

DOI:10.22144/ctu.jsi.2017.020

## PHÂN LOẠI NHẠC VIỆT NAM THEO THỂ LOẠI DỰA TRÊN ÂM SẮC VÀ NHỊP ĐIỆU

Phan Anh Cang<sup>1</sup>, Nguyễn Thị Kim Khánh<sup>2</sup> và Phan Thượng Cang<sup>3</sup>

<sup>1</sup>Khoa Công nghệ Thông tin, Trường Đại học Sư phạm Kỹ thuật Vĩnh Long

<sup>2</sup>Tổ Tin, Trường Trung học Phổ thông Chuyên Nguyễn Bình Khiêm, Vĩnh Long

<sup>3</sup>Khoa Công nghệ Thông tin và Truyền thông, Trường Đại học Cần Thơ

### Thông tin chung:

Ngày nhận bài: 15/09/2017

Ngày nhận bài sửa: 10/10/2017

Ngày duyệt đăng: 20/10/2017

### Title:

Vietnamese music classification by genre based on timbral texture and rhythmic content

### Từ khóa:

Nhạc Việt, phân loại nhạc, rút trích đặc trưng tín hiệu audio, tín hiệu âm nhạc, wavelet rời rạc

### Keywords:

Classification, digital music in Vietnam, extracting specific timbral disposition, wavelet transform

### ABSTRACT

These days, digital music storage systems (DMSS) in Vietnam usually arrange pieces of music according to the composer's name and the song's title, whereas listeners need to search for songs based on genres and contents. This increases the demand for categorizing songs in accordance with genres in DMSS, which enables listeners to search for the most wanted music. However, with a large number of songs collected, the way to classify them for easy management becomes a challenge for all DMSS. Therefore, it is necessary to build up an automatic sorting system. This paper suggests a new method of extracting specific timbral disposition including timbral texture, rhythmic content by using wavelet convert. Thanks to such distinctive features, KNN and SVM methods are utilized to identify types of music files. This study is conducted on four types of music: Bolero, Cai luong (reformed theatre), Cheo (classical theatre) and Hat Boi (traditional opera). The findings show that the reliability is up to 93.75% and 94% corresponding to KNN and SVM on the timbral texture. Moreover, these suggested methods are simple, effective, speedy, and suitable for Vietnamese music sorting systems today.

### TÓM TẮT

Hiện nay, các hệ thống lưu trữ nhạc số Việt nam thường sắp xếp các bản nhạc theo tên nhạc sĩ hoặc theo tên bài hát trong khi người nghe nhạc cũng cần tìm kiếm các bản nhạc theo thể loại và nội dung. Điều này đã nảy sinh nhu cầu phân loại nhạc theo thể loại trong các hệ thống lưu trữ nhạc số để cho phép người nghe nhạc có thể tìm kiếm bản nhạc theo yêu cầu. Tuy nhiên, với số lượng lớn nhạc số sưu tập được, việc phân loại chúng để dễ dàng quản lý trở thành một thách thức đối với các hệ thống lưu trữ nhạc số. Điều này cho thấy việc xây dựng một hệ thống phân loại nhạc tự động là rất cần thiết. Trong luận văn này, chúng tôi đề xuất phương pháp rút trích tập đặc trưng bộ cục âm sắc của tín hiệu audio bao gồm kết cấu âm sắc (timbral texture) và nhịp điệu (rhythmic content) sử dụng phép biến đổi wavelet rời rạc. Dựa trên tập đặc trưng này, phương pháp KNN và SVM được sử dụng để nhận dạng thể loại của các tập tin nhạc. Nghiên cứu của chúng tôi thực hiện minh họa trên bốn thể loại Bolero, Cải lương, Chèo và Hát bội. Kết quả thực nghiệm cho thấy độ chính xác là 93,75 % và 94 % đối với phương pháp phân loại KNN và SVM tương ứng trên tập đặc trưng về bộ cục âm sắc. Hơn nữa, phương pháp đề xuất này đơn giản, hiệu quả và có thời gian thực hiện nhanh phù hợp cho các hệ thống phân loại nhạc Việt hiện nay.

Trích dẫn: Phan Anh Cang, Nguyễn Thị Kim Khánh và Phan Thượng Cang, 2017. Phân loại nhạc Việt Nam theo thể loại dựa trên âm sắc và nhịp điệu. Tạp chí Khoa học Trường Đại học Cần Thơ. Số chuyên đề: Công nghệ thông tin: 145-154.

## 1 GIỚI THIỆU

Trong những năm gần đây, cùng với sự phát triển của công nghệ thông tin, số lượng bản nhạc dưới hình thức dữ liệu audio trong các kho dữ liệu lớn, trên Internet, đang ngày càng gia tăng nhanh chóng. Thông thường ở Việt Nam, muốn biết một bài nhạc thuộc thể loại nhạc nào, chúng ta thường dựa vào kinh nghiệm của người nghe. Tuy nhiên, để biết được điều đó, chúng ta thường tìm tên bài hát, tác giả, ca sĩ hát bài hát để xác định xem bài hát đó thuộc thể loại nhạc nào. Ví như khi ta nghe một bài hát của một tác giả chuyên sáng tác nhạc thuộc thể loại dân ca, ta qui bài hát đó thuộc thể loại nhạc dân ca. Hoặc là, khi ta nghe một bài hát do một nghệ sĩ chuyên hát nhạc kịch - cải lương, thế là ta qui bài hát thuộc thể loại cải lương. Ngoài ra, việc phân loại nhạc có thể dựa vào các loại nhạc cụ được sử dụng trong bản nhạc. Tuy nhiên, với sự phát triển của công nghệ, số lượng bài hát càng nhiều, chúng ta không thể nghe từng bản nhạc để kết luận nó thuộc thể loại nào. Việc nghiên cứu một hệ thống phân loại tự động nhạc Việt Nam để ứng dụng trong các hệ thống sưu tập nhạc Việt theo thể loại (chẳng hạn như Website nhạc Việt online) nhằm giới thiệu nhạc Việt cho các bạn yêu thích âm nhạc trên thế giới.

Trong kho tàng văn hóa nghệ thuật, âm nhạc Việt Nam thực sự đa dạng và phong phú (Phạm Thị Hòa và Ngô Thị Nam, 2006; Phạm Thị Hòa, 2007). Nó là một trong các bộ môn nghệ thuật giáo dục cái đẹp, giáo dục tình cảm thẩm mỹ, làm phong phú thêm đời sống tinh thần, cảm thụ cái đẹp, tạo niềm tin,... cho con người. Bên cạnh đó, âm nhạc đã trở thành một loại hình sinh hoạt văn hóa nghệ thuật quen thuộc của người dân Việt Nam. Bởi nó nuôi dưỡng đời sống tinh thần dân tộc bằng cái chất trữ tình đậm thắm sâu sắc. Nó là sự kết hợp của hàng loạt yếu tố: hát, múa, nhạc, kịch,... mang tính đậm đà bản sắc dân tộc.

