**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI**

**KHOA CÔNG NGHỆ THÔNG TIN**

**A yellow and red sign with a red circle and a star

Description automatically generated**

**ĐỒ ÁN CHUYÊN NGÀNH**

**NGÀNH: KHOA HỌC MÁY TÍNH**

ĐỀ TÀI

**CHẨN ĐOÁN BỆNH NGOÀI DA QUA ẢNH BẰNG MẠNG NƠ-RON TÍCH CHẬP**

**Giảng viên hướng dẫn: T.S. Nguyễn Mạnh Cường**

**Lớp:** 20234IT6052001

**Nhóm thực hiện:** Nhóm 01

**Thành viên:** Nguyễn Mạnh Niên – 2022601951

Phạm Thành Đạt – 2021605356

**Hà Nội, tháng 8 năm 2024**

**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP HÀ NỘI**

**KHOA CÔNG NGHỆ THÔNG TIN**

**A yellow and red sign with a red circle and a star

Description automatically generated**

**ĐỒ ÁN CHUYÊN NGÀNH**

**NGÀNH: KHOA HỌC MÁY TÍNH**

ĐỀ TÀI

**CHẨN ĐOÁN BỆNH NGOÀI DA QUA ẢNH BẰNG MẠNG NƠ-RON TÍCH CHẬP**

**Giảng viên hướng dẫn: T.S. Nguyễn Mạnh Cường**

**Lớp:** 20234IT6052001

**Nhóm thực hiện:** Nhóm 01

**Thành viên:** Nguyễn Mạnh Niên – 2022601951

Phạm Thành Đạt – 2021605356

**Hà Nội, tháng 8 năm 2024**

**MỤC LỤC**

[DANH MỤC HÌNH ẢNH 3](#_Toc174370969)

[DANH MỤC BẢNG BIỂU 4](#_Toc174370970)

[LỜI CẢM ƠN 5](#_Toc174370971)

[LỜI NÓI ĐẦU 6](#_Toc174370972)

[CHƯƠNG 1. BÀI TOÁN PHÂN TÍCH VÀ CHẨN ĐOÁN BỆNH NGOÀI DA QUA HÌNH ẢNH 8](#_Toc174370973)

[1.1. Giới thiệu chung 8](#_Toc174370974)

[1.2. Mục đích và ý nghĩa của đề tài 10](#_Toc174370975)

[1.2.1. Mục đích 10](#_Toc174370976)

[1.2.2. Ý nghĩa khoa học 10](#_Toc174370977)

[1.2.3. Ý nghĩa thực tiễn 10](#_Toc174370978)

[1.3. Dữ liệu đầu vào và đầu ra 10](#_Toc174370979)

[1.4. Phạm vi của bài toán 11](#_Toc174370980)

[1.5. Phương pháp tiếp cận 11](#_Toc174370981)

[1.6. Ý nghĩa của bài toán 12](#_Toc174370982)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 13](#_Toc174370983)

[2.1. Tổng quan về trí tuệ nhân tạo (Artificial Intelligence) 13](#_Toc174370984)

[2.1.1. Các lĩnh vực của trí tuệ nhân tạo 14](#_Toc174370985)

[2.1.2. Một số ứng dụng của trí tuệ nhân tạo 15](#_Toc174370986)

[2.2. Tổng quan về ảnh và một số kỹ thuật xử lý ảnh 16](#_Toc174370987)

[2.2.1. Tổng quan về ảnh 16](#_Toc174370988)

[2.2.2. Một số kỹ thuật xử lý ảnh 18](#_Toc174370989)

[2.3. Trích chọn và biểu diễn đặc trưng hình ảnh 21](#_Toc174370990)

[2.3.1. Đặc trưng ảnh 21](#_Toc174370991)

[2.3.2. Một số phương pháp trích chọn đặc trưng đơn giản 21](#_Toc174370992)

[2.4. Phương pháp phân đoạn ảnh 22](#_Toc174370993)

[2.5. Kỹ thuật học máy 24](#_Toc174370994)

[2.5.1. Tổng quan 24](#_Toc174370995)

[2.5.2. Máy phân loại vector hỗ trợ (Support Vector Machine) 26](#_Toc174370996)

[2.6. Mạng nơ-ron tích chập (CNN) trong chẩn đoán bệnh ngoài da qua hình ảnh 29](#_Toc174370997)

[2.6.1. Giới thiệu 29](#_Toc174370998)

[2.6.2. Cấu trúc 29](#_Toc174370999)

[2.6.3. Ưu điểm 33](#_Toc174371000)

[2.3.4. Các kiến trúc CNN phổ biến 33](#_Toc174371001)

[CHƯƠNG 3. KẾT QUẢ THỰC NGHIỆM 35](#_Toc174371002)

[3.1. Dữ liệu thực nghiệm 35](#_Toc174371003)

[TÀI LIỆU THAM KHẢO 36](#_Toc174371004)

DANH MỤC HÌNH ẢNH

[Hình 1. Minh họa về bài toán 7](#_Toc174107693)

[Hình 2.1. Mô hình mạng nơ-ron tích chập 10](#_Toc174107694)

[Hình 2.2. Mô tả cấu trúc mô hình CNN AlexNet 11](#_Toc174107695)

[Hình 2.3. Hình ảnh mô tả lớp tích chập 12](#_Toc174107696)

[Hình 2.4. Hình ảnh mô phỏng lớp kết nối đầy đủ 14](#_Toc174107697)

DANH MỤC BẢNG BIỂU

[Bảng 2. Bảng minh họa phương thức Max Pooling và Average Pooling 14](#_Toc174107802)

LỜI CẢM ƠN

Lời đầu tiên cho phép chúng em gửi lời cảm ơn sâu sắc tới các thầy cô trong khoa Công nghệ thông tin – Trường Đại học Công nghiệp Hà Nội, những người đã hết mình truyền đạt và chỉ dẫn cho chúng em những kiến thức, những bài học quý báu và bổ ích. Đặc biệt, chúng em xin được bày tỏ sự tri ân và xin chân thành cảm ơn giảng viên Nguyễn Mạnh Cường, người đã trực tiếp hướng dẫn, chỉ bảo chúng em trong suốt quá trình học tập, nghiên cứu và hoàn thành được đồ án.

