**Đánh giá về phân tích và phân loại tiếng khóc trẻ sơ sinh**

**Trừu tượng**: Bài báo này đánh giá những nghiên cứu gần đây trong việc phân loại và phân tích tiếng khóc trẻ sơ sinh. Một lượng lớn các tài liệu được đánh giá chủ yếu từ các khía cạnh thu thập dữ liệu, kỹ thuật xử lý tín hiệu miền chéo và những phương pháp phân loại học máy. Chúng tôi giới thiệu các phươn pháp tiếp cận tiền xử lý và mô tả sự đa dạng của các đặc trưng như MFCC, spectrogram, và tần số cơ bản. Cả những đặc trưng âm thanh và đặc trưng đặc thù được trích xuất từ những miền khác nhau có phân loại các tín hiệu dựa trên khung với nhau và có thê được sử dụng để đào tạo phân loại dữ liệu. Cùng với những phương pháp học máy truyền thống như: KNN, SVM, GMM, những kiến trúc mạng nơ ron mới được phát triển như: CNN, RNN được áp dụng trong nghiên cứu tiếng khóc trẻ sơ sinh. Chúng tôi trình bày một vài kết quả thử nghiệm quan trọng về xác định tiếng khóc bệnh lý, phân loại lý do khóc và phát hiện tiếng khóc với một vài tập dữ liệu điển hình. Khảo sát này nghiên cứu một cách có hệ thống những nghiên cứu trước đây trong tất cả lĩnh vực liên quan về trẻ sơ sinh và cung cấp một cái nhìn sâu sắc trong những công trình tiên hiến hiện nay(the current cutting-edge works) trong phân loại và phân tích tiếng khóc trẻ sơ sinh. Chúng tôi cũng đề xuất những định hướng nghiên cứu tương lai trong xử lý dữ liệu, trích xuất đặc trưng, và lĩnh vực phân loại mạng nơ ron để hiểu rõ hơn, diễn giải và xử lý tiếng khóc trẻ sơ sinh.

**Giới thiệu**

Khoảng 130 triệu trẻ em được sinh ra mỗi năm trên toàn thế giới. Chăm sóc trẻ cẩn thận là một thách thức lớn, đặc biệt với những người lần đầu làm cha mẹ. Theo những gợi ý từ những bố mẹ khác và sách là không đủ để giải quyết những vấn đề trong thực tế. Lý do chính là bởi vì khó để hiểu những ý nghĩ của trẻ sơ sinh khi đang khóc. Những bố mẹ trẻ cảm thấy nản lòng và khó xoa dịu con mình bởi vì tất cả những âm thanh khóc của trẻ đều giống nhau. Việc xác định chính xác tiếng khóc trẻ sơ sinh có thể giúp những bố mẹ có thể chăm sóc tốt hơn con của họ. Nghiên cứu về tiếng khóc trẻ sơ sinh được bắt đầu sớm từ những năm 1960, khi nhóm nghiên cứu Wasz\_Hockert đã xác định được 4 kiểu khóc (đau, đói, sinh nở và vui sướng) bằng thính giác của các cô y tá được đào tạo bài bản. Trong những năm đầu, những nhà nghiên cứu đã xác định rằng những kiểu khác nhau về khóc có thể được nghe bởi thính giác của người lớn đã qua đào tạo. Nhưng việc đào tạo nhận thức của con người với tiếng khóc trẻ sơ sinh khó hơn đào tạo những mô hình học máy. Trong nghiên cứu của Mukhopadhyay, độ chính xác cao nhất trong việc phân loại được đào tạo cho một nhóm người để xác định tiếng khóc là 33.09% trong khi thuật toán học máy dựa trên những tính năng phổ và đặc thù có thể xác định mức độ chính xác lên đến 80.56% trong cùng một tập dữ liệu. Xây dựng máy thông minh để hiểu tiếng khóc trẻ sơ sinh dẫn đén cách để xây dựng những robot bảo mẫu thông minh trong tương lai. Bên cạnh việc hiểu những nhu cầu cần thiết hàng ngày của trẻ sơ sinh, việc dự đoán bệnh tật cũng là một nghiệm vụ khóc trong những nghiên cứu tiếng khóc. Do đường thanh quản và hệ thống thở của trẻ bị ảnh hưởng bởi một vài loại bệnh, tiếng khóc của trẻ không khoẻ chứa những đặc điểm độc đáo khác với những tiếng khóc trẻ khoẻ mạnh. Phân tích tiếng khóc bệnh lý trẻ để xác định bệnh là một phương pháp không xâm lấn và nhanh có thể cứu sống trẻ sơ sinh, đặc biệt trong những khu vực thiếu thiết bị và chuyên môn y tế. Trong những năm đầu về nghiên cứu tiếng khóc trẻ, nhiều công trình đã tập trung phân loại tiếng khóc bình thường và tiếng khóc bệnh lý. Trong đánh giá của Saraswathy, 34 bài báo về phân loại tiếng khóc bình thường và bệnh lý được phát hành từ năm 2003 đến năm 2011. Những công trình bao gồm xác định bênh lý như: giảm âm, ngạt, suy giáp, tăng bilirubin máu, hở hàm ếch,…

Nghiên cứu tiếng khóc trẻ sơ sinh lên quan đến việc thu thập dữ liệu, xử lý dữ liệu, trích xuất đặc trưng lựa chọn và phân loại tiếng khóc. Bởi vì độ nhạy của dữ liệu, nó rất khó cho việc nghiên cứu để có được dữ liệu cần thiết. Những nhà nghiên cứu đã tự mình thu âm những clip khóc hoặc xin quyền trcsh dẫn dữ liệu từ một tác giả khác. Hầu hết dữ liệu được thu tại bệnh viện, NICU, nhà hoặc phòng khám bằng phương pháp ghi âm thời gian thực được lắp đặt với thiết bị ghi âm điện tử để gần trẻ sơ sinh trong một thời gian dài. Xử lý tín hiệu là xoá nhiễu nền và thực hiện phân đoạn tiếng khóc để xây dựng dữ liệu khóc. Một khi dữ liệu đã có sẵn, trích xuất đặc trưng là bước tiếp theo để trích xuất những miền khác nhau của tín hiệu khóc. Trích xuất đặc trưng từ miền thời gian, miền phổ công xuất, hoặc miền prosodic,.. thể hiện những đặc trưng khác nhau của tín hiệu. Lựa chọn đặc trưng phù hợp nhất và giảm chiều đặc trưng là nhiệm vụ khác để xây dựng một mô hình phân đoạn hiệu quả. Áp dụng những mô hình học máy phù hợp cho những đặc trưng riêng là rất quan trọng đối với độ chính xác phát hiện và phân loại. Khi AI được biết đến từ những năm 1990, mạng nơ ron nổi lên như một phương pháp phổ biến trong nghiên cứu tiếng khóc trẻ sơ sinh. Mạng nơ ron là hệ thống máy tính, gồm những nơ ron được liên kết với nhau, lấy cảm hứng từ hệ thống não sinh học. Đầu vào vecto, nơ ron, trọng số, hàm hoạt động, và đầu ra là những thành phần chính trong một mạng nơ ron. Mỗi nơ ron có một giá trị được tính toán trong quá trình chuyển tiếp dựa trên trọng số của mỗi kết nối và độ lệch của mỗi lớp. Hàm hoạt động được sử dụng để đạt được tính phi tuyến trong mạng. Sự lan truyền ngược là một thuật toán chính để đào tạo dữ liệu và tối ưu hàm mất mát và đánh giá mức độ phù hợp của mô hình với tập dữ liệu. Trong những năm 2000, hầu hết các phương pháp được áp dụng trong nghiên cứu tiếng khóc liên quan đến mạng nơ ron, bao gồm : sacled conjugate gradient neural network, multi-layer perceptron, general regression neutral network, evolutionary neural network, probabilistic neural network, neural-fuzzry network, and Time Delay Neural network. Hidden Markov model vaf Support Vector Machine(SVM) cũng được áp dụng trong những năm 2000. Trong thập kỷ gần đây, nhiều phương pháp học máy truyền thống, như SVM, K-Nearest Neighbor(KNN), Gaussian Mixture Model (GMM), fuzzy classifier, logistic regression, K-means clustering, and Random Forest, được áp dụng để phân loại tiếng khóc bệnh lý, phân loại nguyên nhân khóc, và phát hiện tiếng khóc. Trong cùng thời kỳ, các kiến trúc mạng noron mới được sử dụng phổ biến trong ngành công nghiệp và nghiên cứu. Mạng no ron tích chập (CNN), mạng nơ ron tái diễn (RNN), CNN-RNN, Capsule Net, Reservoir Network, và nẻuo-fuzzy networks mở ra một chương mới trong nghiên cứu tiếng khóc trẻ sơ sinh.

