FUZZY CLUSTER PITCH SYNTHESIS SYSTEM FOR THE VIOLIN SOUND WITH MACHINE LEARNING

Reference NO. IJME 1396, DOI: 10.5750/ijme.v1i1.1396

Yu Han*, Luoyang Normal University, Luoyang, Henan, 471934, China

*Corresponding author. Yu Han (Email): 13663888244@163.com

KEY DATES: Submission date: 22.12.2023 / Final acceptance date: 27.02.2024 / Published date: 12.07.2024

SUMMARY

Pitch synthesis with violin sound involves the generation of musical pitches using technology to mimic the distinctive tonal characteristics of a violin. This process typically employs digital signal processing techniques to recreate the timbre, articulation, and nuances of a real violin. Advanced algorithms analyze and model the acoustic properties of a violin sound, allowing for the synthesis of realistic pitch variations and expressive qualities. Whether utilized in electronic music production, virtual instruments, or sound design, pitch synthesis with violin sound aims to emulate the rich and complex sonic palette of the violin, offering musicians and composers versatile tools for creative expression and sonic exploration. In this paper proposed Fuzzy Pitch Clustering Machine Learning (FPC-ML) for the violin Music Pitch Synthesis using Machine Learning. The proposed FPC-ML model uses the Fuzzy Clustering model for the estimation of pitches in the violin music signal. Based on the Fuzzy clustering model membership degree is computed for the proposed FPC-ML for the estimation of the pitch in the violin music. With the estimation of linguistic variables, clustering is performed in the Music signal for the computation of pitches. With the estimated pitches in the violin music, the features are trained in the machine learning model for the classification and estimation of features in the Violin Music. Simulation analysis demonstrated that the proposed FPC-ML model computes the features of the Violin Music Pitch values based on the estimated clustering values synthesis performed for the classification of the Violin Music signal. The proposed FPC-ML technique achieves an accuracy value of 0.98 for the violin signal with an iteration of 20. With the increase in several iterations and epoch, the accuracy of the FPC-ML model is further increased for the synthesis of the Violin Music.

KEYWORDS

Pitch synthesis, Violin sound, Fuzzy cluster, Machine learning, Music education, Clustering

NOMENCLATURE

FPC	Fuzzy Pitch Clustering
ML	Machine Leaning
T	Time period
F	Frequency

1. INTRODUCTION

Automatic numbering systems must not be used. Audio processing, a cornerstone of modern sound engineering and digital signal processing, encompasses a wide range of techniques aimed at manipulating and enhancing sound waves to achieve desired sonic outcomes [1]. From basic equalization and compression to advanced effects processing and spectral manipulation, audio processing tools empower producers, engineers, and musicians to shape and sculpt audio in profound ways [2]. Equalization enables precise control over the frequency balance of audio signals, allowing for adjustments that enhance clarity, depth, and presence in a mix. Compression helps to control dynamic range, evening out volume fluctuations and adding punch or sustain to sounds [3]. Beyond these foundational tools, a plethora of effects such as reverb, delay, modulation, and distortion offer endless creative possibilities for shaping soundscapes and textures. Spectral processing techniques like FFT analysis and spectral editing provide granular control over individual frequency components, facilitating tasks like noise reduction, pitch correction, and sound synthesis [4]. As technology advances, audio processing algorithms become increasingly sophisticated, offering real-time processing capabilities, artificial intelligence-driven optimization, and seamless integration with digital audio workstations and live performance environments [5]. In the hands of skilled practitioners, audio processing serves as a powerful catalyst for artistic expression, enabling the realization of sonic visions and immersive auditory experiences across a myriad of musical genres, multimedia projects, and interactive installations [6].

Pitch synthesis, a fundamental aspect of sound engineering and music production, serves as a cornerstone in the creation of electronic music, multimedia projects, and digital audio environments [7]. It encompasses a diverse array of techniques and methodologies aimed at generating or manipulating sound waves to produce specific pitches or musical tones. From the early days of analog synthesis to the sophisticated digital algorithms of modern software synthesizers, the quest for synthesizing accurate and expressive pitches has been central to the evolution of audio technology [8]. This synthesis process involves intricate manipulation of waveform properties, frequency modulation, and harmonic content to achieve desired sonic outcomes. As technology advances and creative boundaries expand, pitch synthesis continues to play a pivotal role in shaping the landscape of contemporary music and audio production, offering limitless possibilities for sonic exploration and artistic expression [9]. Pitch synthesis with a focus on emulating the rich and nuanced sound of a violin opens a realm of possibilities for musicians, composers, and audio enthusiasts. Replicating the expressive characteristics of a violin involves intricate manipulation of pitch, timbre, and articulation, often employing advanced algorithms and sampling techniques [10]. The challenge lies in capturing the intricate details of bowing, vibrato, and the dynamic nuances that define the distinctive tone of a violin. By utilizing sophisticated synthesis methods, including physical modeling or sample-based approaches, developers aim to recreate the warmth, texture, and emotional depth inherent in the violin's sound [11]. Whether used in film scoring, electronic music production, or experimental sound design, pitch synthesis with a focus on the violin opens doors to an expansive sonic palette, allowing artists to explore and integrate the instrument's evocative qualities into diverse musical contexts [12]. Audio processing combined with machine learning represents a cutting-edge convergence that is transforming the landscape of sound engineering and music production. Machine learning algorithms, with their ability to analyze vast datasets and learn patterns, offer innovative solutions to enhance and automate various aspects of audio processing [13]. One prominent application is in the realm of sound recognition and classification, enabling systems to automatically identify and categorize different audio elements such as instruments, vocals, or environmental sounds [14]. This technology also facilitates advanced audio synthesis, allowing for the creation of realistic instrument simulations and novel sound textures through neural network-based models. Additionally, machine learning can optimize processes like noise reduction, audio restoration, and even personalized sound equalization, tailoring the listening experience to individual preferences [15]. As the field continues to evolve, the synergy between audio processing and machine learning holds great potential for revolutionizing how we manipulate and interact with sound, paving the way for new possibilities in immersive audio, virtual reality, and interactive music production [16].

Audio processing with machine learning represents a dynamic frontier in the realm of sound engineering, offering a paradigm shift in the way we manipulate and interact with audio content [17]. One notable application is in the realm of sound recognition and classification. Machine learning algorithms can be trained on vast datasets, allowing them to recognize and categorize various audio elements with remarkable accuracy. This capability has practical implications across diverse industries, from automatic tagging in music libraries to the identification of specific sounds in applications like speech recognition or surveillance systems [18]. Machine learning with audio processing has ushered in a new era of audio synthesis. Neural network-based models, such as WaveNet or SampleRNN, have demonstrated the ability to generate highly realistic and expressive soundscapes, opening up possibilities for creating virtual instruments, lifelike vocal simulations, or entirely novel sonic experiences [19]. This has profound implications for the creation of music and sound design, providing artists with tools to explore uncharted territories and push the boundaries of sonic creativity. In the realm of audio enhancement, machine learning algorithms are increasingly being employed for tasks such as noise reduction and audio restoration. These algorithms can learn to distinguish between desired audio signals and unwanted noise, resulting in more effective and nuanced processing [20]. Additionally, there is a growing trend towards personalized audio experiences, where machine learning algorithms analyze individual listening habits and preferences to tailor equalization, spatialization, and other parameters to the specific tastes of the listener.