Trải qua thời gian nghiên cứu và làm đề tài, nhóm chúng em đã khám phá được rất nhiều công nghệ mới, tuy nhiên, do năng lực, kiến thức và trình độ chuyên môn còn hạn hẹp nên không thể tránh khỏi một số thiết sót. Vì vậy, chúng em mong nhận được sự đóng góp ý kiến từ các thầy cô để giúp nhóm chúng em hoàn thiện kiến thức và bổ sung thêm thông tin cần thiết cho báo cáo của mình.

Một lần nữa, chúng em xin chân thành cảm ơn thầy và tất cả những người đã giúp đỡ và đồng hành cùng nhóm chúng em trong chặng đường này.

Em xin chân thành cảm ơn!

***Sinh viên thực hiện***

Nguyễn Mạnh Niên

Phạm Thành Đạt

LỜI NÓI ĐẦU

Trong bối cảnh công nghệ ngày càng phát triển mạnh mẽ, đặc biệt là sự tiến bộ vượt bậc của trí tuệ nhân tạo và học máy, việc ứng dụng các kỹ thuật này vào lĩnh vực y tế đã mở ra nhiều cơ hội mới trong việc chẩn đoán và điều trị bệnh. Trong đó, chẩn đoán bệnh ngoài da qua ảnh bằng mạng nơ-ron tích chập (CNN) là một hướng nghiên cứu đầy triển vọng.

Bệnh ngoài da là một trong những vấn đề y tế phổ biến, ảnh hưởng đến chất lượng cuộc sống của nhiều người. Việc chẩn đoán đúng và kịp thời là vô cùng quan trọng để đưa ra các biện pháp điều trị hiệu quả. Tuy nhiên, không phải lúc nào bệnh nhân cũng có thể tiếp cận với các chuyên gia da liễu để nhận được chẩn đoán chính xác. Để khắc phục vấn đề này, việc sử dụng công nghệ CNN trong chẩn đoán bệnh ngoài da qua ảnh đã trở thành một giải pháp tiềm năng.

Nhóm chúng em đã quyết định lựa chọn đề tài "Chẩn đoán bệnh ngoài da qua ảnh bằng mạng nơ-ron tích chập" với mục tiêu xây dựng một hệ thống chẩn đoán tự động, hỗ trợ các bác sĩ và bệnh nhân trong việc nhận diện và phân loại các loại bệnh ngoài da. Đề tài này không chỉ góp phần cải thiện chất lượng chăm sóc y tế mà còn mở ra cơ hội ứng dụng rộng rãi công nghệ AI trong y học.

Nội dung của báo cáo sẽ bao gồm các chương như sau:

* *Chương 1. Phát biểu bài toán:* Trong chương này, nhóm sẽ trình bày tổng quan về bệnh ngoài da, nhu cầu chẩn đoán, cũng như các khó khăn hiện tại trong quá trình chẩn đoán. Đồng thời, nhóm cũng xác định mục tiêu và phạm vi của bài toán.
* *Chương 2. Các kỹ thuật giải quyết bài toán:* Nhóm sẽ trình bày các kỹ thuật học máy và mạng nơ-ron tích chập hiện có, phân tích ưu nhược điểm của từng kỹ thuật và lựa chọn phương pháp phù hợp nhất để giải quyết bài toán chẩn đoán bệnh ngoài da qua ảnh.
* *Chương 3. Kết quả thực nghiệm:* Trong chương này, nhóm sẽ mô tả chi tiết quá trình thu thập dữ liệu, tiền xử lý dữ liệu và các bước thực hiện thực nghiệm. Kết quả đạt được từ các mô hình sẽ được trình bày và đánh giá.
* *Chương 4. Xây dựng hệ thống:* Nhóm sẽ sử dụng các công cụ và thư viện lập trình để xây dựng hệ thống chẩn đoán tự động, bao gồm cả phần giao diện người dùng và phần xử lý ảnh. Các kỹ thuật triển khai và vận hành hệ thống cũng sẽ được mô tả chi tiết.
* *Phần kết luận:* Nhóm sẽ tổng hợp lại những kết quả đạt được, đưa ra những nhận xét và đánh giá về hiệu quả của hệ thống, đồng thời đề xuất những hướng phát triển tiếp theo cho nghiên cứu.

Thông qua việc thực hiện đề tài này, nhóm chúng em không chỉ học hỏi được nhiều kiến thức mới về trí tuệ nhân tạo và ứng dụng trong y học mà còn mong muốn đóng góp một phần vào việc cải thiện chất lượng chăm sóc sức khỏe cộng đồng. Chúng em hy vọng rằng nghiên cứu này sẽ mang lại giá trị thực tiễn và mở ra nhiều cơ hội mới cho các ứng dụng công nghệ trong y tế.

