



ARTICLE

Ordered Clustering-Based Semantic Music Recommender System Using Deep Learning Selection

Weitao Ha¹, Sheng Gang², Yahya D. Navaei³, Abubakar S. Gezawa⁴ and Yaser A. Nanehkaran^{2,5,*}

¹School of Computer Science and Technology, Weinan Normal University, Weinan, 714099, China

²School of Information & Engineering, Yancheng Teachers University, Yancheng, 224002, China

³Department of Computer and Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, 34199-15195, Iran

⁴School of Information Engineering, Sanming University, Sanming, 365004, China

⁵Department of Management Information Systems, Faculty of Economics and Administrative Sciences, Cankaya University, Ankara, 06790, Türkiye

*Corresponding Author: Yaser A. Nanehkaran. Email: yaser@yctu.edu.cn

Received: 22 November 2024; Accepted: 04 March 2025; Published: 16 April 2025

ABSTRACT: Music recommendation systems are essential due to the vast amount of music available on streaming platforms, which can overwhelm users trying to find new tracks that match their preferences. These systems analyze users' emotional responses, listening habits, and personal preferences to provide personalized suggestions. A significant challenge they face is the "cold start" problem, where new users have no past interactions to guide recommendations. To improve user experience, these systems aim to effectively recommend music even to such users by considering their listening behavior and music popularity. This paper introduces a novel music recommendation system that combines order clustering and a convolutional neural network, utilizing user comments and rankings as input. Initially, the system organizes users into clusters based on semantic similarity, followed by the utilization of their rating similarities as input for the convolutional neural network. This network then predicts ratings for unreviewed music by users. Additionally, the system analyses user music listening behaviour and music popularity. Music popularity can help to address cold start users as well. Finally, the proposed method recommends unreviewed music based on predicted high rankings and popularity, taking into account each user's music listening habits. The proposed method combines predicted high rankings and popularity by first selecting popular unreviewed music that the model predicts to have the highest ratings for each user. Among these, the most popular tracks are prioritized, defined by metrics such as frequency of listening across users. The number of recommended tracks is aligned with each user's typical listening rate. The experimental findings demonstrate that the new method outperformed other classification techniques and prior recommendation systems, yielding a mean absolute error (MAE) rate and root mean square error (RMSE) rate of approximately 0.0017, a hit rate of 82.45%, an average normalized discounted cumulative gain (nDCG) of 82.3%, and a prediction accuracy of new ratings at 99.388%.

KEYWORDS: Music recommender system; order clustering; deep learning

1 Introduction

The rapid advancement of technology and the widespread use of online services have enabled users to access a wealth of information effortlessly. This has empowered users to contribute feedback and ratings across a wide spectrum of products and services. The abundance of data has made it increasingly difficult for users to pinpoint specific information, ultimately impacting the overall user experience. To tackle this issue,



new methods with reduced computational demands have emerged to help users efficiently find the content they seek. One such solution is the development of recommender systems, which have garnered substantial attention in recent years. These systems employ algorithms to analyze user preferences and behavior, offering personalized recommendations for content, products, and services [1]. Despite the enhanced accessibility of online information due to recent technological advancements, the current challenge lies in navigating vast amounts of data to find genuinely pertinent and useful content [2].

The concept of information overload in online platforms involves providing users with an extensive array of content and resources. In the context of music, this entails presenting a wide variety of songs, albums, and artists for users to explore. Effective guidance is crucial in assisting users in navigating this extensive content and discovering music that resonates with their preferences. This guidance can take the form of tailored recommendations, curated playlists, or search filters that facilitate easy access to desired content. Ultimately, the objective of information expansion in music and other online platforms is to enrich the user experience by granting access to a diverse and abundant range of content [3].

A music recommendation system is a technological tool designed to offer personalized music suggestions to users based on their unique preferences, listening patterns, and previous interactions. The proliferation of digital music streaming platforms and online music stores has led to a significant increase in the accessibility of music. This abundance of music choices has created challenges for users in discovering new songs that resonate with their individual tastes [3]. In response to this issue, music recommendation systems utilize algorithms to analyze a user's listening history, preferred genres, artists, and songs, along with other pertinent data such as user ratings and reviews. By leveraging this information, the system can generate customized music recommendations that are more likely to align with the user's preferences [4]. These recommendation systems have become integral features for music streaming platforms and online music stores, as they enhance the user experience by aiding individuals in discovering new music that is likely to be enjoyable for them. By providing tailored suggestions, these systems not only assist users in navigating the extensive music library but also contribute to heightened user engagement and satisfaction. In the digital age, music recommendation systems play a pivotal role in addressing the challenge of music discovery, offering users a more personalized and gratifying listening experience [5].

Scalability poses a significant challenge in the realm of music recommendations, particularly due to the vast quantity of music accessible on the Internet. It is imperative to employ robust algorithms capable of managing extensive music collections and delivering personalized recommendations to users. While clustering is a common approach to tackle scalability, existing algorithms often overlook the consideration of music popularity during the clustering process [6]. Furthermore, personalized music recommendation systems frequently struggle to effectively filter music based on a user's listening preferences. Early research in this field typically focuses solely on explicit user preferences, neglecting to integrate user profiles and comments into a comprehensive solution. Additionally, determining the optimal number of music recommendations based on a user's listening behavior presents another challenge that necessitates a novel solution. Addressing these issues is pivotal for the development of effective and personalized music recommendation systems capable of managing the vast volume of music available online [7].

The 'cold start' problem refers to a common challenge in recommendation systems where it is difficult to make accurate recommendations for new users or new items due to a lack of historical interaction data. For new users, the system has no prior information about their preferences, and for new items, there is no user feedback to guide recommendations. This issue is particularly challenging because it hampers the system's ability to provide relevant and personalized suggestions, potentially leading to user dissatisfaction or disengagement. In the proposed method, this problem is addressed by leveraging clustering techniques and convolutional neural networks. By grouping users with similar preferences and analyzing

semantic similarities in user comments, the system can infer preferences for new users based on their cluster membership. Similarly, for new music, the network predicts rankings by learning patterns from similar user clusters, ensuring meaningful recommendations even in the absence of direct interaction data [8,9].

The method to address the challenges by utilizing order clustering [10] and convolutional neural network for the recommender system. The proposed method involves calculating the similarity of users based on the semantic analysis of their comments in the user profile. Providing more details about the “semantic analysis of user comments” can significantly enhance clarity, particularly for readers unfamiliar with the underlying techniques. Semantic analysis typically involves natural language processing (NLP) methods such as sentiment analysis, topic modeling, and word embeddings. Techniques like Word2Vec for capturing contextual relationships between words, and sentiment analysis models to detect users’ opinions and emotions are commonly employed. Including such specifics would offer a better understanding of how user comments are interpreted and leveraged in recommendation systems.

Additionally, Order Clustering (OC) is employed to group the most similar users in terms of meaning into separate clusters, with the first cluster comprising the most similar users, and subsequent clusters containing users with decreasing levels of similarity. The Order Clustering algorithm is an innovative clustering approach that does not fit within the categories of exclusive or fuzzy clustering. This algorithm is designed based on users’ music consumption behavior and the inherent characteristics of music tracks, allowing users to belong to multiple clusters simultaneously. Consequently, a single music track can be recommended to several users, and a user may belong to multiple music categories based on their listening preferences. The number of clusters is determined dynamically during the execution of the algorithm, with clusters defined as subsets of the dataset. The clustering process begins by selecting pairs of users with the highest similarity in the similarity matrix and grouping them into the same cluster. Users are then ordered in descending similarity within each cluster. This method does not assign weights or membership values, making it a non-exclusive clustering approach. Its non-restrictive nature enhances flexibility in categorizing users, accommodating diverse musical tastes and listening behaviors.

Within these clusters, the favorite music of each user is evaluated to identify shared preferences in terms of music ranking. User ratings’ similarity is then computed based on the common favorite music among users. Users exhibiting high semantic and ranking similarity are identified as common users. Consequently, users’ opinions and music ratings are used as training data for the convolutional neural network (CNN) to predict ratings for other music. Machine learning, and especially deep learning, is used in a variety of fields [8–10]. Convolutional Neural Networks are powerful deep learning models designed to capture spatial and temporal patterns within data. In music recommendation systems, CNNs excel at learning complex audio features such as rhythms, melodies, and harmonies from spectrogram representations of music tracks. When combined with user feedback data, including opinions and music ratings, CNNs become even more effective. Users’ ratings act as labeled training data, allowing the network to learn relationships between specific music features and user preferences. Through backpropagation and iterative updates to the model weights, CNNs adjust their filters to better predict unseen ratings. For example, if a user consistently rates songs with strong beats and upbeat melodies higher, the network learns to associate those features with positive ratings. Once trained, the CNN can predict ratings for other music by evaluating how closely the features of a new track match patterns learned from the training data, thereby enabling more personalized and accurate music recommendations.

Furthermore, the authors demonstrate that it is possible to deduce users’ music listening behavior based on their rating behavior across the entire dataset. Additionally, the popularity of each music can be determined based on ratings provided by all users. In addition, in the proposed method, to deal with cold start users, music with a high popularity rate is used. Since these music are very popular, they are likely to be

noticed by the cold start users. This recommendation can lead to assigning a rating or comment to the desired music by cold start users, which will facilitate finding similar users. Ultimately, music recommendations are generated for users based on estimated rankings derived from convolutional neural networks, users' listening patterns, and music popularity.

The main contribution of the article is summarized as follows:

- **Semantic Similarity-Based Clustering:** Developed a semantic similarity matrix for user profiles using text comments related to music, leveraging cosine similarity, followed by sequential clustering to identify semantically similar user groups.
- **User Similarity Ranking:** Introduced a novel similarity ranking method based on the number of common music items among users within the sequential clusters.
- **Enhanced Rating Estimation:** Proposed a convolutional neural network (CNN) model to estimate ratings for new music by combining semantic analysis and existing user ratings as training data.
- **Personalized Music Recommendation:** Designed a recommendation framework integrating estimated ratings, user listening rates, and music popularity scores to deliver personalized music suggestions.

The proposed hybrid recommender system integrates the strengths of both content-based and collaborative filtering approaches to enhance recommendation accuracy and user satisfaction. Content-based methods focus on analyzing user preferences or opinions to recommend music that shares similar attributes with previously liked tracks. Conversely, collaborative filtering methods predict user ratings by analyzing the interactions of multiple users, identifying patterns, and leveraging peer preferences. In this method, user similarity is extracted from comments to capture implicit preferences and opinions beyond mere ratings. Additionally, an order clustering method is employed to group users based on their interactions with specific items, allowing for more targeted and context-aware recommendations. During the recommendation phase, the system also incorporates music popularity and users' listening habits to create a balanced and personalized music experience. By considering these diverse factors, the proposed approach not only improves recommendation precision but also adapts to the dynamic preferences of users.

The rest of paper organized as follow. In [Section 2](#), related works are reviewed. In [Section 3](#), the research methodology will be presented. In [Section 4](#), the implementation and results of the proposed method will be detailed. In [Section 6](#), the conclusion of the paper will be stated.

2 Background and Related Works

Recommender systems are crucial in assisting users in discovering content, products, or services that are relevant to their interests and preferences. These systems are generally classified into three main types: content-based recommender systems, collaborative recommender systems, and hybrid recommender systems. The following subsections will review each type of recommender system.

2.1 Content-Based Recommender Systems

These systems analyze the attributes and characteristics of the items being recommended. They focus on understanding the features of the items and matching them to the user's preferences. Content-based recommenders suggest items that are similar to those that a user has liked or interacted with in the past, based on shared attributes or characteristics.

Content-based recommender systems prioritize the characteristics or features of the items being recommended. This method involves creating item profiles based on their descriptions or features. When a user shows interest in a particular item, the system suggests similar items based on their shared characteristics. For instance, if a user enjoys a specific music genre, a content-based recommender system might propose

other songs or artists with similar musical attributes [11]. In a different approach [12], a novel music recommendation system named Moodify, based on reinforcement learning (RL), has been proposed. This system is capable of inducing emotions in the user to support the interaction process in various usage scenarios (e.g., games, movies, smart spaces). The RL method is designed to learn how to select a list of music pieces that best match the target emotional state, given a desired emotional state and the assumption that an emotional state is entirely determined by a sequence of recently played music tracks. Furthermore, a study [13] introduced the Deep Neural Based Music Recommendation (DNBMR) method, which considers both music attributes and the customer's temporal favourites. Additionally, the role of negative preference in users' music tastes was analysed in a study [14] by comparing music recommendation models with contrastive learning exploiting preference (CLEP) using three different training strategies: exploiting preferences of both positive and negative (CLEP-PN), positive only (CLEP-P), and negative only (CLEP-N). Finally, a music recommendation system framework based on sentiments, named SentiSpotMusic, was proposed in another study [15], utilizing Tableau dashboard and Spotify dataset.

The reviewed methods showcase a range of innovative approaches in music recommendation systems, each addressing specific aspects of user preferences and emotional needs. Traditional content-based systems focus on item features and user preferences, but they often fail to capture contextual and emotional factors, leading to limited diversity in recommendations. Advances like the LSTM-based framework for emotion-based recommendations significantly enhance personalization by aligning music choices with users' moods and emotional stimuli, addressing mental well-being. Similarly, Moodify leverages reinforcement learning to adapt recommendations dynamically based on the user's desired emotional state, demonstrating an innovative integration of psychology and machine learning. The exploration of personal characteristics such as cognition and musical sophistication adds depth to user profiling, though it might face challenges in generalizing across diverse user bases. Deep neural methods like DNBMR advance the field by incorporating temporal preferences and music attributes, but their reliance on extensive data and computational resources could limit practical deployment. Models using contrastive learning, like CLEP, bring a novel perspective by analyzing the role of negative preferences, which is often overlooked in traditional systems. While sentiment-driven frameworks like SentiSpotMusic offer intuitive interfaces, their dependency on external tools like Tableau and specific datasets may constrain their scalability and adaptability. Collectively, these methods reflect the progression toward more nuanced and user-centered music recommendation systems, yet challenges remain in balancing personalization, scalability, and interpretability.

