

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI



LUẬN VĂN THẠC SĨ

**Tối ưu hóa dữ liệu và mô hình phân lớp
nhằm nâng cao chất lượng nhận dạng cảm
xúc sử dụng bộ dữ liệu sinh học của MIT**

NGUYỄN TIẾN NAM

nguyennam291096@gmail.com

Kỹ thuật Điều khiển và tự động hóa

Giảng viên hướng dẫn: TS. Cung Thành Long

Viện: Điện

HÀ NỘI, 3/2021

ĐỀ TÀI LUẬN VĂN

Tối ưu hóa dữ liệu và mô hình phân lớp nhằm nâng cao chất lượng nhận dạng cảm xúc sử dụng bộ dữ liệu sinh học của MIT

Giảng viên hướng dẫn
Ký và ghi rõ họ tên

Lời cảm ơn

Để hoàn thành được luận văn này, thật không thể thiếu sự giúp sức của những người xung quanh. Đầu tiên, em xin gửi lời cảm ơn chân thành tới thầy hướng dẫn, TS. Cung Thành Long, người đã giúp đỡ em rất tận tình trong quá trình làm luận văn, trong việc hướng dẫn đề tài, lên ý tưởng và đưa ra các phương pháp nghiên cứu. Em xin bày tỏ lòng biết ơn sâu sắc đến tập thể thầy cô bộ môn đã giúp em có cái nhìn chuyên sâu về đề tài nghiên cứu, đưa ra những gợi ý và ý tưởng vô cùng đáng giá. Bên cạnh đó cũng là sự hỗ trợ không nhỏ của em Phạm Văn Dũng đã cùng nghiên cứu và thảo luận trong suốt thời gian qua.

Cuối cùng, em xin cảm ơn gia đình và đồng nghiệp, những người luôn giúp đỡ và ủng hộ trong suốt quá trình làm việc và nghiên cứu này.

Tóm tắt nội dung luận văn

- **Lý do chọn đề tài:**

Cảm xúc, ảnh hưởng đến tâm sinh lý của mỗi người, đóng một vai trò rất quan trọng trong cuộc sống của con người. Cảm xúc tích cực giúp cải thiện sức khỏe con người và hiệu quả công việc, trong khi cảm xúc tiêu cực có thể gây ra các vấn đề về sức khỏe, quá nhiều cảm xúc sẽ gây những ảnh hưởng không tốt cho việc đưa ra quyết định. Cảm xúc tương tác với suy nghĩ theo những cách không rõ ràng nhưng lại đóng vai trò quan trọng đối với hoạt động đòi hỏi trí thông minh. Ngược lại, sự tích tụ lâu dài của những cảm xúc tiêu cực là yếu tố dẫn đến các bệnh lý nghiêm trọng. Cảm xúc phản ánh trạng thái tinh thần một cách tự phát chứ không thông qua nỗ lực có ý thức. Cảm xúc đi kèm với những thay đổi về thể chất và sinh lý, liên quan đến các cơ quan của con người như não, tim, da, lưu lượng máu, cơ, nét mặt, giọng nói... Do mức độ phức tạp của sự tương tác lẫn nhau giữa sinh lý và tâm lý trong cảm xúc, nên việc nhận biết cảm xúc của con người một cách chính xác vẫn là một lĩnh vực nghiên cứu có nhiều thách thức.

Nhận biết biểu cảm cũng được cho là đóng một vai trò quan trọng trong việc học tập và phát triển, là một phần quan trọng trong việc đánh giá sự khác biệt giữa sự phát triển bình thường của trẻ em so với trẻ tự kỷ, những người thường bị suy giảm khả năng nhận biết cảm xúc.

Nhận dạng cảm xúc đã bước đầu được áp dụng trong các lĩnh vực như lái xe an toàn, theo dõi sức khỏe con người,...Ngoài ra, nhận dạng cảm xúc con người còn có thể ứng dụng trong phát triển các ứng dụng học từ xa có cảm xúc, phát triển các ứng dụng tương tác có cảm xúc giữa người và máy, phát triển robot thông minh, cùng nhiều ứng dụng an ninh khác.

Với một số vai trò của nhận dạng cảm xúc con người như được đề cập phía trên, học viên nhận thấy đây là hướng nghiên cứu nhiều tiềm năng ứng dụng, và nhiều thách thức về mặt khoa học. Đó cũng là lý do học viên lựa chọn một bài toán cụ thể trong lĩnh vực nghiên cứu này làm đề tài cho luận văn tốt nghiệp.

- **Mục đích nghiên cứu của luận văn, đối tượng, phạm vi nghiên cứu:**

- Nghiên cứu bộ dữ liệu chuẩn của MIT
- Nhận dạng 8 trạng thái cảm xúc khác nhau ở người bao gồm: trung tính, tức giận, ghét bỏ, đau buồn, đồng cảm, lãng mạn, vui vẻ và tôn kính từ bộ dữ liệu nêu trên.
- Tổ chức lại bộ dữ liệu đầu vào của MIT nhằm nâng cao kết quả nhận dạng cảm xúc.

- **Tóm tắt cô đọng các nội dung chính và đóng góp mới của tác giả:**

- Luận văn tốt nghiệp này đã áp dụng một số kỹ thuật học máy truyền thống nhằm nhận dạng các trạng thái cảm xúc khác nhau từ bộ tín hiệu sinh học chuẩn của MIT, qua đó lựa chọn kỹ thuật cho kết quả khả quan...

- Đưa ra phương pháp tổ chức lại dữ liệu đầu vào nhằm nâng cao chất lượng kết quả nhận dạng. Trong đó, thực hiện nhận dạng cảm xúc trên các đoạn dữ liệu ngắn, thực hiện chồng chập dữ liệu vào nhằm nâng cao chất lượng nhận dạng, tìm tỉ lệ chồng chập tối ưu theo nghĩa để đảm bảo có kết quả nhận dạng tốt trở lên.

- **Phương pháp nghiên cứu:**

- Nghiên cứu lý thuyết kết hợp với thực nghiệm để so sánh khả năng của các kỹ thuật khác nhau.

- Đề xuất phương pháp riêng, sử dụng các công cụ đã được kiểm chứng nhằm đánh giá tính đúng đắn và hiệu quả của các đề xuất...

- **Kết luận:**

Nội dung luận văn thực hiện đúng nhiệm vụ nghiên cứu đặt ra, bao gồm tìm hiểu bộ dữ liệu MIT; tìm hiểu các kỹ thuật trích chọn đặc trưng và ra quyết định nhằm thực hiện bài toán nhận dạng cảm xúc; thực hiện ý tưởng chồng chập dữ liệu nhằm nâng cao chất lượng nhận dạng; thử nghiệm nhiều tỉ lệ chồng chập khác nhau để đánh giá tỉ lệ chồng chập tối ưu. Các kết quả bước đầu thu được là khả quan, đáp ứng các yêu cầu đặt ra của luận văn.

HỌC VIÊN

Ký và ghi rõ họ tên

MỤC LỤC

CHƯƠNG 1. TỔNG QUAN VỀ NHẬN DẠNG CẢM XÚC	1
1.1 Nhận dạng cảm xúc thông qua tín hiệu giọng nói	2
1.2 Nhận dạng cảm xúc thông qua biểu cảm khuôn mặt	2
1.3 Nhận dạng cảm xúc thông qua tín hiệu sinh học	3
CHƯƠNG 2. HỌC MÁY VÀ ỨNG DỤNG TRONG NHẬN DẠNG CẢM XÚC	10
2.1 Lựa chọn đặc trưng	11
2.1.1 Recursive Feature Elimination (RFE).....	12
2.1.2 Pearson Correlation và Chi-Squared.....	13
2.2 Phân lớp và ra quyết định nhận dạng	14
2.2.1 Support Vector Machine (SVM).....	14
2.2.2 Linear Discriminant Analysis (LDA)	15
2.2.3 Decision Tree (DT)	17
2.2.4 Random Forest (RF).....	18
2.3 Áp dụng với bộ dữ liệu MIT	19
2.3.1 Các bước thực hiện	19
2.3.2 Thử nghiệm lựa chọn mô hình phân lớp phù hợp.....	22
CHƯƠNG 3. TỔ CHỨC DỮ LIỆU ĐẦU VÀO NÂNG CAO CHẤT LƯỢNG NHẬN DẠNG CẢM XÚC	25
3.1 Tỷ lệ chồng chập 10%	25
3.2 Tỷ lệ chồng chập 20%	27
3.3 Tỷ lệ chồng chập 30%	28
3.4 Tỷ lệ chồng chập 40%	30
3.5 Tỷ lệ chồng chập 50%	31
3.6 Tỷ lệ chồng chập 60%	32
3.7 Tỷ lệ chồng chập 70%	34
3.8 Tỷ lệ chồng chập 80%	35
3.9 Tỷ lệ chồng chập 90%	36
3.10 Phân tích xác định tỷ lệ chồng chập tối ưu	38
3.11 Kết quả nhận dạng với từng trạng thái cảm xúc	39
3.12 Nhận dạng với mô hình ba trạng thái của cảm xúc.....	46
KẾT LUẬN CHUNG VÀ HƯỚNG PHÁT TRIỂN.....	48

Tài liệu tham khảo	1
---------------------------------	----------

DANH MỤC HÌNH VẼ

Hình 1.1 Mô hình trạng thái cảm xúc [1].....	1
Hình 1.2 Một số trạng thái cảm xúc qua biểu hiện khuôn mặt [6]	3
Hình 1.3 Vị trí gắn các cảm biến sinh học [12]	3
Hình 1.4 Các tín hiệu tương ứng với trạng thái cảm xúc trung tính	5
Hình 1.5 Các tín hiệu tương ứng với trạng thái cảm xúc tức giận	5
Hình 1.6 Các tín hiệu tương ứng với trạng thái cảm xúc ghét bỏ	6
Hình 1.7 Các tín hiệu tương ứng với trạng thái cảm xúc đau buồn	6
Hình 1.8 Các tín hiệu tương ứng với trạng thái cảm xúc đồng cảm	7
Hình 1.9 Các tín hiệu tương ứng với trạng thái cảm xúc lãng mạn	7
Hình 1.10 Các tín hiệu tương ứng với trạng thái cảm xúc vui vẻ	8
Hình 1.11 Các tín hiệu tương ứng với trạng thái cảm xúc tôn trọng	8
Hình 2.1 Học máy và ứng dụng của học máy [17]	10
Hình 2.2 Lựa chọn đặc trưng	12
Hình 2.3 Sơ đồ hoạt động thuật toán RFE	13
Hình 2.4 Đường thẳng phân cách điểm dữ liệu của 2 lớp [19]	15
Hình 2.5 Chiếu dữ liệu lên không gian nhiều chiều hơn	15
Hình 2.6 Phép chiếu lên đường thẳng [20]	16
Hình 2.7 Mô hình Decision Tree [22]	17
Hình 2.8 Mô hình Random Forest [23]	19
Hình 2.9 Ảnh hưởng của lựa chọn đặc trưng lên các mô hình phân lớp	21
Hình 2.10 Phương pháp hiệu chỉnh tham số mô hình bằng cross-validation	22
Hình 2.11 Cách chia dữ liệu với $n = 5$	22
Hình 2.12 Kết quả nhận dạng bằng mô hình SVM với dữ liệu không chồng chập	23
Hình 2.13 Kết quả nhận dạng bằng mô hình RF với dữ liệu không chồng chập	24
Hình 3.1 Chia dữ liệu khi không chồng chập	25
Hình 3.2 Chia dữ liệu khi áp dụng kỹ thuật chồng chập	25
Hình 3.3 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 10%	26
Hình 3.4 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 10%	26
Hình 3.5 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 20%	27
Hình 3.6 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 20%	28
Hình 3.7 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 30%	29
Hình 3.8 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 30%	29
Hình 3.9 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 40%	30
Hình 3.10 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 40%	30
Hình 3.11 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 50%	31

Hình 3.12 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 50%	32
Hình 3.13 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 60%	33
Hình 3.14 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 60%	33
Hình 3.15 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 70%	34
Hình 3.16 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 70%	34
Hình 3.17 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 80%	35
Hình 3.18 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 80%	36
Hình 3.19 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 90%	37
Hình 3.20 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 90%	37
Hình 3.21 Tổng hợp kết quả nhận dạng cảm xúc khi thay đổi tỉ lệ chồng chập .	38
Hình 3.22 Kết quả nhận dạng cảm xúc trung tính với từng tỉ lệ chồng chập	39
Hình 3.23 Kết quả nhận dạng cảm xúc tức giận với từng tỉ lệ chồng chập	40
Hình 3.24 Kết quả nhận dạng cảm xúc ghét bỏ với từng tỉ lệ chồng chập	41
Hình 3.25 Kết quả nhận dạng cảm xúc đau buồn với từng tỉ lệ chồng chập	42
Hình 3.26 Kết quả nhận dạng cảm xúc đồng cảm với từng tỉ lệ chồng chập	43
Hình 3.27 Kết quả nhận dạng cảm xúc lãng mạn với từng tỉ lệ chồng chập	44
Hình 3.28 Kết quả nhận dạng cảm xúc vui vẻ với từng tỉ lệ chồng chập	45
Hình 3.29 Kết quả nhận dạng cảm xúc tôn trọng với từng tỉ lệ chồng chập	46
Hình 3.30 Kết quả nhận dạng 3 trạng thái của cảm xúc với SVM	47
Hình 3.31 Kết quả nhận dạng 3 trạng thái của cảm xúc với RF	47

DANH MỤC BẢNG

Bảng 2.1 Kết quả nhận dạng cảm xúc của từng mô hình.....	23
Bảng 3.1 Kết quả nhận dạng của 2 mô hình SVM và RF	38
Bảng 3.2 Tỷ lệ nhận dạng của trạng thái cảm xúc trung tính	39
Bảng 3.3 Tỷ lệ nhận dạng của trạng thái cảm xúc tức giận	39
Bảng 3.4 Tỷ lệ nhận dạng của trạng thái cảm xúc ghét bỏ	40
Bảng 3.5 Tỷ lệ nhận dạng của trạng thái cảm xúc đau buồn	41
Bảng 3.6 Tỷ lệ nhận dạng của trạng thái cảm xúc đồng cảm	42
Bảng 3.7 Tỷ lệ nhận dạng của trạng thái cảm xúc lãng mạn	43
Bảng 3.8 Tỷ lệ nhận dạng của trạng thái cảm xúc vui vẻ	44
Bảng 3.9 Tỷ lệ nhận dạng của trạng thái cảm xúc tôn trọng	45

DANH MỤC TỪ VIẾT TẮT

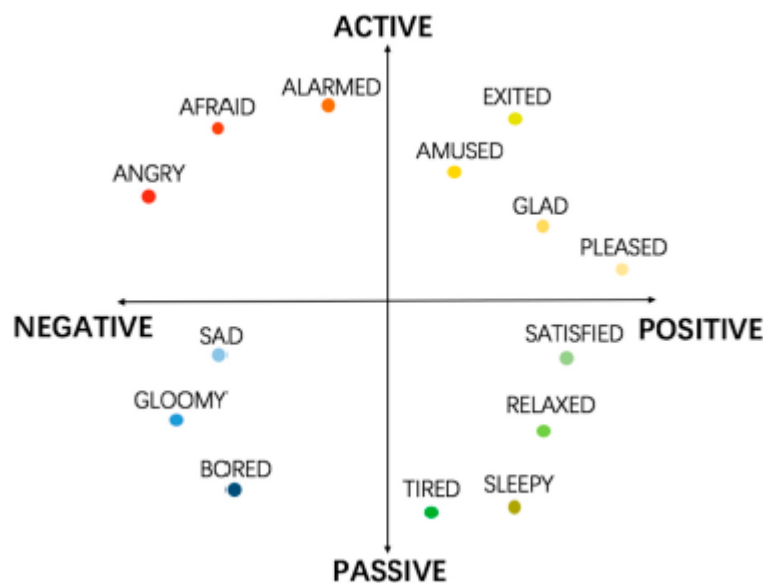
EMG	Electromyogram	Điện cơ đồ
BVP	Blood Volume Pressure	Xung thể tích máu
SC	Skin Conductivity	Độ điện dẫn da
RSP	Respiration	Cảm biến hô hấp
ML	Machine Learning	Học máy
RFE	Recursive Feature Elimination	Loại bỏ đặc trưng
SVM	Support Vector Machine	
LDA	Linear Discriminant Analysis	
DT	Decision Tree	
RF	Random Forest	
HMM	Hidden Markov model	
NN	Neural Network	

CHƯƠNG 1. TỔNG QUAN VỀ NHẬN DẠNG CẢM XÚC

• Bài toán nhận dạng cảm xúc

Cảm xúc, ảnh hưởng đến tâm sinh lý của mỗi người, đóng một vai trò rất quan trọng trong cuộc sống của con người. Cảm xúc tích cực giúp cải thiện sức khỏe con người và hiệu quả công việc, trong khi cảm xúc tiêu cực có thể gây ra các vấn đề về sức khỏe, quá nhiều cảm xúc sẽ gây những ảnh hưởng không tốt cho việc đưa ra quyết định. Cảm xúc tương tác với suy nghĩ theo những cách không rõ ràng nhưng lại đóng vai trò quan trọng đối với hoạt động đòi hỏi trí thông minh. Ngược lại, sự tích tụ lâu dài của những cảm xúc tiêu cực là yếu tố dẫn đến các bệnh lý nghiêm trọng. Cảm xúc phản ánh trạng thái tinh thần một cách tự phát chứ không thông qua nỗ lực có ý thức và đi kèm với những thay đổi về thể chất và sinh lý liên quan đến các cơ quan của con người như não, tim, da, lưu lượng máu, cơ, nét mặt, giọng nói,... Do mức độ phức tạp của sự tương tác lẫn nhau giữa sinh lý và tâm lý trong cảm xúc, việc nhận biết cảm xúc của con người một cách chính xác vẫn là mục tiêu của nghiên cứu khoa học. Nhận biết biểu cảm được cho là đóng một vai trò quan trọng trong việc học tập và phát triển, là một phần quan trọng trong việc đánh giá sự khác biệt giữa sự phát triển bình thường của trẻ em so với trẻ tự kỷ, những người thường bị suy giảm khả năng nhận biết cảm xúc. Nhận diện cảm xúc đã được áp dụng trong nhiều lĩnh vực như lái xe an toàn, theo dõi sức khỏe con người...

Để nhận biết được cảm xúc, trước hết các trạng thái cảm xúc cần được xác định một cách rõ ràng. Định nghĩa về những cảm xúc cơ bản lần đầu tiên được đề xuất cách đây nhiều thập kỷ. Các nhà tâm lý học có xu hướng mô hình hóa cảm xúc theo hai cách khác nhau. Một là chia cảm xúc thành các loại rời rạc. Cách khác là sử dụng nhiều chiều để gắn nhãn cảm xúc. Hình 1.1 là mô hình cảm xúc được đề xuất bởi Lang, P.J, với các trạng thái phân lớp theo trục tích cực - tiêu cực và thụ động - chủ động.



Hình 1.1 Mô hình trạng thái cảm xúc [1]

Các phương pháp nhận biết cảm xúc có thể được phân thành hai loại chính:

Sử dụng các tín hiệu biểu cảm của con người như nét mặt, giọng nói, cử chỉ, tư thế... Các tín hiệu này có ưu điểm là dễ thu thập và đã được nghiên cứu trong nhiều năm. Tuy nhiên, độ tin cậy không thể được đảm bảo vì mọi người tương đối dễ dàng kiểm soát các tín hiệu như nét mặt hoặc giọng nói để che giấu cảm xúc thực của họ, đặc biệt là trong giao tiếp xã hội. Một người có thể tỏ ra vui vẻ ngay cả đang ở trong trạng thái cảm xúc tiêu cực.

Phương pháp thứ hai sử dụng các tín hiệu sinh học để nhận dạng cảm xúc. Các tín hiệu sinh học, bao gồm điện não đồ, nhiệt độ, điện tâm đồ, điện cơ đồ, hô hấp,... Các tín hiệu sinh học sẽ thay đổi theo một cách nhất định khi con người đối mặt với một số tình huống cụ thể. Một trong những lợi thế chính của phương pháp thứ hai là các thông số đo được phần lớn được kích hoạt tự nhiên và do đó con người không thể dễ dàng kiểm soát được. Đã có nhiều nỗ lực nghiên cứu trong lĩnh vực nhận biết cảm xúc bằng cách sử dụng các tín hiệu sinh học và thiết lập một tiêu chuẩn và mối quan hệ cố định giữa những thay đổi cảm xúc và tín hiệu sinh học. Tuy nhiên, nghiên cứu cho thấy rằng tương đối khó để phản ánh chính xác những thay đổi cảm xúc bằng cách sử dụng một tín hiệu sinh học duy nhất. Do đó, nhận dạng cảm xúc sử dụng tín hiệu sinh học thường phải xử lý đồng thời loại tín hiệu khác nhau để đưa ra quyết định nhận dạng. Đây là phương pháp nhận dạng có độ tin cậy cao, có ý nghĩa trong nghiên cứu và ứng dụng thực tế.

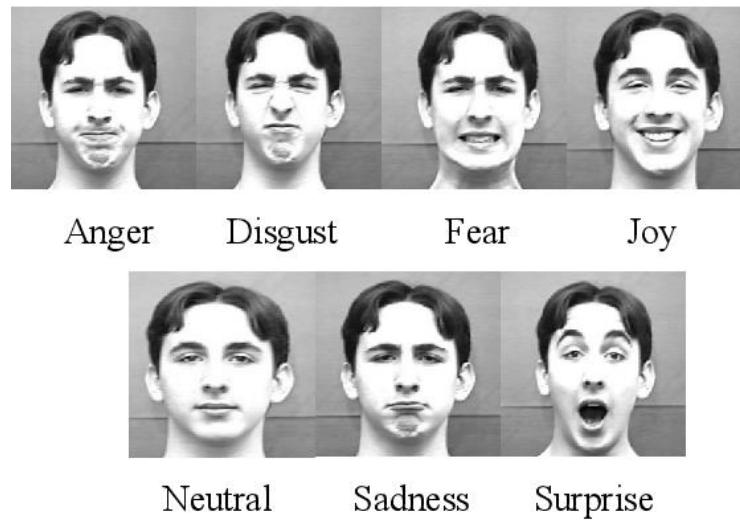
1.1 Nhận dạng cảm xúc thông qua tín hiệu giọng nói

Sử dụng tiếng nói là một trong những cách đầu tiên để nhận biết được cảm xúc của một người một cách tức thời. Khi ở trong trạng thái tức giận đi kèm căng thẳng, giọng nói của người đó có xu hướng to hơn, tốc độ nói nhanh, ngữ điệu lên xuống thất thường. Ngược lại, khi ở trong trạng thái thông thường và không bị chi phối bởi cảm xúc tiêu cực, giọng nói của một người có xu hướng ổn định về âm lượng, tốc độ nói trung bình, biên độ giọng nói không thay đổi đột ngột.