CHƯƠNG 1. BÀI TOÁN PHÂN TÍCH VÀ CHẨN ĐOÁN BỆNH NGOÀI DA QUA HÌNH ẢNH

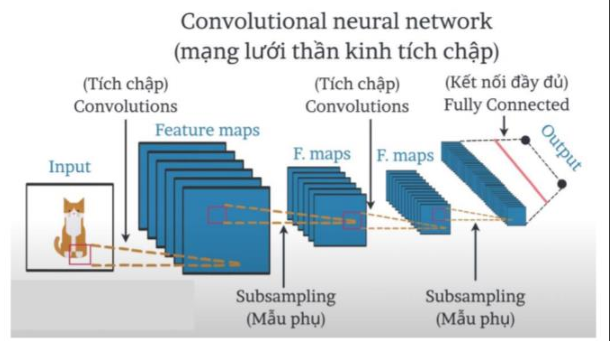
* 1. Giới thiệu chung

Bệnh ngoài da là một nhóm các bệnh lý phổ biến, ảnh hưởng đến mọi lứa tuổi và có thể gây ra nhiều vấn đề về sức khỏe và thẩm mỹ. Việc chẩn đoán chính xác bệnh ngoài da đóng vai trò quan trọng trong việc điều trị hiệu quả và ngăn ngừa biến chứng. Tuy nhiên, việc chẩn đoán dựa vào kinh nghiệm và quan sát của bác sĩ da liễu có thể gặp phải một số hạn chế như:

* + **Tính chủ quan:** Chẩn đoán dựa trên kinh nghiệm của bác sĩ có thể bị ảnh hưởng bởi kiến thức, kỹ năng và kinh nghiệm của từng người.
  + **Khó khăn trong việc phân biệt các bệnh lý có triệu chứng tương tự:** Nhiều bệnh ngoài da có triệu chứng rất giống nhau, khiến cho việc chẩn đoán trở nên khó khăn.
  + **Yêu cầu chuyên môn cao:** Việc chẩn đoán bệnh ngoài da đòi hỏi bác sĩ da liễu phải có chuyên môn và kinh nghiệm cao.

Cùng với sự phát triển nhanh chóng của khoa học công nghệ, các kỹ thuật dựa trên trí tuệ nhân tạo và xử lý ảnh ứng dụng trong các hệ thống thông minh đạt được những kết quả vượt bậc, có nhiều bước đột phá. Cuộc cách mạng công nghiệp lần thứ tư (Industry 4.0) hiện đang diễn ra trên phạm vi toàn cầu, đặc biệt ở các nước có nền khoa học kỹ thuật phát triển. Qua đó, các hệ thống thông minh dần thay thế con người. Nền tảng của các hệ thống thông minh có thể nói bắt nguồn từ lĩnh vực trí tuệ nhân tạo và xử lý ảnh. Trong đó, xử lý ảnh là một trong những giác quan máy quan trọng nhất giúp cho quá trình thu nhận tín hiệu, xử lý, phân tích nhằm đưa ra tri thức phục vụ các hệ thống ra quyết định. Xử lý ảnh là một chuyên ngành có nhiều ứng dựng trong các lĩnh vực khoa học, đời sống. Trong thiên văn học, xử lý ảnh giúp các nhà khoa học thu thập và phân tích hình ảnh vũ trụ; trong địa lý, người ta có thể dựa vào xử lý ảnh để lập chính xác các bản đồ địa hình, địa giới; nén ảnh rất cần thiết cho lĩnh vực thông tin và truyền thông; kỹ thuật nhận dạng hình ảnh được dùng nhiều trong các lĩnh vực liên quan đến kinh tế, quân sự. Đặc biệt, trong y học, xử lý ảnh hỗ trợ rất tốt cho việc chẩn đoán hình ảnh các bệnh về khối u, xương, mạch, ung thư…, tuy nhiên, trong quá trình thu nhận ảnh, ảnh thu được phần nhiều có chất lượng không như ý muốn. Đối với ảnh y học, do đặc trưng thường chụp các bộ phận bên trong cơ thể người bằng các thiết bị chuyên dụng như máy chụp X quang, máy chụp city, máy siêu âm, máy nội soi… nên chất lượng hình ảnh thường bị mờ, nhiễu, không sắc nét… gây khó khăn cho việc chẩn đoán bệnh.

Để khắc phục vấn đề trên, nhóm em sử dụng mạng nơ-ron tích chập (CNNs) để giải quyết. Mạng nơ-ron tích chập thể hiện tính ưu việt hơn các kỹ thuật khác trong việc nhận diện hình ảnh cũng như nhiều lĩnh vực khác, những kết quả nghiên cứu gần đây áp dụng kỹ thuật này cho độ chính xác cao hơn. Trong lĩnh vực y tế có nhiều công ty khởi nghiệp (Enlitic, Freenome, Merck, Atomwise…) về y sinh đã đưa ra những sản phẩm điện toán có khả năng đọc phim X-quang, cộng hưởng từ (MRI) và phim chụp cắt lớp vi tính một cách nhanh chóng và chính xác bệnh chuẩn hơn cả bác sỹ điều trị.



Hình 1.1. Minh họa về bài toán

* 1. Mục đích và ý nghĩa của đề tài
     1. Mục đích

Mục tiêu chính của đề tài là nghiên cứu các kỹ thuật trí tuệ nhân tạo, đặc biệt là kỹ thuật học sâu và các ứng dụng của nó trong thực tế. Nghiên cứu hình ảnh về da, tập trung vào các điểm ảnh bất thường dựa trên các kỹ thuật học sâu mạng tích chập CNNs trong lĩnh vực thị giác máy tính.

Thực nghiệm áp dụng kỹ thuật mạng nơ-ron tích chập CNNs trong trích xuất đặc trưng hình ảnh để dự đoán bệnh ngoài da. Từ đó phân tích, đánh giá hiệu quả và độ chính xác trên các thư viện chuẩn có sẵn.

