HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Bùi Thái Duy

PHÁT HIỆN TIẾNG NGÁY DỰA TRÊN HỌC SÂU

LUẬN VĂN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI- 2020

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Bùi Thái Duy

PHÁT HIỆN TIẾNG NGÁY DỰA TRÊN HỌC SÂU

CHUYÊN NGÀNH : HỆ THỐNG THÔNG TIN

MÃ SỐ : 8.48.01.04

LUẬN VĂN THẠC SĨ KỸ THUẬT

(*Theo định hướng ứng dụng*)

NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS.TS. PHẠM VĂN CƯỜNG

HÀ NỘI- 2020

LỜI CAM ĐOAN

Tôi là Bùi Thái Duy, học viên lớp M18CQIS02 xin cam đoan báo cáo luận văn này được viết bởi tôi dưới sự hướng dẫn của thầy giáo PGS. TS Phạm Văn Cường. Trong toàn bộ nội dung của luận văn, những điều được trình bày là kết quả của cá nhân tôi hoặc là được kế thừa, tổng hợp từ nhiều nguồn tài liệu khác được liệt kê trong danh mục tài liệu tham khảo rõ ràng.

Hà Nội, ngày ..... tháng ..... năm 2020

Học viên

Bùi Thái Duy

LỜI CẢM ƠN

Lời đầu tiên, tôi xin bày tỏ sự cảm ơn chân thành đối với thầy giáo PGS.TS Phạm Văn Cường - Giáo viên hướng dẫn trực tiếp của tôi. Thầy đã giúp tôi tiếp cận những kiến thức về ứng dụng học máy và học sâu cho bài toán phát hiện tiếng ngáy trong suốt quá trình nghiên cứu và hoàn thiện luận văn thạc sĩ kỹ thuật.

Tôi xin gửi lời cảm ơn tới các thầy cô trong khoa Công nghệ Thông tin Học viện Bưu chính viễn thông đã hướng dẫn, chỉ bảo và tạo điều kiện cho chúng tôi học tập và nghiên cứu tại trường trong suốt thời gian qua.

Xin gửi lời biết ơn đến gia đình, bạn học và đồng nghiệp đã luôn quan tâm, động viên, ủng hộ tôi về mặt tinh thần lẫn vật chất trong suốt thời gian thôi tham gia khóa học và thực hiện luận văn này.

Học viên xin gửi lời cảm ơn sự hỗ trợ từ đề tài nghiên cứu độc lập cấp quốc gia “Nghiên cứu thiết kế, chế tạo hệ thống tự động trợ giúp theo dõi hô hấp và vận động bất thường dựa trên nền tảng Internet vạn vật (IoT-Internet of Things) ” mã số ĐTĐLCN-16/18.

Mặc dù đã cố gắng để hoàn thành luận văn nhưng chắc chắn sẽ không thể tránh khỏi những thiếu sót. Kính mong nhận được sự thông cảm và chỉ bảo của các quý thầy cô.

Em xin trân trọng cảm ơn.

Hà Nội, ngày ..... tháng ..... năm 2020

Học viên thực hiện luận văn

Bùi Thái Duy

MỤC LỤC

DANH MỤC CÁC THUẬT NGỮ VÀ CHỮ VIẾT TẮT

Viết tắt	Tiếng Anh	Tiếng Việt
CNN	Convolutional Neural network	Mạng neural tích chập
FFT	Fast Fourier Transform	Biến đổi Fourier nhanh
FT	Fourier transform	Biến đổi Fourier
LSTM	Long short-term memory	Mạng ghi nhớ hồi quy lâu
MFCC	Mel Frequency Cepstral Coefficients	Hệ số biểu diễn phổ của phổ
RNN	Recurrent Neural Network	Mạng neural hồi quy
STFT	Short term fourier transform	Phép biến đổi Fourier thời gian ngắn
SVM	Support Vector Machine	Máy vector hỗ trợ

DANH MỤC CÁC BẢNG

DANH MỤC CÁC HÌNH VẼ VÀ ĐỒ THỊ

MỞ ĐẦU

Trong những năm gần đây với sự phát triển của khoa học kỹ thuật đặc biệt là công nghệ thông tin đã mang lại nhiều hiệu quả đối với khoa học cũng như hỗ trợ con người trong cuộc sống. Nhận dạng hoạt động người là quá trình theo dõi và phân tích các hành vi người dùng nhằm nhận dạng/suy đoán các hành động đang xảy ra.

Sự tiến bộ của công nghệ đã thúc đẩy cộng đồng nghiên cứu chuyển từ truyền, thu nhận và xử lý dữ liệu mức thấp sang nghiên cứu tích hợp thông tin mức cao, xử lý ngữ cảnh, nhận dạng và suy diễn các hoạt động. Thêm vào đó, ngày càng có nhiều bài toán thực tế cần các giải pháp dựa vào nhận dạng hoạt động. Ví dụ như ứng dụng nhận dạng các biển báo giao thông, nhận dạng âm thanh người dùng. Các ứng dụng giúp con người trong cuộc sống hàng ngày cần công nghệ giám sát, phát hiện hoạt động và hỗ trợ con người khi không tỉnh táo như là khi ngủ. Hàng loạt ứng dụng mới như nhà thông minh, theo dõi sức khỏe thời gian thực, phân tích chất lượng giấc ngủ... cũng phụ thuộc vào công nghệ nhận dạng hoạt động để cung cấp nhiều cách thức tương tác đa dạng, chủ động cung cấp các dịch vụ, trợ giúp người dùng hoàn thành công việc.

Bên cạnh tác động tới chất lượng giấc ngủ của con người thì ngáy cũng có dấu hiệu của chứng ngưng thở (OSA) sau khi mất ngủ, tỷ lệ mắc bệnh rối loạn giấc ngủ cao nhất, ảnh hưởng đến khoảng 3 - 7% đàn ông trung niên và 2-5% phụ nữ trung niên trong dân số nói chung. OSA được đặc trưng bởi các đợt lặp đi lặp lại của sự khó khăn một phần hoặc hoàn toàn của đường hô hấp trên trong khi ngủ, gây ra sự trao đổi khí bị suy yếu và rối loạn giấc ngủ.

Là một tình trạng mãn tính gây ra do tắc nghẽn đường hô hấp trên trong khi ngủ, OSA có thể dẫn đến tăng nguy cơ mắc các bệnh về tim mạch và mạch máu não. Một phần không thể thiếu trong điều trị thành công là định vị vị trí tắc nghẽn và rung. Ngoài ra, tiếng ngáy cần được xác định bốn nguồn rung động khác nhau từ các mẫu ngáy âm thanh: biểu mô, vách bên hầu họng, lưỡi và màng khẩu. Các đặc trưng của âm thanh tiếng ngáy của con người thì được đặc trưng qua biên độ, tần số và các sóng

con thông qua các đặc trưng này có thể tạo nên các ảnh phổ của âm thanh, miêu tả các đặc trưng cơ bản nhất của âm thanh. Từ đó, có thể thấy được việc phân lớp âm thanh tiếng ngáy có thể thông qua việc sử dụng ảnh phổ hoặc dựa trên sóng thô của âm thanh.

Những nghiên cứu trong học sâu từ trước tới nay đã và đang được sử dụng để giải quyết nhiều bài toán về nhận dạng, phát hiện đặc biệt trong lĩnh vực thị giác máy tính. Vì đòi hỏi cần một lượng dữ liệu, thời gian, sức mạnh tính toán đáng kể, các nỗ lực nghiên nghiên cứu cách để tận dụng các mạng CNN được đào tạo trước cho các nhiệm vụ khác như mạng CNN được sử dụng trong các hệ thống nhận dạng. Cho đến nay, rất ít các nghiên cứu thực hiện để khám phá biểu diễn đặc trưng của âm thanh với mạng CNN. Trong thử thách INTERSPEECH ComParE 2017 có một thử thách là xác định tiếng ngáy, đó cũng là tiền đề để phát triển các ứng dụng khai thác âm thanh ngáy. Để phát hiện và phân loại âm thanh thông qua phổ của âm thanh dựa trên học sâu là một lĩnh vực nghiên cứu mới. Đến nay, một số bài báo có cách tiếp cận mạng neural tích chập trong vấn đề của Phân loại âm thanh đàn (ASC). Cách tiếp cận việc xử lý âm thanh dưới dạng ảnh phổ có thể kết hợp được những ưu điểm của xử lý hình ảnh và âm thanh từ đó mang lại hiệu quả cao trong việc phát hiện và nhận dạng.

Vì những Đề tài *"Phát hiện tiếng ngáy dựa trên học sâu"* được thực hiện trong khuôn khổ luận văn thạc sĩ chuyên ngành hệ thống thông tin nhằm góp phần đánh giá một số như việc xử lý, lưu trữ âm thanh được thực hiện qua việc xử lý ảnh phổ, kết hợp được việc so sánh, đánh giá các kiến trúc học sâu trong việc phát hiện tiếng ngáy..

CHƯƠNG 1: TỔNG QUAN VỀ PHÁT HIỆN TIẾNG NGÁY

Nội dung chương này sẽ bao gồm giới thiệu chung về bài toán phát hiện tiếng ngáy, những khó khăn và ý nghĩa của bài toán này. Chương này cũng trình bày về các nghiên cứu liên quan với các vấn đề về phát hiện âm thanh, nghiên cứu về học máy cũng như học sâu .Từ những cơ sở nghiên cứu này sẽ xác định rõ hướng nghiên cứu của luận văn.

1.1 Bài toán phát hiện tiếng ngáy

Theo nghiên cứu "Giá trị của sự tỉnh táo: ảnh hưởng của do hạn chế giấc ngủ mãn tính và thiếu ngủ hoàn toàn đối với các chức năng thần kinh và sinh lý" [10] đã nhận xét giấc ngủ là hoạt động hồi phục cho não, ngủ không đủ giấc làm giảm động lực cho các hoạt động thể chất, tăng cân, béo phì và các rốt loạn liên quan khác. Do đó, có nhiều nghiên cứu đã được thực hiện để cải thiện chất lượng giấc ngủ và phân loại giai đoạn giấc ngủ có thể mọi người có thể áp dụng vào thực tế từ đó cải thiện chất lượng cuộc sống. Theo Hiệp hội Y học Giấc ngủ Hoa Kỳ (AASM) đã đưa ra nguyên nhân chính của việc rối loạn giấc ngủ ngon là ngáy.

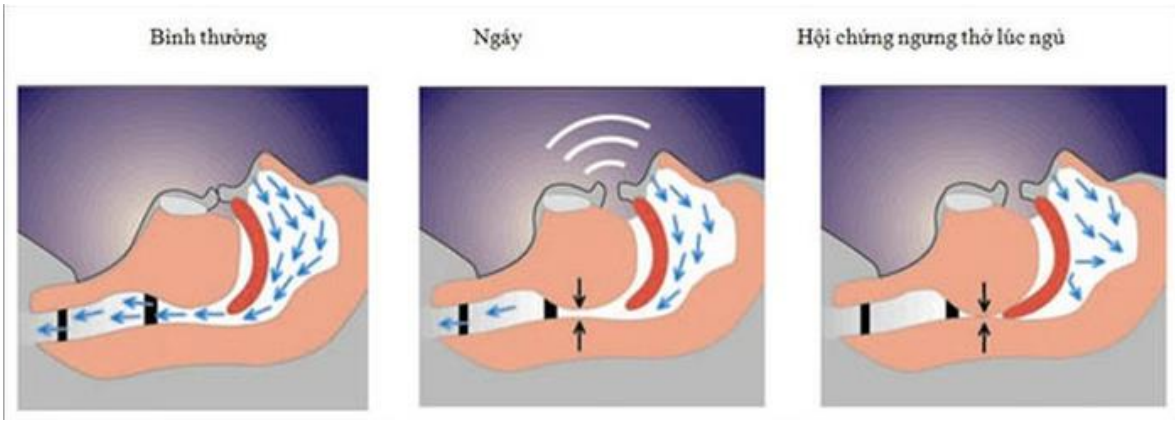
Ngáy thường được mô tả là một âm thanh thô và rung trong khi ngủ do sự tắc nghẽn một phần của trong hầu họng. Tỷ lệ ngáy ngủ thay đổi từ 2% đến 85%. Ngáy đơn giản có thể là khởi đầu của chứng rối loạn giấc ngủ mãn tính (SDB), bao gồm từ sự co hẹp đường thở một phần và tăng nhẹ sự cản đường thở trên đến sự sụp đổ đường thở hoàn toàn và ngưng thở khi ngủ do tắc nghẽn nghiêm trọng (OSA) kéo dài từ 60 giây trở lên . Có bằng chứng tích lũy rằng ngáy có liên quan đến một số vấn đề sức khỏe, bao gồm buồn ngủ, bệnh tim mạch, hội chứng chuyển hóa (MetS) và tử vong do các nguyên nhân khác nhau.

1.1.1 Các bệnh lý liên quan đến tiếng ngáy

Ngáy ngày càng được công nhận là mối quan tâm về sức khỏe cộng đồng. Đây là một vấn đề phổ biến ở người lớn và là dấu hiệu của hội chứng ngưng thở khi ngủ do tắc nghẽn (OSA). Một số nghiên cứu về y tế đã chỉ ra các yếu tố liên quan chính

đến ngáy dựa trên nghiên cứu đó là lão hóa, giới tính nam, tăng huyết áp, buồn ngủ ban ngày, hút thuốc và huyết thống. Các nghiên cứu đã mô tả điều này ở hầu hết các nước phát triển và một số nước đang phát triển như Hàn Quốc, Trung Quốc.. đã chỉ ra những khác biệt phụ thuộc vào yếu tố liên quan này. Cho đến nay vẫn chưa có nghiên cứu nào được công bố về chủ đề này riêng cho người ở Việt Nam. Việc thực hiện nghiên cứu với dữ liệu tiếng ngáy thu thập từ người Việt Nam có sẽ thấy được đặc trưng sự khác biệt với dữ liệu các nước phát triển qua đó có thể so sánh đánh giá từ những sự khác biệt.

Bên cạnh tác động tới chất lượng giấc ngủ của con người thì ngáy cũng có dấu hiệu của chứng ngưng thở (OSA) sau khi mất ngủ, tỷ lệ mắc bệnh rối loạn giấc ngủ cao nhất, ảnh hưởng đến khoảng 3 - 7% đàn ông trung niên và 2-5% phụ nữ trung niên trong dân số nói chung. OSA được đặc trưng bởi các đợt lặp đi lặp lại của sự khó khăn một phần hoặc hoàn toàn của đường hô hấp trên trong khi ngủ, gây ra sự trao đổi khí bị suy yếu và rối loạn giấc ngủ.



Hình 1.1: Mô tả về đường đi của không khí khi ngủ

Là một tình trạng mãn tính gây ra do tắc nghẽn đường hô hấp trên trong khi ngủ, OSA có thể dẫn đến tăng nguy cơ mắc các bệnh về tim mạch và mạch máu não. Một phần không thể thiếu trong điều trị thành công là định vị vị trí tắc nghẽn và rung như trên Hình 1.1. Ngoài ra, tiếng ngáy cần được xác định bốn nguồn rung động khác nhau từ các mẫu ngáy âm thanh: biểu mô, vách bên hầu họng, lưỡi và màng khẩu. Các đặc trưng của âm thanh tiếng ngáy của con người thì được đặc trưng qua biên

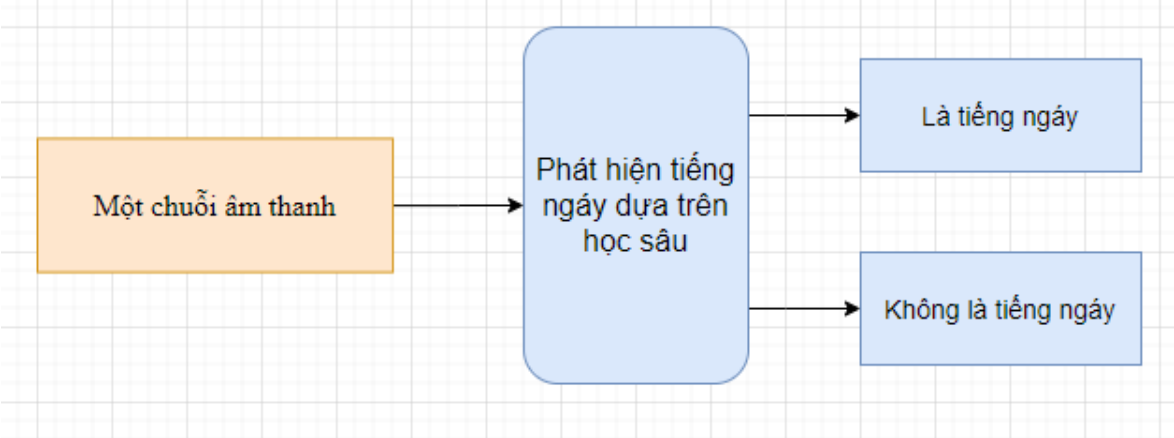
độ, tần số và các sóng con thông qua các đặc trưng này có thể tạo nên các ảnh phổ của âm thanh, miêu tả các đặc trưng cơ bản nhất của âm thanh.

1.1.2 Phát biểu bài toán

Với sự liên kết chặt chẽ của tiếng ngáy tới sức khỏe của con người thì cần thiết phải chọn ra tiếng ngáy với tần số tiếng ngáy và thời gian diễn ra tiếng ngáy trong khi ngủ. Ý tưởng của luận văn sẽ đi vào giải quyết bài toán phát hiện (âm thanh) tiếng ngáy với:

Đầu vào: Một chuỗi âm thanh

Đầu ra: Phát hiện âm thanh là tiếng ngáy hay không



Hình 1.2. Bài toán phát hiện tiếng ngáy

Như ví dụ tại hình 1.2, với đầu vào là "chuỗi âm thanh" hệ thống sẽ đưa ra được trong chuỗi âm thanh đó có tiếng ngáy hay không không phải tiếng ngáy, hay một nhóm các âm thanh vào thì hệ thống sẽ phát hiện được có bao nhiêu âm thanh trong đó là tiếng ngáy. Trong bài toán này có các vấn đề cơ bản cần được quan tâm chú ý như: xác định được đầu vào của hệ thống sẽ là các đặc trưng của chuỗi âm thanh hay là toàn bộ của cả chuỗi âm thanh hoặc một loại biểu diễn thông tin khác, với các dữ liệu đầu vào trên thì các cách xử lý của hệ thống thông qua các phương pháp học sâu.để có thể đưa ra thông tin đầu ra sau khi xử lý âm thanh là thuộc lớp tiếng ngáy hay không phải là tiếng ngáy.

1.1.3 Ý nghĩa bài toán

Dựa vào kết quả của luận văn "Phát hiện tiếng ngáy dựa trên học sâu" thì sẽ có được cái nhìn tổng quan về các đặc trưng âm thanh phù hợp với việc phát hiện âm thanh, đặc biệt là tiếng ngáy, các mô hình học sâu, những ưu điểm của các mô hình này so với các thuật toán học nông khác.

Từ kết quả của bài toán này có thể giải quyết vấn đề cơ bản trong việc thực hiện sản xuất các thiết bị cải thiện sức khỏe của người sử dụng, theo dõi các vấn đề sức khỏe, đưa ra các cảnh báo sớm thông qua sự thay đổi tiếng ngáy.

