Expert Systems With Applications 262 (2025) 125713

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A cognitive few-shot learning for medical diagnosis: A case study on cleft lip and palate and Parkinson's disease

Pei Yin [a,b], Junjie Song [a], Yassine Bouteraa [c,*], Leren Qian [d], Diego Martín [e], Mohammad Khishe [f,g]

[a] Business School, University of Shanghai for Science and Technology, Shanghai 200093, China
[b] School of Intelligent Emergency Management, University of Shanghai for Science and Technology, Shanghai 200093, China
[c] Department of Computer Engineering, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia
[d] School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, 85281, USA
[e] Department of Computer Science, Escuela de Ingeniería Informática de Segovia, Universidad de Valladolid, Segovia 40005, Spain
[f] Innovation Center for Artificial Intelligence Applications, Yuan Ze University, Taiwan
[g] Jadara University Research Center, Jadara University, Irbid, Jordan

ARTICLE INFO

Keywords:
Cognitive few-shot learning approach
Neurodegenerative disorder
Few-shot learning
Medical diagnosis

ABSTRACT

This paper presents a novel cognitive few-shot learning (CFSL) for the diagnosis of cleft lip and palate and Parkinson's diseases. The proposed method utilizes computational analysis of paralinguistic features to expedite the diagnostic process. Unlike other methods that rely on complex and fragmented representations, CFSL trains itself to recognize patterns that are easily interpretable by humans. Rather than learning a single, unstructured metric space, CFSL combines the outputs of individual landmark (LM) learners by mapping LMs into semi-formation spaces. In order to assess the effectiveness of CFSL, we conducted a comparative analysis with seven distinct FSL-based models, including momentum contrastive learning for FSL (MCFSL), self-updating FSL (SUFSL), mutual info multi-attention FSL (MAMIFSL), dual class representation FSL (DCRFSL), Improved FSL (IFSL), *meta*-knowledge for FSL (MKFSL), and prototypical networks (ProtoNet), using three popular datasets, namely GPRS, CIEMPIESS, and PC-GITA. The findings indicate that CFSL demonstrated superior performance compared to the highest-performing baseline frameworks for the 5-shot (5-sh) and 1-shot (1-sh), having an average enhancement of 4.39% and 4.49%, respectively. CFSL demonstrated better performance than the ProtoNet baseline in both 1-sh and 5-sh across all datasets, with an improvement of 12.966% and 11.033%, respectively. In addition, we performed ablation tests to assess the effects of variables such as the density of LMs, the structure of the network, the distance measure used, and the positioning of LMs. The CFSL approach, if adopted in hospitals, has the potential to enhance the precision and efficiency of diagnosis for cleft lip and palate as well as Parkinson's disease.

1. Introduction

Cleft lip and palate (CP) and Parkinson's disease (PAD) are two widespread medical diseases. PAD is a neurological movement condition characterized by tremors, stiffness, and problems with coordination and balance (Zhuo, 2024). CP is a congenital abnormality affecting the lip and/or top of the lip, which can cause speech, nutrition, and oral difficulties (Song et al., 2024). Effective management and treatment of many disorders require a precise and prompt diagnosis (Zou, 2024).

Conventional machine learning and deep learning techniques have proven effective in medical diagnosis, including CT image diagnosis (Xu et al., 2022); cancer-related treatments (Wang, 2024); liver disease (Sun et al., 2023; He, 2024); nervous system diseases (Hui et al., 2024); covid19 diagnosis (Xu et al., 2022); and electroencephalograph (Borlea et al., 2017; Zhao et al., 2021; Borlea et al., 2022). Nonetheless, these algorithms have some restrictions, such as operations on small datasets and the need for large amounts of labeled data to be functional (Jing, 2024; Mohammadzadeh et al., 2024; Mohammadzadeh et al., 2023). Besides, these techniques tend to be black boxes, making it challenging to explain what features of the model account for its predictions

* Corresponding author.
E-mail addresses: pyin@usst.edu.cn (P. Yin), yassine.bouteraa@isbs.usf.tn (Y. Bouteraa), lqian15@asu.edu (L. Qian), diego.martin.andres@uva.es (D. Martín), m_khishe@alumni.iust.ac.ir (M. Khishe).

https://doi.org/10.1016/j.eswa.2024.125713
Received 10 April 2023; Received in revised form 29 October 2024; Accepted 3 November 2024
Available online 4 November 2024
0957-4174/© 2024 Published by Elsevier Ltd.

(Mohammadzadeh et al., 2024; Yan et al., 2023).

Both *meta*-learning (MeL) and few-shot learning (FSL) techniques are promising since they do not require large volumes of data and allow the model to solve new tasks. MeL is simply defined as a model's capability of acquiring a strategy for learning, whereas FSL allows models to assimilate sufficient labeled data to be effective.

1.1. Motivation and research gaps

While there is ample evidence that MeL and FSL have yielded positive results in other areas (Sun et al., 2023); few studies have specifically explored their application in medical diagnostics for CP and PAD. More importantly, the more traditional approaches of MeL are rather opaque, which prevents users from establishing a basic understanding of the factors that affect the model's predictions.

The objective of this research is to demonstrate the flexibility and applicability of the suggested Few-Shot Learning approach, CFSL, in the field of medical diagnosis. The domain of medical diagnosis covers a diverse array of disorders and diseases, each presenting distinct problems and diagnostic information. The objective of this research is to illustrate how CFSL can easily relate to two completely opposite specializations: CP and PAD. Though it may sound odd, CP and PAD present the best-case scenario in proving CFSL's versatility in handling varied medical data and performing different medical tasks. It is important to elaborate at the outset that the coexistence of CP and PAD is not indicative of their structural or clinical relationship. It rather serves the purpose of demonstrating the potential of CFSL as an effective tool for medical and scientific studies in various areas. With the help of the examples from these different fields, we aim to demonstrate that CFSL has the potential to enhance the accuracy or efficiency of diagnosis in many medical disciplines.

1.2. Objective and contributions

The objective of this study is to provide a unique approach for diagnosing CP and PAD utilizing a CFSL. The suggested method leverages the automated evaluation of paralinguistic features in voice signals to facilitate early diagnosis. The generalization ability of CFSL is boosted by obtaining the capacity for learning along LM qualities that are interpretable by humans. CFSL produces abstractions of high-level LMs into semi-formation metric space instead of learning in a single arbitrary metric space, effectively integrating the outcomes of independent LM learners.

The subsequent points delineate the principal contributions of the paper:

- Cognitive FSL (CFSL)

The initial contribution is the development of a new framework called CFSL, which uses few-shot learning approaches to diagnose PD and CLP. CFSL integrates the outcomes of several autonomous LM learners, hence improving the model's ability to apply knowledge to various datasets and clinical situations. The utilization of this composite technique enables the model to generate precise predictions even in scenarios when data are scarce, which is a frequently encountered obstacle in medical diagnosis.

- Paralinguistic Features for Disease Diagnosis

The second contribution of this work is the application of paralinguistic features, which are derived from speech signals, to medical diagnosis contexts. Paralinguistic features, including pitch, prosody, rhythm, etc., offer great insights into the speech of a person and hence assist in determining their health condition. We introduce these characteristics by integrating these general components into our CFSL framework, thereby simplifying and revolutionizing the diagnostic process for medical professionals.

- Improved Diagnostic Accuracy

The final contribution is showcasing CFSL's exceptional diagnostic precision in comparison to established benchmark models. We perform thorough assessments on three separate datasets, encompassing both cases of CLP and PD, in order to demonstrate the strength and efficiency of CFSL. CFSL has the potential to raise the speed and precision of medical diagnosis in clinical settings by significantly improving accuracy compared to current models.

This research introduces the CFSL framework, which utilizes paralinguistic aspects to enhance diagnostic accuracy for CLP and PD. The organization of the paper is as follows: Section 2 provides an overview of the studies that are relevant to the topic. Section 3 explains the technique that is proposed for this research. Section 4 focuses on the experiments conducted. Finally, Section 5 summarizes and closes the study.

2. Related works

Research on medical diagnosis utilizing machine learning (ML) and deep learning (DL) approaches is already abundant (Latif et al., 2019); especially concerning CP and PAD (Yao et al., 2022). Mei et al. (Mei et al., 2021) focused on reviewing several studies related to the theme of the use of ML detection tools in patients with PAD. They noted that a different ML instrument had been implemented in the PAD diagnosis, and it worked well. However, such approaches also face some challenges, such as the necessity of substantial quantities of annotated data and engineered features.

In overcoming the shortcomings of traditional ML techniques, academic specialists have begun using DL in relation to PAD diagnosis. As per the findings of the survey of Loh et al. (Loh, 2021); RNNs and CNNs have been the most preferred DL models in this context. A notable merit of DL is that it relies on its in-built capabilities to elevate features extracted from primary data. All the same, these approaches still rely on a lot of labeled data in order to achieve desired effectiveness.

In their work, Oh et al. (Oh, 2020) developed a deep learning approach for the diagnosis of PAD exploiting EEG signals. The authors employed a CNN architecture that was fitted to a large collection of EEG signals acquired by PAD patients and HLC subjects. The outcome was the ability to diagnose with rather promising accuracy, but it required quite a large amount of fully annotated evidence for training, and its comprehendibility was low.

Vinas et al. presented a DL model for CP diagnosis that utilized 3D facial images (Viñas, 2022). The authors utilized a CNN architecture that was trained on a sizable dataset of facial images collected from both CP patients and healthy individuals. Despite the promising accuracy demonstrated by the results, the approach required a significant amount of labeled data for training and lacked interpretability.

Both traditional ML and DL techniques have been utilized to develop predictive models for medical diagnosis. However, these methods frequently encounter challenges when dealing with small datasets and rely on a significant amount of labeled data for training. Also, one of their obvious limitations is that they fail to explain the attributes that lay behind the model's output.

In this regard, two new approaches, MeL and FSL, have shown a lot of promise in getting models to learn and adapt to new tasks with little data (Xu et al., 2022). Wang et al. (Wang et al., 2021) developed a MeL concept that was used in the diagnosis of pad based on an acoustic investigative approach to the speech. The authors trained a deep neural network structure with a small-sized voice dataset that was derived from strokes as well as healthy PAD voice samples. The DNN was subsequently fine-tuned on a small set of new patients to enhance its generalization ability. The results revealed promising accuracy in diagnosis, but the approach had limited interpretability.

Zhang et al. (Zhang et al., 2021) introduced a novel approach to the diagnosis of Parkinson's disease using speech data. Their method involved using a CNN architecture, which was first trained on a small dataset of facial images of patients with CP and controls, and then the CNN was retrained on a new small cohort of patients to improve the general applicability of the model. While the approach demonstrated encouraging diagnostic accuracy, it was noted that it lacked interpretability.

While previous studies have demonstrated promising results for the application of MeL and FSL in medical diagnosis, there remains a scarcity of research in this area, particularly in relation to CP and PAD. Furthermore, existing MeL techniques often lack interpretability, which can hinder understanding of the specific features contributing to the model's predictions.

Research on PD diagnosis has demonstrated that FSL is useful for deciphering movement and speech data. Early detection of PD has been dramatically facilitated by FSL models' capacity to identify tiny changes in speech, providing a non-invasive diagnostic tool (Ngo et al., 2022). There needs to be better generality, though, because these models frequently fail to account for the fact that speech habits vary among demographics. Like other deep learning-based models, FSL applied to movement data has identified early-stage PD with minimum training data; however, this comes at the tradeoff of high computational requirements (Hossain & Shorfuzzaman, 2023).

Using FSL to diagnose CLP brings its own unique set of advantages and disadvantages. As an example, compared to conventional machine learning methods, the requirement for large image datasets has been drastically diminished when FSL is applied to 3D imaging for CLP diagnosis. Nevertheless, a crucial consideration in clinical contexts is the restricted interpretability of these models' decision-making processes, which somewhat dampens this progress (Aubreville, et al., 2023).

Genetic analysis in CLP cases is another area where FSL shows promise; this is because it helps to grasp complicated genetic markers with less data. Although this is a novel technique, the complexity and diversity of genetic components typically limit the accuracy of these models. This investigation shows a gap in the FSL framework's ability to provide improved accuracy in genetic analysis models (Wurfbain, 2023).

Adding FSL to MRIs for the purpose of diagnosing PD is another exciting development. When it comes to distinguishing PD from other forms of neurodegeneration, these models have proven to be very accurate. Their effectiveness, however, is dependent on the quality of the MRI scans; they can only handle low-quality pictures to a certain extent (van der Hoek-Snieders et al., 2020).

There are significant and complex knowledge gaps in these investigations. An obvious need exists for computationally efficient, interpretable, and flexible FSL models that can provide good results with little training data. Essential issues that require fixing include how well these models generalize to different populations, how well they handle different types of data, and how efficiently they use computational resources (Song et al., 2023).

It is worth noting that the proposed CFSL can bring a remarkable advancement in this area. To compensate for these limitations, the current approach describes a model that can work with various types of information and manage different data quality more effectively. Such an approach has the potential to revolutionize medical diagnostics, particularly in complex and heterogeneous conditions involving CLP and PD.