* + 1. Ý nghĩa khoa học
* Tìm hiểu, mô tả về mô hình hệ thống chẩn đoán bệnh qua ảnh chụp ngoài da.
* So sánh giữa các kỹ thuật được sử dụng trong hệ thống nhận dạng.
* Đánh giá hiệu suất của phương pháp bằng các thực nghiệm trên tập dữ liệu thực.
* Vận dụng trí tuệ nhân tạo trong việc giải quyết bài toán chẩn đoán bệnh qua ảnh chụo ngoài da với khả năng xử lý được nhiều loại bệnh.
  + 1. Ý nghĩa thực tiễn

Thiết kế được mô hình chẩn đoán bệnh qua ảnh chụp ngoài da để xây dựng các hệ thống ứng dụng hỗ trợ bác sỹ trong việc chẩn đoán một số bệnh về da.

* 1. Dữ liệu đầu vào và đầu ra
     1. Dữ liệu đầu vào

Dữ liệu đầu vào là hình ảnh tổn thương da liễu ở định dạng JPEG.

Tất cả hỉnh ảnh tổn thương được đặt tên bằng lược đồ ISIC\_.jpg, trong đó là một định danh duy nhất gồm 7 chữ số. Thẻ EXIF trong ảnh đã bị xóa; mọi thẻ EXIF còn lại sẽ không được dựa vào để cung cấp siêu dữ liệu chính xác.

* + 1. Dữ liệu đầu ra

Dữ liệu đáp ứng là tập hợp các phân loại nhị phân cho mỗi trong số 7 trạng thái bệnh, cho biết chẩn đoán của từng hình ảnh tổn thương đầu vào.

* 1. Phạm vi của bài toán

Bài toán tập trung vào việc nghiên cứu các kỹ thuật trí tuệ nhân tạo, kỹ thuật mạng nơ-ron học sâu để ứng dụng trong phân tích hình ảnh chụp ngoài da và chẩn đoán một số loại bệnh ngoài da phổ biến như:

* + *Mụn trứng cá:* Bệnh lý viêm da mãn tính thường gặp ở tuổi dậy thì.
  + *Vảy nến:* Bệnh lý da mãn tính, gây ra các mảng da đỏ, dày sừng.
  + *Chàm:* Viêm da dị ứng, gây ngứa, nổi mần đỏ.
  + *Nấm da:* Nhiễm trùng da do nấm gây ra, thường gặp ở vùng da ẩm ướt.
  + *Ung thư da:* Khối u ác tính phát triển từ các tế bào da, có thể gây nguy hiểm đến tính mạng.
  1. Phương pháp tiếp cận

Bài toán sẽ được giải quyết bằng cách sử dụng mạng nơ-ron tích chập (CNN) để xây dựng mô hình chẩn đoán. CNN là một loại mạng nơ-ron nhân tạo (ANN) đặc biệt hiệu quả trong việc xử lý hình ảnh. CNN học hỏi các đặc trưng từ hình ảnh thông qua các lớp tích chập và sau đó sử dụng các lớp kết nối đầy đủ để phân loại hình ảnh.

Các bước cụ thể để giải quyết bài toán:

1. *Thu thập dữ liệu:* Xây dựng một bộ dữ liệu ảnh chụp tổn thương ngoài da được gắn nhãn với các loại bệnh tương ứng.
2. *Tiền xử lý dữ liệu:* Chuẩn hóa kích thước, định dạng ảnh và loại bỏ nhiễu.
3. *Xây dựng mô hình CNN:* Chọn kiến trúc CNN phù hợp, huấn luyện mô hình với dữ liệu đã được tiền xử lý.
4. *Đánh giá mô hình:* Đánh giá hiệu suất của mô hình trên tập dữ liệu kiểm tra.
5. *Phân tích và cải thiện mô hình:* Phân tích các yếu tố ảnh hưởng đến hiệu suất của mô hình và cải thiện mô hình bằng cách điều chỉnh kiến trúc, tham số hoặc sử dụng các kỹ thuật nâng cao khác.
   1. Ý nghĩa của bài toán

Bài toán này có ý nghĩa thực tiên cao, mang lại nhiều lợi ích cho ngành y tế và cộng đồng:

* + *Hỗ trợ chẩn đoán sớm bệnh ngoài da:* Giúp phát hiện sớm các loại bệnh ngoài da, tăng khả năng điều trị thành công và giảm nguy cơ biến chứng.
  + *Giảm tải cho hệ thống y tế:* Giảm gánh nặng chẩn đoán cho bác sĩ da liễu, đặc biệt là ở những vùng sâu, vùng xa.
  + *Nâng cao chất lượng chăm sóc sức khỏe:* Cung cấp một công cụ chẩn đoán khách quan, chính xác và hiệu quả cho bệnh nhân.

CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

2.1. Tổng quan về trí tuệ nhân tạo (Artificial Intelligence)

Vào năm 1943, Warren McCulioch và Walter Pitts bắt đầu thực hiện nghiên cứu ba cơ sở lý thuyết cơ bản: Triết học cơ bản và chức năng của các neural thần kinh; phân tích các mệnh đề logic; lý thuyết dự đoán Turing. Các tác giả đã nghiên cứu đề xuất mô hình neural nhân tạo, mỗi neural đặc trưng bởi hai trạng thát “bật”, “tắt” và phát hiện mạng neural có khả năng học.

Trí tuệ nhân tạo (AI) được thiết lập bởi John McCarthy tại hội thảo đầu tiên về chủ đề này vào mùa hè năm 1956[1]. Đồng thời, ông cũng đề xuất ngôn ngữ lập trình Lisp, một trong những ngôn ngữ lập trình hàm tiêu biểu, được sử dụng trong lĩnh vực AI. Sau đó, Alan Turing đưa ra “Turing test” như một phương pháp kiểm chứng hành vi thông minh.