1.2 Các nghiên cứu liên quan

Việc theo dõi và cải thiện sức khỏe của con người đang ngày càng trở nên cấp thiết, việc phát hiện tiếng ngáy đang được quan tâm rất nhiều trong các cộng đồng nghiên cứu cả về y tế và kỹ thuật. Các nghiên cứu về y tế chỉ ra rằng trong lâm sàng thì tiếng ngáy đặc trưng cho sức khỏe của con người, tiếng ngáy được tạo ra khi dòng khí đi qua các vị trí trên bộ phận hô hấp.

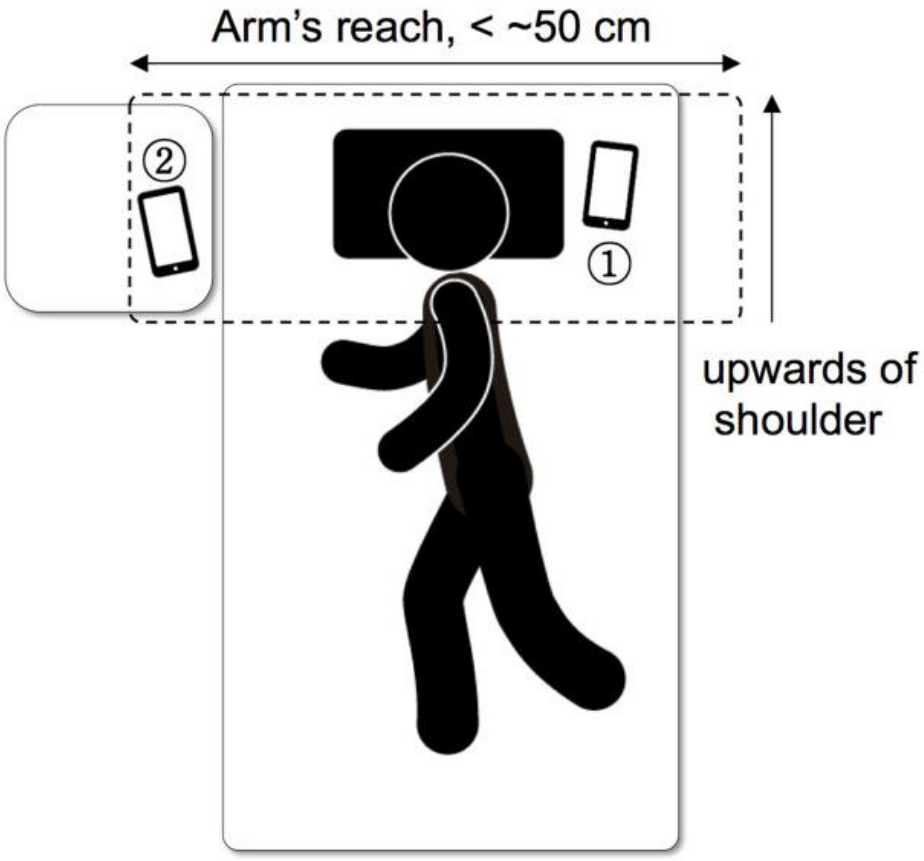
Điều này mang tới thách thức làm thế nào để xây dựng ra được hệ thống mà có thể phát hiện, theo dõi, phân tích và đo lường tiếng ngáy có độ chính xác cao. Trong thời đại số mà lượng thông tin cũng như khả năng xử lý khổng lồ của máy tính thì việc tạo ra một hệ thống như thế hoàn toàn khả thi. Các bài toán của hệ thống đang dần được được thực hiện, kế thừa, cải tiến từ các nghiên cứu trước đó. Bài toán "Phát hiện tiếng ngáy dựa trên học sâu" cũng vậy được phát triển dựa trên các nghiên cứu liên quan sau.

1.2.1 Thiết bị phát hiện tiếng ngáy

Trước đây, các nghiên cứu về phát hiện, phân loại âm nhạc là đề tài thu hút sự quan tâm của cộng đồng nghiên cứu và các cuộc thi về học máy, đây chính là tiền đề cho sự mở rộng của phát hiện và phân lớp các loại âm thanh đặc thù hơn, như nhận dạng bài hát hay phân lớp các âm thanh theo các ngữ cảnh khác nhau. Và cũng theo sự phát triển đó thì nghiên cứu "Ngáy: phân tích, đo lường, ý nghĩa lâm sàng và ứng

dụng”[13] đã nói rằng: “Tiếng ngáy bắt nguồn từ đường hô hấp trên, hoạt động như một ống có thể đóng, mở và có xu hướng đóng lại trong giai đoạn hô hấp. Ngáy là một dấu hiệu quan trọng của rốt loạn hô hấp liên quan tới giấc ngủ, cũng như là một triệu chứng của tắc nghẽn mũi và có liên quan tới csc bệnh tim mạch và hen suyễn về đêm như là một yếu tố gây ra hoặc yếu tố gây bệnh được thể hiện qua âm thanh của tiếng ngáy ” Đây là một nghiên cứu cho thấy sự quan tâm, tính cấp bách của các ứng dụng hỗ trợ cho con người và đặc biệt là sức khỏe. Từ các công trình nghiên cứu về công nghệ lõi thì các sản phẩm áp dụng từ các bài toán gốc này ngày một phát triển hơn.

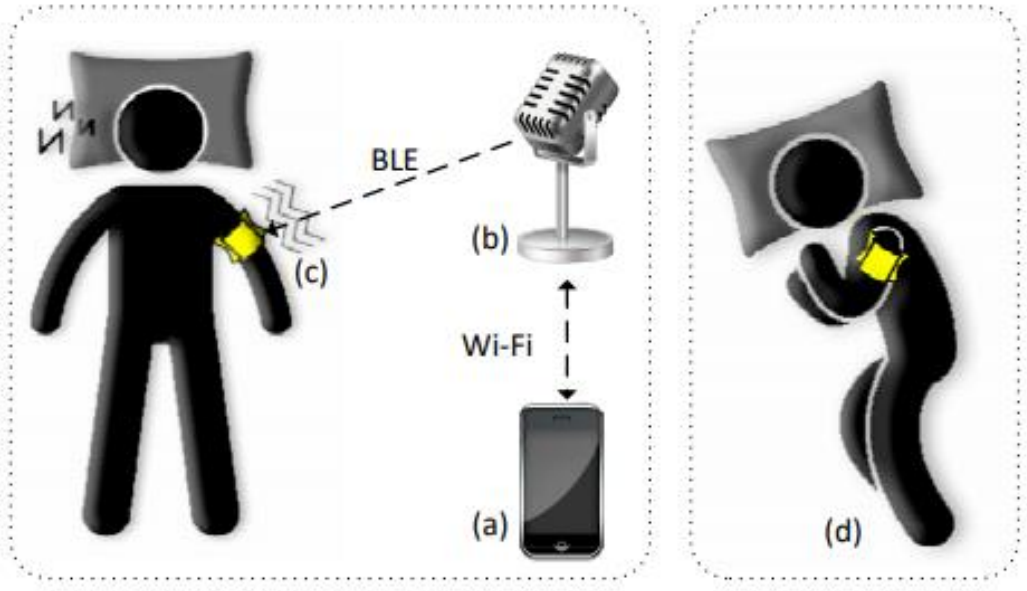
Với các ứng dụng áp dụng vào thực tiễn thì đang được cải tiến như trong ứng dụng tạo ra sản phẩm theo dõi tiếng ngáy với điện thoại thông minh[12] và sản phẩm áp dụng được mô phỏng với mô hình sử dụng như sau:



Hình 1.3. Mô hình theo dõi tiếng ngáy với thiết bị điện thoại thông minh

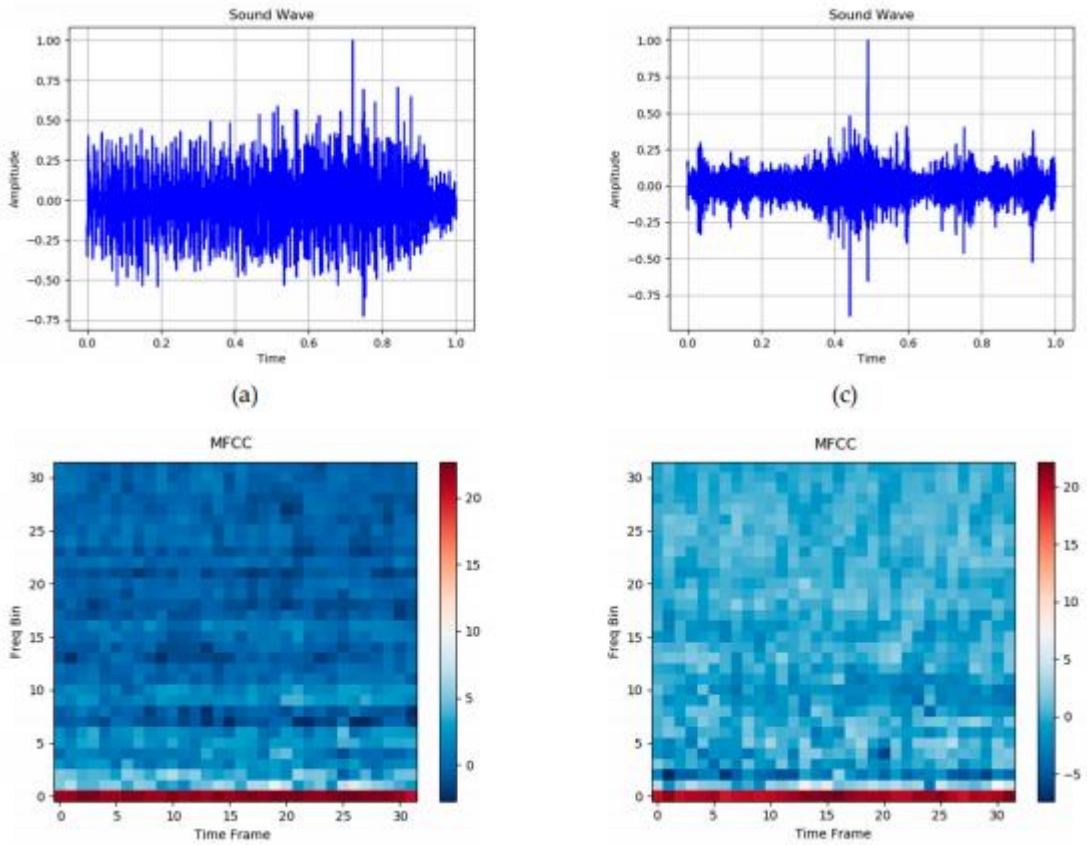
Mô hình theo dõi tiếng ngáy với điện thoại thông minh được thiết kế thành một mô hình hoàn chỉnh với việc thu nhận âm thanh từ điện thoại thông minh, xử lý, nhận biết được nội dung quan tâm và qua đó có thể phát hiện, kiểm soát chất lượng giấc ngủ với tiếng ngáy người sử dụng. Phân tích có tần số cao hơn 80Hz xảy ra ở bệnh nhân mắc OSA.

Sản phẩm sử dụng điện thoại thông minh được thực hiện thì các sản phẩm cải tiến theo phương hướng gọn nhẹ, dễ sử dụng cũng được phát triển lên, đại diện là các thiết bị có khả năng đeo trên người sử dụng[16], sau đây là mô hình của hệ thống



Hình 1.4. Mô hình theo dõi tiếng ngáy với các thiết bị đeo bên người

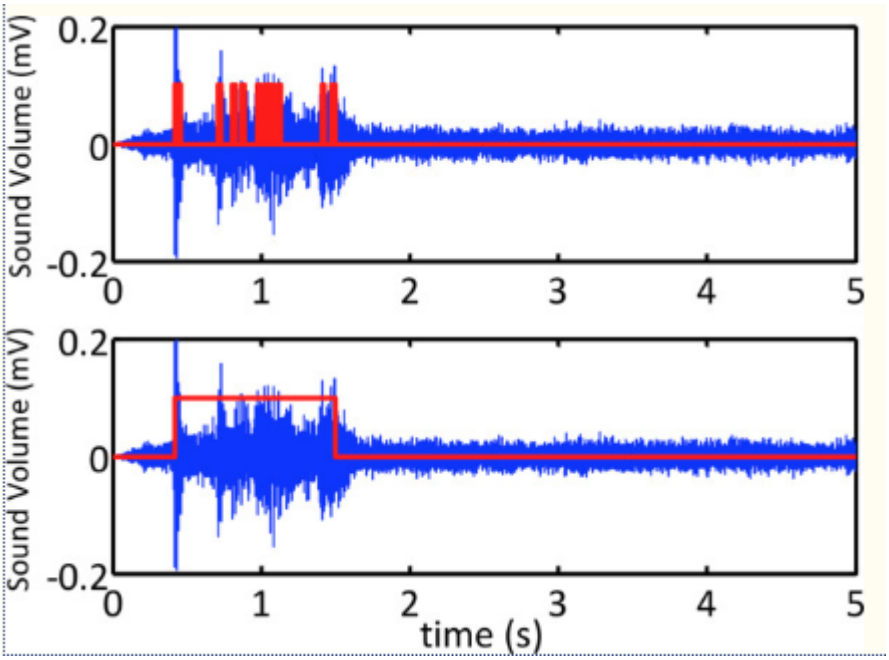
Sự cải tiến về mặt thiết bị thu nhận bên trong hệ thống dần dần được thân thiện với người sử dụng, nhưng về nguyên lý thì vẫn thông qua nhận dạng âm thanh, các âm thanh đầu vào dạng sóng thì sẽ được xử lý và trích các đặc trưng cơ bản của âm thanh, tùy từng bài toán mà các đặc trưng sẽ sự khác biệt với nhau.



Hình 1.5. Sóng âm thanh tiếng ngáy và khi theo dõi đặc trưng

Hình trên là một đề xuất của nghiên cứu [16] khi sử dụng việc theo dõi đặc trưng âm thanh với MFCC, các sóng âm thô sau khi thu nhận được thì sẽ được chuyển về dạng cửa sổ MFCC từ đó tìm ra được quy luật và phát hiện ra tiếng ngáy.

Ngoài việc sử dụng trích đặc trưng thông qua MFCC ra thì khi tiến hành thực nghiệm còn nhận ra là các âm thanh bên ngoài khi ngủ thì thường yên tĩnh, âm thanh khi đặt thiết bị thu gần người cần lấy mẫu gần nhất thì âm thanh ngáy có âm lượng là lớp nhất. Hình phía dưới có mô tả về sóng âm mà có tiếng ngáy thì âm lượng của âm thanh thu được là lớn nhất [17], và từ đó thì có thể lựa chọn được 1 ngưỡng âm lượng nào đó mà trên ngưỡng đó thì âm thanh đó là ngáy. Đôi khi có một số âm thanh môi trường gây nhiễu thì cần phải lọc các nhiễu này.



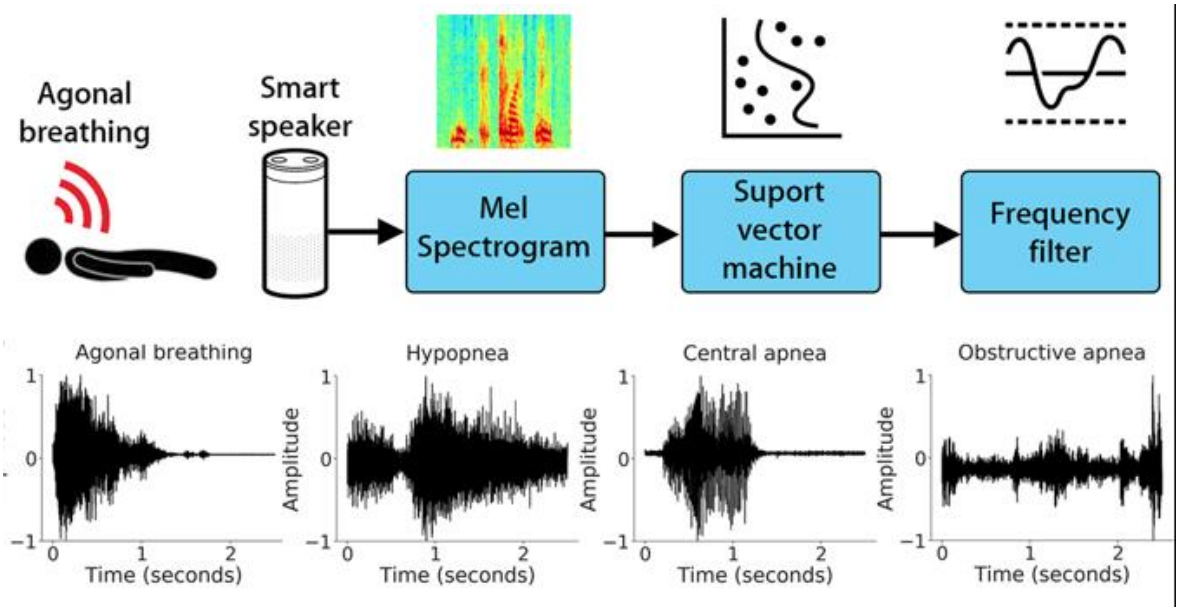
Hình 1.6. Âm thanh tiếng ngáy được thể hiện theo thời gian

Cùng với sự phát triển ngày càng nhanh của tốc độ xử lý máy tính, và các dữ liệu thu thập càng nhiều hơn thì việc phát hiện âm thanh có thể sử dụng các phương pháp học máy hiện đại vào bài toán phát hiện tiếng ngáy, phần sau sẽ trình bày thêm về các nghiên cứu về học máy trong phát hiện tiếng ngáy.

1.2.2 Mô hình học máy cổ điển trong phát hiện tiếng ngáy

Với các phương pháp nghiên cứu để phát hiện tiếng ngáy ở phần trước thì việc tìm ra quy luật hay một công thức nào đó để nhận dạng âm thanh chưa được biết là tiếng ngáy hay không thì thường khó khăn và nhiều khi không được chính xác. Kết hợp với một số giải thuật học máy như học có giám sát, học không giám sát hay học máy tăng cường.. đã đưa ra các mô hình để giải quyết vấn đề đó.

Theo trong nghiên cứu "Phân lớp tiếng ngáy: The Munich-Passau Snore Sound Corpus" [14] đã sử dụng bộ phân loại SVM để đào tạo và dùng để nhận dạng, phát hiện và phân lớp âm thanh. Các tiếng ngáy được phát hiện và phân lớp dựa trên cơ sở dữ liệu âm thanh và theo vị trí kích thích của âm thanh theo các tiêu chí được quy định.