Khảo sát này đánh giá nghiên cứu tiếng khóc chủ yếu tập trung vào các kỹ thuật xử lý tín hiệu và phương pháp học máy được phát triển trong thập kỷ qua. Trước tiên, chúng tôi đánh giá các tập dữ liệu phổ biến được sử dụng trong các nghiên cứu, sau đó giới thiệu những phương pháp tiền xử lý tín hiệu, và mô tả sự đa dạng của các đặc trưng trong cả miền thời gian và miền tần số, cũng như những đặc trưng siêu phân đoạn của tín hiệu khóc. Chúng tôi tập trung vào việc xem xét các phương pháp hiện đại nhất đang sử dụng như KNN, SVM, GMM và những thuật toán dựa trên CNN cho phân loại và phát hiện. Chúng tôi cung cấp một danh sách các tài nguyên cho những nhà nghiên cứu, những người quan tâm tới công việc trong lĩnh vực này, và cuối cùng chúng tôi đưa ra một quan điểm về công việc trong tương lai trong lĩnh vực nghiên cứu này.

**Thu thập dữ liệu**

Như trình bày hình 1, nghiên cứu tiếng khóc tự động liên quan đến 5 giai đoạn: thu thập dữ liệu, tiền xử lý, trích xuất đặc trưng, lựa chọn đặc trưng, và phân loại. Khám phá những phương pháp mới trong bất ký tầng nào có thể giúp cải thiện hiệu suất của độ chính xác phân loại cuối cùng.



Giai đoạn thu thập dữ liệu bao gồm ghi âm tiếng khóc và gán nhãn. Hầu hết dữ liệu được ghi âm ở bệnh viện, nhà ở, được gán nhãn bởi bác sĩ, y tá hoặc bố mẹ. Những máy ghi âm được đặt gần trẻ sơ sinh hoặc hoạt động tại chỗ để thu từng tín hiệu tiếng khóc hoặc để ghi lại các sự kiện âm thanh xung quanh trẻ sơ sinh trong một thời gian dài. Âm thanh trẻ sơ sinh là một tín hiệu tĩnh trong thời gian ngắn, và nó được giả định để . Vì sự giới hạn của tài nguyên và độ nhạy của quy trình thu thập dữ liệu trẻ sơ sinh khóc, số lượng tập dữ liệu trẻ sơ sinh khóc là rất hạn chế. Từ những bài báo đánh giá trước đây, chúng ta có thể thấy rằng hầu hết tệp dữ liệu được sử dụng là Baby Chillanto. Baby Chillanto được thu thập từ Viện Quốc gia về Vật lý thiên văn và Quang điện tử, Mexico. Nó bao gồm 5 kiểu tín hiệu khóc bao gồm: điếc, ngạt, bình thường, đói và đau. Mỗi kiểu khóc được phân đoạn có độ dài bằng nhau là 1s và tổng số mẫu khóc là 2268. Tập dữ liệu khác được sử dụng trong những bài báo khác là Dunstan Baby Language, được trích xuất từ Dunstan Baby Language bởi Priscilla Dunstan, người tạo ra lý thuyết Dunstan Baby Language. Có một vài phiên bản của tập dữ liệu Dunstan Baby Language kể từ khi tác giả trích xuất những đoạn âm thanh theo cách của họ. Phiên bản được mô tả trong [9] bao gồm 315 tệp wave, lấy mẫu tần số 16KHz, với độ dài biến đổi từ 0,3 đến 1.6s. Mỗi lời nói là một từ của trẻ sơ sinh tương ứng với một trong năm “Từ Dunstan”, được dịch như “Neh”=đói, “Eh”= cần phải ợ, “Oah”=mệt, “Eairh”= đau nhẹ, “Heh”= không thoải mái.

Nhiều tập dữ liệu được tự ghi âm cho nghiên cứu. Những nhà nghiên cứu cần kết nối với những tác giả khác để kiểm tra tính khả thi của những dữ liệu. Một tệp dữ liệu tên là Donate A Cry có sẵn online, nhưng nó được gán nhãn không tốt và chỉ một bài báo được tìm thấy sử dụng dữ liệu này. Bảng 1 chỉ ra những tệp dữ liệu thông thường được sử dụng trong những nghiên cứu gần đây. Một vài tập dữ liệu được ghi trong NICU, phòng khám, hoặc những môi trường trẻ ngồi. Một vài tín hiệu khóc trực tuyến cũng được thu thập trong 25. Một vài tập dữ liệu tổng hợp được tạo ra bởi những tác giả để so sánh hiệu xuất của những phương pháp đề xuất trên dữ liệu thật và dữ liệu tổng hợp. Trong công trình của Ferretti, CNN phát hiện tín hiệu khóc trên dữ liệu tổng hợp tốt hơn dữ liệu thực tế. Nó chỉ ra rằng, việc phát hiện và phân loại tự động trên tiếng khóc thời gian thực vẫn là thách thức, bởi vì môi trường thời gian thực có thể tồn tại nhiều kiểu phức tạp, cái mà có thể ảnh hưởng chất lượng của tiếng khóc. Dữ liệu tổng hợp có thể được tạo ra bởi thêm nhiễu để làm sạch cácc bản ghi hoặc kết hợp những kiểu khóc khác nhau. Mô hình đào tạo trên dữ liệu tổng hợp có thể tránh yêu cầu một lượng lớn dữ liệu để trang bị trong những môi trường như NICU.

Từ bảng 1, chúng ta có thể thấy hầu hết tập dữ liệu với số lượng mẫu giới hạn. Độ dài trung bình là 2983 và chỉ một tệp dữ liệu gần tới 20000 mẫu. Bởi vì độ nhạy của dữ liệu khóc thu thập, đặc biệt những tín hiệu khóc bệnh lý, lượng nhỏ dữ liệu là một trong những thách thức trong nghiên cứu tiếng khóc trẻ sơ sinh. Kỹ thuật tăng dữ liệu được sử dụng để tăng tính nhân tạo trong kích thước dữ liệu. Zhang đã tạo những ảnh sóng mới từ tệp dữ liệu đào tạo bằng cách chuyển những ảnh sóng thành sóng nhanh và sóng chậm với mục đích tăng tệp dữ liệu đào tạo để giải quyết vấn đề quá khớp(overfitting). Trong[43], một vài kỹ thuật tăng dữ liệu, như biến đổi nhiễu, biến đổi cường độ tín hiệu, biến đổi âm sắc, và thay đổi kích thước quang phổ được sử dụng để tăng nhân tạo cả số lượng tín hiệu và số lượng quang phổ. Kết quả thử nghiệm đã chỉ ra rằng những phương pháp tăng dữ liệu không thể dẫn đến cải thiện độ chính xác. Nguyên nhân nằm ở chỗ dữ liệu hạn chế không thể nắm bắt được sự đa dạng trong các tín hiệu tiếng khóc của trẻ sơ sinh.

**Xử lý tín hiệu và tạo đặc trưng**

**Tiền xử lý**

Nhiệm vụ chính trong tiền xử lý là khử nhiễu và phân đoạn tín hiệu. Sự kết hợp của môi trường thu âm dẫn đến những tín hiệu trẻ sơ sinh khóc không sạch. Trong một nơi chăm sóc, bên cạnh tiếng khóc trẻ, có thể có nhiều loại âm thanh như: bước chân, người lớn nói chuyện, điều hoà không khí, âm cảnh báo. Để phát hiện hoặc phân loại tín hiệu khóc một cách chính xác, làm sạch bản thu dữ liệu tại bước tiền xử lý là một bước rất quan trọng. Để làm sạch một tín hiệu, nhiệm vụ đầu tiên là phải khử nhiễu, xoá những nhiễu nền như tiếng nói, quạt, bước chân. Turan và Erzin đã áp dụng bộ lọc dải thông cao FIR để xoá âm thanh nói và những nhiễu tần số thấp trong bản ghi[41]. Ferretti giảm nhiễu bằng một bộ lọc chùm tia tổng hợp và sử dụng bộ lọc OMLISA để giảm nhiễu khuếch tán còn sót lại. Trong 16, Gu đã sử dụng tối ưu Blackman window để xử lý từng tín hiệu khung, đây sẽ là kết quả sau khi phát hiện điểm cuối. Nhiễu tín hiệu thực sự giảm sau bộ lọc.