Một số nghiên cứu nhận dạng cảm xúc bằng phương pháp này có thể kể tới, như sử dụng mô hình HMM đạt độ chính xác tới 86% [2], mô hình SVM và NN với 84,2% và 80,8% [3], khi sử dụng GMM đạt độ chính xác 78,14% [4] và nhận dạng 4 trạng thái cảm xúc đạt kết quả 68,2% khi sử dụng kỹ thuật phân lớp SVM [5].

1.2 Nhận dạng cảm xúc thông qua biểu cảm khuôn mặt

Nhận dạng cảm xúc thông qua biểu cảm khuôn mặt là bài toán quen thuộc trong lĩnh vực xử lý ảnh. Để nhận biết nhanh nhất cảm xúc của một người, khuôn mặt là thứ đầu tiên được chú ý đến, với những đặc trưng của lông mày, hình dạng khuôn miệng, mũi, hay những nếp nhăn, độ mở của mắt,... Ví dụ khi ở trong trạng thái vui vẻ, khuôn miệng sẽ mở rộng, hướng lên, má nâng cao và có các vết nhăn xung quanh mắt. Hình 1.2 mô tả các biểu cảm của khuôn mặt của cùng một người với các trạng thái cảm xúc khác nhau.



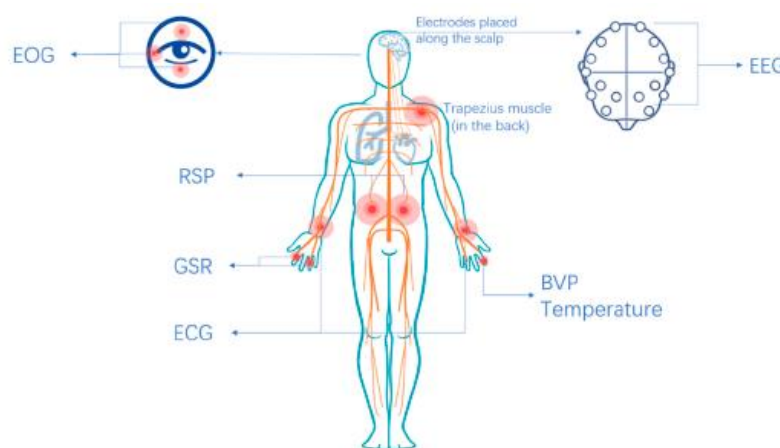
Hình 1.2 Một số trạng thái cảm xúc qua biểu hiện khuôn mặt [6]

Một số nghiên cứu trong lĩnh vực này có thể kể đến là, nghiên cứu nhận dạng 6 loại cảm xúc của B. A. Dixit và A. N. Gaikwad đạt độ chính xác 81,66% [7], hay kết hợp cả dữ liệu ảnh và giọng nói trong nghiên cứu của S.Ashok Kumar với kết quả đạt độ chính xác 60% [8]. Một nghiên cứu khác sử dụng mô hình 3D của khuôn mặt với mô hình phân lớp KNN [9] cho kết quả 82% khi nhận diện 7 loại cảm xúc, hay nhận dạng với mô hình SVM [10] hoặc HMM [11] cho kết quả với độ chính xác lần lượt là 95% và 82,46%.

Phương pháp nhận dạng cảm xúc qua hình ảnh biểu cảm trên khuôn mặt có ưu điểm dễ dàng thu thập dữ liệu, số lượng tập dữ liệu có sẵn lớn so với phương pháp sử dụng tín hiệu sinh học. Tuy nhiên các biểu cảm trên khuôn mặt hoàn toàn có thể làm giả, hoặc phản ứng miễn cưỡng, không tự nhiên làm cho kết quả nhận dạng không thực sự đáng tin cậy.

1.3 Nhận dạng cảm xúc thông qua tín hiệu sinh học

Để thu được các tín hiệu sinh học, người ta gắn các cảm biến trên cơ thể người. Hình 1.3 mô tả các vị trí thường dùng để gắn một số loại cảm biến sinh học.



Hình 1.3 Vị trí gắn các cảm biến sinh học [12]

Các tín hiệu được sử dụng để nhận dạng cảm xúc thường là xung thể tích máu (BVP), điện não đồ (EEG), điện tâm đồ (ECG), điện dẫn da (GSR – SC), nhịp tim (HR), huyết áp (BP), nhịp thở (RSP)... Cảm xúc có thể được kích thích bằng nhiều cách khác nhau như thông qua video, âm nhạc,... Đã có các nghiên cứu về nhận dạng cảm xúc sử dụng tín hiệu sinh học đáng chú ý như sử dụng tín hiệu điện não đồ đạt tỉ lệ chính xác 75,18% [13], sử dụng mô hình LSTM đạt tỉ lệ 74,38% [14]. Nhận diện 4 cảm xúc vui vẻ, giận dữ, buồn chán và hạnh phúc đạt 76% tỉ lệ chính xác [15], hay nhận diện 2 trạng thái tiêu cực và tích cực với 73,10% tỉ lệ chính xác [16] cùng sử dụng mô hình SVM.

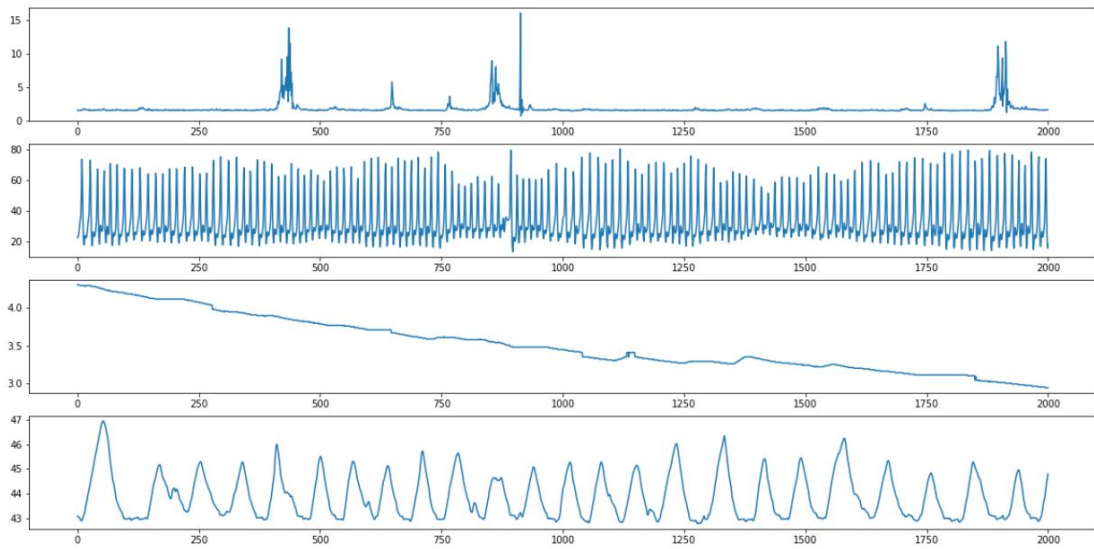
Ưu điểm của phương pháp này là, sử dụng các tín hiệu trên cơ thể con người, chúng thay đổi hoàn toàn tự nhiên, không thể ép buộc hay miễn cưỡng theo các trạng thái cảm xúc, nên kết quả đánh giá có độ tin cậy cao. Ngược lại, nhược điểm của phương pháp này là cần thiết lập kỹ lưỡng hệ đo trước khi tiến hành thu thập dữ liệu, tức quá trình lấy mẫu phức tạp hơn so với phương pháp sử dụng giọng nói hay biểu cảm khuôn mặt. Ngoài ra, việc gán nhãn các trạng thái cảm xúc tương ứng với thể hiện của các tín hiệu không mang tính trực quan, phụ thuộc vào việc kích thích thích hợp để bộc lộ đúng cảm xúc, phụ thuộc vào việc nhận diện đúng trạng thái cảm xúc của chính người tham gia lấy mẫu,... cũng là điểm khó và ảnh hưởng tới độ chính xác của phương pháp này.

• **Bộ dữ liệu sinh học của MIT**

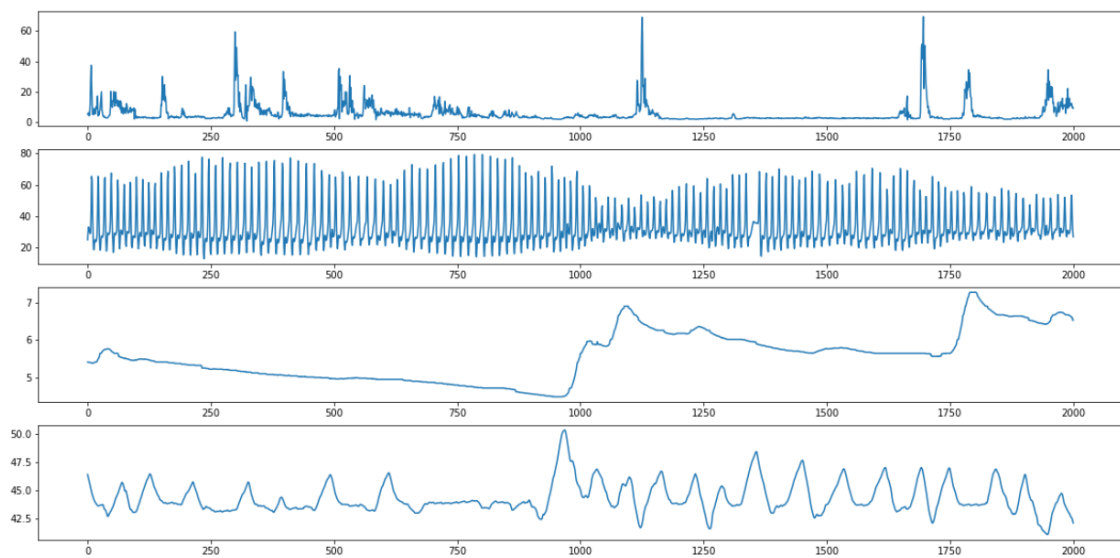
Bộ dữ liệu sinh học của MIT được thiết kế để phân loại tối đa lên đến 8 loại cảm xúc khác nhau bao gồm:

- Trung tính: cảm giác thông thường, trung lập
- Tức giận: giận dữ, mong muốn được giải tỏa bằng các hình thức bạo lực
- Ghét bỏ: trạng thái bị động của cảm xúc tức giận
- Đau buồn: sự mất mát, buồn chán, thất vọng
- Đồng cảm: vui vẻ, cảm thấy yên bình
- Lãng mạn: cảm thấy thu hút, thích thú
- Vui vẻ: cảm giác hạnh phúc
- Tôn trọng: trạng thái ngưỡng mộ, kính trọng

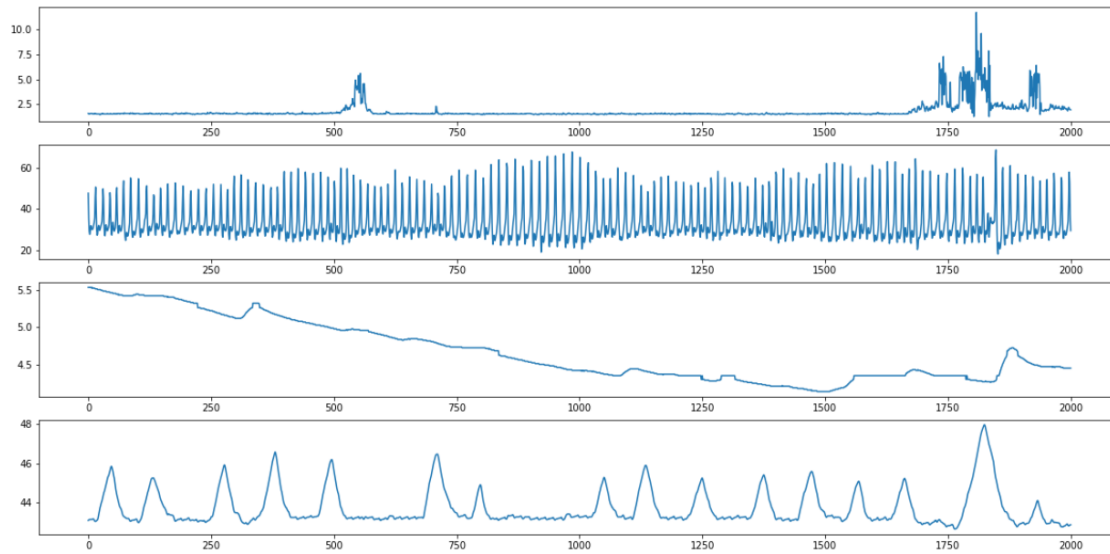
Để tạo ra bộ dữ liệu này, các nhà khoa học tại MIT đã thực hiện đeo 4 loại cảm biến lên đối tượng thử nghiệm: điện cơ đồ, điện dẫn da, xung thể tích máu, cảm biến hô hấp. Thử nghiệm được đo trong nhiều ngày, tuy nhiên có những ngày đo giá trị cảm biến bị lỗi do những lý do khách quan như môi trường hay thiết bị đeo bị lỏng trong quá trình đo, do vậy sau khi loại bỏ đi những ngày giá trị bị lỗi, dữ liệu còn lại bao gồm 20 ngày đo, mỗi ngày bao gồm các giá trị ứng với 8 loại cảm xúc bên trên. Từ Hình 1.4 đến Hình 1.9 biểu diễn các giá trị thu được từ 4 loại cảm biến ứng với 8 loại cảm xúc khác nhau. Mỗi đoạn dữ liệu bao gồm 2000 điểm đo thu được từ 4 cảm biến tương ứng với 100 giây trong thực tế. Một ngày, bao gồm 8 cảm xúc, mỗi một cảm xúc có 4 đoạn dữ liệu ứng với 4 cảm biến khác nhau, tổng cộng thu được 32 đoạn dữ liệu cho một ngày và 640 đoạn dữ liệu cho 20 ngày đo.



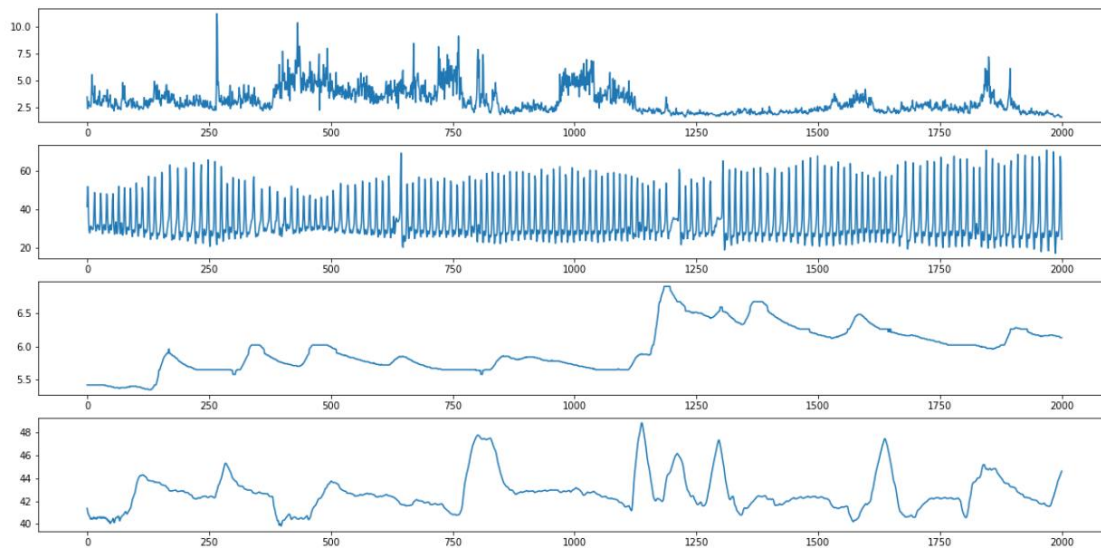
Hình 1.4 Các tín hiệu tương ứng với trạng thái cảm xúc trung tính



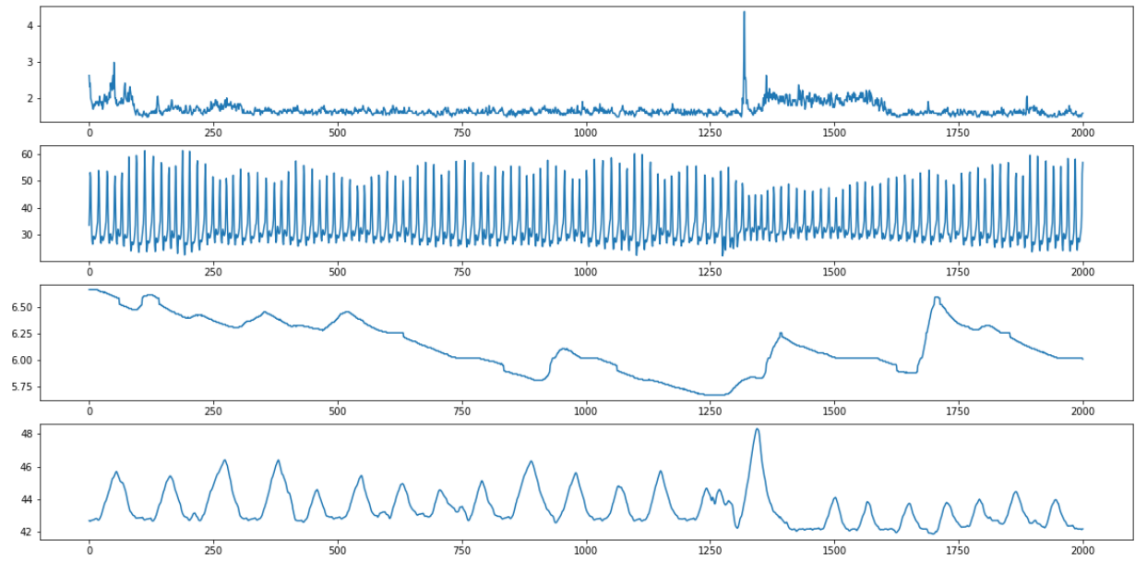
Hình 1.5 Các tín hiệu tương ứng với trạng thái cảm xúc tức giận



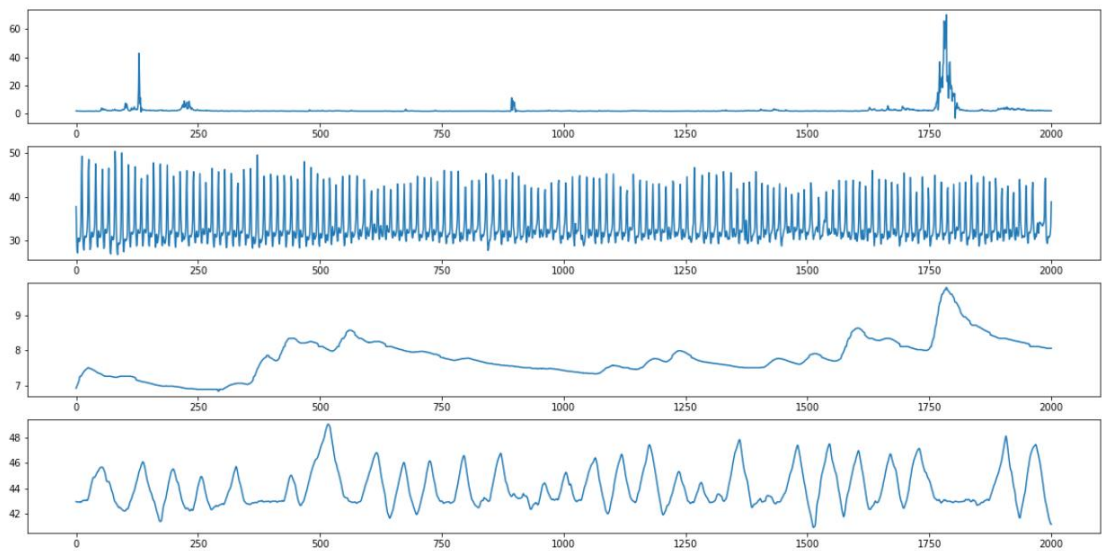
Hình 1.6 Các tín hiệu tương ứng với trạng thái cảm xúc ghét bỏ



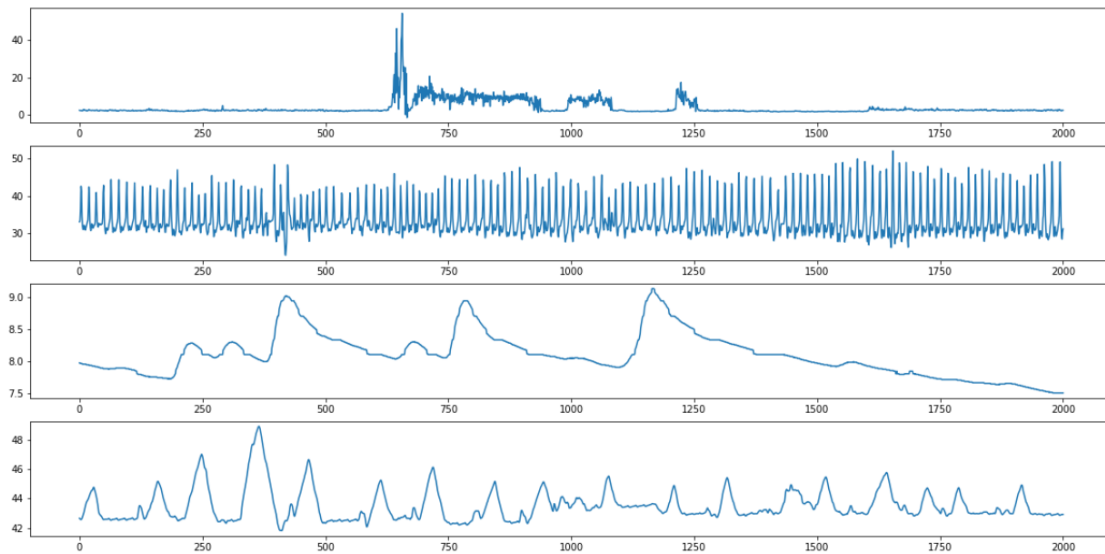
Hình 1.7 Các tín hiệu tương ứng với trạng thái cảm xúc đau buồn