One major barrier to reaching the goal of achieving human-level capabilities comparable to Maslow's is building models that can accurately predict all possible unknown problems with only a handful of labeled examples. Humans can't quickly acquire brand new skills or hold credible beliefs about the world entirely unassisted because they do not possess reusable, structured perceptions. Structures learned without any constraints imposed on them by the current MeL methods enabled MeL algorithms to learn quite complex concepts across many previously labeled tasks. Our proposed MeL method, PML, enhances generalizability by training itself to learn along perception dimensions that humans can understand. As an alternative to learning a shared unstructured metric space, PML efficiently merges the results of separate perception learners by learning maps of high-level observations into semi-structured metric areas.

3. Proposed methodology

The diagnosis of CP and PAD is based on two general steps that are feature extraction and classification. Time-frequency analysis is a common practice used to speed up calculations in the short-time Fourier transform (STFT). In contrast to this, our study utilized a different approach whereby the CP and PAD signals were transformed into images, and categorization employed the spectrogram.

DNNs have gained popularity as a preferred method for identifying CP and PAD signals. Nevertheless, the limited availability of samples can present a substantial obstacle in this effort (Lu, et al., 2023). In order to address this constraint, we have devised the CFSL model to achieve more precise classification. The classification technique consists of three central portions: initial processing, the extraction of LM, and classifying (Li et al., 2022). Fig. 1 illustrates the components of our proposed classification framework.

Fig. 1 illustrates the flow of the CFSL model from input-raw spectrogram to landmarks and classification. This figure illustrates how a short-time Fourier transform processes the raw audio signals from the datasets to obtain the spectrograms. The landmark extraction process, which relates to significant features of the power thresholds, aims to locate the points in the spectrogram. Using clustering algorithms, we cluster the landmarks and embed each spectrogram into a semi-formation metric space. We represent each landmark as a tuple (t,p), where t represents the time when the power level exceeds the threshold, and p represents the frequency of the power level. The final output of the model is a class label, which defines the order of the proximity of the query to the set of support in the landmark's metric space.

3.1. Raw datasets

The proposed diagnostic model was trained and assessed using three

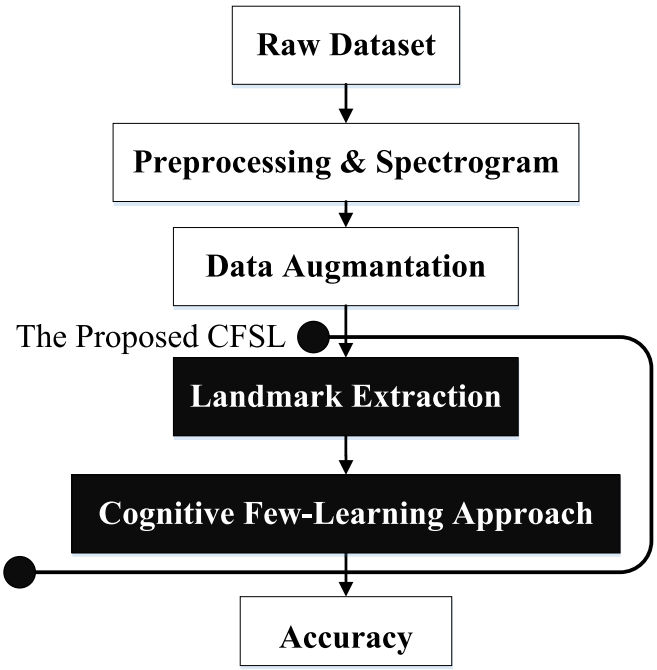


Fig.1. The schematic representation of the suggested system of classification.

well-known datasets: PC-GITA, CIEMPIESS, and GPRS. The subsequent subsections provide comprehensive descriptions of each dataset.

3.1.1. CIEMPIESS dataset[1]

The CIEMPIESS corpus was utilized as the dataset for the genesis of the CFSL, as stated in (Hernández-Mena & Herrera-Camacho, 2014). The aforementioned corpus comprises 16-bit audio records at 16 kHz of sampling rate. The audio files were sampled from a collection of 17 h of Mexican-Spanish podcast material. The corpus lacks background sound, music, or foreign accents. The corpus contains 96 males and 45 females within the recorded files, making a total amount of 16,717 audio records.

3.1.2. GPRS dataset

The additional dataset utilized in this investigation was contributed by the GPRS group of the Colombian National University (CNU). The dataset comprises audio recordings of the speech of 58 children who received HLC repair surgery and 135 children who underwent CP correction surgery, spanning an age range of five to fifteen years. Following cleft repair surgery, patients commonly have hypernasal speech. The participants were instructed to articulate a sequence of Spanish words that encompassed diverse phonemes, each indicating distinct manners and places of articulation, such as /chuzo/, /bola/, /papa/, /coco/, /susi/, /jugo/, /gato/, and /mano/. The database includes 130 hypernasality and 108 healthy control (HLC) recordings for each vowel, with each child producing the vowels twice (Orozco-Arroyave et al., 2016).

3.1.3. PC-GITA dataset

As a means of acquiring the necessary speech utterances associated with PAD, we accessed the PC-GITA database (Orozco-Arroyave et al., 2014). This database consists of audio files from 100 subjects who are native speakers of Spanish from Colombia. Out of these individuals, 50 are in the HLC group, while the remaining 50 are in the PAD group. The subjects were asked to read aloud a text that included a number of thirty-six frequently used words. They were also asked to perform an unstructured speech where they were required to talk about what they did during the course of the day. The data was consistent and controlled for such variables as age, level of education, and gender using MDS-UPADRS-III. A complete characterization of the signal features of the PAD, HLC, and CP groups was given in Table 1 by the authors. Such an analysis incorporated minimum maximum, root mean square (RMS) and mean, crest factor, dynamic range, signal length, and time of correlation, amongst others. The purpose of this analysis was to gain a deeper understanding of the distinctions between these signals.

Although the characteristics of PAD, HLC, and CP signals have been studied, it has been suggested that nonlinear effects may also play a role in their classification. Abnormal vocal fold resonances and an inability to extend the velum may be associated with CP, while PAD may result in dysfunction of muscle control of the limbs and of the speech-making muscles. For the purpose of speech synthesis, nonlinear parameterization is a way of recovering the system dynamics that created a speech signal by rebuilding the phase space. Inverse engineering graphical methods for attractor analysis, also termed phase space plots, find their practical application in various engineering disciplines to help visualize and analyze complex systems over time. These types of graphs can be found in the case of nonlinear dynamical systems, chaotic theory stress, and complicated systems analysis. Studying how Parkinson's disease, epilepsy, or cardiac arrhythmias progress would be invaluable in concentrating on their basic underlying dynamics. Attractor plots illustrate phenomena in motion, such as how physiological variables or brain activity change over time. This plot is of help to understand the nature of disease patterns as well as the state of the disease in the future. The trajectories of attractors constructed in this space should help us greatly to understand its underlying dynamics. Nonetheless, because these phenomena are described by unknown equations, there is no other strategy rather than an indirect approach employed based on Taken's embedding theorem (Takens, 1980). Fig. 2 shows the attractor plots of the subjects with PAD, HLC, and CP using this methodology.

The figure shows three attractors plots, each for different datasets or conditions:

a) HLC: This figure contains the attractor of healthy control patients. The macroscopic phenomenon is labeled with a specific, focused, and often repeating pattern in the phase space, which is interpreted as a time-invariant regular behavior of the graph variables.
b) PAD: In that case the attractor plot of the diseased state with Parkinson's disease is seen, which shows dispersion of pattern and lesser regularity as compared to HLC, which again shows erratic pattern lock dynamics. This plot may indicate the influence of Parkinson's disease on the measured system, suggesting a greater degree of unpredictable behavior.
c) CP: The attractor plot for cleft lip and palate exhibits a distinct distribution of points that differs from both the healthy control and Parkinson's disease. The pattern has a greater level of complexity and is distributed throughout the phase space, suggesting a distinct form of dynamic behavior in the system as compared to the other two circumstances.

In general, these attractor plots are employed to visually represent the intricate dynamics of a system. They can uncover fundamental patterns and disparities in the functioning of a system under different circumstances or ailments, such as PAD and CP, in comparison to a healthy state. Therefore, the following section aims to utilize the preprocessing and spectrogram phases.

3.2. Preprocessing

A proper preprocessing method for CP and PAD analysis is time–frequency analysis. An often-employed method involves utilizing the STFT on the signal $x(t)$ in the following manner (Covell & Richardson, 1991):

$$STFT(x(t)) = \int_{-\infty}^{\infty} x(\upsilon)\varphi(\upsilon - t)e^{-2jf\pi\upsilon}d\upsilon \tag{1}$$

The window function, denoted as $\varphi(t)$, must satisfy the following restriction:

$$\|\varphi(t)\| = 1 \tag{2}$$

In the discrete domain, the formulation of the STFT is given by Eq. (3) (Covell & Richardson, 1991).

$$STFT(x(n)) = \sum_{-\infty}^{\infty} x(k)\varphi(k - n)e^{-2jmk\pi} \tag{3}$$

The STFT is commonly employed in the examination and manipulation of signals related to CP and PAD because it can effectively capture the temporal changes in frequency components. Its advantage lies in its ability to be easily understood and its capability to produce signals with complicated values. Nevertheless, due to the matrix output of the STFT instead of a vector, visualizing the complex-valued information can provide a challenge. In order to tackle this issue, the STFT is commonly represented using a logarithmic spectrum $20log10(\text{STFT}(n,\ k))$. This representation is then shown as a heatmap, also known as a spectrogram. This visualization method allows for the examination of crucial signal characteristics in order to analyze CP and PAD. Fig. 3 illustrates the spectrogram of the genuine CP and PAD signals acquired from the datasets employed in this investigation.

[1] https://ciempiess.org/downloads.

Table 1
The characteristics of PAD, HLC, and CP signals.

Metrics	PAD men	HLC men	PAD women	HLC women	CP child	HLC child
Maximum value	1	1	1	1	1	1
Mean value	−7.77e-18	−8.75e-18	8.68e-19	1.36e-19	−3.49e-18	1.25e-18
Minimum value	−0.84733	−0.84541	−0.91780	−0.66450	−0.83179	−0.92980
RMS value	0.15117	0.092909	0.11490	0.13819	0.15440	0.11164
Crest factor (dB)	16.4059	20.6390	18.7939	17.1890	16.2290	19.0429
Dynamic range (dB)	160.3821	145.1419	100.0169	97.7390	166.3511	107.471
correlation (s)	0.083579	0.018589	0.0559	0.03563	0.050859	0.03580
Signal length (s)	15.8699	16.9149	15.3220	14.7989	20.1810	15.6609

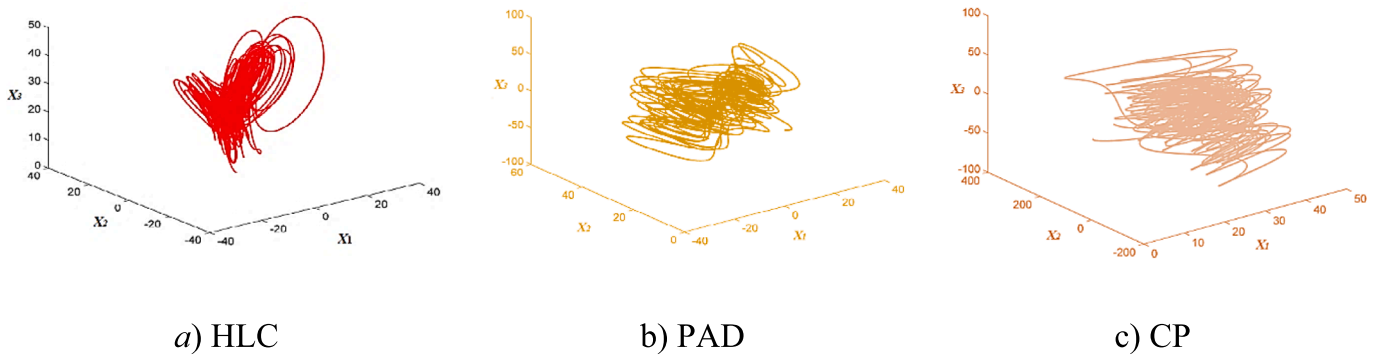


Fig.2. PAD, HLC, and CP attractor plots.

The spectrogram diagram is utilized as input for the models in the suggested classification framework developed in this research. The spectrograms of PAD and CP are examined to identify significant features. It is essential to consider that the time–frequency resolution of the STFT is dependent on the choice of the time window $x(t)$. It is necessary to build a time–frequency accumulation window in order to acquire a correct pattern for data in the combined time–frequency domain. As a result of the fixed product of a signal's bandwidth and its duration, the energy density is constrained by the Heisenberg-Gabor uncertainty principle.

Through a comparative study of signal properties in different domains (time, spectral, and cepstral), it is clear that the differences between each group are tiny and challenging to identify with traditional classification techniques, particularly in small datasets. This classification challenge can be tackled using one of two methods: firstly, by utilizing a more advanced and potentially deep learning network that can accurately identify subtle distinctions with the aid of a significantly larger dataset; secondly, by employing a model specifically designed to isolate and highlight distinguishing characteristics within the data. The latter technique relies on the model's capacity to amplify the importance of particular qualities that are highly indicative of variations across groups. This technique allows for accurate classification even when dealing with complicated signal characteristics and a limited number of data points.

3.3. Data augmentation

In order to create the input power spectrum for the suggested framework, we utilized a setup consisting of 126 temporal intervals and 128 frequency bins based on the Mel scale. The STFT was computed using a window duration of 32 *ms* and a step duration of 4 *ms*, yielding a total of 126 time steps. The input spectrogram was generated by converting 512 frequency points to the Mel-scale using 128 filters. This conversion preserved the fundamental frequency profile, even with substantial frame sizes and a high number of Mel filters. The input power spectrum was created by employing time shifts of 250 ms between each 500 *ms* segment.