Nhiệm vụ phân đoạn tín hiệu được thực hiện thông thường sử dụng Voice Activity Detection(VAD). Kỹ thuật VAD được sử dụng rộng rãi trong xác định tiếng nói để phát hiện tiếng nói con người trong tín hiệu âm thanh. Những nhà nghiên cứu cũng sử dụng nó để phát hiện tiếng khóc trẻ sơ sinh và xoá im lặng trong quá trình thu âm một mẫu. VAD cũng đối mặt thử thách về phân loại tiếng khóc và nhiễu. Pan và cộng sự sử dụng nó để phát hiện sự có mặt hoặc vắng mặt của tiếng khóc trẻ em trong một môi trường nhiễu để cải thiện tỷ lệ nhận dạng tiếng khóc nói chung và nó được sử dụng để phát hiện những đoạn âm thanh với có đủ hoạt động âm thanh. Trong 41, những tác giả triển khai một thuật toán VAD căn bản, sử dụng đặc trưng ngắn hạn của khung âm thanh và chiến lược quyết định cho những khung im lặng và âm thanh xác định. Thỉnh thoảng, các nhà nghiên cứu cũng thường cắt các mẫu để xoá phần im lặng và những âm phần giao thoa âm thanh, và chỉ những phần khóc liên tục được dữ lại.[51]

**Trích xuất đặc trưng**

Tín hiệu khóc trẻ sơ sinh khác tiếng nói người lớn. Hình 2 đưa ra một sự so sánh về phổ âm thanh giữa những âm trẻ sơ sinh và tiếng nói người lớn. Chúng ta có thê thấy rằng sự biến đổi giữa dạng sóng và quang phổ là khá khác nhau, đặc biệt ỏng những khu vực năng lượng, cường độ và các thành phần. Nhìn chung, tiếng khóc trẻ sơ sinh là một sự kết hợp của giọng nói, sự im lặng, ho, nghẹt thở và gián đoạn, bao gồm sự đa dạng về thông tin âm thanh và âm thanh ở cấp độ khác nhau. Nó là cách duy nhất để trẻ em giao tiếp với thế giới bên ngoài.

Trích xuất đặc trưng giai đoạn trích xuất những đặc trưng riêng biệt từ tín hiệu âm thanh và sau đó đưa tới những thuật toán học máy. Nó là một trong những phần quan trọng nhất trong quy trình học máy. Thực hiện nhiệm vụ trích xuất đặc trưng cả trong miền thời gian và tần số giải quyết công việc cơ bản là phân tích và xử lý tiếng khóc trẻ em. Đặc trưng miền thời gian, chẳng hạn như tỷ lệ vượt 0, biên độ, và những đặc trưng dựa trên năng lượng, là đơn giản và tính toán rõ ràng. Trong khi đặc trưng miền thời gian là không đủ mạnh để bao phủ sự biến đổi với tín hiệu khóc trẻ sơ sinh và những đặc trưng nhạy với những nhiễu nền, thì đặc trưng miền tần số có một khả năng mạnh mẽ để mô hình hoá các đặc trưng trong tiếng khóc trẻ em. Thông thường sử dụng MFFCs, LPCCs, và LFCCs có hiệu năng tốt hơn sử dụng đặc trưng miền thời gian. Mặt khác, nó chỉ ra rằng tín hiệu khóc là nhịp ngàng và có sự thay đổi theo chu kỳ vì sự gián đoạn tự nhiên và hơi thở. Thông tin mức cao, như đặc trưng prosodic, là quan trọng để cải thiện khả năng phân biệt tín hiệu. Do đó, những đặc trưng miền prosodic cùng với miền thời gian, tần số có khả năng nắm bắt cả những thông tin vật lý và sinh học. Ngoài ra, quang phổ là một ảnh, là một sự biểu diễn miền thời gian-tần số của một đoạn âm thanh. Nó có một khả năng mạnh mẽ để thể hiện tín hiệu và bao gồm cả thông tin âm thanh và prosodic.

Hình 3 mô tả các dan mục chính của những đặc trưng âm thanh, được áp dụng để nghiên cứu liên quan đến âm thanh, âm nhạc, và âm thanh môi trường. Những đặc trưng âm thanh và âm sắc thông thường được sử dụng cho phân loại và phát hiện tiếng khóc trẻ sơ sinh. Đặc trưng miền quang phổ, đặc trưng prosodic, và những đặc trưng dựa trên hình ảnh được sử dụng rộng rãi trong xử lý tiếng khóc với tỷ lệ hơn 70% các bài báo nghiên cứu. Trong phần này, chúng tôi đánh giá trích xuất đặc trưng dựa trên những công trình mới nhất. Những giải thích chi tiết và những thuật toán về đặc trưng âm thanh có thể tìm thấy trong 58 và 59.

**Cepstral domain features**

Mel-frequency cepstral coefficient (MFFC) được sử dụng rộng rãi trong phân loại tiếng nói. Nó là một đại diện phổ của tín hiệu âm thanh. Những nhà nghiên cứu sử dụng nó để kiểm tra những phương pháp tiếp cận được đề xuất và thường sử dụng nó cho những thử nghiệm cơ bản. Liu và cộng sự đã sử dụng MFFC với 2 đặc trưng khác LPCC vav BFCC để xoá dư thừa từ tín hiệu và cố gắng dự đoán giá trị tiếp theo bằng kết hợp tuyến tính các hệ số đã biết từ trước đó. Nó được sử dụng trong 16 để phát hiện khóc. Quy trình trích xuất lFCC tương tự MFCC. Sự khác biệt là nó sử dụng linear filter-bank thay vì Mel filter bank. Trong 22 và 65, tác giả chỉ ra rằng LFCC hiệu suất tốt hơn MFCC trong phân biệt tín hiệu âm thanh tần số cao như: giọng nói nam và trẻ em khóc. Trong 24, Singh và cộng sự khám phá MFCC và dư và các đặc điểm dư LP ngầm thể hiện thông tin nguồn kích thích. Những nhà nghiên cứu cũng thử những đặc trưng phổ khác như Fast Fourier Transform(FFT), đặc trưng Log-Mel, Mel Scale, Constant-Q Chromagram, Log-mel spectrum, delta spectrum.

Theo những mô hình cảm nhận thính giác, hệ số MFFC là mạnh mẽ hơn hệ số khác như hệ số LPC. Trong công trình trước đây của chúng tôi, đặc trưng MFCC thông thường và bất thường của trẻ sơ sinh trong một khung nhất định kết hợp với 12 thứ tự được vẽ trong một khoảng trống. Quan sát thấy rằng những đặc trưng âm sắc của tín hiệu khóc bình thường khá khác so với các tín hiệu bị ngạt như trong hình 4. Í chỉ ra rằng khoảng giá trị và khuynh hướng của những đặc trưng âm thanh của tín hiệu khóc bình thườn và bị ngạt là khác nhau.

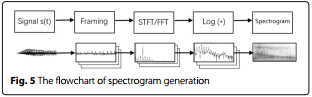
**Các đặc trưng miền prosodic**

Nó chỉ ra rằng tiếng khóc trẻ sơ sinh được tạo từ 4 kiểu âm: một âm từ giai đoạn hết hạn, một khoảng dừng ngắn, một âm từ đoạn truyền cảm hứng, tiếp theo là một khoảng dừng khác. Những biến đổi trong cường độ, tần số cơ bản, công thức và thời lượng là các dấu hiệu âm thanh điển hình mang thông tin về giọng nói và tiếng khóc trẻ sơ sinh. Nó chỉ ra rằng trên những đặc trưng prosodic có ảnh hưởng đến định danh kiểu khóc. Tần số cơ bản người lớn là từ 85 đến 200Hz trong tiếng khóc trẻ sơ sinh là từ 250-700Hz. F0 được tính toán một cách thông thường sử dụng một phương pháp dựa trên sự tương quan đề xuất bởi Praat.

