Notion Meta-Learner: A Technique for Few-Shot Learning in Music Genre Recognition

Jinhong Shi [1], Francisco Hernando-Gallego[2], Diego Martín[2], Mohammad Khishe[3*,4]
[1]Conservatory of Music, Weinan Normal University, Weinan, Shaanxi Province, China, shijinhong-shi@outlook.com
[2]Department of Computer Science Escuela de Ingeniería Informática de Segovia, Universidad de Valladolid, Segovia, Spain, fhernando@uva.es, diego.martin.andres@uva.es
[3*]Applied Science Research Center, Applied Science Private University, Amman, Jordan,
[4]Jadara University Research Center, Jadara University, Irbid, Jordan
m_khishe@alumni.iust.ac.ir, https://orcid.org/ 0000-0002-1024-8822

Abstract: This paper presents the notion of meta-learner (NML), an innovative meta-learning methodology designed to enhance the performance of few-shot learning (FSL) regarding the recognition of music genres. Current FSL techniques frequently encounter difficulties due to the absence of organized representations and low capacity for generalization, which impede their efficacy in practical scenarios. The NML meta-learner overcomes these obstacles by acquiring the ability to learn across notion dimensions that humans can understand, thus improving its capacity for generalization and interpretability. Instead of gaining knowledge in a combined and disorganized metric space, the notion meta-learner acquires knowledge by mapping high-level notions into partially organized metric spaces. This technique allows for the efficient integration of several notion learners. We assessed the performance of NMLFSL by utilizing the GTZAN dataset and comparing employing seven different benchmarks. The experimental outcomes show that the NML performs superior to current FSL approaches in tasks that include recognizing music genres with only one or five examples, thereby demonstrating its potential to improve the current state of the art in this field. In addition, ablation experiments assess the influence of essential variables, offering valuable information about the effectiveness of the suggested method. NMLFSL is a notable advancement in using meta-learning to enhance the reliability and precision of music genre recognition (MGR) systems.

Keywords: Notion meta-learner; music genre recognition; high-level notions; few-shot learning.

1. Introduction

MGR is an evolving area that combines the fields of audio signal processing and machine learning. Conventional methods of categorizing genres have depended on extensive datasets to train models capable of reliably differentiating between various forms of music [1]. Nevertheless, the emergence of FSL has expanded the potential for identifying musical genres with restricted data. FSL is a method that seeks to train models to generate precise predictions with only a limited number of instances, similar to how humans may learn new notions with minimum information [2].

Within MGR, FSL offers a distinct chance to categorize a broader range of musical instruments and emotional recognition without requiring large datasets [3,4]. Utilizing hierarchical structures in musical instrument classes can improve the performance of FSL models. By leveraging the connections between instruments and using a hierarchical loss function, it is feasible to enhance the accuracy of classifying instruments that have limited representation in training datasets [5]. This technique emulates the cognitive process of humans categorizing items by identifying common attributes and hierarchical connections.

Furthermore, the principles of FSL have been effectively implemented in visual perception [6]. In this context, models have been created to acquire knowledge about new categories based on a limited number of examples while still retaining previously learned information. Adapting and retaining information is essential for developing resilient systems that consistently acquire knowledge and progress [7,8]. Applying these principles to the auditory domain, particularly in the context of MGR, has the potential to bring about substantial progress in the discipline [9].

The problem comes in developing systems that possess the capability to not only discern novel musical genres based on a small number of instances but also maintain the capacity to recognize genres that have been previously learned [10]. By using advanced techniques like attention-based classification weight generators and cosine similarity functions for feature representation, FSL models can attain exceptional accuracy in both novel and base categories [11].

To summarize, utilizing FSL in music genre detection shows significant potential for enhancing the capabilities and effectiveness of categorization systems. By using hierarchical structures and leveraging

improvements in visual learning (like game learning) [12–14], we can facilitate the development of groundbreaking systems for recognizing musical genres even when data availability is restricted.

This paper is structured into multiple sections, forming the organizational framework. Section 2 begins by examining relevant literature, which helps to establish the necessary background for the following discussion by digging into related topics. Section 3 provides a detailed explanation of the complexities of the suggested methodology, presenting a thorough summary of the approach used in this research. Subsequently, Section 4 thoroughly chronicles the experimental procedures conducted, providing comprehensive information on the techniques utilized and the results obtained. In Section 5, the conclusions and suggested areas for additional investigation are derived by combining the findings and insights from previous parts. This paper aims to present a well-organized narrative that guides the reader through the study's progression and contributes to the broader discussion in the field.

2. Related works

The discipline of FSL has gained significant attention as researchers aim to develop machine learning models capable of making generalizations from a limited number of instances. FSL is especially pertinent in the field of MGR, where the lack of available data might provide a barrier. This paper provides a concise overview of the latest progress in FSL, explicitly highlighting its applications in the domains of music and related fields.

In the field of music, FSL has been utilized to recognize different musical instruments. A study utilizes hierarchical structures to enhance the accuracy of classifying a broader range of musical instruments, even when there are limited samples available [15]. The researchers improved the accuracy of classifying unfamiliar instrument classes by utilizing a hierarchical loss function and hierarchically aggregating prototypes, which closely align with the predetermined structure of a musical instrument hierarchy.

The principle of FSL is not limited to music but may also be applied to other disciplines, including visual learning. An innovative method in this field includes utilizing an attention-based few-shot classification weight generator and the modification of a ConvNet model's classifier to incorporate cosine similarity between feature representations and classification weight vectors. This approach not

only enhances the ability to identify new categories but also maintains high accuracy while dealing with familiar categories, which is a common difficulty in FSL [16].

Video-based few-shot action recognition is an exciting topic. Research focuses on the difficulties of temporal misalignment and data scarcity. It proposes a particular framework for assessing few-shot action recognition algorithms, an implicit sequence-alignment technique, and an advanced loss function for optimizing pair similarity when data is restricted. This methodology employs long short-term memory (LSTM) after 3D convolutional layers to model and align sequences. Its effectiveness is demonstrated through comprehensive testing [17].

Attribute-guided feature learning has been suggested as a technique to enhance few-shot picture recognition. This method utilizes attribute-related representations and a multi-task learning framework to develop linkages between training and novel categories. As a result, it reduces sensitivity to novel categories and enhances performance [18].

When applied to real-world scenarios, FSL encounters difficulties due to the presence of class distributions that have heavy tails and scenes that are congested. In response to this issue, researchers have implemented parameter-free enhancements, including enhanced training methods, innovative architectures that first identify the location of objects before classifying them, and expansions of the feature space using bilinear pooling. These enhancements have demonstrated the ability to increase the precision of cutting-edge models on challenging benchmarks by two-fold [18].

Visual tempo has been employed in the context of few-shot action recognition. A novel framework, visual tempo contrastive learning (VTCL), has been introduced. This framework consists of two main components: a visual tempo encoding (VTE) module and a visual tempo contrastive encoding (VTCE) loss. This framework aims to improve the distinguishability of visual tempo encoding vectors. This technique has demonstrated encouraging outcomes on multiple few-shot action recognition datasets [19].

Furthermore, the application of contextual cueing in FSL has been investigated for item detection in intricate environments. This method enhances classification accuracy by utilizing scene context semantics and a class-conditioned context attention module (CCAM) to assign weights to the most

significant context elements and establish a connection between visual class representations and context semantics [20].

To summarize, FSL is a rapidly advancing discipline with many applications. Various advanced techniques have been devised to address the difficulties encountered in FSL for MGR and other related tasks. These include incorporating hierarchical structures, attention mechanisms, temporal alignment, attribute guidance, contextual cueing, and visual tempo contrastive learning [21].

The suggested method aims to rectify deficiencies and research voids seen in current FSL techniques for music genre detection.

1. The problem with current meta-learning methods is that they often learn intricate representations across labeled tasks without incorporating any organized structure. The lack of organization might result in inefficiencies in the generalization process. The suggested NML employs a systematic methodology to learn mappings of high-level notions into semi-structured metric spaces. The organized representation improves the generalization ability by offering a more understandable foundation for learning.

2. Limited interpretability: Numerous current techniques lack interpretability in their acquired representations, posing a challenge in comprehending the model's decision-making process. NML improves interpretability by gaining knowledge through human-interpretable notion dimensions. This capability allows for a more comprehensive grasp of the essential aspects that contribute to MGR.

3. Current meta-learning systems frequently depend excessively on unstructured metric spaces, which may inadequately reflect the intrinsic links between various ideas or classes. NML overcomes this constraint by acquiring mappings into semi-structured metric spaces, enabling more intricate representations that accurately depict the connections between elements pertinent to music genre identification.

4. The absence of a well-defined structure in acquired representations might result in inefficiencies in generalization, especially when working with a limited amount of data. NML enhances the generalization capability by efficiently merging the results of separate notion

learners. FSL allows for more precise and reliable identification of music genres, even when only a few training samples are available.

Motivation and Main Contributions of the Paper: The suggested method is motivated by acknowledging the constraints of current FSL algorithms in the field of MGR. The authors recognize the necessity of structured representations that improve the ability to interpret and generalize, especially when labeled data are scarce.

The primary innovation of the study resides in introducing the NML, a pioneering meta-learning technique developed expressly for MGR. NML overcomes the limitations of current approaches by acquiring knowledge through idea dimensions that humans can easily understand and by integrating structured representations into the learning process. Through this approach, NML strengthens the model's capacity to make generalizations based on limited samples and improves its interpretability. As a result, it pushes the boundaries of FSL in MGR, advancing the current state-of-the-art.

3. Proposed methodology

In the standard MGR process, there are usually two primary stages: feature extraction (NML extraction) and classification. Although deep neural networks (DNNs) are commonly used in MGR, the scarcity of samples poses a significant obstacle. To tackle this difficulty, we propose implementing the NMLFSL model, designed to enhance the accuracy of categorization. The whole block diagram of the proposed method is shown in Fig. 1.

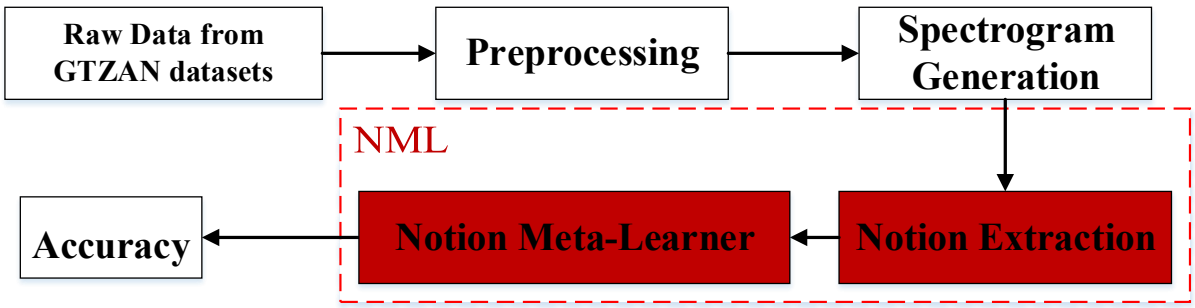


Fig. 1 The whole block diagram of the proposed method

Two primary terms need to be defined first: notion dimensions and partially organized metric spaces:

- **Notion dimensions** describe high-level feature capture, which contains essential semantic information for meta-learning in music genre recognition. These dimensions act as abstract,

explainable features that aid learning by classifying musical features into categories. Unlike conventional feature representations, notion dimensions explain such systems, which include a variation of timbre, rhythmic sophistication, and harmonic sophistication. This allows the model to classify music in a more human-like manner, reasoning across different genres and exploiting commonalities like the sophisticated structure of multi-genre music classification.