In order to improve the model's capacity to generalize and overcome the constraints of the dataset, we utilized data augmentation methods that involve temporal and frequency modifications (Park, et al., 2019). This data augmentation entailed implementing a time-stretched additive noise with a variance of 1 inside the specified parameter range of [0.8, 1]. Additionally, pitch shifting was performed within the specified parameter range of [-2, 2]. In addition, we employed time masking and frequency masking techniques by masking a sequence of t consecutive time steps and f subsequent frequency bins, respectively, using the mean value.

We integrated the previously described changes into a unified time–frequency spectrogram, which was then inputted into the network. Our methodology is resilient and capable of managing partial loss of frequency data and temporal distortions. In order to obtain the Mel-feature power spectrum and feed it into the network, we subjected the input signal to temporal stretching, Gaussian noise, and pitch shifting, as illustrated in Fig. 4. Further outcomes of arbitrarily masking the time–frequency elements of the initial signal can be seen in the figure.

Fig. 4 illustrates the impact of different audio augmentation methods on a spectrogram, which is a graphical depiction of the frequency spectrum of a sound stream as it changes over time. Each panel exhibits the spectrogram of the signal following the application of a specific type of augmentation:

Original: the original spectrogram without any augmentation, assisting as the control for comparison, which is formulated as Eq. (4).

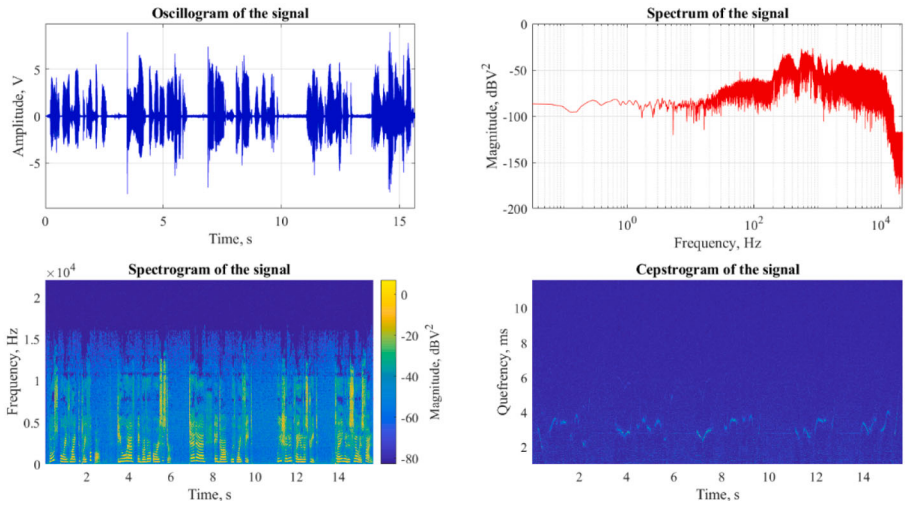
$$S_{original}(t,f) = |STFT(x(t))| \quad x(t):\ \text{Original Signal} \tag{4}$$

Row-Resampling: Modification of the signal's sampling rate, which can impact the temporal resolution of the spectrogram as Eq. (5).

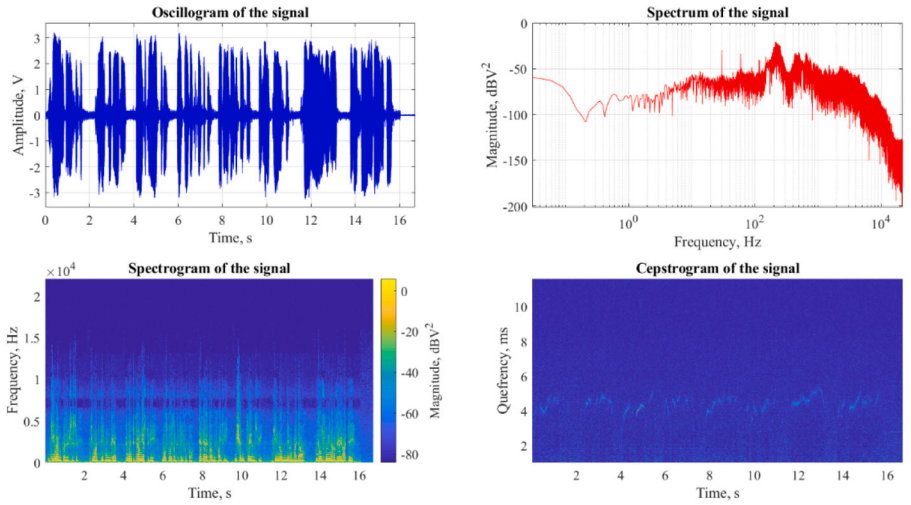
$$x_{resampled}(t) = x(\frac{t}{\alpha}) \quad S_{resampled}(t,f) = \left|STFT(x_{resampled}(t))\right| \tag{5}$$

Where a denotes the time stretching or compressing coefficient.

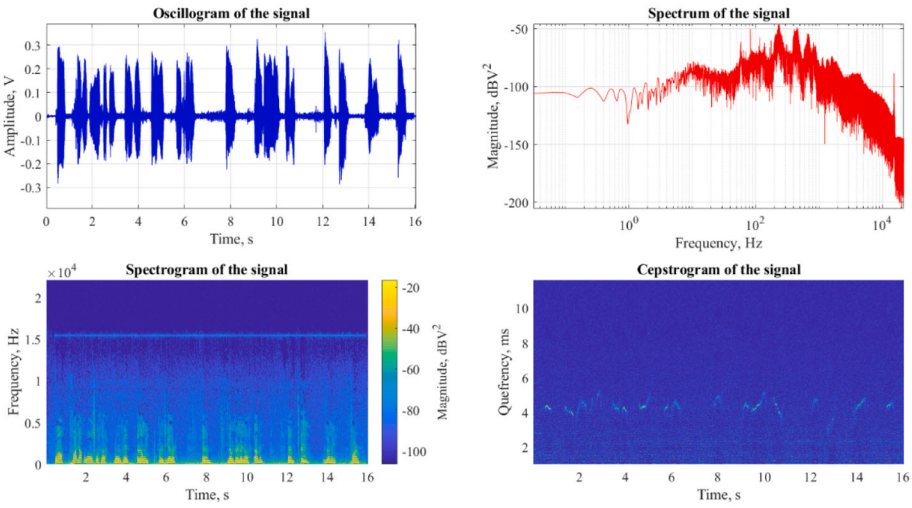
Noise: The introduction of extraneous sound into the signal, resulting in a granular appearance throughout the spectrogram, which can be



a) HLC men

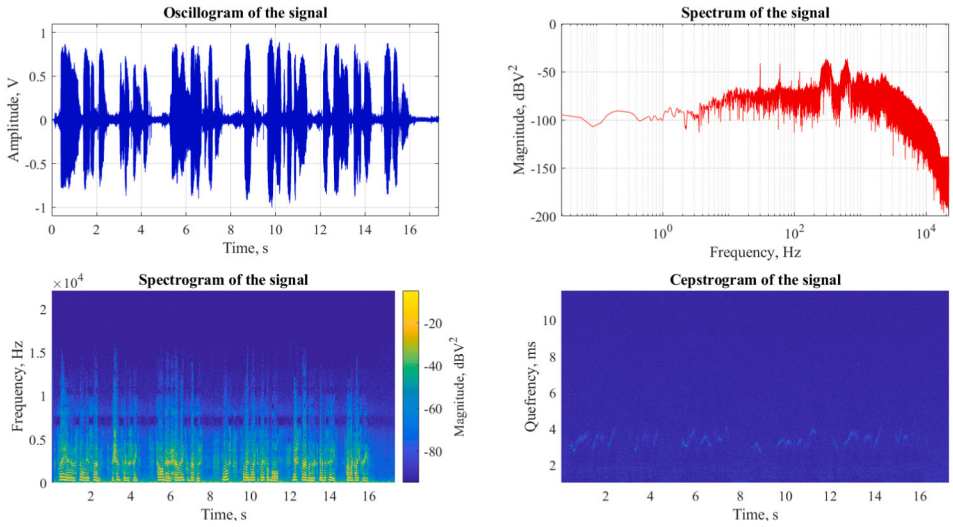


b) PAD men

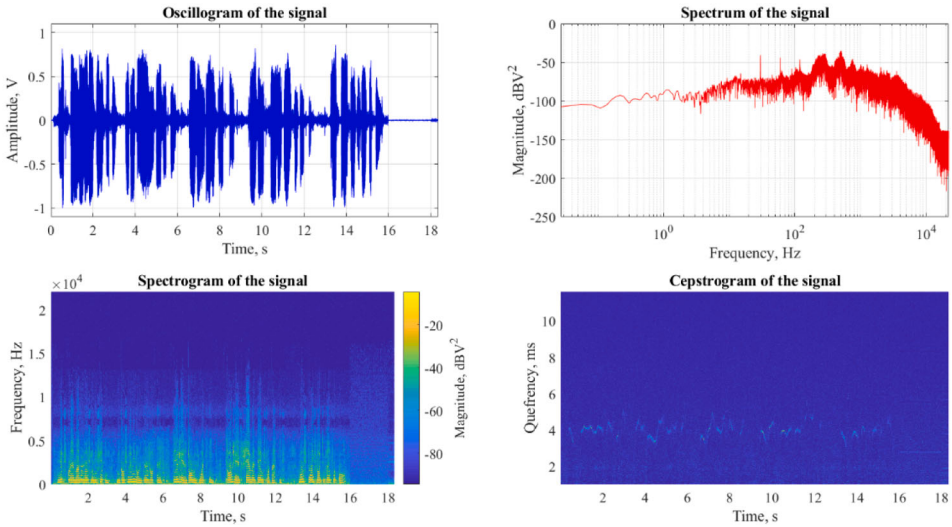


c) HLC women

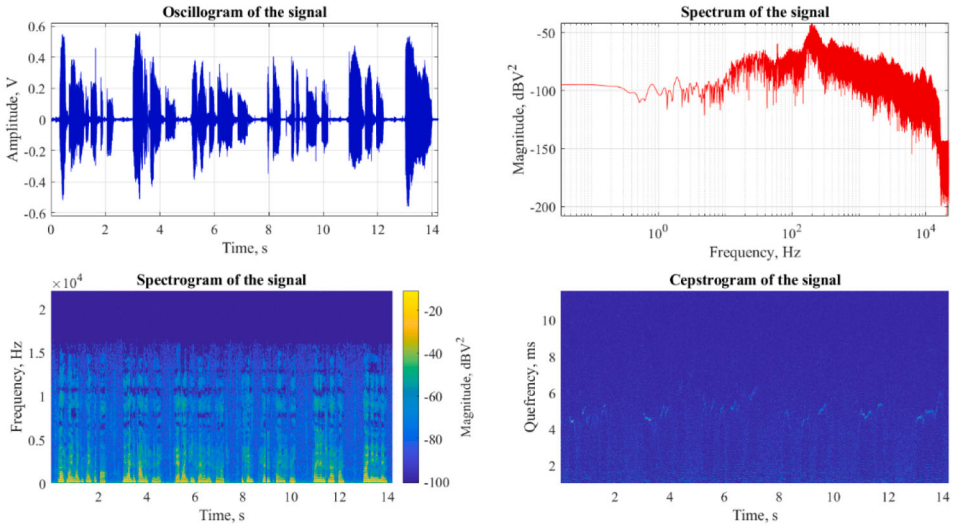
Fig.3. The typical signal and its spectrogram.



d) PAD women



e) HLC child



f) CP child

Fig.3. (*continued*).