Trong công trình trước đây của chúng tôi đã chỉ ra rằng sự kết hợp những đặc trưng prosodic trọng số với đặc trưng MFCC giúp cải thiện sự chính xác phân loại trong mô hình học sâu. Những nghiên cứu khác cũng tìm ra rằng F0 là rất quan trọng trong xác định tiếng khóc trẻ sơ sinh. Chittora và Patil đã sử dụng F0 để tính toán tỷ lệ những phân đoạn không có âm thanh và tìm ra tỷ lệ phần trăm không âm thanh trong một lần khóc là một thông số quan trọng để phân tích tiếng khóc của trẻ sơ sinh. Orlandi và công sự sử dụng trung bình, trung vị, độ lệch chuẩn và tối thiểu, tối đa của FO à F123 để khám phá ra sự khác biệt giữa tiếng khóc đủ tháng và thiếu tháng. Trong 2017, Torres và cộng sự đã sử dụng những đặc trưng thủ công (đếm âm thanh/không âm thanh, liên tiếp F0, tích harmonic ratio accumulation) để chỉ ra hiệu suất phát hiện có thể so sánh được nhưng kết quả trong 20 lần thấp hơn MFCC chuẩn.

**Đặc trưng miền hình ảnh**

Quang phổ là một hình ảnh, là đại diện miền tần số-thời gian của một đoạn âm thanh. Quang phổ có khả năng mạnh mẽ để biểu thị tín hiệu và bao gồm cả thông tin prosodic và acoustic. Quang phổ có thể được trích xuất thông qua khung, FFT, và tính toán logarit của bước lọc phổ được minh hoạ trong hình 5. Đưa quang phổ vào phân tích có thể giải quyết vấn đề về những tiếng khóc khác nhau có thời lượng khác nhau. Thay vì sử dựng thêm 0 để đạt được chiều dài giống với những vecto đặc trưng, normalization được áp dụng trong quy trình tạo quang phổ, tạo ra hình ảnh có cùng kích thước mà không làm thay đổi tín hiệu ban đầu. Bên cạnh đưa quang phổ vào CNN, những nhà nghiên cứu thêm bước để sử dựng hình ảnh quang phổ để nhận thêm đặc trưng như LBP, LPQ, RLBP để giúp cải thiện hiệu xuất phân lớp.



Hình ảnh sóng đại diện những mô hình biên độ áp suất âm than trong miền thời gian. Nó cũng được sử dụng trong những mô hình deep learning như AlexNet để đạt được độ chính xác trên 90% trong xác định tiếng khóc bệnh lý. Trong những công trình trước đây của chúng tôi, chúng tôi sử dụng Praat để tạo ảnh chức những đường đặc trưng prosodic bao gồm F0, cường độ, và formants. Hình ảnh đặc trưng Prosodic mô hình CNN tốt trong việc xác định một số loại tín hiệu tiếng khóc. Kết hợp nó vói CNN quang phổ và CNN sóng tạo độ chính xác tốt hơn 5% trên tập dữ liệu Baby Chillanto và 4% trên Dunstan Baby Language.

**Những đặc trưng liên quan khác**

Những đặc trưng miền khác được sử dụng trong nghiên cứu tiếng khóc trẻ em bao gồm những đặc trưng miền thời gian như zero crossing rate, short-time energy, voiced-unvoiced regions. Zero-crossing rate là tỷ lệ những tín hiệu từ dương, về 0, về âm hoặc từ âm , về 0, về dương. Nó có thể được sử dựng kết hợp với short-time energy để phát hiện điểm cuối của âm thanh nói, do đó để phát hiện sự có mặt của âm thanh khóc từ những âm thanh khác xảy ra trong môi trường. Bởi vì biên độ của tín hiệu thay đổi theo thời gian, short-time energy có thể phục vụ để phân đoạn không âm thanh và có âm thanh. Nó được sử dụng trong 20,57,70 cho việc phát hiện và phân loại tiếng khóc trẻ sơ sinh. Torres và cộng sự sử dụng voiced-unvoiced counter, đếm tất cả những khung có chu ký thay đổi đáng kể, nhưng một trong những đặc trưng trong phát hiện tiếng khóc. Linear Predictive Coding (LPC) đóng vai trò như một thước đo miền thời gian về mức độ gần nhau của hai dạng sóng khác nhau và được sử dụng trong phân loại tiếng khóc trẻ sơ sinh.

Wavelet Transform là một phương pháp để chuyển đổi tín hiệu âm thanh thành miền thời gian-tần số. Biến đổi gói dạng sóng được xử dụng trong nghiên cứu phân loại ngạt và đạt được mức chính xác cao 99% với những mô hình mạng nơ ron. Nó cũng thực hiện tốt trong phân loại nhũng lý do khóc. Đặc trưng Discrete Wavelet Transform MFCC (DWT-MFCC) là việc tốt với SVM và kiến trúc mạng nơ ron.

Những nhà nghiên cứu cũng tính toán những tham số tự nhiên thống kê của dữ liệu như: tần số trung bình, độ lệch chuẩn, third quartile range để giúp việc phát hiện và phân loại tiếng khóc. Trích xuất đặc trung là một bước rất quan trọng trong xử lý âm thanh. Bên cạnh phần mềm Praat, những công cụ trích xuất đặc trưng như thư viện LibROSA và bộ công cụ OpenSMILE làm việc trích xuất đặc trưng trở nên đơn giản hơn.

**Lựa chọn đặc trưng**

Lựa chọn đặc trưng là quy trình lựa chọn một tập những đặc trưng từ những đặc trưng ban đầu được trích xuất từ tín hiệu âm thanh, nó được sử dụng trong kỹ thuật trích xuất đặc trưng. Mục đích là giảm chiều của những dặc trừng mà không làm giảm độ chính xác trong phân lớp. Ít đặc trưng hơn, yêu cầu ít tài nguyền tính toán hơn, do đó việc xây dựng các thiết bị phân loại và phát hiện thông minh trở nên khả thi và giá cả phải chăng hơn trong tương lai. Những đặc trưng ban đầu cũng có thể chức một vài thông tin dư thừa, thể hiện hiệu quả sự khác nhau trong những kiểu khóc khác nhau. Lựa chọn đúng đặc trưng để phù hợp với những yêu cầu cụ thể của những nhiệm vụ có thể cải thiện độ chính xác phân lớp. Phần này đanh giá một vài phương pháp lựa chọn đặc trưng được áp dụng trong nghiên cứu tiếng khóc trẻ em. Phương pháp F-ratio, được sử dụng để lựa chọn những đặc trưng MFCC trong nhóm 20. Những hệ số có tầm quan trọng đánh kể sẽ có điểm F-ratio cao hơn. Trong 2013, Yamamoto và công sự đã sử dụng Principal Component Analysis (PCA) để giảm chiều của đặc trưng FFT. Phương pháp Forward variable Selection (FSM) được áp dụng cho phân loại tiếng khóc bởi Wang năm 2010. Okada và cộng sự đề xuất Iterative FSM (IFSM) dựa trên cross-validation trong 2011. Sau đó, Binary Particle Swarm Optimization (BPSO) được sử dụng để xoá những đặc trưng dư thừa và những đặc trưng đáng kể từ những hệ số MFCC. Orlandi và công sự sử dụng một phần mềm gọi là Biovoice để trích xuất 22 đặc trưng từ tín hiệu khóc và sử dụng một phương pháp tìm kiếm dựa trên thuật toán tổng hợp để lựa chọn những đặc trưng tốt nhất đưa vào phân loại. Trong 2016, Wahid và công sự so sánh 5 phương pháp lựa chọn đặc trưng: OneR, ReliefF, Fast Correlation-Based Feature Selection (CFS). Nó chứng minh rằng kỹ thuật lựa chọn đặc trưng có thể giảm không gian đặc trưng, dẫn đến giảm thời gian tính toán. Hầu hết kỹ thuật lựa chọn cũng có thể cải thiện hiệu xuát của phân lớp mạng nơ ron.