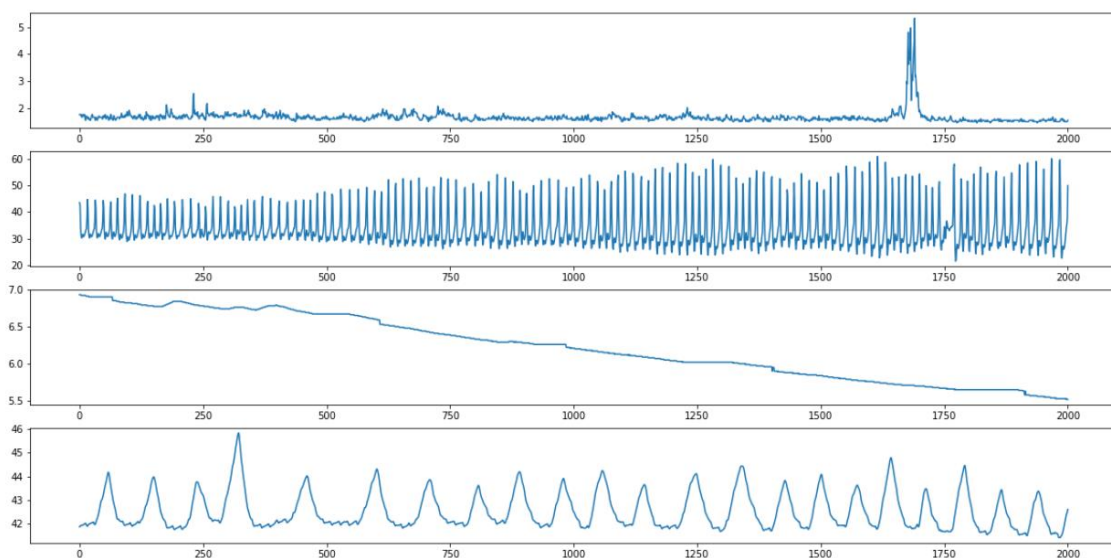
Hình 1.8 Các tín hiệu tương ứng với trạng thái cảm xúc đồng cảm



Hình 1.9 Các tín hiệu tương ứng với trạng thái cảm xúc luyến tiếc



Hình 1.10 Các tín hiệu tương ứng với trạng thái cảm xúc vui vẻ



Hình 1.11 Các tín hiệu tương ứng với trạng thái cảm xúc buồn

Như bước tiếp cận ban đầu, xét 2 trạng thái cảm xúc trung tính và tức giận để quan sát thể hiện của các tín hiệu sinh học:

- Tín hiệu điện cơ đồ: với trạng thái cảm xúc trung tính, biên độ dao động của tín hiệu này là thấp hơn (biên độ cực đại là 25) so với trạng thái cảm xúc

tức giận (biên độ cực đại > 60), đồng thời số lần đạt giá trị đỉnh của cảm xúc tức giận nhiều hơn so với cảm xúc trung tính

- Độ điện dẫn da: giá trị tại các đỉnh của trạng thái cảm xúc tức giận có xu hướng giảm dần, trong khi với trạng thái cảm xúc trung tính giá trị này chênh lệch nhỏ.
- Xung thể tích máu: biên độ dao động của trạng thái cảm xúc trung tính thấp hơn (từ 3 đến 4) và có chiều hướng đi xuống, trong khi đó trạng thái cảm xúc tức giận có biên độ dao động từ 4,5 đến 7.
- Cảm biến hô hấp: biên độ dao động của trạng thái cảm xúc trung tính từ 43 đến 47 và có nhiều lần giá trị đạt đỉnh, trong khi đó trạng thái cảm xúc tức giận có biên độ dao động từ 42 đến 50.

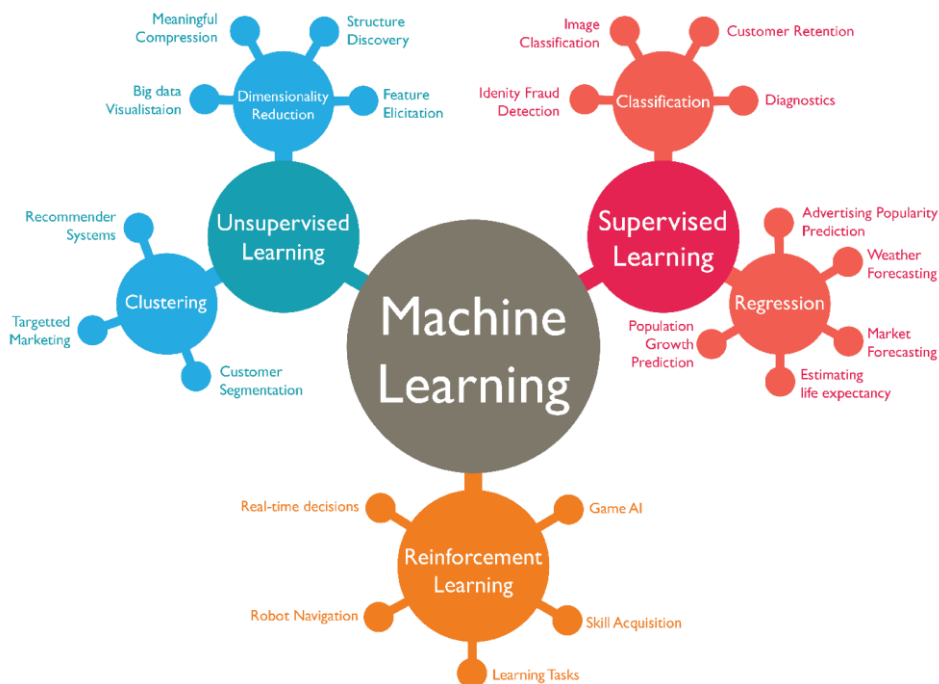
Tương tự, giữa các trạng thái cảm xúc có sự khác nhau về biên độ dao động, số đỉnh hay xu hướng đi lên hoặc xuống,... qua đó cho thấy tính khả thi trong việc phân loại cảm xúc sử dụng bộ dữ liệu sinh học của MIT, từ các đặc trưng tín hiệu, ngay cả trong miền thời gian.

CHƯƠNG 2. HỌC MÁY VÀ ỨNG DỤNG TRONG NHẬN DẠNG CẢM XÚC

• Khái lược về học máy

Học máy (Machine learning - ML) là một phần của trí tuệ nhân tạo. Học máy tập trung vào việc xây dựng các ứng dụng học từ dữ liệu và cải thiện độ chính xác của chúng theo thời gian. Quá trình học bắt đầu từ dữ liệu, mục đích chính là cho phép các máy tính tự động học mà không cần sự can thiệp hay trợ giúp của con người và tự điều chỉnh các hành động tương ứng. Trong học máy, các thuật toán được huấn luyện để tìm ra các mẫu và đặc trưng trong một lượng lớn dữ liệu để đưa ra quyết định và dự đoán trên dữ liệu mới. Thuật toán càng tốt thì các quyết định và dự đoán càng chính xác khi xử lý dữ liệu mới, bên ngoài tập dữ liệu đào tạo.

Ngày nay, học máy được ứng dụng rất nhiều trong đời sống như hệ thống gợi ý sản phẩm dựa trên thói quen người dùng, xử lý ngôn ngữ tự nhiên, phát hiện thư rác giúp ngăn các email không mong muốn, hệ thống phân tích hình ảnh y tế giúp phát hiện khối u,... Các ứng dụng của học máy được mô tả trong Hình 2.1



Hình 2.1 Học máy và ứng dụng của học máy [17]

Nhìn chung, việc xây dựng một mô hình học máy thường trải qua 4 bước sau:

- **Chuẩn bị tập dữ liệu huấn luyện:** Dữ liệu huấn luyện là tập dữ liệu đại diện cho dữ liệu mà mô hình học máy sẽ đưa vào để giải quyết vấn đề. Trong một số trường hợp, dữ liệu huấn luyện được gắn nhãn để nhận biết đặc trưng dữ liệu và phân loại mô hình cần sử dụng. Dữ liệu khác không được gắn nhãn và mô hình sẽ cần trích xuất các đặc trưng đó và xác định phân loại của riêng nó. Dữ liệu được chia thành hai tập: tập huấn luyện, được sử dụng để huấn luyện cho mô hình và tập đánh giá, được sử dụng để kiểm tra và tinh chỉnh.

- *Lựa chọn thuật toán để chạy trên tập dữ liệu huấn luyện*: Thuật toán được sử dụng phụ thuộc vào loại dữ liệu (được gắn nhãn hoặc không được gắn nhãn); lượng dữ liệu trong tập dữ liệu huấn luyện và vấn đề cần giải quyết. Các thuật toán học máy phổ biến được sử dụng cho dữ liệu được gắn nhãn có thể kể tới là thuật toán hồi quy tuyến tính, cây quyết định,... hay các thuật toán học máy phổ biến được sử dụng cho dữ liệu không được gắn nhãn như thuật toán phân cụm K-means, mạng nơ-ron, ...
- *Huấn luyện thuật toán để tạo ra mô hình*: Là một quá trình lặp đi lặp lại việc chạy các biến thông qua thuật toán, so sánh kết quả đầu ra với kết quả mà nó đáng lẽ phải tạo ra, điều chỉnh trọng số và độ lệch trong thuật toán để có thể mang lại kết quả chính xác hơn và chạy lại các biến cho đến khi thuật toán trả về kết quả chính xác trong hầu hết thời gian.
- *Ứng dụng và cải tiến mô hình*: Sử dụng mô hình với dữ liệu mới để mô hình cải thiện độ chính xác và hiệu quả theo thời gian. Dữ liệu mới đến từ đâu sẽ phụ thuộc vào vấn đề đang được giải quyết.

Học máy được chia thành ba loại chính:

+ **Học có giám sát**: Mô hình sẽ được huấn luyện trên tập dữ liệu được gắn nhãn. Tức là, dữ liệu gắn nhãn được sử dụng để huấn luyện mô hình học máy đưa ra các dự đoán phân loại, và khi dự đoán sai thì mô hình sẽ được tinh chỉnh lại cho chính xác hơn. Học máy có giám sát yêu cầu ít dữ liệu huấn luyện hơn so với các phương pháp học máy khác và giúp huấn luyện dễ dàng hơn vì kết quả của mô hình có thể được so sánh với kết quả được dán nhãn thực tế. Tuy nhiên, dữ liệu được gắn nhãn cần được chuẩn bị kỹ lưỡng và trong trường hợp không lý tưởng sẽ tạo ra một mô hình quá chặt chẽ với bộ dữ liệu huấn luyện và làm mô hình không thể xử lý chính xác với các dữ liệu mới.

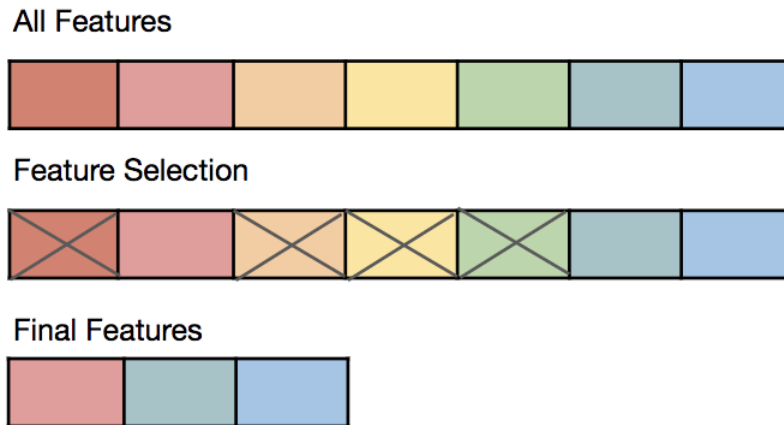
+ **Học không giám sát**: Học máy không giám sát đầu vào là từ lượng lớn dữ liệu không được gắn nhãn, sử dụng các thuật toán để trích xuất các đặc trưng có ý nghĩa nhằm gắn nhãn, sắp xếp và phân loại dữ liệu mà không cần sự can thiệp của con người. Học không giám sát không phải là việc tự động hóa các quyết định và dự đoán, mà thiên về việc xác định các mẫu và mối quan hệ trong dữ liệu.

+ **Học nửa giám sát**: Là sự kết hợp giữa học tập có giám sát và học tập không giám sát. Trong quá trình huấn luyện, thuật toán sử dụng một tập dữ liệu nhỏ có gắn nhãn để phân loại và trích xuất đặc trưng từ một tập dữ liệu lớn hơn, không được gắn nhãn. Học nửa giám sát có thể giải quyết vấn đề không có đủ dữ liệu được gắn nhãn (hoặc không đủ khả năng gắn nhãn đủ dữ liệu) để huấn luyện thuật toán học có giám sát.

2.1 Lựa chọn đặc trưng

Phương pháp lựa chọn đặc trưng dùng để giảm số lượng đặc trưng đầu vào trước khi đưa vào mô hình dự đoán của học máy. Các mô hình dự đoán có quá nhiều đầu vào sẽ làm tiêu tốn bộ nhớ cũng như thời gian huấn luyện, đồng thời các

đặc trưng không cần thiết cũng làm nhiều khả năng dự đoán của mô hình. Việc lựa chọn đặc trưng này nhằm đến việc giảm thiểu thời gian tính toán của mô hình dự đoán, và trong một số trường hợp sẽ làm tăng chất lượng dự đoán của toàn bộ mô hình.



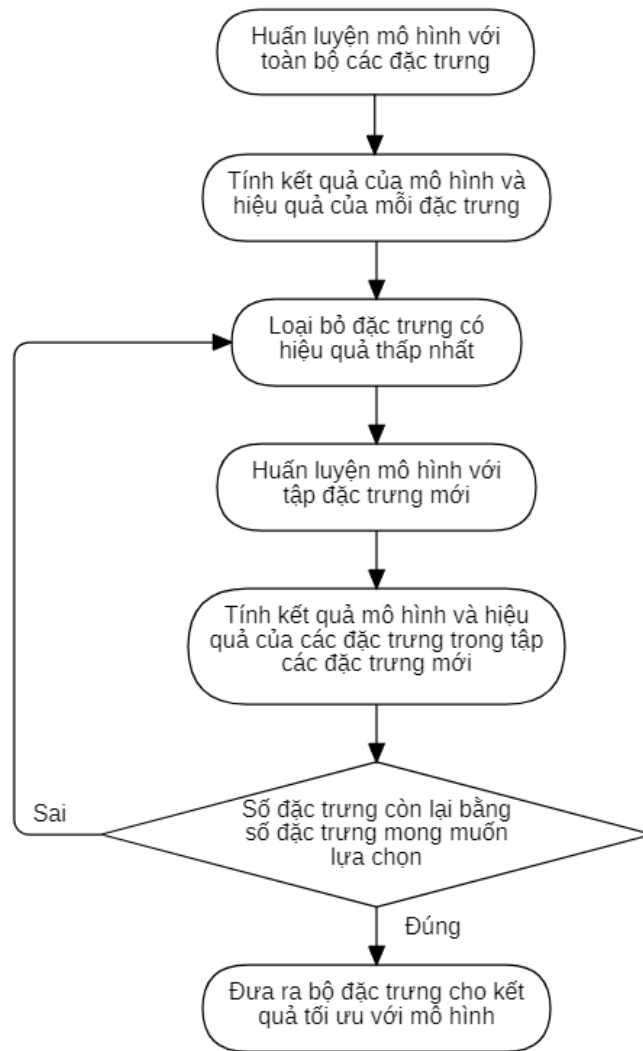
Hình 2.2 Lựa chọn đặc trưng

Các phương pháp lựa chọn sẽ đánh giá sự ảnh hưởng của các đặc trưng đến giá trị dự đoán, sau đó lựa chọn những đặc trưng có ảnh hưởng lớn nhất. Tuy nhiên việc lựa chọn phương pháp phù hợp cũng cần nghiên cứu kỹ lưỡng cho từng yêu cầu bài toán đặt ra. Có 3 cách lựa chọn đặc trưng chính bao gồm:

- Tìm kiếm tập hợp các đặc trưng con tốt nhất từ tập các đặc trưng ban đầu. Thuật toán điển hình của phương pháp này là RFE (Recursive Feature Elimination).
- Tính toán sự tương đồng cũng như ảnh hưởng của từng đặc trưng với giá trị cần đạt được, sau đó lựa chọn ra những đặc trưng có sự tương đồng cũng như ảnh hưởng lớn nhất. Các thuật toán đặc trưng của phương pháp này có thể kể đến như Chi-Squared hay Pearson Correlation.
- Sử dụng các thuật toán mà ở đó có sẵn phân lựa chọn đặc trưng, như Lasso hay Random Forrest

2.1.1 Recursive Feature Elimination (RFE)

RFE là thuật toán lựa chọn đặc trưng kiểu đóng gói, bên trong bao gồm một giải thuật được sử dụng để đánh giá. Thuật toán RFE hoạt động bắt đầu với tất cả các đặc trưng của dữ liệu, đánh giá mức độ quan trọng của từng đặc trưng, sau đó dần loại bỏ các đặc trưng ít ảnh hưởng tới kết quả cần đạt nhất, lặp lại quá trình trên cho tới khi tìm được kết quả tốt nhất hoặc chọn ra được số lượng đặc trưng mong muốn. Hình 2.3 mô tả sơ đồ hoạt động của thuật toán RFE.



Hình 2.3 Sơ đồ hoạt động thuật toán RFE

2.1.2 Pearson Correlation và Chi-Squared

Đây là 2 phương pháp điển hình của lớp các phương pháp trích chọn đặc trưng thứ 2, dựa trên tính toán sự ảnh hưởng của từng đặc trưng đến kết quả cần đạt, bằng công thức Pearson Correlation (2.1), hoặc Chi-Squared (2.2)

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad (2.1)$$

trong đó \bar{x}, \bar{y} là giá trị trung bình tương ứng của x và y

$$X_c^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i} \quad (2.2)$$

với O_i là giá trị nhận được và E_i là giá trị mong muốn của mẫu thứ i

Sau khi tính toán, thuật toán sẽ lựa chọn những đặc trưng có mức độ quan trọng nhất đến kết quả cần đạt được.

2.2 Phân lớp và ra quyết định nhận dạng

2.2.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) [18] là thuật toán học máy có giám sát được sử dụng trong bài toán phân loại. Thuật toán SVM tìm kiếm 1 siêu phẳng (hyper-plane) giúp phân tách các điểm dữ liệu của các lớp khác nhau. Trong bài toán phân loại gồm 2 lớp, tập huấn luyện bao gồm m điểm dữ liệu sao cho:

$$p = (x_i, y_i) | x_i \in R^p, y_i \in (-1, 1)_{i=1}^m \quad (2.3)$$

Trong đó x là 1 vector có số chiều là p , y là nhãn dán của dữ liệu, giá trị là -1 hoặc 1. Siêu phẳng được viết dưới dạng: $w \cdot x + b = 0$. Khi đó khoảng cách từ một điểm bất kì tới siêu phẳng đó là:

$$\frac{y_i(w \cdot x + b)}{\|w\|_2} \quad (2.4)$$

Với $\|w\|_2 = \sqrt{\sum_{i=1}^d w_i^2}$, d là số chiều của không gian

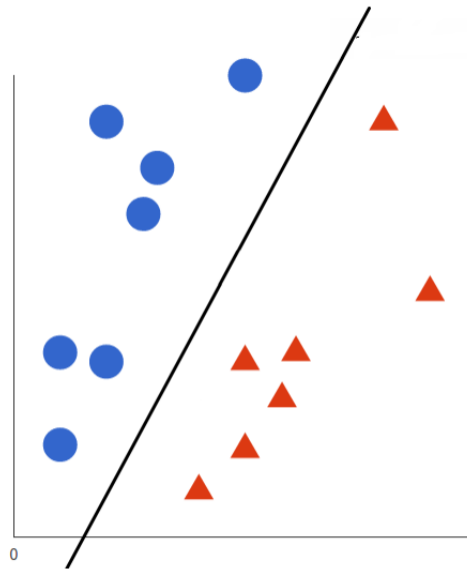
Mục tiêu của thuật toán SVM là đi tìm W và b sao cho khoảng cách từ các điểm gần nhất tới siêu phẳng có giá trị lớn nhất. Trong Hình 2.4, đường thẳng chia cắt dữ liệu của 2 lớp và khoảng cách từ 2 điểm gần nhất mỗi lớp đến đường thẳng này là lớn nhất có thể. Hàm mục tiêu:

$$\begin{aligned} (w, b) &= \operatorname{argmax}_{w, b} \left\{ \min_n \frac{y_n(w^T x_n + b)}{\|w\|_2} \right\} \\ &= \operatorname{argmax}_{w, b} \left\{ \frac{1}{\|w\|_2} \min_n y_n(w^T x_n + b) \right\} \end{aligned} \quad (2.5)$$

Nhận xét thấy $w \cdot x_i + b > 1$ khi i thuộc lớp 1 và $w \cdot x_i + b < -1$ khi i thuộc lớp -1, và y luôn cùng dấu với $w \cdot x + b$, nên $y_i(w \cdot x + b) > 1$ với mọi $1 \leq i < m$. Bài toán tối ưu trở thành:

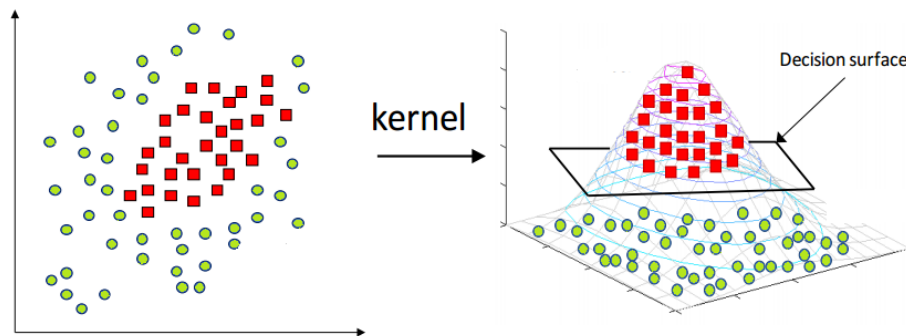
$$(w, b) = \operatorname{argmin}_{w, b} \frac{1}{2} \|w\|_2^2 \quad (2.6)$$

$$\text{với: } 1 - y_n(w^T x_n + b) \leq 0, \forall n = 1, 2, \dots, N$$



Hình 2.4 Đường thẳng phân cách điểm dữ liệu của 2 lớp [19]

Thông thường trong các bài toán phân loại, dữ liệu sẽ không thể phân tách một cách tuyến tính như trong Hình 2.4. Khi đó, thuật toán SVM sẽ sử dụng các hàm nhân để chuyển đổi dữ liệu sang không gian mới mà có thể dễ dàng phân tách dữ liệu hơn. Hình 2.5 mô tả cách thức hoạt động của SVM khi chuyển đổi dữ liệu.