2.2 Collaborative Recommender Systems

Collaborative filtering is a common technique used in these systems, where recommendations are based on users' interactions and preferences. These systems identify patterns and similarities in user behaviors and preferences to recommend items that users with similar tastes have liked or interacted with.

Collaborative recommender systems focus on understanding user behavior and preferences. These systems group users with similar tastes and preferences, referred to as "neighborhoods," and suggest items based on the preferences of users within these groups. For example, if user A shares preferences with a group of users X, the system might recommend popular items among users in group X to user A. The effectiveness of a collaborative algorithm relies on its capacity to accurately identify the user's neighborhood and provide recommendations based on the collective preferences of that group [16,17]. In a related work [18], the common aspects and objectives of explainable recommenders, as well as explainable Artificial Intelligence (XAI), are explored. This study delves into how these concepts can be applied or modified to suit the specific characteristics of music consumption and recommendation. Another study [19] investigates a potential issue

in collaborative-filtering based multimedia recommender systems, particularly the phenomenon of popularity bias leading to the underrepresentation of unpopular items in recommendation lists. In [20], a new method for measuring similarity is introduced, which involves translating intuitive and qualitative conditions into mathematical equations and solving them to achieve the kernel function. In [21], a model called Dynamic evolution of Multi-Graph Collaborative Filtering (DMGCF) is proposed, aiming to extract and reuse side information. In [22], a deep learning method for collaborative recommender systems (DLCRS) is presented, suggesting the use of advanced neural network techniques for recommendation tasks. In [23], an approach to enhancing music recommendation systems is discussed, with potential applications across various platforms and domains. In [24], a fusion method is proposed, which simultaneously takes into account multi-factor similarity and global rating information within a probability matrix factorization framework.

Traditional collaborative filtering (CF) systems excel in leveraging user behavior and preferences to group similar users and recommend items based on neighborhood preferences. However, their reliance on historical data can lead to challenges like popularity bias, as noted in one study, where less popular items are underrepresented. To counteract such biases, the integration of explainable AI (XAI) in recommendation systems has been proposed, enhancing transparency and user trust, particularly in domains like music consumption. Techniques such as kernel-based similarity measurement and multi-graph collaborative filtering (DMGCF) showcase efforts to improve accuracy by incorporating side information and mathematically rigorous similarity metrics. Deep learning-based methods like DLCRS further advance the field by leveraging neural networks to capture complex, non-linear user-item interactions, but their computational intensity may limit scalability. Additionally, multi-factor similarity fusion within probability matrix factorization frameworks offers a promising avenue for balancing localized preferences with global patterns, though ensuring the interpretability of such models remains a challenge. Collectively, these approaches demonstrate progress in addressing CF limitations, such as popularity bias and sparse data, while enhancing personalization and explainability. However, practical deployment requires balancing accuracy, computational efficiency, and user satisfaction.

2.3 Hybrid Recommender Systems

Hybrid systems combine the approaches of content-based and collaborative filtering methods to provide more accurate and diverse recommendations. By leveraging the strengths of both approaches, hybrid recommender systems aim to overcome the limitations of individual methods and offer more personalized and effective recommendations to users.

These different types of recommender systems cater to various user preferences and behaviours, offering a range of approaches to meet the diverse needs of users in discovering relevant and engaging content [2,25].

Hybrid recommender systems blend elements of content-based and collaborative approaches to deliver more precise and diverse recommendations. By harnessing the strengths of both methods, hybrid systems seek to address the limitations of individual recommendation techniques and provide personalized and effective suggestions to users [17,26]. In [27], an algorithm is described that merges collaborative filtering and music genes to develop a personalized music recommendation system. It describes the process of computing recommendation results using collaborative filtering, assigning weights to music based on user preferences, and combining and filtering the results to generate recommendations. In [28], association rules and music genes are integrated into a music collaborative filtering personalized recommendation system to establish a hybrid recommendation model. In [29], a personalized hybrid recommendation algorithm for music based on reinforcement learning (PHRR) is proposed to recommend song sequences that better match listeners' preferences. In [30], a recommendation system is presented based on the similarity of the selected artist, top artists in a genre, using a hybrid model, and artists listened to by users' friends.

The reviewed hybrid recommendation methods showcase the versatility and adaptability of combining content-based and collaborative filtering techniques to enhance recommendation precision and diversity. By addressing the limitations of individual methods, hybrid systems can mitigate issues like data sparsity and cold-start problems while incorporating richer contextual or preference-based insights. For instance, algorithms merging collaborative filtering with music genes demonstrate the potential of weighting music preferences to create refined and accurate recommendations, though their dependence on well-curated meta-data might pose scalability challenges. Approaches integrating association rules with music genes further enrich the recommendation process by uncovering hidden patterns in user behavior, yet the complexity of such integrations may affect real-time applicability. Reinforcement learning-based hybrid systems, such as PHRR, offer a dynamic and adaptive framework for generating personalized music sequences, but they rely on extensive interaction data, which could limit effectiveness for new users or systems with sparse data. Systems that incorporate artist similarity, genre popularity, and social connections provide a comprehensive perspective by blending individual and social preferences, though they risk oversimplifying user tastes or reinforcing echo chambers. Overall, these hybrid approaches effectively enhance recommendation quality, but their success hinges on balancing computational efficiency, user-centric design, and scalability across diverse datasets and user bases.

According to the reviewed articles in this section, it can be seen that content-based methods need to receive user opinions or preferences, while collaborative filter-based methods try to estimate users' ratings. The proposed method tries to exploit the advantages of both cases and hence it is considered a hybrid recommender system.

3 Methodology

As mentioned, in this paper, a music recommendation approach based on ratings using a combination of order clustering and convolutional neural networks is presented. The architecture of the proposed method is shown in Fig. 1.

The proposed method commences by utilizing a dataset containing user ratings for music, encompassing user comments and the rankings they assign to the music. Initially, the method extracts the similarity matrix of users based on their comments, employing the cosine similarity criterion. Cosine similarity is a widely recognized measure for assessing text similarity, with a higher similarity score between user comments denoting greater semantic similarity between users. This similarity measure enables the method to gauge the likeness of user comments, facilitating the identification of users with similar preferences and aiding in the generation of more tailored and accurate music recommendations. Mathematically, the cosine similarity criterion is expressed in Eq. (1).

$$Sim(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}| |\vec{d}|} = \frac{\sum_{k=1}^t w_{qk} \times w_{dk}}{\sqrt{\sum_{k=1}^t (w_{qk})^2} \cdot \sqrt{\sum_{k=1}^t (w_{dk})^2}} \quad (1)$$

where q and d refer to text, and t is the number of words in the text and w is the weight of each word. The similarity matrix serves as the input for the order clustering method. Within the order clustering process, the matching values are initially sorted in descending order. Subsequently, the first value, representing the highest level of similarity between users, is selected as the initial cluster, with users matching this value being placed within the first cluster. The subsequent values in the ordered similarity matrix are then considered as additional clusters. For each similar value, the neighborhood matrix of a cluster is taken into account, and users are ranked based on the proposed similarity function within the clusters. This approach enables the method to organize users into distinct clusters based on their levels of similarity, facilitating the identification

of user groups with comparable preferences and aiding in the generation of more refined and targeted music recommendations.

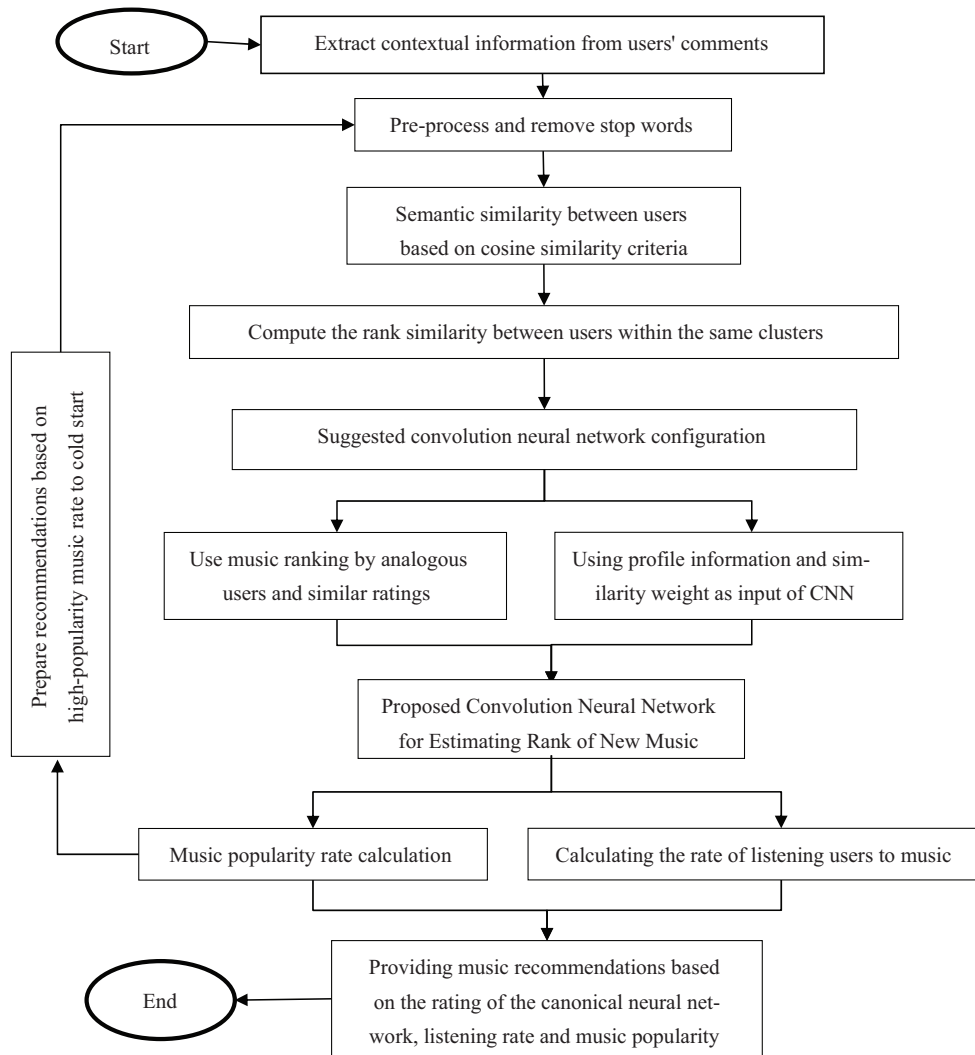


Figure 1: Flowchart of the proposed method

The proposed similarity function is calculated by Eq. (2).

$$Sim_R(i, j) = \frac{\sum_k^K \sum_{i=1}^n \sum_{j=1}^n SameRate_{i,j}^k}{\sum_{i=1}^n Rate_i + \sum_{j=1}^n Rate_j} \quad (2)$$

where n is the number of users in the cluster, K the number of clusters, $SameRate_{i,j}^k$ as the same order of users in each cluster, $Rate_i$ is the actual value of the rank granted by each user. As per the text, the user rating can be likened to the count of music for which two users in a cluster have assigned the same order. Once the semantic similarity and user ranking are established, these values, in conjunction with users' comments, are utilized as input for the proposed convolutional neural networks. Subsequently, the architecture of the deep canola neural networks will be examined.

Clustering Method for Music Recommendation

The proposed music recommendation system employs Order Clustering (OC) to group users with similar preferences. This clustering method begins by constructing a similarity matrix based on user comments and rankings, leveraging semantic similarity measures and user behaviors. The clustering process enhances the system's ability to tailor recommendations by organizing users into meaningful groups that reflect shared interests and tastes.

- Cosine Similarity for Semantic Analysis

To quantify the similarity between user comments, the system uses cosine similarity, a robust metric for comparing text data. Cosine similarity computes the angular difference between vector representations of two textual inputs, effectively capturing their semantic alignment. A higher cosine similarity score indicates a greater resemblance in the content and sentiment of user comments. The similarity metric is expressed in [Eq. \(1\)](#). By applying cosine similarity, the system establishes a numerical measure of semantic similarity between users, serving as the foundation for clustering.

- Order Clustering Process

Once the similarity matrix is computed, the Order Clustering (OC) method organizes users into clusters. The process involves several structured steps:

Sorting: The similarity matrix is ordered in descending order, starting with the highest similarity values.

Initial Cluster Formation: The first value in the sorted matrix is selected to form the initial cluster. Users corresponding to this similarity value are grouped into the first cluster.

Subsequent Clusters: For each subsequent value, the neighborhood matrix of the existing clusters is considered, and additional clusters are formed based on similarity.

Ranking Within Clusters: Users in each cluster are ranked using a custom similarity function, which takes into account both semantic similarity and user rankings.

The similarity function for ranking within clusters is defined in [Eq. \(2\)](#). This method ensures that clusters reflect shared preferences not only in terms of semantic similarity but also in user behavior, resulting in well-defined user groups.

- Integration with Convolutional Neural Networks

After clustering, the semantic similarity values, user rankings, and user comments serve as inputs to the Convolutional Neural Network (CNN). These inputs guide the CNN in learning patterns and predicting rankings for unreviewed music. The integration of clustering with CNN ensures that the recommendation system leverages both user-based and content-based insights for personalized recommendations.

By combining semantic analysis and clustering techniques, the system achieves a robust mechanism to identify user preferences, laying a solid foundation for accurate and efficient music recommendations.

3.1 Convolution Neural Network

In the proposed method, convolutional neural networks are employed for user learning within the music recommender system. Specifically, the music reviewed by similar users in ordered clusters are utilized to train the convolutional neural networks, which operate across multiple layers to facilitate deep learning. A portion of these music is allocated for training data, while the remainder is assigned to validation data. Within the convolutional neural networks, the proposed patterns are derived from user comments, taking into account semantic and user ranking similarity for the music reviewed by users and utilized as training data. Subsequently, these patterns are assessed using the music identified as validation data. Ultimately,

the deep neural networks predict a new ranking as the output for music that have not been reviewed by users and are designated as test data. The system employs the Musical Instruments Reviews, and Music Album Reviews [31,32] datasets, dividing the available music into training and testing data for each user. By classifying individual music reviews and analyzing user sentiments, these datasets can generate an overall rating for specific tracks or albums. This insight allows music platforms to better understand user preferences and make informed decisions on music recommendations or content adjustments. The training set is further partitioned into training and validation data, while the test music is treated as unreviewed music within the model.