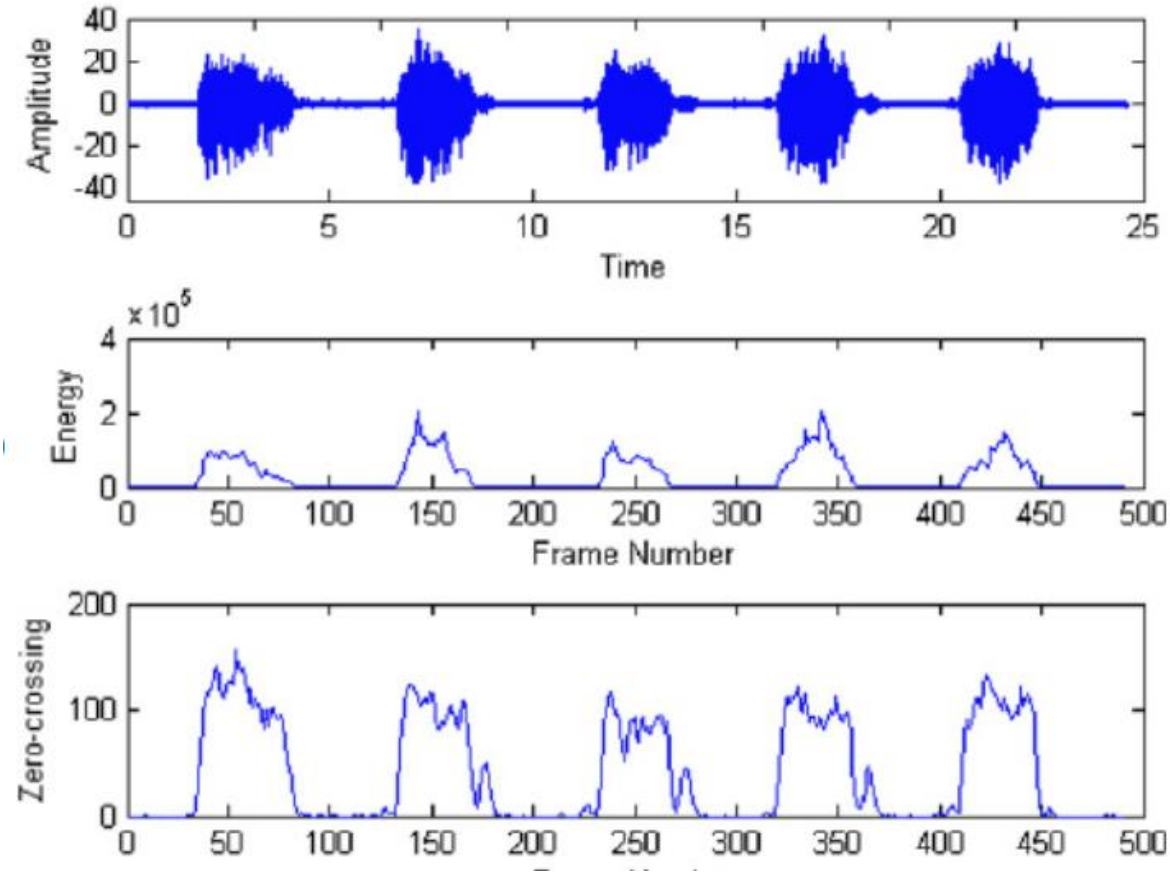


Hình 1.7. Mô hình về phát hiện tiếng ngáy với SVM

Hình phía trên mô tả về các bước có trong việc phát hiện tiếng ngáy sử dụng bộ phân loại SVM, từ các âm thanh gốc sẽ được đưa ra các vector hỗ trợ sau khi đó thì tiến hành lọc tần số và phân lớp đâu là lớp tiếng ngáy, đâu thì không phải. Phương pháp phân loại SVM được mô tả là nằm trong mô hình học nông của học máy, với tín hiệu nhận được thì sẽ được biểu thị thông qua các cửa số quang phổ và từ đó dựa vào những biểu thị từ trên các quang phổ này và so khớp với các lớp đã được huấn luyện thì có thể phát hiện ra được tiếng ngáy.

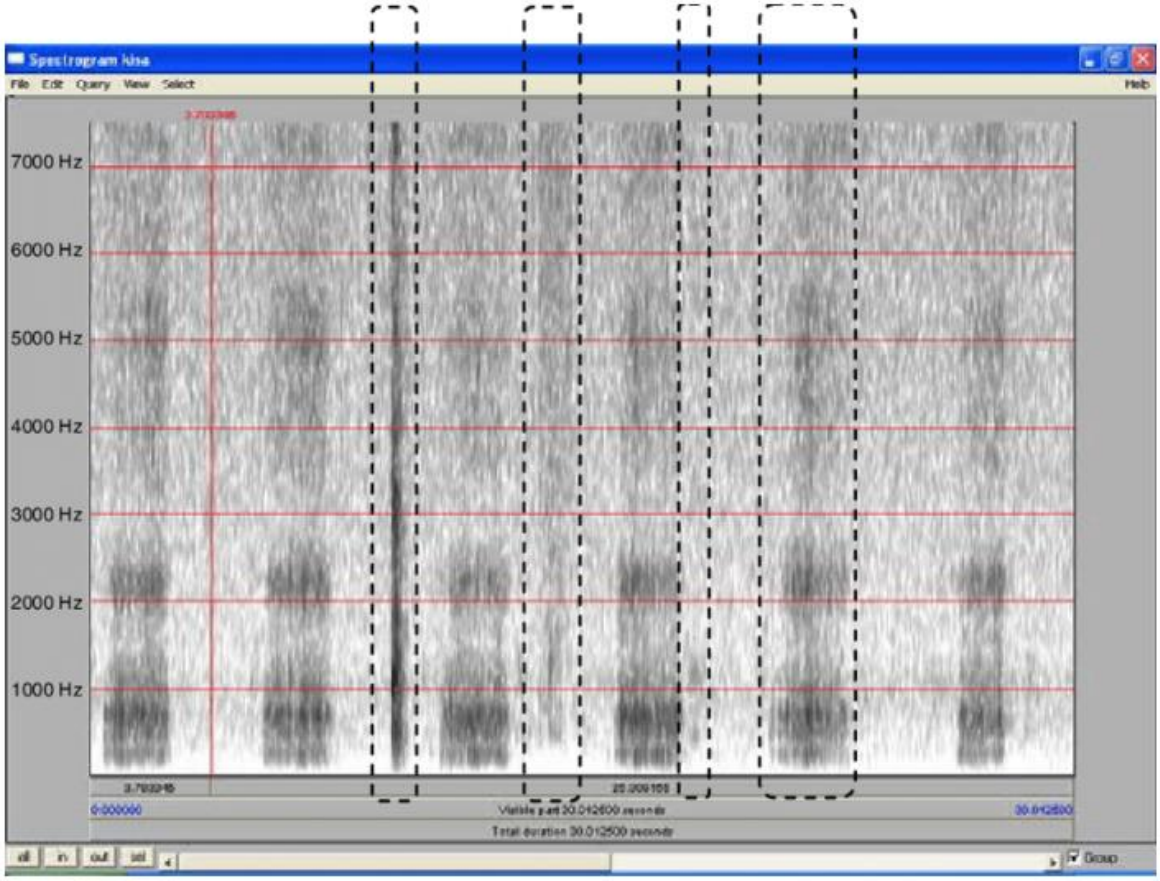
Một số công trình đã được trình bày trong những năm gần đây về các phương pháp phân tích âm thanh đa tính năng với mục đích phân loại và phân chia âm thanh giấc ngủ ngáy / không ngáy.

Trong quá trình nhận dạng và phân loại tỷ lệ vượt quá điểm không (ZCR)[20], được sử dụng để xác định ranh giới của các phân đoạn âm thanh.



Hình 1.8. Phân tách mẫu tín hiệu gốc, mẫu năng lượng và mẫu vượt quá không

Các tập đã được thể hiện một cách hiệu quả thành các tính năng phổ hai chiều bằng cách sử dụng phân tích thành phần chính và được phân loại là ngáy hoặc không ngáy với Hồi quy tuyến tính (RLR). Hệ thống đã được kiểm tra bằng cách sử dụng các nhãn được gán thủ công làm tài liệu tham khảo. Độ chính xác cho những người ngáy được tìm thấy là 97,3% khi hệ thống được huấn luyện chỉ sử dụng dữ liệu của những người ngáy ngủ. Nó giảm xuống 90,2% khi dữ liệu huấn luyện chứa cả bệnh nhân ngáy ngủ đơn giản và dữ liệu bệnh nhân OSA. Trong trường hợp phát hiện tập ngáy với bệnh nhân OSA, độ chính xác là 86,8%.



Hình 1.9. Biểu đồ biên độ của bản ghi mẫu

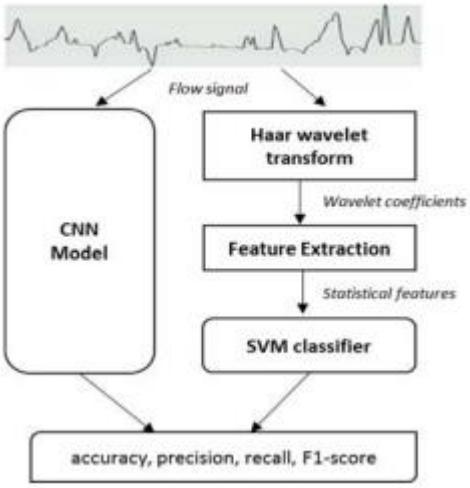
Một phương pháp phân tích tiếng ngáy được tạo ra bằng cách sử dụng biên độ, tần số và các sóng con. Cho thấy ngáy ngủ và ngáy lưỡi khác nhau đáng kể về tần số cao. Trong khi tần số cực đại trung bình trước đây quan sát tại 137Hz, thì tần số này nằm ở 1243 Hz. Trong số các đối tượng được đo ở tần số cao nhất của ngáy có nguồn gốc từ amidan và biểu mô lần lượt là 170Hz và 490Hz. Hơn nữa, cho thấy phổ âm thanh của tiếng ngáy sau khi ngưng thở, thực hiện phân loại bằng cách kết hợp các tính năng âm thanh khác nhau và tìm thấy các tính năng trong phổ là một trong những hoạt động tốt nhất.

Khi mạng neural càng ngày càng được phát triển khéo theo các mô hình mạng neural phát triển theo, và hiệu nặng, độ chính xác tốt hơn nhiều lần bằng chứng thông qua các bài đăng trên các cuộc thi về trí tuệ nhân tạo . Đây cũng là thời kỳ mà học sâu có những bước phát triển nhanh và mạnh, các mạng neural có thể thấy được trong

học sâu mà được nhiều người quan tâm như: mạng neural tích chập CNN, mạng neural hồi quy RNN, mạng neural kết hợp... đã trở thành phổ biến trong nghiên cứu học máy. Các ứng dụng đã phân nhánh tới nhiều nhiệm vụ khác nhau như nhận dạng hình ảnh, phần tích âm thanh, phân tích cảm xúc, phân tích ý kiến .v.v. Vậy sẽ có giả thiết sẽ sử dụng một mạng neural dựa trên âm thanh thu nhận ban đầu và đưa ra được âm thanh đó là gì, tác nhân gây ra âm thanh đó, sự kiện âm thanh dựa vào những dữ liệu đã đào tạo từ trước. Từ đó, có thể phát hiện âm thanh tiếng ngáy cũng như xác định nguyên nhân gây ngáy từ âm thanh.

1.2.3 Mô hình học sâu phát hiện tiếng ngáy

Trong nghiên cứu [15] trước đây thì thường sử dụng với thuật toán máy vector hỗ trợ SVM để phân lớp âm thanh. Đưa ra trong phương thức như sau:



Hình 1.10. Phương pháp học nông và học sâu (mạng neural) trong phát hiện âm thanh

Những nghiên cứu trong học sâu từ trước tới nay đã và đang được sử dụng để giải quyết nhiều bài toán về nhận dạng, phát hiện đặc biệt trong lĩnh vực thị giác máy tính. Vì đòi hỏi cần một lượng dữ liệu, thời gian, sức mạnh tính toán đáng kể, các nỗ lực nghiên nghiên cứu cách để tận dụng các mạng CNN được đào tạo trước cho các nhiệm vụ khác như mạng CNN được sử dụng trong các hệ thống nhận dạng. Cho đến nay, các nghiên cứu thực hiện để khám phá biểu diễn đặc trưng của âm thanh với

mạng CNN. Trong thử thách INTERSPEECH ComParE 2017 có một thử thách là xác định tiếng ngáy, đó cũng là tiền đề để phát triển các ứng dụng khai thác âm thanh ngáy. Để phát hiện và phân loại âm thanh thông qua phổ của âm thanh dựa trên học sâu là một lĩnh vực nghiên cứu mới.

Bài toán phát hiện tiếng ngáy cho thấy được một trong nhiều ứng dụng của phát hiện âm thanh[16]. Bài toán phát hiện tiếng ngáy dựa trên học sâu nhằm phát hiện âm thanh là tiếng ngáy và từ đó có thể phân tích được chất lượng giấc ngủ. Phát hiện tiếng ngáy dựa trên học sâu là sử dụng mô hình mạng neural tích chập (CNN) để phát hiện và phân tích đặc trưng của tiếng ngáy.

Mạng neural nhân tạo, Artificial Neural Network (ANN) là một mô hình xử lý thông tin phỏng theo cách thức xử lý thông tin của các hệ nơron sinh học. Nó được tạo nên từ một số lượng lớn các phần tử (nơron) kết nối với nhau thông qua các liên kết (trọng số liên kết) làm việc như một thể thống nhất để giải quyết một vấn đề cụ thể nào đó. Một mạng nơron nhân tạo được cấu hình cho một ứng dụng cụ thể (nhận dạng mẫu, phân loại dữ liệu,...) thông qua một quá trình học từ tập các mẫu huấn luyện. Về bản chất học chính là quá trình hiệu chỉnh trọng số liên kết giữa các nơron.

Trong một nghiên cứu khác là “Sử dụng mạng LSTM để mô hình hóa chuỗi âm thanh”[23] đã đưa ra kết quả là tỷ lệ mẫu càng cao, dự đoán càng tốt vì các chuỗi âm thanh dài có trình tự chuyển đổi thường xuyên hơn, tốc độ chuyển đổi âm thanh được xử lý mịn hơn so với các chuỗi âm thanh ngắn. Hay trong một công trình khác như : “Phân loại sự kiện âm thanh bằng cách sử dụng các mạng thần kinh sâu”[24] đã sử dụng một mạng neural sâu, sử dụng GMM để lấy điểm đặc trưng nằm trong lớp sau đó sử dụng điểm đó để phân loại các âm thanh

Đến nay, một số bài báo có cách tiếp cận mạng neural tích chập trong vấn đề của Phân loại âm thanh đàn (ASC). Cách tiếp cận việc xử lý âm thanh dưới dạng ảnh phổ có thể kết hợp được những ưu điểm của xử lý hình ảnh và âm thanh từ đó mang lại hiệu quả cao trong việc phát hiện và nhận dạng.

1.2.4 Đánh giá các nghiên cứu

Các nghiên cứu gần đây về học máy hay học sâu đã trở thành xu thế nghiên cứu của các nhà khoa học trên thế giới và trong nước. Cùng với đó là một xu thế mới trong việc phát triển các ứng dụng khác nhau mà có sự hỗ trợ của học máy/học sâu để giải quyết những bài toán mà trước đây vô cùng phức tạp hoặc mất nhiều chi phí. Các kết quả nghiên cứu của học máy đang, đang và sẽ giải quyết vấn đề của các ứng dụng mà có thể hỗ trợ trong đời sống con người như các nghiên cứu về âm thanh ngáy đã là cảm hứng để tạo ra một chuỗi các ứng dụng theo dõi, nhận dạng hành vi con người từ đó tạo nên một cuộc sống tốt đẹp hơn cho con người.

Mặc dù, các ứng dụng của trí tuệ nhân tạo hay học sâu đang dần dần cho thấy tính ưu việt nhưng mà cũng có một số bất cập trong việc thực hiện như các vấn đề về nguồn dữ liệu, các chính sách thu thập thông tin, hành vi của người dùng khi mà thực hiện các ứng dụng. Các ứng dụng về học máy/ học sâu đòi hỏi có một lượng dữ liệu huấn luyện đủ lớn để có thể cho ứng dụng ngày một thông minh, thông suốt hơn. Từ những ngày đầu, các ứng dụng của trí tuệ nhân tạo đã giải quyết các vấn đề đơn nhất, và đến tận ngày nay các ứng dụng này đã phát triển một cách vượt trội qua các ứng dụng phức tạp đòi hỏi việc xử lý thông minh.

1.3 Kết luận chương

Chương 1 đã giới thiệu tổng quan về bài toán phát hiện tiếng ngáy. Tìm hiểu bài toán phát phân loại âm thanh và giới thiệu bài toán phát hiện tiếng ngáy, kèm theo đó là các nghiên cứu liên quan từ các ứng dụng, giải pháp mà được thực hiện từ bài toàn, các mô hình giải quyết bài toán, và các đánh giá về các nghiên cứu qua đó đưa ra những vấn đề cần làm rõ và giải quyết trong luận văn.

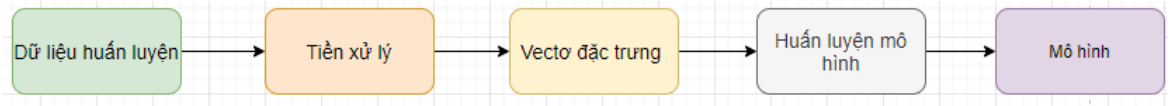
Trong chương 2, luận văn sẽ trình bày về hướng giải quyết cho bài toán phát hiện tiếng ngáy, các bước tiến hành khi giải bài toán nhận dạng, phát hiện tiếng ngáy, các đặc trưng của âm thanh, các thành phần xử lý âm thanh và đi sâu hơn trình bày về phương pháp sẽ áp dụng để giải quyết bài toán. Đây cũng là nền tảng cho phương hướng của việc thực nghiệm giải quyết bài toán đã đề ra.

CHƯƠNG 2: PHƯƠNG PHÁP PHÁT HIỆN VÀ THEO DÕI TIẾNG NGÁY

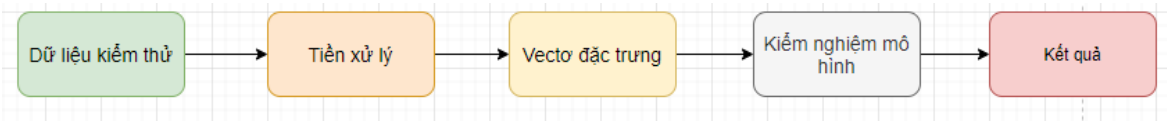
Trình bày một số phương pháp học sâu có có tốc độ tính toán nhanh phù hợp với bài toán phát hiện và theo dõi tiếng ngáy. Các âm thanh được trích rút đặc trưng và đi qua các mô hình học sâu như CNN hoặc mô hình hồi quy RNN.

2.1 Phương pháp giải quyết bài toán

Để giải quyết bài toán phát hiện và theo dõi tiếng ngáy từ "âm thanh ngáy" đầu vào, mục tiêu cần phải phân lớp và đưa những âm thanh này về lớp "Âm thanh ngáy" và những âm thanh còn không phải âm thanh ngáy thì sẽ đưa về lớp "Không phải âm thanh ngáy". Luận văn đã tham khảo và tìm hiểu được các bước thực hiện để xây dựng phương pháp phát hiện và theo dõi tiếng ngáy và được chia làm 2 giai đoạn: huấn luyện và kiểm thử.