Trong 2019, Tuduce và cộng sự sử dụng 3 cách tiếp cận Best Feature Selection(BFS) để loại trừ các tính năng không liên quan và các tính năng dư thừa và kiểm nghiệm chúng với 35 bộ phân loại. Tập đặc trưng được giảm từ hơn 6000 đặc trưng xuống 500 và kết quả chỉ ra rằng BFS có thể cải thiện độ chính xác phân lớp cho một vài phân lớp. Kỹ thuật lựa chọn đặc trưng xoá những đặc trưng không liên quan tới những bài toán, nên nó có thể giảm không gian đặc trưng, tiết kiệm thời gian tính toàn và cải thiện độ chính xác phân lớp.

**Phân loại tiếng khóc trẻ sơ sinh**

Với dữ liệu được làm sạch và phân đoạn và trích xuất đặc trưng, lựa chọn, chuẩn hoá, đưa vào bộ phân lớp thích hợp một giai đoạn quan trọng trong quy trình học máy. Trong phần này, chúng tôi đánh giá những phương pháp học máy phổ biến được sử dụng trong phân loại tiếng khóc trong thập kỷ gần đây.

**Mô hình phân loại tiếng khóc trẻ sơ sinh**

**Mô hình học máy truyền thống**

1. **Support Vector Machine(SVM):** Bộ phân loại xác xuất phổ biến nhất được sử dụng trong phân loại tiếng khóc là SVM. Những kiểu của SVM bao gồm multi-class SVM, linear, RBF kernels binary SVM. Những đặc trưng dưa đến SVM bao gồm những đặc trưng tạm thời, đặc trưng prosodic, đặc trưng phổ. Trong 2017, Onu và công sự so sánh SVM với những phân lớp phi tuyến khác như mạng no ron ở phân loại asphyxia và kết luận tằng SVM được thiết kể để làm việc hiệu quả với những mẫu hạn chế và dữ liệu chiều cao. Trong 2015, Chang và cộng sự sử dụng mô hình học SVM tăng dần, tiếp tục thêm dữ liệu mới vào tập dữ liệu trong mỗi bước đào tạo, đưa kết quả cao hơn 18% với SVM truyền thống trong phân loại dựa trên đặc trưng FFT.
2. **K-Nearest Neighbor (KNN):** là một phương pháp nhận dạng mẫu nổi tiếng được sử dụng trong phân lớp. Có k lân cận gần nhất trong không gian đặc trưng. Mục đích là để chỉ định mẫu để phân lớp tới lớp gần nhất. Trong trường hợp phân loại tiếng khóc, những nhà nghiên cứu sử dụng khoảng cách Euclidean, Minkowski, và những phương pháp khác để đo khoảng cách giữa hai vecto đặc trưng mẫu. Những vecto đặc trưng được lựa chọn thường là MFCC và LFCC. Cohen và Lavner sử dụng thuật toán KNN, trong mỗi khung phân loại khóc và không khóc, và mẫu được phân loại là khóc nếu nhiều khung trong mẫu được xác định là khóc.
3. **Gausian mixture model (GMM):** GMM là một mô hình xác xuất giả định rằng những điểm dữ liệu trong phân phối Gauu của một giá trị trung bình và phương sai. Ý tưởng là tìm hiẻu các tham số để mô hình hoá dữ liệu đào tạo được cung cấp dưới dạng hỗn hợp của một số phân phối Gauss. Thuật toán Expectation Maximization(EM) được sử dụng để tìm kỳ vọng xảy ra lớn nhất của các thông số trong cấu trúc dựa trên GMM. Trong 2016, Banica và cộng sự sử dụng mô hình GMM-UBM để phân loại Dunstan baby cries. The universal background model (UBM) là một mô hình GMM được đào tạo trong một số lượng lớn những tín hiệu khóc chung không được gán nhãn cụ thể. Độ chính xác phân lớp của GMM-UBM với MFCC là 70% trên Dunstan baby và 50,6% trên SPLANN database. GMM-UBM cũng được sử dụng bởi Alaie và cộng sự để phân loại tiếng khóc bệnh lý và tiếng khóc mạnh khoẻ. Phương pháp Boosting Mixture Learning đề xuất vượt xa hiệu xuất của thuật toán MAP. Trong năm 2019, Sharma và cộng sự đã so sánh phân cụm GMM và phân cụm phân cấp và phân cụm K-means trên những đặc trưng khóc và chỉ ra rằng mô hình GMM đạt kết quả tốt nhất với ít điểm dữ liệu chồng chéo với một tập dữ liệu cố định. Điều này chỉ ra rằng dựa trên phân lớp GMM nhạy cảm với môi trường và không thể dẫn đến những kết quả hài lòng với dữ liệu giới hạn.
4. **Phân lớp Fuzzy:** Những hệ thống logic Fuzzy được sử dụng trong nhiều ứng dụng như: hệ thống chuyển đổi, hệ thống năng lượng, và định tuyến mạng không dây. Nó cũng được sử dụng trong phân loại tiếng khóc. Những đặc trưng được lựa chọn được chuyển từ giá trị fuzzy trong bước fuzzycation, những hàm thành viên fuzzy cố định được sử dụng, và quy luật fuzzy được định nghĩa. Trong 66, Kia và cộng sự sử dụng phân lớp fuzzy để phân loại tiếng khóc từ tín hiệu cười. Trong 71, Rosales và cộng sự sử dụng fuzz decision tree, fuzzy decision forest, fuzzy KNN, và phân lớp mạng nơ ron tương quan fuzzy cho phân loại tiếng khóc bệnh lý. Thuật toán đối sánh mẫu fuzzy loại 2 được sử dụng trong 80 để phân loại ngạt, bình thường và tăng máu. Nó cũng vượt hiệu xuất SVM và phân lớp hồi quy logistic trong phân loại đói và đau.
5. **Phân lớp hồi quy logistic:** là một thuật toán giám sát phức tạp thấp, thường được sử dụng như một thử nghiệm tham khảo trong nghiên cứu tiếng khóc trẻ. Lavner và cộng sự sử dụng chúng để chỉ ra rằng CNN thực hiện tốt hơn trong phân loại và Orlandi và cộng sự sử dụng nó để so sánh ới nhiều bộ phân lớp khác, trong đó random forest đạt hiệu xuất cao nhát trong phân loại tiếng khóc trẻ sơ sinh đủ tháng và thiếu tháng.
6. **K-means clustering**: là một thuật toán không giám sát được sử dụng chủ yếu trong phân cụm. Điểm dữ liẹu không gán nhãn có thể được phân loại thành nhóm dự trên giá trị trung bình và di chuyển tâm cụm. Sharma và cộng sự sử dụng phân cụm K-mean để chỉ ra rằng mô hình GMM có hiệu xuất phân biệt tốt hơn những loại khóc khác nhau. Trong 22, phân cụm K-means đã được sử dụng để xây dựng một dữ liệu tiếng nói cho nhận dạng giọng nói.
7. **Bagging, boosted trees, và random forest**: Bagging, boosted trees, và random forest là kỹ thuật thực hiện cây quyết định tổng hợp. Chúng kết hợp cả cây quyết định để đạt hiệu xuất tốt hơn. Thử nghiệm chỉ ra rằng kỹ thuật trên mạnh mẽ trong phân loại tiếng khóc. Osmani và cộng sự chỉ ra bagging và boosted trees vượt trội SVM. Milano và cộng sự so sánh nó ới MLP và SVM, Reservoir Network, GMM, HMM và chỉ ra rằng bộ phân lớp random forest bên cạnh Reservoir Network. Trong 21,45, một phần mềm khai phá dữ liệu mở có tên là Waikato Environment cho Knowledge Analysis (WEKA) được sử dụng. Giữa hơn 100 thuật toán phân loại được triển khai trong WEKA, random forest vượt hiệu xuất SVM, MLP, hồi quy logistic, BayesNet. Tuduce và cộng sự đã kiểm tra 40 bộ phân lớp trong WEKA và bộ phân lớp cây đưa ra hiệu xuất tốt nhất so với Bayes, lazy, function, rule.

