش تعداد داده‌ها نتایج بهتری به دست آید. در این پژوهش برای کم‌رنگ کردن این مساله از یادگیری انتقالی استفاده شده و وزن‌های آموزش داده شده با پایگاه داده‌ی Imagenet به شبکه‌ی پیشنهادی انتقال داده شده است. با توجه به پیشرفت تکنولوژی و ابزارهای پوشیدنی که قابلیت نظارت بر شرایط حیاتی افراد را دارد پیشنهاد می‌شود به جای استفاده از شبکه‌ی رزنت از شبکه‌های سبک و کم‌مصرف استفاده شود که قابل استفاده در ابزارهای پوشیدنی باشد. برای آموزش وزن‌های این

۴-۵- آزمایش ۵: رزنت توجه زمانی

آخرین آزمایش برای فرم نهایی شبکه‌ی پیشنهادی که توجه زمانی به آن اضافه شده صورت گرفته و ورودی مانند آزمایش ۴ است. در این شبکه به دلیل اضافه شدن توجه زمانی، زمان بیشتری صرف آموزش شده و هر اپوک چند برابر حالت‌های قبلی زمان برده و با تعداد اپوک‌های بیشتر نتیجه‌ی بهتری نسبت به آزمایش‌های قبلی به دست آمده است. بیش‌ترین مقدار دقت تقسیم‌بندی داده‌های تست در یک مرحله‌ی آزمایش برابر با ۹۷ درصد و خطای ارزیابی آموزش شبکه برابر با ۰/۰۸۹۳ بوده که کاهش خطای ارزیابی آموزش و دقت بالای تست به این معنی بوده که شبکه به گونه‌ای وزن‌ها را آموزش داده که عمومیت آن برای داده‌های جدید از دست نرفته است.

۴-۶- یافته‌ها

با توجه به ماتریس درهم‌ریختگی که در جدول (۳) نشان داده شده می‌توان نتیجه گرفت که شبکه داده‌های مربوط به دسته‌ی سایر بیماری‌ها را به خوبی دیگر دسته‌ها یاد نگرفته که این موضوع به دلیل تشکیل شدن این دسته از چندین بیماری و کم بودن داده‌ها، قابل پیش‌بینی است. در واقع این شبکه برای طبقه‌بندی دو دسته‌ی سالم و COPD دقت بالاتری دارد اما به منظور این که بتواند COPD را از سایر بیماری‌ها متمایز کند دسته‌ی سوم که شامل نمونه‌های چندین بیماری بوده اضافه شده است با این حال می‌توان با داشتن داده‌های بیشتر این مشکل را رفع کرد و نتایج بهتری به دست آورد.

جدول (۳) - ماتریس درهم‌ریختگی

ماتریس درهم‌ریختگی	سالم	COPD	سایر بیماری‌ها
سالم	٪ ۹۷/۱۴	٪ ۱/۴	٪ ۱/۴
COPD	۰	٪ ۹۲/۸۵	٪ ۱۱/۴
سایر بیماری‌ها	٪ ۲/۸	٪ ۵/۷	٪ ۸۷/۱۴

۵- نتیجه‌گیری

در بخش قبل مشاهده شد که شبکه‌ی پیشنهادی بدون توجه زمانی برای ورودی MFCCs به دقت ۸۰/۹۵ درصد، برای ورودی RP به دقت ۸۵/۷۱ درصد و برای ورودی Log mel به دقت ۸۸/۸۹ درصد دست یافته و زمانی که هر سه ورودی به شبکه داده شده به دقت ۹۶/۸۳ درصد رسیده که نشان می‌دهد مجموع ویژگی‌های استخراج شده برای سه ورودی از تک‌تک ورودی‌ها نتیجه‌ی بهتری داشته است. در واقع شبکه از تلفیقی از ویژگی‌های زمانی و فرکانسی استفاده کرده که در مقایسه با حالتی که فقط از ویژگی‌های فرکانسی و یا زمانی استفاده کند