Hình 2.1. Giai đoạn huấn luyện mô hình



Hình 2.2. Giai đoạn kiểm thử mô hình

Hai giai đoạn huấn luyện và kiểm thử trong phát hiện tiếng ngáy được mô tả như các hình phía trên. Các bước thực hiện trong luận văn sẽ gồm các bước từ trái sang phải sau:

1. Chia dữ liệu thành 2 phần: dữ liệu huấn luyện và dữ liệu kiểm thử
2. Tiền xử lý dữ liệu huấn luyện và kiểm thử trước khi lựa chọn ra các vector đặc trưng, điều này sẽ loại bỏ đi các thông tin có giá trị thấp.

3. Vectơ đặc trưng trích đặc trưng cho tập dữ liệu đã qua tiền xử lý, tại đây sẽ có các đặc trưng riêng của các bài toán được thể hiện ra.
4. Áp dụng các mô hình học sâu (mô hình CNN, mô hình LSTM, mô hình CNN-LSTM) để giải quyết bài toán và so sánh với mô hình học nông
5. Đưa ra mô hình sau khi huấn luyện và kết quả sau khi kiểm thử qua mô hình, từ đó đưa ra được kết quả và đánh giá bài toán

Tại bước 1, luận văn sẽ áp dụng phương pháp cross validation và chia dữ liệu thành 2 phần gồm phần dữ liệu huấn luyện 90%, phần dữ liệu kiểm thử 10% . Cụ thể về phương pháp cross validation sẽ được luận văn trình bày tại mục 3.1 về thu thập dữ liệu.

Trong bước 2, tiền xử lý, dữ liệu đầu vào là âm thanh cần phải loại bỏ các yếu tố dư thừa của dữ liệu như các đoạn không có thu nhận được âm thanh...

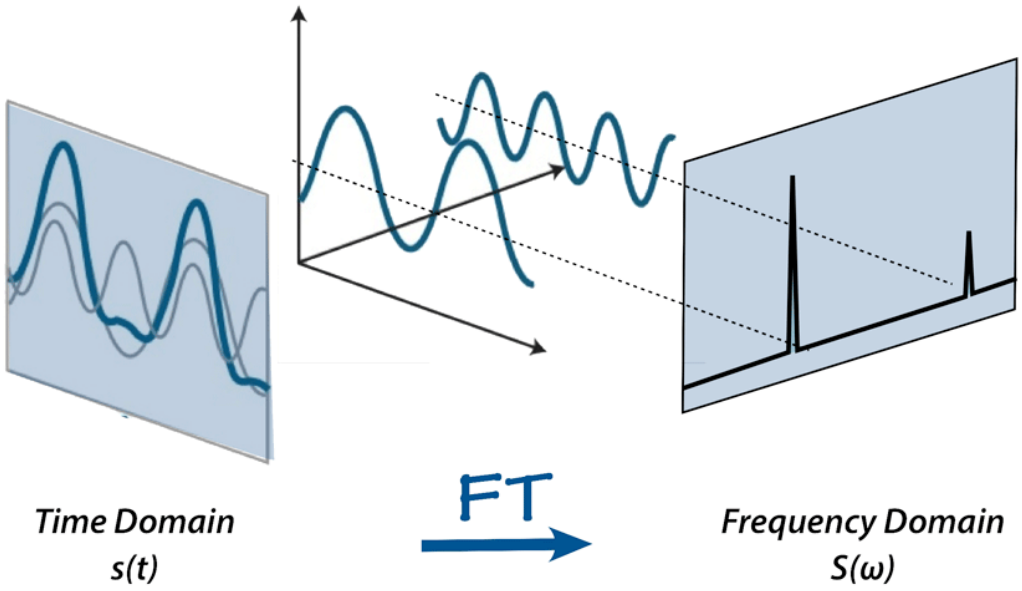
Các phần tiếp theo của chương sẽ trình bày chi tiết về các phương pháp, mô hình và đề xuất lựa chọn và áp dụng vào việc phát hiện tiếng ngáy trong hệ thống phát hiện, theo dõi tiếng ngáy.

2.2 Xử lý âm thanh

2.2.1 Biến đổi Fourier (FT)

Âm thanh là một chuỗi tín hiệu dài biến thiên theo thời gian, nhưng hàm lượng thông tin trong đó không nhiều. Và âm thanh được kết hợp từ các sóng có tần số khác nhau, vậy hãy suy nghĩ ngược lại, tại sao không tìm cách phân giải 1 đoạn âm thanh được biểu diễn trong miền thời gian thành các sóng với tần số và biên độ cụ thể trong miền tần số. Điều đó được minh hoạ trong hình 2.3 phía dưới miêu tả đoạn âm thanh trong miền thời gian được kết hợp từ 2 sóng tuần hoàn. Do 2 sóng này có tính chất tuần hoàn, thay vì phải lưu giá trị theo thời gian, thì chỉ cần lưu lại tần số, biên độ và pha giao động của các sóng âm thanh này. Kết quả sau cách biến đổi thì sẽ nhận được biểu diễn giàu thông tin hơn so với cách biểu diễn thông tin thông thường.

Như vậy, với biến đổi Fourier từ dữ liệu nghèo thông tin đã được chuyển đổi thông tin từ miền thời gian sang miền tần số. Ngược lại, biến đổi Fourier ngược (inverse Fourier transform) sử dụng để chuyển đổi thông tin từ miền tần số về miền thời gian. Biến đổi Fourier được ứng dụng nhiều trong lĩnh vực xử lý tín hiệu (âm thanh, ảnh, thông tin) do những đặc điểm và lợi ích của phương pháp này.



Hình 2.3. Phép biến đổi Fourier

Công thức biến đổi Fourier cho hàm f(x) liên tục trong công thức (2.1) :

$$f(x) = \int_{-\infty}^{\infty} F(k) e^{2\pi i k x} dk \quad \textbf{(2.1)}$$

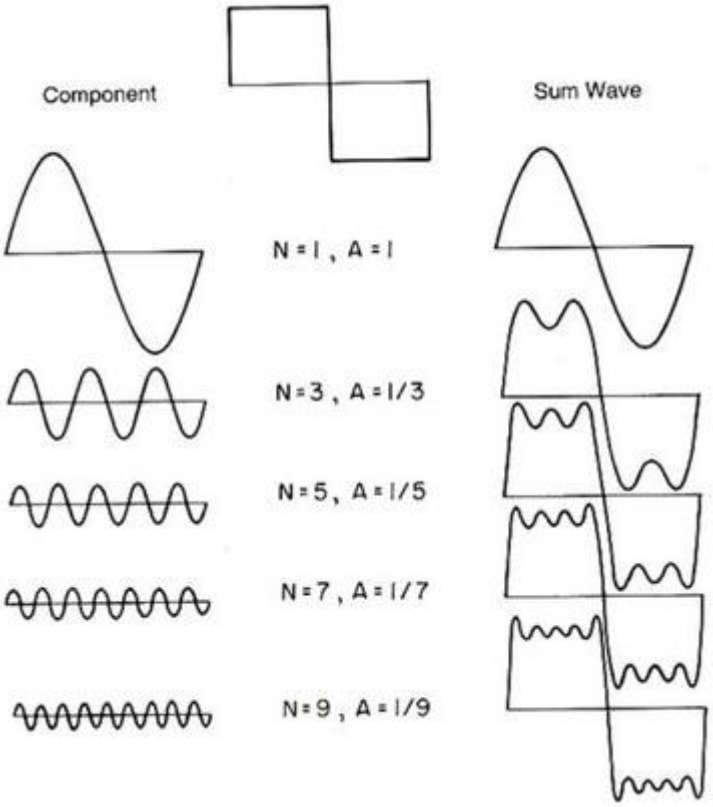
$$F(k) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i k x} dx \quad \textbf{(2.2)}$$

Trong đó, F(k) là công thức biến đổi fourier ngược trong công thức 2.2

Công thức biến đổi Fourier rời rạc (DFT) trong công thức 2.3

$$X(k) = \sum_{n=-\infty}^{\infty} x[n] e^{-jkn} \quad \textbf{(2.3)}$$

Biến đổi Fourier là phép biến đổi đối xứng, tức một thông tin được biến đổi Fourier từ miền thời gian sang miền tần số, có thể biến đổi Fourier ngược để chuyển đổi thông tin từ miền tần số lại về miền thời gian như dạng ban đầu của thông tin. Dưới đây là minh hoạ cho sóng vuông được phân giải thành các sóng Sin. Có thể thấy với giá trị n càng cao, độ chính xác càng lớn.



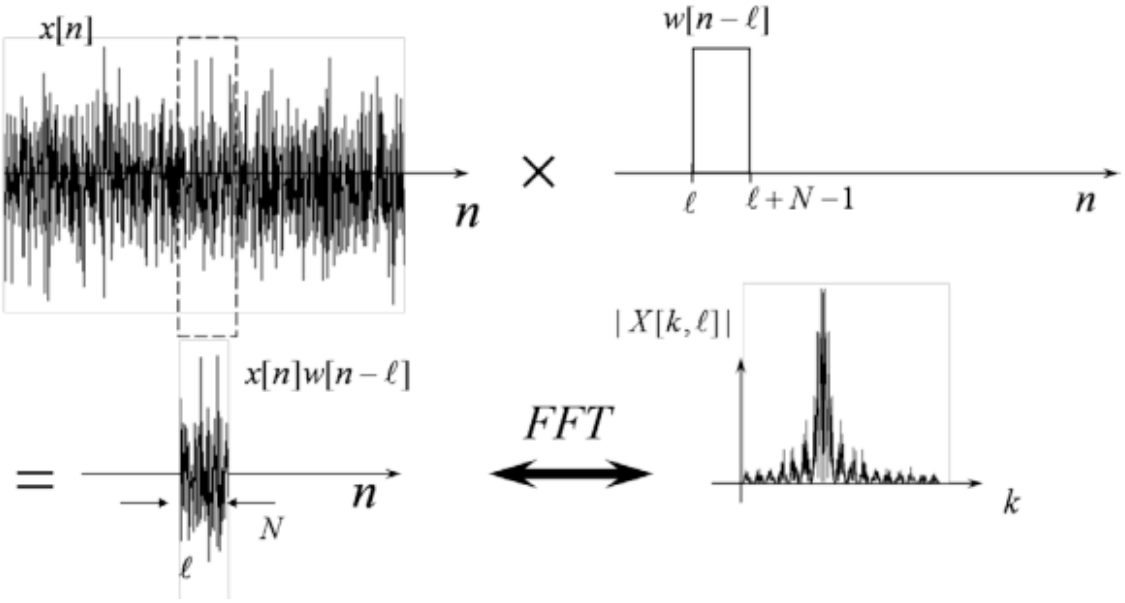
Hình 2.4. Biến đổi Fourier rời rạc

Phép biến đổi Fourier do đặc trưng là việc chuyển đổi của các sóng tuần hoàn trên miền thời gian sang miền tần số nên thường được dùng cho phân tích các tín hiệu trong đó có tín hiệu âm thanh. Tuy nhiên, phép biến đổi này có hạn chế là 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 Fourier thời gian ngắn STFT (Short time Fourier transform). Theo đó, tín hiệu được chia thành các khoảng nhỏ và được biến đổi Fourier trong từng khoảng đó.

2.2.2 Biến đổi Fourier thời gian ngắn (STFT)

Nguyên tắc của phương pháp này là phân chia tín hiệu ra thành từng đoạn đủ nhỏ sao cho có thể xem tín hiệu trong mỗi đoạn là tín hiệu ổn định, sau đó, thực hiện biến đổi Fourier trên từng đoạn tín hiệu này.

Các giá trị DFT và IDFT được tính toán hiệu quả bằng thuật toán FFT.



Hình 2.5. Mô tả STFT được biến đổi từ FT

DFT làm việc với tín hiệu rời rạc có giới hạn độ dài (n). Thực tế, rất nhiều tín hiệu trong khoảng thời gian dài. Rất khó tính toán DFT công thức (2.3) với n rất lớn. Để giải quyết vấn đề này, ý tưởng sử dụng biến đổi Fourier thời gian ngắn STFT (Short Time Fourier Transform) dựa trên phép biến đổi Fourier rời rạc, chia tín hiệu thành nhiều đoạn nhỏ, được thực hiện để giải quyết từng phần. Trong đó, tín hiệu với độ dài tùy ý được chia thành các khối gọi là frame và DFT áp dụng cho từng frame. Frame được hình thành bằng cách nhân tín hiệu gốc với hàm cửa sổ. Thông thường độ dài frame khoảng 10 đến 20 ms được sử dụng vào phân tích không gian

2.2.3 Phương pháp hệ số biểu diễn phổ của phổ (MFCC)

MFCC (Mel Frequency Cepstral Coefficients) là các hệ số biểu diễn phổ của phổ của đoạn âm thanh được xem xét. Phương pháp này được thực hiện dựa trên việc biến đổi chuyển dữ liệu âm thanh đầu vào (đã được biến đổi Fourier cho phổ âm thanh) về thang đo tần số Mel, một thang đo được cho là có sự thể hiện tốt hơn sự nhạy cảm của tai người đối với âm thanh.

MFCC có thể hình dung được tính theo luồng xử lý: cắt chuỗi tín hiệu âm thanh thành các đoạn ngắn bằng nhau (25ms) và overlap lên nhau (10ms). Mỗi đoạn âm thanh này được biến đổi, tính toán để thu được 39 features. Mỗi danh sách 39 đặc trưng có tính độc lập cao, ít nhiễu, đủ nhỏ để đảm bảo tính toán, đủ thông tin để đảm bảo chất lượng cho các thuật toán nhận dạng.

. Kỹ thuật trích chọn đặc trưng này gồm các bước biến đổi liên tiếp, trong đó kết quả đầu ra của bước biến đổi trước sẽ là giá trị đầu vào của bước biến đổi sau.

Mạch tăng cường

Do các âm ở tần số thấp có mức năng lượng cao, các âm ở tần số cao lại có mức năng lượng khá thấp. Trong khi đó, các tần số cao này vẫn chứa nhiều thông tin về âm vị.. Do đó, nhấn mạnh trước được sử dụng để tăng năng lượng từ thấp đến cao, được thể hiện trong công thức 2.4

$$\tilde{x}(n) = x(n) - \alpha x(n-1) \quad \textbf{(2.4)}$$

Trong đó x(n) là tín hiệu và, n là số lượng mẫu lấy, và α là tham số bộ lọc số bậc thấp để phổ âm nhận được giá trị đồng đều, giảm ảnh hưởng gây ra cho các xử lý tín hiệu bước (n) phía sau và nhận giá trị trong khoảng từ 0.9 tới 1.0.

Khung

Khung được sử dụng để chia $\tilde{x}(n)$ thành N thời gian của khung với các khung liền kề được phân tách bằng dịch chuyển khung P. Giả định rằng tồn tại một thuộc tính tín hiệu không đổi trong mỗi khung, tuy nhiên, việc phân chia tín hiệu đột ngột

(ở cả hai đầu) bằng cách tạo khung dẫn đến mất thông tin hoặc mất đặc trưng. Dựa trên thời gian đo N, phạm vi từ 10 đến 30ms và thời gian trùng khớp < 0,5. Mỗi khung được ước lượng giá trị như sau: Công thức (2.5) là tính số lượng của khung trong tín hiệu. công thức (2.6) thể hiện giá trị ước lượng của khung f

$$\eta = \frac{p + [\tau - N]}{p} \quad \textbf{(2.5)}$$

$$\widetilde{f_j}(n) = \tilde{x}(p_j + n) \quad \textbf{(2.6)}$$

Trong đó có $0 \leq n \leq N-1, 0 \leq j \leq \eta.\eta$ là số lượng của khung trong tín hiệu, τ là tổng số mẫu của tín hiệu đó.

Cửa sổ Hamming

Cửa sổ Hamming được sử dụng để tránh quá trình mất thông tin có thể xảy ra trong quá trình đóng khung. Hơn nữa, nó được sử dụng để ngăn chặn sự cắt giảm liên tục khung hình ở cả hai đầu của tín hiệu (âm thanh ngáy). Để thực hiện cửa sổ trên tín hiệu, các khung được thực hiện bởi cửa sổ hamming theo công thức (2.7) như sau

$$f_j = \omega(n) \times f_j(n), 0 \leq n \leq N-1 \quad \textbf{(2.7)}$$

$$\omega(n) = \left[-\beta\cos\left(\frac{2\pi n}{N-1}\right) - (\beta - 1)\right] \quad \textbf{(2.8)}$$

$$0 \leq n \leq N-1$$

Giá trị của β tham số tác động tới việc ngăn chặn sự cắt giảm liên tục tín hiệu ở các khung và được đặt giá trị là 0.46

Biến đổi Fourier nhanh

Biến đổi Fourier nhanh sử dụng tín hiệu liên tục và định kỳ trong một khung và chuyển đổi từng tín hiệu trong miền thời gian sang miền tần số

Biến đổi Fourier nhanh (FFT) được thực hiện để chuyển đổi mỗi khung với N mẫu từ miền thời gian sang miền tần số. Tín hiệu gốc cần được thực hiện biến đổi Fourier qua bộ lọc thông dải để xử lý độ lệch tần số Mel. Biến đổi Fourier chuẩn

không được sử dụng do tín hiệu âm thanh không xác định trên toàn miền thời gian. Thông thường hay sử dụng biến đổi DFT.

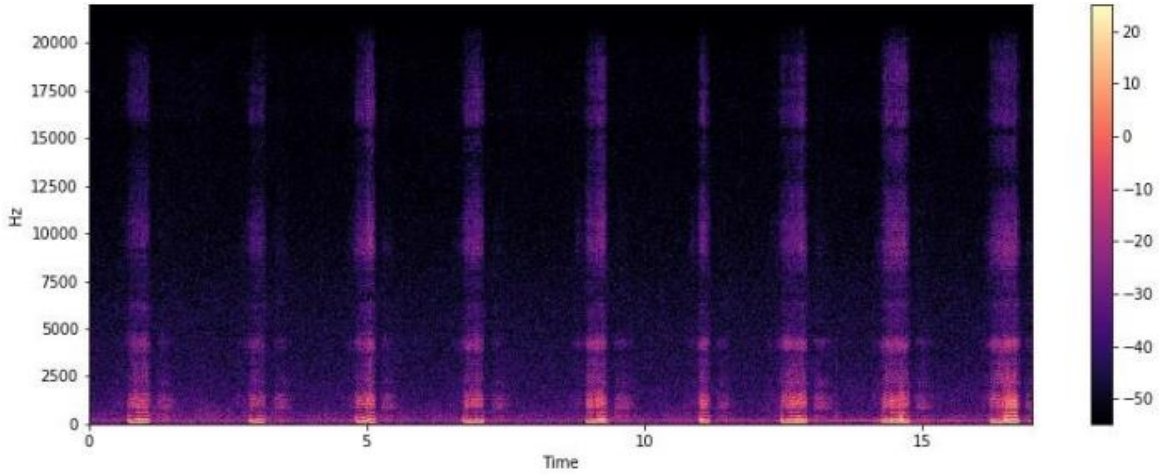
Mel filter DCT

Để mô tả chính xác sự tiếp nhận tần số của hệ thống thính giác, người ta xây dựng một thang khác – thang Mel. Việc chuyển đổi sang miền tần số Mel làm nhẵn phổ và làm nổi lên các tần số cảm thụ có nghĩa. Biến đổi Fourier lên tín hiệu qua bộ lọc thông dải để làm đơn giản phổ mà không làm mất dữ liệu. Điều này được thực hiện bằng cách tập hợp các thành phần phổ thành một dải tần số. Phổ được làm đơn giản hóa do sử dụng một dàn bộ lọc để tách phổ thành các kênh. Các bộ lọc được đặt cách đều nhau trên thang Mel và lấy logarit trên thang tần số, các kênh có tần số thấp là không gian tuyến tính trong khi các kênh có tần số cao là không gian logarit.