Hình 2.5 Chiếu dữ liệu lên không gian nhiều chiều hơn

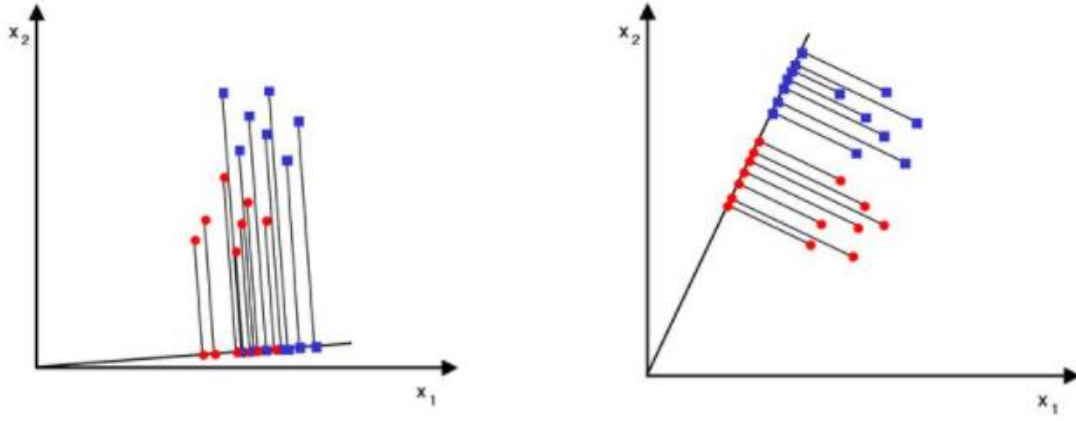
Hàm nhân giúp chuyển đổi dữ liệu được định nghĩa:

$$K(x_i, x_j) = \theta(x_i) * \theta(x_j) \quad (2.7)$$

Trong đó $\theta(x)$ là hàm chiếu vector dữ liệu x lên không gian mới. Các hàm chiếu phổ biến thường gặp gồm: polynomial, chi-square, RBF, ...

2.2.2 Linear Discriminant Analysis (LDA)

Thuật toán Linear Discriminant Analysis là thuật toán giảm chiều dữ liệu cho bài toán phân loại. Số chiều của dữ liệu mới sẽ có giá trị tối đa là $C-1$ với C là số lượng nhóm phân loại



Hình 2.6 Phép chiếu lên đường thẳng [20]

Trong Hình 2.6, dữ liệu được chiếu lên 2 đường thẳng khác nhau là a (bên trái) và b (bên phải). Khi chiếu lên đường thẳng a, việc phân loại 2 lớp sẽ khó khăn vì các điểm của 2 lớp chồng lấn lên nhau. Trong khi đó, khi chiếu xuống đường thẳng b, các điểm dữ liệu của mỗi lớp được phân tách riêng biệt, bài toán phân loại trong trường hợp này trở nên thuận tiện hơn, do vậy đường thẳng b là đường thẳng lý tưởng cần tìm trong bài toán phân loại áp dụng thuật toán LDA này.

Mục tiêu của thuật toán LDA là đi tìm giá trị lớn nhất của tỉ lệ giữa độ lệch chuẩn giữa mỗi lớp và độ lệch chuẩn giữa các điểm dữ liệu của mỗi lớp. Phép chiếu dữ liệu của dữ liệu X lên không gian Y mới được định nghĩa:

$$Y = \{y_{ik} = w^T x_{ik} | i = (1, \dots, c); k = (1, \dots, n_i)\} \quad (2.8)$$

Giá trị trung bình của lớp là $\mu_k = \frac{1}{N_k} \sum_{n \in C_k} x_n$ với N là số lượng điểm của lớp đó

Độ lệch chuẩn giữa các điểm trong 1 lớp được định nghĩa:

$$S_k^2 = \sum_{n \in C_k} (y_n - \mu_k)^2 \quad (2.9)$$

Ta có công thức ma trận hiệp phương sai giữa các lớp:

$$S_W^y = \sum_{i=1}^c \sum_{k=1}^{n_i} (y_{ik} - \mu_i)(y_{ik} - \mu_i)^T \quad (2.10)$$

Và ma trận hiệp phương sai giữa các dữ liệu trong một lớp:

$$S_B^y = \sum_{i=1}^c (\mu_i - \mu)(\mu_i - \mu)^T \quad (2.11)$$

Từ đó ta có:

$$S_W^y = w^T S_W^x w \quad (2.12)$$

$$S_B^y = w^T S_B^x w \quad (2.13)$$

Với S_W^x và S_B^x được tính theo công thức:

$$S_W^x = \sum_{i=1}^c \sum_{k=1}^{n_i} (x_{ik} - \mu_i^{(x)})(x_{ik} - \mu_i^{(x)})^T \quad (2.14)$$

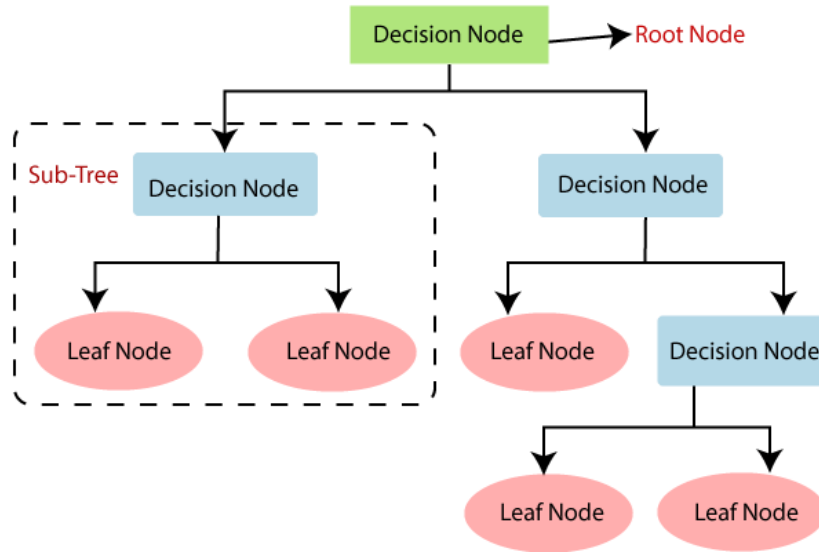
$$S_W^x = \sum_{i=1}^c \sum_{k=1}^{n_i} (x_{ik} - \mu_i^{(x)})(x_{ik} - \mu_i^{(x)})^T \quad (2.15)$$

Hàm mục tiêu của thuật toán LDA có công thức:

$$w^* = \operatorname{argmax}_w \frac{\operatorname{trace}(S_B^y)}{\operatorname{trace}(S_W^y)} = \operatorname{argmax}_w \frac{\operatorname{trace}(w^T S_B^x w)}{\operatorname{trace}(w^T S_W^x w)} \quad (2.16)$$

2.2.3 Decision Tree (DT)

Decision Tree (DT) [21] là mô hình huấn luyện có giám sát có thể dùng cho bài toán phân loại. Decision Tree tạo ra mô hình dự đoán bằng việc đưa ra các lựa chọn ở mỗi điểm trong mô hình. Mô hình sẽ bắt đầu ở rễ của cây, sau đó phụ thuộc vào giá trị các đặc trưng và đi vào các nhánh bên dưới, cuối cùng là đến các lá là nơi đưa ra giá trị dự đoán. Hình 2.7 mô tả cấu tạo của một mô hình Decision Tree.



Hình 2.7 Mô hình Decision Tree [22]

Thuật toán ID3 là thuật toán xây dựng lên Decision Tree sử dụng công thức Entropy. Cho phân phối xác suất của biến rời rạc x có thể nhận các giá trị x_1, x_2, \dots, x_n . Khi đó xác suất x nhận các giá trị đó là $p_i = p(x = x_i)$ với $0 \leq p_i \leq 1$ và $\sum_{i=1}^n p_i = 1$. Entropy của phân phối này được định nghĩa:

$$H(p) = - \sum_{i=1}^n p_i \log(p_i) \quad (2.17)$$

Trong bài toán phân loại C lớp khác nhau, ở một node bất kì không phải là lá, tập các dữ liệu S có N phần tử, mỗi lớp có N_c phần tử, entropy tại node này là:

$$H(S) = - \sum_{c=1}^C \frac{N_c}{N} \log \left(\frac{N_c}{N} \right) \quad (2.18)$$

Với thuộc tính x , các điểm dữ liệu chia thành K node con S_1, S_2, \dots, S_K với số lượng điểm dữ liệu là m_1, m_2, \dots, m_K . Ta có:

$$H(x, S) = \sum_{k=1}^K \frac{m_k}{N} H(S_k) \quad (2.19)$$

là tổng trọng số entropy mỗi node con. Định nghĩa thông tin mang lại của thuộc tính x :

$$G(x, S) = H(S) - H(x, S) \quad (2.20)$$

Bài toán tại mỗi node trở thành:

$$x^* = \underset{x}{\operatorname{argmax}} G(x, S) = \underset{x}{\operatorname{argmin}} H(x, S) \quad (2.21)$$

Ngoài thuật toán ID3 thì thuật toán CART (Classification And Regression Tree) cũng được dùng trong mô hình Decision Tree sử dụng công thức Gini:

$$Gini = 1 - \sum_{i=1}^C (p_i)^2 \quad (2.22)$$

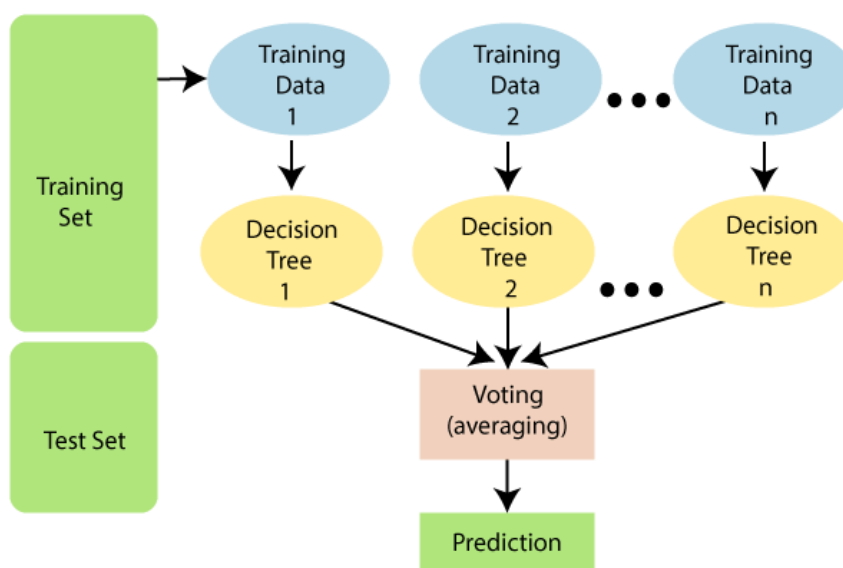
Mô hình Decision Tree có ưu điểm dễ sử dụng, tuy nhiên rất dễ bị tình trạng overfitting dữ liệu.

2.2.4 Random Forest (RF)

Random Forest (RF) là thuật toán phân loại chứa các Decision Tree ứng với 1 tập các dữ liệu huấn luyện đầu vào, sau đó lấy dự đoán được chọn nhiều nhất để làm dự đoán cuối cùng. Số lượng cây sử dụng càng lớn sẽ cho độ chính xác càng cao và tránh được vấn đề overfitting thường gặp khi sử dụng Decision Tree. Hình 2.8 mô tả cách hoạt động của mô hình Random Forest.

Các bước thực hiện khi xây dựng một mô hình Random Forest bao gồm:

- Chọn K dữ liệu từ tập huấn luyện
- Xây dựng mô hình Decision Tree với K dữ liệu đó
- Lặp lại 2 bước trên cho đến khi đạt được số lượng Decision Tree mong muốn
- Khi dự đoán dữ liệu mới, sử dụng các Decision Tree đã xây dựng để dự đoán, dự đoán nào có số lần được dự đoán nhiều nhất là kết quả dự đoán cuối cùng



Hình 2.8 Mô hình Random Forest [23]

So với mô hình Decision Tree, mô hình Random Forest cho kết quả cao hơn, tránh được tình trạng overfitting dữ liệu của Decision Tree. Tuy nhiên thời gian xây dựng mô hình Random Forest lại lâu hơn do cần phải tạo ra nhiều Decision Tree, và không thích hợp khi dữ liệu có quá nhiều đặc trưng.

2.3 Áp dụng với bộ dữ liệu MIT

2.3.1 Các bước thực hiện

Để áp dụng các thuật toán phân lớp để nhận dạng lên bộ dữ liệu MIT, cần tiến hành trải qua các bước:

- *Phân chia dữ liệu*

Bộ dữ liệu nhận dạng cảm xúc của MIT bao gồm 20 ngày đo, dữ liệu mỗi ngày chứa 8 loại cảm xúc, mỗi cảm xúc có 4 đoạn giá trị đo trong 100 giây (tương ứng với 2000 điểm dữ liệu) đo từ 4 cảm biến.

Dữ liệu mỗi ngày của từng loại cảm biến sẽ được chia thành các đoạn ứng với 10 giây tách rời (gồm 200 điểm dữ liệu). Một mẫu sẽ bao gồm 4 đoạn dữ liệu 10 giây tách rời của 4 loại cảm biến, tổng cộng 1 ngày sẽ có 80 mẫu, và tổng cộng là 1600 mẫu cho 20 ngày đo.

Trường hợp tổ chức dữ liệu đầu vào với 10 giây có chồng chập thì lượng mẫu sẽ lớn hơn, tùy thuộc vào tỉ lệ chồng chập được sử dụng. Việc xây dựng dữ liệu có chồng chập sẽ được thực hiện, thử nghiệm và trình bày trong chương sau. Trong chương này, sẽ tập trung thử nghiệm các thuật toán phân lớp khác nhau, trên cùng bộ dữ liệu 10 giây không chồng chập. Kết quả thu được sẽ là cơ sở cho việc chọn dùng thuật toán học máy “tốt nhất” cho việc thử nghiệm nâng cao độ chính xác nhận dạng qua xử lý dữ liệu huấn luyện.

- *Tính toán các giá trị đặc trưng*

Sau khi chia tách dữ liệu, mỗi đoạn giá trị thu được cần được tính ra các đặc trưng cho mỗi đoạn. Các giá trị đặc trưng bao gồm:

+ Giá trị trung bình:

$$\mu_X = \frac{1}{N} \sum_{n=1}^N X_n \quad (2.23)$$

+ Giá trị lớn nhất: X_{\max}

+ Giá trị nhỏ nhất: X_{\min}

+ Trung vị:

$$Med(x) = \begin{cases} X \left[\frac{n}{2} \right] & \text{nếu } n \text{ chẵn} \\ \frac{(X \left[\frac{n-1}{2} \right] + X \left[\frac{n+1}{2} \right])}{2} & \text{nếu } n \text{ lẻ} \end{cases} \quad (2.24)$$

+ Độ lệch chuẩn:

$$\sigma_X = \left(\frac{1}{N-1} \sum_{n=1}^N (X_n - \mu_X)^2 \right)^{1/2} \quad (2.25)$$

+ Phương sai: σ_X^2

+ Biên độ: $X_{\max} - X_{\min}$

+ Độ lệch:

$$g_1 \frac{m_3}{m_2^{3/2}} \quad (2.26)$$

trong đó

$$m_i = \frac{1}{N} \sum_{n=1}^N (x[n] - \bar{x})^i \quad (2.27)$$

+ Chuẩn hóa tín hiệu:

$$\tilde{X}_n = \frac{X_n - \mu_X}{\sigma_X} \quad (2.28)$$

+ Giá trị chênh lệch trung bình tuyệt đối 2 tín hiệu liên tiếp của dữ liệu:

$$\delta_X = \frac{1}{N} \sum_{n=1}^{N-1} |X_{n+1} - X_n| \quad (2.29)$$

+ Giá trị chênh lệch trung bình tuyệt đối 2 tín hiệu liên tiếp của dữ liệu chuẩn hóa:

$$\tilde{\delta}_X = \frac{1}{N-1} \sum_{n=1}^{N-1} |\tilde{X}_{n+1} - \tilde{X}_n| = \frac{\delta_X}{\sigma_X} \quad (2.30)$$

+ Giá trị chênh lệch trung bình tuyệt đối 2 tín hiệu cách nhau của dữ liệu:

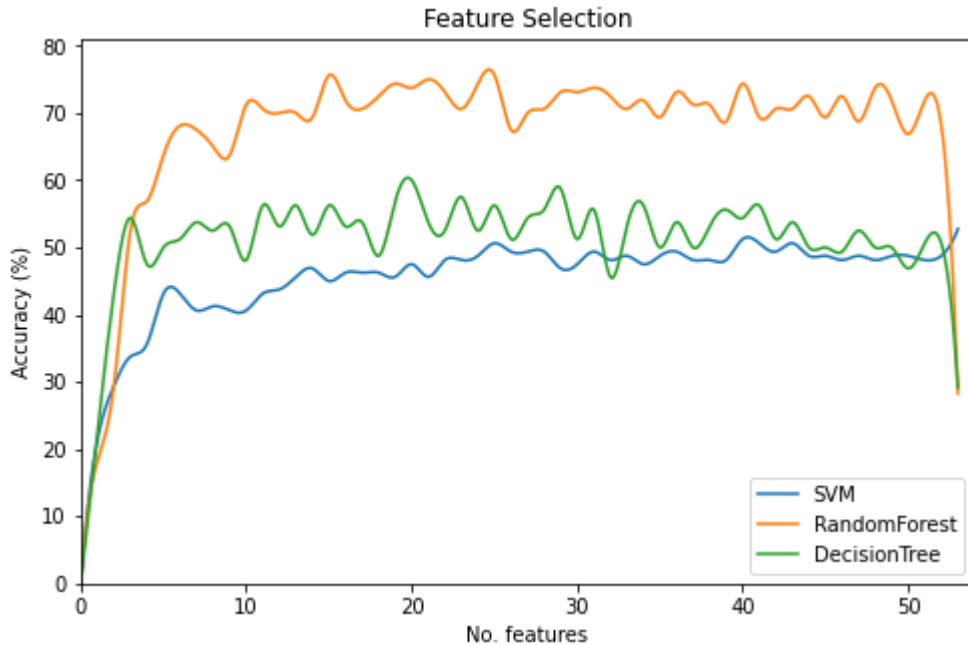
$$\gamma_X = \frac{1}{N-2} \sum_{n=1}^{N-2} |X_{n+2} - X_n| \quad (2.31)$$

- + Giá trị chênh lệch trung bình tuyệt đối 2 tín hiệu cách nhau của dữ liệu chuẩn hóa:

$$\tilde{\gamma}_X = \frac{1}{N-2} \sum_{n=1}^{N-2} |\tilde{X}_{n+2} - \tilde{X}_n| = \frac{\gamma_X}{\sigma_X} \quad (2.32)$$

- *Lựa chọn đặc trưng cho mô hình*

Khi đã tính toán các đặc trưng cho từng mẫu dữ liệu, bước tiếp theo cần phải lựa chọn số lượng và các đặc trưng nào là tốt nhất cho từng mô hình nhận dạng. Hình 2.9 thể hiện mối quan hệ giữa số lượng đặc trưng được chọn với tỉ lệ nhận dạng đúng của 3 mô hình là Random Forest, SVM và Decision Tree.



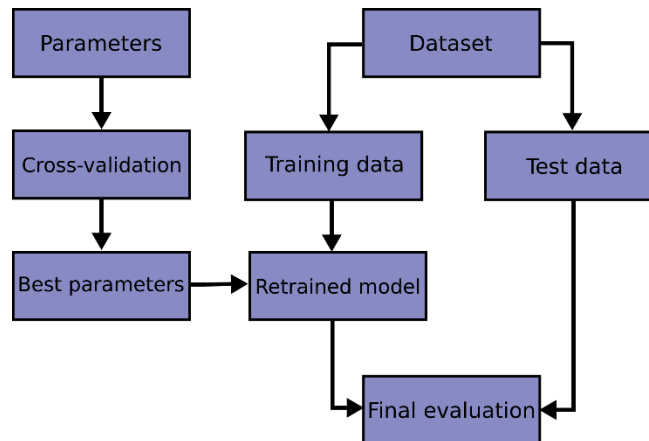
Hình 2.9 Ảnh hưởng của lựa chọn đặc trưng lên các mô hình phân lớp

Dựa vào Hình 2.9, các mô hình nhận dạng cho kết quả tốt nhất khi số lượng đặc trưng được chọn nằm trong khoảng 20 đặc trưng, lựa chọn đặc trưng quá ít hoặc quá nhiều sẽ làm giảm độ chính xác của cả 3 mô hình.

- *Hiệu chỉnh tham số cho các mô hình*

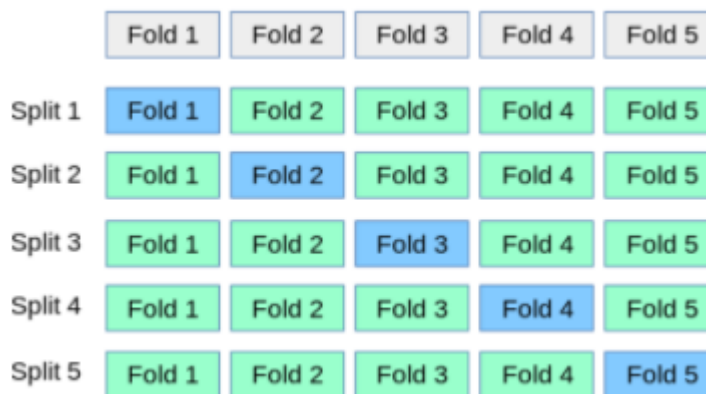
Để kết quả nhận dạng đạt được kết quả tốt hơn, các mô hình phân lớp cần được hiệu chỉnh tham số đầu vào. Thông thường, tập dữ liệu sẽ được chia làm 3 tập con: tập huấn luyện (train test), tập kiểm thử (validation set) và tập đánh giá (test set). Việc chia dữ liệu này sẽ đảm bảo được tập đánh giá là hoàn toàn độc lập, không có thông tin nào bị rò rỉ trong quá trình huấn luyện cũng như kiểm thử. Với

các giá trị tham số, mô hình sẽ được huấn luyện với tập huấn luyện, sau đó được kiểm tra trên tập kiểm thử và cuối cùng là đưa ra các tham số của mô hình cho kết quả tốt nhất. Mô hình khi đã xác định được tham số phù hợp sẽ được huấn luyện lại và đánh giá trên tập đánh giá. Trong trường hợp dữ liệu không đủ lớn để chia thành 3 tập dữ liệu khác nhau, phương pháp được sử dụng là cross-validation (Hình 2.10).