Marvin Minsky và Seymour Papert đưa ra các chứng minh đầu tiên về giới hạn của các mạng neural đơn giản. Ngôn ngữ lập trình logic Prolog ra đời và được phát triển bởi Alan Colmerauer. Ted Shortliffe xây dựng thành công một số hệ chuyên gia đầu tiên trợ giúp chẩn đoán y học, các hệ thống này sử dụng ngôn ngữ luật để biểu diễn tri thức và suy diễn.

Vào đầu năm 1980, những nghiên cứu thành công liên quan đến AI như các hệ chuyên gia (expert systems), một dạng của chương trình AI mô phỏng tri thức và các kỹ năng phân tích của một hoặc nhiều chuyên gia con người. AI được áp dụng trong logic, khai phá dữ liệu, chẩn đoán y học và nhiều lĩnh vực ứng dụng khác trong công nghiệp. Sự thành công dựa vào nhiều yếu tố: Tăng khả năng tính toán của máy tính, tập trung giải quyết các bài toán con cụ thể, xây dựng các mối quan hệ giữa AI và các lĩnh vực khác giải quyết các bài toán tương tự và một sự chuyển giao mới của các nhà nghiên cứu cho các phương pháp trong toán học vững chắc và chẩn đoán khoa học chính xác.

2.1.1. Các lĩnh vực của trí tuệ nhân tạo

* *Lập luận, suy diễn tự động:* Khái niệm lập luận và suy diễn được sử dụng rất phổ biến trong lĩnh vực AI. Lập luận là suy diễn logic, dùng để chỉ một tiến trình rút ra kết luận (tri thức mới) từ những giả thiết đã cho (được biểu diễn dưới dạng cơ sở tri thức). Như vậy, để thực hiện lập luận người ta cần có các phương pháp lưu trữ cơ sở tri thức và các thủ tục lập luận trên cơ sở tri thức đó.
* *Biểu diễn tri thức:* Muốn máy tính có thể lưu trữ và xử lý tri thức thì cần có các phương pháp biểu diễn tri thức. Các phương pháp biểu diễn tri thức ở đây bao gồm các ngôn ngữ biểu diễn và các kỹ thuật xử lý tri thức. Một ngôn ngữ biểu diễn tri thức được đánh giá là “tốt” nếu nó có tính biểu đạt cao và tính hiệu quả của thuật toán lập luận trên ngôn ngữ đó. Tính biểu đạt của ngôn ngữ thể hiện khả năng biểu diễn một phạm vi rộng lớn các thông tin trong một miền ứng dụng. Tính hiệu quả của các thuật toán lập luận thể hiện chi phí về thời gian và không gian dành cho việc lập luận.
* *Lập kế hoạch:* Khả năng suy ra các mục đích cần đạt được đối với các nhiệm vụ đưa ra và xác định dãy các hành động cần thực hiện để đạt được mục đích đó.
* *Xử lý ngôn ngữ tự nhiên:* Là một nhánh của AI, tập trung vào các ứng dụng trên ngôn ngữ của con người. Các ứng dụng trong nhận dạng tiếng nói, nhận dạng chữ viết, dịch tự động, tìm kiếm thông tin…
* *Hệ chuyên gia:* Cung cấp các hệ thống có khả năng suy luận để đưa ra những kết luận. Các hệ chuyên gia có khả năng xử lý lượng thông tin lớn và cung cấp các kết luận dựa trên những thông tin đó. Có rất nhiều hệ chuyên gia nổi tiếng như các hệ chuyên gia y học MYCIN, đoán nhận cấu trúc phân tử từ công thức hóa học DENDRAL…

2.1.2. Một số ứng dụng của trí tuệ nhân tạo

Ngày nay, AI ngày càng được ứng dụng nhiều trong các lĩnh vực khác nhau, từ việc phục vụ đời sống hàng ngày của con người cho đến giáo dục, tài chính ngân hàng, y học, rô-bôt, ôtô tự hành… và thực tế chứng minh rằng việc ứng dụng AI giúp nâng cao hiệu suất lao động, cải thiện chất lượng cuộc sống của con người, phát triển kinh doanh cho các doanh nghiệp và nó cũng sẽ là nền tảng của rất nhiều các ứng dụng và dịch vụ mới khác trong tương lai.

Nhờ sự phát triển khoa học vật lý lượng tử giúp cho việc tính toán và xử lý song song của các hệ thống nhanh hơn đáng kể, vì thế việc áp dụng các phương pháp học máy vào xử lý các bài toán thực tế ngày một thuận lợi hơn. Trong đó, kỹ thuật học sâu (deep learning) đã được quan tâm và phát triển mạnh giúp cho máy tính giải quyết các bài toán trong lĩnh vực học máy ngày càng tốt hơn, mà cụ thể đó là các bài toán tương tác người - máy trong lĩnh vực thị giác máy tính, nhận thức sự vật, gợi ý trong các hệ thống lớn, chẩn đoán các bệnh hiếm gặp...

Một số ứng dụng phổ biến hiện nay như[2]:

* *Lĩnh vực giáo dục:* Mô hình trường học thông minh hay học trực tuyến ngày một phát triển nhờ ứng dụng AI vào quá trình tương tác học tập giữa nhà trường và học sinh, sinh viên tạo ra một hệ sinh thái về giáo dục mà ở đó nhà trường và học sinh tương tác với nhau một cách thuận lợi và nhanh chóng thông qua hệ sinh thái này.
* *Lĩnh vực y tế:* Chẩn đoán, điều trị và theo dõi bệnh giúp nâng cao chăm sóc sức khỏe con người, giảm chi phí chữa bệnh cho người dân, trong đó kể cả các dự án như điều trị ưng thư bằng AI, Rô-bôt chăm sóc y tế…
* *Công nghiệp:* Nhận dạng hình ảnh, giọng nói, rô-bôt thông minh, xe tự hành, hệ thống tương tác thực ảo… đã phát triển mạnh mẽ nhờ sự phát triển của mạng Neural học sâu, hệ thống xử lý phân tán song song (Parallel Distributed processing).