Thang tần số Mel được định nghĩa như sau với giá trị của f được lấy từ công thức (2.7) ta được giá trị của tần số Mel trong công thức (2.9)

$$Mel(f) = 2595 \log_{10}\left(\frac{f}{700} + 1\right) \quad \textbf{(2.9)}$$

Trong giá trị của Mel(f) thu được sau công thức (2.9) thì sẽ được biểu diễn trên biểu đổ phổ như trong hình (2.6)



Hình 2.6. Biểu diễn của MFCC từ âm thanh tiếng ngáy.

Sau khi tín hiệu âm thanh được biểu diễn phổ phổ của phổ âm thanh thông qua MFCC thì được biểu diễn như trong hình 2.6. Trên hình 2.6 có thể thấy được những thời gian ngáy thì có một đường thẳng kéo dài từ dưới đi lên trên.

2.3 Mô hình học nông

2.3.1 Trích đặc trưng của âm thanh

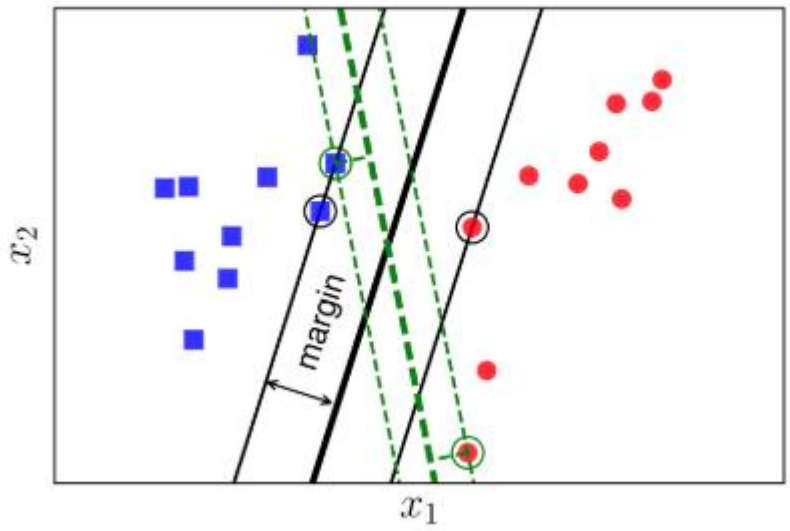
Trích chọn đặc trưng bao gồm hai phần: tách/trích xuất đặc trưng (feature extraction) và lựa chọn đặc trưng (feature selection). Trích chọn đặc trưng nhằm rút gọn các tín hiệu thành các đặc trưng để phân biệt các hoạt động đang có loại bỏ các thông tin dư thừa và sau đó được sử dụng làm dữ liệu đầu vào cho bước phân lớp. Tùy thuộc vào từng hệ thống cụ thể mà lựa chọn đặc trưng có thể được thực hiện hoặc không.

Các đặc trưng có thể được trích xuất tự động hoặc dựa trên tri thức chuyên gia. Tập các đặc trưng có được từ dữ liệu được gọi là không gian đặc trưng. Nói chung, khi các hoạt động được phân tách càng rõ ràng trong không gian đặc trưng thì hiệu suất nhận dạng của hệ thống càng cao. Lý tưởng nhất là các đặc trưng của cùng hoạt động có thể nhóm thành một nhóm trong không gian đặc trưng và ngược lại các đặc trưng của các hoạt động khác nhau cần phân biệt càng xa càng cho ra kết quả tốt.

Qua quá trình tham khảo và tìm hiểu, luận văn nhận thấy các đặc trưng cần trích để phân lớp âm thanh thì được sử dụng gồm đặc trưng biến đổi Fourier thời gian ngắn và sử dụng hệ số biểu diễn phổ của phổ âm thanh.

2.3.2 Mô hình học máy SVM

Mô hình học máy SVM là mô hình kinh điển trong bài toán phân loại. Tư tưởng của SVM là định nghĩa ra một siêu mặt phẳng có thể phân tách các tập dữ liệu cần phân loại sao cho khoảng cách (margin) từ siêu mặt phẳng đến các tập cần phân loại là tương đương nhau và lớn nhất. Thuật toán SVM ban đầu được thiết kế để giải quyết bài toán phân lớp nhị phân với ý tưởng chính như sau:



Hình 2.7. Khoảng cách của 2 phân lớp bằng nhau và lớp nhất

Trong không gian Euclid có cách tính khoảng cách từ một điểm có tọa độ (x_0, y_0) tới đường thẳng có phương trình $w_1x + w_2y + b = 0$ được tính bằng:

$$h = \frac{|w_1x_0 + w_2y_0 + b|}{\sqrt{w_1^2 + w_2^2}} \quad \textbf{(2.10)}$$

Trong không gian ba chiều khoảng cách từ một điểm có tọa độ (x_0, y_0, z_0) tới một mặt phẳng có phương trình $w_1x + w_2y + w_3z + b = 0$ được tính bằng

$$h = \frac{|w_1x_0 + w_2y_0 + w_3\,z_0 + b|}{\sqrt{w_1^2 + w_2^2 + w_3^2}} \quad \textbf{(2.11)}$$

Nhận thấy nếu bỏ dấu giá trị tuyệt đối thì có thể xác định được điểm đang xét nằm phía nào của đường thẳng hay mặt phẳng. Từ đó, có thể tổng quát cho rằng nếu biểu thức bỏ dấu giá trị tuyệt đối thì những điểm nào cùng mang dấu với nhau thì nằm cùng phía với nhau và có được công thức tính khoảng cách trong không gian có n số chiều mà trong đó có khoảng cách được tính bằng:

$$h = \frac{|w^Tx_0 + b|}{\sqrt{\sum_{i=1}^{n} w_i^2}} \quad \textbf{(2.12)}$$

Giả sử với xét các cặp dữ liệu đào tạo là $(x_1, y_1), (x_2, y_2)$ ..... (x_n, y_n) tượng trưng cho dữ liệu đầu vào của một điểm dữ liệu

Bài toán SVM trở thành đi tìm w và b sao cho khoảng cách này đạt giá trị lớn nhất.

Đối với bài toán phân lớp mà có số lớp n > 2 thì có thể sử dụng bằng cách chuyển bài toán phân lớp nhị phân giữa 1 lớp và (n-1) lớp còn lại. Tức là sẽ phải thực hiện n lần giữa phân lớp giữa lớp thứ i và (n-i) lớp còn lại.

Khoảng cách từ chiều tới mặt

2.3.3 Đánh giá mô hình học máy SVM

Mô hình học máy SVM có những ưu điểm như sau:

- SVM là một phương pháp phân lớp khá phổ biến, SVM có nhiều ưu điểm trong đó việc tính toán hiệu quả trên các tập dữ liệu lớn.
- Xử lý trên không gian có số chiều cao: SVM là một công cụ tính toán hiệu quả trong không gian chiều cao, trong đó đặc biệt áp dụng cho các bài toán phân loại văn bản và phân tích quan điểm khi mà có số chiều lớn
- Tiết kiệm bộ nhớ: Do chỉ có một tập hợp con của các điểm được sử dụng trong quá trình huấn luyện và ra quyết định thực tế cho các điểm dữ liệu mới nên chỉ có những điểm cần thiết mới được lưu trữ trong bộ nhớ khi thực hiện
- Tính linh hoạt - phân lớp là phi tuyến tính. Khả năng áp dụng các phương pháp tính mới cho phép linh động giữa các phương pháp tuyến tính và phi tuyến tính từ đó khiến cho hiệu suất phân loại lớn hơn.

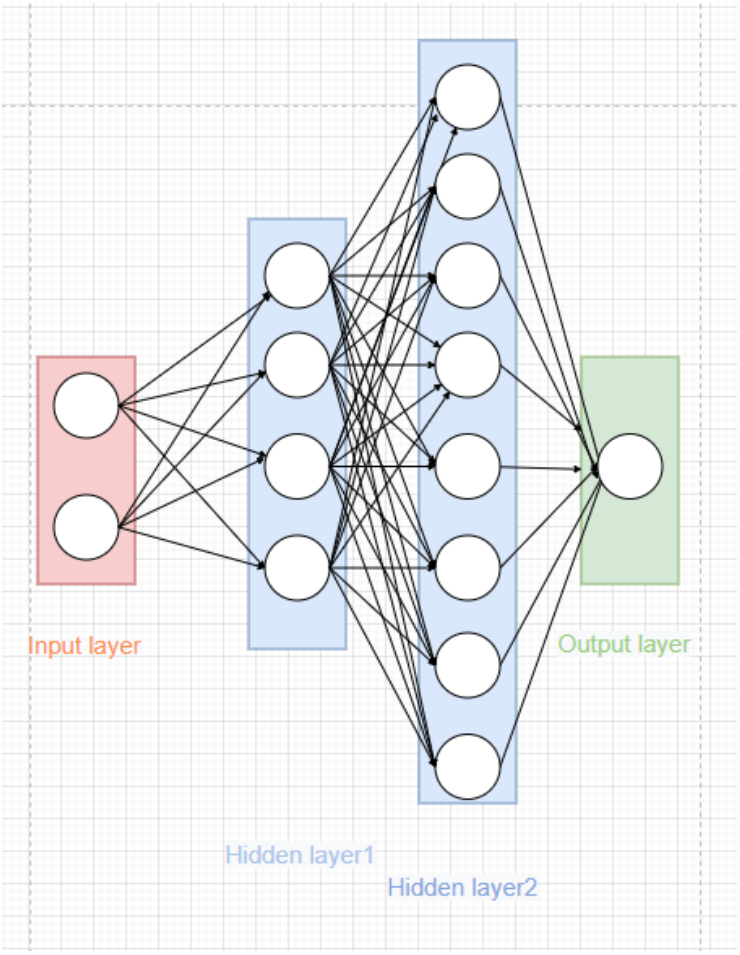
Nhược điểm của SVM

- Bài toán số chiều cao: Trong trường hợp số lượng thuộc tính của tập dữ liệu lớn hơn rất nhiều so với số lượng dữ liệu thì SVM cho kết quả chưa được tốt

- Chưa thể hiện rõ tính xác suất: Việc phân lớp của SVM chỉ là việc cố gắng tách các đối tượng vào hai lớp được phân tách bởi siêu phẳng SVM. Điều này chưa giải thích được xác suất xuất hiện của một phần tử có trong lớp. Tuy nhiên hiệu quả của việc phân lớp có thể được xác định dựa vào khái niệm khoảng cách từ điểm dữ liệu mới đến siêu phẳng phân lớp mà luận văn đề cập ở trên.

2.4 Mô hình CNN cho phát hiện tiếng ngáy

Mạng neural được lấy cảm hứng từ cấu tạo về não của con người, khi mà từ thông tin tiếp nhận và được xử lý khi đi qua các neural và khi đến cuối của neural thì thông tin đã được xử lý xong hoàn toàn. Mô hình mạng neural được mô tả thông qua hình sau:



Hình 2.8. Mô hình về mạng neural

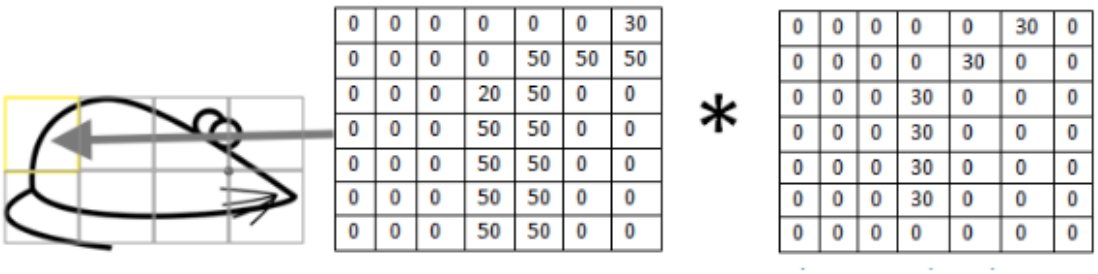
Lớp đầu tiên là lớp input, các layer ở giữa được gọi là các lớp ẩn, lớp cuối cùng là lớp đầu ra. Các hình tròn được gọi là node

2.4.1 Kiến trúc mạng CNN

Mạng neural tích chập (CNN) là một trong những mô hình mạng neural học sâu tiên tiến giúp cho việc xây dựng những hệ thống trí tuệ nhân tạo thông minh với độ chính xác cao. Thường được sử dụng trong tín hiệu số (Signal Processing), phân lớp ảnh (Image Classification).

2.4.2 Tích chập trong mạng neural

Tích chập được sử dụng đầu tiên trong xử lý tín hiệu số (Signal processing). Theo nguyên lý biến đổi thông tin bằng tích các ma trận ảnh số, các nhà khoa học đã áp dụng kĩ thuật này vào xử lý ảnh và video số.



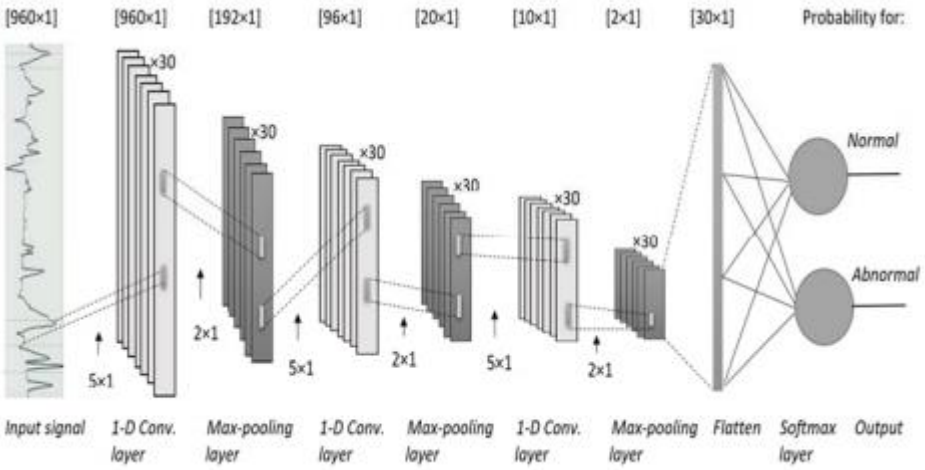
Hình 2.9. Ma trận trong ảnh số

Ma trận bên trái là một bức ảnh đen trắng. Mỗi giá trị của ma trận tương đương với một điểm ảnh (pixel), 0 là màu đen, 1 là màu trắng (nếu là ảnh grayscale thì giá trị biến thiên từ 0 đến 255), hay nói cách khác khi với ảnh grayscale thì các điểm ảnh được biểu diễn bằng các chuỗi 8bit).

Sliding window còn có tên gọi là kernel, filter hay feature detector. Ở đây, ta dùng một ma trận filter 3×3 nhân từng thành phần tương ứng (element-wise) với ma trận ảnh bên trái. Gía trị đầu ra do tích của các thành phần này cộng lại. Kết quả của

tích chập là một ma trận sinh ra từ việc trượt ma trận filter và thực hiện tích chập cùng lúc lên toàn bộ ma trận ảnh bên trái.

CNN gồm một vài layer của tích chập kết hợp với các hàm kích hoạt phi tuyến (nonlinear activation function) như ReLU hay tanh để tạo ra thông tin trừu tượng hơn (abstract/higher-level) cho các layer tiếp theo.



Hình 2.10. Mô hình mạng neural trong xử lý âm thanh

Trong mô hình Feedforward Neural Network (mạng neural truyền thẳng), các layer kết nối trực tiếp với nhau thông qua một trọng số w (weighted vector). Các layer này còn được gọi là có kết nối đầy đủ (fully connected layer) hay affine layer. Trong mô hình CNNs thì ngược lại. Các layer liên kết được với nhau thông qua cơ chế convolution. Layer tiếp theo là kết quả convolution từ layer trước đó, nhờ vậy mà ta có được các kết nối cục bộ. Nghĩa là mỗi neural ở layer tiếp theo sinh ra từ filter áp đặt lên một vùng ảnh cục bộ của neural layer trước đó. Mỗi layer nhờ vậy được áp đặt các filter khác nhau, thông thường có vài trăm đến vài nghìn filter như vậy. Một số layer khác như pooling/subsampling layer dùng để chắt lọc lại các thông tin hữu ích hơn (loại bỏ các thông tin nhiễu).

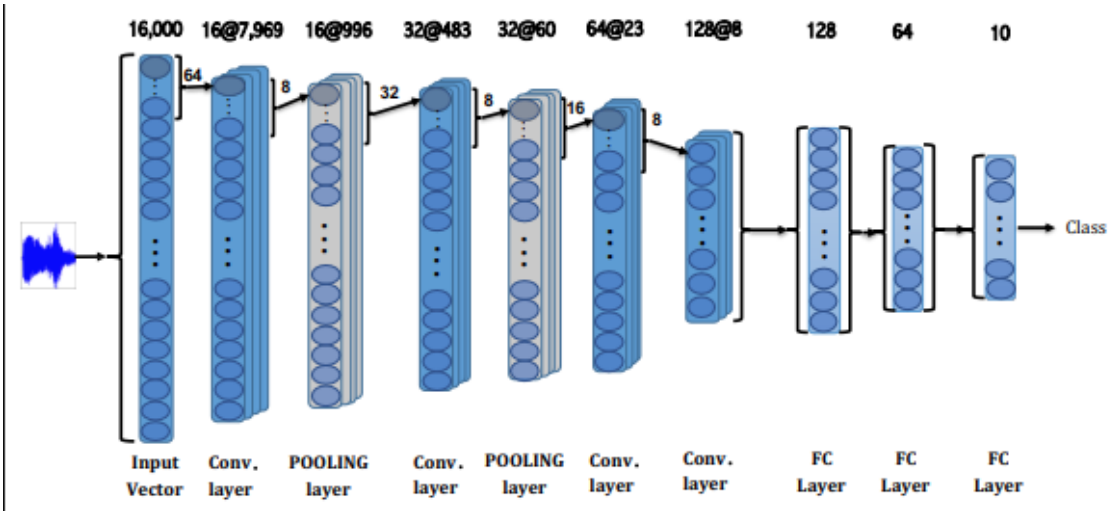
Có ba tầng chính để xây dựng kiến trúc cho một mạng nơron tích chập:

1. Tầng tích chập

2. Tầng gộp (pooling layer)

3. Tầng được kết nối đầy đủ (fully-connected).

Tầng kết nối đầy đủ giống như các mạng nơron thông thường, và tầng chập thực hiện tích chập nhiều lần trên tầng trước. Tầng gộp có thể làm giảm kích thước mẫu trên từng khối 2x2 của tầng trước đó. Ở các mạng nơron tích chập, kiến trúc mạng thường chồng ba tầng này để xây dựng kiến trúc đầy đủ. Ví dụ minh họa về một kiến trúc mạng nơron tích chập đầy đủ:



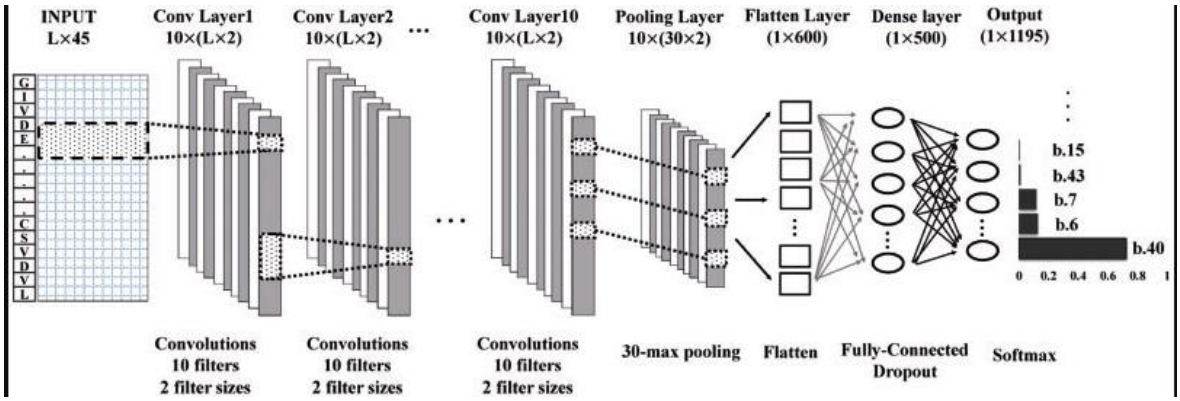
Hình 2.11. Một mô hình phân lớp âm thanh sử dụng mạng neural tích chập

Tuy nhiên, ta sẽ không đi sâu vào khái niệm của các layer này. Trong suốt quá trình huấn luyện, CNNs sẽ tự động học được các thông số cho các filter. Ví dụ trong tác vụ phân lớp ảnh, CNNs sẽ cố gắng tìm ra thông số tối ưu cho các filter tương ứng theo thứ tự raw pixel > edges > shapes > facial > high-level features. Layer cuối cùng được dùng để phân lớp ảnh.