Nhiều nghiên cứu đã đưa ra các ý tưởng phát triển một hệ thống mà nó có thể truy tìm thông tin nhạc trên Internet một cách tự động sao cho các bản nhạc tìm thấy tương tự với bộ sưu tập của người sử dụng, tìm các bản nhạc có các đặc trưng gần giống với các đặc trưng mà người sử dụng mong muốn, chú thích tự động các tập tin nhạc với những mô tả về thể loại,... Chẳng hạn, Y.M.D. Chathuranga và K.L. Jayaratne Musical (2013) đã xây dựng hệ thống phân loại nhạc theo thể loại và tập trung vào việc phân tích thông tin từ các tín hiệu audio. Nhóm nghiên cứu này trình bày cách tiếp cận trích lọc các đặc trưng từ tín hiệu audio và máy học phục vụ cho việc phân loại tự động thể loại nhạc, trong đó phương pháp phân loại Support Vector Machine (SVM) (Đỗ Thanh Nghi, 2008) (Tao *et al.* 2010) với

hàm nhân đa thức được sử dụng. Hệ thống này có thể nhận được một bộ sưu tập nhạc theo thể loại bằng cách tra các bản nhạc vào hệ thống phân loại nhạc tự động. Các tín hiệu audio được phân loại một cách tự động thuộc 1 trong 10 thể loại nhạc phổ biến trên thế giới với bộ dữ liệu GTZAN và ISMIR2004 (Tzanetakis *et al.*, 2001; Tzanetakis and Cook, 2002). Kết quả cho thấy việc sử dụng phương pháp SVM với độ chính xác đã thu được 81%. Bên cạnh đó, Rini Wongso and Diaz D. Santika (2014) nghiên cứu kết hợp tính năng nâng Tree Complex Wavelet Transform (TCWT) và SVM. Nghiên cứu này tập trung vào việc phân loại bốn thể loại nhạc: Pop, Classical, Jazz và Rock bằng cách sử dụng các chỉ số thống kê về trung bình, độ lệch chuẩn, phương sai, và entropy của các đặc trưng tín hiệu nhạc. Dữ liệu được sử dụng trong nghiên cứu được lấy từ tập GTZAN. Dựa trên các kết quả thử nghiệm, phương pháp này đạt được độ chính xác 88,33%. Ngoài ra, nhiều nghiên cứu khác dựa trên tập các đặc trưng liên quan nhịp điệu, âm sắc, độ cao thấp nốt nhạc,... nhằm nâng cao tỷ lệ nhận dạng, phân loại.

Trong bài báo này, chúng tôi tập trung nghiên cứu xây dựng hệ thống phân loại nhạc tự động để phân loại các bản nhạc thuộc một trong các thể loại: nhạc Bolero Việt, Cải lương, Hát bội và Chèo dựa trên các đặc trưng về bố cục âm sắc. Vì khi đưa các tín hiệu âm sắc vào dao động điện từ ta sẽ được trên màn hình những đường cong liên tục có cùng tần số nhưng có dạng khác nhau (Anan *et al.* 2011). Ngoài ra, các loại nhạc Cải lương, Hát bội, Chèo là các loại nhạc đặc trưng cho ba vùng miền (Cải lương- miền Nam, Hát bội- miền Trung, Chèo- miền Bắc) và nhạc Bolero Việt là dòng nhạc đang được nhiều tầng lớp dân chúng ở các vùng miền yêu thích. Chúng tôi sử dụng phép biến đổi wavelet rời rạc (DWT) để phân tích tín hiệu audio dùng cho việc xác định các đặc trưng về nhịp điệu. Nó có thể áp dụng mở rộng đối với các thể loại nhạc khác hoặc xây dựng các hệ thống truy vấn thông tin nhạc dựa vào nội dung, kiểm tra việc sao chép bản quyền nhạc,... Chúng tôi cũng trình bày việc lựa chọn các đặc trưng phù hợp vì chúng ảnh hưởng đáng kể đến độ chính xác phân loại.

## 2 CÁC NGHIÊN CỨU LIÊN QUAN

### 2.1 Đặc trưng nhạc Việt

Mỗi thể loại nhạc là một tập các đặc trưng chung mà người nghe có thể phân biệt nó với những loại nhạc khác từ những bản nhạc khác nhau. Những đặc trưng có thể kể như: độ cao thấp nốt nhạc (pitch), âm sắc (timbres), nhịp điệu (rhythm) của bản nhạc hoặc những đặc trưng liên quan đến bố cục nhạc (music texture). Một trong những thách thức trong phân loại thể loại nhạc tự động là tìm ra các đặc

trung đó. Như chúng ta đã biết, thính giác của ta phân biệt được những âm thanh có tính nhạc và âm thanh có tính chất tiếng động (Phạm Thị Hòa và Ngô Thị Nam, 2006) như tiếng sấm, tiếng gõ, tiếng nhạc cụ... Âm sắc (timbres) là một thuộc tính của âm thanh. Mỗi nhạc cụ hoặc mỗi giọng hát đều có âm sắc riêng. Sự khác biệt của âm sắc phụ thuộc vào thành phần của các âm thanh như giọng hát và nhạc cụ (nhạc cụ dây, dụng cụ gió, các nhạc cụ gõ...).

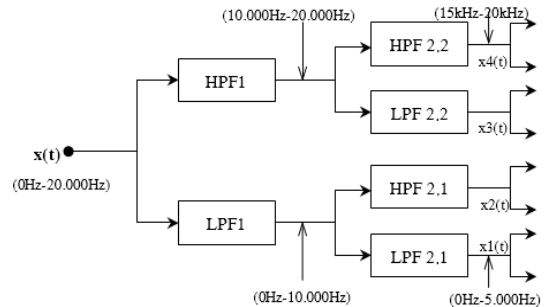
Trong bài báo này, do giới hạn thời gian, chúng tôi tập trung nghiên cứu xây dựng hệ thống phân loại nhạc tự động để phân loại các bản nhạc thuộc một trong các thể loại: nhạc Bolero Việt, Cải lương, Hát bội và Chèo dựa trên các đặc trưng về bố cục âm sắc và nhịp nhạc. Vì khi đưa các tín hiệu âm sắc và nhịp nhạc vào dao động điện tử ta sẽ được trên màn hình những đường cong liên tục có cùng tần số nhưng có dạng khác nhau. Việc phân loại các bản nhạc không đơn thuần chỉ xác định thuộc một trong các thể loại nhạc của thể giới như: Rock, Classical, Jazz,... mà các bản nhạc có thể có sự pha trộn của nhiều thể loại thậm chí có những thể loại nhạc Việt không thuộc các thể loại trên (như cải lương, chèo, dân ca,...). Đây thật sự là thách thức đối với các hệ thống phân loại nhạc Việt. Ngoài ra, vấn đề đặt ra đối với chúng ta là: cần tìm ra tập các đặc trưng về âm nhạc, đặc biệt là đối với nhạc Việt Nam, từ đó đưa ra các thuật toán rút trích các đặc trưng từ tín hiệu audio phục vụ cho việc phân loại nhạc. Để giải quyết bài toán này, việc nghiên cứu phân loại tự động nhạc Việt Nam theo thể loại là hết sức cần thiết và đáp ứng nhu cầu thực tiễn.