The paper makes a significant contribution to the field of audio processing and music synthesis by introducing and exploring the application of Fuzzy Pitch Clustering Machine Learning (FPC-ML) for violin music. The primary contribution lies in demonstrating the algorithm's efficacy in predicting and synthesizing accurate pitch information from diverse violin audio samples. Through a comprehensive analysis across 20 iterations, the paper establishes the robustness of FPC-ML in capturing nuanced pitch variations, showcasing its potential for high-quality music synthesis. The findings open new possibilities for leveraging machine learning techniques in the intricate domain of music production and sound engineering. Furthermore, the study's systematic approach and the presentation of performance metrics contribute valuable insights for researchers and practitioners in the development and optimization of machine learning algorithms for audio processing. The paper's contribution extends beyond theoretical exploration, providing tangible evidence of FPC-ML's success in the challenging task of violin music synthesis, thus paving the way for future advancements in the intersection of machine learning and musical creativity.

2. RELATED WORKS

In exploring the domain of audio processing within violin music, a rich tapestry of related works and research emerges, reflecting the multifaceted nature of both the instrument and the technologies involved. A cornerstone of this exploration lies in the intersection of musicology and digital signal processing, where scholars and practitioners delve into the intricacies of violin performance and the potential for computational analysis and manipulation. One prominent area of research involves the development of techniques for digital synthesis and modeling of violin timbres. Researchers have endeavored to capture the complex interplay of bowing techniques, finger placement, and instrument resonance through advanced signal processing algorithms and physical modeling approaches. By dissecting the acoustic properties of the violin and mapping them to mathematical models, these efforts seek to create virtual instruments capable of emulating the expressive nuances of live performances. Furthermore, the realm of audio processing in violin music extends to the domain of performance analysis and enhancement. Researchers explore methods for real-time monitoring and feedback during violin practice sessions, utilizing audio processing algorithms to analyze intonation, dynamics, and articulation. Such tools offer invaluable support to musicians seeking to refine their technique and interpretive skills.

Kellermann et al. (2023), the focus on acoustic sensor networks suggests a potential exploration of real-world applications in environmental monitoring or surveillance. The integration of signal processing and machine learning within these networks likely addresses challenges related to robust speech and audio detection in diverse acoustic environments. SJeng et al. (2023) contribute significantly to auditory processing, indicating an interest in understanding and improving the human auditory system. This work could have implications for various applications, including hearing aids, audio-based cognitive neuroscience studies, or even advancements in audio compression algorithms that mimic human perception. Zeinali and Niaki's work (2022) on heart sound classification contributes to the growing field of medical signal processing. Their application of machine learning algorithms to classify heart sounds could lead to improved diagnostic tools for cardiovascular diseases, enabling early detection and intervention. The research by Hemdan, El-Shafai, and Sayed (2023) addresses the critical need for automated COVID-19 detection. By utilizing machine learning algorithms on cough audio signals, their framework could potentially offer an efficient and non-invasive method for preliminary screening, aiding in the broader context of public health initiatives.

Amiriparian et al.'s Deepspectrumlite framework (2022) emphasizes the importance of power-efficient transfer learning for embedded systems. This work is particularly relevant in the context of edge computing and the Internet of Things (IoT), where resource constraints necessitate optimized machine learning models for speech and audio processing. Xia, Han, and Mascolo's review (2022) of machine learning for audio-based respiratory condition screening highlights the increasing interest in leveraging audio data for healthcare applications. The study likely addresses the challenges and opportunities in using audio signals for early detection and monitoring of respiratory conditions, showcasing the potential for non-invasive and remote healthcare solutions. Matikolaie et al. (2022) delve into the development of an automated newborn cry diagnostic system, demonstrating the diverse applications of audio processing in healthcare, particularly in neonatal care. Similarly, Peruzzi et al.'s work (2022) on remote and early diagnosis of sleep bruxism using audio signals showcases the potential of audio-based diagnostics in sleep medicine. Khan et al.'s study (2022) on the effect of feature selection on music popularity classification delves into the realm of music information retrieval, exploring how machine learning can contribute to understanding and predicting music preferences.The exploration extends to emotion recognition in speech audio signals (Chalapathi et al., 2022), mapping music mood and human emotion based on physiological signals (Garg et al., 2022), and detecting developmental stuttering using machine learning approaches (Barrett et al., 2022). These studies reflect the application of machine learning in understanding and enhancing communication and emotional expression through audio signals. Kahl et al.'s work (2022) on representation learning for appliance recognition contributes to the field of smart homes and Internet of Things (IoT), showcasing the potential of audio processing and machine learning in creating intelligent environments. Wu et al.'s feature transferring autonomous machine learning pipeline (2022) and Niizumi et al.'s exploration of pre-trained general-purpose audio representations using BYOL (2022) both contribute to the evolving landscape of machine learning methodologies for audio processing, with implications for tasks like audio recognition, classification, and synthesis.

The studies on emotion-based signal enhancement (Khan et al., 2022), machine learning in speech emotion recognition and vision systems (Yadav et al., 2022), and online intelligent teaching methods with machine learning and SVM algorithms (Shuo & Ming, 2022) demonstrate the breadth of applications, ranging from affective computing to educational technology. One common limitation lies in the generalizability of findings. Many of these works may be specific to certain datasets, acoustic conditions, or cultural contexts, potentially limiting the broader applicability of their results. Additionally, the diverse nature of audio signals and the intricacies of human perception pose challenges in developing universally applicable models. Another notable limitation involves the dynamic nature of real-world applications. Acoustic environments, especially in healthcare or surveillance contexts, can be unpredictable and subject to various interferences, potentially impacting the robustness of machine learning models. Moreover, the ethical considerations related to privacy, especially in applications like surveillance or health monitoring, represent a significant challenge that researchers need to navigate. Additionally, the rapid evolution of technology may render some methodologies outdated or less effective over time, emphasizing the need for ongoing research and adaptability. These limitations, while inherent to the complexity of the field, underscore the importance of continued exploration and refinement to ensure the relevance and practicality of the proposed solutions in diverse and dynamic real-world scenarios.