2.4.3 Mô hình mạng CNN trong phát hiện tiếng ngáy

CNNs có tính bất biến và tính kết hợp cục bộ (Location Invariance and Compositionality). Với cùng một đối tượng, nếu đối tượng này được chiếu theo các gốc độ khác nhau (translation, rotation, scaling) thì độ chính xác của thuật toán sẽ bị

ảnh hưởng đáng kể. Pooling layer biểu hiện được tính bất biến đối với phép dịch chuyển (translation), phép quay (rotation) và phép co giãn (scaling).

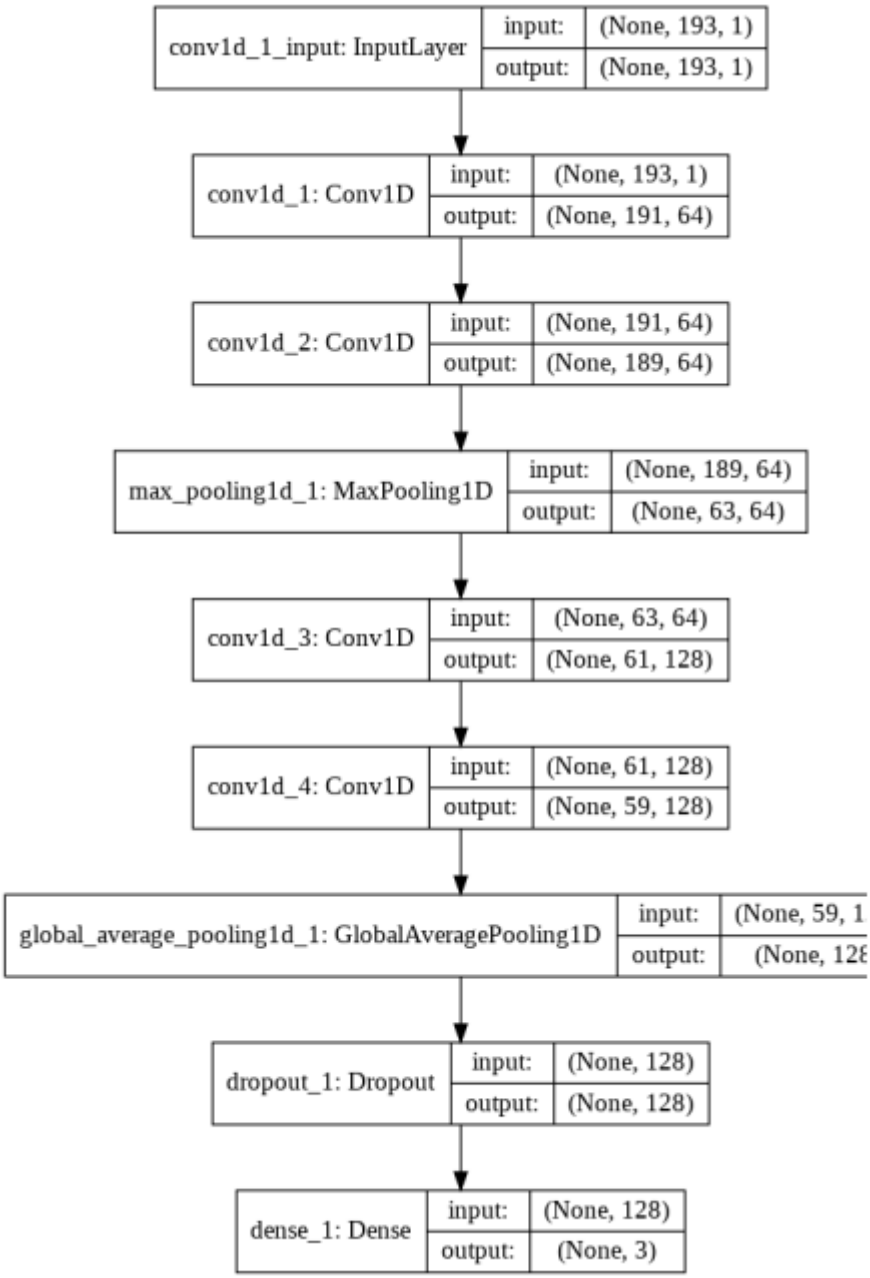


Hình 2.12. Phương pháp phát hiện tiếng ngáy trên mô hình mạng neural CNN

Sau quá trình tìm hiểu và tham khảo, với điều kiện thực nghiệm còn hạn chế với kiến trúc CNN, luận văn quyết định áp dụng 4 lớp tích chập với các thông số như sau:

Bảng 2.1. Các lớp tích chập trong mô hình CNN nhận dạng tiếng ngáy

	Feature maps	**Patch size**	**Pool size**
Conv layer 1	64	193x1	191x64
Conv layer 2	64	191x64	189x64
Conv layer 3	128	189x64	63x64
Conv layer 4	128	61x128	59x128



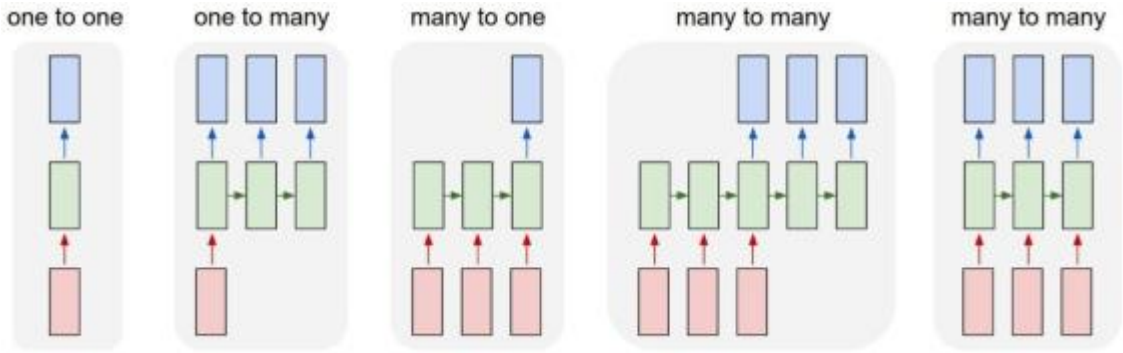
Hình 2.13. Mô hình CNN luận văn sử dụng

Trong mô hình mạng CNN luận văn sử dụng thì từ dữ liệu đầu vào thông qua lớp đầu vào với kích thước là 193x1 sau đó từ ma trận từ lớp đầu vào này thông qua các bước tích chập với các đặc trưng được quan tâm thì kết quả thu được sẽ mang tới cách phân loại của loạt âm thanh này.

2.5 Mô hình LSTM cho phát hiện tiếng ngáy

2.5.1 Giới thiệu về mạng neural hồi quy

Mạng neural hồi quy là một trong những mô hình Deep Learning được đánh giá có nhiều ưu điểm trong các tác vụ xử lý ngôn ngữ tự nhiên. và các ứng dụng nhận diện giọng nói, các chuỗi thông tin đóng vai trò rất quan trọng. Các dạng mạng neural hồi quy



Hình 2.14. Các dạng bài toán RNN

One to one: mẫu bài toán cho Neural Network (NN) và Convolutional Neural Network (CNN), 1 input và 1 output, ví dụ với CNN input là ảnh và output là ảnh được segment.

One to many: bài toán có 1 input nhưng nhiều output, ví dụ: bài toán caption cho ảnh, input là 1 ảnh nhưng output là nhiều chữ mô tả cho ảnh đấy, dưới dạng một câu.

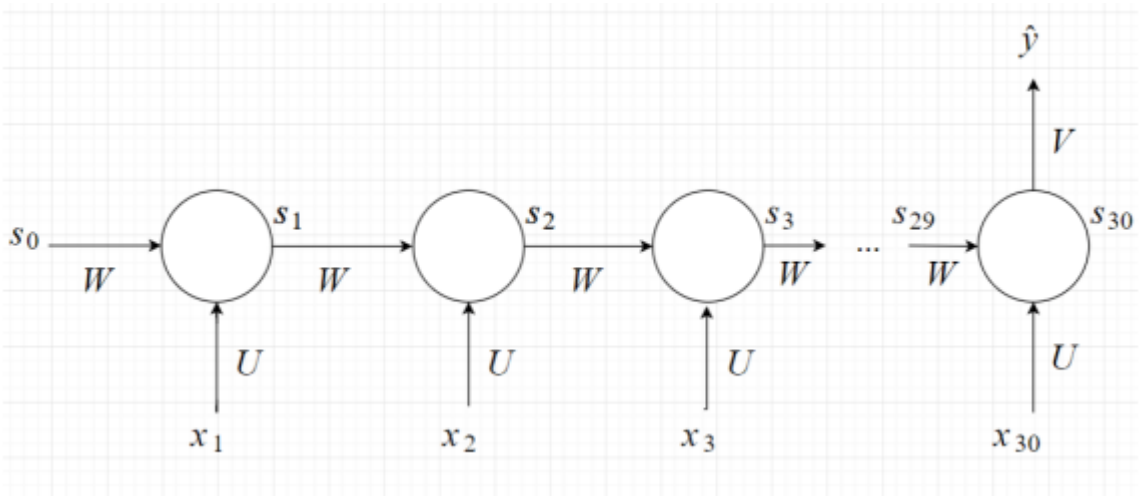
Many to one: bài toán có nhiều input nhưng chỉ có 1 output, ví dụ bài toán phân loại hành động trong video, input là nhiều ảnh (frame) tách ra từ video, ouptut là hành động trong video

Many to many: bài toán có nhiều input và nhiều output, ví dụ bài toán dịch từ tiếng anh sang tiếng việt, input là 1 câu gồm nhiều chữ: "I love Vietnam" và output cũng là 1 câu gồm nhiều chữ "Tôi yêu Việt Nam".

2.5.2 *Hồi quy trong mạng neural và mô hình LSTM*

Recurrent có nghĩa là thực hiện lặp lại cùng một tác vụ cho mỗi thành phần trong chuỗi. Trong đó, kết quả đầu ra tại thời điểm hiện tại phụ thuộc vào kết quả tính toán của các thành phần ở những thời điểm trước đó.

Quan sát sơ đồ biểu diễn RNNs, ta thấy rằng mô hình này có khả năng biểu diễn mối quan hệ phụ thuộc giữa các thành phần trong chuỗi



Hình 2.15. Mô hình RNN

Ta có:

• Mô hình có 30 đầu vào và 1 đầu ra, các input được cho vào mô hình đúng với thứ tự ảnh trong video x1;x2;:::x30.

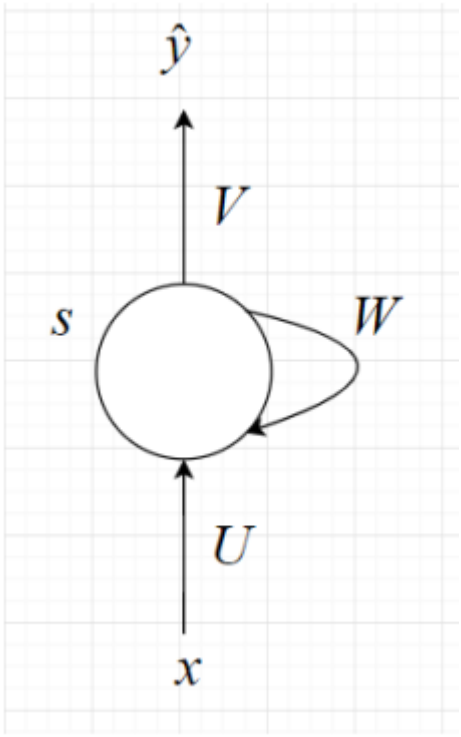
Mỗi hình tròn được gọi là 1 trạng thái, trạng thái t có đầu vào là xt và st−1 (output của state trước); đầu ra là st = f(U ∗xt +W ∗st−1). f là activation function thường là Tanh hoặc ReLU.

• Có thể thấy st mang cả thông tin từ trạng thái trước (st−1) và đầu vào của trạng thái hiện tại => st giống như bộ nhớ các đặc điểm của các đầu vào từ x1 đến xt

• s0 được thêm vào chỉ cho chuẩn công thức nên thường được gán bằng 0 hoặc giá trị ngẫu nhiên. Có thể hiểu là ban đầu chưa có dữ liệu gì để học thì bộ nhớ rỗng.

• Do chỉ có 1 đầu ra, nên sẽ được đặt ở state cuối cùng, khi đó s30 học được thông tin từ tất cả các đầu vào. yˆ = g(V ∗s30). g là activation function, phân loại sử dụng softmax.

Ta thấy là ở mỗi trạng thái nếu các hệ số W, U là giống nhau nên mô hình có thể được viết lại thành:



Hình 2.16. Mô hình RNN rút gọn

Nói cách khác, RNNs là một mô hình có bộ nhớ (memory), có khả năng nhớ được thông tin đã tính toán trước đó. Không như các mô hình Neural Network truyền thống đó là thông tin đầu vào (input) hoàn toàn độc lập với thông tin đầu ra (output)..

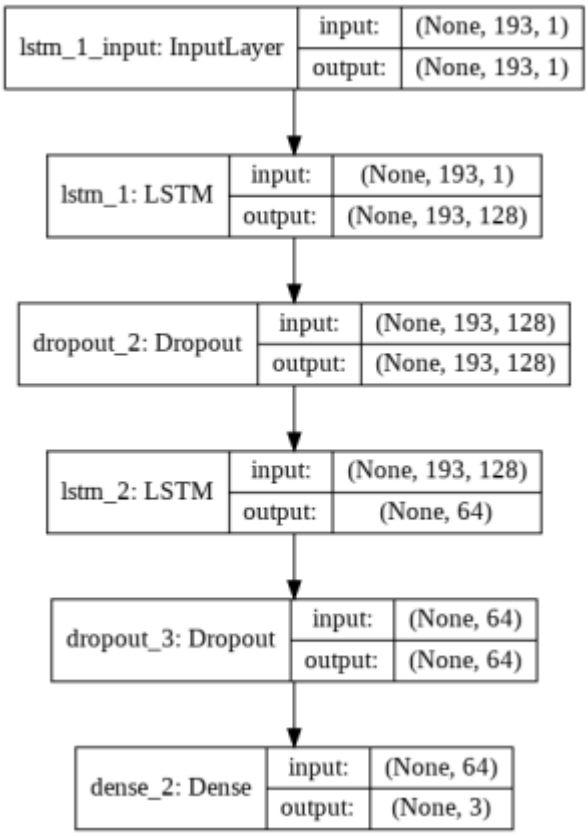
2.5.3 Mô hình mạng LSTM trong phát hiện tiếng ngáy

Như đã giới thiệu ở phần trên về mạng RNN, RNN có thể xử lý thông tin dạng chuỗi, như dự đoán hành dộng trong chuỗi ảnh, hay số tăng giảm giá nhà từ trong các dữ liệu trong lịch sử. RNN mang thông tin của các trạng thái trước tới các trạng thái

sau, rồi ở trạng thái cuối là sự kết hợp của các trạng thái đã diễn ra để dự đoán kết quả.

Về lý thuyết, RNNs có thể nhớ được thông tin của chuỗi có chiều dài bất kì, nhưng trong thực tế mô hình này chỉ nhớ được thông tin ở vài bước trước đó

Với ưu điểm về lưu trữ phụ thuộc dài, model sử dụng để huấn luyện trong luận văn này là mô hình LSTM. Mô hình được luận văn sử dụng được mô tả gồm hai lớp LSTM sau đó là một lớp hồi quy đa thức để khái quát hồi quy cho các lớp..