2.2. Tổng quan về ảnh và một số kỹ thuật xử lý ảnh

2.2.1. Tổng quan về ảnh

* **Một số khái niệm cơ bản**

- *Ảnh số:* Ảnh số là tập hợp hữu hạn các điểm ảnh với mức xám phù hợp dùng để mô tả ảnh gần với ảnh thật. Ảnh số bao gồm một tập hữu hạn các phần tử được biểu diễn bởi giá trị số. Ảnh số có thể được biểu diễn dưới dạng ma trận hai chiều, mỗi phần tử của ảnh số gọi là điểm ảnh (pixel). Số điểm ảnh xác định độ phân giải của ảnh. Ảnh có độ phân giải càng cao thì càng thể hiện rõ nét các đặt điểm của tấm ảnh và càng làm cho tấm ảnh trở nên thực và sắc nét hơn.

A black and white image of a person

Description automatically generated

Hình 2.1. Ảnh ký tự chữ A và ma trận số của vùng chọn

*- Điểm ảnh:* Điểm ảnh (Pixel) là một phần tử của ảnh số tại toạ độ (x, y) với độ xám hoặc màu nhất định. Kích thước và khoảng cách giữa các điểm ảnh đó được chọn thích hợp sao cho mắt người cảm nhận sự liên tục về không gian và mức xám (hoặc màu) của ảnh số gần như ảnh thật. Mỗi phần tử trong ma trận được gọi là một phần tử ảnh. Trong hình 1.1 mỗi điểm ảnh là một ô mang một giá trị số biểu thị mức xám từ 0 đến 255.

- *Mức xám của ảnh:* Là kết quả của sự biến đổi tương ứng 1 giá trị độ sáng của 1 điểm ảnh với một giá trị nguyên dương. Thông thường nó xác định trong [0, 255] tuỳ thuộc vào giá trị mà mỗi điểm ảnh được biểu diễn. Các thang giá trị mức xám thông thường: 2, 16, 32, 64, 128. Ảnh đa mức xám thường dùng là 256, như vậy mức xám thường xác định trong khoảng [0, 255] tùy thuộc vào giá trị mà mỗi điểm ảnh được biểu diễn.

A table with numbers and a few squares

Description automatically generated with medium confidence

Hình 2.2. Ma trận số biếu thị mức xám của các điểm ảnh

- *Độ phân giải ảnh:* Độ phân giải (Resolution) của ảnh là mật độ điểm ảnh được ấn định trên một ảnh số được hiển thị. Theo định nghĩa, khoảng cách giữa các điểm anh phải được chọn sao cho mắt người vẫn thấy được sự liên tục của ảnh. Việc lựa chọn khoảng cách thích hợp tạo nên một mật độ phân bổ, đó chính là độ phân giải và được phân bố theo trục x và y trong không gian hai chiều.

* **Một số dạng ảnh**

- *Ảnh màu:* thường là các ảnh chứa thông tin về đối tượng được biểu diễn dưới dạng màu sắc mà mắt thường có thể quan sát được. Mỗi điểm ảnh có cấu trúc gồm nhiều kênh khác nhau, thông thường trong máy tính, nó biểu diễn 3 lớp màu cơ bản RGB (Red, Green, Blue)[2].

- *Ảnh đa mức xám:* thường biểu diễn thông tin liên quan đến cường độ đa xám của đối tượng trong không gian mà không được thể hiện bởi màu sắc thực của nó[2].

- *Ảnh nhị phân:* Giá trị xám của tất cả các điểm ảnh chỉ nhận giá trị 1 hoặc 0 như vậy mỗi điểm ảnh trong ảnh nhị phân được biểu diễn bởi 1 bit. Ảnh nhị phân thường được dùng để biểu diễn, phân biệt sự xuất hiện đối tượng và nền trong mỗi bức ảnh.

2.2.2. Một số kỹ thuật xử lý ảnh

* **Lược đồ ảnh (Histogram)**

Lược đồ ảnh là một đồ thị biểu diễn tần số xuất hiện của cường độ sáng điểm ảnh theo các mức hay nói cách khác lược đồ Histogram của một hình ảnh biểu diễn lược đồ về sự phân bố các mức cường độ xám của một bức ảnh[4].

A comparison of a person's face

Description automatically generated

Hình 2.3. Lược đồ ảnh đa mức xám kích thước 64x6 pixels

Lược đồ ảnh thể hiện mức cường độ sáng theo chiều ngang ảnh là kết quả của việc tính tổng giá trị của các điểm ảnh theo từng cột của ảnh. Ví dụ: cho ảnh xám đầu vào có kích thước 600x200 pixels, khi đó lược đồ sẽ có 600 cột và độ cao mỗi cột là tổng giá trị (từ 0 đến 255) của 200 điểm ảnh tương ứng với 200 dòng.