- **Partially organized metric spaces** are feature spaces that are at least partially hierarchically structured within themselves, where the distances between elements are preserved as relationships of category differences. Partially organized metric spaces represent a middle ground between fully unstructured metric spaces, which deal with randomly scattered data points, and structured spaces that impose strict hierarchy. In such spaces, closely related genres like jazz and blues are placed next to each other, whereas well-separated distinct genres are ensured to be distinct. Such an organization aids the meta-learner's performance by improving classification based on similarities and the generalization capabilities of the model. A visual representation of this structure has been added to the revised manuscript to fulfill this purpose.

3.1 GTZAN datasets

The GTZAN dataset is widely recognized as a standard in the field of automatic MGR [22]. It has been extensively used to evaluate the effectiveness of algorithms for genre-based music classification. The GTZAN dataset has been a crucial resource for studying MGR for a considerable period. However, it has only recently been thoroughly examined for its composition and integrity, revealing issues such as duplications, incorrect labels, and distortions [23].

Feature extraction plays a crucial role in audio retrieval systems, particularly those that utilize the GTZAN dataset. An advantageous attribute for this objective is the utilization of weighted mel-frequency cepstral coefficients (WMFCC). The WMFCC algorithm has achieved high precision values for audio file retrieval on the GTZAN dataset [24].

Ultimately, despite the recent identification of its shortcomings, the GTZAN dataset has played a crucial role in advancing music genre detection systems. In order to achieve a high level of accuracy, it is crucial to employ feature extraction techniques such as WMFCC when utilizing this dataset for audio

retrieval tasks. The study highlights the essentiality of effective feature extraction and maintaining the integrity of datasets in music information retrieval. The GTZAN audio collection consists of 1000 recordings, each lasting 30 seconds. Each of the ten genres consists of one hundred tracks, all of which are 22050Hz Mono 16-bit WAV audio files. Genres include pop, reggae, rock, hip-hop, jazz, blues, country, disco, and metal.

To sum up, the GTZAN dataset needed some preprocessing to adapt to the FSL challenges outlined below.

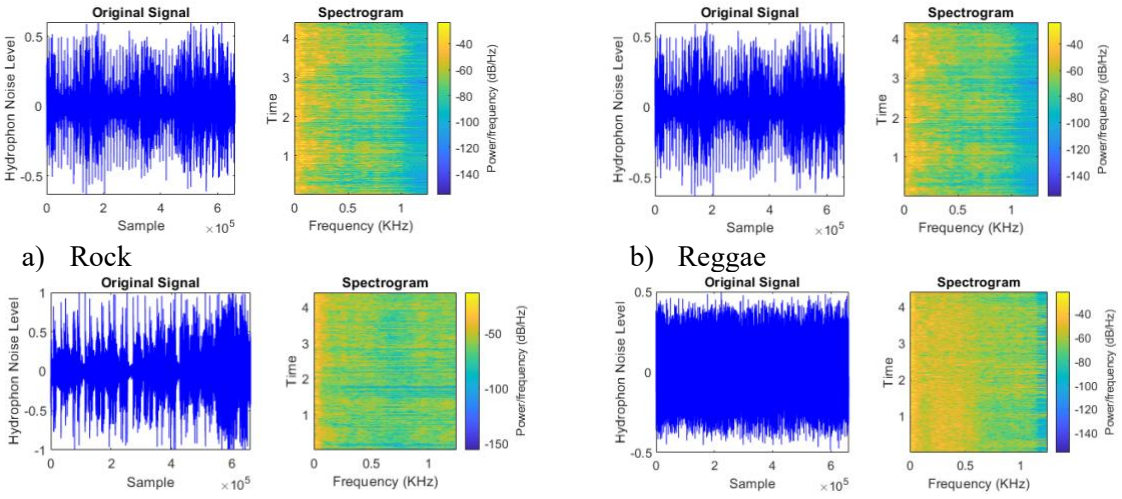
Resampling: All audio files were transformed to have the same sampling rate of 22.05 kHz.

Segmentation: Audio files were split into thirty-second clips and divided again with three-second overlapping windows for task creation [25].

Feature Extraction: From each segment, 13 Mel-frequency cepstral coefficients (MFCCs) were extracted, with a frame size of 25 ms and a hop size of 10 ms. These clips had zero mean and unit variance, normalized individually for every audio clip.

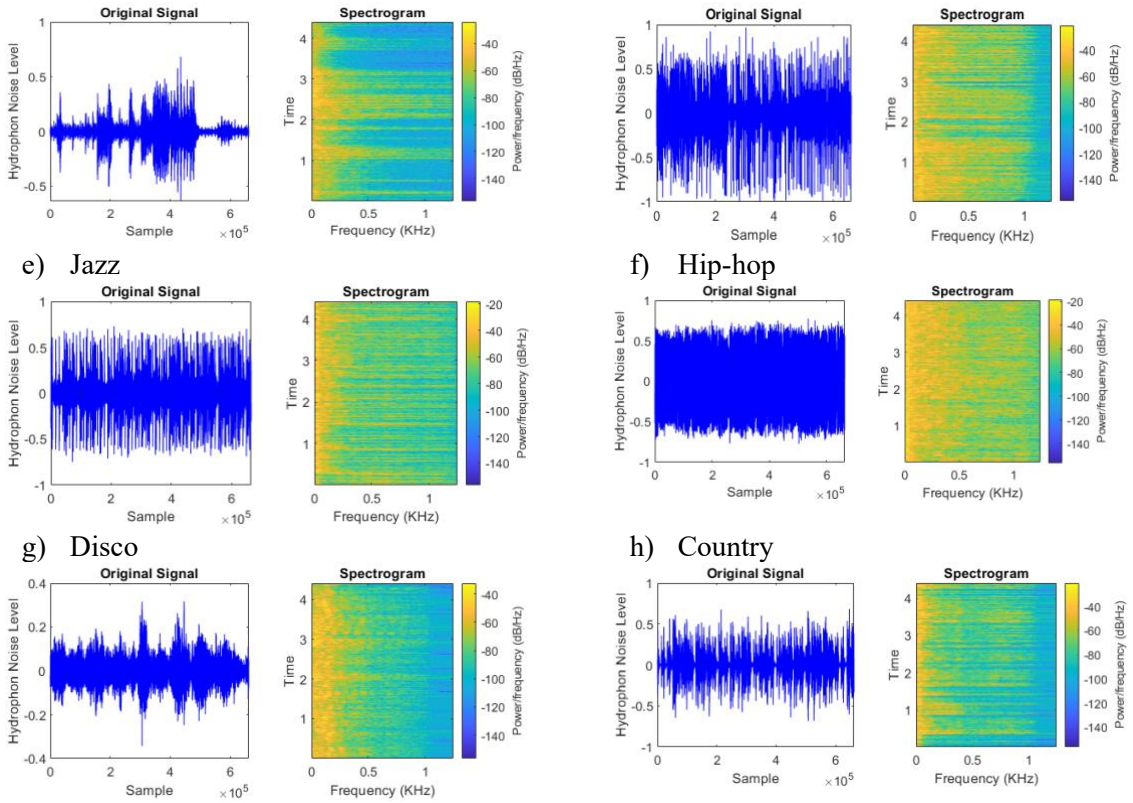
Augmentation: To decrease data sample bias, augmentation strategies such as pitch shifting, time stretching, and noise injection were used.

The dataset was divided randomly into sets with 70%, 15%, and 15% for training, validation, and testing, respectively, ensuring the genres in the test set were not used in training [26,27]. Fig. 2 depicts a standard portrayal of each genre.



a) Rock b) Reggae

c) Pop d) Metal



e) Jazz f) Hip-hop

g) Disco h) Country

i) Classical j) Blue

Fig. 2 An exemplar depiction of each genre alongside its corresponding spectrogram.

3.2 Notion meta-learner (NML)

Section 2 of the paper centers on the difficulty of precisely identifying music genres using deep neural networks (DNNs) when data are scarce [28,29]. To tackle this challenge, we introduce the NMLFSL, which comprises two stages: NML customization and recognition. Within the NML framework, "notion learners" are defined as dedicated sub-modules or models that capture high-lever domain-specific features (so-called "notions") from raw data. These "notions" are used as primitive abstractions on top of which few-shot learning tasks are built. Unlike standard feature extractors, notion learners aim at features relevant to the specific problem and domain, emphasizing interpretability. Algorithm 1 presents a comprehensive framework for the proposed NML technique, elucidating its intricacies. More information about this method will be discussed in later sections.

Algorithm1: NML Extraction for MGR

Input: *Spectrograms*
Output: *NMLs*
For *each* $\mu = 1,2, \ldots, N$:
    Calculate *Two points with a maximum difference in threshold values*
    Locate *NMLs by identifying locations in which the amplitude difference surpasses the threshold*

Calculate *coordinates of NMLs (vectors of coordinates)*
Clustering on the NMLs is performed *using similarity measures to classify them into various categories.*
Determine *the centroid of every group as the NML for that category*
Apply *MGR to Spectrograms by calculating the Euclidean distance between the NML of each image*
Allocate *every sound to the group that has the nearest center*
Return *NMLs (k_n; A_n)*

In summary, our suggested method for MGR entails utilizing NMLs derived from their Spectrograms. These NMLs are acquired by detecting power fluctuations that exceed a specific threshold and grouping them according to their similarity. To classify the signals, we assess the distance between every NML and the centroid of each cluster. The suggested technique for NML extraction entails identifying particular points in time and thresholds, facilitating the accurate determination of amplitude fluctuations' exact location. The NML locations are recorded as (k_n; A_n) coordinates, where A_n and k_n are the spatial and temporal coordinates at which the amplitude is above the threshold. Our approach provides a thorough method of collecting essential data from MGR spectrograms. Clustering NMLs can aid in the identification of unique characteristics and patterns that distinguish these genres, hence enhancing the precision of their classification.

3.3 FSL

The NMLFSL method employs NMLs to generate distinct metric spaces for every NML using a limited collection of annotated real-world data. With the use of NML prototypes as class-level disparity markers in the assessment space of each higher-level dimension and NML incorporating functions as learners, NMLFSL may discover minor differences. The NMLFSL model improves the primary learner's generalization capacity by combining many precise NML learners.

NMLFSL uses three data sets in its operation: a support set (*SS*), unlabeled query data (*UQD*), and labeled training data (*LTD*). Unlike training and query sets located in separate label spaces, the support set comprises tagged datasets that share the same space. In the NML framework, a data point is expressed as a pair of integers (k, A), where k represents the label and A represents the actual data point. Labeling the query set with the help of the labeled support and training sets is the goal of the NMLFSL technique.

In an effort to successfully generalize new issues, prototypical FSL is a popular machine learning approach. This method uses episodes, which are tiny subsets of data, during the training phase. The number of categories form the training data are included in each episode, and the data points within each one are labeled appropriately [30].