- [11] Orjuela-Cañón, D. Alvaro, F. D. Gómez-Cajas, and R. Jiménez-Moreno, "Artificial Neural Networks for Acoustic Lung," *Prog. Pattern Recognition, Image Anal. Comput. Vision, Appl. Springer Int. Publ.*, pp. 214–221, 2014.
- [12] H. Chen, X. Yuan, Z. Pei, M. Li, and J. Li, "Triple-Classification of Respiratory Sounds Using Optimized S-Transform and Deep Residual Networks," *IEEE Access*, vol. 7, pp. 32845–32852, 2019.
- [13] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A Lightweight CNN Model for Detecting Respiratory Diseases from Lung Auscultation Sounds using EMD-CWT-based Hybrid Scalogram," vol. XX, no. Xx, pp. 1–9, 2020, [Online]. Available: <http://arxiv.org/abs/2009.04402>.
- [14] R. X. A. Pramono, S. Bowyer, and E. Rodriguez-Villegas, "Automatic adventitious respiratory sound analysis: A systematic review," vol. 12, no. 5, 2017.
- [15] ICBHI dataset: https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge
- [16] A. K. Dwivedi and G. S. Member, "Algorithms for Automatic Analysis and Classification of Heart Sounds — A Systematic Review," *IEEE Access*, vol. 7, pp. 8316–8345, 2019.
- [17] H. Salman, N. Ahmadi, R. Mengko, A. Z. R. Langi, and T. L. R. Mengko, "Performance comparison of denoising methods for heart sound signal," in *2015 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 435–440, IEEE, Nusa Dua, Indonesia, 2015.
- [18] J. P. Eckmann, O. Oliffson Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Epl*, vol. 4, no. 9, pp. 973–977, 1987.
- [19] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Phys. Rep.*, vol. 438, no. 5–6, pp. 237–329, 2007.
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." [Online]. Available: <http://image-net.org/challenges/LSVRC/2015/>.
- [21] Z. Zhang, S. Xu, S. Zhang, T. Qiao, and S. Cao, "Learning Attentive Representations for Environmental Sound Classification," *IEEE Access*, vol. 7, pp. 130327–130339, 2019.
- [22] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [23] AudioSet dataset: <https://research.google.com/audioset/dataset/index.html>

شبکه می‌توان از پایگاه داده‌ی [۲۳] Audioset استفاده کرد که حاوی دو میلیون صدای ۱۰ ثانیه‌ای در ۵۲۷ کلاس است. همچنین برای تنظیم مجدد وزن‌ها می‌توان از داده‌های صدای ریه استفاده کرد. علاوه بر این می‌توان با ایجاد سیستم نرم‌افزاری یا فضای ابری در اینترنت این امکان را فراهم کرد که هر شخص از هر مکان به راحتی با وارد کردن صدای ریه، وضعیت سلامت خود را بررسی کند.

۷- مراجع

- [1] X. Xu et al., "A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, 2020, doi: 10.1016/j.eng.2020.04.010.
- [2] L. Shi, K. Du, C. Zhang, H. Ma, and W. Yan, "Lung Sound Recognition Algorithm Based on VGGish-BiGRU," *IEEE Access*, vol. 7, pp. 139438–139449, 2019.
- [3] F. Demir, A. Sengur, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," *Heal. Inf. Sci. Syst.*, vol. 8, no. 1, 2020.
- [4] Altan, G., Kutlu, Y., & Allahverdi, N. Deep learning on computerized analysis of chronic obstructive pulmonary disease. *IEEE Journal of Biomedical and Health Informatics*, 24(5), 1344–1350. 2020.
- [5] A. A. Saraiva et al., "Classification of images of childhood pneumonia using convolutional neural networks," *BIOIMAGING 2019 - 6th Int. Conf. Bioimaging, Proceedings; Part 12th Int. Jt. Conf. Biomed. Eng. Syst. Technol. BIOSTEC 2019*, no. Biostec, pp. 112–119, 2019.
- [6] D. Perna and A. Tagarelli, "Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks," *Proc. - IEEE Symp. Comput. Med. Syst.*, vol. 2019-June, pp. 50–55, 2019.
- [7] R. Liu, S. Cai, K. Zhang, and N. Hu, "Detection of adventitious respiratory sounds based on convolutional neural network," in *2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*. IEEE, 2019, pp. 298–303.
- [8] S. I. Khan and R. B. Pachori, "Automated classification of lung sound signals based on empirical mode decomposition," *Expert Syst. Appl.*, vol. 184, no. October 2020, p. 115456, 2021.
- [9] S. İçer and Ş. Gengeç, "Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds," *Digit. Signal Process. A Rev. J.*, vol. 28, no. 1, pp. 18–27, 2014.
- [10] M. Bahoura, "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes," *Comput. Biol. Med.*, vol. 39, no. 9, pp. 824–843, 2009.