A black and white image of a line

Description automatically generated

Hình 2.4. Lược đồ cường độ sáng theo chiều ngang ảnh

* **Nhị phân hóa ảnh**

Để chuyển sang ảnh nhị phân, trước tiên cần chuyển ảnh về dạng ảnh xám với các mức xám có giá trị từ 0 đến 255 dựa trên ba giá trị red, green, blue theo công thức:

Greycolor = r \* 0.299 + g \* 0.587 + b \* 0.114

Sử dụng ngưỡng thích hợp để chuyển các điểm ảnh về dạng 0 (màu đen) hoặc 255 (màu trắng) theo nguyên tắc:

g(x, y) = 1 nếu f(x, y) > T và g(x, y) = 0 nếu f(x, y) T

Trong đó: g(x, y) là giá trị kết quả tại điểm ảnh có toạn độ (x, y);

f(x, y) là giá trị điểm ảnh (x, y) của ảnh cần xử lý;

T là ngưỡng

* **Làm mịn ảnh**

Mịn ảnh được thực hiện dựa trên bộ lọc trơn (Smoothing filter) nhằm loại nhiễu, bước này dùng trong quá trình tiền xử lý (Pre-processing) khi phải giảm bớt một số chi tiết không cần thiết của một đối tượng nào đó trong ảnh. Một hướng áp dụng phổ biến để giảm nhiễu là lọc tuyến tính, những bộ lọc tuyến tính theo hướng này được biết đến như là lọc thông thấp.

Ý tưởng cho những bộ lọc thông thấp là thay thế giá trị mức sáng của mọi điểm ảnh bằng giá trị mức sáng trung bình của các hàng xóm, định nghĩa theo mặt nạ lọc. Kết quả trên dẫn tới ảnh số văn bản mất đi những chi tiết nhiễu, ma trận của một bộ lọc làm mịn ảnh thường sử dụng có các hệ số như sau:

hoặc

* **Phép co ảnh – Erosion**

Xét tập hợp A và tập hợp B (phần tử cấu trúc), phép co ảnh nhị phân của tập hợp A bởi phẩn tử cấu trúc B được ký hiệu là A ⊝ B và viết dưới dạng công thức như sau[2]:

A ⊝ B = {c | (B)C ⊆ A}

Trong đó: A: Ma trận điểm ảnh của ảnh nhị phân

B: Phần tử cấu trúc

A number and number equation

Description automatically generated with low confidence

Hình 2.5. Minh họa phép co ảnh trên ảnh nhị phân

A black and white image of a person

Description automatically generated

Hình 2.6. Ảnh nhị phân ký tự “A” trước và sau khi co ảnh

2.3. Trích chọn và biểu diễn đặc trưng hình ảnh

2.3.1. Đặc trưng ảnh

* **Đặc trưng màu sắc**

Màu sắc là một đặc trưng nổi bật và được sử dụng phổ biến nhất trong tìm kiếm ảnh theo nội dung. Mỗi một điểm ảnh (thông tin màu sắc) có thể được biểu diễn như một điểm trong không gian màu sắc ba chiều. Các không gian màu sắc thường dùng là: RGB, Munsell, CIE, HSV.

* **Đặc trưng kết cấu**

Kết cấu hay còn gọi là vân (texture) là một đối tượng dùng để phân hoạch ảnh ra thành những vùng được quan tâm và để phân lớp những vùng đó. Vân cung cấp thông tin sự sắp xếp về mặt không gian của màu sắc và cường độ của một ảnh.

* **Đặc trưng hình dạng**

Hình dạng của một ảnh hay một vùng là một đặc trưng quan trọng trong việc xác định và phân biệt ảnh trong nhận dạng mẫu. Mục tiêu chính của biểu diễn hình dạng trong nhận dạng mẫu là đo thuộc tính hình học của một đối tượng được dùng trong phân lớp, so sánh và nhận dạng đối tượng.

2.3.2. Một số phương pháp trích chọn đặc trưng đơn giản

* **Trích chọn đặc trưng ảnh dựa trên màu sắc**

- *Lược đồ màu (Histogram):* Là đại lượng đặc trưng cho phân bố màu cục bộ của ảnh. Độ đo tính tương tự về màu sắc được tính bằng phần giao của 2 lược đồ màu ảnh truy vấn H(IQ) và ảnh trong cơ sở dữ liệu ảnh H(ID). Kết quả sẽ là một lược đồ màu thể hiện độ giống nhau giữa 2 ảnh trên. Tuy nhiên vì lược đồ màu chỉ thể hiện tính phân bố màu toàn cục của ảnh mà không xét đến tính phân bố cục bộ của điểm ảnh nên có thể có 2 ảnh xem rất khác nhau nhưng lại có cùng lược đồ màu.

- *Vector liên kết màu (Color Coherence Vector):* Là lược đồ tinh chế lược đồ màu, chia mỗi ô màu thành 2 nhóm điểm ảnh: nhóm liên kết màu (coherence pixels) và nhóm không liên kết màu (noncoherence pixels). Vector liên kết màu còn giúp giải quyết khuyết điểm về tính không duy nhất của lược đồ màu đối với ảnh. Hai ảnh có thể có chung lược đồ màu nhưng khác nhau hoàn toàn, đây là khuyết điểm của lược đồ màu. Nhưng với tìm kiếm theo đặc trưng vector liên kết màu thì nó sẽ giải quyết được khuyết điểm không duy nhất này.

* **Trích chọn đặc trưng ảnh dựa trên kết cấu**

Kết cấu hay còn gọi là vân (texture) là một đối tượng dùng để phân hoạch ảnh ra thành những vùng được quan tâm và để phân lớp những vùng đó. Vân cung cấp những thông tin sự sắp xếp về mặt không gian của màu sắc và cường độ của một ảnh.