Each episode of the NMLFSL model is devoted to maintaining a constant sample group while minimizing loss in the query set. When there is insufficient data available during the testing phase, this strategy helps to increase the generalizability of the model. An example of a training set utilized in this method is the so-called "balanced episodes," where "Way" denotes the number of classes in each episode, whereas "shot" denotes the number of support points each class [31].

An NML learner IF_{ψ}^{θ}, or non-linearly parameterized NML integrating function, is a component of the NMLFSL method. Every NML learner θ generates NML prototypes PN_{φ}^{θ} for each class φ, which are calculated by taking the average of the observed merging of the data points in the support set. By averaging, the model obtains a more advanced depiction of each class, which in turn improves its ability to generalize as Eq. (1).

$$PN_{\varphi}^{\theta} = \frac{1}{\Gamma_{\varphi}} \sum_{(k_n; A_n) \in \Gamma_{\varphi}} IF_{\psi}^{\theta}(\mathrm{k}_n \circ \boldsymbol{\lambda}^{\alpha}) \tag{1}$$

In this case, the NML number is indicated by the symbol θ, and the Hadamard product is shown via the symbol $\circ$. Therefore, every class, represented by φ, is signified by a collection of T NML prototypes, which are labeled as $\{PN_{\varphi}^{\theta}\}_{\theta=1}^{T} = 1$. Furthermore, $\{\boldsymbol{\lambda}^{\theta}\}_{\theta=1}^{T} = \Gamma$ denotes the set of "T" NMLs obtained through the suggested NML extraction procedure. The NML functions as preexisting knowledge to aid the model in constructing a more accurate depiction of each category. By contrasting independent NML learners with NML prototypes, Fig. 3 illustrates how NMLFSL learns NML by integrating all dimensions and allocating NML significance grades to each dimension.

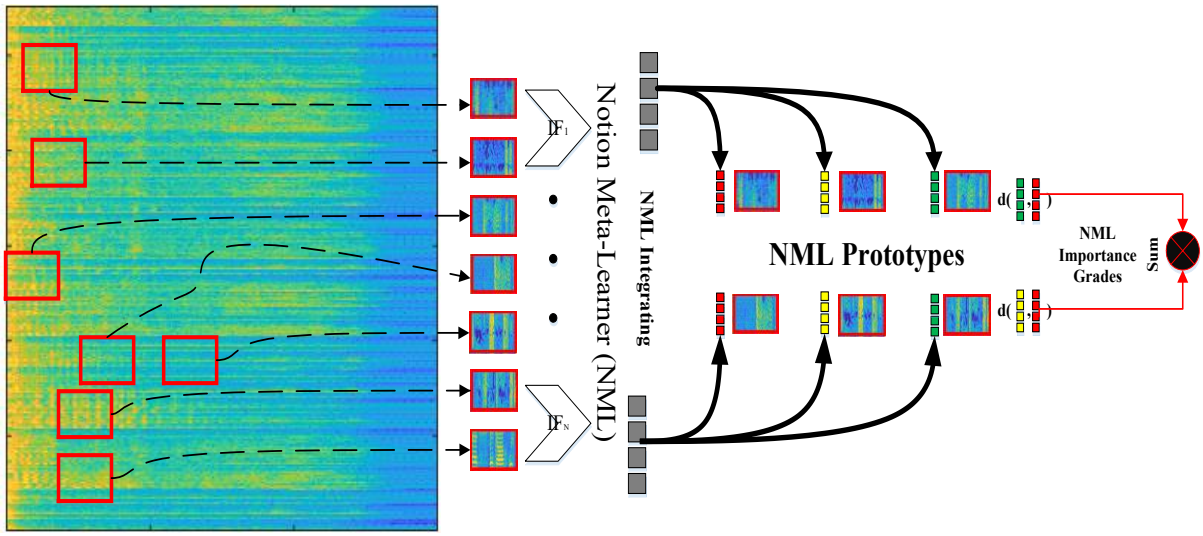


Fig. 3 NMLFSL model's ability to learn to merge NMLs along all dimensions.

NMLFSL uses an interpretability mechanism to give each class's local and global NML an importance grade. When NML incorporates the query point of data into its products, it contributes more significantly to the query point's classification, which is why it gets higher marks.

For every prediction, NMLFSL provides accurate locations by using the local grades. The model determines the global NML significance grade by calculating the inverse average distance between the NML and integrating the query points of interest with the NML prototype. This distance is then used for global explanations. By assigning a grade to the NML and organizing them accordingly, it becomes feasible to identify crucial NML among a collection of samples and acquire valuable insights. Also, NMLFSL can sort points by their similarity to a given NML. This technique helps find locally similar examples in or out of identical classes. Finding instances that are similar to or different from the NML prototype is easier with this helpful tool.

4. Experimentation

We investigated using the NMLFSL approach on two datasets, one with 25 NMLs each. However, several photographs in the datasets were devoid of unique NML. In these instances, we substituted the missing NML with a prototype counterpart. Utilizing part coordinates, we generated an NML mask that populated the vicinity-bound rectangular with a specified dimension.

4.1 Experimental configuration

The experiments were performed on an NVIDIA RTX 3090 GPU, Intel Core i9, and 32 GB RAM machine. The implementation was done in PyTorch 1.10.1 on Python. Random seeds were set to 42 for

reproducibility across different runs. A comparison analysis is undertaken to evaluate the efficacy of NMLFS against seven distinct benchmarks, including adapting with memory for probabilistic FSL (AWMP-FSL) [32], improved prototypical networks for FSL (IPN-FSL) [33], scaling FSL (CFSL) [34], distribution-agnostic probabilistic FSL (DAPFSL) [35], conditional diffusion FSL (CDFSL) [36], large-scale FSL (LSFSL) [37], and ProtoNet [38]. These methods were selected because they represent a broad spectrum of approaches in FSL, ranging from metric-based to optimization-based techniques.

We employed the widespread 2-way classification issue to evaluate the models in this study. The query set comprised five samples without labels from each class in the support set. After selecting the best model according to its validation accuracy, its performance was evaluated using a distinct test set that included additional categories. More precisely, we employed the evaluation methodology outlined in [39], which involved dividing the data into three sets: 50% for basic training, 25% for validation, and 25% for testing. Notably, the same division was maintained consistently throughout the process.

Our solution utilized the Conv4 structure, comprising four convolutional layers and an input dimension of 84 × 84, as suggested by [38]. We utilized Adam as the optimizer for all datasets, with a starting learning rate of 0.001; 40000 episodes were used for training the five-shot and 50000 episodes for the one-shot tasks [39].

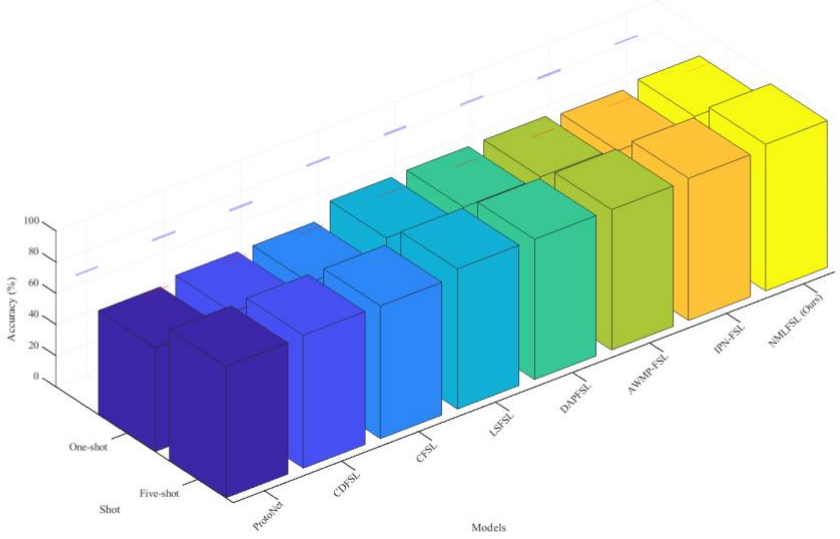
We took advantage of the fact that NML learners share network parameters to make training the NMLFSL method faster. The convolutional structure analyzes the complete image k_n, resulting in a merged set of features $IF^{\theta}(k_n)$. From this, the j-th NML merging is obtained as $IF^{\theta}(k_n \circ \boldsymbol{\lambda}^{\theta})$. Applying the mask at the beginning or end of the procedure does not substantially impact performance, but the latter approach reduces the duration of the training. When specific sections of an image lack annotations, we substitute the absent NML with a representative NML corresponding to the full image.

4.2 Assessment of Performance

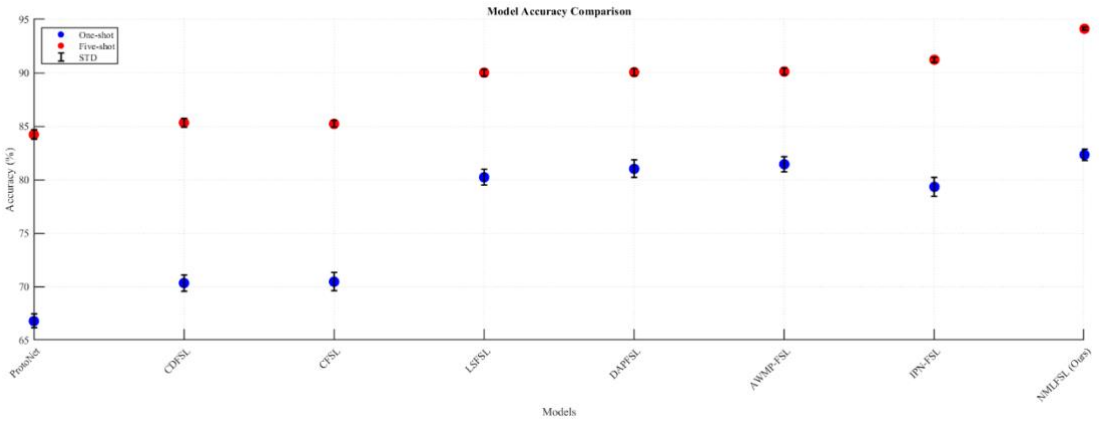
Table 1 shows the results of our effectiveness assessment on the given datasets, with NML as the domain prior. The NMLFSL outperforms all other comparable models by a significant margin. Fig. 4 shows a more graphically thorough representation of the data.

