Musical genre classifier – a comparison of different methods

Lucian-Daniel Jimon  
UTCN – TM 1st yearjimon.mi.lucian@student.utcluj.ro

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords—component, formatting, style, styling, insert (key words)

# Introduction

Musical genre classification is a complex, but important problem in the domain of MIR (music information retrieval), being crucial for several other tasks like music recommendations, audio catalogs, playlist generation etc.

The current paper proposes to compare popular machine learning algorithms for solving the mentioned task, namely the SVM, Random Forrest and CNN algorithms, all which will be used on the GTZAN dataset, which is possibly the most used free online dataset for musical genre classification. This paper aims to provide a comprehensive comparison of SVM, RF, and CNN models for the task of musical genre classification using the well-known GTZAN dataset. We employ both traditional machine learning techniques and modern deep learning approaches to evaluate their performance in terms of classification accuracy, training time, and computational efficiency. The study involves several key steps, which are:

* Feature Extraction: A rich set of features is extracted from the audio files, including Mel-frequency cepstral coefficients (MFCCs), chroma features, spectral contrast etc. These features are then used to train SVM and RF classifiers
* Deep Learning Approach: For the CNN model, the spectrograms are used as input to capture the temporal and spectral characteristics of the audio signals
* Model Training and Evaluation: Each model is trained on the GTZAN dataset, and their performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. Additionally, an analysis of misclassifications is performed

By comparing these methodologies, the advantages and trade-offs associated with each model are aimed. This comparison could even provide valuable insight for future works.

# State-of-the-art

As being a crucial part of other more complex music related tasks like music generation, there are already well-based studies on genre classification, some of them are presented below.

## Music genre classification with ResNet

When talking about complex classification problems which require a robust CNN architecture, a starting point can be the use of pretrained ResNet models which can be deployed and fine-tuned for the current task []. In [] the authors managed to combine ResNet with BiGRU (Bi-gated recurrent unit), obtaining a final accuracy of 90%, which situates it as one of the best performances for genre classification.

## Comparison of music genre classifiers

Another paper treats the topic in a similar manner to the current one: comparing SVM, RF and CNN []. In this paper, a multi-model CNN is used based on spectrograms of the GTZAN dataset, obtaining a result for accuracy of 87%, 90% and 88% for CNN, SVM and RF.

## Comparison of music genre classifiers

The paper [] is treating the topic of music genre classification once again in an investigation manner, presenting a 3-layern CNN model, a SVM using an RBF (radial basis function) as kernel and the classic KNN (K-Nearest Neighbors) algorithms. The results show a significant difference in the preparation of the data, e.g. for the CNN the final accuracy was 82% if the data was preprocessed, and 53% otherwise.

# Theoretical foundations

## Introduction to Musical Genres

Musical genres play a pivotal role in the classification and organization of music. Each genre encapsulates distinct musical characteristics, including specific rhythmic patterns, harmonic structures, melodic contours, and instrumentation. These genres are crucial for various applications, including music recommendation systems, automated playlist generation, and enhancing user experience in digital music platforms. Understanding and correctly identifying musical genres enable efficient navigation and enjoyment of vast music libraries.

## Data Preprocessing and Feature Extraction

The preprocessing of audio data is essential to convert raw audio signals into a suitable format for machine learning models. Raw audio is typically high-dimensional and contains noise, which can degrade the performance of classifiers. Preprocessing steps such as normalization, silence removal, and noise reduction improve the signal-to-noise ratio, making the data more tractable.

* Feature extraction transforms raw audio data into a set of numerical features that capture the essential characteristics of the audio signal. Some of the key features include:
* Mel-frequency Cepstral Coefficients (MFCCs): MFCCs are derived from the Fourier transform of the audio signal and represent the short-term power spectrum. They are particularly effective in capturing the timbral texture of the audio.
* Chroma Features: Chroma features represent the 12 different pitch classes of the musical octave, capturing harmonic and tonal content.
* Spectral Features: These include the spectral centroid (the center of mass of the spectrum), spectral bandwidth (the width of the spectrum), spectral roll-off (the frequency below which a certain percentage of the total spectral energy lies), and spectral flatness (a measure of how noise-like a signal is).
* Temporal Features: Temporal features such as the zero-crossing rate (the rate at which the signal changes sign) and root mean square (RMS) energy (a measure of the signal’s power) describe the signal's temporal dynamics.

## Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a type of deep learning model particularly well-suited for grid-like data, such as images and spectrograms. CNNs consist of several layers, including convolutional layers that apply filters to the input data, pooling layers that reduce dimensionality, and fully connected layers that perform classification.

CNNs are advantageous for musical genre classification because they can automatically learn complex feature representations from spectrograms, which are visual representations of the spectrum of frequencies in a sound signal as they vary with time. The ability to learn hierarchical features enables CNNs to capture both low-level features (such as edges and textures) and high-level features (such as shapes and objects) from the input data.

## Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are supervised learning models used for classification and regression tasks. SVMs work by finding the optimal hyperplane that separates data points of different classes with the maximum margin. The kernel trick allows SVMs to operate in a high-dimensional space by implicitly mapping the input features into higher-dimensional feature spaces, enabling the separation of data that is not linearly separable in the original feature space.

SVMs are effective in high-dimensional spaces and are robust to overfitting, especially when the number of features exceeds the number of samples. This makes them suitable for audio classification tasks, where the feature space can be large and complex.

## Random Forests (RFs)

Random Forests (RFs) are an ensemble learning method that combines the predictions of multiple decision trees to improve classification performance. Each decision tree is trained on a bootstrap sample of the training data, and the final prediction is obtained by aggregating the predictions of all individual trees (e.g., by majority voting).

RFs are advantageous for musical genre classification because they provide insights into feature importance, are robust to noisy data and outliers, and can model complex interactions between features. The ensemble approach reduces the risk of overfitting and enhances the model's generalization ability.

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

## Units

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
* Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.

Identify applicable funding agency here. If none, delete this text box.

* Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.