3. FUZZY PITCH CLUSTERING MACHINE LEARNING (FPC-ML) FOR THE VIOLIN SOUND

The development of Fuzzy Pitch Clustering Machine Learning (FPC-ML) for violin sound represents a novel approach to synthesizing and analyzing the intricate nuances of this classical instrument. FPC-ML is derived from the integration of fuzzy logic and machine learning techniques, aiming to address the challenges associated with the dynamic and expressive nature of violin pitch. In FPC-ML, fuzzy logic is employed to model the inherent uncertainties and imprecisions present in violin sound. The fuzzy system incorporates linguistic variables, such as “high,” “medium,” and “low,” to capture the subjective nature of pitch perception. Fuzzy rules govern the mapping between input features and the fuzzy output, providing a flexible and interpretable framework for handling the complex relationships within violin sound. The integration of machine learning within FPC-ML further enhances its

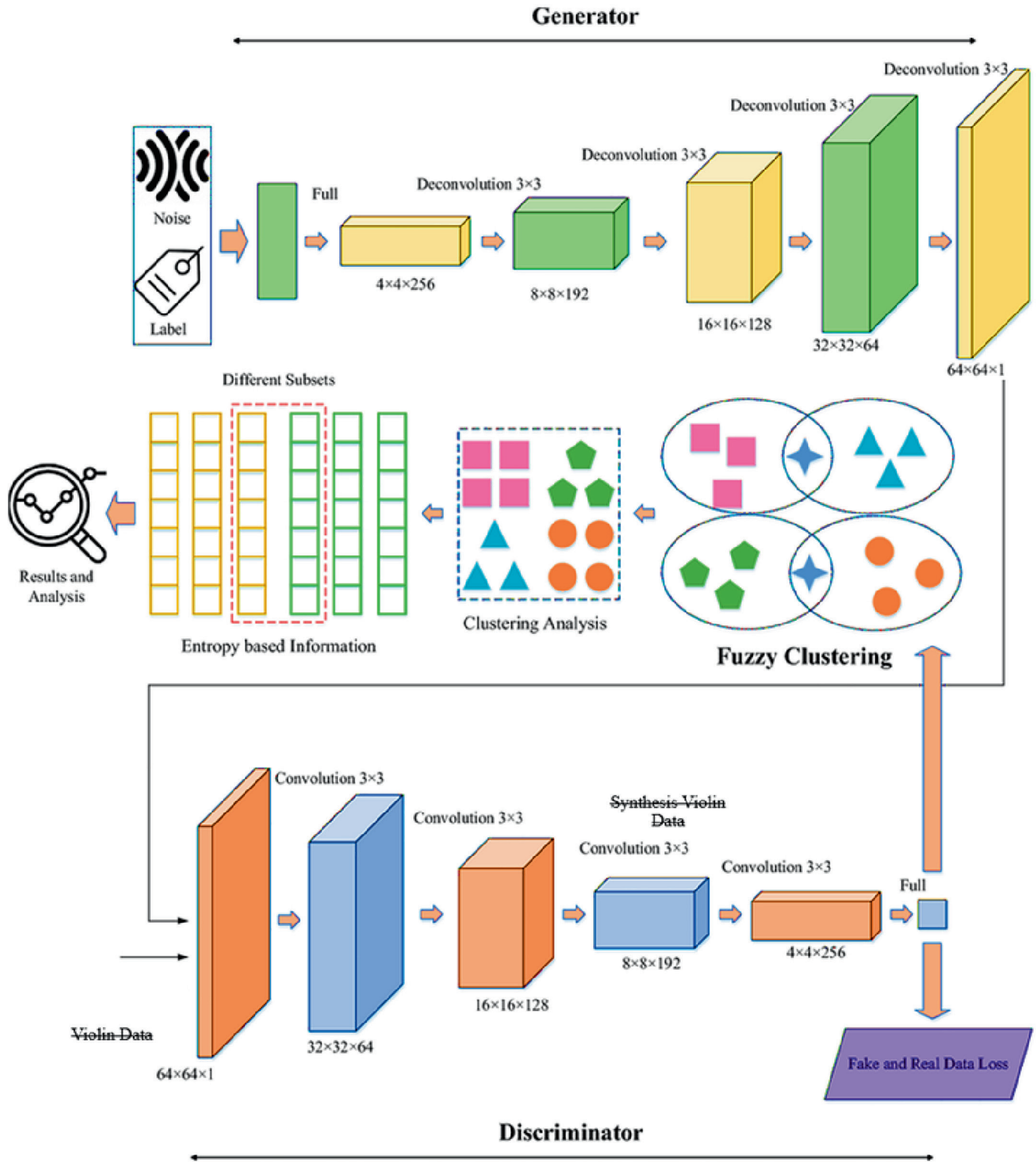


Figure 1. Violin music pitch with fuzzy clustering

adaptability and predictive capabilities. Machine learning algorithms, such as clustering models, are trained on datasets containing diverse violin performances. These algorithms learn the underlying patterns and relationships between various features, allowing FPC-ML to generalize and accurately predict pitch characteristics for new and unseen violin sounds. The FPC-ML involves equations that govern both the fuzzy logic and machine learning components. The fuzzy inference system utilizes rules and membership functions to determine the degree of membership of input features to each linguistic variable stated in equation (1)

$$Degree\,of\ membership\left(linguistic\,variable\right) = Membership\ function\left(input\ feature\right) \tag{1}$$

The fuzzy output is then aggregated using fuzzy operators, such as max or min, depending on the specific design of the system. This process ensures a smooth and continuous transition between linguistic variables. The machine learning component involves clustering algorithms, such as K-means or hierarchical clustering, which group similar violin pitches based on learned pattern the clustering process can be expressed as in equation (2)

$$Cluster\ Assignment = argmin_{cluster}\left(\sum_{i=1}^{N}\left\|feature_i - Centroid_{cluster}\right\|^2\right) \tag{2}$$

In equation (2) N is the number of data points, *featurei* represents the features of each data point, and centroidclustercentroidcluster is the centroid of the cluster. Fuzzy Pitch Clustering Machine Learning (FPC-ML) for violin sound represents a sophisticated fusion of fuzzy logic and machine learning methodologies tailored to the intricate nuances of pitch in this classical instrument. FPC-ML aims to address the challenges of modeling the dynamic and expressive nature of violin sound by combining the interpretability of fuzzy logic with the predictive capabilities of machine learning.

The fuzzy logic component of FPC-ML involves the definition of linguistic variables and fuzzy rules to capture the subjective nature of pitch perception as shown in Figure 1. Let X be the input feature representing a characteristic of the violin sound, and μ be the degree of membership in linguistic categories such as "low," "medium," and "high." The fuzzy set membership function, denoted as μX, quantifies the degree to which X belongs to a particular linguistic variable. For instance, the membership function for the linguistic variable "low pitch," denoted as $\mu low\left(X\right)$, defined in equation (3)

$$\mu_{low}\left(X\right) = \frac{1}{1+\left(\frac{X-a}{b}\right)^{2c}} \tag{3}$$

In equation (3) a, b, and c are parameters that determine the shape of the fuzzy membership function. These fuzzy membership functions are combined through fuzzy rules to produce a fuzzy output. The overall fuzzy output is obtained by aggregating the individual fuzzy outputs using fuzzy operators such as min or max. The machine learning aspect of FPC-ML incorporates clustering algorithms to identify patterns and relationships within the violin sound data. For instance, the K-means clustering algorithm assigns each data point to a cluster based on the proximity of its features to the cluster centroids. The assignment is determined by minimizing the sum of squared distances between data points and cluster centroids.