**2.2 Phép biến đổi wavelet rời rạc**

Phép biến đổi Fourier thường dùng cho phân tích các tín hiệu audio. Tuy nhiên, nó có hạn chế là ta không thể biết được tại một thời điểm sẽ xuất hiện những thành phần tần số nào. Để khắc phục nhược điểm này, các nhà khoa học sử dụng biến đổi STFT (Short time Fourier transform). Theo đó, tín hiệu được chia thành các khoảng nhỏ và được biến đổi Fourier trong từng khoảng đó. Phương pháp này có hạn chế là việc chọn độ rộng của các khoảng tín hiệu phân chia sao cho phù hợp vì nếu độ rộng này càng nhỏ thì độ phân giải thời gian càng tốt nhưng phân giải tần số càng kém và ngược lại. Để khắc phục cả 2 phương pháp trên, biến đổi wavelet ra đời. Biến đổi wavelet (WT) được thực hiện như sau: tín hiệu được nhân với hàm Wavelet (trung tự như nhân với hàm cửa sổ trong biến đổi STFT), sau đó thực hiện phân tích riêng rẽ cho các khoảng tín hiệu khác nhau trong miền thời gian tại các tần số khác nhau.

Phép biến đổi wavelet rời rạc (DWT) là một trường hợp đặc biệt của WT. Biến đổi Wavelet đưa ra giải pháp linh hoạt như sau: thành phần tín hiệu

tần số cao sẽ phân giải tốt hơn trong miền thời gian, còn thành phần tín hiệu tần số thấp, sẽ phân giải tốt hơn ở miền tần số. Nó cung cấp một cách biểu diễn tín hiệu dưới dạng nén trong miền thời gian-tần số giúp cho việc tính toán một cách nhanh chóng và hiệu quả. DWT thực hiện phân tích đa phân giải một tín hiệu audio  $x$  thành 2 thành phần: thành phần tín hiệu thô A (coarse approximation) tương ứng với thành phần tần số thấp  $y_{low}$  và thành phần tín hiệu chi tiết D (detail) tương ứng với thành phần tần số cao  $y_{high}$  (Tzanetakis *et al.*, 2001). Sau đó, thành phần tín hiệu thô tiếp tục được phân tích tương tự. Như vậy, một tín hiệu có thể sẽ chứa các tần số và tín hiệu khác sẽ chứa các tần số trong phạm vi lớn hơn. Tiếp tục đưa hai tín hiệu này qua bộ lọc Lowpass và Highpass sẽ cho kết quả là 4 tín hiệu thành phần. Tiếp tục cách làm này, cuối cùng chúng ta được một số tín hiệu thành phần mà mỗi tín hiệu chứa 1 vùng tần số xác định. được biểu diễn dưới dạng tổng của thành phần tín hiệu thô và các thành phần tín hiệu chi tiết. Quá trình phân tích này được thực hiện bởi các bộ lọc băng tần cao và thấp đối với tín hiệu  $x$  như biểu diễn trong Hình 1.



**Hình 1: Tín hiệu x(t) được đưa qua các bộ lọc Lowpass và Highpass**

Việc tính toán các hệ số Wavelet của tín hiệu audio là một công việc hết sức phức tạp. Để giảm thiểu công việc tính toán người ta chỉ chọn ra một tập nhỏ các giá trị và các vị trí để tiến hành tính toán cụ thể như DWT chia tín hiệu thành hai thành phần: thành phần xấp xỉ (tần số thấp) và thành phần chi tiết (tần số cao) trong ngưỡng nghe được của con người. Công việc này là hoàn toàn có thể thực hiện được nhờ phép biến đổi Wavelet rời rạc (Discrete wavelet transform - DWT). Do đó, việc tính toán DWT thực chất là sự rời rạc hóa biến đổi Wavelet liên tục của tín hiệu.

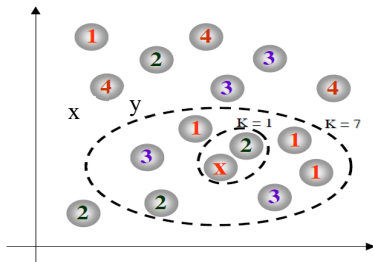
Trong giai đoạn huấn luyện, chúng ta sẽ rút trích đặc trưng từ các tập tin nhạc trong bộ sưu tập nhạc mẫu đã được biết trước thể loại và lưu trữ các đặc trưng đó để huấn luyện nhằm giảm chi phí tính toán cho việc rút trích đặc trưng các tập tin nhạc mẫu mỗi khi nhận dạng.

**2.3 Phương pháp phân loại KNN**

Có nhiều phương pháp phân lớp như: KNN, SVM, Bayes, HMMs, Gaussian,... Trong nghiên cứu này, chúng tôi sử dụng phương pháp K-NN và SVM vì nó đơn giản và được sử dụng phổ biến trong các bài toán phân lớp. Phương pháp KNN cho phép bổ sung mẫu huấn luyện mới vào bộ huấn luyện dễ dàng và hiệu quả khi tập huấn luyện lớn. Bên cạnh đó, bộ huấn luyện được huấn luyện từ chính các vector đặc trưng rút trích từ tín hiệu audio. Nó xử lý tốt với tập dữ liệu nhiễu do dựa trên khoảng cách giữa các vector đặc trưng để quyết định phân lớp, do đó nó phù hợp với hệ thống phân loại nhạc.

Phương pháp K-NN xem các mẫu (vector đặc trưng) như là các điểm biểu diễn trong không gian đặc trưng n chiều (Hình 2). Khoảng cách giữa mẫu cần phân loại x và k mẫu láng giềng y là d(x, y) được xác định dựa trên khoảng cách không gian. Thông thường, người ta dùng khoảng cách Euclide để xác định khoảng cách giữa các mẫu trong không gian đặc trưng được xác định bởi công thức (1).

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{1}$$



**Hình 2: Mô hình phân lớp K-NN**

Xác suất mẫu x thuộc vào thể loại ci được xác định bởi công thức (2):

$$p(c_i | x) = \frac{\sum_{y \in K, y_c=c_i} w_y}{\sum_{y \in K} w_y} \tag{2}$$

Trong đó:  $w_y = (1/d(x,y))$ ; K là một tập hợp k mẫu láng giềng gần x nhất;  $y_c$  là thể loại của y;  $c_i$  là thể loại thứ i.

**2.4 Phương pháp phân loại SVM**

SVM (Support Vector Machine) (Đỗ Thanh Nghị, 2008; Tao et al. 2010) là một khái niệm trong thống kê và khoa học máy tính cho một tập hợp các phương pháp học có giám sát liên quan đến nhau để phân loại và phân tích hồi quy. SVM là một thuật toán phân loại nhị phân, SVM nhận dữ liệu vào và phân loại chúng vào hai lớp khác nhau. Với một bộ các ví dụ huấn luyện thuộc hai thể loại cho

trước, thuật toán huấn luyện SVM xây dựng một mô hình SVM để phân loại các ví dụ khác vào hai thể loại đó. Việc sử dụng phương pháp máy tựa vector SVM trong việc phân loại dữ liệu hiện đang được áp dụng trong rất nhiều lĩnh vực.