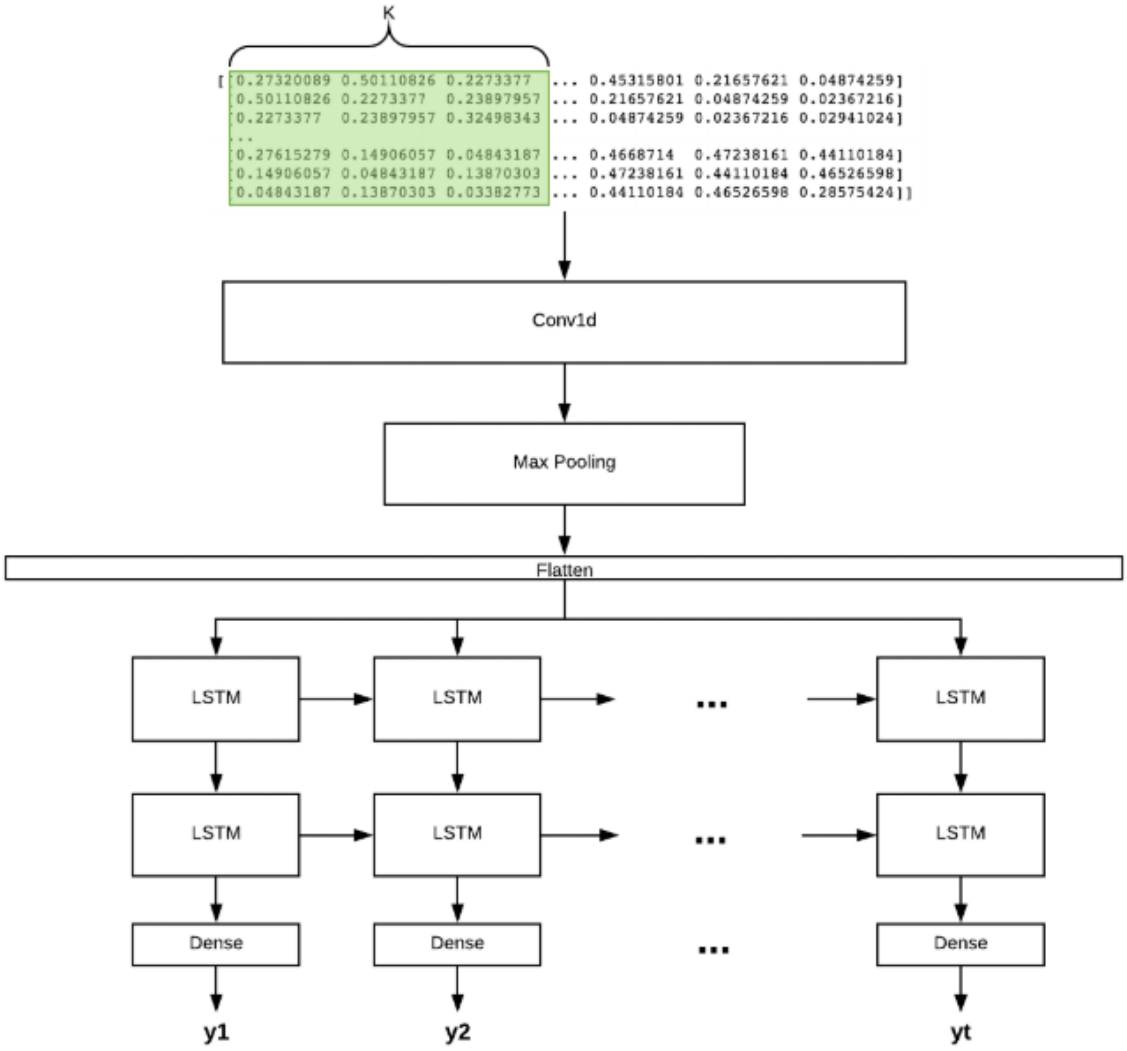


Hình 2.17. Mô hình LSTM luận văn sử dụng

Trong mô hình LSTM cho phát hiện âm thanh thì có lớp đầu vào là đặc trưng đã được trích chọn của các thông tin đầu vào có kích thước là 193x1, dữ liệu này được đưa qua một số lớp LSTM và đưa ra được kết quả cuối cùng là các lớp âm thanh được phân lớp ra.

2.6 Mô hình CNN-LSTM cho phát hiện tiếng ngáy

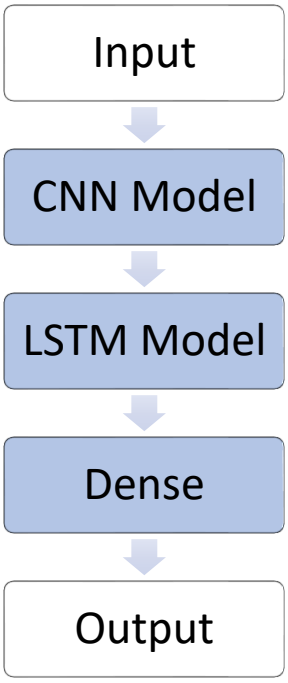
Mô hình CNN LSTM được miêu tả là mô hình lai dựa trên sự kết hợp giữa mô hình neural tích chập CNN và mô hình LSTM để phân lớp dữ liệu. Mô hình CNN LSTM sẽ được đào tạo chung trong Keras. Một LSTM CNN có thể là được xác định bằng cách thêm các lớp CNN ở mặt trước và sau đó là các lớp LSTM với lớp Mật độ ở đầu ra. Kiến trúc này được minh họa như hình sau:



Hình 2.18. Minh họa mô hình mạng CNN-LSTM

Kiến trúc này khi xác định hai mô hình con: Mô hình CNN để trích xuất đặc trưng và Mô hình LSTM để diễn giải các tính năng theo các bước thời gian. Điều này,

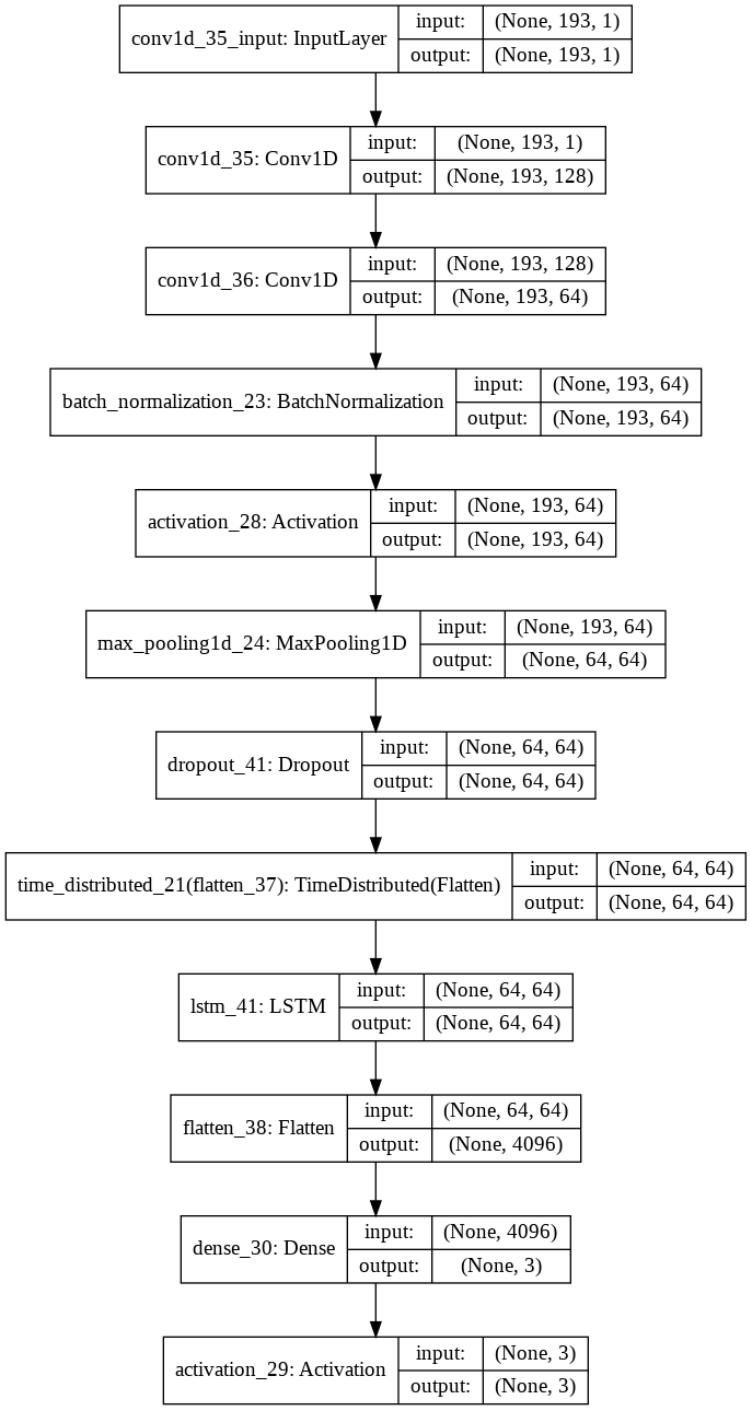
sẽ mang lại cho mô hình tận dụng được ưu điểm của từng mô hình con[22] đối với kết quả sau khi học tập. Sau quá trình tham khảo và nghiên cứu luận văn nhận thấy kiến trúc phát hiện tiếng ngáy sử dụng mô hình học sâu CNN-LSTM sẽ được mô tả như sau:



Hình 2.19. Kiến trúc mô hình học sâu với CNN LSTM cho nhận dạng tiếng ngáy

Trên kiến trúc của mô hình học sâu CNN-LSTM thì từ dữ liệu đầu vào thì sẽ được đi qua mô hình CNN, sau đó đi tới mô hình LSTM để thực hiện phân tích và tiếp theo đi qua lớp Dense để kết thúc mô hình.

Dựa trên mô hình kiến trúc học sâu CNN-LSTM đã tham khảo và tìm hiểu luận văn đưa ra quyết định triển khai mô hình thử nghiệm và cài đặt với các lớp được mô tả như hình 2.19.



Hình 2.20. Mô hình CNN-LSTM cho phát hiện tiếng ngáy

Trong mô hình CNN-LSTM mà luận văn sử dụng có tham khảo từ các mô hình CNN và LSTM mà luận văn đã lựa chọn ở trong hai phần đã được trình bày ở trên. Mô hình mạng CNN-LSTM trong các nghiên cứu gần đây về nhận dạng cảm

xúc lời nói, mạng này được thiết kế và được thực nghiệm đưa ra được kết quả tốt so với việc sử dụng riêng rẽ một mạng neural.

2.7 Kết luận chương

Trong chương này đã trình bày về quá trình tìm hiểu và áp dụng mô hình học nông SVM các mô hình học sâu CNN, LSTM, CNN-LSTM. Bênh cạnh đó chương này cũng trình bày giới thiệu về thuật toán SVM, mạng neural tích chập, mạng neural hồi quy và mạng neural tích chập và hồi quy đẻ phân lớp dữ liệu.

Với những kiến thức đã tìm hiểu và trình bày tại chương, luận văn sẽ áp dụng kiến trúc mạng neural tích chập, kiến trúc mạng neural hồi quy – LSTM và so sánh với SVM.

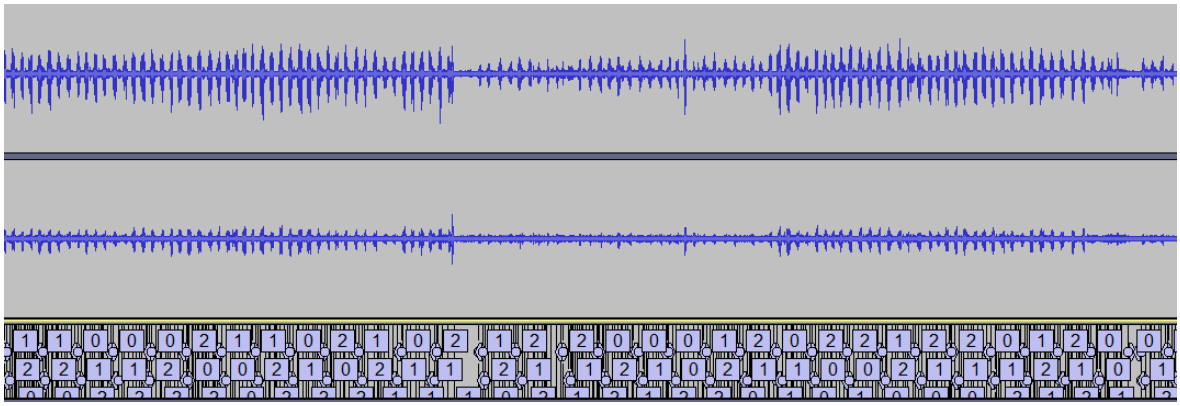
Chương 3 sẽ tiến hành thực nghiệm dữ liệu với phương pháp đã đề xuất dựa trên các kịch bản khác nhau, sau đó sẽ đánh giá độ chính xác và đưa ra đề xuất định hướng tiếp theo

CHƯƠNG 3: THỬ NGHIỆM VÀ ĐÁNH GIÁ

Trong chương này sẽ trình bày các vấn đề: thu thập dữ liệu tiếng ngày; thử nghiệm mô hình CNN hoặc mô hình hồi quy RNN phân tích các âm thanh qua đó có thể đánh giá được các kiến trúc học sâu trong việc phát hiện tiếng ngáy.

3.1 Thu thập dữ liệu

Luận văn sử dụng dữ liệu thử nghiệm từ những người có tình trạng ngáy và dữ liệu tiếng ngáy thu thập được trên Kaggle. Quá trình gán nhãn cho tệp dữ liệu gồm 2 bạn tham gia, 1 bạn gán nhãn và 1 bạn kiểm tra lại phần gán nhãn.



Hình 3.1. Một âm thanh ngáy đã được đánh nhãn

Sau khi thực hiện gán nhãn cho các âm thanh mẫu nhận được thì sẽ được chia thành 2 tệp dữ liệu khác biệt với nhau gồm, tệp dữ liệu huấn luyện và tệp dữ liệu kiểm thử theo tỉ lệ 90%, 10%. Dữ liệu đã được chia ra 2 tệp này sẽ được đưa vào thực nghiệm và cho ra kết quả sẽ được trình bày tại phần 3.2.

Tổng hợp các tập dữ liệu đã được gán nhãn đầy đủ với các lớp âm thanh ngáy, và không ngáy số lượng cụ thể thu được sau quá trình gán nhãn âm thanh ngáy được mô tả tại bảng sau.

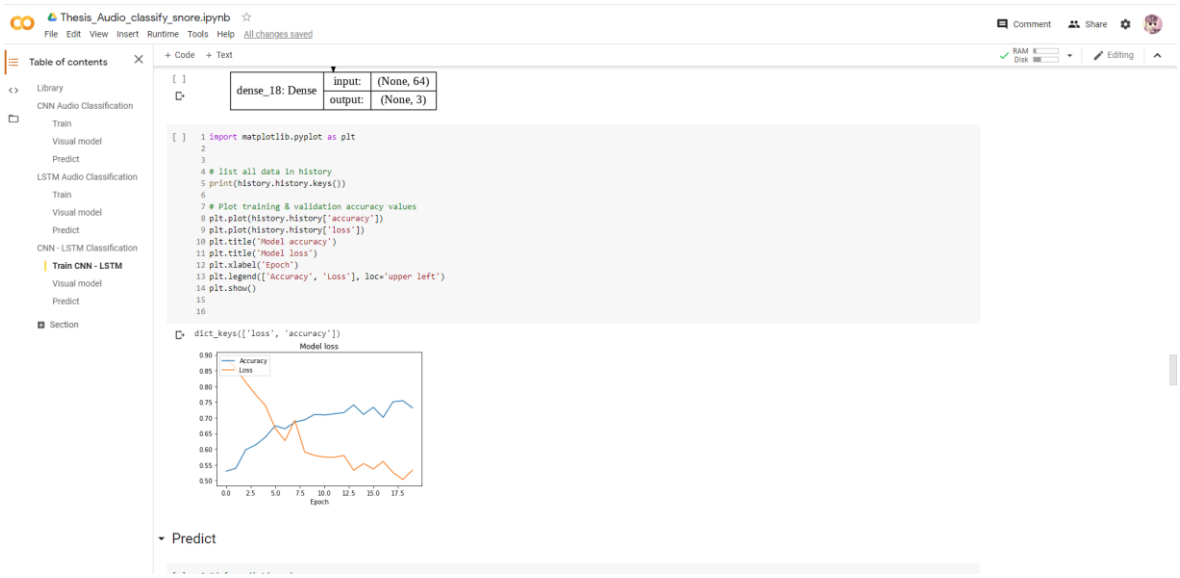
Bảng 3.1. Thống kê dữ liệu thực nghiệm

Dữ liệu âm thanh ngáy			
	Thời gian ngáy	Tổng thời gian	Tỉ lệ tiếng ngáy/ tổng thời gian
Ngáy 1	36 phút	40 phút	0.9
Ngáy 2	25 phút	30 phút	0.83
Dữ liệu Kaggle	8 phút	16 phút	1
Tổng cộng	69 phút	86 phút	

Với dữ liệu thực nghiệm như trên thì có đủ các âm thanh ngáy/ không ngáy từ những người xuất hiện tình trạng ngáy khi ngủ và thêm vào đó có thêm các dữ liệu của Kaggle về các lớp ngáy/ không ngáy được thu thập trên trang mạng chia sẻ âm thanh.

3.2 Kết quả thử nghiệm

Môi trường thử nghiệm các mô hình học sâu được tìm hiểu thông qua Google Colab hay Colaboratory notebooks. Google Colab cung cấp cho chúng ta khả năng tính toán mạnh hơn với Tesla K80 GPU, thay vì phải code và train model với máy tính, laptop cá nhân. Google Colab cũng hỗ trợ khá toàn diện các thư viện trong python, phiên bản mới nhất của tensoflow, keras, PyTorch,. Cv2 .. trong việc cài đặt các mô hình.



Hình 3.2. Môi trường thực nghiệm Google Collab

Trong môi trường Collab Hình 3.2 thì ngôn ngữ sử dụng để cài đặt các thực nghiệm là ngôn ngữ lập trình python có thể được phân cấp ra để chạy từng phân vùng, kèm theo các chú thích của các vùng trong đó. Việc sử dụng Google Collab có đưa ra 3 tùy chọn là sử dụng CPU, GPU có sẵn của Google Collab, hoặc có thể kết nối tới tài nguyên của máy tính.

Để đánh giá các mô hình thì luận văn sử dụng 2 độ đo là Precision và Reall trong đó:

TP: là số âm thanh tiếng ngáy mà mô hình đoán là tiếng ngáy.

FP: là số âm thanh tiếng ngáy mà mô hình đoán là không phải tiếng ngáy.

FN: là số âm thanh không phải là tiếng ngáy mà mô hình dựa đoán là tiếng ngáy.

Precision được định nghĩa là tỉ lệ số điểm TP trong số những điểm được phân loại là chủ động của mô hình (TP+FP) với công thức (3.1) được tính như sau:

$$Precision = \frac{TP}{TP + FP} \quad \textbf{(3.1)}$$

Recall được định nghĩa là tỉ lệ số điểm TP trong số những điểm thực sự là do mô hình dự đoán ra (TP+FN) với công thức (3.2) được tính như sau:

$$Recall = \frac{TP}{TP + FN} \quad \textbf{(3.2)}$$

Ngoài ra, hai độ đo trên không phải lúc nào cũng tăng giảm tương ứng với nhau, có trường hợp Recall cao còn Precision thấp và ngược lại, để cho đánh giá tổng quát hơn thì F-measure là trung bình điều hòa của 2 độ đó trên với hệ số 0.5 (tầm quan trọng của 2 hệ số ngang nhau) được tính với công thức (3.3) như sau:

$$F_1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} = 2\frac{precision\,.recall}{precision + recall} \quad \textbf{(3.3)}$$

3.2.1 Kết quả học nông SVM

Với mô hình học nông SVM với các tham số được xác định khi chạy thực nghiệm là tham số C và gamma là hai tham số rất quan trọng trong việc huấn luyện SVM

C là tham số trong bài toán khoảng cách mềm giúp đưa các điểm dữ liệu nằm trong khoảng hai siêu phẳng được phân loại đúng vào lớp của chúng hay giúp kiểm soát các lớp khác nhau. Khi C càng lớn mô hình được huấn luyện sẽ càng gần với với dữ liệu trong tập đào tạo. Điều này cũng đồng nghĩa với việc mô hình có thể bị quá vừa với dữ liệu đào tạo tạo nên kết quả tốt với dữ liệu dược đào tạo nhưng với các dữ liệu khác cần thực hiện thì sẽ không có giá trị đúng. Vậy có quy luật được rút ra:

+ C tăng, cho phép sai lệch giảm, thu được khoảng cách tăng.