The input layer of the neural network feeds similar user rankings as the target and user reviews as input features. Subsequently, the pooling layer aggregates the input features of the training data and conveys them as training parameters to the middle layer.

Within the multi-layer fully connected layer, each input feature is transformed into a parameter and assigned a label class and weight based on the data. This process determines the weight of each feature for each music, forming the basis of the recommendation model. The multiple layers are fully connected to ensure comprehensive error detection and rectification within the middle layer.

The classification layer allows the neural network model to learn necessary parameters from the training data and categorize each example based on user reviews. The middle layer also ranks music based on the weight of user reviews, with the weight of the features determined in the convolutional layers, shaping patterns for ranking the music.

Finally, the output layer utilizes the output of the previous layer and compares the validation methods with the models related to music rating, expressing the matching or non-matching of the validation data. This layer compares the matching rate of the influence factor of the sample characteristics validation methods with the coefficient of influence of the characteristics of the training samples, specifying the error of that stage in the proposal model based on the matching of each validation sample with the training sample.

In the proposed music recommendation system, convolutional neural networks (CNN) play a crucial role in modeling user preferences and predicting rankings for unreviewed music. The CNN architecture leverages ordered clusters of music reviewed by users with similar preferences, enabling it to capture semantic patterns and ranking similarities. The dataset is divided into training, validation, and test sets. The training data consists of user reviews and ratings for music that have been previously assessed, while the test data represents unreviewed music for which predictions are generated. Validation data is used during the training process to refine the network parameters and assess intermediate performance.

3.1.1 CNN Architecture and Input Features

The CNN architecture begins with an input layer that receives user reviews and rankings. These inputs are processed through convolutional layers designed to extract localized features, such as semantic themes in reviews or rating trends. The convolutional filters scan these inputs, generating feature maps that emphasize significant patterns relevant to user preferences. The pooling layer follows, aggregating these feature maps to reduce dimensionality and retain essential characteristics, ensuring efficient processing without losing critical information. This hierarchical feature extraction process is integral to identifying patterns that influence music rankings.

3.1.2 Fully Connected Layers and Feature Weighting

After feature extraction, the fully connected layers transform the extracted features into parameterized representations. Each feature is assigned a weight and label class, reflecting its importance in predicting

music preferences. This stage involves complex computations to optimize feature weights, ensuring that the model accurately captures user behavior. By fully connecting all neurons in this layer, the model achieves comprehensive error detection and correction, enhancing its ability to generalize across diverse user inputs. These fully connected layers also facilitate ranking music by assigning weighted scores based on the extracted features.

3.1.3 Classification and Output Layers

The classification layer categorizes the music based on the learned patterns from the training data. It determines how closely a piece of music aligns with user preferences by analyzing the aggregated weights from previous layers. Finally, the output layer evaluates the model's predictions against the validation data, identifying discrepancies and refining parameters accordingly. This iterative validation process ensures that the model's predictions for unreviewed music are both accurate and consistent with user preferences. By integrating CNNs into the recommendation system, the proposed method achieves a sophisticated analysis of user behavior, enabling personalized music suggestions tailored to individual tastes.

3.2 Recommendation Module

The recommendation module in the provided method represents the final stage where the actual recommendations are generated. The output of the convolutional neural network includes a rating for test music, which is essentially unexplored by users. When providing music advice to users, two key factors are taken into consideration: the music audience rate and the music rate. Users who have only listened to a few music groups may have a large number of unreviewed music in the dataset. If all unreviewed music is presented to such users, it is likely that a significant portion of the recommendations will be disregarded, potentially diminishing the effectiveness of the recommender system. Therefore, tailoring music recommendations according to each user's listening rate can mitigate this challenge and ensure the delivery of a number of effective recommendations for each user. The rate of user listening to music is calculated by the Eq. (3).

$$Listen_{Rate_i} = \frac{\sum_{i=1}^N \sum_{m=1}^M Num_{Rate_i}}{M} (i, j) = \frac{\sum_k^K \sum_{i=1}^n \sum_{j=1}^n SameRate_{i,j}^k}{\sum_{i=1}^n Rate_i + \sum_{j=1}^n Rate_j} \quad (3)$$

where M the total number of music in the data set, Num_{Rate_i} is the number of ratings that one user has granted to the music. In the proposed method, the approach to providing recommendations takes into account the musical listening habits of users, allowing for the calculation of a user's preferences. By understanding a user's listening habits, the system can offer tailored and pertinent recommendations that align with the user's music preferences.

Moreover, the method also incorporates the consideration of the popularity rate of music. This popularity rate is determined based on the number of maximum ratings given to a particular piece of music by users. By factoring in the popularity rate, the system can ensure that the recommendations provided are not only aligned with the user's preferences but also take into account the general appeal and popularity of the music, enhancing the overall quality of the recommendations. This comprehensive approach enables the system to offer a more personalized and effective recommendation experience for each user. The popularity rate of music can be calculated based on Eq. (4).

$$Popularity_{Rate_m} = \frac{\sum_{m=1}^M \max(Rate_m)}{\sum_{m=1}^M Rate_m} \quad (4)$$

where $\sum_{m=1}^M \max(Rate_m)$ is the highest amount of rating given by users to a music.

As previously mentioned, cold-start users are new users who have no prior experience with music. Consequently, no information is available regarding their opinions or ratings of Yindi's music. To gather insights from these users, it is necessary to provide them with music recommendations and collect their feedback.

In the proposed method, highly popular music—frequently listened to by many users and highly rated—is recommended to cold-start users. This approach allows for the collection of their opinions, which can then be utilized to identify similar users through structured clustering. Additionally, these users' feedback can be leveraged to predict the preferences of other new users using convolutional neural networks.

To generate recommendations, the proposed method selects music from the unreviewed tracks in the test dataset that meet specific conditions.

The model has predicted the highest rating for them.

The most popular music.

The number of music should be equal to the music listening rate for each user.

If the number of music with the highest rating and the most popularity is less than the number of music according to the listening rate of the user's music, we can use music with a lower rating and high popularity.

4 Implementation Results

The method presented for implementing the music recommender system is based on the analysis of user comments and utilizes the large-scale dataset “Musical instruments reviews”, and “Music Album Reviews”. This datasets contains valuable information extracted from users' profiles about music in text form, as well as the ratings provided by users for available music. Each user is uniquely identified within the dataset. [Table 1](#) presents the properties of the dataset used, providing insights into the specific attributes and characteristics of the data that are leveraged for the music recommender system.

Table 1: Specification related to the dataset used

Dataset	Music count	Number of users	Ranking values set
Musical instruments reviews	900	1429	{1, 2, 3, 4, 5}
Music album reviews	73	1027	{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5}

These datasets have 2 fields in which “Review” could refer to user-generated free text comments where listeners share their opinions, emotions, or critiques about specific songs, albums, or playlists. On the other hand, “Rating” might represent a numeric value (described in “Ranking Values Set” column) that reflects the user's overall assessment of the music. This distinction is essential for effective data processing and meaningful feedback analysis in music recommendation systems.

The method described involves a data pre-processing step, followed by the application of cosine similarity to analyse user comments in a music recommendation system. Firstly, the dataset is partitioned into training and test sets, with 70% and 30% of the data, respectively. The training set includes user ratings for music and their corresponding textual comments, which convey the users' emotional responses and sentiments towards the music. These comments, when coupled with the user ratings, can be utilized for sentiment analysis, providing insights into user behavior.

The text preprocessing stage involves identifying words, phrases, sentences, removing stop-words, stemming, and more to reduce text size. Techniques such as indexing, keyword removal, stemming, phrase

extraction, word sense disambiguation, query refinement, and using knowledge bases enhance information retrieval system performance. Removing stop words, which are common but insignificant in natural language processing tasks, improves system efficiency. Providing detailed explanations about the specific stop words removed and the techniques used in text preprocessing would significantly clarify the methodology. For instance, specifying whether common stop words like “the” “is”, and “but” are excluded can help readers understand the data cleaning process. Additionally, indicating whether text processing was implemented using popular libraries such as NLTK or spaCy would add transparency and reproducibility to the research. Such libraries offer predefined lists of stop words and advanced text processing functions, making them valuable tools for efficiently preparing data for analysis.

In the proposed method, user profile data is preprocessed into meaningful words to measure user similarities using the cosine similarity metric. This widely-used technique in information retrieval models textual documents as term vectors, calculating the cosine of the angle between vectors to determine similarity. In recommender systems, cosine similarity generates a symmetric $n \times n$ matrix, where n represents the number of users, and values range from 0 to 1. Each element indicates the similarity between two users, with a higher value implying greater similarity. The diagonal elements always equal 1, as a user is most similar to themselves. Fig. 2 shows the cosine similarity matrix for 10 users from the training dataset as an example.

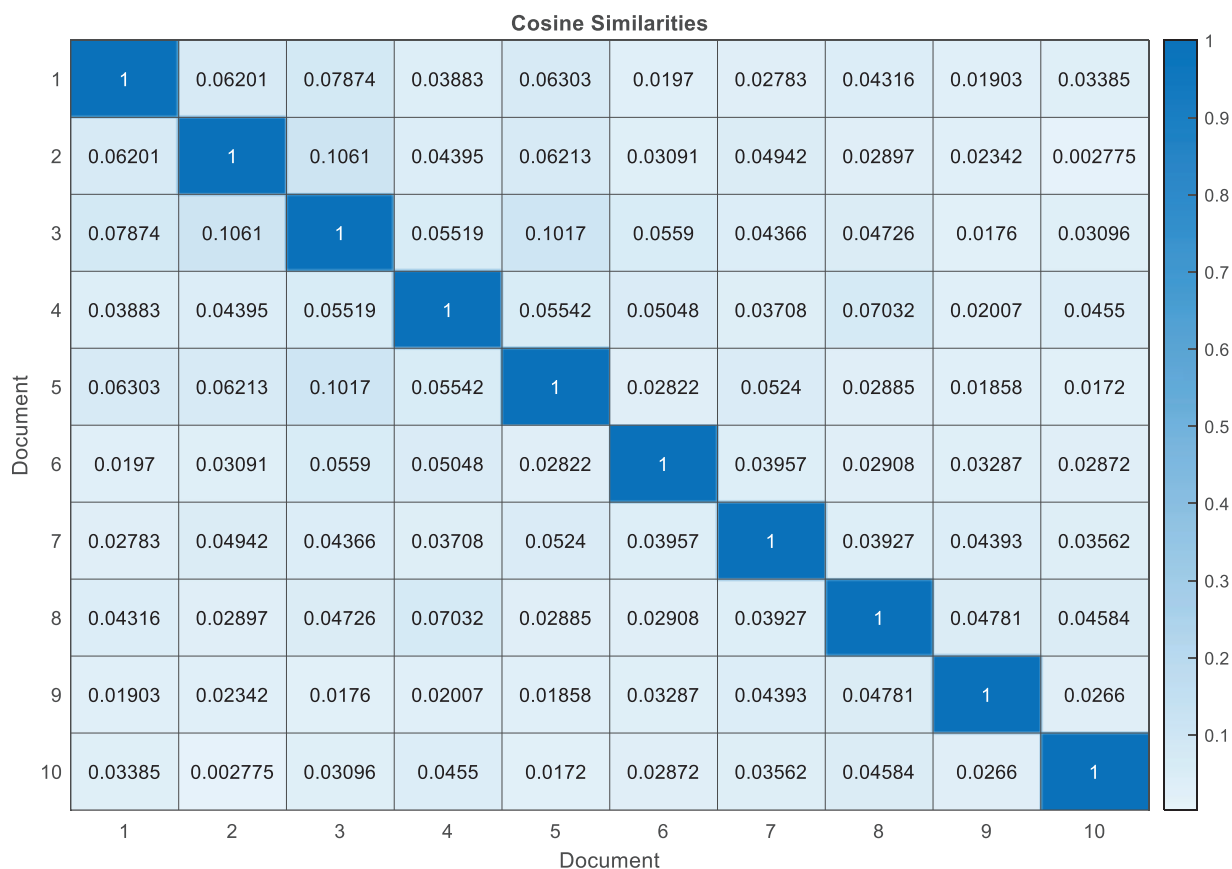


Figure 2: Similarity matrix between users based on the cosine similarity criterion

Given that the users' comments are in the form of text strings, it is crucial to preprocess the data by removing stop words to enhance its analytical potential. Following the pre-processing step, the similarity

between users' comments is computed using the cosine similarity measure by creating similarity matrix. This measure allows for the quantification of the similarity between textual expressions, enabling the identification of commonalities in users' emotional responses and sentiments towards the music. Each entry in the similarity matrix reflects the degree of similarity between a pair of users, with values ranging between 0 and 1. This numerical value quantifies the similarity between two users in terms of their comments, where a higher value indicates a greater similarity in the users' expressed sentiments and emotions.

In the presented method, the process of identifying similar users involves the establishment of a threshold. The threshold serves as a criterion for determining the level of similarity required for users to be considered as "similar" to each other based on their comments. Users whose similarity values exceed this threshold are deemed sufficiently similar to be grouped together for personalized recommendation purposes. Therefore, the proposed threshold is defined in the form of the Eq. (5).

$$Popularity_{Rate_m} = \frac{\sum_{m=1}^M \max(Rate_m)}{\sum_{m=1}^M Rate_m} \quad (5)$$

In Eq. (5), i and j are the index of users and n is the total number of users in the training dataset. After calculating the similarity between users' comments, in the presented method, order clustering is performed in which users are assigned to each cluster according to the similarity value. Table 2 shows an example of order clustering of users based on the similarity of their opinions.