Phương pháp tựa vector ánh xạ các vector đầu vào x sang không gian đặc trưng có số chiều cao hoặc vô hạn chiều ( $z = \phi(x)$ ) sau đó xây dựng một siêu phẳng tối ưu  $w.z + b = 0$  để phân loại dữ liệu thành hai lớp. Trong đó,  $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$  là hàm hạt nhân (kernel function) thực hiện ánh xạ phi tuyến.

Một số hàm hạt nhân thường được sử dụng là:

- Gaussian kernel:

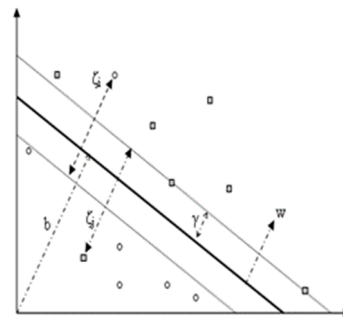
$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{3}$$

- Polynomial kernel:

$$k(x_i, x_j) = (1 + x_i \cdot x_j)^d \tag{4}$$

- RBF kernel:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{5}$$



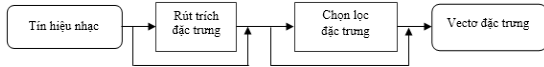
**Hình 3: Phương pháp phân loại SVM**

Chúng tôi sử dụng kết quả phân loại khi sử dụng SVM với hàm nhân RBF làm giá trị của hàm mục tiêu. Khi sử dụng SVM với hàm nhân RBF có hai tham số cần được thiết lập trước đó là tham số C và tham số  $\gamma$ . Chúng tôi sử dụng phương pháp thực nghiệm để xác định các tham số C và  $\gamma$  tối ưu cho từng tập dữ liệu đầu vào.

**3 XÂY DỰNG HỆ THỐNG PHÂN LOẠI NHẠC THEO THỂ LOẠI**

Trên thực tế, tất cả các đặc trưng của tín hiệu audio khi đưa trực tiếp vào các mô hình phân loại sẽ làm giảm đi rõ rệt tốc độ huấn luyện và phân loại. Rút trích đặc trưng là một trong những kỹ thuật tiên xử lý tín hiệu nhạc được sử dụng phổ biến trong việc phân loại. Quá trình rút trích sẽ khử nhiễu tín hiệu

và chỉ chọn các thông tin cần thiết cho việc phân loại nhạc. Ngoài ra, việc chọn lọc đặc trưng được dùng để tạo ra một tập con đặc trưng từ dữ liệu đầu vào nhằm làm tăng hiệu quả về mặt thời gian trong việc nhận dạng vì nó là tiến trình tự động hoá được dùng để giảm số chiều dữ liệu sao cho dữ liệu đầu vào được chuyển đổi sang dạng đơn giản và nhỏ hơn trước khi đưa vào mô hình phân loại.



**Hình 4: Sơ đồ rút trích đặc trưng từ một tín hiệu nhạc**

Nhiều nghiên cứu đã đề xuất các đặc trưng của tín hiệu audio để nhận dạng, phân loại trong các hệ thống nhận dạng, phân loại khác nhau. Mỗi nghiên cứu đều đưa ra một số các đặc trưng của tín hiệu audio và phương thức sử dụng để phân loại. Các đặc trưng của tín hiệu audio thường được chia làm hai nhóm chính: các đặc trưng trong miền thời gian – tần số và các đặc trưng cảm thụ âm thanh của con người (nhịp điệu, cao độ) (Wongso and Santika, 2014). Trong bài báo này, chúng tôi xây dựng hệ thống phân loại nhạc dựa trên hai tập đặc trưng như sau:

- Các đặc trưng về âm sắc (Timbral Texture Features).
- Các đặc trưng về nhịp điệu (Rhythmic Content Features).

**3.1 Đặc trưng về âm sắc**

Tập đặc trưng về âm sắc được sử dụng để biểu diễn các đặc trưng của âm nhạc liên quan đến tiết tấu, âm sắc và nhạc cụ. Vector đặc trưng về âm sắc được sử dụng trong hệ thống phân loại của chúng tôi bao gồm 19 chiều với các đặc trưng: (Trung bình và độ lệch chuẩn của Spectral Centroid, Rolloff, Flux, ZeroCrossing, LowEnergy, và Trung bình và độ lệch chuẩn của 5 hệ số MFCC đầu tiên). Trung bình và độ lệch chuẩn của các đặc trưng này được xác định dựa trên STFT với các cửa sổ phân tích chia tín hiệu đầu vào có độ dài 1s thành các đoạn nhỏ khoảng 20ms. Sau đây là các đặc trưng được xác định trên mỗi cửa sổ phân tích:

*a. Đặc trưng 1: Spectral Centroid*

Spectral Centroid là một độ đo liên quan hình dáng của phổ tần số. Nó xác định điểm cân bằng của phổ tần số. Giá trị Centroid cao tương ứng với phổ có độ sáng chói hơn và chứa nhiều tần số cao. Spectral Centroid được xác định bởi công thức (6):

$$C_t = \frac{\sum_{n=1}^N M_t[n] * n}{\sum_{n=1}^N M_t[n]} \tag{6}$$

Trong đó:  $M_t[n]$  là biên độ của tần số thứ  $n$  trong phổ tần số tương ứng với cửa sổ  $t$ .

*b. Đặc trưng 2: Rolloff*

Rolloff cũng là một độ đo liên quan hình dáng của phổ tần số. Điểm Rolloff của phổ tần số ( $R_t$ ) được định nghĩa như tần số biên mà ở đó 85% phân bố năng lượng được tập trung trong phổ là dưới điểm này. Công thức (7) xác định  $R_t$  - điểm Rolloff của phổ tần số.

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 \sum_{n=1}^N M_t[n] \tag{7}$$

*c. Đặc trưng 3: Flux*

Flux được xem là độ biến thiên phổ, cho biết sự thay đổi về biên độ tần số của phân phối quang phổ giữa hai cửa sổ phân tích liên tiếp. Nó được xác định là bình phương hiệu giữa các biên độ chuẩn của tần số trong phổ và được xác định bởi công thức (8).

$$F_t = \sum_{n=1}^N (N_t[n] - N_{t-1}[n])^2$$

$$N_t[n] = \frac{M_t[n]}{\sqrt{\sum_{i=1}^N (M_t[i])^2}} \tag{8}$$

Với  $N_t[n]$  và  $N_{t-1}[n]$  là biên độ chuẩn của tần số thứ  $n$  trong phổ tần số ở cửa sổ  $t$  và  $t-1$  tương ứng.

*d. Đặc trưng 4: Zero-crossings*

Zero Crossings cho biết mức độ ồn (noisiness) của âm thanh trong tín hiệu. Nó xuất hiện khi các mẫu kề nhau trong tín hiệu khác dấu. Nó được xác định bởi số lần tín hiệu audio vượt qua trục zero trên một đơn vị thời gian và được tính bởi công thức :

$$Z_t = \frac{1}{2} \sum_{n=1}^N |sign(x[n]) - sign(x[n-1])| \quad \text{và}$$

$$sign(x[n]) = \begin{cases} 1 & x[n] > 0 \\ 0 & x[n] \leq 0 \end{cases} \tag{9}$$

$x[n]$  là tín hiệu trong miền thời gian đối với cửa sổ  $t$ .