Hình 2.10 Phương pháp hiệu chỉnh tham số mô hình bằng cross-validation

Tập dữ liệu huấn luyện sẽ được chia thành n nhóm (n thường là 5 hoặc 10), mỗi lần lặp sẽ bỏ riêng ra 1 nhóm dùng để kiểm thử, và $n-1$ nhóm còn lại được đưa vào để huấn luyện mô hình. Kết quả cuối cùng là trung bình của các kết quả thu được của n lần riêng lẻ đó. Hình 2.11 mô tả cách phương pháp cross-validation chia dữ liệu khi $n = 5$.



Hình 2.11 Cách chia dữ liệu với $n = 5$

2.3.2 Thử nghiệm lựa chọn mô hình phân lớp phù hợp

Với phương pháp chia tách dữ liệu 10 giây riêng biệt, tổng cộng thu được 1600 mẫu. Trong đó, 160 mẫu (tương ứng 10%) được tách ra làm tập dữ liệu kiểm tra, 1440 mẫu còn lại được dùng để huấn luyện mô hình. Áp dụng kỹ thuật lựa chọn đặc trưng, hiệu chỉnh tham số như đã trình bày trong mục 2.3.1, sau đó sử dụng các kỹ thuật phân lớp LDA, SVM, DT và RF để đưa ra quyết định nhận dạng. Bảng 2.1 Kết quả nhận dạng cảm xúc của từng mô hình thống kê kết quả nhận dạng 8

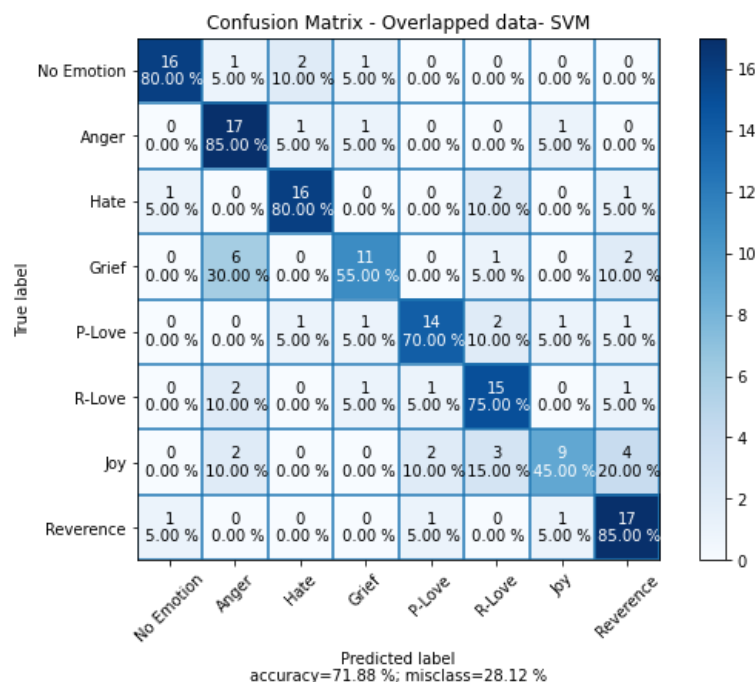
trạng thái cảm xúc trên bộ dữ liệu MIT của từng mô hình phân lớp, sử dụng dữ liệu có độ dài 10s, không chồng chập.

Bảng 2.1 Kết quả nhận dạng cảm xúc của từng mô hình

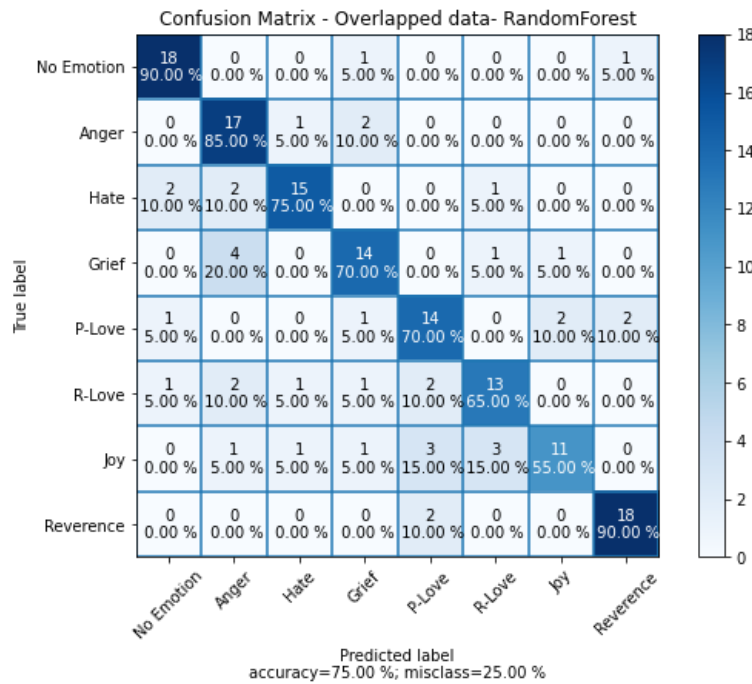
Mô hình	SVM	RF	DT	LDA
Kết quả (%)	71.88	75.00	65.62	61.88

Từ Bảng 2.1, ta nhận thấy 2 mô hình SVM và Random Forest cho tỉ lệ nhận dạng cao hơn so với 2 mô hình còn lại (lần lượt là 71,88% và 75,00%). Vì vậy, 2 mô hình này được chọn để nghiên cứu phương pháp nâng cao chất lượng nhận dạng dựa trên xử lý dữ liệu đầu vào sẽ được trình bày trong Chương 3.

Chi tiết hơn, Hình 2.12 và Hình 2.13 tổng hợp kết quả nhận dạng từng trạng thái cảm xúc khi sử dụng hai kỹ thuật SVM và Random Forest. Các kết quả trên đường chéo chính thể hiện tỉ lệ nhận dạng đúng của từng trạng thái cảm xúc. Các phần tử không thuộc đường chéo chính thể hiện tỉ lệ nhận dạng nhầm của loại cảm xúc này sang loại cảm xúc khác.



Hình 2.12 Kết quả nhận dạng bằng mô hình SVM với dữ liệu không chồng chập



Hình 2.13 Kết quả nhận dạng bằng mô hình RF với dữ liệu không chồng chập

Tỉ lệ nhận dạng đúng (trung bình cho cả 8 loại cảm xúc khác nhau) của mô hình SVM và Random Forest lần lượt là 71,88% và 75,00%. Kết quả của mô hình Random Forest tốt hơn so với mô hình SVM ($\approx 4\%$). Đối với mô hình SVM, 2 trạng thái cảm xúc giận dữ và tôn trọng cho kết quả nhận dạng tốt nhất (90%), trạng thái cảm xúc vui vẻ cho tỉ lệ nhận dạng thấp nhất (45%). Tương tự, mô hình Random Forest nhận dạng tốt nhất với 2 trạng thái cảm xúc là trung tính và tôn trọng (90%) và thấp nhất với trạng thái cảm xúc vui vẻ (55%). Cả 2 mô hình phân lớp đều chưa đạt được tỉ lệ nhận dạng cao.

Để nâng cao chất lượng nhận dạng cho một kỹ thuật phân lớp chọn trước, thì cần có dữ liệu đào tạo hợp lý hơn. Từ đó, trong chương 3 sẽ tập trung thử nghiệm giải pháp chồng chập dữ liệu đầu vào. Mục đích là giúp tăng số lượng mẫu huấn luyện, đồng thời tăng tính đồng nhất về đặc điểm của dữ liệu đặc trưng cho mỗi loại cảm xúc. Qua đó, làm giàu thông tin hữu ích cho các tập mẫu huấn luyện và thử nghiệm.

CHƯƠNG 3. TỔ CHỨC DỮ LIỆU ĐẦU VÀO NÂNG CAO CHẤT LƯỢNG NHẬN DẠNG CẢM XÚC

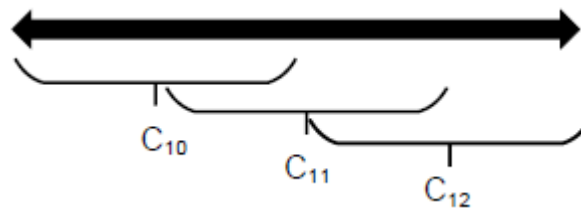
Cách chia tách dữ liệu đã được trình bày trong chương 2, trong đó chuẩn hóa các đoạn dữ liệu nhận dạng có độ dài 10s. Có hai cách tổ chức dữ liệu là 10s không chồng chập và 10s có chồng chập với các tỉ lệ khác nhau.

Hình 3.1 mô tả cách biểu diễn cách chia dữ liệu kiểu không chồng chập. Đoạn dữ liệu gốc dài 100s của MIT sẽ được cắt làm 10 đoạn nhỏ hơn, mỗi đoạn 10s liên tiếp nhau. Như vậy, mỗi đoạn dữ liệu 10s không có thông tin chung với các đoạn khác...



Hình 3.1 Chia dữ liệu khi không chồng chập

Hình 3.2 mô tả cách chia dữ liệu kiểu chồng chập. Trong đó, mỗi đoạn dữ liệu 10s chỉ có một phần khác so với đoạn liền kề nó. Với cách làm như vậy, từ bộ dữ liệu mẫu của MIT sẽ thu được bộ dữ liệu đào tạo lớn hơn nhiều, đồng thời có tính đồng nhất cao hơn khi chia tách không chồng chập.



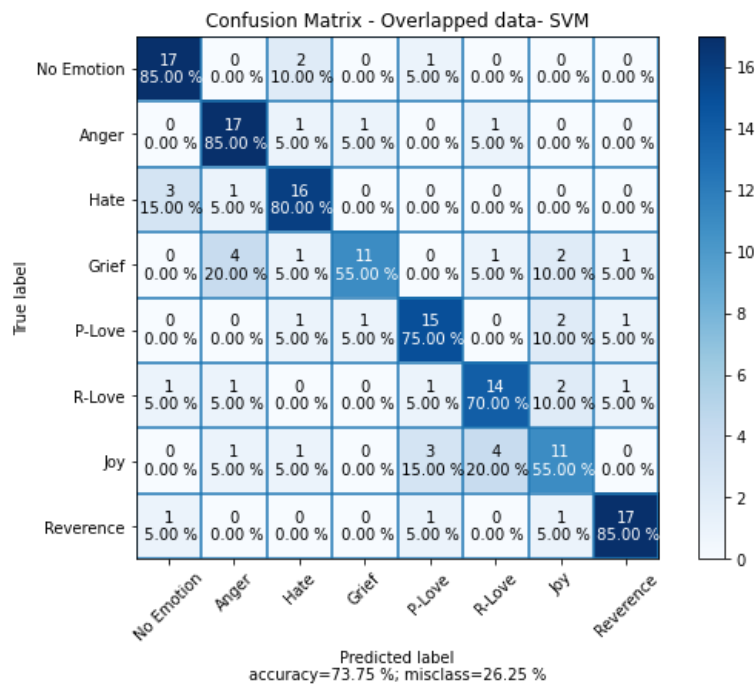
Hình 3.2 Chia dữ liệu khi áp dụng kỹ thuật chồng chập

Nội dung chương này sẽ tập chung trình bày kết quả nhận dạng cảm xúc của hai mô hình phân lớp SVM và RF, khi sử dụng dữ liệu huấn luyện và thử nghiệm được làm mới từ tập dữ liệu MIT, theo các tỉ lệ khác nhau. Việc so sánh chất lượng nhận dạng ở từng tỉ lệ chồng chập để có thể xác định được tỉ lệ chồng chập tốt nhất, hay nói cách khác là tìm được phương án tổ chức dữ liệu tối ưu nhằm nâng cao chất lượng nhận dạng cảm xúc.

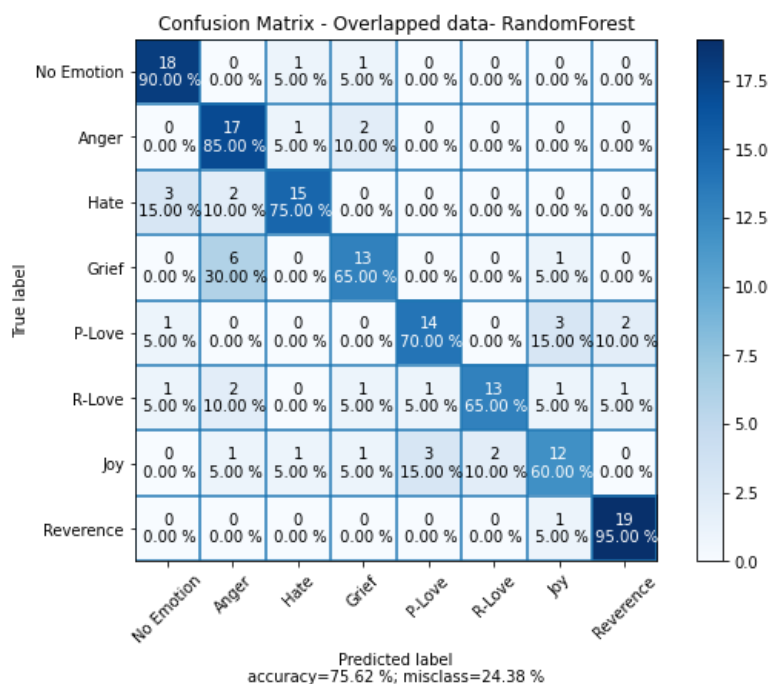
Các mục tiếp theo trong chương này sẽ trình bày kết quả nhận dạng với các tỉ lệ chồng chập từ 10% tới 90%.

3.1 Tỉ lệ chồng chập 10%

Khi sử dụng giải pháp chồng chập dữ liệu với tỉ lệ 10%, số lượng mẫu thu được tổng cộng là 1760 mẫu. Sử dụng dữ liệu đã được chồng chập đó để huấn luyện và kiểm tra của hai mô hình phân lớp SVM và Random Forest. Hình 3.3 và Hình 3.4 thể hiện kết quả nhận dạng thu được của hai kỹ thuật SVM và Random Forest trên tập dữ liệu này.



Hình 3.3 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 10%



Hình 3.4 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 10%

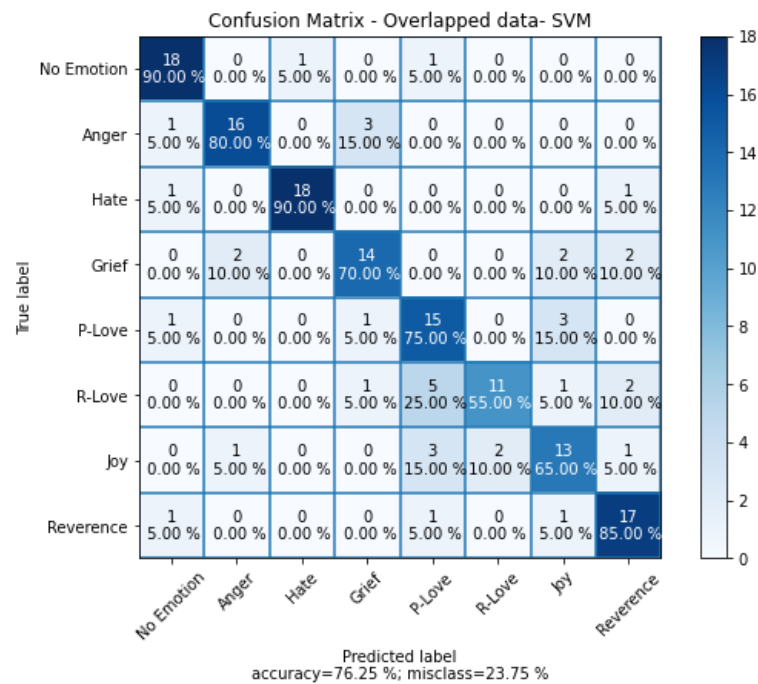
Tỉ lệ nhận dạng đúng (trung bình cho cả 8 trạng thái cảm xúc khác nhau) của mô hình SVM và Random Forest lần lượt là 73,75% và 75,62%. Đối với mô hình SVM, 3 trạng thái cảm xúc trung tính, giận dữ và tôn trọng cho kết quả nhận dạng tốt nhất (85%). Hai trạng thái cảm xúc đau buồn và vui vẻ cho tỉ lệ thấp nhất (55%). Tương tự, mô hình Random Forest cho kết quả nhận dạng tốt nhất với trạng thái cảm xúc tôn trọng (95%) và thấp nhất với trạng thái cảm xúc vui vẻ (60%).

Như vậy, với tỉ lệ chồng chập 10%, kết quả nhận dạng của cả hai kỹ thuật SVM và RF đều cao hơn khi sử dụng dữ liệu không chồng chập. Song, trung bình

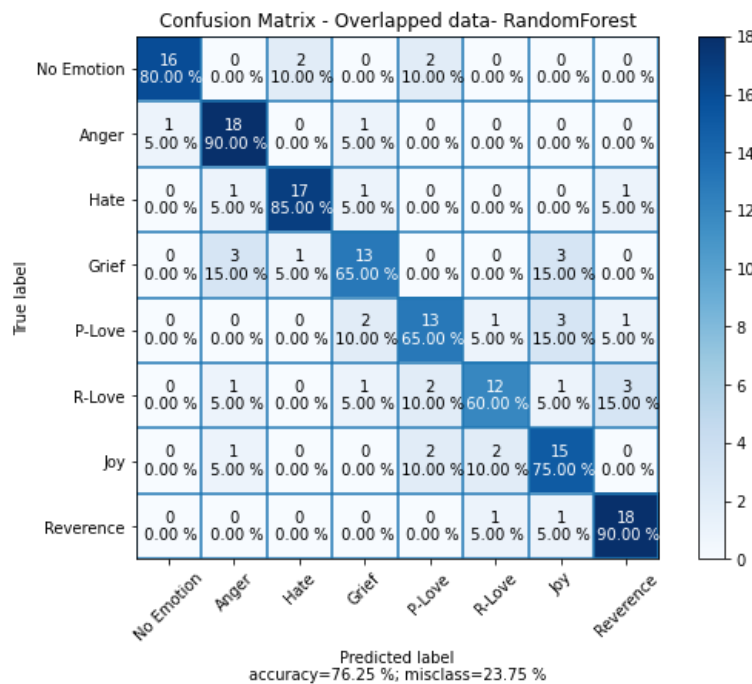
cho cả 8 trạng thái cảm xúc, kết quả nhận dạng chưa đủ tốt, khi chưa đạt độ chính xác từ 80% trở lên.

3.2 Tỷ lệ chồng chập 20%

Khi sử dụng giải pháp chồng chập dữ liệu với tỷ lệ 20%, số lượng mẫu thu được tổng cộng là 1920 mẫu. Sử dụng dữ liệu đã được chồng chập đó để huấn luyện và thử nghiệm hai mô hình phân lớp SVM và Random Forest. Hình 3.5 và Hình 3.6 biểu diễn kết quả nhận dạng thu được, sau khi sử dụng mô hình nêu trên với tập dữ liệu chồng chập 20%.



Hình 3.5 Kết quả nhận dạng bằng mô hình SVM tỷ lệ chồng chập 20%

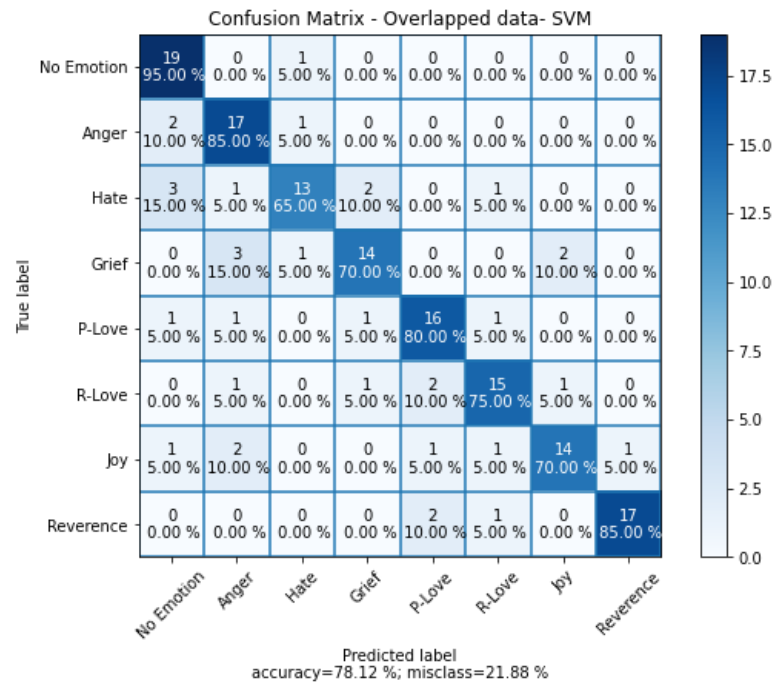


Hình 3.6 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 20%

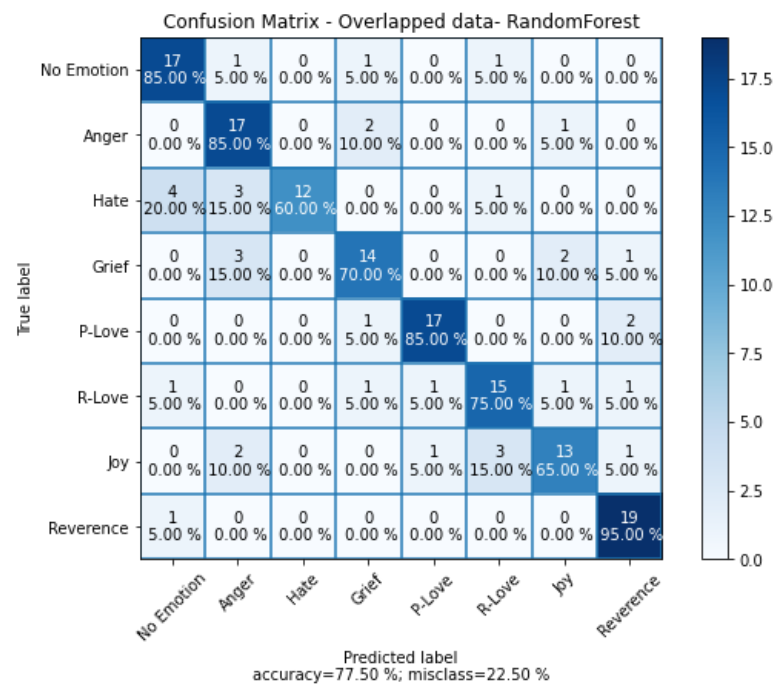
Tỉ lệ nhận dạng đúng (trung bình cho cả 8 trạng thái cảm xúc khác nhau) của mô hình SVM và Random Forest cùng là 76,25%. Đối với mô hình SVM, 2 trạng thái cảm xúc trung tính và ghét bỏ cho kết quả nhận dạng tốt nhất (90%), trạng thái cảm xúc lãng mạn cho tỉ lệ thấp nhất (55%). Tương tự, mô hình Random Forest nhận dạng tốt nhất với 2 trạng thái cảm xúc tức giận và tôn trọng (90%), thấp nhất với trạng thái cảm xúc lãng mạn (60%).