Table 1 The outcomes of the 1-shot and 5-shot

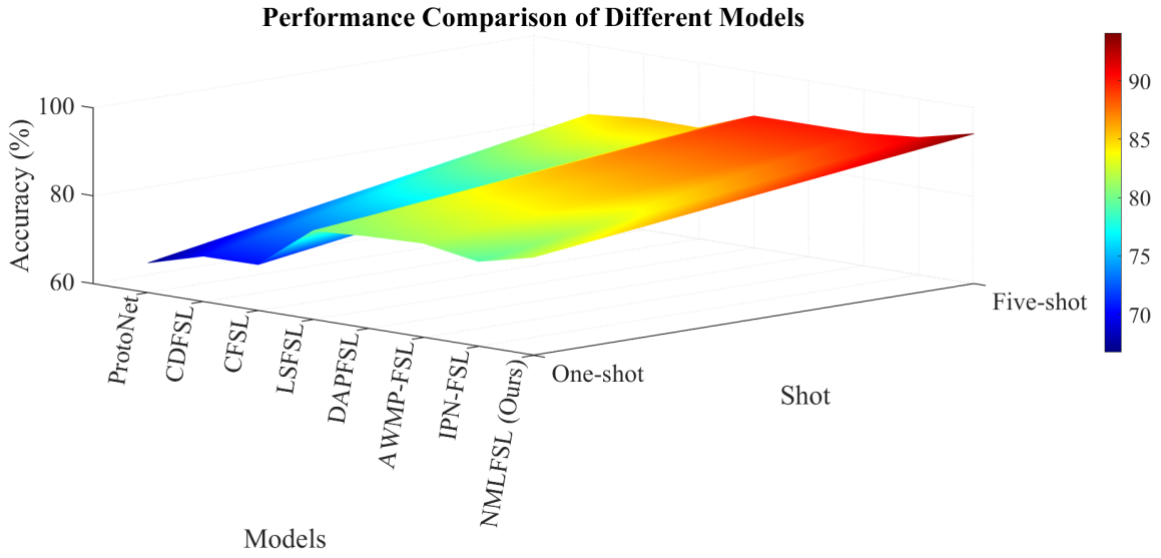
Methods	one-shot	five-shot
IPN-FSL	79.33 ± 0.89	91.22 ± 0.25
CDFSL	70.33 ± 0.77	85.33 ± 0.41
CFSL	70.45 ± 0.85	85.22 ± 0.33
LSFSL	80.22 ± 0.74	90.01 ± 0.34
DAPFSL	81.01 ± 0.83	90.05 ± 0.34
AWMP-FSL	81.44 ± 0.69	90.11 ± 0.33
ProtoNet	66.77 ± 0.65	84.22 ± 0.45
NMLFSL (Ours)	**82.33 ± 0.55**	**94.11 ± 0.14**



a) 3D bar plot representation



b) Scatter plot representation



c) 3D plot representation for various shots

Fig. 4 The outcomes of the one and 5-shot tasks

Table 1 succinctly presents findings from our assessment of different FSL models on a standardized dataset. Performance significantly improves when transitioning from one-shot to five-shot settings across all models, suggesting that having more training data enhances generalization.

NMLFSL routinely performs better than previous methods in both one-shot and five-shot scenarios. These outcomes imply that the model is more proficient in acquiring knowledge from restricted data and extrapolating to unfamiliar categories. ProtoNet and CDFSL models exhibit comparatively inferior performance compared to other models, suggesting possible constraints in their capacity to utilize provided information effectively.

There is a noticeable pattern where more intricate models, such as DAPFSL, AWMP-FSL, and NMLFSL, have superior levels of accuracy. Nevertheless, a trade-off exists as using this technology leads to higher computing complexity, which must be considered when implementing it in real-world applications.

The notable increase in performance attained by NMLFSL highlights the significance of methodological breakthroughs in FSL. The model may include superior feature representations or more efficient learning algorithms compared to current methods. NMLFSL's strong performance in both one-shot and five-shot scenarios indicates its resilience and capacity for generalization, which are vital for practical applications that require the ability to adapt to new tasks and classes.

Additional research is necessary to comprehend the precise mechanisms that contribute to NMLFSL's exceptional performance. In addition, investigating methods to reduce computing complexity while maintaining high accuracy would improve the practical usability of advanced FSL techniques. The reported enhancements in performance have significant ramifications for various applications, such as image recognition, natural language processing, and medical diagnosis, where the availability of labeled data is frequently restricted or costly to acquire.

To summarize, our study showcases the effectiveness of NMLFSL in tackling the difficulties of FSL, which can lead to enhanced performance in a range of practical applications. Additional investigation in this field holds the potential for ongoing progress in machine learning frameworks.

We conducted experiments to ascertain if the improvements in NMLFSL performance are predominantly due to the use of NML learners rather than additional weights and to determine how a more robust Conv6 backbone affects NMLFSL performance. Our research indicates that NMLFSL continues to exhibit substantial performance improvements even when using a more complex underlying structure. Furthermore, we compared NMLFSL and a ProtoNet ensemble and discovered that NMLFSL achieves superior performance, even when the weights are shared among NML. Moreover, we evaluated NMLFSL's performance with standard weights NML and saw no considerable drop in performance; Table 2 tabulated the obtained results. The 3-dimensional plot of these models for various shots (1 to 5) is presented in Fig. 5.

Table 2 Performance evaluations of NMLFSL models and an ensemble of standard networks show that NML uses the same weights.

Models	1-shot	5-shot
Ensemble ProtoNet	69.42±0.36	85.37±0.22
Shared weight NML	80.55±0.25	93.44±0.19
NML	**82.34±0.14**	**94.02±0.11**

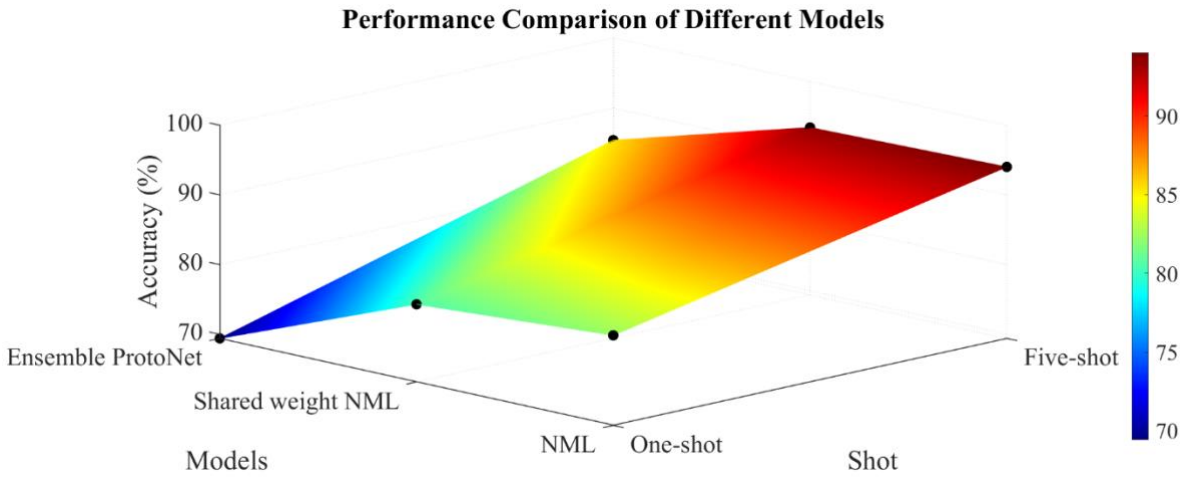


Fig. 5 3D plot of the comparison models for various shots (1 to 5) is presented in

By comparing three different models in 1 and 5-shot scenarios, we may gain insight into their effectiveness in FSL tasks. Considering the Ensemble ProtoNet first, its performance, even though praiseworthy, is still the lowest compared to its counterparts. Within the one-shot situation, the Ensemble ProtoNet achieves an accuracy score of 69.42% ± 0.36, demonstrating its ability to learn from limited data. However, in the scenario with more data, precisely five shots, the model's accuracy climbs to 85.37% ± 0.22, which suggests a proportional improvement. However, despite the advancements made, the overall precision of this model is significantly lower compared to the other models being evaluated.

On the other hand, the shared weight NML model appears to make substantial improvements regarding performance in both cases. This model derives considerable improvements in accuracy from the ensemble of shared weights. In a one-shot situation, the shared weight NML model can get an 80.55% ± 0.25 accuracy, a much more significant boost than the ensemble Prototypical Network. In the five-shot scenario, this tendency continues, with its accuracy increasing to 93.44% ± 0.19. The shared weight mechanism proves effective as it improves the model's accuracy, showing its robustness in many situations.

Nonetheless, the topmost results seem to originate from the performance of the MML model. The MML model surpasses its rivals in terms of accuracy by utilizing potential meta-model learning mechanisms. The one-shot scenario obtains an 82.34% ± 0.14 accuracy, outperforming the Ensemble ProtoNet and the Shared weight NML model. In the five-shot scenario, the model's accuracy increases even further

to 94.02% ± 0.11, demonstrating its extraordinary capacity to derive knowledge from a small amount of data.

These findings strengthen the case for using more advanced methods in FSL challenges. Ensemble ProtoNet is the simplest model, while accuracy increased even more in shared-weight NML and MML models, particularly in scenarios with little available information. The shared-weight mechanism in NML enhances performance, whereas the MML model's meta-learning technique exhibits exceptional effectiveness in utilizing limited data resources. Additional investigation into the precise mechanisms underlying these models could reveal valuable insights into their excellent performance and guide future developments in FSL approaches.

4.3 The effect of NML number

Fig. 6 shows the relationship between the number of NMLs and the performance of NMLFSL on the test samples. We begin with the results generated from ProtoNet, which uses a single NML that covers all the input sizes. Consequently, we gradually increased the quantity of NML, which trained and evaluated NMLFSL. The results indicate that augmenting the amount of NML improves the performance of NMLFSL across all datasets.

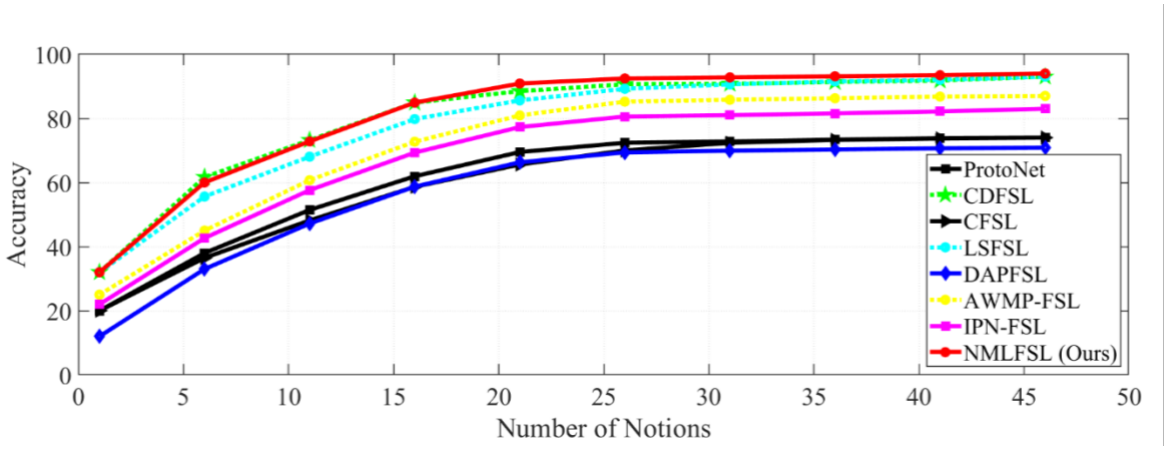


Fig. 6 The effect of NML number

As can be inferred from this figure, when just one NML corresponding to a 10 dB threshold of the total NML is introduced, ProtoNet's performance is improved by 12% in the 1-shot and 5% in the 5-shot task. We have increased the number of NMLs to 50, including many repeated connections, to evaluate NMLFSL's ability to withstand a substantial collection of superfluous NML. Fig. 6 shows that compared to using 25 NML, using 50 NML in NMLFSL slightly improves performance. Even when

faced with a small amount of NML, inadequate annotations, and a large amount of identical and overlapped NML, NMLFSL outperforms other techniques.