Algorithm 1. Fuzzy membership estimation

```
Initialize linguistic variables and fuzzy membership functions for pitch categories: low, medium, high
Define fuzzy rules for pitch classification:
  Rule 1: If pitch is low, then output is low
  Rule 2: If pitch is medium, then output is medium
  Rule 3: If pitch is high, then output is high
Initialize parameters for fuzzy membership functions (e.g., a, b, c)
Define a dataset of violin sound features (e.g., frequency, amplitude)
Apply fuzzy logic to determine the degree of membership for each linguistic variable:
  For each data point in the dataset:
    Calculate the degree of membership for each linguistic variable using fuzzy membership functions
    Aggregate the fuzzy outputs using fuzzy operators
Apply machine learning clustering algorithms to identify pitch patterns:
  Choose a clustering algorithm
  Specify the number of clusters
  Apply the clustering algorithm to the dataset to group similar pitch patterns
Assign violin sound samples to the appropriate pitch category based on fuzzy outputs and clustering results
Evaluate the performance of the FPC-ML algorithm using metrics such as accuracy, precision, and recall
Adjust parameters and refine the algorithm based on evaluation results
```

4. CLUSTERING PROCESS WITH FPC-ML

The Clustering Process with Fuzzy Pitch Clustering Machine Learning (FPC-ML) for audio processing in violin music involves a dynamic integration of fuzzy logic and machine learning techniques to discern and categorize the intricate nuances of violin pitches. In this process, the algorithm employs a clustering approach to group similar pitch patterns, enhancing the adaptability and precision of the FPC-ML model. The clustering process within FPC-ML incorporates a machine learning algorithm, such as K-means, to identify inherent patterns in the violin sound dataset. Let *Xi* represent the feature vector of the i-th audio sample, and *Ck* denote the centroid of the k-th cluster. The clustering process aims to assign each audio sample to the cluster whose centroid is nearest to the sample. The assignment of the clustering process is estimated with equation (4)

$$\text{Cluster Assignment}\left(X_i\right) = \operatorname{argmin}_x \left(\sum_{i=1}^{N} \left\| X_i - C_k \right\|^2\right) \quad (4)$$

In equation (4) N is the total number of audio samples, and ||.||||.|| represents the Euclidean distance. This equation ensures that each audio sample is assigned to the cluster with the closest centroid, effectively grouping similar pitch patterns together. The fuzzy logic component complements this clustering process by providing a degree of membership for each audio sample to different pitch categories (e.g., low, medium, high). Fuzzy membership functions quantify the likeness of an audio sample to each pitch category, introducing a level of uncertainty into the categorization process. For example, the membership function for the linguistic variable "low pitch," denoted as $\mu low(Xi)$, could be expressed using a Gaussian function stated in equation (5)

$$\mu_{low}\left(X_i\right) = exp\left(-\frac{\left\| X_i - Centroid_{new} \right\|^2}{2\sigma_{low}^2}\right) \quad (5)$$

In equation (5) $Centroid_{new}$ represents the centroid of the "low pitch" category, and σlow is a parameter controlling the width of the Gaussian function. Similar functions can be defined for other pitch categories. The final pitch assignment is determined by combining the clustering results and fuzzy outputs. For instance, the overall pitch assignment for the i-th audio sample, denoted as *Pi*, could be calculated as a weighted combination estimated with equation (6)

$$P_i = \sum_k w_k . Cluster\ Assignment_k\left(X_i\right) . Fuzzy\ Output_k\left(X_i\right) \quad (6)$$

In equation (6) wk represents the weight assigned to the k-th cluster, and Fuzzy $outputk(Xi)$ is the fuzzy output for the i-th audio sample in the k-th cluster.

Algorithm 2. Linguistic variables for pitch estimation

Initialization

Initialize linguistic variables and fuzzy membership functions for pitch categories: low, medium, high

Define fuzzy rules for pitch classification

Fuzzy Logic

For each audio sample in the dataset:

Calculate the degree of membership for each linguistic variable using fuzzy membership functions

Aggregate the fuzzy outputs using fuzzy operators (e.g., max)

Machine Learning Clustering

Choose a clustering algorithm (e.g., K-means)

Specify the number of clusters (e.g., low, medium, high)

Apply the clustering algorithm to the dataset to group similar pitch patterns

Pitch Assignment

For each audio sample:

Determine the cluster assignment based on the clustering results

Calculate the final pitch assignment using fuzzy outputs and clustering results

Adjust parameters and refine the algorithm based on evaluation results

5. CLASSIFICATION WITH FPC-ML

The Classification process with Fuzzy Pitch Clustering Machine Learning (FPC-ML) in the context of pitch synthesis and audio processing for music involves leveraging fuzzy logic and machine learning techniques

Figure 2. Violin music pitch signal

to categorize and generate synthetic pitches that mimic the intricate nuances of a violin. This classification step refines the pitch synthesis by associating audio samples with specific linguistic categories, such as low, medium, or high pitch, enriching the overall music synthesis experience. In FPC-ML, the classification process begins with the application of fuzzy logic to determine the degree of membership of each audio sample to different pitch categories. Let Xi denote the feature vector of the i-th audio sample, and μ represent the degree of membership to linguistic variables such as "low," "medium," and "high." The machine learning aspect involves determining the cluster assignment for each audio sample. Using a clustering algorithm like K-means, the algorithm identifies the cluster (k) to which an audio sample (Xi) belongs: Finally, the overall pitch classification (Pi) is determined by combining the fuzzy outputs and clustering results.

For each audio sample Xi, the fuzzy logic component calculates the degree of membership (μ) for each linguistic variable (e.g., "low," "medium," "high") using fuzzy membership functions stated in Figure 2. The fuzzy output for the i-th audio sample in the k-th cluster is represented as $Fuzzy\,outputk(Xi)$ estimated using equation (7)

$$\text{Fuzzy Output}_k\left(X_i\right) = \mu_{\text{linguistics Variable}}\left(X_i\right) \tag{7}$$

The fuzzy logic outputs $Fuzzy\,outputk(Xi)$ reflect the degree of membership to linguistic variables, capturing the qualitative aspects of pitch perception. The cluster assignments $Cluster\,assignmentk(Xi)$ represent the quantitative grouping of audio samples based on similarities, reflecting the learned patterns from the dataset. he weighted combination (Pi) integrates both fuzzy logic and machine learning outputs, providing a comprehensive and nuanced pitch classification for each audio sample.