3.3 Tỉ lệ chồng chập 30%

Khi sử dụng giải pháp chồng chập dữ liệu với tỉ lệ 30%, số lượng mẫu thu được tổng cộng là 2080 mẫu. Sử dụng dữ liệu đã được chồng chập này để huấn luyện và thử nghiệm hai mô hình là SVM và Random Forest. Hình 3.7 và Hình 3.8 biểu diễn kết quả nhận dạng thu được khi sử dụng tập dữ liệu với tỉ lệ chồng chập 30%.



Hình 3.7 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chông chấp 30%

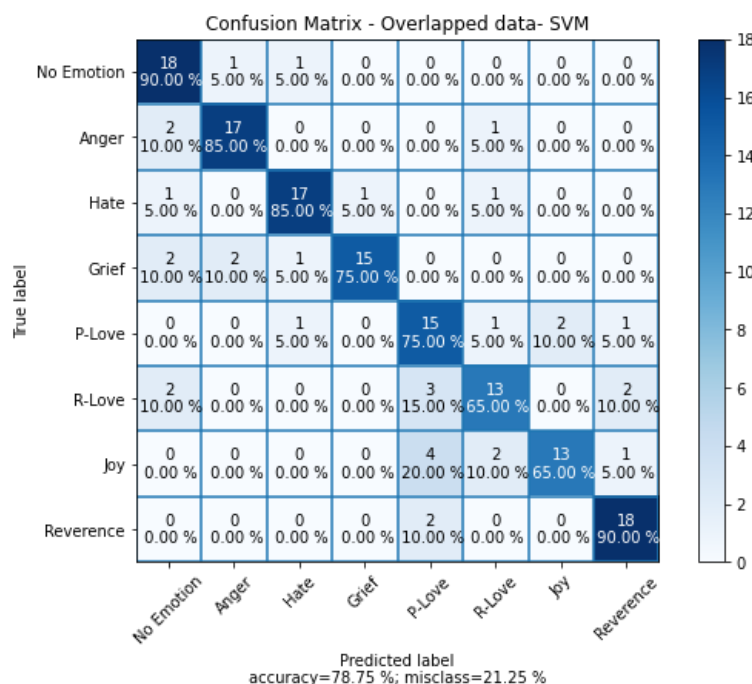


Hình 3.8 Kết quả nhận dạng bằng mô hình RF tỉ lệ chông chấp 30%

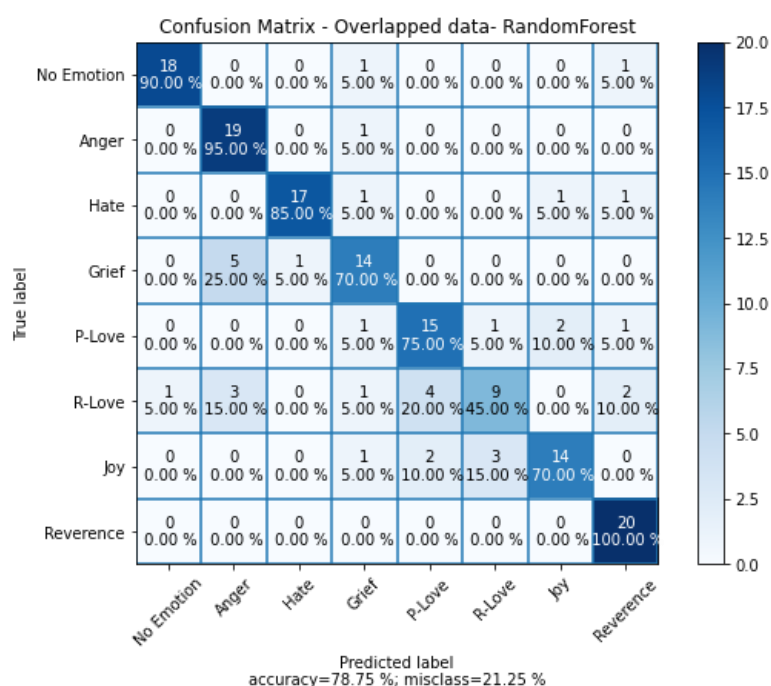
Tỉ lệ nhận dạng đúng (trung bình cho cả 8 trạng thái cảm xúc khác nhau) của mô hình SVM và Random Forest lần lượt là 78,12% và 77,5%. Đối với mô hình SVM, trạng thái cảm xúc trung tính cho kết quả nhận dạng tốt nhất (95%), trạng thái cảm xúc ghét bỏ cho tỉ lệ thấp nhất (65%). Tương tự, mô hình Random Forest nhận dạng tốt nhất với trạng thái cảm xúc tôn trọng (95%) và thấp nhất với trạng thái cảm xúc ghét bỏ (60%).

3.4 Tỷ lệ chồng chập 40%

Khi sử dụng giải pháp chồng chập dữ liệu với tỷ lệ 40%, số lượng mẫu thu được tổng cộng là 2400 mẫu. Sử dụng tập dữ liệu đã được chồng chập này để huấn luyện và thử nghiệm hai mô hình là SVM và Random Forest. Hình 3.9 và Hình 3.10 biểu diễn kết quả nhận dạng thu được khi sử dụng bộ dữ liệu với tỷ lệ chồng chập 40%.



Hình 3.9 Kết quả nhận dạng bằng mô hình SVM tỷ lệ chồng chập 40%

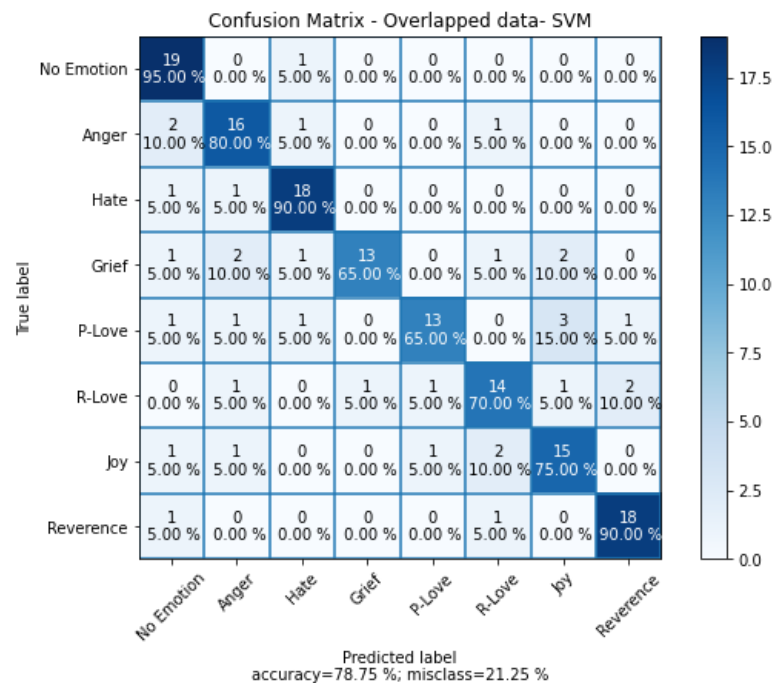


Hình 3.10 Kết quả nhận dạng bằng mô hình RF tỷ lệ chồng chập 40%

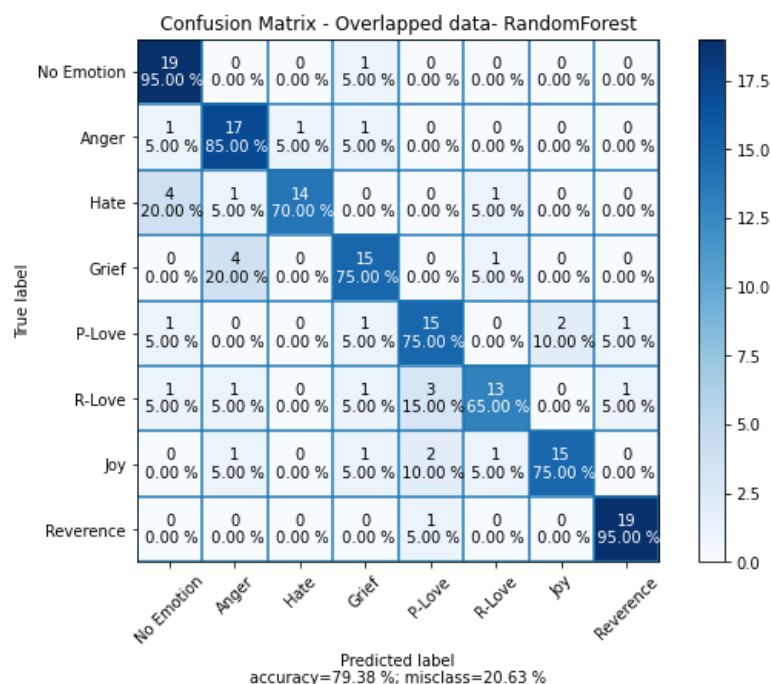
Tỉ lệ nhận dạng đúng (trung bình cho cả 8 loại cảm xúc khác nhau) của mô hình SVM và Random Forest đều là 78,75%. Đối với mô hình SVM, hai trạng thái cảm xúc trung tính và tức giận cho kết quả nhận dạng tốt nhất (90%), hai trạng thái cảm xúc lãng mạn và vui vẻ cho tỉ lệ thấp nhất (65%). Tương tự, mô hình Random Forest nhận dạng tốt nhất với trạng thái cảm xúc tôn trọng (100%) và thấp nhất với trạng thái cảm xúc lãng mạn (45%).

3.5 Tỉ lệ chồng chập 50%

Khi sử dụng giải pháp chồng chập dữ liệu với tỉ lệ 50%, số lượng mẫu thu được tổng cộng là 2880 mẫu. Sử dụng tập dữ liệu đã được chồng chập này để huấn luyện và thử nghiệm hai mô hình là SVM và Random Forest. Hình 3.11 và Hình 3.12 biểu diễn kết quả nhận dạng thu được khi sử dụng bộ dữ liệu với tỉ lệ chồng chập 50%.



Hình 3.11 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 50%

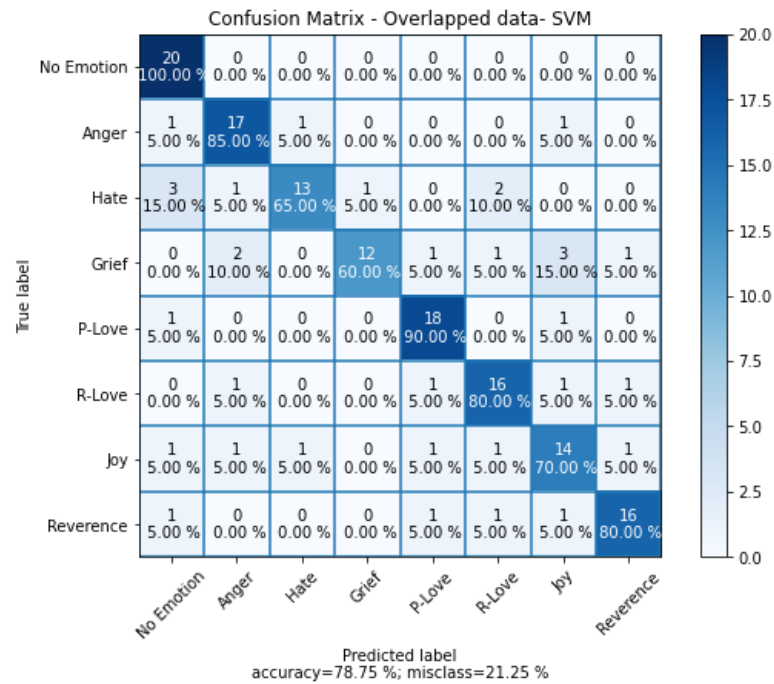


Hình 3.12 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 50%

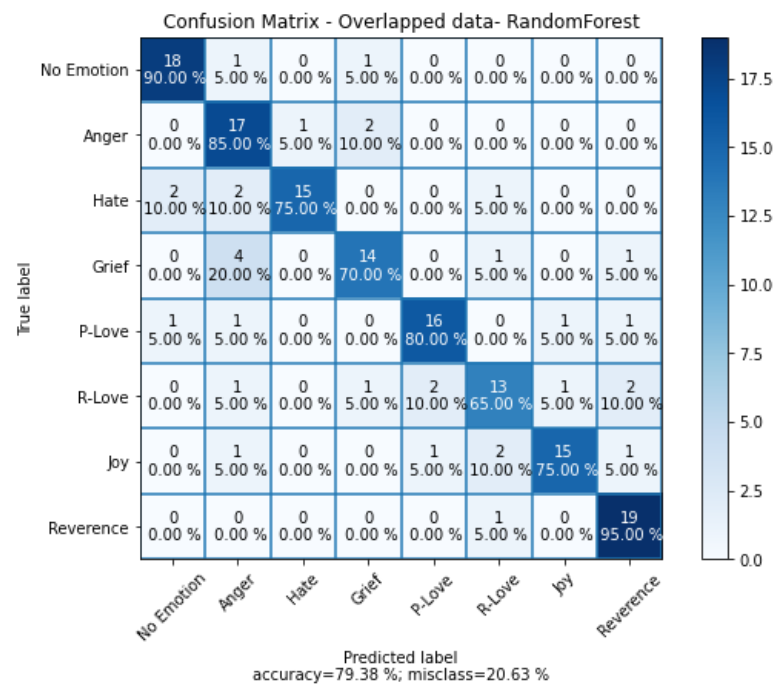
Tỉ lệ nhận dạng đúng (trung bình cho cả 8 loại cảm xúc khác nhau) của mô hình SVM và Random Forest lần lượt là 78,75% và 79,38%. Đối với mô hình SVM, trạng thái cảm xúc trung tính cho kết quả nhận dạng tốt nhất (95%), hai trạng thái cảm xúc đau buồn và đồng cảm cho tỉ lệ thấp nhất (65%). Tương tự, mô hình Random Forest nhận dạng tốt nhất với hai trạng thái cảm xúc trung tính và tôn trọng (95%) và thấp nhất với trạng thái cảm xúc lãng mạn (65%).

3.6 Tỉ lệ chồng chập 60%

Khi sử dụng giải pháp chồng chập dữ liệu với tỉ lệ 60%, số lượng mẫu thu được tổng cộng là 3200 mẫu. Sử dụng tập dữ liệu đã được chồng chập này để huấn luyện và thử nghiệm hai mô hình là SVM và Random Forest. Hình 3.13 và Hình 3.14 biểu diễn kết quả nhận dạng thu được khi sử dụng bộ dữ liệu với tỉ lệ chồng chập 60%.



Hình 3.13 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 60%

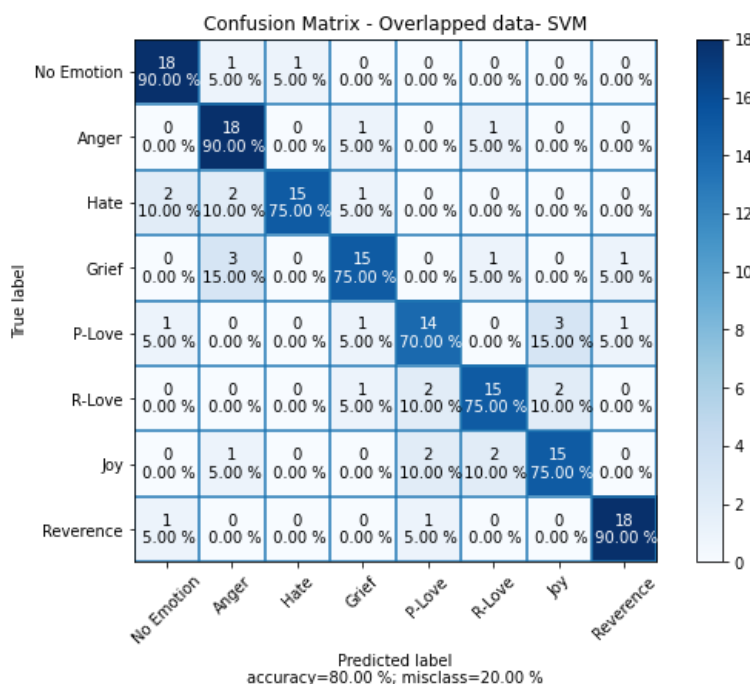


Hình 3.14 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 60%

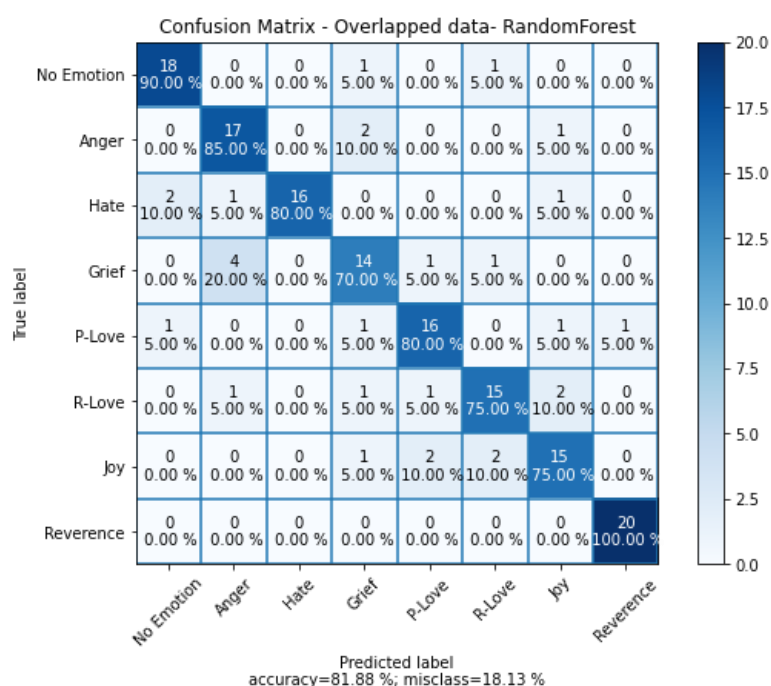
Tỉ lệ nhận dạng đúng (trung bình cho cả 8 loại cảm xúc khác nhau) của mô hình SVM và Random Forest lần lượt là 78,75% và 79,38%. Đối với mô hình SVM, trạng thái cảm xúc trung tính cho kết quả nhận dạng tốt nhất (100%) và trạng thái cảm xúc đau buồn cho tỉ lệ thấp nhất (60%). Tương tự, mô hình Random Forest nhận dạng tốt nhất với trạng thái cảm xúc tôn trọng (95%) và thấp nhất với trạng thái cảm xúc lãng mạn (65%).

3.7 Tỷ lệ chồng chập 70%

Khi sử dụng giải pháp chồng chập dữ liệu với tỷ lệ 70%, số lượng mẫu thu được tổng cộng là 4480 mẫu. Sử dụng tập dữ liệu đã được chồng chập này để huấn luyện và thử nghiệm hai mô hình là SVM và Random Forest. Hình 3.15 và Hình 3.16 biểu diễn kết quả nhận dạng thu được khi sử dụng bộ dữ liệu với tỷ lệ chồng chập 70%.



Hình 3.15 Kết quả nhận dạng bằng mô hình SVM tỷ lệ chồng chập 70%

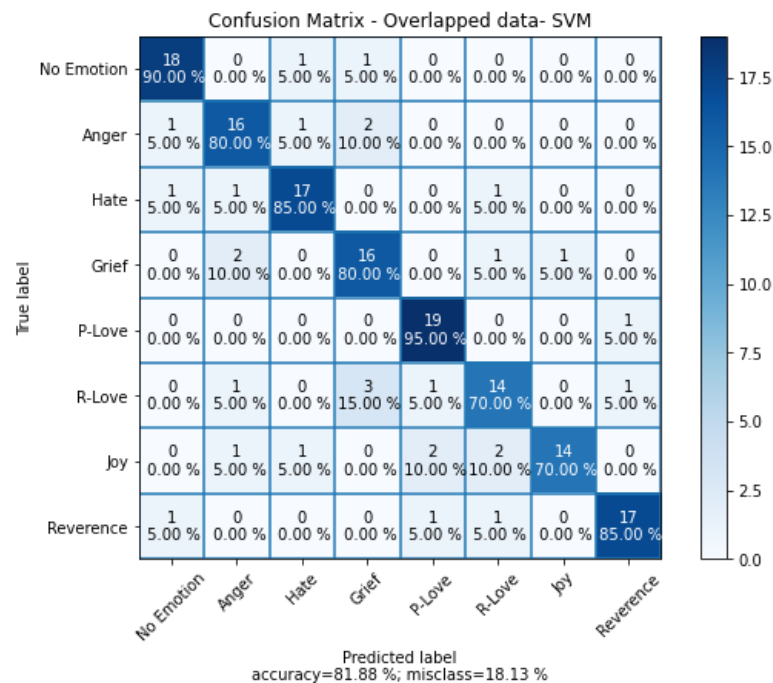


Hình 3.16 Kết quả nhận dạng bằng mô hình RF tỷ lệ chồng chập 70%

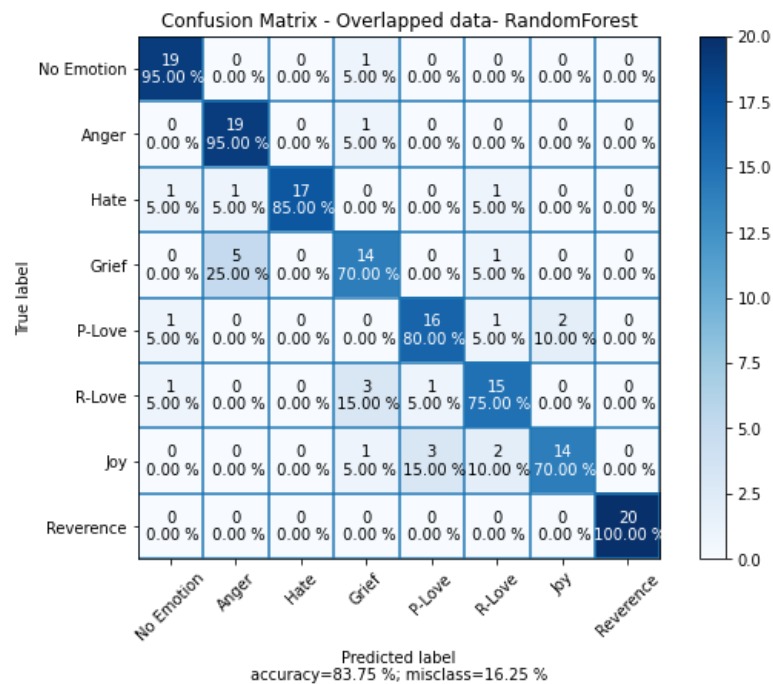
Tỉ lệ nhận dạng đúng (trung bình cho cả 8 loại cảm xúc khác nhau) của mô hình SVM và Random Forest lần lượt là 80% và 81,88%. Đối với mô hình SVM, ba trạng thái cảm xúc trung tính, tức giận và tôn trọng cho kết quả nhận dạng tốt nhất (90%), trạng thái cảm xúc lãng mạn cho tỉ lệ thấp nhất (70%). Tương tự, mô hình Random Forest nhận dạng tốt nhất với trạng thái cảm xúc tôn trọng (100%) và thấp nhất với trạng thái cảm xúc đau buồn (70%).