4.4 Ablation analysis to test the effect of distance functions

A critical application of distance measures in FSL is to find the degree of similarity among two samples or the degree of similarity of a new sample with a few known samples. To assess the distance measure's effect on the model's overall performance, we conducted ablation experiments. The distance measures employed in the study were varied to ascertain the model's accuracy. Using one of the ablation studies as a case study, we could highlight the importance of the measurement function and its impact on overall performance. FSL employs various distance measures, including, but not limited to, Manhattan, Euclidean, Mahalanobis, and cosine distances.

- **Euclidean distance**: Locations A and B are separated by a straight line distance; for example, Euclidean distance is the straight line metric that separates two locations. The roots of this concept lie in Euclidean geometry, which bears the name of Euclid, an eminent mathematician from Greece. Mathematically, the Euclidean distance of two vectors, x and y, is expressed in Eq. (2), and a Cartesian space is represented as three-dimensional spatial coordinates:

$$d_{Euc}(\mathbf{x},\mathbf{y}) = \sqrt{\sum_i (x_i - y_i)^2} \tag{2}$$

Where x_i and y_i represent the i-th elements of vectors **x** and **y**, respectively, the i value falls inside the vector's length range.

- **Manhattan distance**: The Manhattan distance, which is also called taxicab distance or city block distance, is a distance metric that can be illustrated using the vectors **x** and **y** wherein the pairwise distance between the vectors is computed using Eq. (3):

$$d_{Man}(\mathbf{x},\mathbf{y}) = \sum_i |x_i - y_i| \tag{3}$$

- **Cosine distance**: The cosine distance, sometimes called the cosine similarity distance, is a metric used to assess how similar two vectors are within a multi-dimensional space. The computation requires the use of the cosine of the angle made by the vectors, which expresses how close the two vectors point out to the same direction rather than the magnitude of the vectors [40]:

$$d_{Cos}(\mathbf{x},\mathbf{y}) = 1 - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}} \quad (4)$$

- **Mahalanobis distance**: Mahalanobis distance is a measurement that considers how specific points in a dataset interrelate with each other, in essence, measuring how far a particular point is from the distribution. In contrast to Euclidean distance, which applies equal and independent treatment to all dimensions, Mahalanobis distance can be described by the particular equations [41]:

$$d_{Mah}(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x}-\mathbf{y})^T \cdot inv(R) \cdot (\mathbf{y}-\mathbf{x})} \quad (5)$$

Where R represents the data covariance matrix, while $inv(R)$ denotes its inverse.

We conducted an analysis and evaluation to determine which distance measurements are most effective in accurately classifying NMLFSL. The results of this analysis are summarized in Table 3 and Fig. 7. The findings demonstrated that the Euclidean distance outperformed other measures in terms of accuracy in all studies.

Table 3. The influence of various similarity criteria on the NMLFSL performance.

Distance	1-shot	5-shot
Manhattan	81.33±0.68	93.22±0.35
Mahalanobis	79.25±0.74	92.11±0.40
Cosine	78.33+0.82	92.37±0.41
Euclidean	**83.44±0.56**	**94.04±0.24**

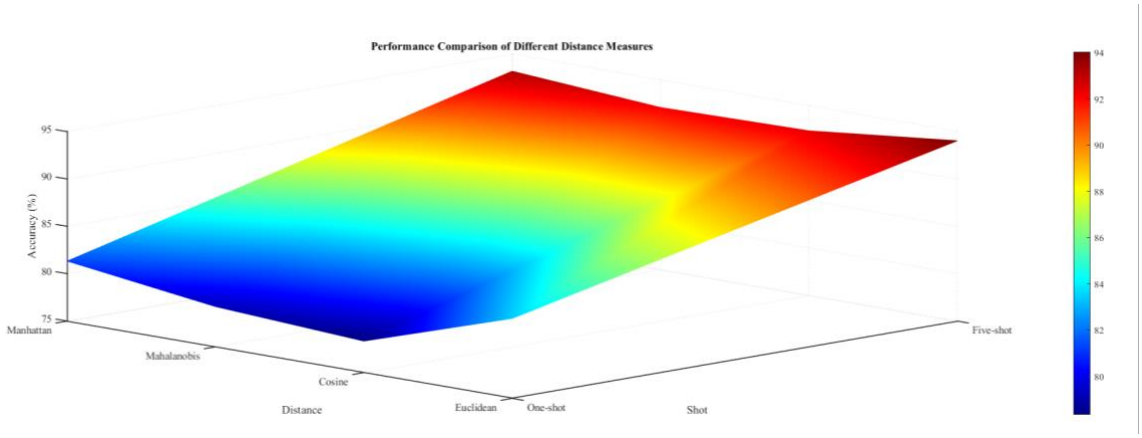


Fig. 7. The influence of various similarity criteria on the NMLFSL performance according to multiple shots (1 to 5).

4.5 Ablation-based experiment to assess the network's backbone

Ablation analysis was conducted on the network's backbone to evaluate the effectiveness of NMLFSL compared to benchmark approaches. For the GTZAN dataset, we employed the more complicated Conv6 instead of the more straightforward Conv4 backbone. Furthermore, part-based markers were used for precise NML identification. Fig. 8 and Table 4 illustrate the average and standard deviation of 600 episodes chosen randomly from the research.

Table 4 The outcomes of the Conv6 backbone's evaluation of performance.

Models	1-shot	5-shot
SUFSL	68.22±0.22	84.25±0.30
AWMP-FSL	68.33±0.25	80.22±0.42
CDFSL	68.25±0.41	80.89±0.41
CFSL	66.11±0.32	82.67±0.39
LSFSL	67.38±0.45	85.38±0.20
DAPFSL	68.86±0.40	83.75±0.34
ProtoNet	68.85±0.36	84.44±0.21
NMLFSL/1 NML	71.26±0.27	85.89±0.18
NMLFSL	**74.45±0.11**	**90.34±0.09**

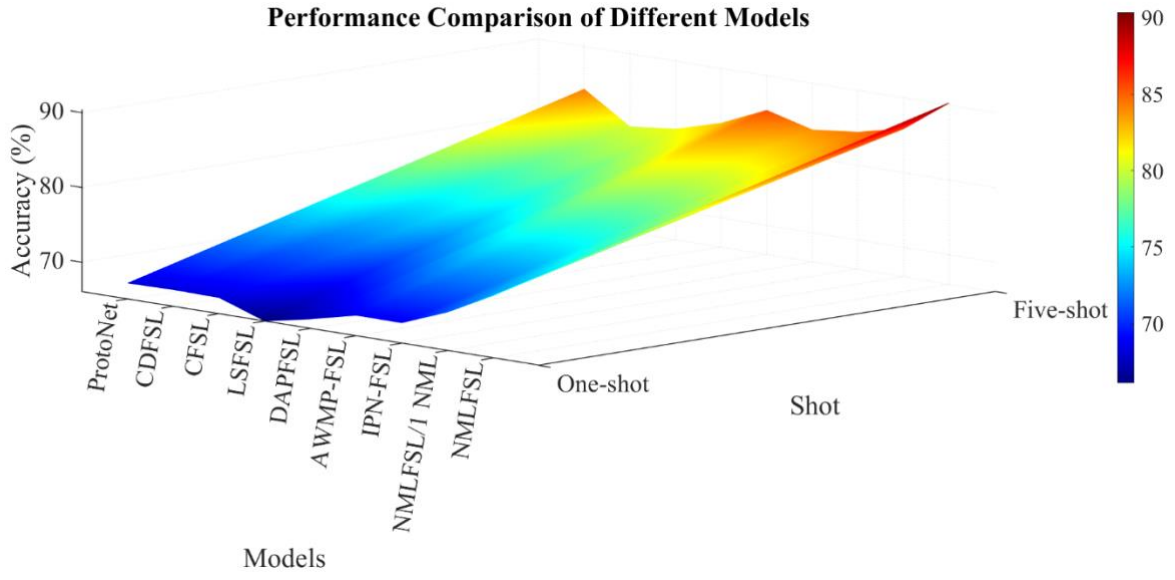


Fig. 8 shows the results for various comparison models using the Conv6 backbone according to different shots (1 to 5).

The performed ablation study attempted to assess the performance of NMLFSL compared to various benchmark approaches. In this test, we used a Conv6 backbone rather than a Conv4. Part-based markers

were employed to further identify NML within the GTZAN dataset. Observations recorded in Fig. 8 and Table 4 outline the average and dispersion of a sample comprising 600 randomly chosen episodes. The table summarizes the performance of models SUFSL, AWMP-FSL, CDFSL, CFSL, LSFSL, DAPFSL, ProtoNet, and NMLFSL in one-shot and five-shot experiments with the Conv6 backbone. The results show that NMLFSL performs better than the rest in the one- and five-shot scenarios, with accuracy rates of 74.45% and 90.34%, respectively. What is significant is that NMLFSL proposes a method that outstands most methods, further validating the incorporation of NML within the FSL framework.

Fig. 8 discusses the performance results obtainable from comparing different proposed models using the Conv-6 backbone. The graph supports the conclusions made in Table 4, in particular that NMLFSL is superior to other methods in one- and five-shot paragraphs. This gap between NMLFSL and the remaining models illustrates the potential and effectiveness of NMLFSL in handling FSL tasks as this gap continues to grow with increasing shifts.

As a result, the evidence indicates that correcting for neighborhood manifold learning in the NMLFSL framework by applying a more sophisticated Conv6 backbone and using part-based markers significantly boosts FSL's performance on the GTZAN dataset. The results show the promise of NMLFSL as an efficient method for addressing problems on FSL tasks, especially for big and heterogeneous datasets like GTZAN.

4.6 The assessment of the NML's positions

To assess NMLFSL's effectiveness on the datasets using physically extracted NML, we used the auto-encoding feature-finding methods proposed in [42]. We used the default parameters and implementation provided by the authors and picked 25 features. The estimated feature coordinates produced by the encoding module created an NML mask by enclosing the detected features in a bounding box. We compared 1 and 5-shot tasks, employing randomly selected and human-determined masks for NML. The results are tabulated in Table 5 and Fig. 9. Furthermore, Fig. 10 illustrates instances of retrieved characteristics from ten distinct photos across all classes.

Table 5 Randomly chosen masks and masks specified by humans using a threshold.