Table 2: Order clustering of users based on the similarity of opinions

Cluster number	Similarity value		Similar users		Predictive rank	
	Music instruments	Music album	Music instruments	Music album	Music instruments	Music album
1	0.1960	0.5164	[869; 918]	[339; 547]	5	4.5
2	0.1910	0.4048	[515; 1195]	[81; 488]	2	4.5
3	0.1860	0.4019	[849; 1019]	[214; 885]	4	5
4	0.1840	0.3939	[46; 393]	[249; 707]	5	3.5
5	0.1580	0.3921	[245; 377; 576]	[110; 323]	5	4

In the presented method, Table 2 illustrates the assignment of users to clusters based on the similarity of their comments. In this clustering approach, users are grouped based on their preferences or opinions about specific items. Users who express similar views or exhibit similar interaction patterns with the same items are categorized into the same cluster. This method essentially aligns user clusters with individual items, implying that the number of clusters is directly determined by the number of distinct items in the dataset. As a result, each item forms the basis for a unique cluster, ensuring that users with comparable preferences are grouped together, thereby facilitating more targeted analysis, personalization, or recommendation strategies in applications such as collaborative filtering. Each user is uniquely assigned to a specific cluster, with the clustering process accommodating scenarios where a cluster may consist of only one user. This clustering approach is founded on the premise that users allocated to the same cluster exhibit a high probability of sharing similar behaviors and preferences, specifically in terms of their expressed sentiments and emotions toward music.

Furthermore, the method utilizes the identified user clusters to calculate the similarity of users based on the ratings they have assigned to music. Users who have both semantic similarity in their comments and demonstrate greater similarity in their music ratings are considered to be prime candidates for input into convolutional neural networks (CNNs). The rationale behind this approach is rooted in the notion that users exhibiting both semantic similarity in their comments and alignment in their rating behaviors are likely to possess comparable preferences and emotional responses toward music.

4.1 Evaluation

In the context of evaluating the recommendation system, the study employs two distinct criteria to assess the accuracy and effectiveness of the proposed method. The first criterion pertains to the statistical accuracy measure, which serves as a means to evaluate the precision of the predicted ranking for new music and the accuracy of music recommendations provided to users. The second criterion focuses on evaluating the performance of convolutional neural networks (CNNs) in predicting the ranking of new music based on user feedback, and subsequently comparing this performance with that of other classification methods. Furthermore, the study adopts a systematic approach, first emphasizing the evaluation of the proposed method based on statistical accuracy. Subsequently, the accuracy of CNN with various hyper-parameters such as learning rate, batch size, and number of layers will be evaluated. Finally, the study delves into a comparative analysis, wherein the accuracy of classification methods, including CNNs, is contrasted.

4.1.1 Statistical Accuracy Evaluation

For evaluating the accuracy of the recommendation system, statistical accuracy measures serve as fundamental tools for quantitatively assessing the alignment between predicted music rankings and users' actual ratings. The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are prominent statistical accuracy measures employed to gauge the precision and reliability of the method's predictions. The Mean Absolute Error (MAE) is widely recognized as a key statistical accuracy measure due to its simplicity and interpretability. It quantifies the average magnitude of errors between predicted rankings and actual ratings, providing a clear indication of the method's accuracy in capturing users' preferences. A lower MAE value signifies greater accuracy, as it indicates that the predicted rankings closely correspond to users' actual ratings. Essentially, a lower MAE value denotes a higher level of precision in the method's ability to forecast music rankings, thereby enhancing the overall accuracy of the recommendation system. Similarly, the Root Mean Square Error (RMSE) is another essential statistical accuracy measure utilized to assess the predictive performance of the method. RMSE quantifies the average magnitude of the discrepancies between predicted rankings and actual ratings, while also penalizing larger errors to a greater extent compared to MAE. A lower RMSE value signifies enhanced accuracy, indicating that the method's predictions closely align with users' actual preferences and ratings. These measures are calculated as Eq. (6).

$$MAE = \frac{1}{N} \sum_{u,i} |p_{u,i} - r_{u,i}| \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{u,i} (p_{u,i} - r_{u,i})^2}{N}} \quad (7)$$

where $p_{u,i}$ is defined as the predicted rating for user u on product i and $r_{u,i}$ is the actual rating of user u on product i and N is the total number of ratings in the data set. Fig. 2 shows the mean absolute error value and the graph of the cumulative mean absolute error value per test samples in the data set. Also, Fig. 3 shows the

root mean square error value and cumulative graph of the root mean square error per test samples in the data set.

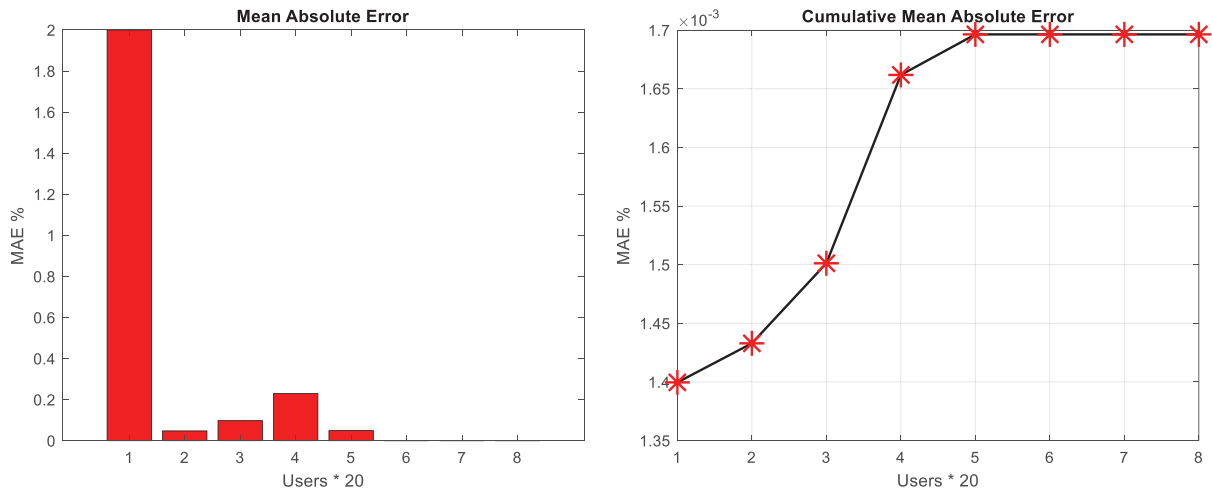


Figure 3: Mean absolute error value for test users in music instruments dataset

The findings presented in Figs. 3–6 illustrate the cumulative graph of MAE and RMSE values for test samples in both datasets, providing insights into the predictive accuracy of the proposed CNNs in forecasting ratings for new music. The gradual increase in the cumulative error values, as depicted by the gentle slope in the graphs, signifies the high accuracy of the CNN-based method in predicting ratings for new music. The observation that the cumulative MAE and RMSE values exhibit a gentle slope suggests that the CNNs consistently yield accurate predictions for a wide range of test samples, thereby demonstrating the method's robustness and reliability in capturing user preferences. This trend underscores the method's efficacy in consistently delivering accurate and precise predictions, as evidenced by the incremental but controlled increase in cumulative error values across the test samples. Furthermore, the culmination of the cumulative MAE and RMSE values at the maximum value of 0.0017 is indicative of the method's exceptional predictive performance. This minimal error value signifies the near-optimal accuracy of the CNN-based approach in forecasting ratings for new music, underscoring its capability to closely align predicted rankings with users' actual preferences. The attainment of such a small error value underscores the method's high precision and reliability, indicating its ability to achieve near-perfect predictions for new music ratings.

4.1.2 Recommendations Accuracy Evaluation

The accuracy of the recommendations is evaluated by identifying useful recommendations from the dataset, which are defined as music items recommended to a user and rated with the highest value by that user. Since the rating value of test music is known, determining useful recommendations is straightforward. The accuracy of the recommendations is then calculated as the proportion of useful recommendations to the total recommendations for the users. This ratio serves as a measure of the accuracy of the provided recommendations. Figs. 7 and 8 present the accuracy of the recommendations in music instruments and musicalbums datasets, respectively.

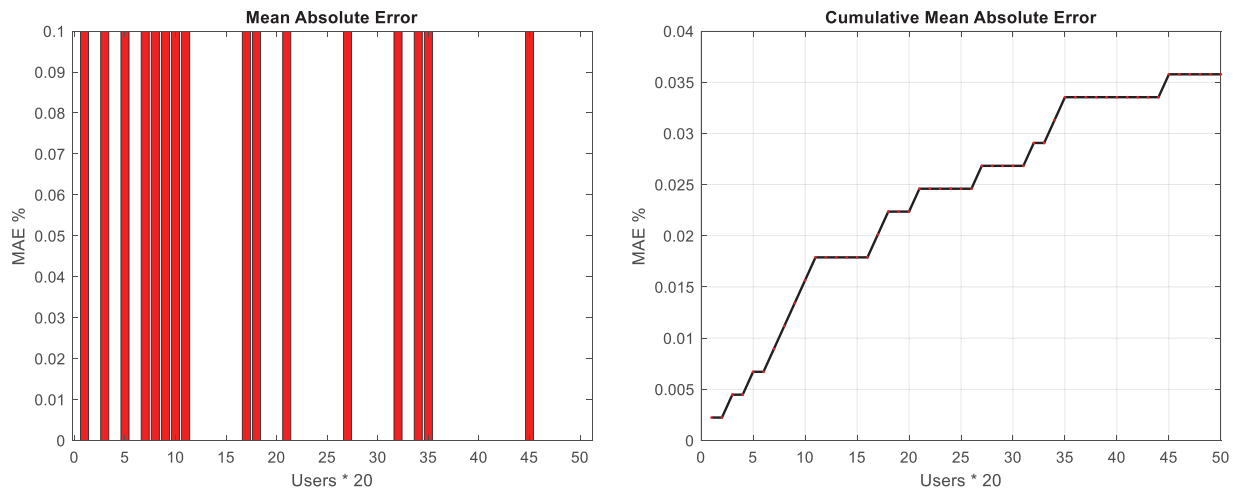


Figure 4: Mean absolute error value for test users in music album dataset

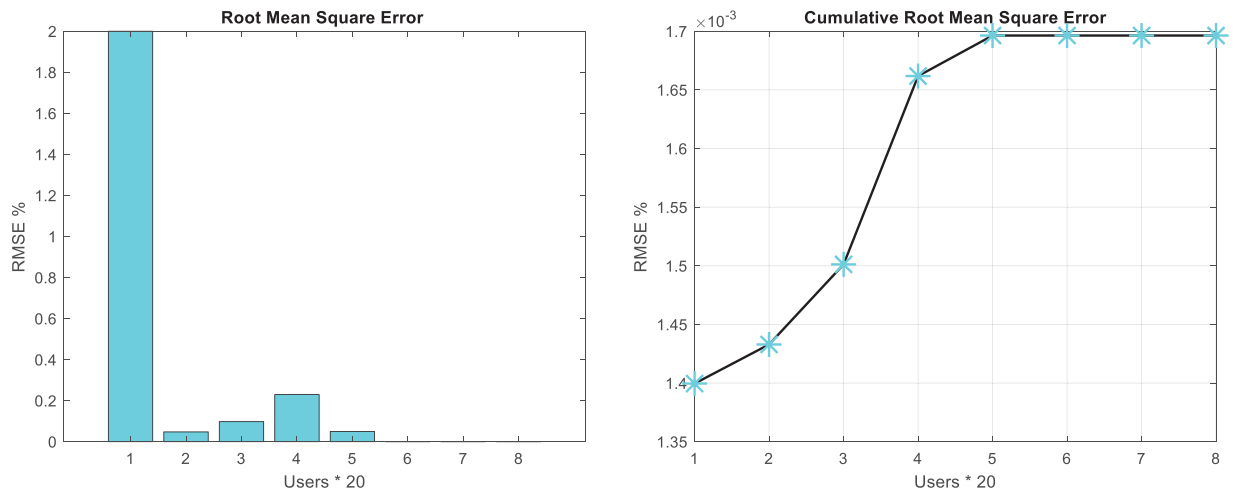


Figure 5: The value of mean squared error for test users in music instruments dataset