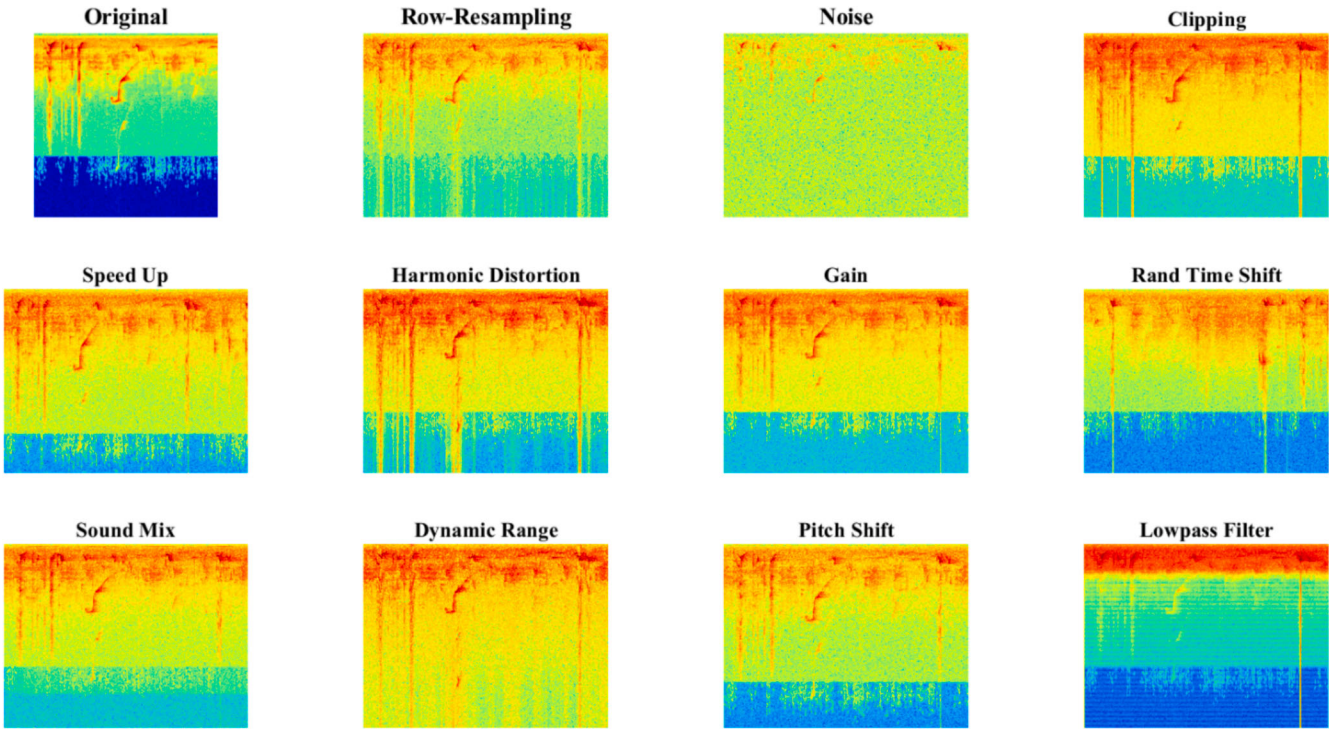


Fig.4. The augmentation process.

mathematically expressed as Eq. (6).

$$\begin{aligned} x_{noise}(t) &= x(t) + \eta(t) \\ S_{noise}(t,f) &= |STFT(x_{noise}(t))| \end{aligned} \tag{6}$$

Where η(t) denotes the noise term.

Clipping: The process of reducing the amplitude of a signal, which leads to the flattening of peaks on the spectrogram, as presented in Eq. (7).

$$\begin{aligned} clipped\,(t) &= \begin{cases} Aifx(t) > A \\ -xAifx(t) < A \\ x(t)otherwise \end{cases} \\ S_{clipped}(t,f) &= \left|STFT(x_{clipped}(t))\right| \end{aligned} \tag{7}$$

Clipping restricts the peaks of the signal to a maximum amplitude of A, thereby creating a 'flat top' effect.

Speed Up: Enhancing the playing speed, resulting in the compression of the spectrogram horizontally and the upward shift of frequencies, which is expressed as Eq. (8).

$$\begin{aligned} x_{speed}(t) &= x(\frac{t}{\beta}) \\ S_{speed}(t,f) &= \left|STFT(x_{speed}(t))\right| \end{aligned} \tag{8}$$

Where β denotes the speeding-up coefficient.

Harmonic Distortion: The addition of harmonic overtones that modify the frequency components of the signal, which can be mathematically expressed as Eq. (9).

$$\begin{aligned} x_{harmonic}(t) &= x(t) + \varphi(x(t)) \\ S_{harmonic}(t,f) &= |STFT(x_{harmonic}(t))| \end{aligned} \tag{9}$$

Where $\varphi(x(t))$ denotes harmonic distortion function.

Gain: The augmentation of the signal's amplitude is reflected by enhanced colors on the spectrogram, which can be mathematically expressed as Eq. (10).

$$\begin{aligned} x_{gain}(t) &= x(t) + G(x(t)) \\ S_{gain}(t,f) &= \left|STFT(x_{gain}(t))\right| \end{aligned} \tag{10}$$

Where gain increases the strength of a signal by G coefficient.

Rand Time Shift: Over time, the signal undergoes random variations which can be depicted as horizontal displacement in the spectrogram. This can be represented mathematically as illustrated in Eq. (11).

$$\begin{aligned} x_{shifted}(t) &= x(t-\tau) \\ S_{shifted}(t,f) &= \left|STFT(x_{shifted}(t))\right| \\ \tau:&\ \text{Shift Factor} \end{aligned} \tag{11}$$

Sound Mix: The process of superimposing or mixing one sound with another, replacing or modifying existing frequencies, is defined mathematically in Eq. (12).

$$\begin{aligned} x_{mixed}(t) &= x(t) + Mix(t) \\ S_{mixed}(t,f) &= |STFT(x_{mixed}(t))| \\ Mix(t):&\ \text{Another Sound} \end{aligned} \tag{12}$$

Dynamic Range: The result of signal expansion or compression's effect on the spectrogram's dynamic range of contrast is presented in Eq.(13).

$$\begin{aligned} x_{comp}(t) &= Comp(x(t)) = sgn(x(t))\cdot|x(t)|^{\gamma} \\ &\begin{cases} Compress\ \text{if}\ \gamma < 1 \\ Expand\ \text{if}\ \gamma > 1 \end{cases} \\ S_{comp}(t,f) &= \left|STFT(x_{comp}(t))\right| \end{aligned} \tag{13}$$

Pitch Shift: Shifting the spectrogram by changing the frequency of the sound while still preserving the time, can be written mathematically as Eq. (14).

$$\begin{aligned} x_{pitch}(t) &= Pich\_Shift(x(t),\lambda) \\ S_{pitch}(t,f) &= \left|STFT(x_{pitch}(t))\right| \end{aligned} \tag{14}$$

The time period is stable, but the fundamental frequency of the signal has been changed which has the impact of λ on its frequency content.

Lowpass Filter: To get rid of the former high-frequency elements,

the lowpass filter is implemented so that the spectrogram would only display the lower frequencies as seen mathematically in Eq. (15).

$$\begin{aligned} x_{lowpass}(t) &= STFT^{-1}(H(f).STFT(x(t)) \\ S_{lowpass}(t,f) &= \left|STFT(x_{lowpass}(t))\right| \\ H(f) &: \text{Filter Transfer Function} \end{aligned} \tag{15}$$

These enhancements are usually used to increase the tolerance of audio-processing systems to a wider environment of sound conditions and thus improve their ability to generalize beyond training data. Each method produces specific changes that can be readily seen in the spectrogram and therefore provide useful feedback concerning the possible benefits of such enhancements in the performance of audio classification or recognition systems.

3.4. Perceptual approach

Section 2 highlighted the challenge of effectively classifying CP and PAD using DNNs due to the limitations posed by short datasets. In order to address this constraint, we suggest employing the CFSL technique, which consists of two stages: the identification of LM points and the classification of CP/PAD.

The suggested CFSL framework utilizes LM points taken from the spectrogram of CP and PAD data to classify them into several groups accurately. The LM spots are acquired by detecting the spots where the power exceeds a specific threshold and subsequently grouped into distinct categories based on their proximity through clustering. The categorization is executed by calculating the distance between the LM point of each signal and the centroid of each cluster and then allocating it to the category with the nearest centroid.

The suggested method for identifying LMs from the spectrograms of HLC, CP, and PAD signals entails the identification of two parameters: time points and thresholds. This LM facilitates the identification of the precise location where the power transitions beyond the threshold at a certain point. The CFSL approach is employed to extract and save the precise coordinates of LM points as LMs. The approach considers both the spectrogram count (λ) and the total spectrogram count (TOT) in the dataset. The LM point is marked by the coordinates (t_n; p_n), where t_n is time and p_n is the time-specific instant when power cuts the threshold.

Such methodology gives a complete way of accumulating important information from the CP or PAD signal spectrogram. Using the LM points clustering process, such patterns and features of the diseases can be highlighted, and they can also help in more accurately classifying these diseases.

3.5. Few-shot learning

Generating separate measuring regions for each LM attribute from a small set of labeled real-world data is a key component of the CFSL process. The CFSL approach allows for the identification of minor variances by utilizing LM merging functions as learners and LM prototypes as markers for every measurement space of the higher-level dimension. The CFSL paradigm improves the base learner's generalization capabilities by integrating many accurate LM learners.

CFSL operates using three types of data: support set (δ), query dataset (D_Q), and training dataset (D_L). In contrast to the training and query sets, which use different label spaces, the support set uses the same label space across all labeled datasets. A tuple (t, p) is used to indicate an LM data point, and t serves as the label. By drawing on the annotated training and support sets, the CFSL method attempts to label the query set.

Standard FSL is a widely used machine learning that focuses on effectively adjusting to novel challenges. During the training process, this method utilizes small subsets of data called episodes. Each episode consists of a selection of classes from the training data, and the data points inside each episode are appropriately labeled. Two sets of marked data points are created: one for use in calculating the prediction error (the query set) and another for use in providing support (the support set). The objective of the CFSL is to optimize the loss within the query set to ensure a consistent sample set across each episode. This strategy is extremely helpful for enhancing the framework's potential to generalize to novel data during the testing phase, especially when there is a scarcity of available data. The common name for training sets utilized in the process is "balanced episodes," with "β shot" denoting the number of support samples per class and "α way" indicating the number of classes in each episode.

The CFSL approach incorporates non-linearly parametric LM merging functions, also known as LM learners EF_{ξ}^{α}. The average observed integration of the data from the support set is used to build one LM prototype $\boldsymbol{CP}_{\beta}^{\alpha}$ for every β LM. By using this method, the model is able to improve its generalizability by obtaining a better representation of each class.

$$\boldsymbol{CP}_{\beta}^{\alpha} = \frac{1}{\Gamma_{\beta}} \sum_{(t_n;p_n)\in\Gamma_{\beta}} EF_{\xi}^{\alpha}(\boldsymbol{t}_n \circ \boldsymbol{\mu}^{\alpha}) \tag{16}$$

The LM number is denoted by the α, and the Hadamard product is represented by the symbol $\circ$. As a result, for each of the three classes (represented by the numbers β and $\left\{\boldsymbol{CP}_{\beta}^{\alpha}\right\}_{\alpha=1}^{N} = 1$), a set of N feature prototypes is used. Furthermore, $\{\boldsymbol{\mu}^{\alpha}\}_{\alpha=1}^{N} = \Gamma$ represents the total number of "N" LMs that were retrieved using the proposed approach of LM extraction. These LMs provide the model with background information that it can use to represent each category better. In Fig. 5, we can see how CFSL compares individual LM learners and LM prototypes to learn about integrating LMs in every dimension and how it rates the importance of each dimension.

CFSL utilizes a technique that assigns significance ratings to global and local LMs for every class with the aim of enhancing interpretability.

- **Local importance grades**

The local importance grades are determined by evaluating the similarity between a query data point and LM prototypes and assigning higher grades to LMs that have a more significant impact on the classification of the query point. The "*local importance grades*" can be mathematically modeled as Eq. (17):

$$w_i^{local}(q) = \frac{1}{d(q, LM_i)} \tag{17}$$

where q denotes the query point, LM_i represents the landmark model prototype for a given class, and $d(q, LM_i)$ denotes the distance (e.g., Euclidean distance) between the query point and the prototype.

- **Global importance grade**

By employing the local grades, CFSL is able to provide precise reasons for each forecast. The model computes the mean distance between the prototype of an LM and the integrating query points to provide a measure of the "*global importance grade*" of the LMs, which is formulated in Eq. (18).

$$w_i^{gobal} = \frac{1}{\frac{1}{N}\sum_{j=1}^{N} d(q_i, LM_i)} \tag{18}$$

- **Integration of local and global grades**

The importance of a particular LM can be expressed as the sum of the local importance grade and the global importance grade. One of the methods of combination points out taking a linear combination of the local grade and the global grade with a specified weight θ, as shown in

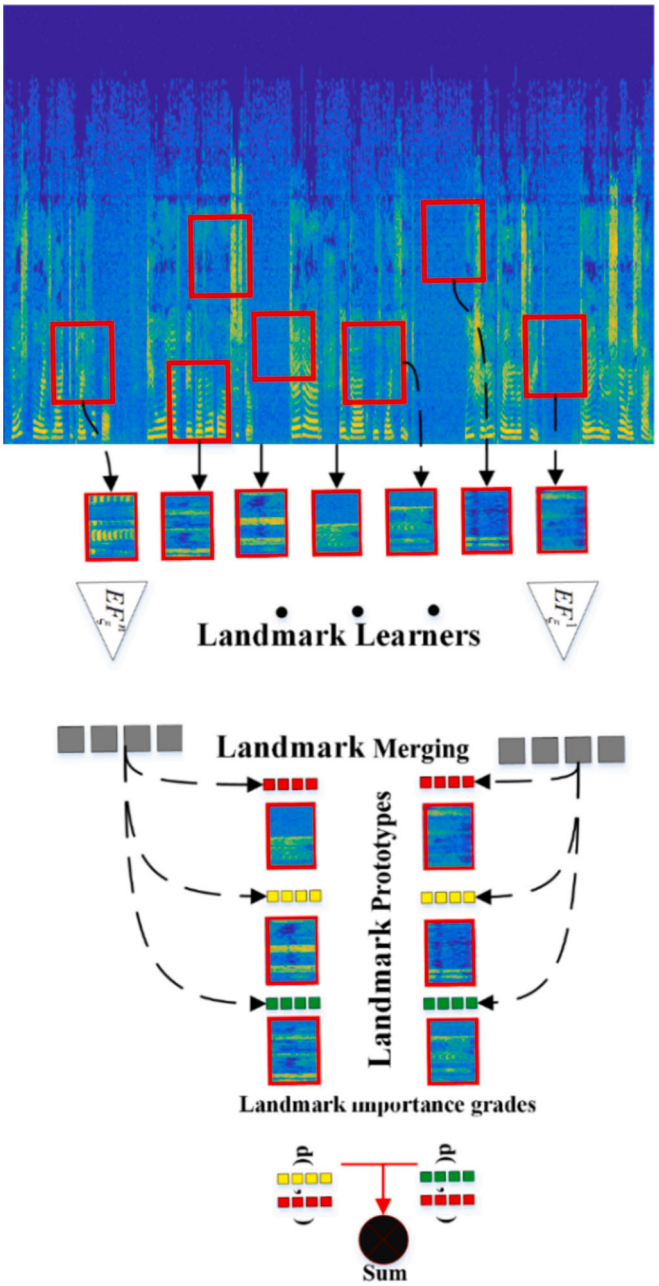


Fig.5. Detailing the process by which the model acquires the ability to integrate LMs.