Models	one-shot	five-shot

Randomly masks	72.44±0.89	91.35±0.24
Threshold-Determined (human)	**73.74±0.71**	**93.04±0.11**

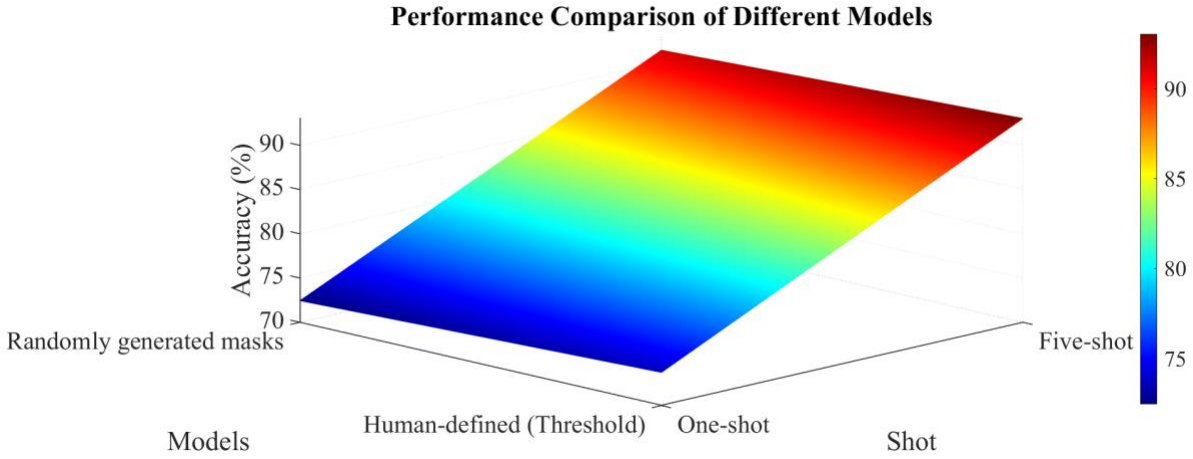
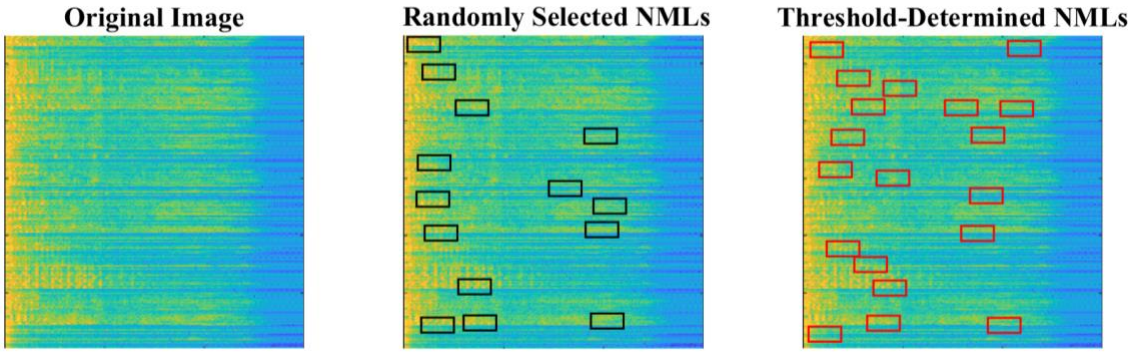
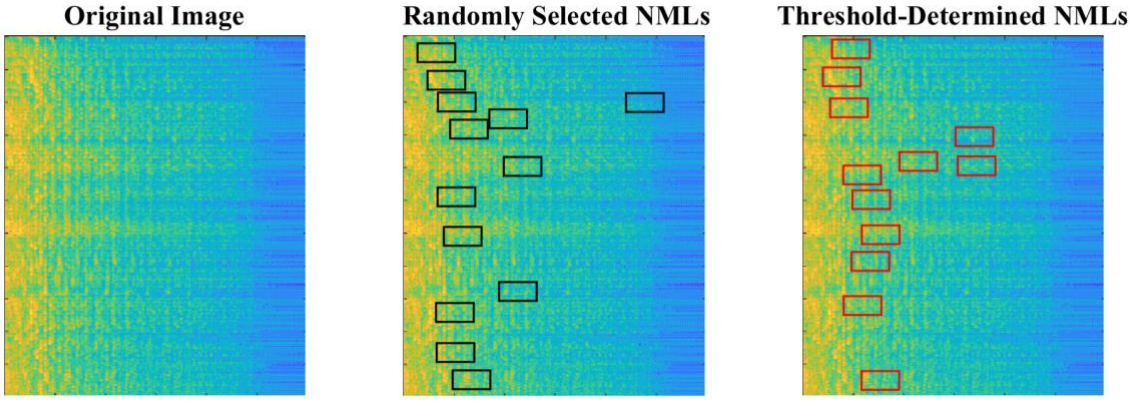


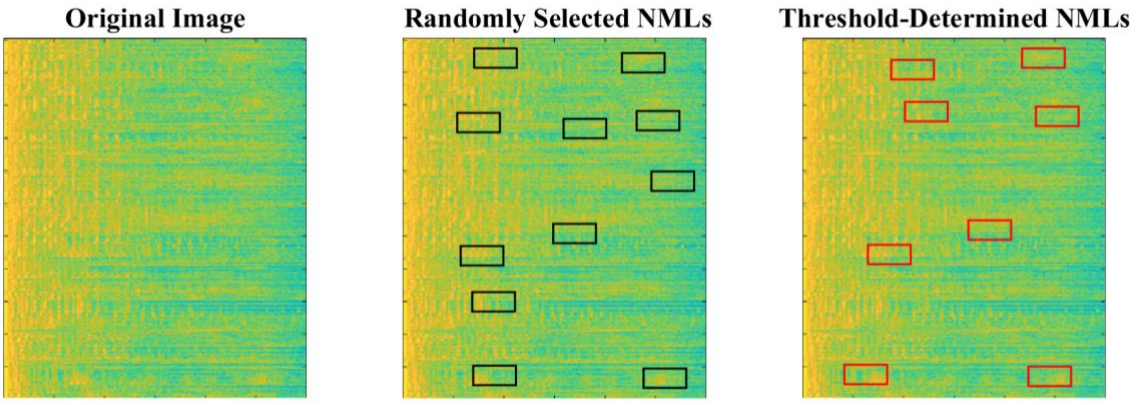
Fig. 9 A performance comparison between selected masks and specified masks using a threshold for different shots (1 to 5)