**Mô hình dựa trên mạng no ron**

Mạng no ron nhân tạo (ANN) là phương pháp học máy. Trong 1995, Petroni và cộng sự đã thực hiện những nỗ lực đầu tiên về ANN trong phân loại tiếng khóc.

1. **Feed Forward Neural Network(FFNN):** là mạng nơ ron đơn giản nhất và Multi Layer Perceptron(MLP) là một dạng của FFNN, chứa ít nhất 3 lớp. Thử nghiệm trong 37 và 13 đã chỉ ra rằng hiệu suất của FFNN không tốt bằng phân lớp lân cận gần nhất dựa trên đặc trưng MFCC. MLP được sử dụng trong 52,61-63 với MFCC để xác định những tiếng khóc bệnh lý. Để phân loại ngạt, Hariharan và cộng sự sử dụng Probabilistic Neural Network(PNN), General Regression Neural Network (GRNN), và Time-Delay Neural Network (TDNN) và đạt độ chính xác trên 97%.
2. **Convolutional Neural Network:** là một thuật toán học sâu, đã sử dụng thành công trong thị giác máy tính, xử lý ngôn ngữ, và những lĩnh vực khác đạt được độ chính xác cao chưa từng có. CNN đa kênh, chấp nhận đầu vào đa kênh được áp dụng trong 11. Manikanta và cộng sự sử dụng CNN 1D trên những đặc trưng MFCC cho phát hiện khóc và kết quả vượt trội so với mạng no ron chuyển tiếp và bộ phân lớp SVM. Trong 2019, Le và cộng sự áp dụng học chuyển đổi với CNN trên quang phổ trên Baby Chillanto và đạt được kết quả đầy hứa hẹn.
3. **Long Short-Term Memory (LSTM):** là một dạng của Recurrent Neural Network(RNN), có những trạng thái bên trong để chấp nhận trình tự của dữ liệu khả thi và được biết đến như mạng nơ ron tốt nhất cho dữ liệu chuỗi thời gian như dịch ngôn ngữ,, và phân biệt giọng nói. Mark Huckvale đã đưa những đặc trưng tạm thời tín hiệu mức thấp tới mô hình LSTM bidirectional và sau đó kết hợp với mạng nơ rơn 2 lớp ẩn khác trong 40. Cả mạng kết hợp và LSTM đều đạt hiệu suất vượt trội so với mô hình SVM đường cơ sở.
4. **CNN**-RNN: là một kiến trúc deep learning kết hợp CNN với RNN, là một công cụ mạnh trong phân loại và phát hiện âm thanh. Trong phân loại và phát hiện về Âm thanh, Lim và cộng sự sử dụng nó để thắng giải nhất trong phát hiện âm thanh mục tiêu (tiếng khóc trẻ, vỡ kính, tiếng súng) với nhiễu nền. Trong 2019, Maghfira và cộng sự sử dụng CNN-RNN để phân loại 5 kiểu của tiếng khóc và đạt được độ chính xác cao nhất 94.97% trên tập dữ liệu Dunstan Baby Languages.
5. **Neuro-fuzzy Network:** Kết hợp fuzzy logic với mạng nơ ron và nó được sử dụng thành công bởi những nhà nhiên cứu trong phân loại tiếng khóc. Năm 2009, Santiago-Sanchez và cộng sự đã sử dụng fuzzy kiểu 1 để phân loại ngạt và thiếu máu. Trong 2012, Molaeezadeh và cộng sự đề xuất bộ phân lớp đối sánh mẫu fuzzy loại 2 và nó vượt trội SVM và bộ phân lớp hồi quy logistic trong phân loại đói và đau. Trong những năm gần đây, nhận thấy rằng, kết hợp hệ thống fuzzy với mạng no ron có thể hợp nhất những ưu điểm của chúng và né tránh những nhược điểm của cả hai phương pháp. Hệ thống Fuzzy yêu cầu nhưng quy tắc trong khi mạng nơ ron trực tiếp học từ dữ liệu. Neuro-fuzzy được sử dụng để phân loại kiểu khóc trong tập dữ liệu Dunstan baby. Mạng no ron được đào tạo và Mandani fuzzy logic được chấp nhận sau khi chuẩn hoá dữ liệu để tạo mới “dữ liẹu đã được chuyển đổi”, được sử dụng cho bước cuối của KNN. Độ chính xác phân lớp đạt tới 86,25% tốt hơn mô hình mạng no ron thông thường, SVM và GMM.
6. **Capsule Network:** là một cấu trúc liên kết học sâu bổ sung một cấu trúc được gọi là capsule và mô hình CNN. Như maxpooling trong CNN chỉ chọn những giá trị lớn nhất trong vùng và ném các thông tin vào các vị trí nhất định, capsule mức cao hơn phue vùng rộng hơn của ảnh và thay vào đó thực hiện định tuyến theo thoả thuận. CapsNet được áp dụng để phân loại khóc cảm xúc của trẻ trong môi trường gia đình và độ chính xác cải thiện hơn 10% mô hình CNN với quan phổ.
7. **Reservoir Network (RN):** là một mô hình mạng no ron nguồn gốc từ RNN. Những nốt đầu vào của nó kết nối tới một non-trainable reservoir, chứa những phần phi tuyến đã được kết nối với những trọng số đã được tạo ra ngẫu nhiên. Ntalampiras sử dụng RN trong phân loại nhiều tiếng khóc với những đặc trưng hợp nhất và chỉ ra rằng RN vượt trội so với MLP, SVM, random forest, phân cụm GMM, etc.

Nhiều phương pháp học máy đã thử nghiệm trong nghiên cứu trẻ sơ sinh. Mỗi phương pháp có những ưu nhược điểm và không thuật toán nào hoàn hảo cho mọi tập dữ liệu và nhiệm vụ. Chọn một mô hình phù hợp để đạt hiệu suất cao là một thử thách. Để xác định khả năng phân loại của những mô hình khác nhau, Furh và cộng sự thử nghiệm khóc trẻ sơ sinh khoẻ mạnh khác nhau và khóc của trẻ bị bệnh bằng cách sử dụng 12 bộ phân loại gồm SVM, cây quyết định, KNN, MLP. Kết quả chỉ ra rằng cây quyết định và KNN đạt độ chính xác trên 90%. Áp dụng nhiều thuật toán vào nhiệm vụ trước khi lựa chọn thuật toán để sử dụng impractical. So sánh những thuật toán học máy được sử dụng trong nghiên cứu tiếng khóc, chúng tôi phân tích chúng từ các khía cạnh. Người đọc có thể chọn thuật toán phù hợp theo dữ liệu và nhiệm vụ của họ.

* Độ phức tạp thời gian. Nó bao gồm thời gian đào tạo và thời gian phân lớp phản hồi trên kích thước dữ liệu, không gian tìm hiếm, và những độ phức tạp của hệ số. Nhìn chung, những phương pháp truyền thống như SVM, K-mean clustering, GMM là tương đối đơn giản và mạch lạc. Kích thước mẫu nhỏ hơn được chấp nhận, khác so với những phương pháp mạng nơ ron. Do dó, thời gian đào tạo, thời gian tìm kiếm, và thời gian phân lớp nhỏ hơn nhiều những phương pháp mạng no ron. Ngoài ra, viẹc tinh chỉnh trong các mô hình mạng no ron cũng đòi hỏi nhiều thời gian phát triển hơn
* Độ phức tạp mẫu: Nó cho biết mô hình có yêu cầu kích thước dữ liệu lớn hay không để học. Nó phụ thuộc vào độ phức tạp của dữ liệu và độ phức tạp của thuật toán. Để đạt hiệu suất cao hơn, những phương pháp mạng no ron nhìn chung yêu cầu kích thước mẫu lớn hơn cho không gian tìm kiếm phức tạp hơn những thuật toán truyền thống khác. Kích thước dữ liệu lớn hơn cần cho những mạng no ron sâu.
* Tham số. Nó cho biết nếu số lượng tham số được sử dụng trong mô hình được sửa hoặc nó thay đổi theo khi dữ liệu mới được đưa vào. Hồi quy tuyến tính, GMM, mạng nơ ra là những phương pháp tham số, trong khi KNN và SVM là những phương pháp không tham số.
* Độ phức tạp đặc trưng: Những đặc trưng được trích xuất từ cả miền thời gian và miền tần số có khả năng giống nhau để thể hiện những đặc tính khác nhau của tin hiệu khóc trong những mô hình khác nhau. Không có sự khác nhau về độ phức tạp đặc trưng liên quan đến các mô hình truyền thống và những mô hình dựa trên mạng no ron. Nhưng sử dụng nhiều đặc trưng để đại diện một mẫu có thể dẫn đến vấn đề quá khớp, do đó, lựa chọn những đặc trưng phù hợp nhất cho các mô hình cụ thể là rất quan trọng.
* Parallelizability. Là một đặc trưng quan trọng để tiết kiệm thời gian đào tạo cho những mô hình học máy. Một lượng lớn dữ liệu trong những mạng no ron được liên kết với chi phí tính toán cao trong cả thời gian và không gian. Paralleizability với GPU tính toán giảm thời gian đào tạo và tạo mô hình khả thi. Những phương pháp khác như KNN dễ dàng để song song nhưng song song là khó nếu bước tiếp theo dựa trên kết quả của bước trước như cây quyết định.