+ C giảm, cho phép sai lệch tăng, thu được khoảng cách giảm.

Gamma không phải là tham số của SVM mà là tham số của hàm kernel RBF. Gamma ảnh hưởng tới mô hình theo quy luật sau:

+ gamma tăng, cho phép sai lệch tăng, thu được khoảng cách giảm

+ gamma giảm, cho phép sai lệch giảm, thu được khoảng cách tăng.

Điều này hoàn toàn trùng khớp với phần đã được trình bày tại chương 2 về đặc điểm của việc học máy cổ điển với SVM

Dựa vào quy luật đó luận văn thử nghiệm với các tham số C vào gamma có giá trị như sau: clf = SVC(C=20.0, gamma=0.00001)

Kết quả thực nghiệm của SVM thu được:

Bảng 3.2. Kết quả của phương pháp học nông SVM

SVM			
Acc (%)	0.724637681		
	Presion	**Recall**	**F1**
Tiếng ngáy	0.71559633	0.75	0.732394366
Không ngáy	0.734693878	0.699029126	0.71641791

Dựa trên bảng kết quả của mô hình SVM ta có thể nhận thấy trong SVM thì tỉ lệ phát hiện tiếng ngáy/ không ngáy gần như bằng nhau. Tỉ lệ chính xác khoảng gần 0.724.

3.2.2 Kết quả của phương pháp CNN

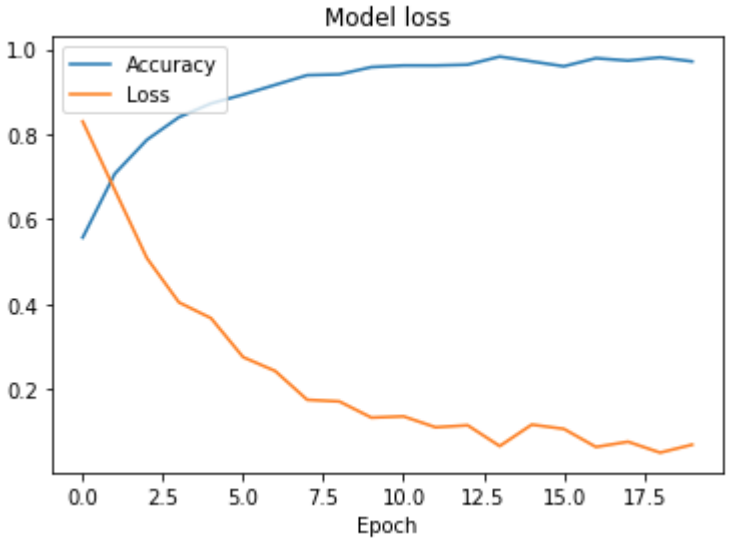
Mô hình học sâu với mạng mô hình CNN đã được lựa chọn trong phần 2.4 mô hình mạng CNN trong phát hiện tiếng ngáy.

Kết quả thực nghiệm của CNN được thể hiện như sau.

Bảng 3.3. Kết quả của mô hình CNN

CNN			
Acc	0.768115942		
	Presion	**Recall**	**F1**
Tiếng ngáy	0.689189189	0.980769231	0.80952381
Không ngáy	0.966101695	0.80952381	0.703703704

Mô hình học CNN đánh giá mô hình



Hình 3.3. Thực nghiệm độ chính xác của mô hình CNN qua số lần epoch

Thời gian mà mô hình đào tạo hết tổng cộng 17 giây, kiểm tra độ chính xác đạt, 0.968 và đạt điểm 0.12452.

Dựa trên bảng kết quả của mô hình học sâu CNN, kết quả thực nghiệm, kết quả đo đánh giá mô hình, kết quả huấn luyện mô hình ta có thể nhận thấy mô hình mạng CNN có tỉ lệ chính xác vượt trội hơn so với phương pháp học sâu với độ chính

xác lên tới 0.768. Các độ đo về độ chính xác khi phát hiện âm thanh ngáy là 0.689 nhỏ hơn nhiều so với việc phát hiện ra âm thanh đó không phải tiếng ngáy là 0.9661.

Điều này chứng tỏ mô hình mạng neural học sâu đưa ra kết quả tốt hơn nhiều so với mô hình phương pháp học máy bình thường mà ở đây là phương phám học nông SVM. Mạng CNN phát hiện cho phát hiện ra âm thanh ngáy có kết quả cao như vậy một phần là do mạng sử dụng mạng tích chập Cov1D thích hợp có các đặc trưng thể hiện tuyến tính theo thời gian và các đặc trưng này đi qua lớp tích chập trong mô hình đã thể hiện rõ các lớp.

3.2.3 Kết quả của phương pháp LSTM

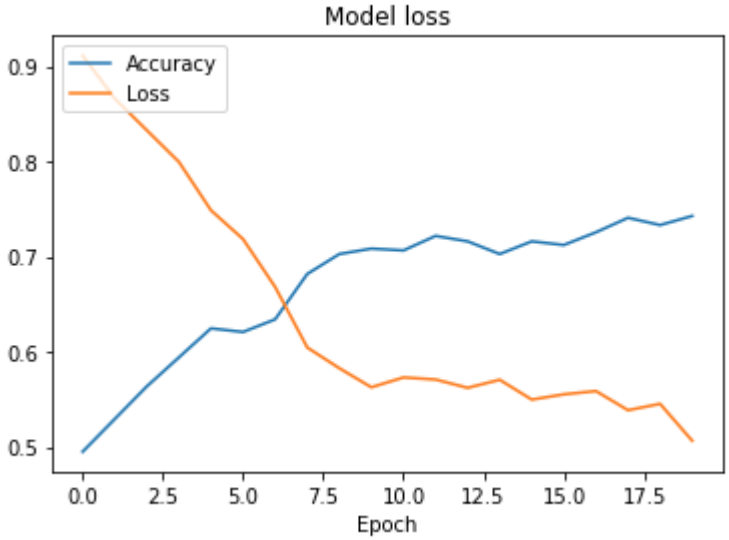
Mô hình học sâu với mạng mô hình LSTM đã được lựa chọn trong phần 2.5 mô hình mạng LSTM trong phát hiện tiếng ngáy.

Kết quả thực nghiệm của LSTM được thể hiện như sau:

Bảng 3.4. Kết quả của mô hình LSTM

LSTM			
Acc (%)	0.753623188		
	Presion	**Recall**	**F1**
Tiếng ngáy	0.702290076	0.884615385	0.782978723
Không ngáy	0.842105263	0.621359223	0.715083799

Mô hình học LSTM đánh giá mô hình



Hình 3.4. Thực nghiệm độ chính xác mô hình LSTM qua số lần epoch

Thời gian mà mô hình đào tạo hết tổng cộng 205 giây, kiểm tra độ chính xác đạt, 0.7635 và đạt điểm : 0.466

Dựa trên bảng kết quả của mô hình học sâu LSTM, kết quả thực nghiệm, kết quả đo đánh giá mô hình, kết quả huấn luyện mô hình ta có thể nhận thấy mô hình mạng LSTM có tỉ lệ chính xác vượt trội hơn so với phương pháp học sâu với độ chính xác lên tới 0.7536. Các độ đo về độ chính xác khi phát hiện âm thanh ngáy là 0.7022 nhỏ hơn nhiều so với việc phát hiện ra âm thanh đó không phải tiếng ngáy là 0.8421.

Điều này chứng tỏ mô hình mạng neural LSTM đưa ra kết quả tốt hơn so với mô hình phương pháp học máy bình thường mà ở đây là phương phám học nông SVM và không bằng được với mô hình học sâu CNN. Mạng LSTM phát hiện cho phát hiện ra âm thanh ngáy có kết quả như vậy một phần là do đặc điểm của mô hình mạng sử dụng các dữ liệu từ quá khứ để đoán kết quả tiếp theo và dần được cải thiện sau thời gian, có thể nhận thấy rõ ràng trong Hình 3.4, khi mà độ chính xác của mô hình được phát tiển từ từ không có sự bứt phá về độ chính xác rõ ràng như trong mạng CNN.

3.2.4 Kết quả của phương pháp CNN-LSTM

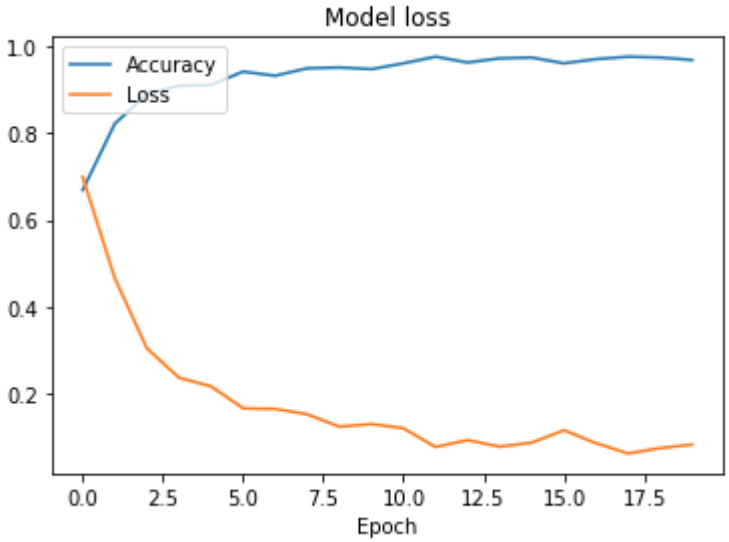
Mô hình học sâu với mạng mô hình CNN-LSTM đã được lựa chọn trong phần 2.6 mô hình mạng CNN-LSTM trong phát hiện tiếng ngáy.

Kết quả thực nghiệm của CNN-LSTM được thể hiện như sau:

Bảng 3.5. Kết quả của mô hình CNN-LSTM

CNN-LSTM			
Acc (%)	**0.917874396**		
	Presion	**Recall**	**F1**
Tiếng ngáy	0.871794872	0.980769231	0.923076923
Không ngáy	0.977777778	0.854368932	0.911917098

Mô hình học CNN-LSTM đánh giá mô hình



Hình 3.5. Thực nghiệm độ chính xác mô hình CNN-LSTM qua số lần epoch

Dựa trên bảng kết quả của mô hình CNN-LSTM ta có thể nhận thấy thời gian mà mô hình đào tạo hết tổng cộng 52 giây, kiểm tra độ chính xác đạt, 0.9772 và đạt điểm: 0.0489

Dựa trên bảng kết quả của mô hình học sâu CNN-LSTM, kết quả thực nghiệm, kết quả đo đánh giá mô hình, kết quả huấn luyện mô hình ta có thể nhận thấy mô hình mạng CNN-LSTM có tỉ lệ chính xác vượt trội hơn so với phương pháp học sâu với độ chính xác lên tới 0.9178. Các độ đo về độ chính xác khi phát hiện âm thanh ngáy là 0.8717 nhỏ hơn so với việc phát hiện ra âm thanh đó không phải tiếng ngáy là 0.9777.

Điều này chứng tỏ mô hình mạng neural CNN -LSTM đưa ra kết quả tốt nhất so với mô hình phương pháp học máy bình thường mà ở đây là phương phám học nông SVM và không bằng được với mô hình học sâu CNN, mạng học sâu LSTM. Mạng CNN -LSTM phát hiện cho phát hiện ra âm thanh ngáy có kết quả như vậy một phần là do đặc điểm của mô hình mạng sử dụng các dữ liệu tích chập từ các đặc trưng và từ quá khứ để đoán kết quả tiếp theo và dần được cải thiện sau thời gian, có thể nhận thấy rõ ràng trong Hình 3.5, khi mà độ chính xác của mô hình được phát tiển từ từ có sự bứt phá về độ chính xác rõ ràng, có thể nhận thấy mô hình đạt được độ chính xác lớn ngay từ những epoch đầu và qua những lần sau thì càng được tăng độ chính xác lên.

3.3 Phân tích và đánh giá

Dựa vào kết quả của các đánh giá trên thì nhận thấy được các mạng học sâu đều cho kết quả phát hiện âm thanh ngáy tốt hơn nhiều so với mạng học nông mà cụ thể ở đây là SVM.

Độ chính xác, đánh giá qua các độ đo được nên ra trong phần kết quả thử nghiệm gồm Pression, Recall, F1-score thì có thể thấy được các phương pháp có kết quả được xếp từ thấp lên cao như sau:

Bảng 3.6. Độ chính xác của các mô hình

Mô hình	Độ chính xác
Mô hình học nông SVM	0.724637681
Mô hình mạng CNN	0.768115942
Mô hình mạng LSTM	0.753623188
Mô hình mạng CNN-LSTM	**0.917874396**

Kết quả của các mô hình được thực nghiệm trong luận văn có thể nhận thấy rằng, mô hình mạng học sâu có kết quả tốt hơn hẳn so với mô hình mạng học nông như SVM, kết quả của mô hình mạng học sâu CNN-LSTM cho ra là kết quả tốt nhất, nhờ có sự kết hợp giữa ưu điểm của mô hình CNN và LSTM điều này có sự tương đồng với các nghiên cứu về phân lớp âm thanh có liên quan. Mô hình CNN – LSTM cũng đã khắc phục được các thiếu sót của từng loại mô hình khi sử dụng riêng rẽ của các mô hình học sâu khác trên cơ sở dữ liệu đã chọn

3.4 Kết luận chương

Trong chương này sẽ trình bày các vấn đề: thu thập dữ liệu tiếng ngày; thử nghiệm mô hình CNN hoặc mô hình hồi quy RNN phân tích các âm thanh qua đó có thể đánh giá được các kiến trúc học sâu trong việc phát hiện tiếng ngáy. Sau quá trình thử nghiệm với tập dữ liệu và cài đặt với các mô hình, phương pháp học máy khác nhau thì thu được kết quả tốt nhất thuộc về mô hình mạng học sâu kết hợp CNN-LSTM với kết quả tốt hơn nhiều so với các phương pháp còn lại.

KẾT LUẬN

Nghiên cứu về phát hiện âm thanh nói chung, về bài toán phát hiện tiếng ngáy dựa trên học sâu nói riêng với tôi là công nghệ mới, thời gian nghiên cứu còn ngắn nên vẫn còn nhiều vấn đề chưa thực sự nắm bắt tốt. Tuy nhiên, qua quá trình nghiên cứu, luận văn đã tìm hiểu sâu về các giai đoạn từ tiền xử lý dữ liệu đến các phướng pháp xử lý âm thanh, các phương pháp học máy mà đặc biệt là các mô hình học sâu với mạng neural, phương pháp học sâu để xây dựng mô hình phân lớp dữ liệu (mô hình hình CNN, LSTM, CNN-LSTM) và so sánh với mô hình học nông SVM.

Sử dụng các mạng neural nói chung hay CNN, LSTM và CNN-LSTM nói riêng trong học sâu là một hướng đi có kỹ thuật và hiệu quả trong các bài toán xử lý chuỗi và hiện đang trở thành xu thế của các nhà nghiên cứu.

Trong tương lai, luận văn có thể được phát triển nghiên cứu các mô hình khác, giải quyết các bài toán khác về theo dõi, nhân diện âm thanh, hoặc phát triển thành những ứng dụng y tế mà có thể hỗ trợ cho nhiều người trong cộng đồng..

DANH MỤC CÁC TÀI LIỆU THAM KHẢO

[1] PGS.TS. Lê Hữu Lập, Bài giảng Phương pháp nghiên cứu khoa học, Học viện Công nghệ BCVT.

[2] PGS.TS. Từ Minh Phương, Giáo trình trí tuệ nhân tạo, Học viện Công nghệ BCVT.

[3] J. Dennis, H. D. Tran, and H. Li, "Spectrogram Image Feature for Sound Event Classification in Mismatched Conditions,"

[4] Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, Kevin Wilson, Cnn architectures for large-scale audio classification ,

[5] Shahin Amiriparian, Maurice Gerczuk, Sandra Ottl, Nicholas Cummins, Michael Freitag, Sergey Pugachevskiy, Alice Baird, Björn Schulle, Snore Sound Classification Using Image-based Deep Spectrum Features ,

[6] Jonathan William DennisPublished, (2014), Sound Event Recognition in Unstructured Environments using Spectrogram Image Processing,

[7] Zixing Zhang, ,Snore-GANs: Improving Automatic Snore Sound Classification with Synthesized Data

[8] Alex Krizhevsky , Ilya Sutskever , Geoffrey E. Hinton ImageNet Classification with Deep Convolutional Neural Networks,

[9] Janott, Christoph & Schmitt, Maximilian & Zhang, Yue & Qian, Kun & Pandit, Vedhas & Zhang, Zixing & Heiser, Clemens & Hohenhorst, Winfried & Herzog, Michael & Hemmert, Werner & Schuller, Björn. (2018). Snoring classified: The Munich-Passau Snore Sound Corpus.

[10] Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation.

[11] Thorpy MJ. (1990), Classification of sleep disorders. J Clin Neurophysiol.

[12] Shin, H., Cho, J. Unconstrained snoring detection using a smartphone during ordinary sleep. *BioMed Eng OnLine* **13,** 116 (2014).

[13] F. Dalmasso, R. Prota, Snoring: Analysis,Measurement, Clinical Implications and Applications, European Respiratory Journal.

[,14] Weng, Chih-Wen & Lin, Cheng Yuan & Jang, Jyh-Shing. (2004). Music Instrument Identification Using MFCC: Erhu as an Example.

[15] Kang, Bingbing & Dang, Xin & Wei, Ran. (2017). Snoring and apnea detection based on hybrid neural networks. 57-60. 10.1109/ICOT.2017.8336088.

[16] Khan, Tareq Hasan. (2019). A Deep Learning Model for Snoring Detection and Vibration Notification Using a Smart Wearable Gadget. Electronics.

[17] Cavusoglu, Mustafa & Poets, Christian & Urschitz, Michael. (2017). Acoustics of snoring and automatic snore sound detection in children. Physiological Measurement.

[18] Zhang, Zixing & Han, Jing & Qian, Kun & Janott, Christoph & Guo, Yanan & Schuller, Björn. (2020). Snore-GANs: Improving Automatic Snore Sound Classification With Synthesized Data. IEEE Journal of Biomedical and Health Informatics. 24. 300-310. 10.1109/JBHI.2019.2907286.

[19] Kim, T., Kim, J. & Lee, K (2018). Detection of sleep disordered breathing severity using acoustic biomarker and machine learning techniques. BioMed Eng OnLine 17, 16

[20] Cavusoglu, M & Kamasak, Mustafa & Eroğul, O. & Çiloglu, Tolga & Serinagaoglu Dogrusoz, Yesim & Akcam, T. (2007). An efficient method for snore/nonsnore classification of sleep sounds. Physiological measurement.

[21] Jason Brownlee (2017), Long Short-Term Memory Networks With Python

[22] Yang, Yang & Zheng, Xiangwei & Yuan, Feng. (2018). A Study on Automatic Sleep Stage Classification Based on CNN-LSTM. ICCSE'18: Proceedings of the 3rd International Conference on Crowd Science and Engineering. 1-5. 10.1145/3265689.3265693.

[23] Adrien Ycart and Emmanouil Benetos. “A study on LSTM networks for polyphonic music sequence modelling”, 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017.

[24] Kons, Zvi & Toledo-Ronen, Orith. (2013). Audio event classification using deep neural networks. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH. 1482-1486.