Eq. (19):

$$w_i(q) = \theta \cdot w_i^{local}(q) + (1-\theta) w_i^{global} \tag{19}$$

- **Ranking query points based on resemblance to a predetermined reference point**

CFSL can rank points according to their similarity to a predetermined reference point (r). This helps to identify instances that are similar to others within or outside the same category. In order to rank the query points qj with respect to a single point reference point r, we compute the distance d(q_j,r) for all query points $j = 1,2,\ldots, N$ and sort the query points in ascending order based on this distance. The ranking function $Rank(q_j)$ for the query point q_j can be defined as Eq. (20).

$$Rank(q_j)\text{argsort}_j(d(q_j, r)) \tag{20}$$

This distance is representative of the distance between query point q_j and a reference point as defined in the overall ranking. This hierarchy classifies query points relative to the referenced location, with lower ranks indicating similarity and higher ranks showing dissimilarity.

- **Interpretability and Insights**

CFSL can put the classification into perspective by assigning rates to each LM and ranking query points in relation to reference points. Local scores, or graduations, present arguments for LMs that were decisively active in the prediction of the outcome under investigation. On the other hand, global grades address the issue of the importance of all LMs in the dataset examined.

This enhances the model's interpretability by allowing for the logical and comprehensible explanation of all predictions, as well as the easy establishment of significant landmarks that characterize a class or a patient's case.

4. Experimentation and discussion

In order to carry out our study, we utilized the CFSL approach on three separate datasets: GPRS, CIEMPIESS, and PC-GITA. Each dataset comprises 25 individual sections or LMs. Nevertheless, a portion of the photos in the databases lacked specific markers. In these cases, we replaced the absent LM with a usual one. Using part coordinates, we created an LM mask that completely encompassed the bounding box of the surrounding area.

4.1. Description of the experiment's configuration

The efficacy of CFSL was evaluated through a comparative analysis of its performance with seven distinct FSL-based models, including SUFSL (Xu et al., 2023); MCFSL (Chen et al., 2021); DCRFSL (Singh & Mazumder, 2022); MAMIFSL (Wang et al., 2022); IFSL (Oreshkin et al., 2018); ProtoNet (Snell et al., 2017); and MKFSL (Tian et al., 2021).

In this study, the evaluation methodology revolves around the most common classification task, which contains two stages: the one that involves a k-way classification with k shots. This task calls for k instances to represent every category in the support set, where k is called the number of shots. The construction of the query had five unlabeled samples, where each sample was taken from one of the classes. In order to assess the best model, a test set of novel classes is used, and the selection of the best model is based on validation accuracy. To ensure consistency, we followed the evaluation protocol outlined in (Luo et al., 2023); and the dataset was partitioned using the same split method, consisting of twenty-five percent validation, twenty-five percent test, and fifty percent baseline sets.

For the implementation, we followed the advice in (Snell et al., 2017) and used the Conv-4 design, which has four layers of convolution and an input dimension of 84 × 84. The Adam optimizer with an initial learning rate of 10^{-3} and a weight decay of 0 is employed for all datasets. In contrast to the 60,000 episodes used to train the 1-sh tasks, we conducted 40,000 episodes of training on the 5-sh tasks, following the methodology described in the literature (Luo et al., 2023).

In order to accelerate the training operation of the CFSL, we distributed the system characteristics among LM learners. More precisely, the convolutional network processes the entire picture $\boldsymbol{t}_n$, resulting in a merged feature $EF^{\alpha}(\boldsymbol{t}_n)$. This merged feature is then used to produce the j-th LM merging $EF^{\alpha}(\boldsymbol{t}_n{}^{\circ}\boldsymbol{\mu}^{\alpha})$. There is minimal performance variation when the mask is put at the start or finish of the procedure. Nevertheless, the latter method dramatically accelerates the duration of training. When there are unannotated parts, we substitute the missing landmark with the archetypal landmark that corresponds to the complete image.

We utilized the implementations derived from RelationNet, IFSL,

MCFSL, ProtoNet, and SUFSL (Luo et al., 2023). We utilized the implementations outlined in the corresponding papers for DCRFSL and MKFSL. We present the average accuracy that was utilized by selecting 600 episodes at random throughout the *meta*-testing or fine-tuning phase.

4.2. Evaluation of results and performance

Table 2 summarizes the results of the analysis of the performance for the PC-GITA, CIEMPIESS, and GPRS datasets, utilizing LMs as previous domain knowledge. CFSL consistently outperforms all comparison classifiers on all datasets. The CFSL exhibits a mean improvement of 4.43 % and 4.50 % in the 5-sh and 1-sh tasks, accordingly, when compared to the most superior baseline systems. CFSL outperforms ProtoNet's scores by 11.033 % and 12.966 % in the 5-sh and 1-sh tasks, respectively, for all datasets.

We performed further analysis to evaluate the influence of using a more intricate Conv-6 framework on CFSL's efficiency. The results of our study confirm that significant improvements in performance are still present. To ascertain that the enhancements in performance are solely attributed to LM learners and not extra supplementary weights, an evaluation between CFSL and a conglomerate of ProtoNets is carried out. In addition, we assessed the efficacy of CFSL when utilizing shared weights for all LMs. According to the data in Table 3, CFSL outperforms the ProtoNet ensemble, even when the weights are shared among LMs. Using standard weights across LMs has a minimal impact on CFSL's efficiency.

Table 2
The outcomes (average ± standard deviation) of the 1-sh and 5-sh challenges for the GPRS, CIEMPIESS, and PC-GITA.

Dataset	Model	5-sh (%)	1-sh (%)
GPRS	ProtoNet (Snell et al., 2017)	83.60 ± 0.68	65.60 ± 0.98
	MAMIFSL (Wang et al., 2022)	84.44 ± 0.67	69.99 ± 0.99
	IFSL (Oreshkin et al., 2018)	84.09 ± 0.68	69.84 ± 1.01
	DCRFSL (Singh & Mazumder, 2022)	89.76 ± 0.68	79.17 ± 0.98
	MKFSL (Tian et al., 2021)	89.23 ± 0.68	79.15 ± 0.99
	SUFSL (Xu et al., 2023)	87.50 ± 0.68	76.60 ± 0.98
	MCFSL (Chen et al., 2021)	89.05 ± 0.67	79.76 ± 0.99
	CFSL (Proposed)	**92.80 ± 0.36**	**80.11 ± 0.97**
CIEMPIESS	ProtoNet (Snell et al., 2017)	76.19 ± 0.88	57.18 ± 1.10
	MAMIFSL (Wang et al., 2022)	84.79 ± 0.88	68.95 ± 1.10
	IFSL (Oreshkin et al., 2018)	83.30 ± 0.88	66.90 ± 1.08
	DCRFSL (Singh & Mazumder, 2022)	84.83 ± 0.87	68.78 ± 1.08
	MKFSL (Tian et al., 2021)	84.12 ± 0.87	68.15 ± 1.08
	SUFSL (Xu et al., 2023)	82.49 ± 0.87	64.28 ± 1.07
	MCFSL (Chen et al., 2021)	83.68 ± 0.87	67.69 ± 1.07
	CFSL (Proposed)	**86.30 ± 0.79**	**69.29 ± 1.04**
PC-GITA	ProtoNet (Snell et al., 2017)	76.19 ± 0.80	59.58 ± 1.02
	MAMIFSL (Wang et al., 2022)	80.09 ± 0.77	64.80 ± 1.01
	IFSL (Oreshkin et al., 2018)	80.99 ± 0.77	65.01 ± 1.01
	DCRFSL (Singh & Mazumder, 2022)	80.46 ± 0.76	64.98 ± 1.00
	MKFSL (Tian et al., 2021)	81.26 ± 0.76	65.07 ± 1.00
	SUFSL (Xu et al., 2023)	87.14 ± 0.76	67.07 ± 1.00
	MCFSL (Chen et al., 2021)	80.87 ± 0.76	64.78 ± 1.00
	CFSL (Proposed)	**90.27 ± 0.72**	**72.63 ± 1.00**

Table 3
Comparative evaluation of ensemble prototypical networks and CFSL models with shared weights across LMs.

Technique	CIEMPIESS		GPRS		PC-GITA	
	1-sh	5-sh	1-sh	5-sh	1-sh	5-sh
PML (Shared weight)	68.2 ± 1.1	85.9 ± 0.61	80.0 ± 0.9	91.9 ± 0.7	69.9 ± 0.8	90.0 ± 0.5
Ensemble ProtoNet	66.2 ± 1.1	84.0 ± 0.68	68.3 ± 1.1	84.7 ± 0.8	63.5 ± 0.8	80.4 ± 0.5
PML	**69.3 ± 1.0**	**86.2 ± 0.60**	**80.0 ± 1.0**	**92.8 ± 0.6**	**72.6 ± 0.8**	**90.3 ± 0.4**

4.3. The impact of LM quantity

Fig. 6 evaluates the impact of the quantity of LMs on the performance of CFSL on GPRS, CIEMPIESS, and PC-GITA. The investigation commences by examining the outcomes of ProtoNet, which uses a solitary reference point in CFSL that encompasses all input dimensions. Subsequently, the quantity of LMs is systematically augmented, and the CFSL model is trained and evaluated using the specified number of LMs. The findings indicate that CFSL consistently enhances performance across all datasets as the number of LMs is augmented.

When a single LM is introduced, ProtoNet's efficiency on CIEMPIESS demonstrates an enhancement of 6 % and 11 % in the 5-sh and 1-sh tasks, respectively. This enhancement is equivalent to a 10-decibel threshold for the entire picture luminance map. CFSL outperforms all models in GPRS by utilizing just 8 LMs, resulting in an enhancement of 8 % and 18 % above ProtoNet in the 1-sh and 5-sh, respectively. In order to showcase the resilience of CFSL against a substantial number of repetitive LMs, we increased the total count of LMs to 50 by gathering all thresholds, leading to a significant number of duplicate associations. According to Fig. 6, the performance improvement of CFSL is minimal while using 50 LMs compared to 25 LMs. These data suggest that CFSL is superior to other approaches, even in scenarios with few LMs, insufficient annotations, and a high number of redundant and overlapping LMs.

The figure provides an illustration of the effectiveness of a model in accuracy for three datasets: CIEMPIESS, GPRS, and PC-GITA. In the graphs, the x-axis demonstrates the number of perceptions, while the y-axis demonstrates the value portrayed in the percentage of the model performance. Two settings were compared in the experiment: a 1-shot scenario and a 5-shot scenario with a black line and diamond markers, respectively, and a blue line and circular markers, respectively. Such settings indicate the minimal number of training examples per class, with the 5-shot situation being less restrictive than the 1-shot case in this respect.

Regarding the CIEMPIESS dataset, it is noted that the performance of the models on perceptions increases with the increase in perceptions. It can be said that the 5-shot model will always dominate the 1-shot model. The former almost always exceeds the latter across all ranges. After approximately 30 perceptions, the 5-shot accuracy climbs to about 95 % and remains at that level, while the 1-shot, in comparison, settles at about 85 % as well but at a different pace. This means that, indeed, more training examples allow a more accurate model to be created, but to an extent, as does the other setting, rendering such a model also to have little return on getting more assets, such as 30 perceptions or raising once the model is trained.

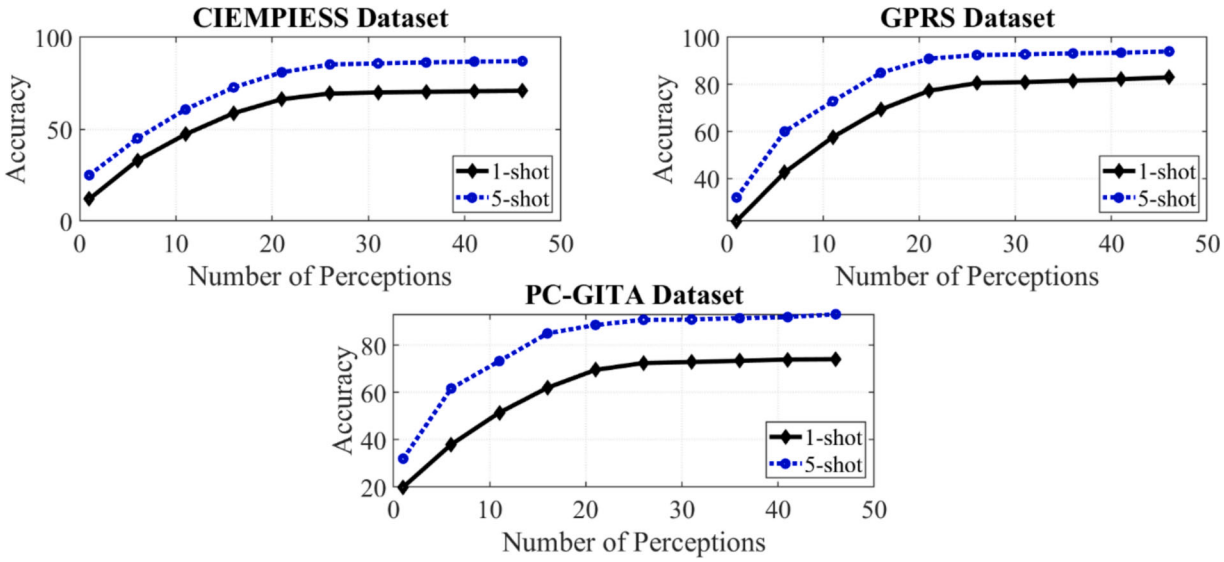


Fig.6. The Impact of LM quantity.