Với môi trường tính toán mạnh mẽ hiện nay, những phương pháp sử dụng trong phân loại tiếng khóc có thể đạt được dự đoán thời gian thực. Vì lượng mẫu hạn chế trong dữ liệu phân trẻ khóc hiện nay, thời gian đào tạo và kiểm tra chưa được đánh dấu như một vấn đề. Không có những mô hình rất sâu với những dữ liệu lớn liên quan trong nghiên cứu. Hiện tại, lượng dữ liệu lớn nhất nhỏ hơn 20000 mẫu. Dữ liệu nhỏ và không cân bằng dẫn đến độ chính xác phân loại cao nhưng độ tin cậy thấp với một vài nhiệm vụ. Để đạt được hiệu suất tin cậy cao hơn, dữ liệu thưc với những mô hìn học sâu thực cần được khám phá.

**Ứng dụng tiếng khóc trẻ sơ sinh**

Những nhà nghiên cứu sử dụng các bộ phân loại khác nhau để thực hiện nhiệm xử lý tiếng khóc trẻ em. Trong thập kỷ gần đây, một số công trình nghiên cứu tiêp tục cải thiện độ chính xác phân lớp về tín hiệu khóc trẻ sơ sinh bao gồm những tiếng khóc bệnh lý khác nhau từ khóc thông thường và hiểu ý nghĩa những tín hiệu khóc. Trong phần này, chúng ta đánh giá những công trình đáng kể trong phân loại và phát hiện tiếng khóc trẻ em.

**Phân loại lý do khóc trẻ sơ sinh**

Trong những năm gần đây về nghiên cứu tiếng khóc trẻ em, nhiều công trình đã được thực hiện một cách tự động những tiếng khóc mạnh khoẻ từ những tiếng khóc bệnh lý. Trong những năm gần đây, khám phá ý nghĩa của tiếng khóc thu hút nhiều quan tâm nghiên cứu. Bảng 2 chỉ ra, một vài công trình đáng kể được hoàn thành trong chủ đề này. Nó đáng chú ý rằng những nhà nhiên cứu sử dụng những tập dữ liệu khác nhau, hầu hết là tự ghi âm. Với những tập dữ liệu khác trong nghiên cứu tương tự, thậm chí những kiểu phân loại là giống nhau, nó không công bằng để tạo so sánh trực tiếp trên những phương pháp đề xuất. Phân loại tiếng khóc trẻ em vẫn đang là thử thách vì thiếu dữ liệu chuẩn công khai và độ chính xác vẫn tương đối thấp.

**Phân loại tiếng khóc bệnh lý trẻ sơ siinh**

Những tín hiệu khóc trẻ sơ sinh đã sử dụng để xác định nhiều bệnh lý như ngạt, thiếu máu, .. Người đọc có thể tìm những công trình liên quan trong phân loại tiếng khóc bệnh lý trước 2011 trong 3. Trong thập kỷ gần đây, những nhà nghiên cứu tiếp tục áp dụng những phương pháp mới để phân loại tiếng khóc bình thường và khóc bệnh lý. Khóc ngạt là bệnh lý phổ biến nhất trong nghiên cứu. Bảng 3 chỉ ra rằng những công trình mới nhất trong phân loại tiếng khóc bình thường từ khóc ngạt. Những nhà nghiên cứu sử dụng tập dữ liệu Baby Chillanto để thực hiện phân loại nhị phân. Trong 2012, Probabilistic Neural Network(PNN) và General Regression Neural Network (GRNN) đạt tới độ chính xác 99%, mô hình SVM mới nhất đạt tới 97.7% và mô hình học sâu FFNN đạt tới 96.64%.

Bên cạnh xác định ngạt, các kiểu bệnh lý khác cũng được nghiên cứu. Theo đánh giá của Esposotp, chỉ ra rằng tín hiệu khóc của trẻ sơ sinh hữu ích cho chuẩn đoán sớm rối loạn tự kỷ (ASD). Trong 2012, Orlandi và cộng sự đã phân tích tiếng khóc của trẻ nguy cơ cao, những người có anh chị em được chuân đoán mắc ASD. Người ta nhận thấy rằng các giai đoạn khóc ít xảy ra hơn, F0 nhỏ hơn, và Formants chạm tới giá trị cao trong những trẻ có nguy cơ cao hơn những trẻ khoẻ mạnh. Mặc dù một vài em bé sinh ra đã mắc ASD, bệnh thường được chuẩn đoán khi trẻ 2-3 tuổi khi những chuẩn đoán liên quan giám sát hành vi của trẻ. Điều này dẫn đến sự khó khăn về thu nhận tiếng khóc trẻ tự kỉ. Trong 2019, Wu và cộng sự đã thu 20 mẫu âm thanh về trẻ tự kỉ trong độ tuổi từ 2-3 tuổi. Chúng đạt độ chính xác đến 96% bằng cách sử dụng phân lớp SVM với đặc trưng MFCC. Xác định tiêng khóc giảm âm đã thành công trong những năm gần đây. Trong 2011, GRNN đã đạt 99% trên tập dữ liệu Baby Chillanto, và năm 2009, Reyes và cộng sự sử dụng evolutionary neural network đã đạt tới 100% trên tập dữ liệu Mexican-Cuba. Năm 2014, rosale và cộng sự sử dụng mô hình fuzzy và thuật toán genetic đạt được 99,42% trên tập đữ liệu Baby Chillanto. Những kiểu bệnh lý khác như suy giáp, hội chứng suy hô hấp, hở hàm ếch đã được nghiên cứu trong những năm đầu và được xem xét trong 3. Trong 2014, Feier và cộng sự đã nghiên cứu tiếng khóc trẻ sơ sinh trong vòng vài phút sau khi sinh. Các phương pháp Random tree và random forest có thể phân loại tiếng khóc trẻ sơ sinh khoẻ mạnh từ những trẻ sơ sinh thiếu tháng, trẻ sơ sinh với dây rốn thắt cổ trong khi sinh và trẻ sơ sinh với những bệnh lý khác đạt độ chính xác trên 95%.

**Phát kiện tiếng khóc**

Phát hiện tiếng khóc được xem xét như một phân loại nhị phân với khóc và không khóc. Nó là một chủ đề nghiên cứu hấp dẫn khác trong thập kỷ gần đây. Mục đích là để phát hiện tiếng khóc hiệu quả, chính xác trong môi trường phức tạp như xe ô tô, nhà, bệnh viện, trong khi những âm thanh khác xảy ra cùng thời điểm. Dữ liệu được ghi âm trong suốt một quá trình dài về thời gian tròn một môi trường cố định như nhà hoặc bệnh viện. Thuật toán phát hiện cần có thể phát hiện âm thanh khóc cho dù âm thanh nền đang xảy ra trong môi trường. Những nhà nghiên cứu cũng đề xuất những phương pháp khác nhau để xâu dựng những nôi thông minh, có thể phát hiện tiếng trẻ em khóc và cảnh báo cha mẹ trong khi họ đi vắng. Những phương pháp được đề xuất không chỉ nhắm đến độ chính xác phát hiện cao hơn mà còn xem xét giá cả những hệ thống giám sát trẻ em để làm nó với mức giá hợp lý cho những gia đình thu nhập thấp.