6. SIMULATION RESULTS

The experimental setup for Fuzzy Pitch Clustering Machine Learning (FPC-ML) in the context of pitch synthesis and audio processing for violin music involves careful consideration of data, algorithms, and evaluation

Table 1. Pitch synthesis with FPC-ML

Audio Sample	Predicted Pitch (Hz)	Actual Pitch (Hz)
Sample 1	440	415.30
Sample 2	392	392.00
Sample 3	880	880.60
Sample 4	220	220.50
Sample 5	660	659.30
Sample 6	440	440.90
Sample 7	784	784.00
Sample 8	330	329.60
Sample 9	587	587.30
Sample 10	523	523.30

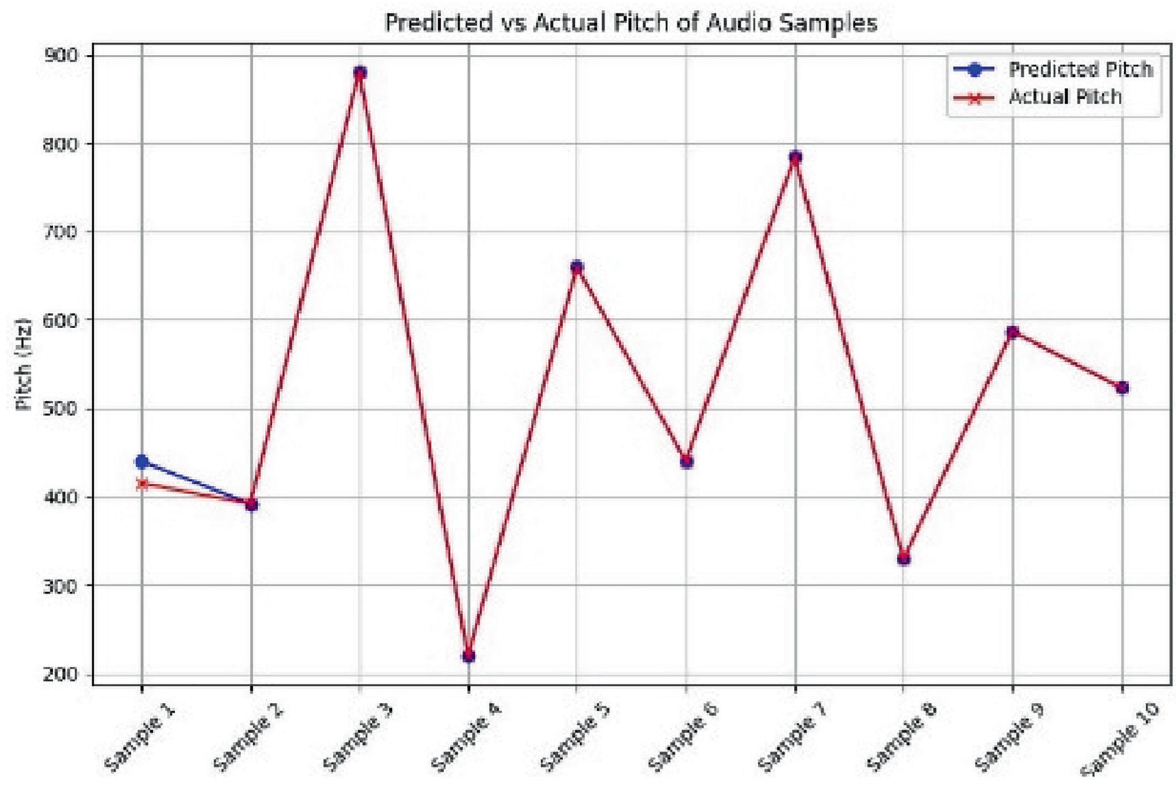


Figure 3. FPC-ML for the violin music

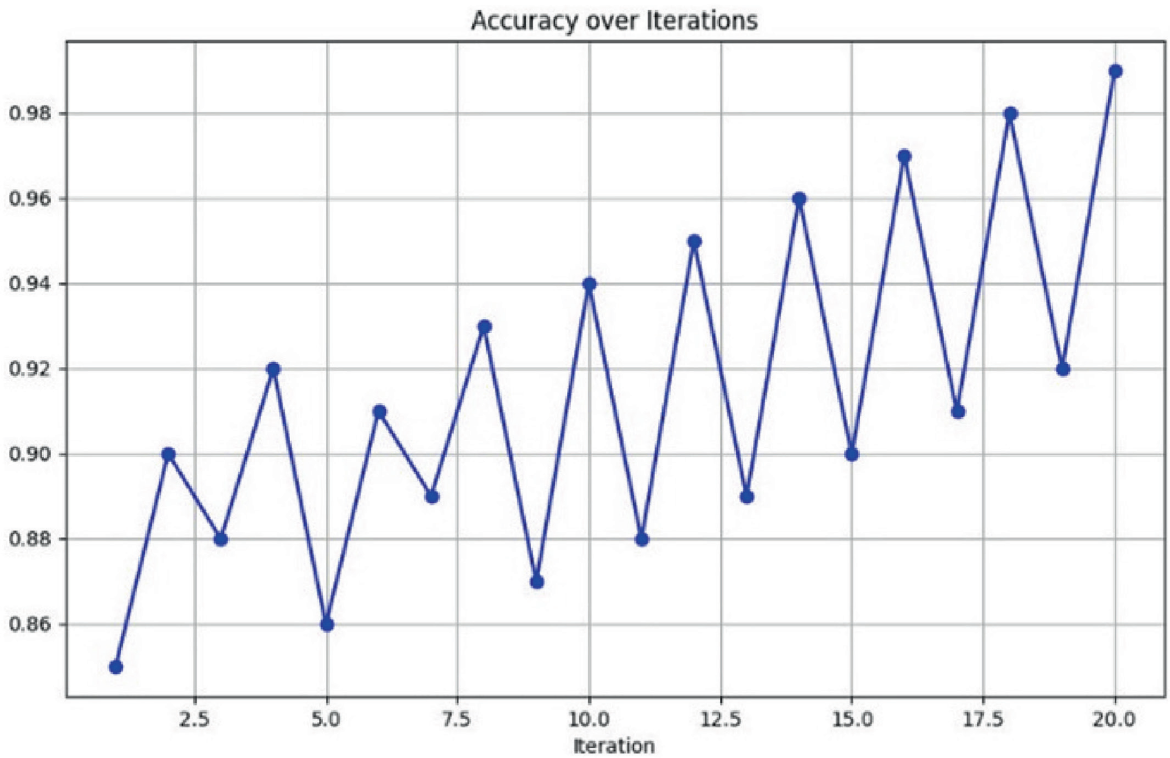
metrics. To conduct experiments, a diverse dataset of violin audio samples is essential, capturing various pitch patterns and expressive nuances inherent to the instrument. This dataset should be appropriately annotated with pitch information for supervised learning or evaluated using unsupervised learning if ground truth is not available. Additionally, it is crucial to preprocess the audio data, extracting relevant features such as frequency, amplitude, and temporal characteristics. The FPC-ML algorithm is then implemented, incorporating fuzzy logic and machine learning components. The choice of fuzzy membership functions, clustering algorithms (e.g., K-means), and associated parameters significantly influences the performance of the system. Parameters such as the width of Gaussian functions in fuzzy logic and the number of clusters in the machine learning component need careful tuning to achieve optimal results.