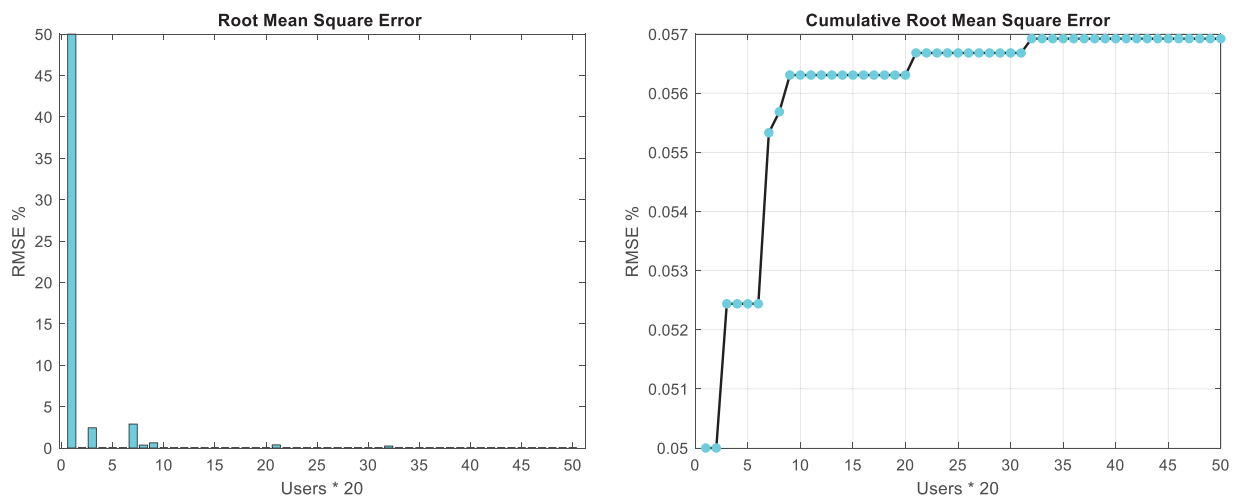


Figure 6: The value of mean squared error for test users in music album dataset

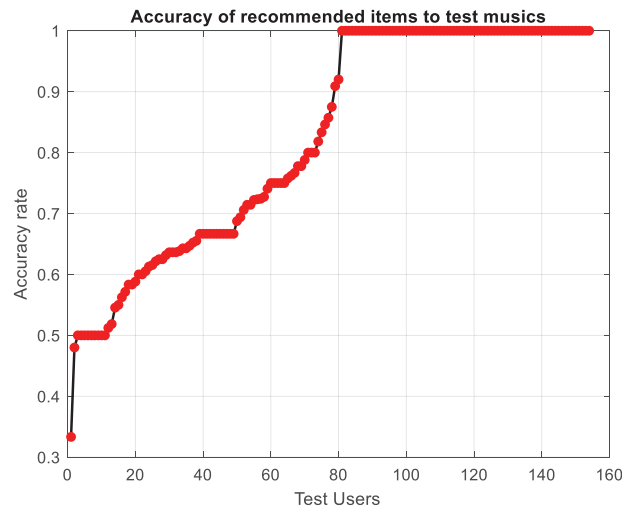


Figure 7: Accuracy of the provided recommendations in music instruments dataset

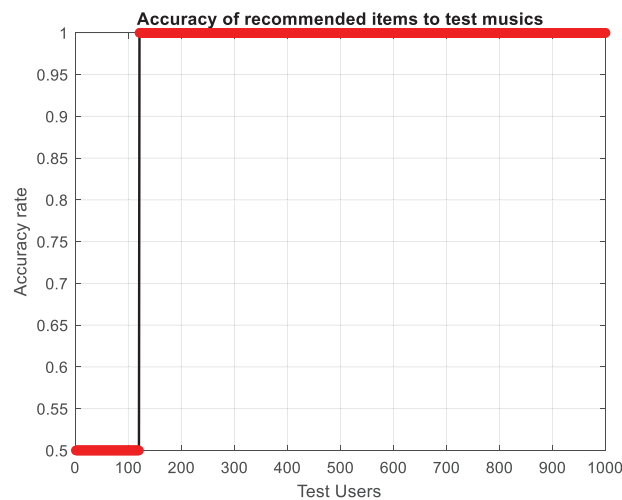


Figure 8: Accuracy of the provided recommendations in music album dataset

The analysis of Figs. 7 and 8 reveals a consistent upward trend in the accuracy of the provided recommendations in both datasets. This trend signifies the precise prediction of ratings for new music and underscores the influence of music popularity and user listening habits on the accuracy of music recommendations. The average accuracy of music recommendations in the test dataset is determined to be 82.45% in the music instruments dataset and 94% in music album dataset.

4.1.3 Performance Evaluation of Classifiers

The evaluation of the proposed method involves assessing its performance in predicting the rankings of new samples that were not included in the training process. This assessment utilizes classification methods, which categorize data into different classes, and employs various measurement criteria based on comparing predicted rankings against actual rankings, as determined by the confusion matrix. The confusion matrix yields four key parameters—True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN)—which are essential for calculating evaluation metrics.

In the context of the deep neural network's performance evaluation for users in the test dataset, the predicted rankings for the music in the dataset by the test users are measured. Test music for which users have assigned the maximum rank and the neural network also predicts the maximum rank are classified as True Positive (TP). Test users who have not assigned the maximum rank to the music and for which the neural network also does not predict the maximum rank are classified as True Negative (TN). Test users who have assigned the maximum rank to the music, but the neural network does not predict the maximum rank, are classified as False Positive (FP). Lastly, test users who have not assigned the maximum rank to the music, but the neural network predicts the maximum rank for this music, are classified as False Negative (FN). These parameters collectively constitute the components of the confusion matrix. Based on these parameters, the following evaluation criteria can be defined as Eqs. (8)–(11).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (11)$$

The evaluation metrics are considered superior to other metrics due to their ability to measure the performance of a classifier in diverse contexts. These standardized metrics enable comparisons of the presented classifier with others in the same field.

Accuracy of CNN with Various Hyper-Parameters

Hyper-parameters in a Convolutional Neural Network (CNN) can have a significant impact on the accuracy of classification tasks. Here are some ways in which different hyper-parameters can affect the accuracy of a CNN model:

Learning rate: The learning rate determines how quickly the model adapts to the training data. A too-high learning rate may cause the model to overshoot the optimal solution, while a too-low learning rate may result in slow convergence. Finding the right learning rate can help improve the accuracy of the model.

Batch size: The batch size determines how many samples are processed before updating the model's parameters. A larger batch size can lead to faster training but may result in less accurate models. On the other hand, a smaller batch size can lead to more accurate models but may require more training time. Finding the optimal batch size is crucial for improving accuracy.

Number of layers: The depth of a CNN, determined by the number of layers, can impact the model's ability to learn complex patterns in the data. Adding more layers can help the model capture intricate features but may also lead to over-fitting. Finding the right balance in the number of layers can improve the accuracy of classification tasks.

By tuning these hyper-parameters and conducting a sensitivity analysis, one can optimize the CNN model for classification tasks, leading to improved accuracy and performance. The values of evaluation criteria for CNN based on different parameters are shown in Table 3.

Table 3: Evaluation criteria for various hyper-parameters for CNN

Hyper-parameters		Accuracy		Recall		Precision		F-measure	
		Music instruments	Music album	Music instruments	Music album	Music instruments	Music album	Music instruments	Music album
Learning rate	0.05	99.15	94.3	99.15	90.42	99.34	95.19	99.24	92.74
	0.02	99.23	97.34	99.22	93.04	99.37	95.27	99.30	94.14
	0.01	99.32	95.91	99.28	95.34	99.27	91.32	99.28	93.29
	0.009	99.21	97.66	99.25	94.87	99.20	94.88	99.22	94.88
	0.008	99.24	92.39	99.30	91.31	99.21	92.67	99.25	91.99
	0.005	99.23	95.05	99.17	96.74	99.24	94.32	99.20	95.51
	0.002	99.39	96.36	99.23	91.26	99.26	93.12	99.24	92.18
	0.001	99.32	92.77	99.35	96.1	99.23	93.88	99.29	94.98
Batch size	25	99.21	95.66	99.21	96.11	99.18	92.83	99.19	94.44
	35	99.22	96.62	99.22	90.98	99.36	96.86	99.29	93.83
	50	99.36	94.1	99.34	93.8	99.11	91.01	99.22	92.38
	70	99.27	97.5	99.22	92.78	99.27	96.73	99.24	94.71
	100	99.28	93.74	99.29	96.47	99.28	93.86	99.28	95.15
Number of layers	50	99.23	97.82	99.28	91.28	99.27	94.58	99.27	92.9
	70	99.29	92.07	99.34	92.19	99.17	93.97	99.25	93.07
	80	99.25	91.71	99.21	90.14	99.30	94.71	99.25	92.37
	90	99.40	94.74	99.30	90.41	99.30	97.92	99.30	94.02
	100	99.34	95.69	99.32	91.45	99.24	94.62	99.28	93.01

As can be seen in Table 3, the changes in the hyper-parameters in the CNN structure can contribute to the accuracy of predicting the ranking of new users. Therefore, in the proposed method, we choose the best combination for CNN hyper-parameters, which includes a learning rate equal to 0.002, a batch size equal to 50, and a number of layers equal to 90.

Accuracy of New Rating Prediction

In the proposed method, the convolutional neural networks are utilized alongside three well-known classifiers—K-Nearest Neighbor, Decision Tree, Support Vector Machine, and Naive Bayes classifier—to predict the ranking of test users. These classifiers are then compared with the convolutional neural networks. Fig. 5 illustrates the accuracy metrics and presents a bar chart showcasing the average accuracy across different classifiers.

Figs. 9 and 10 depicts a convolutional neural network model based on user comments and ratings, demonstrating superior performance in accuracy compared to other classification models in music instruments and music albums datasets, respectively. Fig. 9 presents 450 and Fig. 10 represents 300 test samples, with accuracy predictions for every 10 tests displayed to enhance clarity and avoid overcrowding on the accuracy plot. This approach aims to provide a clear representation of accuracy and other evaluation metrics. Furthermore, Fig. 6 introduces the sensitivity metric, which quantifies the ratio of correctly predicted maximum rankings among all maximum rankings in the test data. The figure showcases the sensitivity metric and features a bar chart illustrating the average sensitivity across different classification models.

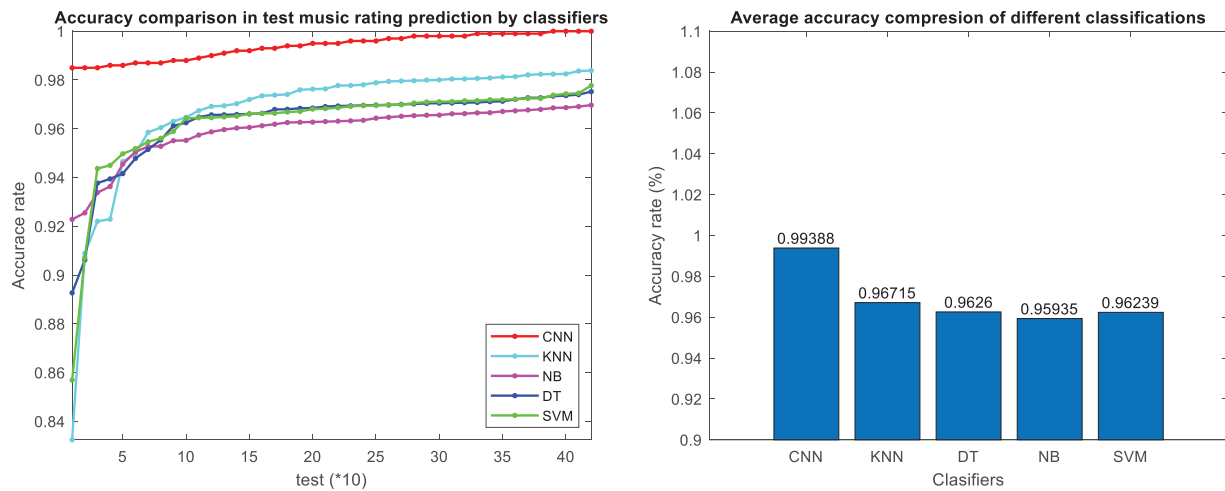


Figure 9: Comparison of accuracy criteria in classification methods in music instruments dataset

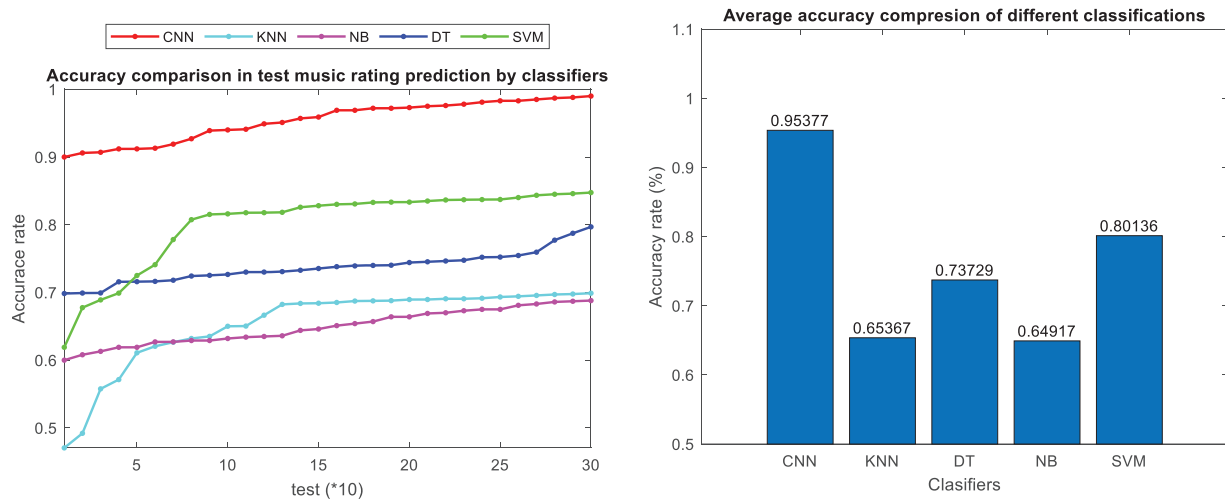


Figure 10: Comparison of accuracy criteria in classification methods in music album dataset

The convolutional neural network approach, utilizing users' comments and ratings, demonstrates superior recall performance compared to other classification methods, as illustrated in Figs. 11 and 12 for music instruments and musicalbums datasets, respectively. Additionally, the figure introduces the accuracy metric, which signifies the correctness of predicted maximum rankings among the maximum rankings of test samples within the proposed classification model. Figs. 13 and 14 present the precision metric and the average accuracy across different classification models for music instruments and musicalbums datasets, respectively.

As seen in Figs. 13 and 14, the presented convolutional neural network approach based on users' comments and ratings in the presented approach has shown superior performance in terms of precision metrics compared to other classification methods in both datasets. Additionally, the F-measure, which is the harmonic mean of precision and sensitivity, has been utilized to assess the overall performance of the model in ranking music in the test dataset. Figs. 15 and 16 present the F-measure metric and a bar chart depicting the average F-measure across different classification models for music instruments and musicalbums datasets, respectively.