Bảng 4 chỉ ra một vài công trình nghiên cứu đáng kể trong phát hiện tiếng khóc. Nó cho thấy rằng tiếp cận dựa trên mạng nơ ron đạt hiệu quả tốt trong điều kiện sạch và hạn chế. Mặt khác, vớ môi trường nhiễu và dữ liệu đào tạo giới hạn, các bộ phân lớp là nhạy với ranh giới và dễ bị nhầm lần và chồng chéo với các tín hiệu nhiễu.

**Các hướng trong tương lai và những thách thức**

Với sự cải thiện của khả năng tính toán tiếp cận học sâu, những thách thức vẫn còn trong nghiên cứu tiếng khóc trẻ sơ sinh.

* Thiếu dữ liệu và khả năng mở rộng của nghiên cứu. Những nhà nghiên cứu dựa trên những tập dữ liệu khác nhau tự thu bởi các tác giả. Do đó, khó để so sánh hiệu suất của những phương pháp đã được thử nghiệm trên các tập dữ liệu khác nhau. Chỉ những tập dữ liệu đã được chia sẻ bởi một vào nhà nghiên cứu như tập Baby Chillanto, đã được khoảng 2 thập kỷ. Tổng dữ liệu của Baby Chillanto là 2287 và tập dữ liệu cá nhân lớn nhất là dưới 20000 mẫu, không đủ cho các mô hình no ron học sâu. Dữ liệu là chìa khoá chính trong học máy đặc biệt học sâu. Chúng tôi lưu ý rằng một số phương pháp học sâu như CNN và CNN-RNN thường được sử dụng trong nghiên cứu tiếng khóc trẻ sơ sinh, kiến trúc của mô hình là không sâu. Lý do chính là những mô hình sâu không phù hợp với dữ liệu đào tạo nhỏ và dẫn đến hiệu suất kém. Để lấy ưu điểm của học sâu, dữ liệu quy mô lớn với những mẫu đầy đủ bao gồm những thay đổi đa dạng trong các đặc trưng âm thanh, âm vị của trẻ khác nhau là cần thiết.
* Thu thập dữ liệu và gán nhãn là một quy trình tốn nhiều thời gian và yêu cầu nhưng người được đào tạo. Hầu hết dữ liệu được sử dụng là tự ghi âm bởi tác giả và riêng tư cho một số người nhất định hoặc tổ chức. Mặc dù một số nguồn trực tuyến có sẵn như video trên Youtube, Google, hầu hết đoạn tín hiệu không có gán liên quan nhãn và nhiều bản ghi có tạp âm nền. Để tăng tốc quy trình xây dựng bộ phân lớp tự động, hệ thống nôi thông minh và hơn nữa để xây dựng bảo mẫu tự động, cần nỗ lực để tạo tập dữ liệu được gán nhãn và có cấu trúc tốt được công khai là cần thiết. Hơn nữa, dữ liêu bao gồm những mẫu từ trẻ em cụ thể có thể kiểm tra tiếng khóc của chúng ở những độ tuổi khác nhau là cần thiết. Kiểu của dữ liệu là thiết yếu để nghiên cứu đặc tính của tiếng khóc trẻ sơ sinh cùng với sự phát triển của chúng. Thiết lập những thiết bị ghi âm trong nôi của trẻ sơ sinh và thi âm thời gian thực sử dụng điện thoại di động bởi bảo mẫu là phương pháp chính được sử dụng thu thập dữ liệu. Ứng dụng di động phiên dịch tiếng trẻ khóc như ChatterBaby giúp dự đoán lý do trẻ khóc và làm việc thu thập dữ liệu dễ dàng hơn. Nó sẽ lợi ích cho sự phát triển trong nghiên cứu tiếng khóc trẻ em nếu một vài tập dữ liệu mới được thu thập có thể được công khai.
* Sự kết nối ngặt nghèo giữa chuyên gia y khoa và những nhà nghiên cứu làm giảm dần khả năng thúc đẩy lẫn nhau giữa các ngành. Những nhà nghiên cứu chứng minh rằng phân loại tiếng khóc trẻ em là một phương pháp không xâm lấn và có thể rất hữu ích trong một vài chuẩn đoán bệnh sớm như ngạt, … Nhưng hầu hết những nghiên cứu về bệnh lý với trẻ khóc được thực hiện trước năm 2010, và kích thước dữ liệu rất nhỏ. Sự khó khăn về thu thập dữ liệu có thể là thách thức lớn nhất trong lĩnh vực nghiên cứu này. Những vẫn đề về pháp lý và đạo đức liên quan đến quy trình thu thập dữ liệu cản trở sự phát triển trong nghiên cứu tiếng khóc trẻ em. Sự hợp tác giữa những chuyên gia y khoa và những nhà khoa học máy tính có thể tạo ra một vài cơ hội trong đề tài nghiên cứu này.

Chúng tôi gần đây xây dựng một tập dữ liệu lớn chứa những tiếng khóc trẻ sơ sinh từ 0 đến 9 tháng tuổi. Những đoạn khóc được ghi âm và gán nhãn bởi cha mẹ tại nhà và bởi bác sĩ, y tá tại bệnh viện sử dụng điện thoại di động. Giai đoạn thu thập dữ liệu gần đây và nó được kỳ vọng để dữ liệu chứa hơn 30000 mẫu sau 50h ghi âm, phù hợp với yêu cầu của mạng no ron sâu. Chúng tôi cũng áp dụng GNN để phân loại tiêng khóc. GNN đã được sử dụng trên nhiều lĩnh vực khác nhau va đồ thị có thể đại lại dữ liệu non-Eucliden với mối quan hệ phức tạp giữa những đối tượng. Kết hợp với học sâu, đã chứng minh hiệu quả thành công trong dữ liẹu Euclidean, mô hình học sâu GNN có thể tận dụng nhiều đặc trưng và có nhiều khả năng phân loại cho nhiệm vụ phân loại tiếng khóc. Hơn nữa, những kiến trúc học sâu nhúng với kiến thức trước đó có thể được khám phá. Với nhiều dữ liệu có sẵn trong tương lai, chúng tôi tin rằng phương pháp học máy có thể được khám phá trong lĩnh vực này. Kết hợp những phương pháp xử lý tín hiệu âm thanh mới và những phương pháp học máy mới sẽ dẫn nghiên cứu này tới tương lai đáng chú ý , sẽ thay đổi cuộc sống con người bằng cung cấp dịch vụ chăm sóc trẻ sơ sinh tự động giá cả hợp lý.

**Kết luận**

Trong bài báo này, chúng tôi mô tả những công trình nghiên cứu đáng kể trong phân loại và phân tích tiếng khóc trẻ em, cung cấp chi tiết và những tài nguyên hữu ích cho cả những nhà nghiên cứu và những chuyên gia y tế làm việc trong lĩnh vực này. Nó cũng chỉ ra rằng nguồn dữ liệu ạn chế cản trở sự phát triển trong nghiên cứu tiếng khóc trẻ sơ sinh. Dữ liệu lớn với số lượng mẫu đa dạng phù hợp với sự cần thiết của mạng nơ ron sâu. Xu hướng hiện tại cho trích xuất đặc trưng là tạo ra một tập đặc trưng hỗn hợp và tận dụng lợi thế các lĩnh vực khác nhau để đạt được khả năng phân biệt tốt hơn. Kết quả nghiên cứu liên quan chỉ ra rằng sự cải thiện đầy hứa hẹn với những đặc trưng kết hợp. Thêm nữa những kiến trúc dựa trên mạng no ron mới trở thành những phương pháp chính. Nó chứng minh hiệu suất mạnh mẽ hơn những tiếp cận học máy truyền thống. Trong tương lai, chúng tôi quan tâm trong việc tạo một dữ liệu lớn, trích xuất đặc trưng mạnh mẽ hơn, kết hợp những đặc trưng với tỷ lệ tốt, thành lập những kiến trúc mạng no ron mới với những kiến thức sử dụng trước đó cũng như thông tin khác từ các lĩnh vực liên ngành.