3.8 Tỉ lệ chồng chập 80%

Khi sử dụng giải pháp chồng chập dữ liệu với tỉ lệ 80%, số lượng mẫu thu được tổng cộng là 6720 mẫu. Sử dụng tập dữ liệu đã được chồng chập này để huấn luyện và thử nghiệm hai mô hình là SVM và Random Forest. Hình 3.17 và Hình 3.18 biểu diễn kết quả nhận dạng thu được khi sử dụng bộ dữ liệu với tỉ lệ chồng chập 80%.



Hình 3.17 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chồng chập 80%

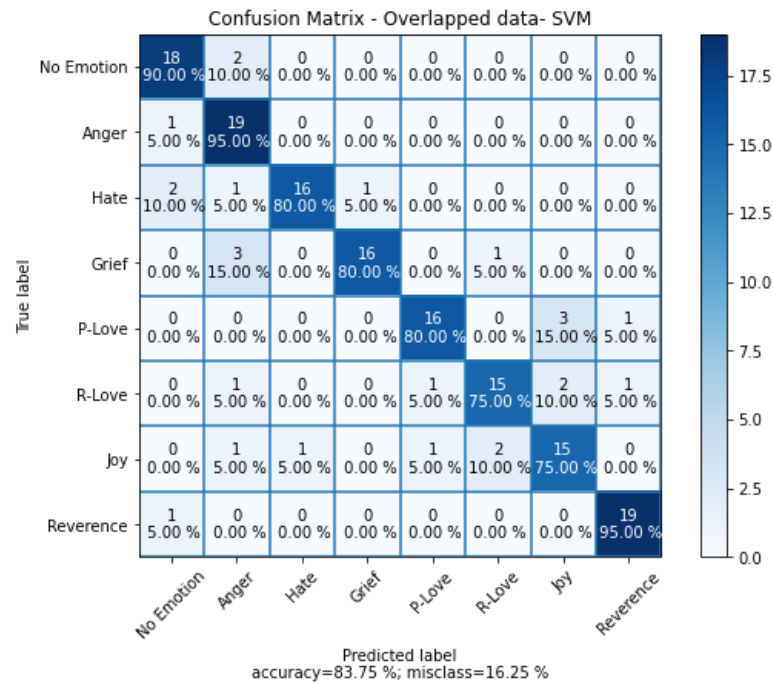


Hình 3.18 Kết quả nhận dạng bằng mô hình RF tỉ lệ chồng chập 80%

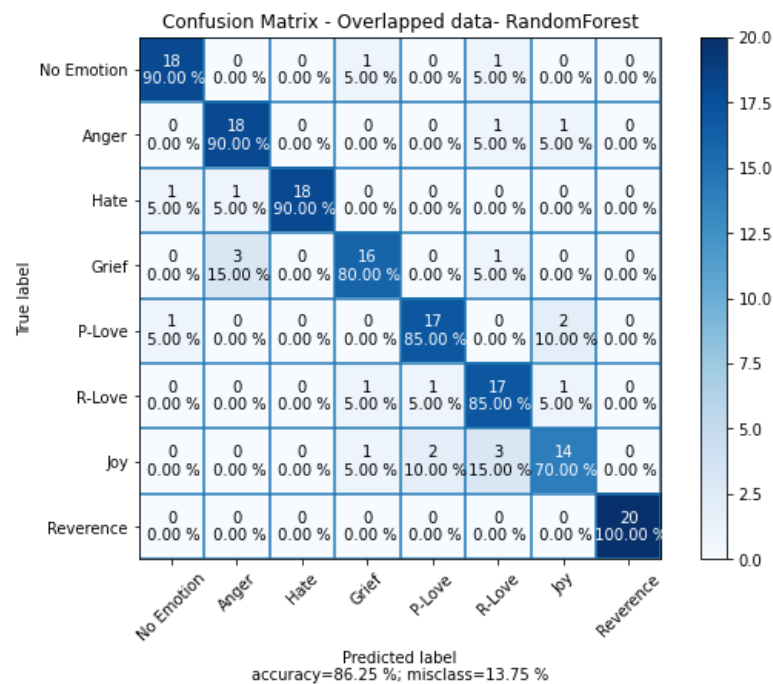
Tỉ lệ nhận dạng đúng (trung bình cho cả 8 loại cảm xúc khác nhau) của mô hình SVM và Random Forest lần lượt là 81,88% và 83,75%. Đối với mô hình SVM, trạng thái cảm xúc đồng cảm cho kết quả nhận dạng tốt nhất (95%), hai trạng thái cảm xúc lãng mạn và vui vẻ cho tỉ lệ thấp nhất (70%). Tương tự, mô hình Random Forest nhận dạng tốt nhất với trạng thái cảm xúc tôn trọng (100%) và thấp nhất với hai trạng thái cảm xúc đau buồn và vui vẻ (70%).

3.9 Tỉ lệ chồng chập 90%

Khi sử dụng giải pháp chồng chập dữ liệu với tỉ lệ 90%, số lượng mẫu thu được tổng cộng là 13120 mẫu. Sử dụng tập dữ liệu đã được chồng chập này để huấn luyện và thử nghiệm hai mô hình là SVM và Random Forest. Hình 3.19 và Hình 3.20 biểu diễn kết quả nhận dạng thu được khi sử dụng bộ dữ liệu với tỉ lệ chồng chập 90%.



Hình 3.19 Kết quả nhận dạng bằng mô hình SVM tỉ lệ chông chấp 90%



Hình 3.20 Kết quả nhận dạng bằng mô hình RF tỉ lệ chông chấp 90%

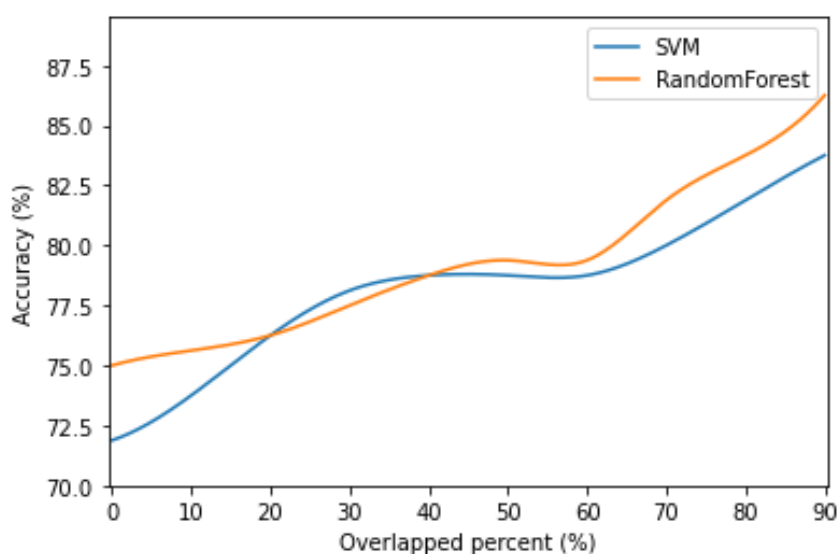
Tỉ lệ nhận dạng đúng (trung bình cho cả 8 loại cảm xúc khác nhau) của mô hình SVM và Random Forest lần lượt là 83,75% và 86,25%. Đối với mô hình SVM, hai trạng thái cảm xúc tức giận và tôn trọng cho kết quả nhận dạng tốt nhất (95%), hai trạng thái cảm xúc lãng mạn và vui vẻ cho tỉ lệ thấp nhất (75%). Tương tự, mô hình Random Forest nhận dạng tốt nhất với trạng thái cảm xúc tôn trọng (100%) và thấp nhất với trạng thái cảm xúc vui vẻ (70%).

3.10 Phân tích xác định tỉ lệ chồng chập tối ưu

Kết quả nhận dạng đúng trung bình cho cả 8 trạng thái cảm xúc thu được từ các thử nghiệm tỉ lệ chồng chập khác nhau như đã trình bày ở các mục từ 3.1 tới 3.9, được tổng hợp qua Bảng 3.1 và dạng biểu đồ Hình 3.21.

Bảng 3.1 Kết quả nhận dạng của 2 mô hình SVM và RF

Tỉ lệ (%)	0	10	20	30	40	50	60	70	80	90
SVM (%)	71.88	73.75	76.25	78.12	78.75	78.75	78.75	80.00	81.88	83.75
RF (%)	75.00	75.62	76.25	77.50	78.75	79.38	79.38	81.88	83.75	86.25



Hình 3.21 Tổng hợp kết quả nhận dạng cảm xúc khi thay đổi tỉ lệ chồng chập

Từ biểu đồ có thể thấy khi tỉ lệ chồng chập tăng lên thì chất lượng nhận dạng của cả 2 mô hình cũng tăng theo, từ tỉ lệ 71,88% khi chưa chồng chập lên đến 83,75% tương ứng với tỉ lệ chồng chập từ 10% tới 90% đối với mô hình SVM. Với mô hình Random Forest, tỉ lệ nhận dạng đúng tăng từ 75,00% lên 86,25% khi tỉ lệ chồng chập dữ liệu vào tăng từ 10% lên 90% đối với mô hình Random Forest. Khi tỉ lệ chồng chập từ 50% trở lên, chất lượng nhận dạng của 2 mô hình khá cao (trên 78%), trong đó mô hình Random Forest luôn cho kết quả cao hơn mô hình SVM từ 1-3%. Khi tỉ lệ chồng chập từ 70% trở lên, kết quả nhận được của cả hai mô hình đều lớn hơn 80%. Như vậy, để có kết quả nhận dạng tốt (từ 80% trở lên) cho cả hai mô hình phân lớp SVM và RF, dữ liệu vào cần có tỉ lệ chồng chập từ 70% trở lên.

Xét trên các đặc trưng sinh lý học, thì việc xuất hiện và duy trì các trạng thái cảm xúc thường không bất chợt mà diễn tiến từ từ, ngoại trừ các trạng thái “giật mình” khi có kích thích đột ngột xảy ra. Do đó, việc duy trì tỉ lệ chồng chập đủ lớn (từ 70% trở lên) khiến các đặc trưng tín hiệu được thể hiện tốt và thống nhất trong

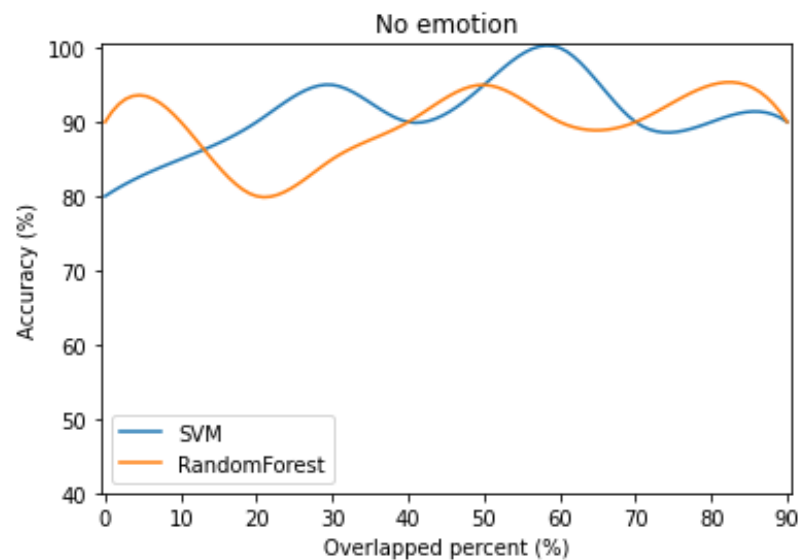
tập dữ liệu huấn luyện, và vì thế kết quả nhận dạng tốt hơn. Song cần cân nhắc sử dụng tỉ lệ chồng chấp lớn tới 90% để phù hợp với bài toán cụ thể.

3.11 Kết quả nhận dạng với từng trạng thái cảm xúc

- Trạng thái cảm xúc trung tính

Bảng 3.2 Tỉ lệ nhận dạng của trạng thái cảm xúc trung tính

Tỉ lệ (%)	0	10	20	30	40	50	60	70	80	90
SVM (%)	80.00	85.00	90.00	95.00	90.00	95.00	100	90.00	90.00	90.00
RF (%)	90.00	90.00	80.00	85.00	90.00	95.00	90.00	90.00	95.00	90.00



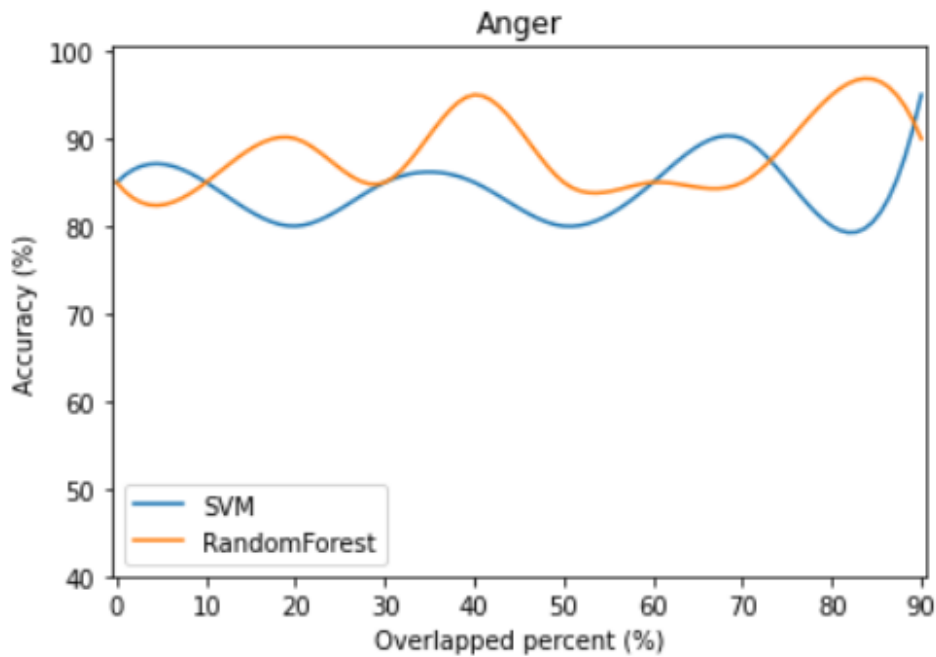
Hình 3.22 Kết quả nhận dạng cảm xúc trung tính với từng tỉ lệ chồng chấp

Với trạng thái cảm xúc trung tính, tỉ lệ nhận dạng trung bình của mô hình SVM là 90,5%, cao hơn so với tỉ lệ nhận dạng trung bình của mô hình Random Forest với 89,5%. Đây là trạng thái cảm xúc có tỉ lệ nhận dạng rất tốt (luôn > 80%).

- Trạng thái xầm xúc tức giận

Bảng 3.3 Tỉ lệ nhận dạng của trạng thái cảm xúc tức giận

Tỉ lệ (%)	0	10	20	30	40	50	60	70	80	90
SVM (%)	85.00	85.00	80.00	85.00	85.00	80.00	85.00	90.00	80.00	95.00
RF (%)	85.00	85.00	90.00	85.00	95.00	85.00	85.00	85.00	95.00	90.00



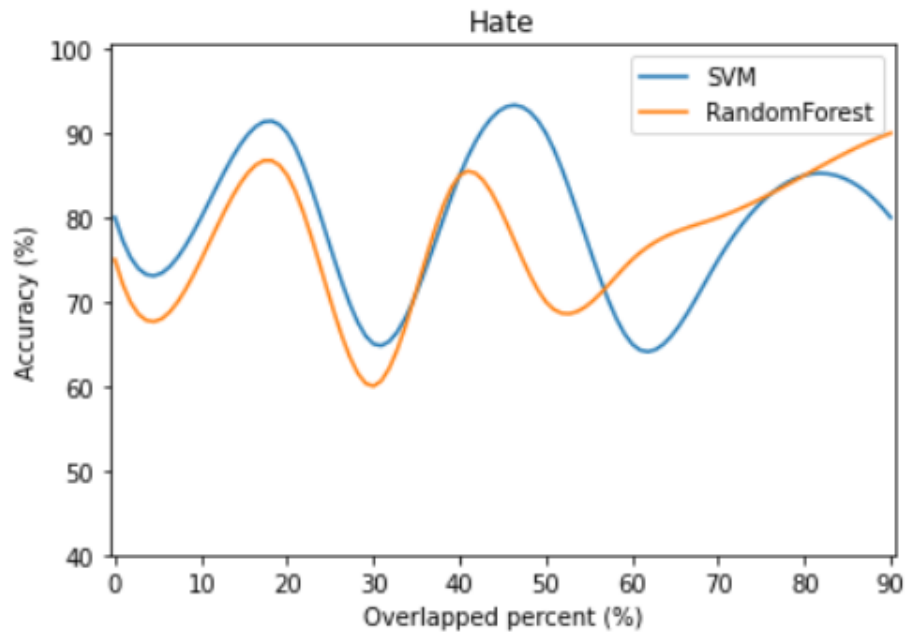
Hình 3.23 Kết quả nhận dạng cảm xúc tức giận với từng tỉ lệ chồng chập

Với trạng thái cảm xúc tức giận, tỉ lệ nhận dạng trung bình của mô hình SVM là 85,0%, thấp hơn so với tỉ lệ nhận dạng trung bình của mô hình Random Forest với 88,0%. Trạng thái cảm xúc này có tỉ lệ nhận dạng tốt, dao động trong ngưỡng từ 80% đến 95%.

- Trạng thái cảm xúc ghét bỏ

Bảng 3.4 Tỉ lệ nhận dạng của trạng thái cảm xúc ghét bỏ

Tỉ lệ (%)	0	10	20	30	40	50	60	70	80	90
SVM (%)	80.00	80.00	90.00	65.00	85.00	90.00	65.00	75.00	85.00	80.00
RF (%)	75.00	75.00	85.00	60.00	85.00	70.00	75.00	80.00	85.00	90.00



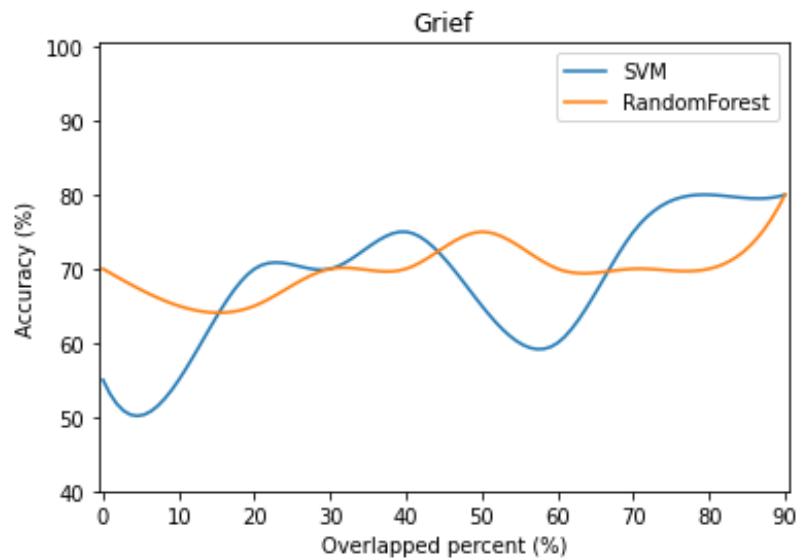
Hình 3.24 Kết quả nhận dạng cảm xúc ghét bỏ với từng tỉ lệ chồng chập

Với trạng thái cảm xúc ghét bỏ, tỉ lệ nhận dạng trung bình của mô hình SVM là 79,5%, cao hơn so với tỉ lệ nhận dạng trung bình của mô hình Random Forest với 78,0%. Trạng thái cảm xúc này có tỉ lệ nhận dạng trung bình, dao động lớn trong ngưỡng từ 60% đến 90%.

- Trạng thái cảm xúc đau buồn

Bảng 3.5 Tỉ lệ nhận dạng của trạng thái cảm xúc đau buồn

Tỉ lệ (%)	0	10	20	30	40	50	60	70	80	90
SVM (%)	55.00	55.00	70.00	70.00	75.00	65.00	60.00	75.00	80.00	80.00
RF (%)	70.00	65.00	65.00	70.00	70.00	75.00	70.00	70.00	70.00	80.00



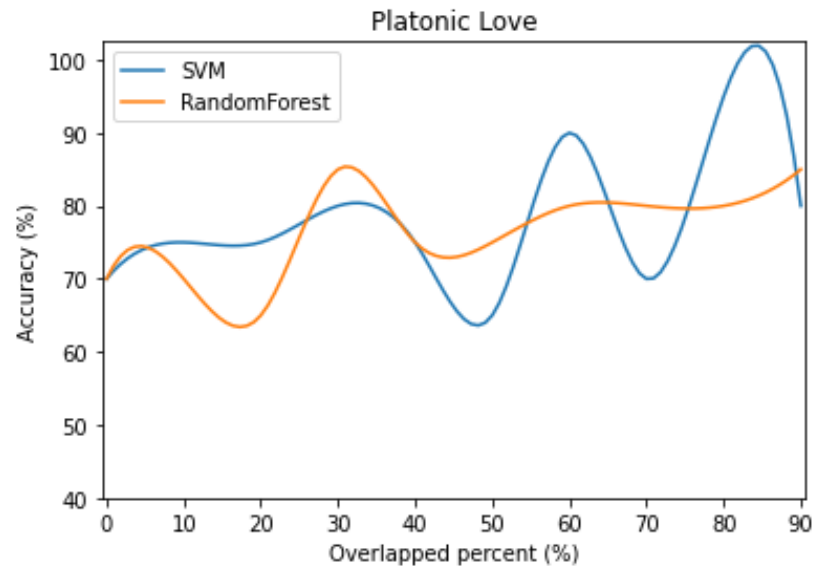
Hình 3.25 Kết quả nhận dạng cảm xúc đau buồn với từng tỉ lệ chồng chập

Với trạng thái cảm xúc đau buồn, tỉ lệ nhận dạng trung bình của mô hình SVM là 68,5%, thấp hơn so tỉ lệ nhận dạng trung bình của mô hình Random Forest với 70,0%. Trạng thái cảm xúc này cho tỉ lệ nhận dạng thấp, thấp nhất 55,0% (khi tỉ lệ chồng chập là 10%). Với RF, tỉ lệ nhận dạng đúng gần như không phụ thuộc vào tỉ lệ chồng chập, khi dao động quanh ngưỡng 70%.