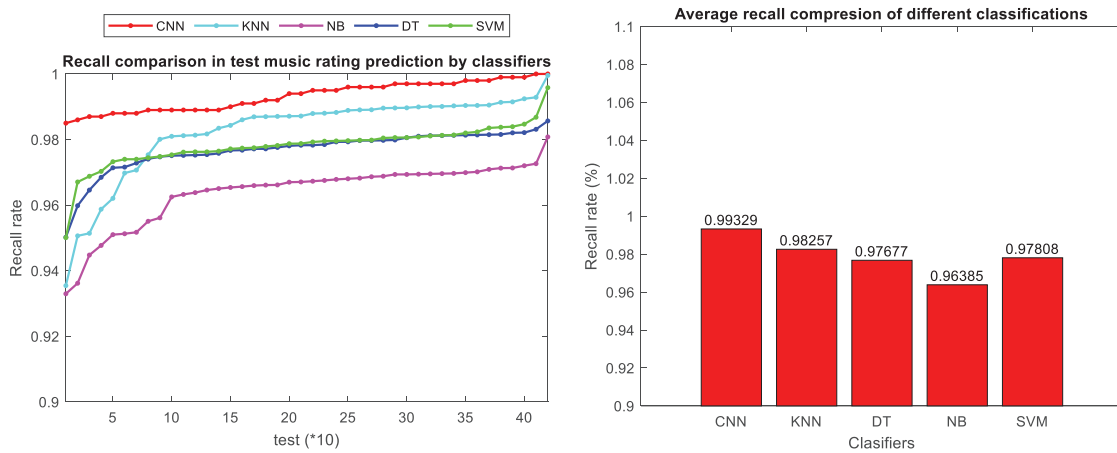


Figure 11: Comparison of recall criteria in classification methods in music instruments dataset

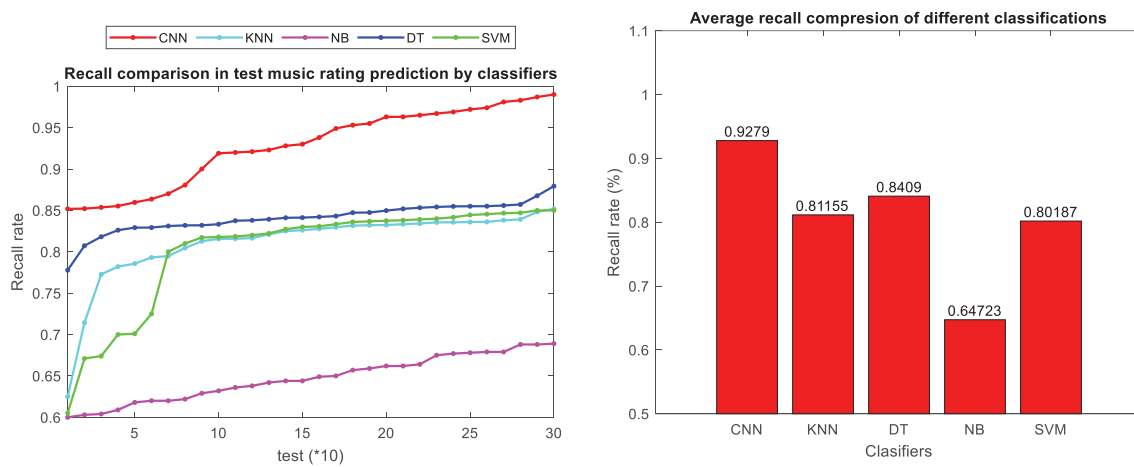


Figure 12: Comparison of recall criteria in classification methods in music album dataset

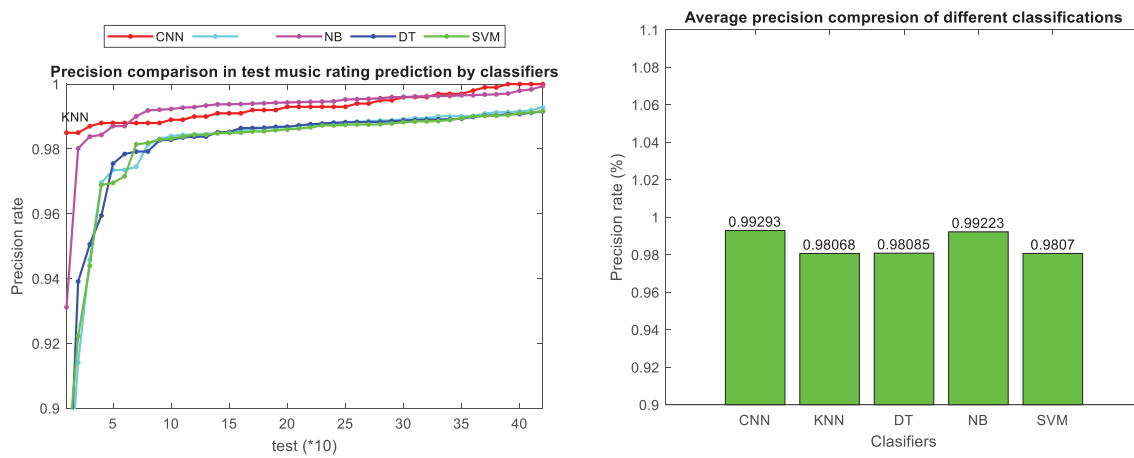


Figure 13: Comparison of precision criteria in classification methods in music instruments dataset

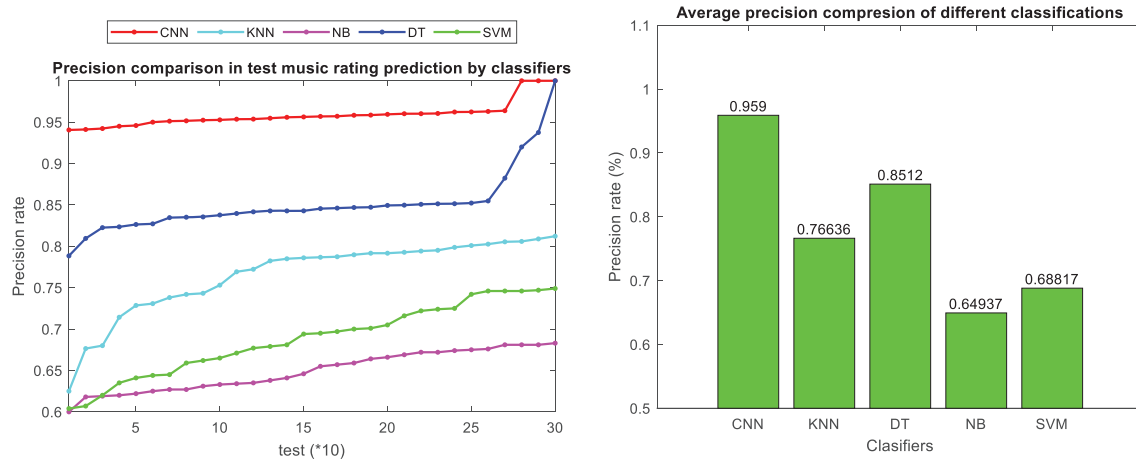


Figure 14: Comparison of precision criteria in classification methods in music album dataset

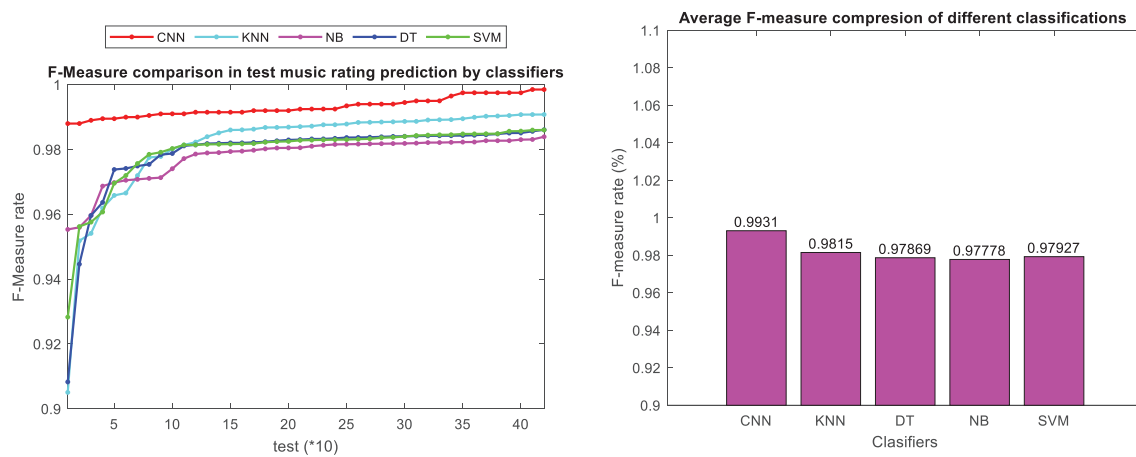


Figure 15: Comparison of F-measure in classification methods in music instruments dataset

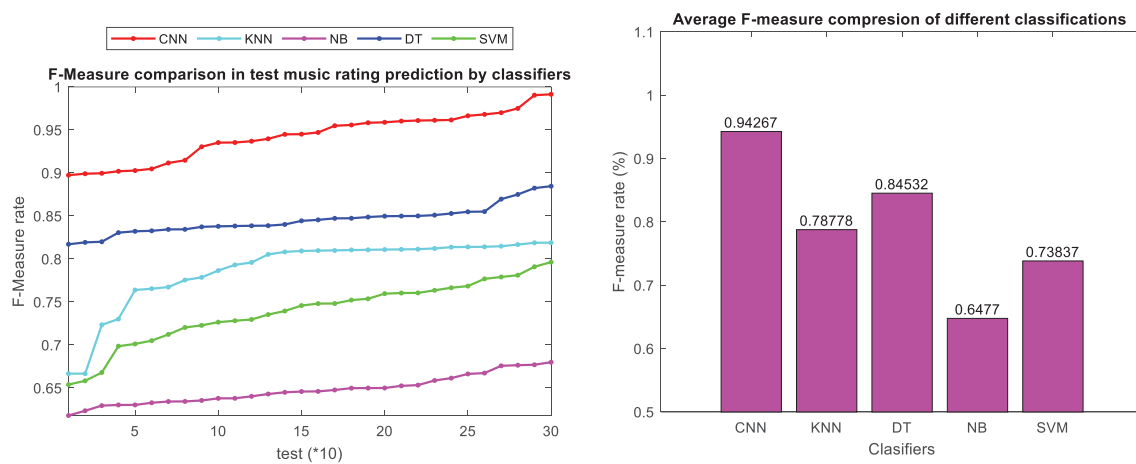


Figure 16: Comparison of F-measure in classification methods in music album dataset

The presented convolutional neural network, utilizing users' comments and ratings, has demonstrated superior performance in terms of the F-measure metric compared to other classification methods, as depicted in Figs. 15 and 16 for music instruments and musicalbums datasets, respectively. The results indicate that the approach employing users' ratings and rankings has provided highly suitable training data for convolutional neural network classification. Consequently, the convolutional neural network has achieved high accuracy and precise predictions on the specific features of this dataset due to the appropriateness of the training data. Among the classification methods, the convolutional neural network has produced the most suitable results for classifying the training instances and predicting user rankings in the test dataset. Additionally, Figs. 17 and 18 present a bar chart comparing the aggregated evaluation metrics for different classification methods for music instruments and musicalbums datasets, respectively.

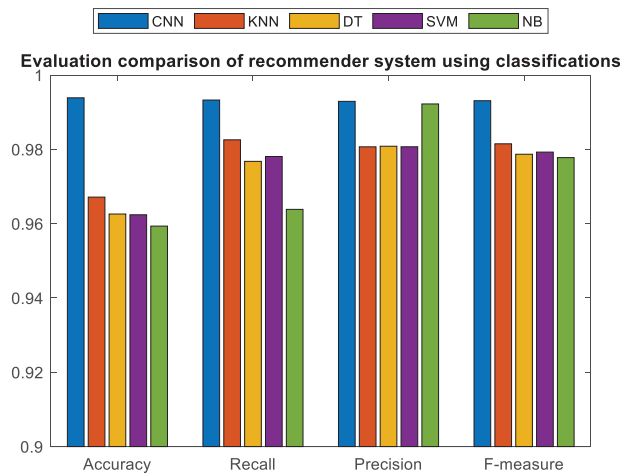


Figure 17: Bar chart of cumulative comparison of evaluation criteria in music instruments dataset

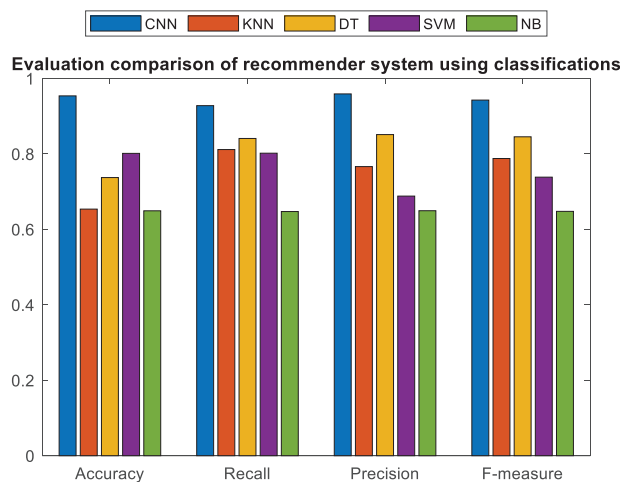


Figure 18: Bar chart of cumulative comparison of evaluation criteria in music album dataset

In Figs. 17 and 18, convolutional neural networks utilizing deep learning techniques have shown superior accuracy in predicting product ratings for new users compared to other classification methods in both

datasets. The deep learning layers and completion of the training process in the middle layer contribute to its high performance in evaluation metrics

Normalized Discounted Cumulative Gain for the Proposed Method

Normalized Discounted Cumulative Gain (nDCG) is a metric used to evaluate the effectiveness of a ranking algorithm in information retrieval tasks, such as search engine results or recommendation systems. nDCG takes into account both the relevance of the documents retrieved by the algorithm and their position in the ranking. It is based on the idea that highly relevant documents should be ranked higher and that the relevance of a document decreases as its position in the ranking decreases. To calculate *nDCG*, the Discounted Cumulative Gain (DCG) is first calculated by summing the relevance scores of the documents at each position in the ranking, discounted by a logarithmic function of their position. The DCG is then normalized by dividing it by the ideal DCG, which is the DCG of a perfect ranking where all relevant documents are ranked at the top. The formula for *nDCG* is [33]:

$$nDCG = \frac{DCG}{IDCG} \quad (12)$$

$$DCG = \sum_{k=1}^K \frac{rel_i}{\log_2(i+1)} \quad (13)$$

where *IDCG* is the ideal *DCG*, *K* is the number of users, and *i* is the number of music. A higher *nDCG* score indicates a better ranking algorithm as it means that the algorithm is retrieving more relevant documents and ranking them higher in the list. It is a widely used metric in information retrieval and machine learning to evaluate and compare different ranking algorithms. In the proposed method, to generalize the proposed idea, various scenarios have been used with the presence of a portion of users to evaluate *nDCG*. In these scenarios, the α value represents a part of the total users who are selected as test users. Fig. 10 shows the diagrams of *nCGD* related to the proposed method for different values of α .