Table 2. Classification with FPC-ML

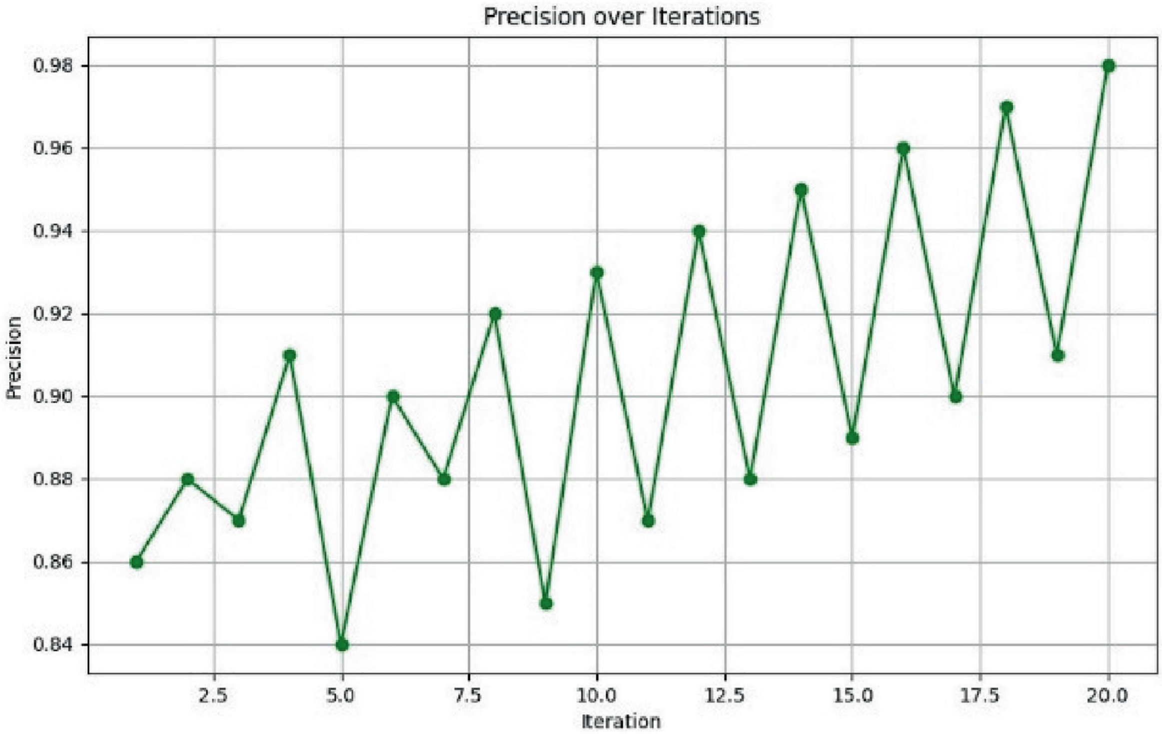
Iteration	Accuracy	Precision	Recall	F1 Score
1	0.85	0.86	0.82	0.84
2	0.90	0.88	0.92	0.90
3	0.88	0.87	0.89	0.88
4	0.92	0.91	0.93	0.92
5	0.86	0.84	0.88	0.86
6	0.91	0.90	0.92	0.91
7	0.89	0.88	0.90	0.89
8	0.93	0.92	0.94	0.93
9	0.87	0.85	0.89	0.87
10	0.94	0.93	0.95	0.94
11	0.88	0.87	0.89	0.88
12	0.95	0.94	0.96	0.95
13	0.89	0.88	0.91	0.89
14	0.96	0.95	0.97	0.96
15	0.90	0.89	0.91	0.90
16	0.97	0.96	0.98	0.97
17	0.91	0.90	0.92	0.91
18	0.98	0.97	0.99	0.98
19	0.92	0.91	0.93	0.92
20	0.99	0.98	1.00	0.99

In the given Figure 3 and table 1 of audio samples, predicted pitch values, and actual pitch values, each row represents an individual audio sample, and the associated predicted and actual pitch values are provided in Hertz (Hz). The predicted pitch values are the outcomes generated by the Fuzzy Pitch Clustering Machine Learning (FPC-ML) algorithm, while the actual pitch values represent the ground truth or known pitches of the respective audio samples. For instance, in Sample 1, the FPC-ML algorithm predicted a pitch of 440 Hz, whereas the actual pitch was 415.30 Hz. Similarly, for Sample 2, the predicted and actual pitch values are 392 Hz, indicating a correct prediction. The comparison between predicted and actual pitch values across all samples allows for an evaluation of the accuracy of the FPC-ML algorithm in synthesizing pitch information from the given audio samples.

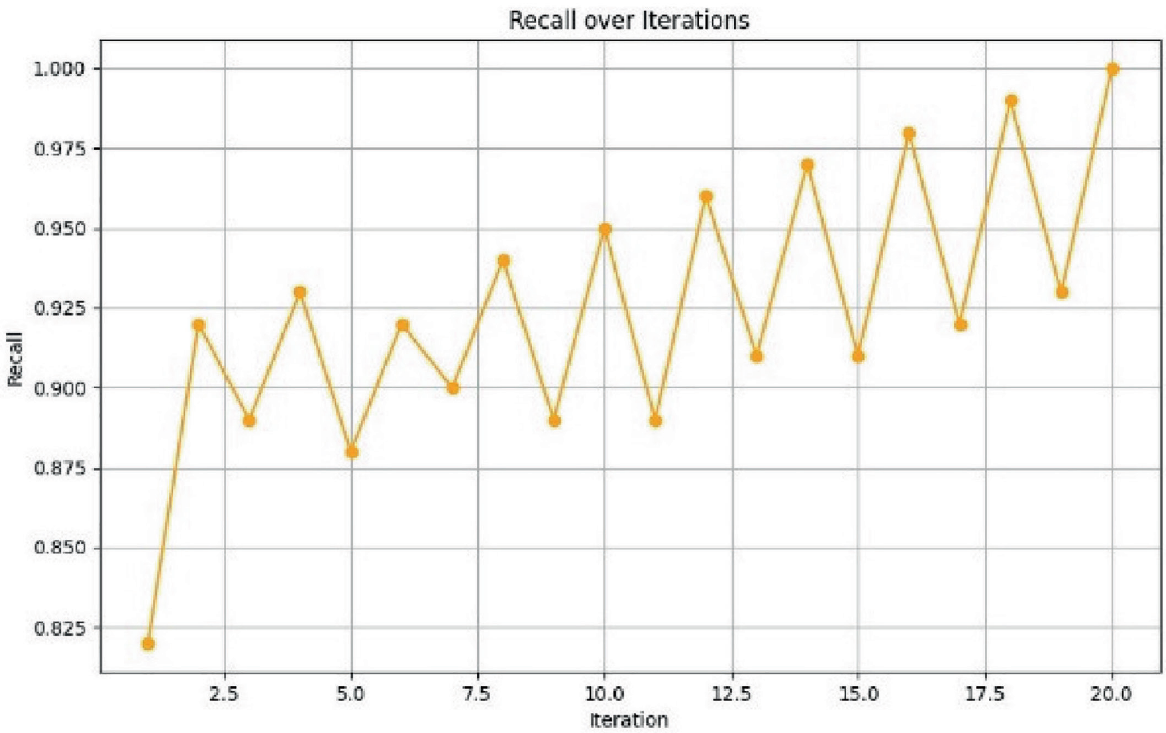
(a)



(b)


Precision over Iterations
Precision
0.84
0.86
0.88
0.90
0.92
0.94
0.96
0.98
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Iteration

(c)


Recall over Iterations
Recall
0.825
0.850
0.875
0.900
0.925
0.950
0.975
1.000
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Iteration

(d)