- Trạng thái cảm xúc đồng cảm

Bảng 3.6 Tỉ lệ nhận dạng của trạng thái cảm xúc đồng cảm

Tỉ lệ (%)	0	10	20	30	40	50	60	70	80	90
SVM (%)	70.00	75.00	75.00	80.00	75.00	65.00	90.00	70.00	95.00	80.00
RF (%)	70.00	70.00	65.00	85.00	75.00	75.00	80.00	80.00	80.00	85.00



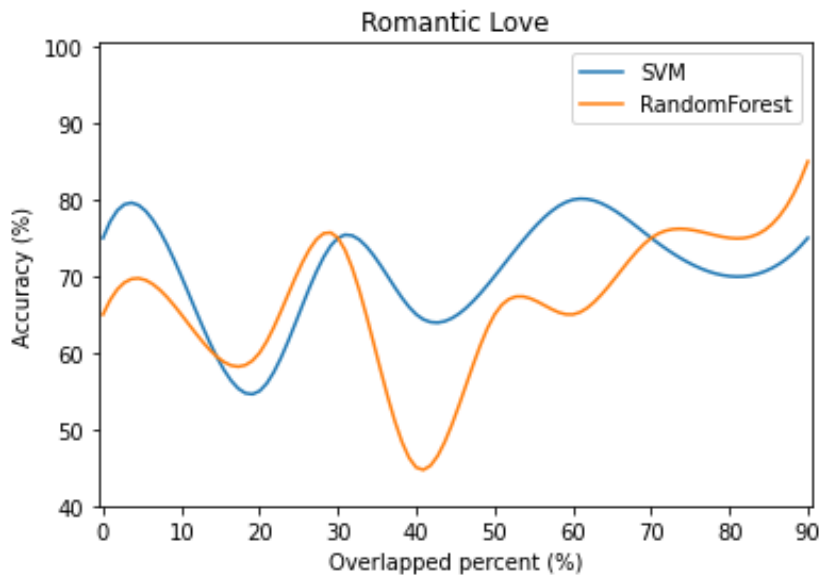
Hình 3.26 Kết quả nhận dạng cảm xúc đồng cảm với từng tỉ lệ chồng chập

Với trạng thái cảm xúc đồng cảm, tỉ lệ nhận dạng trung bình của mô hình SVM là 77,5%, cao hơn so với tỉ lệ nhận dạng trung bình của mô hình Random Forest với 76,5%. Trạng thái cảm xúc này cho tỉ lệ nhận dạng trung bình, dao động chủ yếu trong ngưỡng từ 65% đến 85%.

- Trạng thái cảm xúc lãng mạn

Bảng 3.7 Tỉ lệ nhận dạng của trạng thái cảm xúc lãng mạn

Tỉ lệ (%)	0	10	20	30	40	50	60	70	80	90
SVM (%)	75.00	70.00	55.00	75.00	65.00	70.00	80.00	75.00	70.00	75.00
RF (%)	65.00	65.00	60.00	75.00	45.00	65.00	65.00	75.00	75.00	85.00



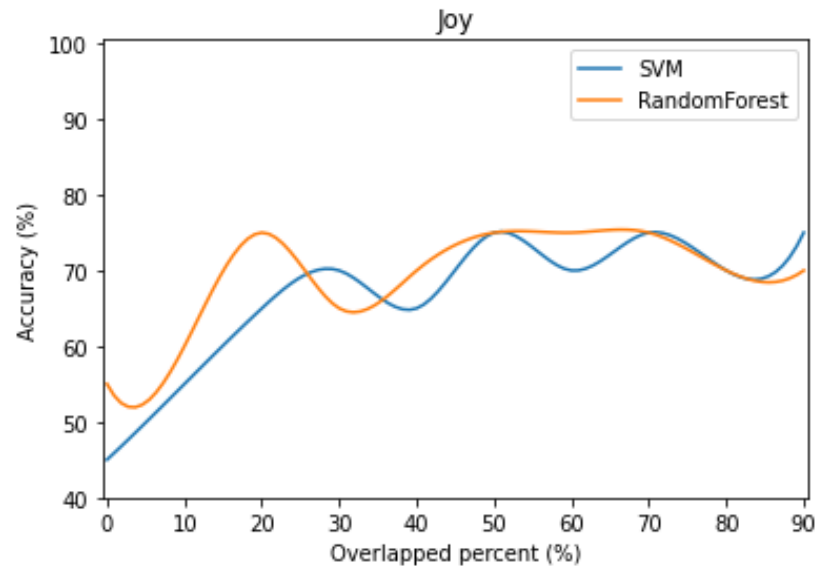
Hình 3.27 Kết quả nhận dạng cảm xúc lãng mạn với từng tỉ lệ chồng chập

Với trạng thái cảm xúc lãng mạn, tỉ lệ nhận dạng trung bình của mô hình SVM là 71,0%, tốt hơn đáng kể so với tỉ lệ nhận dạng trung bình của mô hình Random Forest với 67,5%. Trạng thái cảm xúc này cho tỉ lệ nhận dạng trung bình với mô hình SVM (dao động trong khoảng 65-80%) nhưng lại rất thấp với mô hình Random Forest (từ 50% đến 70%).

- Trạng thái cảm xúc vui vẻ

Bảng 3.8 Tỉ lệ nhận dạng của trạng thái cảm xúc vui vẻ

Tỉ lệ (%)	0	10	20	30	40	50	60	70	80	90
SVM (%)	45.00	55.00	65.00	70.00	65.00	75.00	70.00	75.00	70.00	75.00
RF (%)	55.00	60.00	75.00	65.00	70.00	75.00	75.00	75.00	70.00	70.00



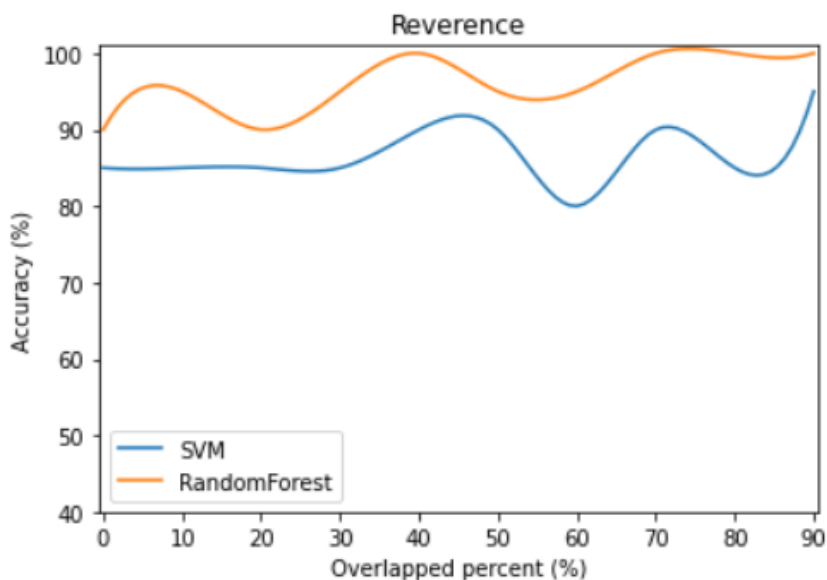
Hình 3.28 Kết quả nhận dạng cảm xúc vui vẻ với từng tỉ lệ chồng chập

Với trạng thái cảm xúc vui vẻ, tỉ lệ nhận dạng trung bình của mô hình SVM là 66,5%, thấp hơn so với tỉ lệ nhận dạng trung bình của mô hình Random Forest là 69,0%. Trạng thái cảm xúc này cho tỉ lệ nhận dạng thấp với cả 2 mô hình SVM và Random Forest (chỉ từ 45% tới 75%, tùy vào tỉ lệ chồng chập dữ liệu). Với RF, khi tỉ lệ chồng chập dữ liệu trong khoảng 50%-70%, kết quả nhận dạng tốt nhất và bằng nhau (bão hòa với tỉ lệ chồng chập này).

- Trạng thái cảm xúc tôn trọng

Bảng 3.9 Tỉ lệ nhận dạng của trạng thái cảm xúc tôn trọng

Tỉ lệ (%)	0	10	20	30	40	50	60	70	80	90
SVM (%)	85.00	85.00	85.00	85.00	90.00	90.00	80.00	90.00	85.00	95.00
RF (%)	90.00	95.00	90.00	95.00	100	95.00	95.00	100	100	100



Hình 3.29 Kết quả nhận dạng cảm xúc tôn trọng với từng tỉ lệ chồng chập

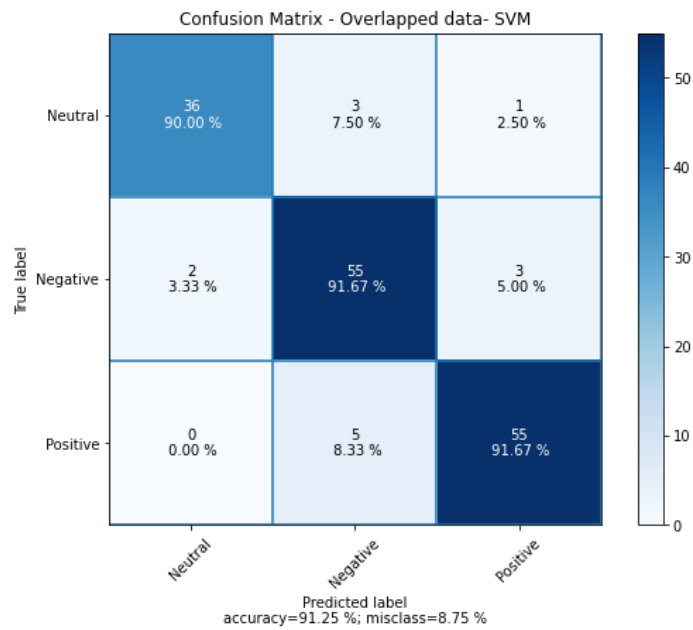
Với trạng thái cảm xúc tôn trọng, tỉ lệ nhận dạng trung bình (tính trên tất cả các tỉ lệ chồng chập, từ 0% tới 90%) của mô hình SVM là 87,0%, còn với mô hình Random Forest là rất cao, tới 96,0%. Trạng thái cảm xúc này cho tỉ lệ nhận dạng rất tốt với cả hai mô hình SVM và Random Forest, trong đó mô hình Random Forest cho thấy khả năng nhận dạng vượt trội (lớn hơn 9%) so với mô hình SVM. Mặt khác, trạng thái cảm xúc này có tỉ lệ nhận dạng đúng gần như không phụ thuộc vào tỉ lệ chồng chập dữ liệu.

Qua nhận xét tỉ lệ nhận dạng đúng với từng trạng thái cảm xúc, khi thay đổi tỉ lệ chồng chập, cho thấy hiệu quả tổng quát của giải pháp chồng chập dữ liệu, nhằm làm giàu thông tin và thống nhất hóa các đặc trưng quan trọng cho nhận dạng. Song ở một số trạng thái cảm xúc, gần như việc thay đổi tỉ lệ chồng chập không ảnh hưởng nhiều tới kết quả nhận dạng. Điều đó mở ra bài toán tìm hiểu kỹ hơn về các đặc trưng của tín hiệu sinh học tương ứng với các trạng thái cảm xúc khác nhau.

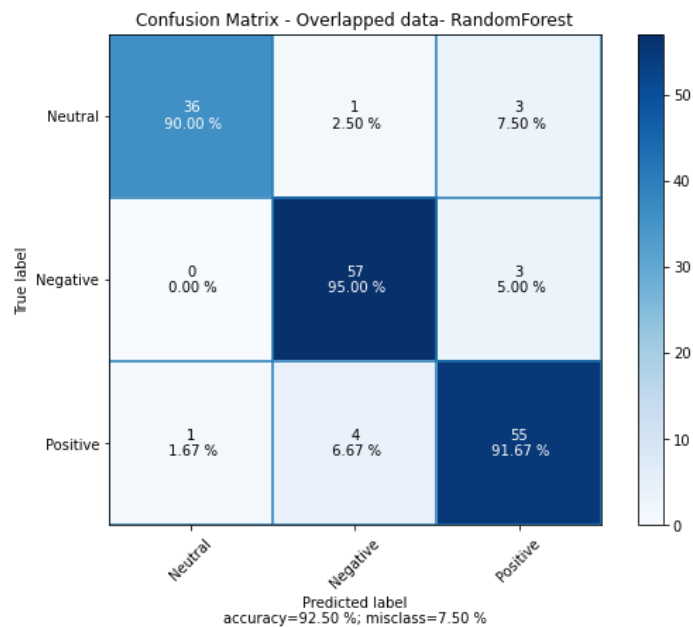
3.12 Nhận dạng với mô hình ba trạng thái của cảm xúc

Từ 8 trạng thái cảm xúc riêng biệt trong bộ dữ liệu MIT, chúng tôi gộp lại và đề xuất mô hình cảm xúc ba trạng thái, gồm: tích cực, trung lập và tiêu cực. Trong đó, ba cảm xúc đồng cảm, lãng mạn và vui vẻ được nhóm vào trạng thái cảm xúc tích cực; hai cảm xúc trung tính và tôn trọng được xếp vào trạng thái trung lập; ba cảm xúc còn lại là tức giận, ghét bỏ và đau buồn thuộc về trạng thái tiêu cực.

Kết quả nhận dạng của 2 mô hình SVM và Random Forest với 3 trạng thái cảm xúc trong mô hình chúng tôi đề xuất được thể hiện trong Hình 3.30 và Hình 3.31. Ở đây, chúng tôi sử dụng dữ liệu huấn luyện có độ dài 10s, với tỉ lệ chồng chập 90%.



Hình 3.30 Kết quả nhận dạng 3 trạng thái của cảm xúc với SVM



Hình 3.31 Kết quả nhận dạng 3 trạng thái của cảm xúc với RF

Với mô hình ba trạng thái cảm xúc, tỉ lệ nhận dạng đúng trung bình là rất cao, 91,25% khi sử dụng SVM và 92,50% khi sử dụng Random Forest. Với mô hình cảm xúc ít trạng thái riêng biệt hơn, khiến việc nhận dạng dễ dàng hơn so với khi phải nhận dạng 8 loại cảm xúc khác nhau. Mặt khác, các kết quả thu được cũng thể hiện tính hợp lý trong đề xuất của chúng tôi về việc nhóm các trạng thái cảm xúc khác nhau vào một mô hình ba trạng thái (tích cực tiêu cực và trung lập).

KẾT LUẬN CHUNG VÀ HƯỚNG PHÁT TRIỂN

- Kết luận chung:

Luận văn đã đưa ra một giải pháp nâng cao chất lượng nhận dạng cảm xúc sử dụng tín hiệu sinh học bằng cách tổ chức lại dữ liệu đầu vào. Kết quả nghiên cứu đã đảm bảo những yêu cầu đặt ra ban đầu, bao gồm lựa chọn kỹ thuật phân loại cảm xúc thích hợp, tổ chức lại dữ liệu huấn luyện nhằm nâng cao chất lượng nhận dạng, thử nghiệm và đánh giá kết quả thu được.

Các kết quả thử nghiệm cho thấy, khi tỉ lệ chồng chập từ 70% trở lên, kết quả nhận dạng đúng, trung bình luôn lớn hơn 80% cho cả 8 trạng thái cảm xúc khác nhau. Với tỉ lệ chồng chập 90% cho tỉ lệ nhận dạng đúng lên tới với 86.25%.

Bên cạnh đó, học viên cũng tổ chức phân chia lại 8 trạng thái cảm xúc trong bộ dữ liệu MIT thành 3 lớp cảm xúc xét theo tính chất, bao gồm tích cực, tiêu cực và trung tính. Kết quả nhận dạng 3 trạng thái cảm xúc này khi áp dụng nguyên tắc chồng chập dữ liệu với tỉ lệ 90% đã thu được tỉ lệ chính xác lên đến 92,5%.

Luận văn cho thấy tính khả thi của phương pháp đề ra trong việc nhận dạng cảm xúc trong thời gian ngắn. Các kết quả này cũng mở ra khả năng xây dựng các ứng dụng nhận dạng cảm xúc thời gian thực.

- Hướng phát triển đề tài:

Do hạn chế về mặt thời gian, phương pháp nghiên cứu trong luận văn mới chỉ được áp dụng trên bộ dữ liệu cảm xúc của MIT, cùng với đó là 4 thuật toán phân lớp khác nhau trong học máy. Có thể mở rộng hướng nghiên cứu này theo một số đề xuất dưới đây:

- + Áp dụng các thuật toán lựa chọn đặc trưng và phân lớp khác trên bộ dữ liệu này của MIT.
- + Mỗi mô hình cho kết quả ưu thế hơn các mô hình còn lại đối với một số loại cảm xúc. Từ đó, nghiên cứu kết hợp các mô hình một cách thích hợp để nâng cao chất lượng nhận dạng cho tất cả các loại cảm xúc khác nhau.
- + Áp dụng phương pháp đề ra lên các tập dữ liệu mẫu khác, nhằm đánh giá kỹ hơn tính toàn diện của giải pháp đã đề xuất.
- + Thiết kế bộ thu dữ liệu đa kênh, thử nghiệm khả năng nhận dạng cảm xúc online trên các đoạn dữ liệu ngắn.

Tài liệu tham khảo

- [1] P.J.Lang, "The emotion probe: Studies of motivation and attention. Am. Psychol. 1995, 50, 372–385".
- [2] "B. Schuller, G. Rigoll and M. Lang, "Hidden Markov model-based speech emotion recognition," 2003 International Conference on Multimedia and Expo. ICME '03. Proceedings (Cat. No.03TH8698), Baltimore, MD, USA, 2003, pp. I-401, doi: 10.1109/ICME.2003.1220939".
- [3] "Tsang-Long Pao, Yu-Te Chen, Jun-Heng Yeh and Pei-Jia Li, "Mandarin Emotional Speech Recognition Based on SVM and NN," 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 2006, pp. 1096-1100, doi: 10.1109/ICPR.2006.780".
- [4] "T. L. Nwe, N. T. Hieu and D. K. Limbu, "Bhattacharyya distance based emotional dissimilarity measure for emotion classification," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 2013, pp. 7512-7516, do".
- [5] "E. Mower, M. J. Matarić and S. Narayanan, "A Framework for Automatic Human Emotion Classification Using Emotion Profiles," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 5, pp. 1057-1070, July 2011, doi: 10.1109/TASL.2010.207".
- [6] "<https://www.semanticscholar.org/paper/Emotion-Recognition-via-Facial-Expression%3A-of-in-Kwong-Garcia>," [Online].
- [7] "B. A. Dixit and A. N. Gaikwad, "Statistical moments based facial expression analysis," 2015 IEEE International Advance Computing Conference (IACC), Bangalore, India, 2015, pp. 552-557, doi: 10.1109/IADCC.2015.7154768".
- [8] "Kumar, S.Ashok & K K, Thyagarajan. (2013). Facial Expression Recognition with Auto-Illumination Correction. 843-846. 10.1109/ICGCE.2013.6823551".
- [9] "M. Żarkowski, "Identification-driven emotion recognition system for a social robot," 2013 18th International Conference on Methods & Models in Automation & Robotics (MMAR), Miedzyzdroje, Poland, 2013, pp. 138-143, doi: 10.1109/MMAR.2013.6669895".
- [10] "S. Liu and W. Wang, "The Application Study of Learner's Face Detection and Location in the Teaching Network System Based on Emotion Recognition," 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing, Wuh".
- [11] "Cohen, Ira & Garg, Ashutosh. (2000). Emotion Recognition from Facial Expressions using Multilevel HMM".
- [12] "Shu, L.; Xie, J.; Yang, M.; Li, Z.; Li, Z.; Liao, D.; Xu, X.; Yang, X. A Review of Emotion Recognition Using Physiological Signals. Sensors 2018, 18, 2074".
- [13] "S. Wu, X. Xu, L. Shu and B. Hu, "Estimation of valence of emotion using two frontal EEG channels," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, USA, 2017, pp. 1127-1130, doi: 10.1109/BIBM.2017.8217815".
- [14] "Li, Z.; Tian, X.; Shu, L.; Xu, X.; Hu, B. Emotion Recognition from EEG Using RASM and LSTM. In Proceedings of the International Conference on Internet Multimedia Computing and Service, Tsingtao, China, 23–25 August 2017; pp. 310–318".

- [15] "Zong, C.; Chetouani, M. Hilbert-Huang transform based physiological signals analysis for emotion recognition. In Proceedings of the IEEE International Symposium on Signal Processing and Information Technology, Ajman, UAE, 14–17 December 2010; pp. 334–339."
- [16] "Lan, Z.; Sourina, O.; Wang, L.; Liu, Y. Real-time EEG-based emotion monitoring using stable features. *Vis. Comput.* 2016, 32, 347–358."
- [17] "https://therbootcamp.github.io/ML_2019May/_sessions/WhatIsML/WhatIsML.html," [Online].
- [18] "Support Vector Machine, <https://machinelearningcoban.com/2017/04/09/smv/>," [Online].
- [19] "<https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm>," [Online].
- [20] "Linear Discriminant Analysis, <https://machinelearningcoban.com/2017/06/30/lda/>," [Online].
- [21] "Decision Tree, <https://machinelearningcoban.com/2018/01/14/id3/>," [Online].
- [22] "<https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>," [Online].
- [23] "<https://www.javatpoint.com/machine-learning-random-forest-algorithm>," [Online].