According to Figs. 19 and 20, it can be seen that with the increase in the number of test users, the growth trend of *nDCG* is still upward in both datasets. These graphs show that the proposed method of finding music related to users' preferences is performing well. The recommendations given to users have been accepted to a large extent. Table 4 shows the average values of *nDCG* according to the number of users in different scenarios.

Based on the average values of *nDCG* in Table 4, it can be seen that the proposed method has provided good recommendations to the test users and the average number of favorite music of the users is high.

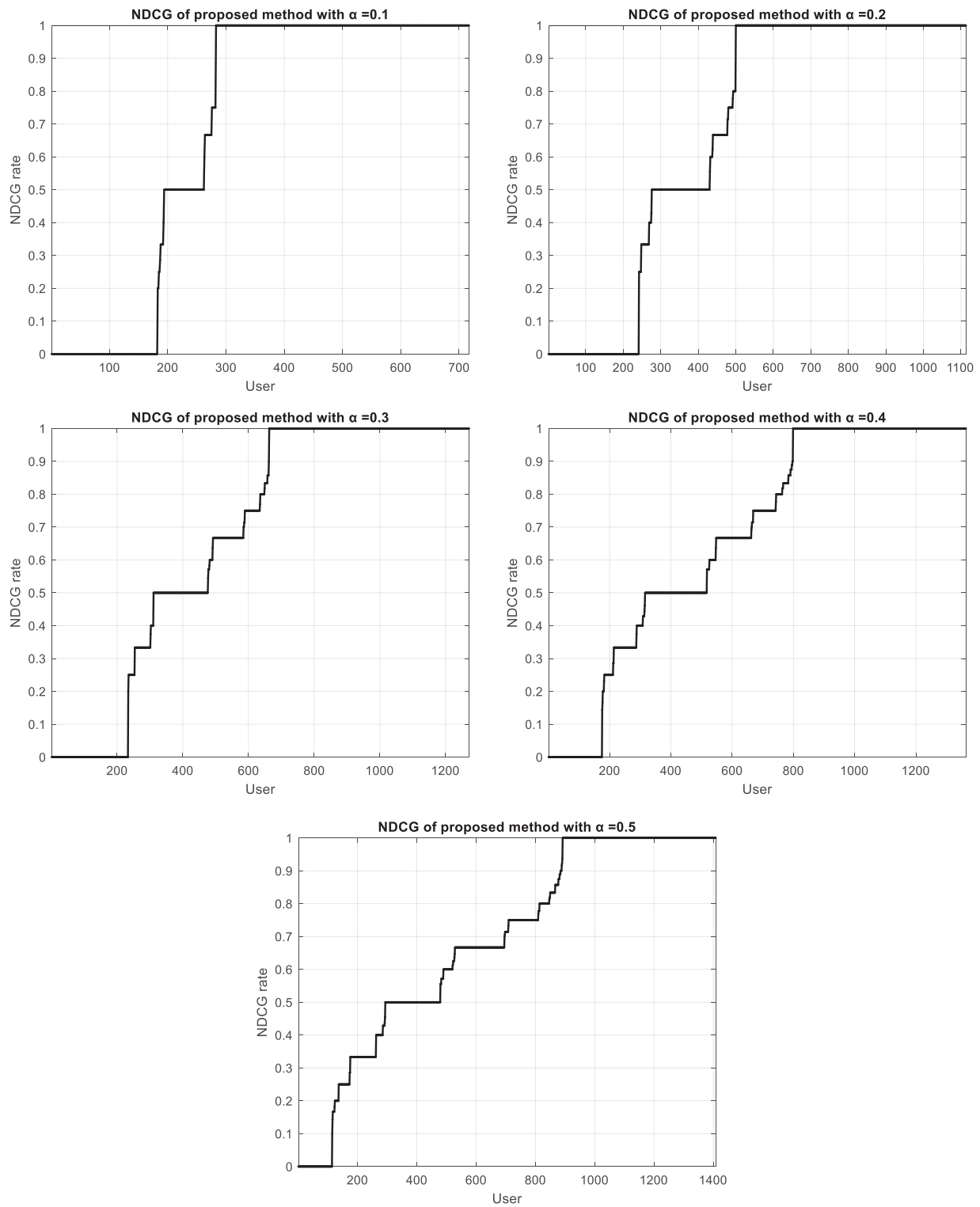


Figure 19: The n DCG of the proposed method for different values of α in music instruments dataset

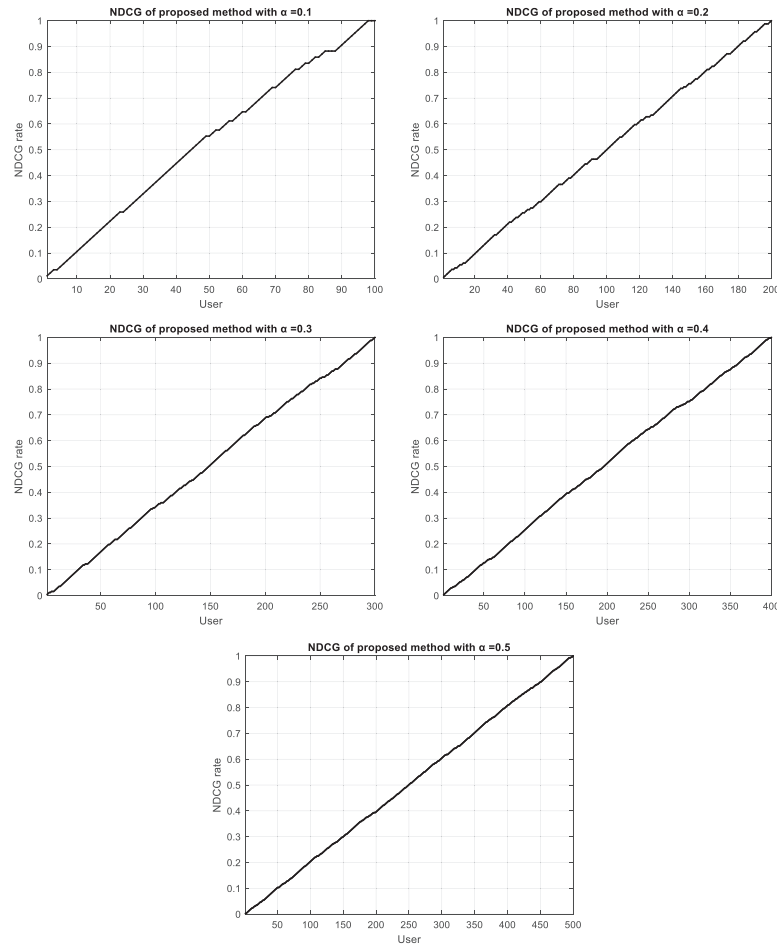


Figure 20: The $nDCG$ of the proposed method for different values of α in music album dataset

Table 4: Average $nDCG$ of the proposed method for different values of α

0.1	0.8229	0.7432
0.2	0.8211	0.7617
0.3	0.8053	0.7490
0.4	0.7726	0.7492
0.5	0.7414	0.7549

4.2 Comparison of the Proposed Method with Previous Methods

The study plans to validate the proposed method by comparing it with other recommender systems. The most famous of these methods include multimedia recommender systems based on collaborative filtering [19], convolutional neural networks [34], recurrent networks [35] and content-based recurrent neural networks [36], genre personalization [37], popularity [38], sequential and content-based sequential models [39,40], and dynamic time-based models [41].

The results obtained from the proposed method will be compared with other methods to assess the improvement resulting from the proposed idea. The comparison will utilize metrics such as mean absolute

error, hit rate, and ranking accuracy, given the use of convolutional neural learning in the proposed method. Fig. 21 illustrates the comparison of the proposed method with previous methods in terms of the mean absolute error metric.

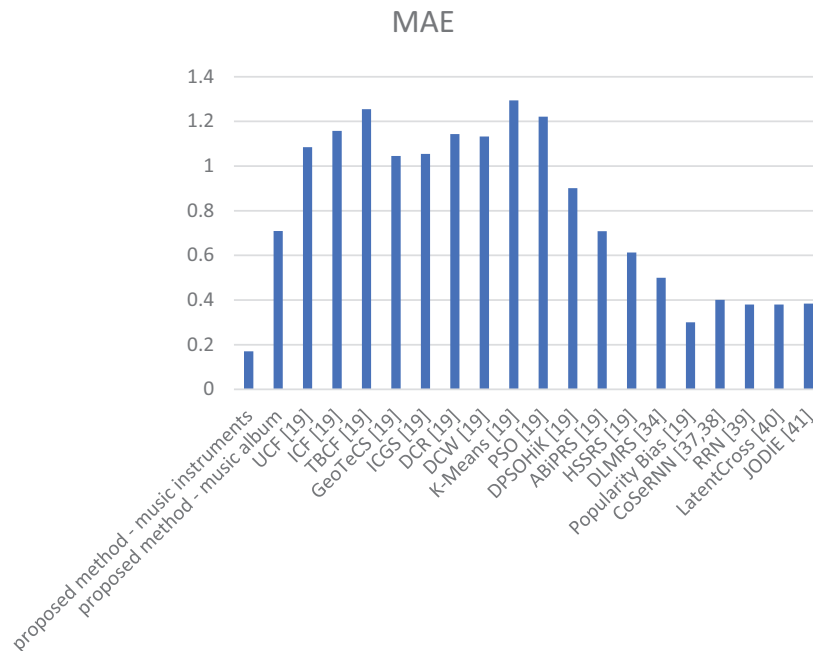


Figure 21: Comparison of the proposed method with previous methods in terms of mean absolute error [19,34,37–41]

Fig. 21 demonstrates that the proposed method has achieved a lower mean absolute error compared to previous methods. This outcome is attributed to the deep learning of user review features, accurate identification of similar users through order clustering, and precise rating prediction for new music using the proposed convolutional neural network. Consequently, it is anticipated that the accuracy of recommendations provided by the proposed method, measured as hit rate, will surpass that of other existing methods due to the low mean absolute error. In Fig. 22, a comparison of the proposed method with previous methods in terms of hit rate or accuracy of recommendations provided to users will be presented.

Fig. 22 illustrates that the proposed method, leveraging deep learning in convolutional neural networks, has effectively trained user review features, user listening rates, and music popularity rates. This has led to a reduction in mean absolute error in providing recommendations and an increase in hit rate or accuracy of recommendations provided. The proposed method has achieved a higher hit rate compared to previous recommender system methods. Additionally, the proposed method can be compared to previous methods in terms of the accuracy of ranking prediction, considering the utilization of deep learning and other classification methods in users' rating for new music. In the proposed method, alongside the convolutional neural network, other classification methods such as k-nearest neighbor (KNN), Decision Tree (DT), Naïve Bayes (NB), and support vector machine (SVM) have been employed to predict ratings based on user reviews. The comparison of the proposed method with other classification methods and previous methods in terms of ranking prediction accuracy is presented in Fig. 23.

The findings depicted in Fig. 23 indicate that the proposed method, leveraging deep learning on user features, opinions, and ratings, has achieved superior accuracy in predicting new ratings compared to other classification methods and previous methods.

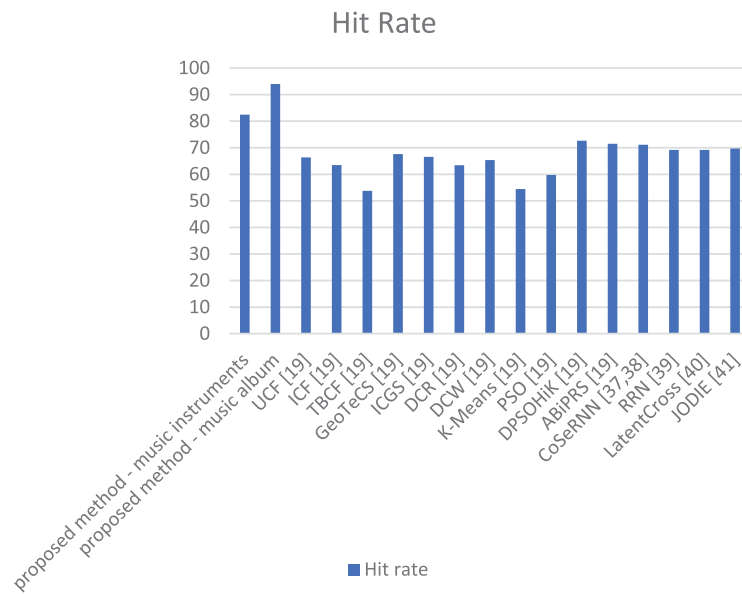


Figure 22: Comparison of the proposed method with previous methods in terms of hit rate [19,37–41]

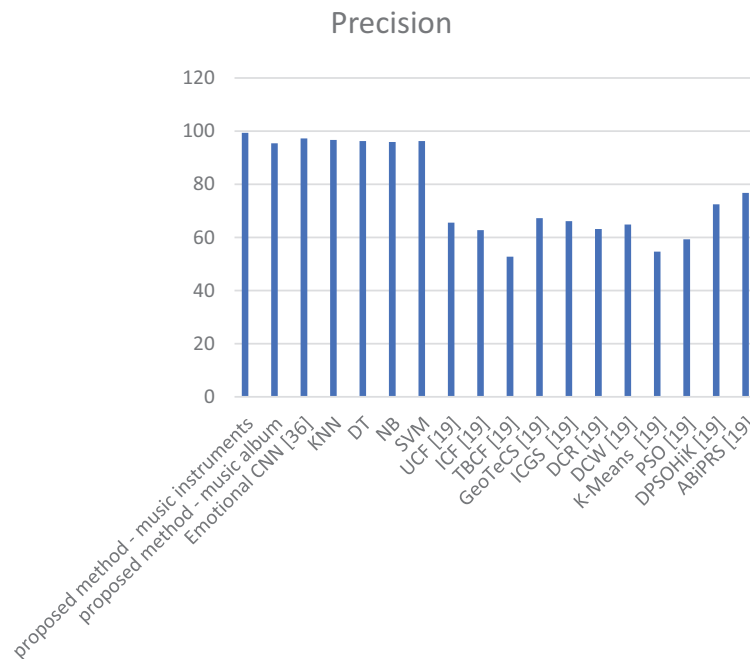


Figure 23: Comparison of the proposed method with the previous methods in terms of accuracy of rating prediction [19]

5 Discussion

In the proposed method, particular attention has been given to analyzing users' music listening habits to align music recommendations with their typical behavior. Consequently, high-popularity tracks are recommended to regular users only if they are compatible with their established listening patterns. This approach enables the system to provide recommendations that not only reflect general popularity but also

resonate with the users' individual listening preferences. However, for cold-start users, the absence of data on their music listening habits poses a significant challenge. As a default solution, the system recommends popular tracks to this group of users. While this strategy may be effective in some cases, it might not align with the actual preferences of cold-start users, potentially resulting in a suboptimal user experience.