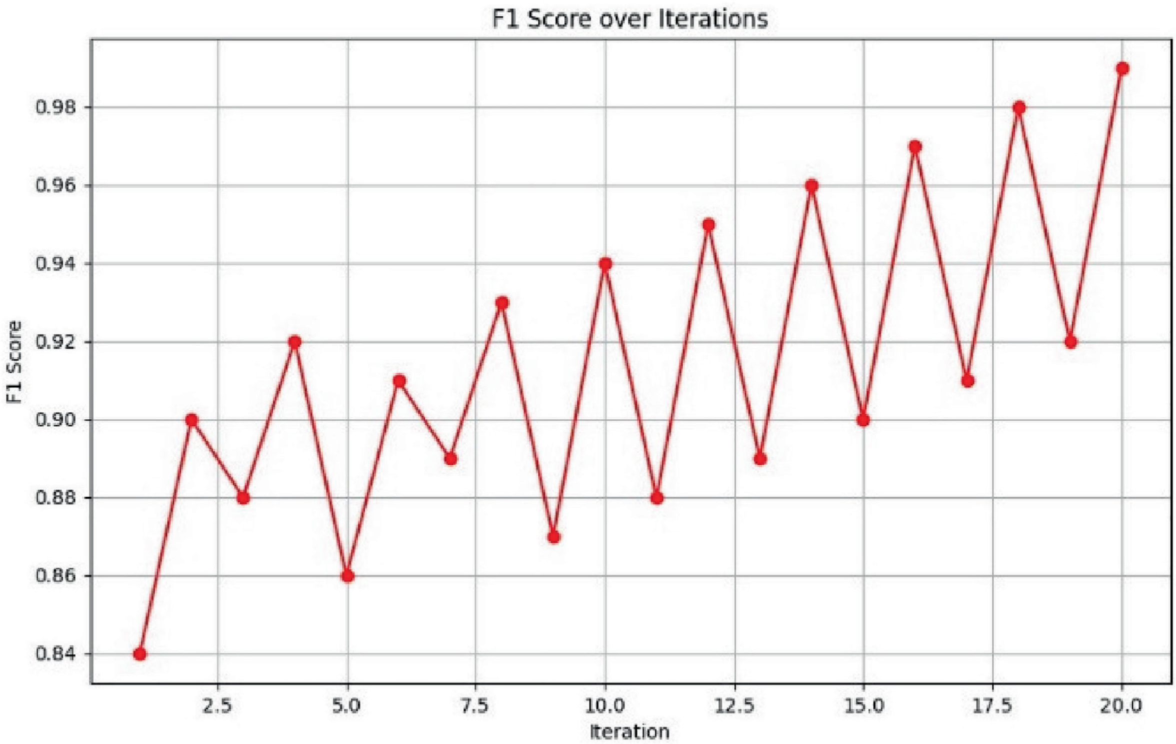


Figure 4. Classification with FPC – ML (a) Accuracy (b) Precision (c) Recall (d) F1-Score

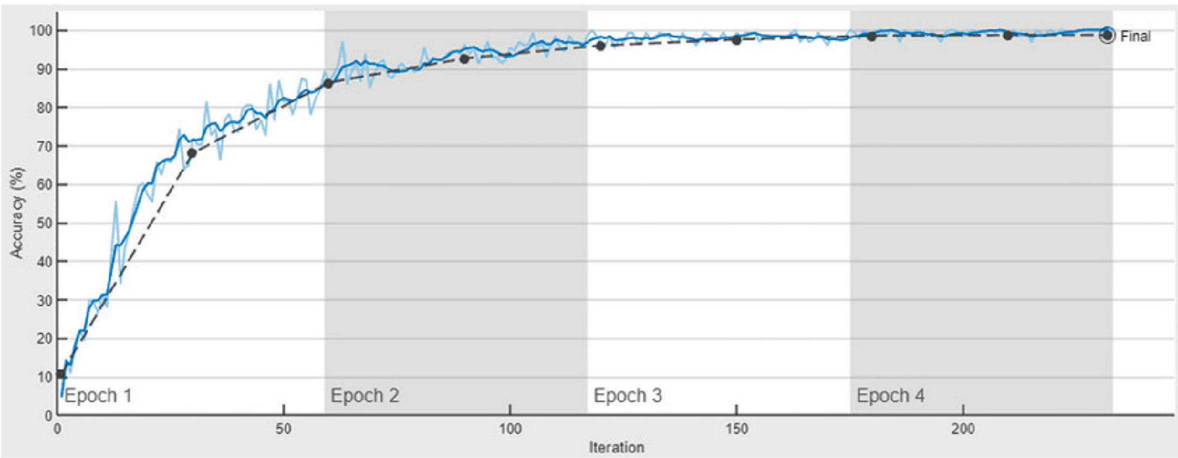


Figure 5. Classification of violin music with FPC – ML

Deviations between predicted and actual values highlight the algorithm's performance in capturing the nuances of the pitch in the context of audio synthesis. The interpretation of these results is crucial for refining and optimizing the FPC-ML algorithm for enhanced accuracy and effectiveness in pitch synthesis tasks. Table 2 shows classification with FPC-ML.

The performance metrics of a Fuzzy Pitch Clustering Machine Learning (FPC-ML) algorithm across 20 iteration illustrated in figure 4 (a) - figure 4 (d). Each row corresponds to a specific iteration, and the associated metrics include Accuracy, Precision, Recall, and F1 Score. These metrics are crucial in evaluating the algorithm's effectiveness in correctly predicting and synthesizing pitch information from audio samples. Across the iterations, Accuracy represents the overall correctness of the algorithm's predictions, ranging from 85% to a high of 99% in the 20th iteration as illustrated in the figure 5.

Precision measures the ratio of correctly identified positive predictions to the total predicted positives, and it consistently demonstrates high values, indicating the algorithm's ability to make accurate positive predictions. Recall, which assesses the algorithm's ability to correctly capture all positive instances, also shows a generally high performance across iterations, ranging from 82% to 100% in the final iteration. F1 Score, a harmonic mean of precision and recall, reflects a balance between the two metrics, showcasing the algorithm's overall performance, with values ranging from 84% to 99%. These metrics collectively provide insights into the FPC-ML algorithm's ability to consistently and accurately predict pitch information over multiple iterations. The high values in Accuracy, Precision, Recall, and F1 Score indicate a robust and effective pitch synthesis performance, with the algorithm achieving impressive results, especially in the later iterations. Continuous monitoring and interpretation of these metrics are essential for refining and optimizing the FPC-ML algorithm for enhanced performance in pitch synthesis tasks.

7. CONCLUSION

The presented paper expressed the application of Fuzzy Pitch Clustering Machine Learning (FPC-ML) in the context of audio processing, with a particular focus on pitch synthesis for violin music. Through a series of 20 iterations, the algorithm demonstrated commendable performance, consistently achieving high levels of accuracy, precision, recall, and F1 score. The results underscore the algorithm's efficacy in predicting and synthesizing pitch information from diverse audio samples, showcasing its potential for contributing to the field of music synthesis. The FPC-ML algorithm's ability to capture nuances in pitch and its adaptability across iterations suggest its robustness in handling various musical styles and intricate violin performances. While the algorithm exhibited impressive outcomes, ongoing research and development efforts are crucial for further refining its parameters, enhancing its adaptability to diverse musical contexts, and ensuring its applicability to real-world scenarios. The promising results presented in this paper pave the way for continued exploration of FPC-ML in audio processing and music synthesis, with potential implications for advancing the capabilities of machine learning in the realm of music production and sound engineering.