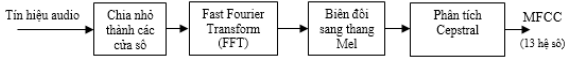
e. Đặc trưng 5: Low-Energy

Khác với các đặc trưng trên, đặc trưng Low-Energy được xác định trên toàn bộ tín hiệu miền thời gian. Nó là tỉ lệ phần trăm của các cửa sổ phân tích có RMS (Root-Mean-Square) năng lượng thấp hơn RMS trung bình năng lượng của các tín hiệu trong các cửa sổ phân tích. Trong đó, RMS năng lượng của tín hiệu ở cửa sổ t được xác định bởi công thức (10):

$$RMS_t = \sqrt{\frac{\sum_{i=1}^N (M_t[i]^2)}{N}} \tag{10}$$

f. Đặc trưng 6: Các hệ số MFCC (Mel-Frequency Cepstral Coefficients)

MFCC là một trong các tập đặc trưng được dùng phổ biến trong các hệ thống nhận dạng giọng nói, truy tìm thông tin nhạc,... Nó cung cấp cách biểu diễn nén tín hiệu audio dưới dạng phổ sao cho hầu hết năng lượng của tín hiệu được tập trung vào các hệ số đầu tiên. Hình 4 mô tả các bước thực hiện rút trích đặc trưng MFCC từ tín hiệu audio. Chi tiết về phương pháp rút trích đặc trưng MFCC (Logan and Beth, 2000) mô tả trong Hình 5.



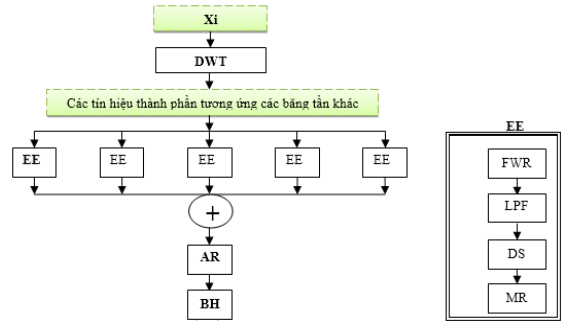
Hình 5: Sơ đồ rút trích đặc trưng MFCC

Kết quả thu được là một tập đặc trưng MFCC gồm 13 hệ số. Tuy nhiên, nhiều nghiên cứu (Li et al., 2003) cho thấy 5 hệ số MFCC đầu tiên cung cấp khá đầy đủ thông tin cho việc phân loại nhạc theo thể loại. Vì vậy, để giảm số chiều cho vector đặc trưng, chúng tôi chọn 5 hệ số MFCC đầu tiên cho hệ thống phân loại nhạc theo thể loại của chúng tôi.

3.2 Đặc trưng về nhịp điệu nhạc

Vector đặc trưng về nhịp điệu cung cấp rất nhiều thông tin có ích về đặc điểm của các thể loại nhạc. Hầu hết các hệ thống dò tìm nhịp điệu nhạc cung cấp các thuật toán xác định nhịp điệu của bản nhạc và cường độ của chúng. Bên cạnh đó, chúng còn cho biết mối liên hệ giữa các nhịp của bản nhạc. Trong bài báo này, chúng tôi sử dụng phương pháp xác định tập đặc trưng về nhịp điệu nhạc được đề xuất bởi George Tzanetakis (Tzanetakis et al., 2001) trong việc phân loại nhạc theo thể loại. Phương pháp này dựa trên việc dò tìm các chu kỳ (đơn vị: bpm - số nhịp/phút) có biên độ lớn nhất của tín hiệu. Tín hiệu audio X được chia nhỏ thành các tín hiệu thành phần Xi bởi cửa sổ phân tích có kích thước 65536 mẫu với tần số lấy mẫu (sampling rate) là 22050 Hz tương ứng xấp xỉ 3s. Sau đó, thuật toán xác định nhịp điệu nhạc được áp dụng đối với mỗi Xi như biểu diễn trong Hình 5.

Quá trình xác định nhịp điệu nhạc trên tín hiệu audio được áp dụng lặp đi lặp lại trên các tín hiệu thành phần Xi và tích lũy vào trong biểu đồ nhịp điệu BH. Tập các đỉnh cao nhất của hàm tự tương quan tạo nên biểu đồ nhịp điệu nhạc được sử dụng làm cơ sở cho việc xác định các đặc trưng về nhịp điệu. Trong đó, các đỉnh cao nhất trong BH tương ứng với các chu kỳ khác nhau của tín hiệu audio là các nhịp chính của bản nhạc.

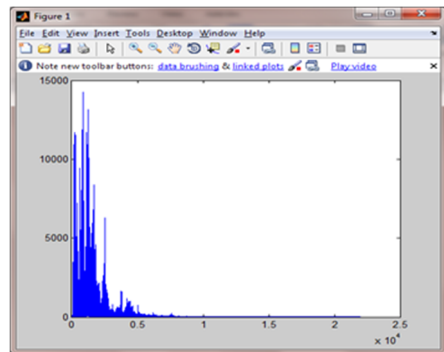


Hình 6: Sơ đồ khối xác định Histogram nhịp điệu nhạc

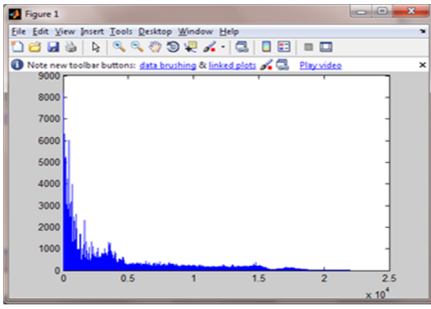
Xác định các đặc trưng về nhịp điệu:

Vector đặc trưng về nhịp điệu là một vector 6 chiều gồm các đặc trưng:

- A1, A2: Đặc trưng này là độ đo sự khác nhau về nhịp so với các nhịp còn lại của tín hiệu. Nó được xác định bởi tỉ số giữa biên độ của lần lượt 2 đỉnh Đ1 và Đ2 với tổng biên độ của tất cả các đỉnh trong BH.
- RA: là tỷ số giữa biên độ của đỉnh Đ2 với biên độ của đỉnh Đ1. Đặc trưng này biểu diễn mối quan hệ giữa nhịp chính và nhịp phụ đầu tiên.
- P1, P2: Chu kỳ của đỉnh Đ1 và Đ2 được tính bằng số nhịp trong 1 phút (đơn vị tính: bpm).
- SUM: Tổng biên độ của các đỉnh trong BH. Đặc trưng này cho biết độ mạnh của nhịp nhạc.



Hình 7: Quang phổ về âm thanh của 1 bản Chèo



Hình 8: Quang phổ về âm thanh của 1 bản Cải lương

4 KẾT QUẢ THỰC NGHIỆM

4.1 Tập dữ liệu dùng cho huấn luyện và kiểm tra

Trong nghiên cứu này, nghiên cứu thử nghiệm trên dữ liệu được tải từ các trang web có chứa nhạc Việt Nam <http://youtube.com>, trang nhạc <http://nhacvui.com>.