The trend captured in the GPRS dataset is almost similar to the trend observed in previous models and datasets. In general, the 5-shot model has outperformed all others from the onset; however, the accuracy level of the model experiences a steep rise initially, followed by a plateau of sorts. At the 50 perceptions mark, the 5-shot model has nearly 90 % accuracy, while the 1-shot model dies at a sketchy 80 %. While one can observe an increase in the level of both models when more perceptions are tested, the rise in the level of the models is relatively slow after the initial steep rise, particularly in the case of the five-shot model.

In the case of the PC-GITA dataset, however, the situation is quite the opposite; the overall accuracy is generally lower in the initial stages in comparison to the other datasets, but the overall trend remains the same. The general trend is, however, consistent; the 5-shot model managed to overshadow the 1-shot model by fully utilizing its potential in this dataset, too, with over 80 % accuracy after 50 perceptions, while the 1-shot model remained limited at 70 %. The nature of dataset GITA is quite distinct; the rate of improvement is unlike the other datasets; the 5-shot model has improved steadily to higher accuracy levels than the 1-shot model.

To conclude, the figure shows that as the number of perceptions increased, the model's performance was improved on all three data sets. It has also been shown that the marginal benefit from adding the subsequent perception decreases after approximately thirty or forty perceptions. Furthermore, what matters the most is that when taking into account the 5-shot scenario, this situation 5-shot scenario tends to be more beneficial in terms of accuracy than the 1-shot scenario for every single dataset, suggesting that it is beneficial to have a more significant number of images per class when training the model. In spite of these general observations, the dataset CIEMPIESS has the best performance, followed by the GPRS and PC-GITA datasets.

4.4. Ablation study for the distance function

Fig. 6 examines how the number of LMs affects the performance of CFSL on GPRS, CIEMPIESS, and PC-GITA. The study begins with a consideration of the results of ProtoNet, which considers a single point of reference in CFSL that encapsulates all image input dimensions. Thereafter, the number of LMs is gradually increased, and the specified number of LMs is used to train and assess the CFSL model. The findings indicate that CFSL consistently enhances performance across all datasets as the number of LMs is augmented.

When a single LM is incorporated, ProtoNet's performance on CIEMPIESS improves by 6 % and 11 % in the 5-shot and 1-shot, respectively. This enhancement equates to a 10-decibel threshold for the entire picture luminance map. CFSL outperforms all benchmark models in GPRS with only 8 LMs, leading to an enhancement of 18 % and 8 % over ProtoNet in the 5-sh and 1-sh, respectively. In order to showcase the resilience of CFSL against a substantial number of repetitive LMs, we increased the total count of LMs to 50 by gathering all thresholds, leading to a significant number of duplicate associations. According to Fig. 6, the performance improvement of CFSL is minimal when 50 LMs are used compared to 25 LMs. These data suggest that CFSL is superior to other approaches, even in scenarios with few LMs, insufficient annotations, and a high number of redundant and overlapping LMs.

- **Euclidean distance**: The formula for calculating Euclidean distance for **u** and **v** vectors is shown below (Yuan, 2023):

$$d_{Euc}(\boldsymbol{v},\boldsymbol{u}) = \sqrt{\sum_i (v_i - u_i)^2} \tag{21}$$

- **Manhattan distance**: As a measure of distance, the city block distance—sometimes called the Manhattan distance—is calculated by adding together the absolute differences of two points' coordinates, as follows (Chang et al., 2024):

$$d_{Man}(\boldsymbol{v},\boldsymbol{u}) = \sum_i |v_i - u_i| \tag{22}$$

- **Cosine distance**: This kind of distance can be calculated using Eq. (23) (Qian et al., 2024).

$$d_{Cos}(\boldsymbol{v},\boldsymbol{u}) = 1 - \frac{\boldsymbol{v}\cdot\boldsymbol{u}}{\|\boldsymbol{v}\|\|\boldsymbol{u}\|} = \frac{\sum_i u_i v_i}{\sqrt{\sum_i u_i^2}\sqrt{\sum_i v_i^2}} \tag{23}$$

- **Mahalanobis distance**: This distance is calculated by taking the square root of the sum of the squared differences between the two vectors. Specifically, given two vectors **u** and **v**, and the distribution's covariance matrix *Cov*, the Mahalanobis distance between them is computed as (Li et al., 2023):

Table 4
Exploring the effects of various similarity benchmarks on the performance of CFSL models.

distance	CIEMPIESS		GPRS		PC-GITA	
	1-sh	5-sh	1-sh	5-sh	1-sh	5-sh
Cosine	67.1 ± 1.0	84.2 ± 0.7	76.6 + 1.1	91.5 ± 0.7	70.9 ± 0.8	88.1 ± 0.5
Mahalanobis	66.4 ± 1.0	83.2 ± 0.8	76.3 ± 1.2	91.1 ± 0.8	69.9 ± 0.9	68.4 ± 0.5
Manhattan	68.2 ± 1.0	86.1 ± 0.8	79.6 ± 1.0	92.6 ± 0.7	71.4 ± 0.8	90.2 ± 0.5
Euclidean	**69.3 ± 1.0**	**86.2 ± 0.6**	**80.5 ± 1.0**	**92.8 ± 0.6**	**72.6 ± 0.8**	**90.3 ± 0.4**

$$d_{Mah}(\boldsymbol{v},\boldsymbol{u}) = \sqrt{(\boldsymbol{v}-\boldsymbol{u})^{T}\cdot inv(Cov)\cdot(\boldsymbol{v}-\boldsymbol{u})} \tag{24}$$

A study was done to analyze the influence of various distance measures on the efficiency of CFSL. The results of this investigation are displayed in Table 4. The Euclidean distance continuously demonstrated higher accuracy than other measures.

4.5. Ablation study for backbone network

An ablation investigation was performed on the leading network to assess the efficacy of CFSL compared to comparison models (Seow & Qian, 2024). To identify LMs, CIEMPIESS applied a Conv-6 backbone and not a Conv-4 one or part-based identifiers. Results from 600 episodes can be seen in Table 5 along with the average and the variance values. The results indicate that the CFSL is better than all other models' architecture even if complex backbone is used. The effect of the commonly used LM of 20 dB limit and the complete vision LM on the ProtoNet's efficiency increased it by 5.5 % for the 5-sh job and 5.6 % for the 1-sh task.

4.6. Assessing the effect of the location of LMs

Following the guidelines in (Zhang et al., 2018); we employed the auto-encoding method for feature discovery to assess how well CFSL performed when applied to the datasets in the query in order to extract information using visual LMs. We used the setup and default parameters recommended by the authors, choosing a set of 25 features. An LM mask is built using the encoding component, which generates an enclosure that contains the observed characteristics. The parameters for training the auto-encoder were identical to those listed in (Zhang et al., 2018). We performed a comparative examination of the results obtained from 1-shot and 5-shot classification methods on three datasets. This analysis involved the utilization of randomly generated masks as well as human-defined LMs. The statistics presented in Table 6 consist of the AVE and STD calculated from a sample of 600 randomly selected episodes. Fig. 7 displays retrieved features for six distinct photos across all datasets.

The table shows the outcomes of an experiment that compared three different datasets—GPRS, CIEMPIESS, and PC-GITA—on the task of differentiating between human-created masks and randomly selected ones. The table shows the results of each method's accuracy in detecting the masks employing two different protocols, namely 1-sh and 5-sh. As predicted, human-defined masks outperformed arbitrarily produced masks in every single example. Human-defined masks exhibited higher accuracy compared to random masks across all three approaches and both 1-shot and 5-shot procedures. Statistically significant variations were observed in certain circumstances, as evidenced by the standard deviations.

Table 5
Results obtained using the Conv-6 on the CIEMPIESS.

Technique	CIEMPIESS	
	1-sh	5-sh
IFSL	68.0 ± 1.1	80.0 ± 0.8
SUFSL	66.8 ± 1.1	78.9 ± 0.8
MCFSL	66.9 ± 1.1	80.0 ± 0.8
MKFSL	66.6 ± 1.3	84.2 ± 0.9
MAMIFSL	65.5 ± 1.0	81.5 ± 0.7
DCRFSL	67.9 ± 1.0	82.3 ± 0.8
ProtoNet	67.7 ± 1.1	83.1 ± 0.7
CFSL/1 LM	70.0 ± 1.0	84.9 ± 0.7
CFSL	**73.3 ± 1.0**	**88.8 ± 0.6**

Table 6
Comparative analysis of performance: stochastically selected masks versus human-established thresholds.

Technique	GPRS		CIEMPIESS		PC-GITA	
	5-sh	1-sh	5-sh	1-sh	5-sh	1-sh
Chosen stochastically masks	90.1 ± 0.5	71.2 ± 1.0	90.9 ± 0.6	78.3 ± 1.1	87.4 ± 0.9	69.3 ± 1.2
Human-established threshold	**91.0 ± 0.4**	**72.7 ± 0.7**	**930 ± 0.5**	**79.9 ± 0.9**	**90.0 ± 0.8**	**70.0 ± 1.1**

4.7. Computational efficiency

The time complexity of the suggested CFSL model is estimated in this section, and it is contrasted with that of various baseline models. It is crucial to highlight that these estimations are derived from the mathematical properties of the models and may not accurately represent real execution times, as they can differ based on hardware and software configurations.

4.7.1. Time complexity estimations

We analyze CFSL with particular focus on its training and testing phases, outlining their time complexity. Time complexity may be represented using Big O notation, informally referred to as O notation, which describes computational time devoted to a specific algorithm (Yu et al., 2024; Zhang et al., 2024; Cai, 2023). The method allocates computation resources based on the size of its input.

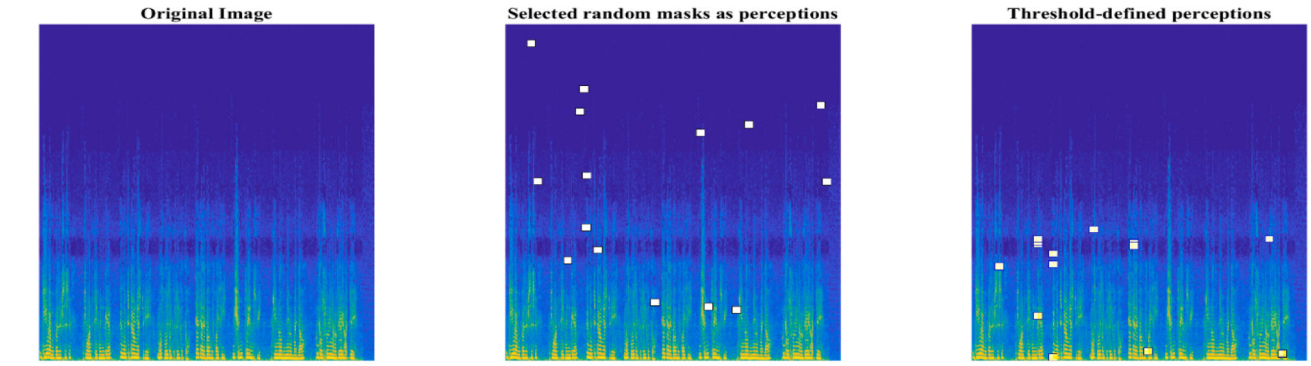
- **Training Time Complexity (O)**: CFSL involves the learning of deep neural networks used in few-shot learning. The number of episodes and the model's learning speed determine the training time. A rough estimation could be due to O (E × N): E for the number of episodes, N for the number of data points per episode.
- **Testing Time Complexity (O)**: In tests, CFSL computes similarity scores between query samples and the support set. Such metrics are complex numbers dependent on the volume of support samples and the size of the feature vectors and can be approximated as O (S × D), where S is the volume of the support samples and D is the size of the feature vector.