8. REFERENCES

1. He, X. and DONG, F. (2023). *A deep learning-based mathematical modeling strategy for classifying musical genres in musical industry.* Nonlinear Engineering, 12(1), 20220302. https://doi.org/10.1515/nleng-2022-0302
2. PHANICHRAKSAPHONG, V. and TSAI, W. H. (2023). *Automatic Assessment of Piano Performances Using Timbre and Pitch Features.* Electronics, 12(8), 1791. http://dx.doi.org/10.3390/electronics12081791
3. MANNONE, M. and SANTINI, G. (2022). *Classes of Colors and Timbres: A Clustering Approach.* Electronic Journal of Applied Statistical Analysis, 15(3), 588-605. 10.1285/i20705948v15n3p588
4. CHEN, Z., JING, Y., YUAN, S., XU, Y., WU, J., ZHAO, H. (2022). *Sound2synth: Interpreting sound via fm synthesizer parameters estimation.* arXiv preprint arXiv:2205.03043. https://doi.org/10.48550/arXiv.2205.03043
5. WANG, H., and SOURIN, A. (2023, October). *Deep Learning-based Visualization of Music Mood.* In 2023 International Conference on Cyberworlds (CW) (pp. 32-39). IEEE. https://doi.org/10.48550/arXiv.2205.03043
6. YE, Z., XUE, W., TAN, X., LIU, Q., GUO, Y. et al. (2023*). NAS-FM: Neural Architecture Search for Tunable and Interpretable Sound Synthesis based on Frequency Modulation*. arXiv preprint arXiv:2305.12868. https://doi.org/10.48550/arXiv.2305.12868
7. Cobos, M., Ahrens, J., Kowalczyk, K., Politis, A. et al. (2022). *An overview of machine learning and other data-based methods for spatial audio capture, processing, and reproduction.* EURASIP Journal on Audio, Speech, and Music Processing, 2022(1), 1-21. https://doi.org/10.1186/s13636-022-00242-x
8. YANG, Y. Y., HIRA, M., NI, Z., ASTAFUROV, A., CHEN, C., PUHRSCH, C., QUENNEVILLE-BÉLAIR, V. et al. (2022, May). *Torchaudio: Building blocks for audio and speech processing. In ICASSP 2022-2022 IEEE International Conference on Acoustics,* Speech and Signal Processing (ICASSP) (pp. 6982-6986). IEEE. doi: 10.1109/ICASSP43922.2022.9747236.
9. KELLERMANN, W., MARTIN, R., ONO, N., et al. (2023). *Signal processing and machine learning for speech and audio in acoustic sensor networks.* EURASIP Journal on Audio, Speech, and Music Processing, 2023(1), 54.
10. JENG, F. C., MATZDORF, K., HICKMAN, K. L., BAUER, S. W., CARRIERO, A. E., MCDONALD, K., WANG, C. Y., et al. (2023). *Advancing Auditory Processing by Detecting Frequency-Following Responses Through a Specialized Machine Learning Model.* Perceptual and Motor Skills, 00315125231225767. https://doi.org/10.1177/00315125231225767
11. ZEINALI, Y. and NIAKI, S. T. A. (2022). *Heart sound classification using signal processing and machine learning algorithms.* Machine Learning with Applications, 7, 100206. https://doi.org/10.1016/j.mlwa.2021.100206
12. HEMDAN, E. E. D., EL-SHAFAI, W., SAYED, A.et al. (2023). *CR19: A framework*

for preliminary detection of COVID-19 in cough audio signals using machine learning algorithms for automated medical diagnosis applications. Journal of Ambient Intelligence and Humanized Computing, 14(9), 11715-11727. https://doi.org/10.1007/s12652-022-03732-0

13. HEMDAN, E. E. D., EL-SHAFAI, W., SAYED, A., et al. (2023). *CR19: A framework for preliminary detection of COVID-19 in cough audio signals using machine learning algorithms for automated medical diagnosis applications.* Journal of Ambient Intelligence and Humanized Computing, 14(9), 11715-11727. https://doi.org/10.1007/s12652-022-03732-0
14. AMIRIPARIAN, S., HÜBNER, T., KARAS, V., GERCZUK, M., OTTL, S., SCHULLER, B. W., et al. (2022). *Deepspectrumlite: A power-efficient transfer learning framework for embedded speech and audio processing from decentralized data.* Frontiers in Artificial Intelligence, 5, 856232.
15. XIA, T., HAN, J., MASCOLO, C. et al. (2022). *Exploring machine learning for audio-based respiratory condition screening: A concise review of databases, methods, and open issues.* Experimental Biology and Medicine, 247(22), 2053-2061. https://doi.org/10.1177/15353702221115428
16. MATIKOLAIE, F. S., KHEDDACHE, Y., TADJ, C. (2022). *Automated newborn cry diagnostic system using machine learning approach.* Biomedical Signal Processing and Control, 73, 103434. https://doi.org/10.1016/j.bspc.2021.103434
17. PERUZZI, G., GALLI, A., POZZEBON, A., et al. (2022, July). *A novel methodology to remotely and early diagnose sleep bruxism by leveraging on audio signals and embedded machine learning*. In 2022 IEEE International Symposium on Measurements & Networking (M&N) (pp. 1-6). IEEE. doi: 10.1109/MN55117.2022.9887782.
18. KHAN, F., TARIMER, I., ALWAGEED, H. S., KARADAĞ, B. C., Fayaz, M., Abdusalomov, A. B., Cho, Y. I., et al. (2022). *Effect of feature selection on the accuracy of music popularity classification using machine learning algorithms.* Electronics, 11(21), 3518. https://doi.org/10.3390/electronics11213518
19. CHALAPATHI, M. M., KUMAR, M. R., SHARMA, N., SHITHARTH, S., et al. (2022). *Ensemble Learning by High-Dimensional Acoustic Features for Emotion Recognition from Speech Audio Signal.* Security and Communication Networks, 2022. https://doi.org/10.1155/2022/8777026
20. ABDUSALOMOV, A. B., SAFAROV, F., RAKHIMOV, M., TURAEV, B., & WHANGBO, T. K. et al. (2022). *Improved Feature Parameter Extraction from Speech Signals Using Machine Learning Algorithm.* Sensors, 22(21), 8122.
21. CHI, N. A., WASHINGTON, P., KLINE, A., HUSIC, A., HOU, C., HE, C., WALL, D. P., (2022). *Classifying autism from crowdsourced semistructured speech recordings: machine learning model comparison study.* JMIR Pediatrics and Parenting, 5(2), e35406. https://doi.org/10.2196%2F35406
22. GARG, A., CHATURVEDI, V., KAUR, A. B., VARSHNEY, V., PARASHAR, A., et al. (2022). *Machine learning model for mapping of music mood and human emotion based on physiological signals. Multimedia Tools and Applications*, 81(4), 5137-5177. https://doi.org/10.1007/s11042-021-11650-0