Dữ liệu tải về được lưu lại dưới dạng file \*.mp3, \*.mp4, thuộc 4 loại nhạc đặc trưng truyền thống của Việt Nam (mỗi loại có trên 300 file), tạo thành tập dữ liệu trên 1200 file nhạc, dữ liệu được chuyển về dạng file \*.wav bằng phần mềm chuyển đổi Total Video Converter và được lưu vào các thư mục tương ứng: Bolero, Cailuong, Cheo, Hatboi. Với tên thư mục là tên của loại nhạc đã được xác định trước. Do các file có thời lượng từ 5 phút đến 45 phút (trích đoạn cải lương, hát bội) nên để thống nhất dữ liệu được cắt thành file có thời lượng 15 giây và 30 giây để làm tập huấn luyện và nhận dạng. Tên các thư mục, file nhạc gắn liền với tên file gốc trước đó được lưu trữ phục vụ đánh giá, lựa chọn mô hình phù hợp nhất.

Tập dữ liệu các file nhạc thuộc 4 chủ đề khác nhau như sau:

1. Bolero: các file là các bài hát theo dòng nhạc bolero...
2. Cailuong: các file là các bài ca cổ, tân cổ, trích đoạn cải lương, vọng cổ hay một đoạn nhạc đờn ca tài tử ...
3. Cheo: các file là các bài hát dòng nhạc chèo chủ yếu của Đoàn hát Chèo Thái Bình...
4. Hatboi: các file là các bài trích đoạn hát bội của Nhà hát Thành phố Hồ Chí Minh ...

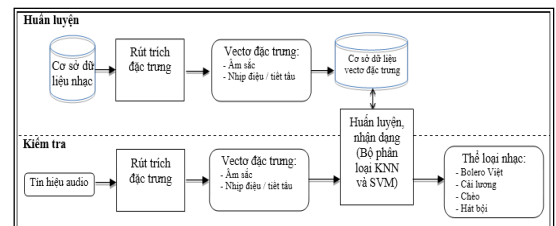
Trong phương pháp của chúng tôi, nguồn dữ liệu được chia thành 2 tập dữ liệu: huấn luyện và kiểm tra. Tập dữ liệu huấn luyện được sử dụng để huấn luyện cho bộ phân loại KNN để đưa ra các quyết định cho hệ thống phân loại nhạc theo thể loại trong khi tập dữ liệu kiểm tra sẽ được sử dụng để đánh giá hiệu quả của phương pháp đề xuất. Số tập tin audio sử dụng trong tập huấn luyện và kiểm tra tương ứng từng thể loại được trình bày trong Bảng 1.

Bảng 1: Số lượng tập tin audio dùng cho huấn luyện và kiểm tra

STT	Tên thể loại	Số lượng tập tin audio		
		Huấn luyện		Kiểm tra
		15 giây	30 giây	
1	Bolero	100	100	100
2	Cải lương	100	100	100
3	Chèo	100	100	100
4	Hát bội	100	100	100
Tổng		400	400	400

4.2 Mô hình tổng quát hệ thống phân loại nhạc theo thể loại

Chúng tôi đề xuất hệ thống phân loại nhạc theo thể loại gồm 2 pha: rút trích đặc trưng và huấn luyện hoặc phân loại. Kết quả sau khi rút trích đặc trưng của tín hiệu audio là một tập gồm các đặc trưng về âm sắc, nhịp điệu. Chi tiết việc rút trích đặc trưng được trình bày trong phần III. Chúng tôi sử dụng phương pháp biến đổi wavelet rời rạc (DWT) để rút trích đặc trưng về nhịp điệu. Phương pháp phân loại KNN và SVM được sử dụng để nhận dạng các thể loại nhạc. Quá trình huấn luyện bao gồm việc sử dụng các vector đặc trưng đã được gán nhãn thể loại để huấn luyện cho bộ phân loại KNN. Từ đó, bộ phân loại sẽ gán nhãn thể loại cho các vector đặc trưng mới một cách tự động. Mô hình tổng quát hệ thống phân loại nhạc theo thể loại được minh họa trong Hình 6.



Hình 9: Mô hình tổng quát hệ thống phân loại nhạc theo thể loại

Tập các đặc trưng sử dụng cho hệ thống phân loại nhạc trong nghiên cứu này bao gồm các đặc trưng sau đây:

**Các đặc trưng về âm sắc:** Gồm 19 đặc trưng: Trung bình và phương sai của Centroid, Rolloff, Flux, ZeroCrossing (8), LowEnergy (1); Trung bình và phương sai của 5 hệ số MFC đầu tiên (10).

**Các đặc trưng về nhịp điệu / tiết tấu:** Gồm 6 đặc trưng: A1, A2, RA, P1, P2, SUM được xác định từ biểu đồ nhịp điệu.

4.3 Ma trận đánh giá độ chính xác phân loại

Việc đánh giá phương pháp đề xuất được thực hiện bởi các file audio trong tập dữ liệu kiểm tra. Kết

quả phân loại của hệ thống sẽ được trình bày trong ma trận đánh giá độ chính xác phân loại như Bảng 2.

Trong ma trận này, các giá trị trong ma trận là số lượng tập tin audio trong tập dữ liệu kiểm tra. Các phần tử trong ma trận được giải thích như sau:

B, Ca, C, H: số tiên đoán đúng đối với các file nhạc có nhãn thể loại Bolero, Cải lương, Chèo, Hát bội tương ứng.

Bi, Cai, Ci, Hi (i = 1,...,4): số tiên đoán sai đối với các file nhạc được gán nhãn thể loại Bolero, Cải lương, Chèo, Hát bội tương ứng.

**Bảng 2: Ma trận đánh giá độ chính xác phân loại**

Thể loại thực tế	Thể loại	Thể loại tiên đoán (Kết quả tiên đoán từ hệ thống đề xuất)				Tổng cộng
		Bolero	Cải lương	Chèo	Hát bội	
	Bolero	B	B1	B2	B3	100
	Cải lương	Ca1	Ca	Ca2	Ca3	100
	Chèo	C1	C2	C	C3	100
	Hát bội	H1	H2	H3	H	100

Như vậy, dòng tương ứng với thể loại thật sự của các file nhạc và cột tương ứng với thể loại tiên đoán của các file nhạc sau khi hệ thống đề xuất thực hiện phân loại. Số tập tin nhạc được gán nhãn thể loại đúng nằm trên đường chéo của ma trận (các giá trị in đậm: C, R, J, P). Đề đánh giá hiệu quả của phương pháp đề xuất, độ chính xác phân loại A (Accuracy) được sử dụng và được xác định bởi công thức (16):

$$A(\%) = \frac{B + Ca + C + H}{(B + Ca + C + H) + \sum_{i=1}^4 Bi + Cai + Ci + Hi} \times 100\% \quad (16)$$

Phương pháp của chúng tôi được thực hiện trong môi trường Visual C++ trên máy tính để thực hiện cài đặt hệ thống phân loại nhạc theo thể loại. Việc phân loại nhạc được thực hiện chủ yếu dựa vào 2 tập đặc trưng được rút trích từ tín hiệu audio: âm sắc và nhịp nhạc. Nghiên cứu thực nghiệm trên 3 trường hợp:

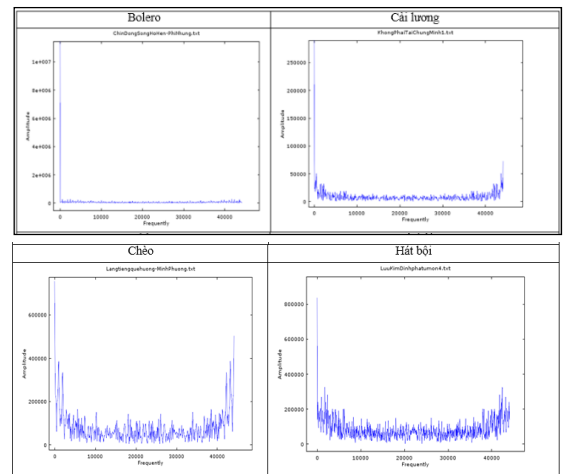
– Trường hợp 1: Huấn luyện và nhận dạng nhạc Việt Nam theo thể loại dựa trên các đặc trưng liên quan âm sắc bằng phương pháp KNN và SVM. Nghiên cứu chỉ dùng 9 đặc trưng trong tập đặc trưng âm sắc: Trung bình và phương sai của Spectral Centroid, Rolloff, Flux, ZeroCrossing (8), LowEnergy(1) (chưa tính các đặc trưng MFCC) trong việc phân loại nhạc theo thể loại.

– Trường hợp 2: Huấn luyện và nhận dạng nhạc Việt Nam theo thể loại dựa trên các đặc trưng liên quan nhịp điệu bằng phương pháp KNN và SVM. Nghiên cứu chỉ dùng 6 đặc trưng liên quan đến nhịp điệu (vector đặc trưng 6 chiều) trong việc phân loại nhạc theo thể loại.

– Trường hợp 3: Huấn luyện và nhận dạng nhạc Việt Nam theo thể loại dựa trên các đặc trưng liên quan bố cục âm sắc (nhịp điệu và âm sắc) bằng

phương pháp KNN và SVM. Nghiên cứu kết hợp 2 tập đặc trưng liên quan đến bố cục âm sắc và nhịp điệu (vector đặc trưng 25 chiều) trong việc phân loại nhạc theo thể loại. Chúng tôi kiểm tra trên hệ thống với việc rút trích đặc trưng dựa trên một trong các tập đặc trưng trên hoặc kết hợp chúng với nhau và sau đó tìm giá trị tham số k (số láng giềng gần nhất) sao cho hệ thống đạt hiệu quả về độ chính xác phân loại cao nhất.

Kết quả biểu đồ nhịp điệu (BH) của bốn thể loại nhạc: bolero, cải lương, chèo và hát bội với các bài nhạc sử dụng tương ứng các loại trên là:



**Hình 10: Biểu đồ nhịp điệu của bốn thể loại nhạc**

Chúng tôi thực nghiệm trên hệ thống với việc phân loại dựa trên chỉ một hoặc 2 tập đặc trưng. Chúng tôi cũng kiểm tra trên 1 số giá trị tham số k (k = 1, 2, 3, ..., 7). Trong đó, với giá trị k = 4, hệ thống cho kết quả phân loại tốt nhất. Vì vậy, chúng tôi chọn trình bày trong trường hợp này. Sau đây là kết quả đánh giá độ chính xác của việc phân loại.



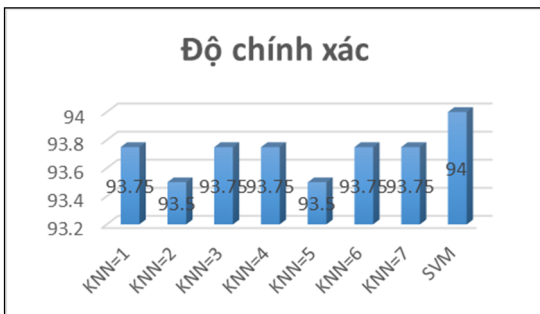
**Bảng 3: Độ chính xác phân loại dựa trên 1 tập đặc trưng với giá trị tham số k = 4**

A (%)	Tập các đặc trưng			
	ĐT1 (âm sắc)		ĐT2 (nhịp điệu)	
	KNN	SVM	KNN	SVM
	93 %	93.5 %	92.25 %	92.5 %

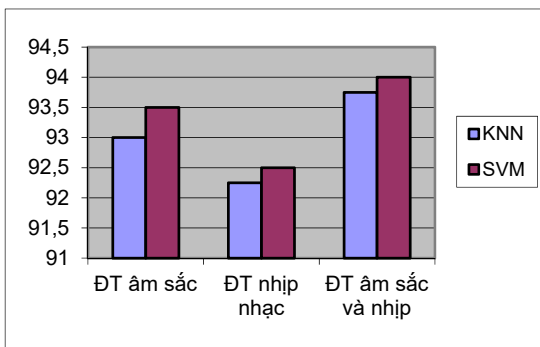
**Bảng 4: Độ chính xác phân loại dựa trên 2 tập đặc trưng với giá trị tham số k = 4**

A (%)	Tập các đặc trưng ĐT1 và ĐT2	
	KNN	SVM
	93,75 %	94 %

Từ kết quả trình bày trong Bảng 3 và Bảng 4, chúng tôi nhận xét: Nếu chúng tôi chỉ sử dụng 1 tập đặc trưng thì việc phân loại nhạc theo thể loại từ tín hiệu audio đạt độ chính xác thấp. Kết quả này khó có thể chấp nhận được. Vì vậy, chúng tôi tiếp tục kiểm tra trên hệ thống mà trong đó sử dụng cả 2 tập đặc trưng 1 và 2 trong việc phân loại nhạc (tạo nên một vector đặc trưng 25 chiều) với mong muốn làm tăng độ chính xác của việc phân loại. Trong đó, hệ thống cho kết quả phân loại tốt nhất với k = 4. Nguyên nhân là với k = 4 hệ thống phân loại nhạc theo thể loại đề xuất đạt độ chính xác là: 93%. Với các giá trị khác của k, kết quả độ chính xác phân loại thấp hơn. Chẳng hạn: k = 2 độ chính xác chỉ đạt 92,75 %.



**Hình 11: Đồ thị biểu diễn độ chính xác phân loại sử dụng kết hợp cả 2 tập đặc trưng**



**Hình 12: Đồ thị biểu diễn độ chính xác trung bình phân loại nhạc dựa vào các tập đặc trưng**

Từ các kết quả thực nghiệm trên tập dữ liệu kiểm tra biểu diễn trong Hình 8 cho thấy nếu hệ thống chỉ sử dụng một trong 2 tập đặc trưng về âm sắc hoặc nhịp điệu, thì việc phân loại nhạc theo thể loại từ tín hiệu audio được thực hiện nhanh hơn (thời gian thực hiện trung bình là 3 giây) do số chiều của vector đặc trưng nhỏ hơn, nhưng độ chính xác của việc phân loại sẽ thấp hơn (đạt khoảng 92,25% - 93%) so với trường hợp phân loại nhạc dựa trên cả 2 tập đặc trưng này. Vì vậy, việc sử dụng kết hợp cả 2 tập đặc trưng âm sắc, nhịp điệu là rất cần thiết đối với hệ thống phân loại nhạc theo thể loại vì nó cho kết quả phân loại khá chính xác.