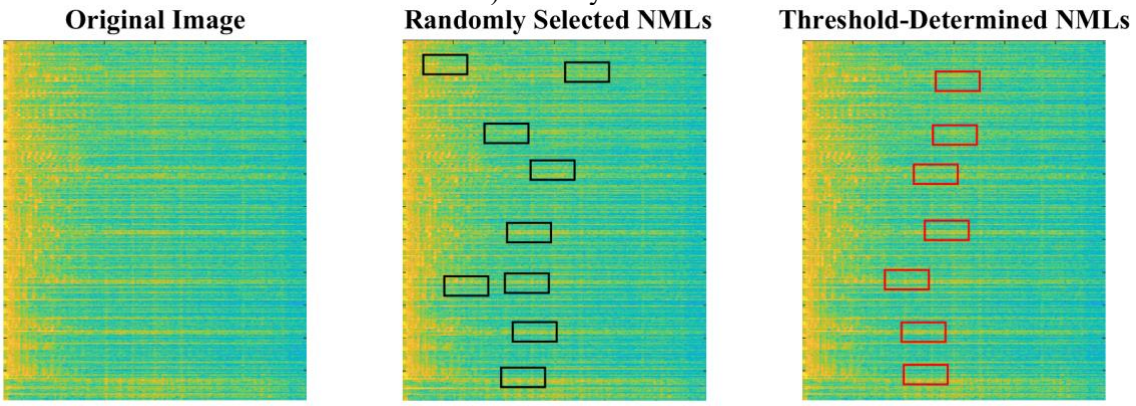
a) Blues



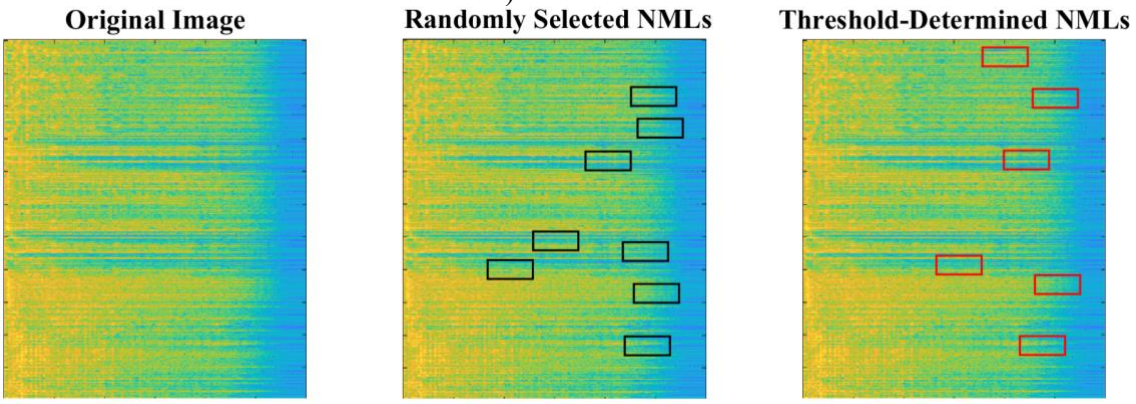
b) Classical



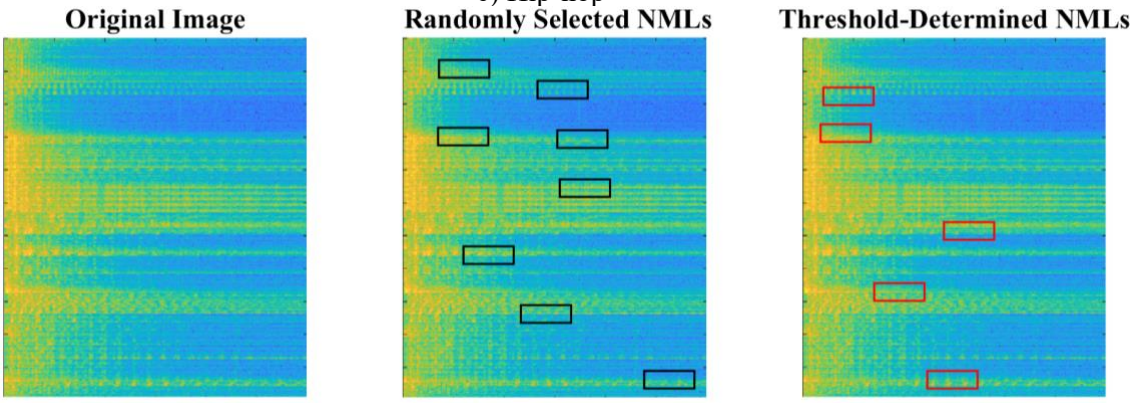
c) Country



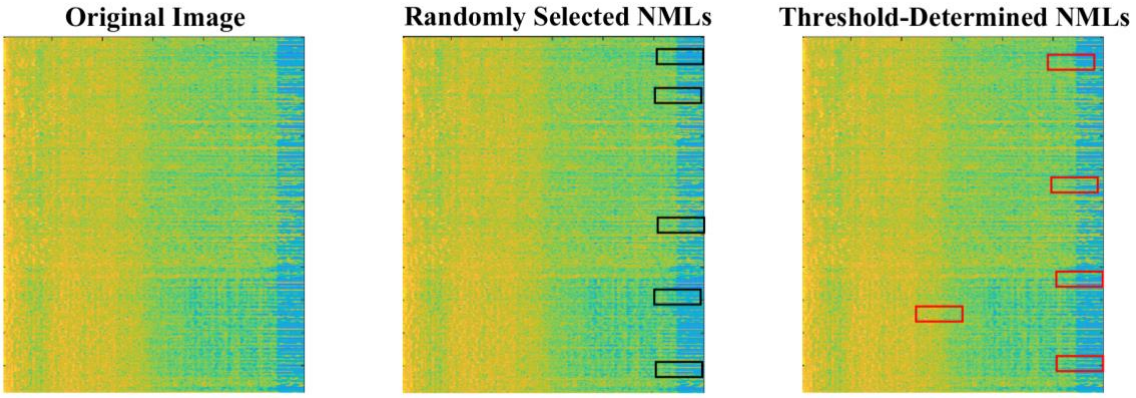
d) Disco



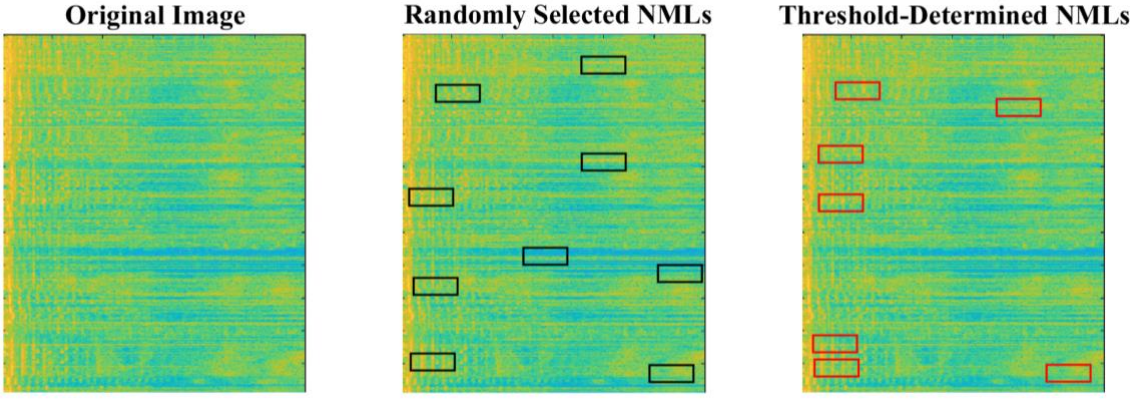
e) Hip-hop



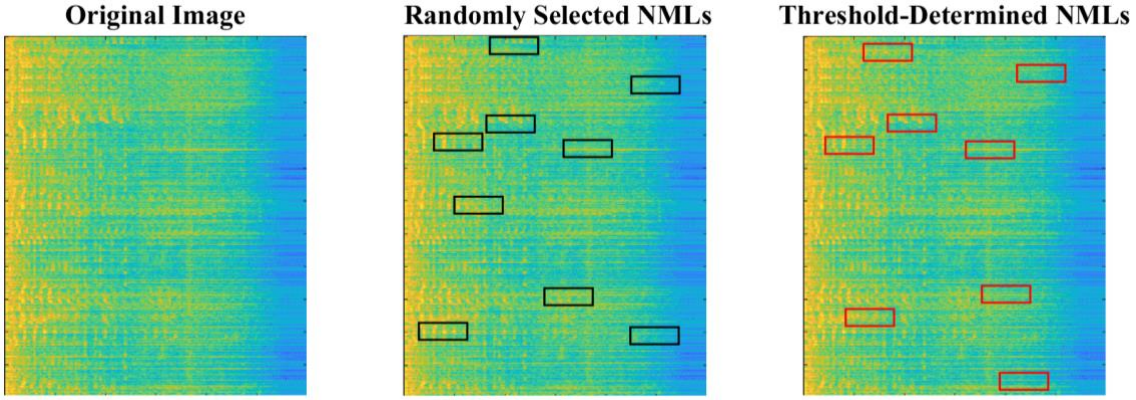
f) Jazz



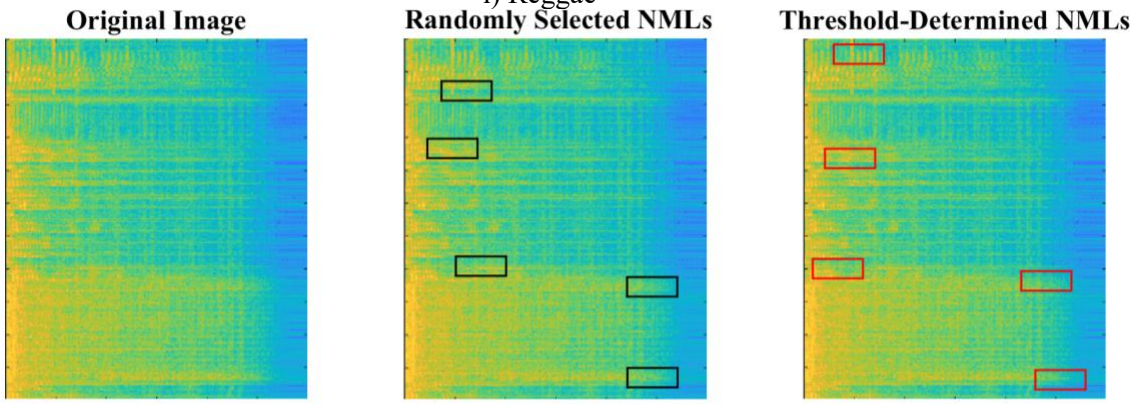
g) Metal



h) Pop



i) Reggae



j) Rock

Fig. 10 displays typical instances of extracted NML for ten different photos across all classes.

Analyzing the performance of randomly generated masks and human-defined (Threshold) models in 1 and 5-shot situations provides valuable insights into the efficacy of various segmentation methods.

The one-shot scenario yielded an accuracy of 72.44% ± 0.89 for the randomly generated masks model, while the human-defined (Threshold) model achieved a slightly higher accuracy of 73.74% ± 0.71. This outcome implies that both methods can produce satisfactory segmentation outcomes when training data are scarce. However, the human-constructed model appears to have a slight advantage because it highlights the importance of human craftsmanship in creating accurate masks from a limited data set.

As expected, both models in the five-shot situation significantly enhance accuracy as they have additional training data. The masks model generated randomly achieves an accuracy of 91.35% ± 0.24; however, the human-defined (Threshold) model obtains a better accuracy of 93.04% ± 0.11. This achievement emphasizes the exceptional efficacy of the human-defined method, even with a slight augmentation in training samples, indicating that the manual intervention in mask generation resulted in more precise segmentation outcomes.

These results stress the undeniable role of human involvement in developing accurate masks. While random masks can be effective in segmentation tasks, research demonstrates that human-developed methods are superior and require fewer training samples to produce reliable and correct segmentation masks. Future work has to focus on improving computerized algorithms for segmentation so that the results equal human-defined mask accuracy.

4.7 Time complexity analysis

For NMLFSL, the Big O notation must estimate time complexity by understanding the algorithm components' working principles, including algorithm complexity, data, and learned algorithms. From a macroscopic viewpoint, the analysis of NMLFSL's time complexity does not only focus on some general principles and examples. First, for the study of time consumption relative to the complexity of images, wherein d denotes the dimensionality/complexity of the data, it is assumed that the time expenditure per image is O(d).

Thus, the time to compute the entire dataset of size N in NMLFSL will be O(Nd). Note that the model's structure also contributes to the time complexity. Assume that the model has L layers and each layer has complexity O(m); then, the time complexity is due to the structure of the model being O(Lm).

Moreover, the adaptive learning aspect in FSL makes the scheme more complex. Let us assume that k shots are taken when adapting the system and O(p) computations are done for each investment so that the complexity of learning becomes O(kp). Considering all these aspects, we can express the time complexity of NMLFSL as O(Nd + Lm + kp).

However, this portrait reduces the true process to mere action by presuming free will and a fixed order of operations. Moreover, the actual time for the program to execute may differ significantly from the anticipated time due to the combined effects of concurrent execution, algorithm optimizations, and the performance of the computational unit. It is essential to note that in dynamic environments, rather than ones requiring real-time interactions, the inference time is more critical than the training time. This enables using the NMLFSL model for real-time analysis without internet connectivity, utilizing a small training sample to illustrate the need for effective execution.

4.8 Discussion

The NML framework offers a new angle for interpreting FSL by employing the concepts of "notion dimensions" and "partially organized metric spaces" that fundamentally increase system interpretability and serve to broaden the generalization capability of the FSL systems. Though the findings speak to the efficacy of this approach, particularly in music genre classification, additional essential issues need to be analyzed to appreciate the full scope of the contributions and limitations of this work.

- **Key Contributions**

NML attempts to deal with two significant problems of FSL: too few labeled examples and the generalization problem where a system must perform several unrelated tasks. The principles of the methodology broaden semantic machine learning by framing the learning process around notion dimensions, which is more typical for people. This new and efficient model structure captures relationships within the data at a much higher level, automatically improving the performance of many tasks with little training data available. The experimental results on the GTZAN dataset, which show significant improvement in classification accuracy over seven known FSL benchmarks, reinforce the efficacy of this approach.

Moreover, the completeness of ablation studies reflects the most essential parts of the NML framework and their quantitative merits. These studies show how idea representation, organization of the metric

space, and the classifier's design are integrated, laying the groundwork for enhancing the refinement of the methodology.

- **Shortcomings**

The NML framework does present new concepts and insights; however, it does have certain shortcomings that require more investigation:

The GTZAN Dataset Dependency

The GTZAN dataset is one of the most popular datasets for music genre classification; however, it has some issues… such as limited sample size, class overlap, and other data fusion problems. All these aspects lower the range of applicability of the results obtained. The efficacy of the NML framework should also be tested using datasets such as FMA and the Million Song Dataset to cement the claim of its effectiveness and scalability across multiple music genres and other contexts.

Real-World Application of NML And Its Challenges

The existing NML implementation faces high computational costs, especially regarding logic optimization and notion extraction. This computational cost presents difficulties for near real-time implementation systems with a premium on latency and scalability. Although the research notes this limitation, a comprehensive evaluation of time complexity in conjunction with some form of algorithmic approximation or even hardware acceleration needs to be conducted to make this approach seem feasible in practice.

Human Dependency In Notion Extraction

Using human-generated masks in the Notion Extraction stage incurs bias, making the process less scalable. This semi-automated approach of Notion Mask Learning ensures quality is sustained during the learning process. However, it presents challenges for large-scale and fully automated scenarios. Developing automatic segmentation strategies that are unsupervised or weakly supervised could unclog this bottleneck, making the NML framework much more flexible.

Reproducibility and Methodological Clarity

Going through the experimental setup, it is very vague how the training, preprocessing of the dataset, and selection of the benchmarks were done. Such gaps make it challenging to reproduce the results and restrict other researchers from verifying and expanding upon the results established. A more detailed

explanation of hyperparameter values, evaluation measures, and training procedures is fundamental to enhance reproducibility and transparency.

- **Practical Implications**

The NML framework that is being proposed has the potential for several models, especially those that need high-accuracy classification while having very few labeled data. For instance, personalized music recommendation systems can be used because the model can quickly adapt with little trained data. Likewise, the framework can be extended to other areas like image FSL for classification, medical FSL for diagnostics, or even FSL for natural language processing which is progressively gaining traction.

However, addressing these constraints is needed for adequate NML integration. Lightweight models with real-time system requirements, particularly for efficient and accurate systems, are extremely important to ensure strict latency requirements are met. The effective practical use of NML requires its more advanced versions to be made, which may be done by adding hardware for acceleration or compressing the model.

- **Future Directions**

To improve the reach and utility of the NML framework, the following future approaches are suggested:

Dataset Expansion

Evaluating the method against more extensive and heterogeneous datasets would enhance understanding of its generalizability and robustness. Moreover, datasets with differing class imbalances and noise levels would assess how the method's approach can be adapted to real-world problems.