4.7.2. Experimental train and test time comparison

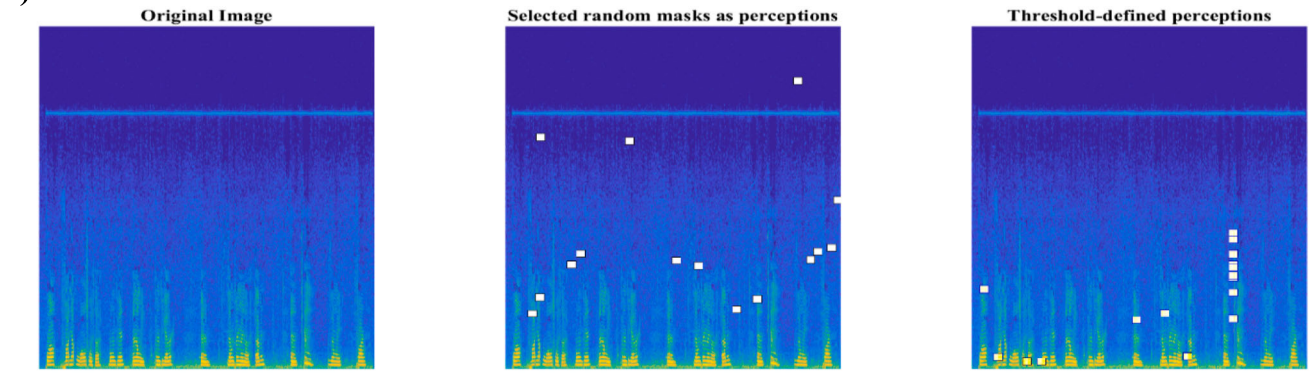
We estimated the average training and testing times of the CFSL model using a standard computer platform together with the benchmark model in order to have a clearer view of resource consumption for the proposed solution. Table 7 displays the average test times for the GPRS, CIEMPIESS, and PC-GITA datasets, aligning them with the deep learning few-shot techniques used against CFSL.

CFSL demonstrates superior time complexity compared to most comparison algorithms except ProtoNet. Despite CFSL's higher accuracy, it operates within a timeframe of 0.012, 0.0023, and 0.0011 s, while ProtoNet achieves slightly lower time complexity at 0.019, 0.0101, and 0.0011 for GPRS, CIEMPIESS, and PC-GITA datasets, respectively.

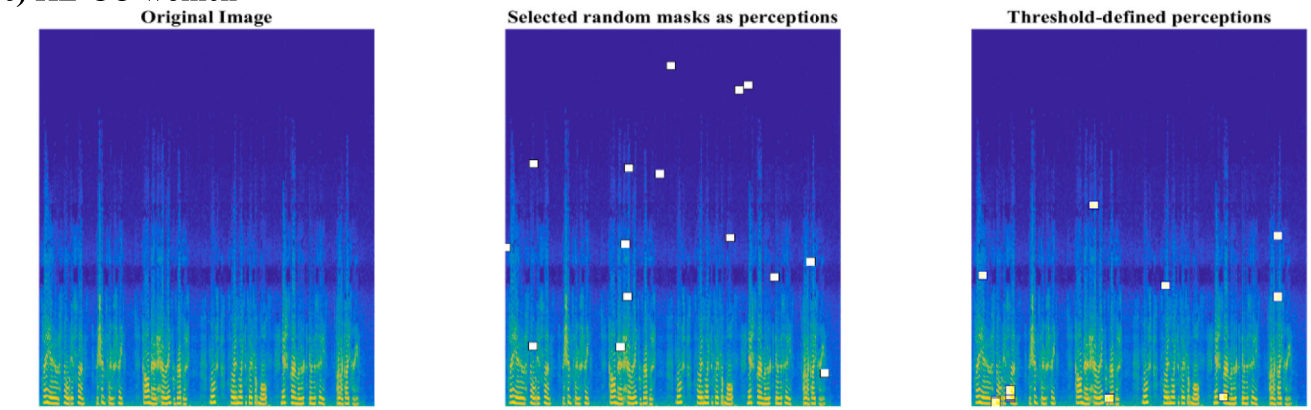
CFSL consistently exhibited competitive inference times across all



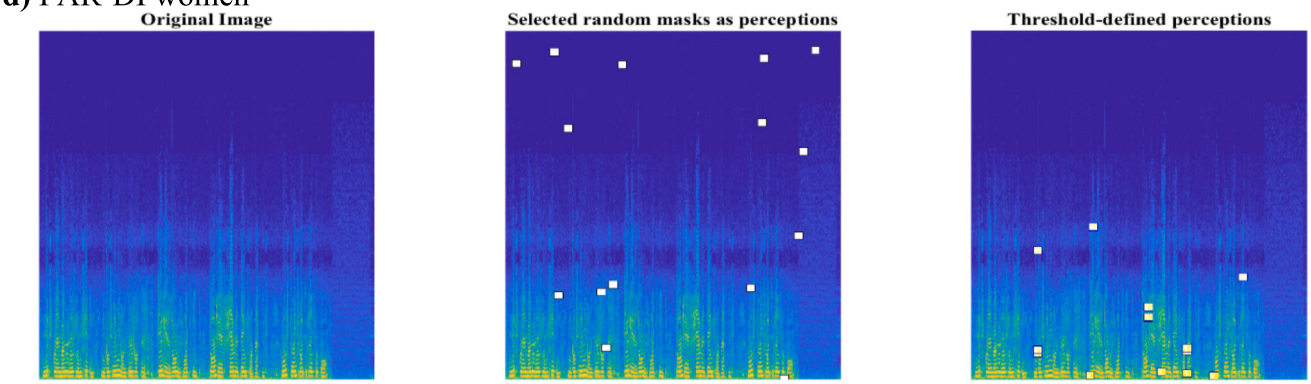
b) PAR-DI men



c) HL-CO women



d) PAR-DI women



e) HL-CO child

Fig.7. Exemplary instances of retrieved perceptions for six distinct images across all datasets.

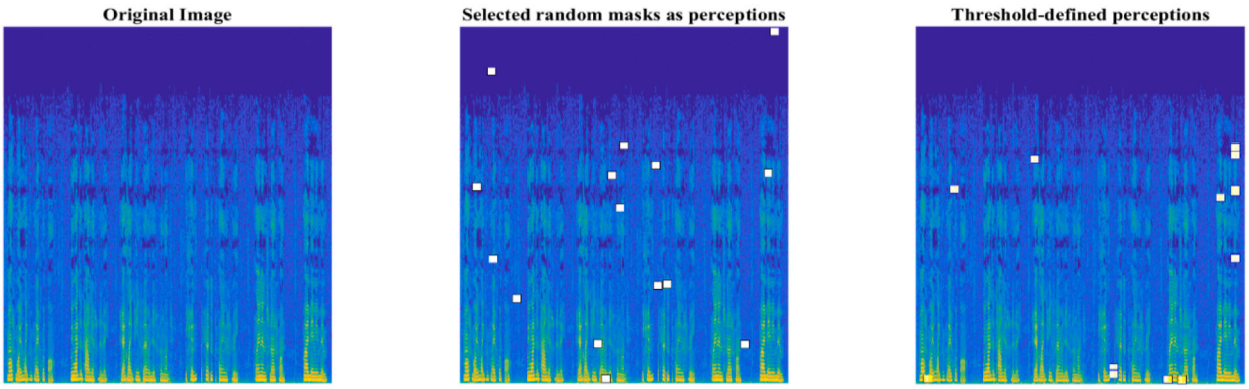


Fig.7. (*continued*).

Table 7
Average inference time for the 5-sh.

Dataset	Model	Training Time	Test Time (second)
GPRS	ProtoNet (Snell et al., 2017)	**39.01 min**	**0.0019**
	IFSL (Oreshkin et al., 2018)	50.11 min	0.0032
	MAMIFSL (Wang et al., 2022)	50.01 min	0.0031
	MKFSL (Tian et al., 2021)	43.21 min	0.0027
	DCRFSL (Singh & Mazumder, 2022)	41.33 min	0.0025
	MCFSL (Chen et al., 2021)	42.41 min	0.0026
	SUFSL (Xu et al., 2023)	40.01 min	0.0025
	CFSL (Ours)	39.81 min	0.0023
CIEMPIESS	ProtoNet (Snell et al., 2017)	**3 h 02 min**	**0.0101**
	IFSL (Oreshkin et al., 2018)	3 h 54 min	0.0152
	MAMIFSL (Wang et al., 2022)	4 h 25 min	0.1701
	MKFSL (Tian et al., 2021)	4 h 11 min	0.0167
	DCRFSL (Singh & Mazumder, 2022)	3 h 47 min	0.0143
	MCFSL (Chen et al., 2021)	3 h 44 min	0.0142
	SUFSL (Xu et al., 2023)	3 h 26 min	0.0121
	CFSL (Ours)	3 h 24 min	0.0120
PC-GITA	ProtoNet (Snell et al., 2017)	**18.12 min**	**0.0011**
	IFSL (Oreshkin et al., 2018)	24.45 min	0.0021
	MAMIFSL (Wang et al., 2022)	22.13 min	0.0017
	MKFSL (Tian et al., 2021)	21.04 min	0.0014
	DCRFSL (Singh & Mazumder, 2022)	20.11 min	0.0014
	MCFSL (Chen et al., 2021)	20.03 min	0.0013
	SUFSL (Xu et al., 2023)	20.03 min	0.0013
	CFSL (Ours)	19.01 min	**00.0011**

datasets while preserving comparatively shorter training durations compared to numerous other models. The results highlight the effectiveness of CFSL in performing inference tasks for few-shot learning on various datasets, indicating its potential for real-time applications where efficiency is paramount. Nevertheless, ProtoNet exhibited marginally reduced inference times in certain instances, highlighting the inherent trade-offs between the duration of training and the efficiency of inference that are characteristic of any model.

4.8. Discussion

This study is limited in scope since only two protocols, that is, 1-shot and 5-shot, have been studied, and the applicability of the findings to other protocols remains unknown. The study further did not probe the variable factors that have the potential to affect the efficacy of the different techniques, such as the size and complexity of the masks or the expertise of the human annotators.

In conclusion, human masks achieve high efficiency, whereas randomly selected masks do not perform as effectively. Other factors, such as resource availability and work demands, may influence the chosen decision-making strategy. We need to conduct further studies to investigate the success determinants of various methodologies and expand the findings to encompass other processes and procedures.

We carried out an extensive evaluation of our CFSL model on three datasets: GPRS, CIEMPIESS, and PC-GITA. The efficiency of CFSL was then compared with that of comparison models, and from the results, CFSL outperformed all the comparison models by a wide margin, as seen in Table 2. CFSL demonstrated a notable average enhancement of 4.50 % and 4.43 % compared to the top competitors in the 1-sh and 5-sh tasks, respectively. Furthermore, CFSL outperformed the most cutting-edge framework, ProtoNet, by a margin of 13.004 % in the 1-sh examination and 10.929 % in the 5-sh challenge.

We evaluated CFSL's efficiency to see how using a more sophisticated Conv-6 basis would affect things. The findings indicated that CFSL's significant enhancements were maintained. Table 3 shows the results of comparing CFSL's efficiency to a set of typical networks. Additionally, we assessed CFSL's performance while using shared weights in every LM. Despite the distribution of weights among LMs, CFSL outperformed the combined set of ProtoNets. Moreover, the efficiency of CFSL was only slightly impacted when employing standard weights across language models.

Our investigations focused on understanding how the performance of CFSL (Context-Free Syntax Learning) varies with changes in the number of LMs on three datasets: GPRS, CIEMPIESS, and PC-GITA. A representation of the results appears in Fig. 6 and illustrates that increasing the number of LMs gives a relatively better CFSL performance for all datasets. Adding a single LM, which represents a 10 dB threshold for the complete LM, improved ProtoNet's performance for CIEMPIESS in 5-sh tasks by 6 % and in 1-sh tasks by 11 %. CFSL demonstrated exceptional performance on the GPRS, surpassing all benchmarks with just 8 LMs. It surpassed ProtoNet by 8 % and 18 % in the 1-sh and 5-sh challenges, respectively. Our CFSL model exhibited exceptional performance compared to the latest models on three distinct datasets, thereby concluding our proposal. According to the results, CFSL shows potential as a framework for classification problems and is robust against changes in the number of LMs and backbone design.

5. Conclusion

The present study developed a new approach for detecting and diagnosing CP and PAD in early stages with the help of CFSL which employs machine-assisted evaluation of paralinguistic aspects of speech. A major strength of the CFSL method is that it enhances the

generalization of models by learning from human-consistent LM features that allow it to integrate takeaway from independent LM learners. Seven benchmark models were utilized to compare CFSL's performance on three famous datasets respectively. CFSL also posted a good average improvement of 4.50 % and 4.43 % over the competition in the 1-sh and 5-sh tasks as compared to the best competitors. Besides, there was an overall improvement over the most advanced framework, ProtoNet, by 13.004 % in the first and 10.929 % in the fifth challenge when tested using a 1-sh method. In conclusion, the application of CFSL provides a method for improving the accuracy of diagnosis of these diseases.

The presented CFSL approach in this work has its own inadequacies that need to be looked into. Paralinguistic features in speech signals are essential to their functionality but may not necessarily be available in a real life situation. Furthermore, CFSL has been evaluated with only three data sets, which may be insufficient to reach any conclusive statements regarding the efficacy of the model in other environments. The future research can impose itself to search the limits of generalization of CFSL to more comprehensive and more heterogeneous datasets and to check its effectiveness with various sorts and modalities of information as well. As well, there is a need to improve the comprehensibility of the representations acquired in the course of CFSL training for such will be critical to comprehending the performance of the model in the first place.

Lastly, it could be useful to combine CFSL with other already known methods, for example deep learning or transfer learning, in order to increase the diagnostic possibilities of the first one – this direction can be turned to in further research.