Như vậy, phương pháp đề xuất của chúng tôi là kết hợp cả 2 tập đặc trưng âm sắc và nhịp điệu trong việc phân loại nhạc theo thể loại bởi vì hệ thống đưa ra kết quả phân loại với độ chính xác cao (trung bình 93,75%).

Thời gian phân loại (thời gian huấn luyện + thời gian rút trích đặc trưng + thời gian xác định thể loại) 1 tập tin nhạc cụ thể sử dụng phương pháp KNN:

**Bảng 5: Bảng thời gian phân loại 1 tập tin nhạc cụ thể**

STT	Trường hợp	Thời gian phân loại (giây)
1	TH1	4 giây
2	TH2	5 giây
3	TH3	6 giây

Trong đó: TH1: chỉ sử dụng 9 đặc trưng trong tập đặc trưng liên quan âm sắc (chưa tính các hệ số MFCC), TH2: chỉ sử dụng tập đặc trưng liên quan nhịp điệu, TH3: sử dụng cả 2 tập đặc trưng âm sắc và nhịp điệu.

**5 KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

Một phương pháp phân loại nhạc theo thể loại nhanh và chính xác là rất cần thiết đối với các hệ thống quản lý một số lượng lớn nhạc số. Tuy nhiên, đây là một công việc không đơn giản vì các thể loại nhạc vẫn còn là một khái niệm mờ, tùy thuộc vào ý kiến chủ quan của con người. Trong nghiên cứu thực nghiệm này, chúng tôi đề xuất sử dụng các tập đặc trưng được rút trích bởi các công cụ STFT, DWT, bộ phân loại KNN và SVM. DWT là một kỹ thuật phân tích tín hiệu, cung cấp một cách biểu diễn tín hiệu trong miền thời gian và tần số dưới dạng nén làm cho việc tính toán nhanh và hiệu quả hơn. Nghiên cứu này tập trung vào việc phân loại 4 thể loại nhạc: Bolero Việt, Cải lương, Chèo và Hát bội bằng cách sử dụng kết hợp 2 tập đặc trưng về âm sắc và nhịp điệu. Tập dữ liệu được sử dụng trong nghiên cứu này được sưu tập từ các nguồn nhạc Việt Nam. Dựa trên các kết quả thực nghiệm, phương pháp đề

xuất của chúng tôi đạt độ chính xác trung bình là 93,75% và 94 % đối với phương pháp phân loại KNN và SVM tương ứng trên tập đặc trưng về bố cục âm sắc. Hơn nữa, phương pháp đề xuất này đơn giản, hiệu quả và có thời gian thực hiện nhanh phù hợp cho các hệ thống phân loại nhạc Việt hiện nay.

Việc phân loại nhạc theo thể loại được thực hiện một cách tự động bằng máy tính và cho kết quả khá chính xác là hoàn toàn có thể. Nghiên cứu này cung cấp cơ sở khoa học cho phát triển các hệ thống: truy vấn thông tin nhạc dựa vào nội dung, phát hiện sao chép bản quyền nhạc, tìm các bản nhạc có các đặc trưng gần giống với các đặc trưng mà người sử dụng mong muốn, phân tích nhạc và lời bài hát, phân loại bản nhạc theo ca sĩ - nhạc sĩ, chú thích tự động các tập tin nhạc với những mô tả,... Nó có thể áp dụng cho việc phân loại thêm nhiều loại nhạc truyền thống của Việt Nam như: dân ca Bắc Bộ, dân ca Nam Bộ, nhạc trẻ,... Hệ thống đề xuất cũng có thể áp dụng với các bộ phân loại kết hợp khác như: Gaussian, mạng Neural,... Chúng tôi dự định thực nghiệm hệ thống đề xuất trên một tập dữ liệu lớn (Big Data) và thời gian phát file dài hơn; nghiên cứu và sử dụng thêm tập đặc trưng cao độ nốt nhạc nhằm nâng cao độ chính xác phân loại. Hệ thống thực hiện phân loại đối với các bản nhạc có sự pha trộn các thể loại, bổ sung thêm các thể loại nhạc Việt chưa được nghiên cứu ở đề tài này góp phần hình thành kho dữ liệu về âm nhạc Việt Nam. Đó chính là những hướng nghiên cứu của chúng tôi trong thời gian sắp tới.

## TÀI LIỆU THAM KHẢO

Anan, Yoko, Hatano, Kohei, Bannai, Hideo, and Takeda, Masayuki, "Music Genre Classification Using Similarity Functions", Proceedings of the 12th International Society for Music Information Retrieval Conference (Miami (Florida), USA), pp. 693-698, 2011.

Đỗ Thanh Nghị, "*Khai mở dữ liệu – Minh học bằng ngôn ngữ R*", Nhà xuất bản Đại học Cần Thơ, trang 7-52, Cần Thơ 2008.

G. Tzanetakis and P. Cook, "Musical genre classification of audio signals", IEEE Trans. on

speech and audio process, vol. 10, no. 5, pages 293–302, July 2002.

George Tzanetakis, Georg Essl and Perry Cook, "Automatic Musical Genre Classification of Audio Signals", 2nd Annual International Symposium on Music Information Retrieval 2001 ( ISMIR 2001), pages 1-6, 2001.

Li, Tao and Tzanetakis, G. , "Factors in automatic musical genre classification of audio signals", Applications of Signal Processing to Audio and Acoustics, IEEE Workshop, pp. 143-146, 2003.

Logan, Beth. "Mel Frequency Cepstral Coefficients for Music Modeling", Proceedings of the 1st International Conference on Music Information Retrieval (Plymouth (Massachusetts), USA October 23, 2000.

Phạm Thị Hòa và Ngô Thị Nam, "*Giáo dục âm nhạc*", Tập 1-Nhạc lý cơ bản – xướng âm, Nhà xuất bản Đại học Sư phạm, trang 7 -8, ĐH 2006.

Phạm Thị Hòa, "*Giáo dục âm nhạc*", Tập 2- Phương pháp giáo dục âm nhạc, Nhà xuất bản Đại học Sư phạm, trang 7 -9, ĐH 2007.

R. Tao, Z. Li, Y. Ji, and E. Bakker, "Music genre classification using temporal information and support vector machine", Proceedings of the Sixteenth annual conference of the Advanced School for Computing and Imaging (ASCI 2010), pages 1-7, 2010.

Rini Wongso and Diaz D. Santika, "Automatic music genre classification using dual tree complex wavelet transform and support vector machine", in Journal of Theoretical and Applied Information Technology, Vol. 63 No.1, pages 1-8, May 2014.

Tzanetakis, George, Essl, Georg, and Cook, Perry, "Automatic Musical Genre Classification of Audio Signals", Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (Bloomington (Indiana), USA), pp. 205-210, 2001.

Y.M.D. Chaturanga and K.L. Jayaratne, "Automatic Music Genre Classification of Audio Signals with Machine Learning Approaches", in GSTF International Journal on Computing (JoC), Vol. 3 No.2, pages 1-12, July 2013.