Extraction Automation

A computer vision and deep learning-based fully automated notion extraction mechanism would drastically decrease the requirement for human input, thereby facilitating the broad utilization of the technology.

Real-time Application Optimization

Investigating deep learning’s NML computational efficiency, using approximation algorithms or parallelization, will solve the existing time complexity boundaries and make deep learning feasible for latency-sensitive tasks.

NML Adaptability Expansion

Moving beyond the scope of deep learning visual recognition or even text NML would evaluate this methodology's boundlessness and serve as an eye-opener for its undeniable practicality in FSL.

Integrating the meta-learner framework's construction effectively revolutionizes the few-shot learning paradigm. While these results are encouraging and hold great value, specific constraints, like the dataset's diversity, scalability, and computational efficiency, must be addressed to harness their full potential. I believe that by focusing on the strengths and weaknesses of the current methodology, the NML framework can potentially change the vision of few-shot learning.

5. Conclusion

To summarize, the NMLFSL method provides hope for overcoming the challenges faced in FSL tasks, particularly for music genre recognition in the case of thin data. That uses the function to combine available NML extractions and augments the overall precision and generalization of classification. The results show that NMLFSL has been much better than the benchmark models in several shooting scenarios. This proves its strength and the usefulness of learning from a few examples. The performance of the technology confirms the results of ablation studies, distance function studies, and the ability of NML derived visually from the image to 'evaluate' the technology. The precision of mask generation, which requires some degree of human involvement, also emphasizes the requirement of more work on auto-segmentation procedures. Indeed, a review of the time complexity of NMLFSL raises critical practical problems concerning its deployment in real time, where the duration to predict is more important than the duration for model building. Overall, the evaluation findings demonstrated that the NMLFSL adds value for performance enhancement in several real-world scenarios and calls for more focus in the future.

The adage "the scholar has it all" was evident in the NMLFSL performance ingenuity that the FSL models promised to challenge. Therefore, future NMLFSL research must focus on model and purpose specificity. So, once better classifiers are developed, the NMLFSL performance will also increase. Thus, research into novel distance metrics might give vital classification insights into the model's performance maximization. Consequently, improving segmentation in automation will optimize music image analysis by having closer accuracy to human-defined masks and, in the same breath, increase

reliability. There is a clear advantage in exploring the practical scope of the NMLFSL NN by expanding its applicability and domains in that it can be used.

Once fine-tuned with multi-task learners, cross-domain missions, and a sufficient number of clusters, the NMLFSL will reclaim its position as the once-favored algorithm by always suggesting discrimination boundaries when insufficient data is present. Using this approach to, for example, target close target detection should greatly assist industries that require a quick update. In conclusion, more research into NMLFSL and its derivatives will fill a clear gap. NMLFSL derivatives could be game-changers for many machine-learning problems and help break down barriers in the real world.

Declaration

Funding: Not Applicable

Conflicts of interest/ Competing interests: The authors declare that there is no conflict of interest

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Data availability statement: The dataset used is available from https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification.

References

[1] N. Narkhede, S. Mathur, A. Bhaskar, M. Kalla, Music genre classification and recognition using convolutional neural network, Multimed. Tools Appl. (2024) 1–16.

[2] W. Zeng, Z. Xiao, Few-shot learning based on deep learning: A survey, Math. Biosci. Eng. 21 (2024) 679–711.

[3] C. Zhu, Research on Emotion Recognition-Based Smart Assistant System: Emotional Intelligence and Personalized Services, J. Syst. Manag. Sci. 13 (2023) 227–242.

[4] S. Pan, G.J.W. Xu, K. Guo, S.H. Park, H. Ding, Cultural insights in souls-like games: analyzing player behaviors, perspectives, and emotions across a multicultural context, IEEE Trans. Games (2024).

[5] S. Tian, L. Li, W. Li, H. Ran, X. Ning, P. Tiwari, A survey on few-shot class-incremental learning, Neural Networks 169 (2024) 307–324.

[6] H. Pan, Y. Wang, Z. Li, X. Chu, B. Teng, H. Gao, A complete scheme for multi-character

classification using EEG signals from speech imagery, IEEE Trans. Biomed. Eng. (2024).

[7] C. Zuo, X. Zhang, L. Yan, Z. Zhang, GUGEN: Global User Graph Enhanced Network for Next POI Recommendation, IEEE Trans. Mob. Comput. (2024).

[8] W. Song, X. Wang, S. Zheng, S. Li, A. Hao, X. Hou, TalkingStyle: Personalized Speech-Driven 3D Facial Animation with Style Preservation, IEEE Trans. Vis. Comput. Graph. (2024).

[9] M.G. San-Emeterio, A survey on few-shot techniques in the context of computer vision applications based on deep learning, in: Int. Conf. Image Anal. Process., Springer, 2022: pp. 14–25.

[10] S. Chen, W. Wang, X. Chen, M. Zhang, P. Lu, X. Li, Y. Du, Enhancing Chinese comprehension and reasoning for large language models: an efficient LoRA fine-tuning and tree of thoughts framework, J. Supercomput. 81 (2025) 50.

[11] X. Long, C. Tian, Spatial and channel attention-based conditional Wasserstein GAN for direct and rapid image reconstruction in ultrasound computed tomography, Biomed. Eng. Lett. 14 (2024) 57–68.

[12] W. Xu, Q.-W. Xing, Y. Yu, L.-Y. Zhao, Exploring the influence of gamified learning on museum visitors' knowledge and career awareness with a mixed research approach, Humanit. Soc. Sci. Commun. 11 (2024) 1–13.

[13] D. Li, W. Jianxing, The effect of gamified learning monitoring systems on Students' learning behavior and Achievement: An empirical study, Entertain. Comput. (2024) 100907.

[14] H. Xie, Z. Gao, G. Jia, S. Shimoda, Q. Shi, Learning rat-like behavioral interaction using a small-scale robotic rat, Cyborg Bionic Syst. 4 (2023) 32.

[15] H.F. Garcia, A. Aguilar, E. Manilow, B. Pardo, Leveraging hierarchical structures for few-shot musical instrument recognition, ArXiv Prepr. ArXiv2107.07029 (2021).

[16] S. Gidaris, N. Komodakis, Dynamic few-shot visual learning without forgetting, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018: pp. 4367–4375.

[17] C. Cao, Y. Li, Q. Lv, P. Wang, Y. Zhang, Few-shot action recognition with implicit temporal alignment and pair similarity optimization, Comput. Vis. Image Underst. 210 (2021) 103250.

[18] Y. Zhu, W. Min, S. Jiang, Attribute-guided feature learning for few-shot image recognition,

IEEE Trans. Multimed. 23 (2020) 1200–1209.

[19] G. Wang, W. Ye, X. Wang, R. Jin, H. Wang, Visual tempo contrastive learning for few-shot action recognition, in: 2022 IEEE Int. Conf. Image Process., IEEE, 2022: pp. 1096–1100.

[20] S. Feng, M.F. Duarte, Few-shot learning-based human activity recognition, Expert Syst. Appl. 138 (2019) 112782.

[21] H.-J. Ye, H. Hu, D.-C. Zhan, F. Sha, Few-shot learning via embedding adaptation with set-to-set functions, in: Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2020: pp. 8808–8817.

[22] B.L. Sturm, An analysis of the GTZAN music genre dataset, in: Proc. Second Int. ACM Work. Music Inf. Retr. with User-Centered Multimodal Strateg., 2012: pp. 7–12.

[23] N. Karunakaran, A. Arya, A scalable hybrid classifier for music genre classification using machine learning concepts and spark, in: 2018 Int. Conf. Intell. Auton. Syst., IEEE, 2018: pp. 128–135.

[24] R.C.P. Kumar, J. Elfreda, Audio retrieval based on cepstral feature, Int. J. Comput. Appl. 107 (2014) 28–33.

[25] F. Han, P. Yang, H. Du, X.-Y. Li, Accuth $^+ $: Accelerometer-based Anti-Spoofing Voice Authentication on Wrist-worn Wearables, IEEE Trans. Mob. Comput. (2023).

[26] A. Mohammadzadeh, H. Taghavifar, Y. Zhang, W. Zhang, A Fast Nonsingleton Type-3 Fuzzy Predictive Controller for Nonholonomic Robots Under Sensor and Actuator Faults and Measurement Errors, IEEE Trans. Syst. Man, Cybern. Syst. (2024).

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

[28] S.-R. Yan, W. Guo, A. Mohammadzadeh, S. Rathinasamy, Optimal deep learning control for modernized microgrids, Appl. Intell. 53 (2023) 15638–15655.

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

[30] F. Pahde, M. Puscas, T. Klein, M. Nabi, Multimodal prototypical networks for few-shot

learning, in: Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis., 2021: pp. 2644–2653.

[31] Y. Gong, Y. Yue, W. Ji, G. Zhou, Cross-domain few-shot learning based on pseudo-Siamese neural network, Sci. Rep. 13 (2023) 1427.

[32] L. Zhang, L. Zuo, Y. Du, X. Zhen, Learning to adapt with memory for probabilistic few-shot learning, IEEE Trans. Circuits Syst. Video Technol. 31 (2021) 4283–4292.

[33] Z. Ji, X. Chai, Y. Yu, Y. Pang, Z. Zhang, Improved prototypical networks for few-shot learning, Pattern Recognit. Lett. 140 (2020) 81–87.

[34] Z. Wang, P. Ma, Z. Chi, D. Li, H. Yang, W. Du, Multi-attention mutual information distributed framework for few-shot learning, Expert Syst. Appl. 202 (2022) 117062.

[35] D. Wang, X. Xian, H. Li, D. Wang, Distribution-Agnostic Probabilistic Few-Shot Learning for Multimodal Recognition and Prediction, IEEE Trans. Autom. Sci. Eng. (2024).

[36] B. Zhang, C. Luo, D. Yu, X. Li, H. Lin, Y. Ye, B. Zhang, MetaDiff: Meta-Learning with Conditional Diffusion for Few-Shot Learning, in: Proc. AAAI Conf. Artif. Intell., 2024: pp. 16687–16695.

[37] S. Wang, J. Yue, J. Liu, Q. Tian, M. Wang, Large-scale few-shot learning via multi-modal knowledge discovery, in: Comput. Vision–ECCV 2020 16th Eur. Conf. Glas. UK, August 23–28, 2020, Proceedings, Part X 16, Springer, 2020: pp. 718–734.

[38] J. Snell, K. Swersky, R. Zemel, Prototypical networks for few-shot learning, Adv. Neural Inf. Process. Syst. 30 (2017).

[39] X. Luo, H. Wu, J. Zhang, L. Gao, J. Xu, J. Song, A Closer Look at Few-shot Classification Again, ArXiv Prepr. ArXiv2301.12246 (2023).

[40] X. Li, M. Khishe, L. Qian, Evolving deep gated recurrent unit using improved marine predator algorithm for profit prediction based on financial accounting information system, Complex Intell. Syst. (2023) 1–17.

[41] L. Qian, J. Bai, Y. Huang, D.Q. Zeebaree, A. Saffari, D.A. Zebari, 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 87 (2024) 105492.

[42] Y. Zhang, Y. Guo, Y. Jin, Y. Luo, Z. He, H. Lee, Unsupervised discovery of object landmarks as structural representations, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018: pp. 2694–2703.