To address this challenge and improve the accuracy of the recommendation system, it is suggested to incorporate demographic data such as age, gender, or location. These data points can serve as indirect indicators of cold-start users' initial preferences. For instance, analyzing user profiles on social networks may provide insights into their interests, enabling the system to deliver more personalized recommendations. This approach not only enhances the system's accuracy but also offers a more satisfying user experience for cold-start users. Therefore, extending research in this direction represents a promising avenue for the future development of recommendation systems.

To make the methodology more dynamic, it is crucial to address system adaptability and the cold start problem. As users provide more points or interactions, the system can utilize incremental learning techniques to update clustering models without requiring a complete retraining. Adaptive algorithms, such as online clustering or dynamic centroid updating, can efficiently accommodate new user data, ensuring that the clustering structure remains relevant. To overcome the cold start problem, which arises when new users or items lack sufficient data, hybrid recommendation strategies that combine content-based and collaborative filtering methods can be employed. Additionally, using demographic information or pre-trained models can help generate initial recommendations, gradually improving as more user-specific interactions are collected.

6 Conclusion

Music recommender systems are designed to employ algorithms and machine learning models to provide personalized music suggestions to users based on their individual tastes and preferences. These systems leverage data gathered from users, including their listening habits and music ratings, to generate music recommendations. A key advantage of music recommender systems is their ability to introduce users to new music that aligns with their tastes, potentially exposing them to songs or artists they may not have encountered otherwise. By analyzing users' music preferences and behaviors, these systems facilitate the discovery of new music that resonates with their individual preferences, enhancing their overall music listening experience. This paper discusses the significance of music recommender systems in leveraging algorithms and machine learning models to suggest music aligned with users' preferences. These systems enable users to discover new music that suits their tastes based on their listening rates and music popularity. The study introduces a hybrid music recommender system that combines sequential clustering and convolutional neural networks using user reviews and rankings. Initially, users are sequentially clustered based on semantic similarity, followed by the utilization of convolutional neural networks to predict rankings for unexplored music based on users' reviews. The method also considers the user's music listening behavior and music popularity to recommend the highest-ranked and most popular unexplored music to each user. The proposed method outperforms other classification methods and previous recommender systems, achieving a mean absolute error and mean squared error of approximately 0.0017, a hit rate of 82.45%, and a new ranking prediction accuracy of 99.388%.

Acknowledgement: Not applicable.

Funding Statement: This research was funded by the National Nature Sciences Foundation of China with Grant No. 42250410321.

Author Contributions: Conceptualization: Weitao Ha, Yahya D. Navaei. Methodology: Weitao Ha, Yahya D. Navaei. Software: Weitao Ha, Sheng Gang. Validation: Sheng Gang, Yahya D. Navaei. Formal Analysis: Weitao Ha, Abubakar S. Gezawa, Yaser A. Nanekharan. Investigation: Weitao Ha, Sheng Gang. Resources: Weitao Ha, Yaser A. Nanekharan. Data Curation: Weitao Ha, Yahya D. Navaei. Writing—Original Draft Preparation: Weitao Ha. Writing—Review & Editing: Weitao Ha, Yahya D. Navaei, Yaser A. Nanekharan. Visualization: Weitao Ha, Sheng Gang, Abubakar S. Gezawa. Supervision: Sheng Gang, Yahya D. Navaei. Project Administration: Sheng Gang, Yahya D. Navaei. Funding Acquisition: Sheng Gang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in Musical instruments reviews at <https://www.kaggle.com/datasets/eswarchandt/amazon-music-reviews> and Music Album Reviews in <https://www.kaggle.com/datasets/michaelbryantds/78k-music-album-reviews> (accessed on 1 January 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Galety MG, Thiagarajan R, Sangeetha R, Vignesh LB, Arun S, Krishnamoorthy R. Personalized music recommendation model based on machine learning. In: 2022 8th International Conference on Smart Structures and Systems (ICSSS); 2022 Apr 21–22; Chennai, India.
2. Roy D, Dutta M. A systematic review and research perspective on recommender systems. *J Big Data*. 2022;9(1):59. doi:10.1186/s40537-022-00592-5.
3. Nagaraj P, Muneeswaran V, Kumar BM, Rao KRK, Nagaraju MS, Kumar MVH. Comparative analysis of different approaches to the music recommendation system. In: 2023 International Conference on Computer Communication and Informatics (ICCCI); 2023 Jan 23–25; Coimbatore, India.
4. Deldjoo Y, Schedl M, Knees P. Content-driven music recommendation: evolution, state of the art, and challenges. *arXiv:2107.11803*. 2021.
5. Gundabatini DSG, Bindu SR, Vamsi M, Yagna YN, Krishna VK. Music recommendation system using machine learning. *Int J Innov Eng Manag Res*. 2023;12(4).
6. Navaei YD, Rezvani MH, Moghaddam AME. A novel neighborhood-based importance measure for social network influence maximization using NSGA-III. In: 2024 10th International Conference on Artificial Intelligence and Robotics (QICAR); 2024 Feb 29; Qazvin, Islamic Republic of Iran.
7. Schedl M, Knees P, McFee B, Bogdanov D. Music recommendation systems: techniques, use cases, and challenges. In: *Recommender systems handbook*. New York, NY, USA: Springer; 2021. p. 927–71.
8. Guan J, Chen B, Yu S. A hybrid similarity model for mitigating the cold-start problem of collaborative filtering in sparse data. *Expert Syst Appl*. 2024;249:123700. doi:10.1016/j.eswa.2024.123700.
9. Shurrah M, Mahboobeh D, Mizouni R, Singh S, Otrók H. Overcoming cold start and sensor bias: a deep learning-based framework for IoT-enabled monitoring applications. *J Netw Comput Appl*. 2024;222:103794. doi:10.1016/j.jnca.2023.103794.
10. Darvishy A, Ibrahim H, Sidi F, Mustapha A. HYPNER: a hybrid approach for personalized news recommendation. *IEEE Access*. 2020;8:46877–94. doi:10.1109/ACCESS.2020.2978505.
11. Bai X, Wang M, Lee I, Yang Z, Kong X, Xia F. Scientific paper recommendation: a survey. *IEEE Access*. 2019;7:9324–39. doi:10.1109/ACCESS.2018.2890388.
12. De Prisco R, Guarino A, Malandrino D, Zaccagnino R. Induced emotion-based music recommendation through reinforcement learning. *Appl Sci*. 2022;12(21):11209. doi:10.3390/app122111209.
13. Singh J, Sajid M, Yadav CS, Singh SS, Saini M. A novel deep neural-based music recommendation method considering user and song data. In: 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI); 2022 Apr 28–30; Tirunelveli, India.
14. Park M, Lee K. Exploiting negative preference in content-based music recommendation with contrastive learning. In: *Proceedings of the 16th ACM Conference on Recommender Systems*; 2022 Sep 18–23; Seattle, WA, USA.

15. Sarin E, Vashishtha S, Kaur S. SentiSpotMusic: a music recommendation system based on sentiment analysis. In: 2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCSST); 2022 Feb 11–12; Jamshedpur, India.
16. Milano S, Taddeo M, Floridi L. Recommender systems and their ethical challenges. *AI Soc.* 2020;35:957–67. doi:10.1007/s00146-020-00950-y.
17. Velankar M, Kulkarni P. Music recommendation systems: overview and challenges. *Adv Speech Music Technol Comput Asp Appl.* 2022;51–69.
18. Afchar D, Melchiorre A, Schedl M, Hennequin R, Epure E, Moussallam M. Explainability in music recommender systems. *AI Mag.* 2022;43(2):190–208. doi:10.1002/aaai.12056.
19. Kowald D, Lacic E. Popularity bias in collaborative filtering-based multimedia recommender systems. In: International Workshop on Algorithmic Bias in Search and Recommendation; 2022 Apr 10; Stavanger, Norway. Berlin/Heidelberg, Germany: Springer.
20. Gazdar A, Hidri L. A new similarity measure for collaborative filtering based recommender systems. *Knowl-Based Syst.* 2020;188:105058. doi:10.1016/j.knosys.2019.105058.
21. Tang H, Zhao G, Bu X, Qian X. Dynamic evolution of multi-graph based collaborative filtering for recommendation systems. *Knowl-Based Syst.* 2021;228:107251. doi:10.1016/j.knosys.2021.107251.
22. Aljunid MF, Dh M. An efficient deep learning approach for collaborative filtering recommender system. *Procedia Comput Sci.* 2020;171:829–36. doi:10.1016/j.procs.2020.04.090.
23. Fessahaye F, Perez L, Zhan T, Zhang R, Fossier C, Markarian R, et al. T-RECSYS: a novel music recommendation system using deep learning. In: 2019 IEEE International Conference on Consumer Electronics (ICCE); 2019 Jan 11–13; Las Vegas, NV, USA.
24. Feng C, Liang J, Song P, Wang Z. A fusion collaborative filtering method for sparse data in recommender systems. *Inf Sci.* 2020;521:365–79. doi:10.1016/j.ins.2020.02.052.
25. Zhang Q, Lu J, Jin Y. Artificial intelligence in recommender systems. *Complex Intell Syst.* 2021;7:439–57. doi:10.1007/s40747-020-00212-w.
26. Da'u A, Salim N. Recommendation system based on deep learning methods: a systematic review and new directions. *Artif Intell Rev.* 2020;53(4):2709–48. doi:10.1007/s10462-019-09744-1.
27. Wu D. Music personalized recommendation system based on hybrid filtration. In: 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS); 2019 Jan 12–13; Changsha, China.
28. Wang WZ. Personalized music recommendation algorithm based on hybrid collaborative filtering technology. In: 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA); 2019 Aug 10–11; Xiangtan, China.
29. Wang Y. A hybrid recommendation for music based on reinforcement learning. In: Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020; 2020 May 11–14; Singapore.
30. Dhruv A, Kamath A, Powar A, Gaikwad K. Artist recommendation system using hybrid method: a novel approach. In: Emerging research in computing, information, communication and applications: ERCICA 2018. Vol. 1. Berlin/Heidelberg, Germany: Springer; 2019.
31. Reviews Ma. Dataset. [cited 2025 Jan 25]. Available from: <https://www.kaggle.com/datasets/michaelbryantds/78k-music-album-reviews>.
32. Manco I, Weck B, Bogdanov D, Tovstogan P, Won M. Song describer dataset (1.0.0). [cited 2025 Jan 1]. Available from: <https://zenodo.org/records/10072001>.
33. Jeunen O, Potapov I, Ustimenko A. On (Normalised) discounted cumulative gain as an off-policy evaluation metric for top-n recommendation. *arXiv:230715053*. 2023.
34. Priyanka VT, Reddy YR, Vajja D, Ramesh G, Gomathy S. A novel emotion based music recommendation system using CNN. In: 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS); 2023 May 17–19; Madurai, India.
35. Wu CY, Ahmed A, Beutel A, Smola AJ, Jing H. Recurrent recommender networks. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining; 2017 Feb 6–10; New York, NY, USA.

36. Beutel A, Covington P, Jain S, Xu C, Li J, Gatto V, et al. Latent cross: making use of context in recurrent recommender systems. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining; 2018 Feb 5–9; Los Angeles, CA, USA.
37. Liang Y, Willemsen MC. Promoting music exploration through personalized nudging in a genre exploration recommender. *Int J Hum-Comput Interact.* 2023;39(7):1495–518. doi:10.1080/10447318.2022.2108060.
38. Kowald D, Schedl M, Lex E. The unfairness of popularity bias in music recommendation: a reproducibility study. In: *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020*; 2020 Apr 14–17; Lisbon, Portugal.
39. Hansen C. Sequential modelling with applications to music recommendation, fact-checking, and speed reading. arXiv:210906736. 2021.
40. Hansen C, Hansen C, Maystre L, Mehrotra R, Brost B, Tomasi F, et al. Contextual and sequential user embeddings for large-scale music recommendation. In: *Proceedings of the 14th ACM Conference on Recommender Systems*; 2020 Sep 22–26; New York, NY, USA.
41. Kumar S, Zhang X, Leskovec J. Predicting dynamic embedding trajectory in temporal interaction networks. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*; 2019 Aug 4–8; New York, NY, USA.