CRediT authorship contribution statement

Pei Yin: Data curation, Visualization, Writing – review & editing, Investigation. **Junjie Song:** Formal analysis, Validation, Visualization, Software. **Yassine Bouteraa:** Conceptualization, Supervision, Project administration. **Leren Qian:** Formal analysis, Resources. **Diego Martín:** Software, Writing – review & editing. **Mohammad Khishe:** Validation, Writing – review & editing.

Funding

This work was supported by Humanities and Social Science Foundation of Ministry of Education (23YJCZH281), Shanghai Philosophy and Social Science Planning Project (2022ZGL010), Key Lab of Information Network Security, Ministry of Public Security) (C23600). The financial support is gratefully acknowledged. This study is supported via funding from Prince sattam bin Abdulaziz University project number (PSAU/2024/R/1445).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eswa.2024.125713.

Data availability

The source code of the models can be available via https://www.mathworks.com/matlabcentral/fileexchange/174190-cfsl-framework. The datasets used in the paper can be available via: **PC-GITA Dataset:** https://www.kaggle.com/datasets/dipayanbiswas/parkinsons-disease-speech-signal-features. **CIEMPIESS Dataset**: https://ciempiess.org/downloads. **GPRS Dataset:** https://ultrasuite.github.io/data/cleft/.

References

Zhuo, Y., et al. (2024). TGF-β1 mediates hypoxia-preconditioned olfactory mucosa mesenchymal stem cells improved neural functional recovery in Parkinson's disease models and patients. *Military Medical Research, 11*(1), 48.

Song, W., Wang, X., Guo, Y., Li, S., Xia, B., & Hao, A. (2024). CenterFormer: A novel cluster center enhanced transformer for unconstrained dental plaque segmentation. *IEEE Transactions on Multimedia*.

Zou, G.-J., et al. (2024). Microglial activation in the medial prefrontal cortex after remote fear recall participates in the regulation of auditory fear extinction. *European Journal of Pharmacology*, Article 176759.

Xu, B., Martín, D., Khishe, M., & Boostani, R. (2022). COVID-19 diagnosis using chest CT scans and deep convolutional neural networks evolved by IP-based sine-cosine algorithm. *Medical & Biological Engineering & Computing, 60*(10), 2931–2949.

Wang, Y., et al. (2024). Tumor cell-targeting and tumor microenvironment-responsive nanoplatforms for the multimodal imaging-guided photodynamic/photothermal/chemodynamic treatment of cervical cancer. *International Journal of Nanomedicine*, 5837–5858.

Sun, T., Lv, J., Zhao, X., Li, W., Zhang, Z., & Nie, L. (2023). In vivo liver function reserve assessments in alcoholic liver disease by scalable photoacoustic imaging. *Photoacoustics, 34*, Article 100569.

He, J., et al. (2024). Graveoline attenuates D-GalN/LPS-induced acute liver injury via inhibition of JAK1/STAT3 signaling pathway. *Biomedicine & Pharmacotherapy, 177*, Article 117163.

Hui, Z., Lai-Fa, W., Xue-Qin, W., Ling, D., Bin-Sheng, H., & Li, J.-M. (2024). Mechanisms and therapeutic potential of chinonin in nervous system diseases. *Journal of Asian Natural Products Research*, 1–16.

Cai, C., et al. (2023). Improved deep convolutional neural networks using chimp optimization algorithm for Covid19 diagnosis from the X-ray images. *Expert Systems with Applications, 213*, Article 119206.

Borlea, I.-D., Precup, R.-E., Dragan, F., & Borlea, A.-B. (2017). Centroid update approach to K-means clustering. *Advances in Electrical and Computer Engineering, 17*(4).

Zhao, Y., Yu, Y., Chen, R., Li, Y., & Tian, A. (2021). Toward transparent and accountable privacy-preserving data classification. *IEEE Network, 35*(4), 184–189.

Borlea, I.-D., Precup, R.-E., & Borlea, A.-B. (2022). Improvement of K-means cluster quality by post processing resulted clusters. *Procedia Computer Science, 199*, 63–70.

Jing, N. (2024). Neural network-based pattern recognition in the framework of edge computing. *Science and Technology, 27*(1), 106–119.

Mohammadzadeh, A., Taghavifar, H., Zhang, Y., & Zhang, W. (2024). A fast nonsingleton type-3 fuzzy predictive controller for nonholonomic robots under sensor and actuator faults and measurement errors. *IEEE Transactions on Systems, Man, and Cybernetics System*.

Mohammadzadeh, A., Zhang, C., Alattas, K. A., El-Sousy, F. F. M., & Vu, M. T. (2023). Fourier-based type-2 fuzzy neural network: Simple and effective for high dimensional problems. *Neurocomputing, 547*, Article 126316.

Mohammadzadeh, A., Taghavifar, H., Zhang, C., Alattas, K. A., Liu, J., & Vu, M. T. (2024). A non-linear fractional-order type-3 fuzzy control for enhanced path-tracking performance of autonomous cars. *IET Control Theory and Applications, 18*(1), 40–54.

Yan, S.-R., Guo, W., Mohammadzadeh, A., & Rathinasamy, S. (2023). Optimal deep learning control for modernized microgrids. *Applied Intelligence, 53*(12), 15638–15655.

Sun, L., Zhang, M., Wang, B., & Tiwari, P. (2023). Few-shot class-incremental learning for medical time series classification. *IEEE Journal of Biomedical and Health Informatics*.

Latif, J., Xiao, C., Imran, A., & Tu, S. (2019). Medical imaging using machine learning and deep learning algorithms: a review, in 2019 2nd International conference on computing, mathematics and engineering technologies (iCoMET), IEEE, pp. 1–5.

Yao, D., Chi, W., & Khishe, M. (2022). Parkinson's disease and cleft lip and palate of pathological speech diagnosis using deep convolutional neural networks evolved by IPWOA. *Applied Acoustics, 199*, Article 109003.

Mei, J., Desrosiers, C., & Frasnelli, J. (2021). Machine learning for the diagnosis of Parkinson's disease: A review of literature. *Frontiers in Aging Neuroscience, 13*, Article 633752.

Loh, H. W., et al. (2021). Application of deep learning models for automated identification of Parkinson's disease: A review (2011–2021). *Sensors, 21*(21), 7034.

Oh, S. L., et al. (2020). A deep learning approach for Parkinson's disease diagnosis from EEG signals. *Neural Computing and Applications, 32*, 10927–10933.

Viñas, M. J., et al. (2022). Craniofacial and three-dimensional palatal analysis in cleft lip and palate patients treated in Spain. *Scientific Reports, 12*(1), 18837.

Xu, X., Niu, S., Wang, Z., Li, D., Yang, H., & Du, W. (2022). Client selection based weighted federated few-shot learning. *Applied Soft Computing, 128*, Article 109488.

Wang, D., Yu, J., Wu, X., Sun, L., Liu, X., & Meng, H. (2021). Improved end-to-end dysarthric speech recognition via meta-learning based model re-initialization, in 2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP), IEEE, pp. 1–5.

Zhang, X., Ma, J., Li, Y., Wang, P., & Liu, Y. (2021). Few-shot learning of Parkinson's disease speech data with optimal convolution sparse kernel transfer learning. *Biomedical Signal Processing and Control, 69*, Article 102850.

Ngo, Q. C., Motin, M. A., Pah, N. D., Drotár, P., Kempster, P., & Kumar, D. (2022). Computerized analysis of speech and voice for Parkinson's disease: A systematic review. *Computer Methods and Programs in Biomedicine*, Article 107133.

Hossain, M. S., & Shorfuzzaman, M. (2023). MetaParkinson: A cyber-physical deep meta-learning framework for n-shot diagnosis and monitoring of Parkinson's patients. *IEEE Systems Journal*.

Aubreville, M., *et al.*, (2023). Few Shot Learning for the Classification of Confocal Laser Endomicroscopy Images of Head and Neck Tumors," *arXiv Prepr. arXiv2311.07216*.
Wurfbain, L. F., et al. (2023). Diagnostic gene panel testing in (non)-syndromic patients with cleft lip, alveolus and/or palate in the Netherlands. *Molecular Syndromology, 14*(4), 270–282.
van der Hoek-Snieders, H. E. M., van den Heuvel, A. J. M. L., van Os-Medendorp, H., & Kamalski, D. M. A. (2020). Diagnostic accuracy of fetal MRI to detect cleft palate: A meta-analysis. *European Journal of Pediatrics, 179*, 29–38.
Song, Y., Wang, T., Cai, P., Mondal, S. K., & Sahoo, J. P. (2023). A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys*.
Lu, S., et al. (2023). Soft tissue feature tracking based on deepmatching network. *CMES-Computer Modeling in Engineering & Sciences, 136*(1).
Li, H., Huang, Q., Huang, J., & Susilo, W. (2022). Public-key authenticated encryption with keyword search supporting constant trapdoor generation and fast search. *IEEE Transactions on Information Forensics and Security, 18*, 396–410.
Hernández-Mena, C. D., & Herrera-Camacho, J. A. (2014). CIEMPIESS: A new open-sourced mexican spanish radio corpus. *LREC*, 371–375.
Orozco-Arroyave, J. R., Vargas-Bonilla, J. F., Vásquez-Correa, J. C., Castellanos-Domínguez, C. G., & Nöth, E. (2016). Automatic detection of hypernasal speech of children with cleft lip and palate from spanish vowels and words using classical measures and nonlinear analysis. *Revista Facultad de Ingenieria-Universidad de Antioquia, 80*, 109–123.
Orozco-Arroyave, J. R., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Gonzalez-Rátiva, M. C., & Nöth, E. (2014). New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. *LREC*, 342–347.
Takens, F. (1981), Detecting strange attractors in turbulence, in Dynamical systems and turbulence, Warwick 1980, Springer, pp. 366–381.
Covell, M., & Richardson, J. (1991). A new, efficient structure for the short-time Fourier transform, with an application in code-division sonar imaging, in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, IEEE Computer Society, pp. 2041–2044.
Park, D.S., *et al.*, (2019). Specaugment: A simple data augmentation method for automatic speech recognition, *arXiv Prepr. arXiv1904.08779*.
Xu, X., Wang, Z., Chi, Z., Yang, H., & Du, W. (2023). Complementary features based prototype self-updating for few-shot learning. *Expert Systems with Applications, 214*, Article 119067.
Chen, X., Yao, L., Zhou, T., Dong, J., & Zhang, Y. (2021). Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images. *Pattern Recognition, 113*, Article 107826.
Singh, P., & Mazumder, P. (2022). Dual class representation learning for few-shot image classification. *Knowledge-Based Syst., 238*, Article 107840.
Wang, Z., Ma, P., Chi, Z., Li, D., Yang, H., & Du, W. (2022). Multi-attention mutual information distributed framework for few-shot learning. *Expert Systems with Applications, 202*, Article 117062.
Oreshkin, B., Rodríguez López, P., & Lacoste, A. (2018). Tadam: Task dependent adaptive metric for improved few-shot learning. *Advances in Neural Information Processing Systems, 31*.
Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems, 30*.
Tian, P., Li, W., & Gao, Y. (2021). Consistent meta-regularization for better meta-knowledge in few-shot learning. *IEEE Transactions on Neural Networks and Learning Systems, 33*(12), 7277–7288.
Luo, X., Wu, H., Zhang, J., Gao, L., Xu, J., & Song, J. (2023). A closer look at few-shot classification again, *arXiv Prepr. arXiv2301.12246*.
Yuan, J., et al. (2023). Biomimetic peptide dynamic hydrogel inspired by humanized defensin nanonets as the wound-healing gel coating. *Chemical Engineering Journal*, Article 144266.
Chang, L., Li, M., Qian, L., & de Oliveira, G. G. (2024). Developed multi-objective honey badger optimizer: Application to optimize proton exchange membrane fuel cells-based combined cooling, heating, and power system. *International Journal of Hydrogen Energy, 50*, 592–605.
Qian, L., Bai, J., Huang, Y., Zeebaree, D. Q., Saffari, A., & Zebari, D. A. (2024). Breast cancer diagnosis using evolving deep convolutional neural network based on hybrid extreme learning machine technique and improved chimp optimization algorithm, *Biomed. Signal Process. Control*, vol. 87, p. 105492, 2024, [Online]. Available: https://doi.org/10.1016/j.bspc.2023.105492.
Li, X., Khishe, M., & Qian, L. (2023). Evolving deep gated recurrent unit using improved marine predator algorithm for profit prediction based on financial accounting information system. *Complex & Intelligent Systems*, 1–17. https://doi.org/10.1007/s40747-023-01183-4
Seow, M.-J., & Qian, L. (2024). Knowledge augmented intelligence using large language models for advanced data analytics. *SPE Eastern Regional Meeting SPE*. p. D021S001R003.
Zhang, Y., Guo, Y., Jin, Y., Luo, Y., He, Z., & Lee, H. (2018). Unsupervised discovery of object landmarks as structural representations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2694–2703).
Yu, M., Khishe, M., Qian, L., Martín, D., Abualigah, L., & Ghazal, T. M. (2024). Quantum chimp optimization algorithm: A Novel integration of quantum mechanics into the chimp optimization framework for enhanced performance. *Journal of Artificial Intelligence and Soft Computing Research, 14*(4), 321–359.
Zhang, Z., Khishe, M., Qian, L., Martín, D., Abualigah, L., & Ghazal, T. M. (2024). Evolving chimp optimization algorithm using quantum mechanism for engineering applications: A case study on fire detection. *Journal of Computational Design and Engineering, 11*(5), 143–163.