2.4. Phương pháp phân đoạn ảnh

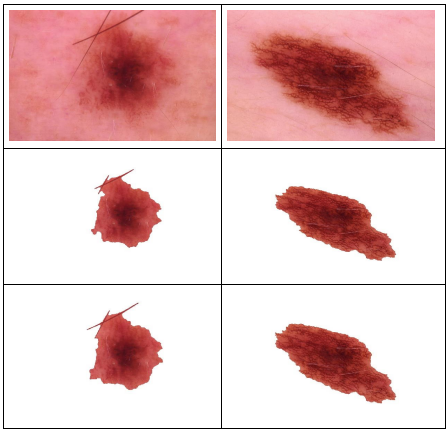
Trong lĩnh vực xử lý ảnh, phân đoạn ảnh (segmentation) được hiểu là quá trình nhóm điểm ảnh thành các vùng sao cho các điểm ảnh trong cùng một vùng có chung đặc điểm, tính chất nào đó, ví dụ như sự tương tự về màu sắc, cường độ hoặc kết cấu. Nói cách khác, phân đoạn ảnh là một quá trình chia một ảnh số thành nhiều vùng khác nhau, mỗi vùng gồm tập hợp các điểm ảnh (có thể gọi là superpixels hoặc region) có cùng tính chất. Trong một số tài liệu, phân đoạn hình ảnh còn được gọi là phân vùng ảnh. Mục tiêu của việc phân đoạn là để đơn giản hóa hoặc thay đổi cách biểu diễn một ảnh theo tiêu chuẩn nào đó nhằm làm cho nó ý nghĩa hơn và dễ dàng trong phân tích, trích xuất đặc trưng [3]. Kết quả của quá trình phân đoạn ảnh là tập hợp các vùng (region) hay còn gọi là các phân đoạn (segments) có thể bao gồm toàn bộ ảnh hoặc tập hợp các vùng đối tượng được trích xuất ra từ ảnh.

Phân đoạn ảnh có ý nghĩa quan trọng trong việc phát hiện các đặc tính và cấu trúc đối tượng trong ảnh vì mỗi vùng thường miêu tả cấu trúc tương ứng với đối tượng trong vùng đó. Ví dụ số lượng các vùng trong ảnh thể hiện rằng trong ảnh đó có chứa nhiều đối tượng khác nhau và ngược lại mỗi đối tượng chứa các vùng trong nó tương ứng với các phần khác nhau của cùng đối tượng. Phân đoạn ảnh có nhiều ứng dụng quan trọng trong nhiều lĩnh vực khác nhau như phân tích hình ảnh y học (medical imaging), phát hiện và nhận dạng đối tượng, phát hiện hành động trong các hệ thống camera giám sát thông minh, hệ thống điều khiển và giao thông thông minh, xe không người lái, xử lý hình ảnh vệ tinh…

Phân đoạn ảnh được xem là quá trình tiền xử lý, có ý nghĩa quan trọng trong hệ thống xử lý ảnh. Phân đoạn tốt sẽ giúp cho quá trình xử lý ở các bước tiếp theo đạt hiệu quả cao hơn, nâng cao độ chính xác và cũng như tiết kiệm chi phí tính toán.

Các vùng phân tách được dùng trong nhiều ngữ cảnh khác nhau và đặc biệt có thể được sử dụng làm đại diện thay thế cho đối tượng và biểu diễn dưới dạng nhiều hình thức khác nhau [4]. Việc lựa chọn mô tả vùng dữ liệu có nhiều cách thức khác nhau tùy thuộc vào tính chất, mục tiêu và đảm bảo phù hợp cho từng ứng dụng cụ thể. Nhiều ứng dụng không chỉ đòi hỏi mô tả các vùng riêng lẻ mà còn cần thể hiện được mối quan hệ với nhau giữa các vùng trong ảnh.

Ngưỡng (threshold) là một khái niệm phổ biến trong lĩnh vực phân đoạn đối tượng của kỹ thuật xử lý ảnh. Ngưỡng được dùng để chỉ một giá trị nào đó mà ta dùng để phân tách một tập hợp các điểm ảnh thành các vùng phân biệt. Ví dụ sử dụng một ngưỡng th để phân đoạn ảnh thành các vùng có cường độ sáng lớn hơn th và các vùng còn lại có giá trị cường độ sáng nhỏ hơn hoặc bằng th. Giá trị ngưỡng thường được xác định dựa vào giá trị đặc biệt như giá trị trung bình, giá trị trung vị hoặc dựa vào khảo sát thống kê theo kinh nghiệm. Có nhiều cách phân loại khác nhau, tuy nhiên nếu căn cứ vào số lượng ngưỡng dùng cho phân đoạn thì có phương pháp phân đoạn sử dụng ngưỡng đơn, phân đoạn ngưỡng kép hoặc phân đoạn đa ngưỡng. Nếu căn cứu vào tính chất biến thiên của giá trị ngưỡng thì có thể phân thành phương pháp dùng ngưỡng cố định (constant threshold) và ngưỡng động hay ngưỡng thích ứng (adaptive threshold). Trong đó, ngưỡng động được hiểu theo nghĩa giá trị của ngưỡng sẽ thay đổi tùy thuộc vào sự biến thiên của tập dữ liệu ảnh theo không gian và thời gian. Thông thường giá trị ngưỡng được xác định bằng cách áp dụng phương pháp phân tích thống kê.



Hình 2.7. Sử dụng các phương pháp phân đoạn khác nhau trong xử lý ảnh y tế

2.5. Kỹ thuật học máy

2.5.1. Tổng quan

Học máy (Machine Learning) là một ngành khoa học nghiên cứu các thuật toán cho phép máy tính có thể học được các khái niệm (concept). Có hai loại phương pháp học máy chính:

*- Phương pháp quy nạp:* máy học hoặc phân biệt các khái niệm dựa trên dữ liệu đã thu thập được trước đó. Phương pháp này cho phép tận dụng được nguồn dữ liệu rất nhiều và sẵn có.

*- Phương pháp suy diễn:* máy học hoặc suy diễn dựa vào các luật. Phương pháp này cho phép tận dụng được các kiến thức chuyên ngành để hỗ trợ máy tính.

Machine Learning được chua thành 3 nhánh chính: supervised learning (học có giám sát), unsupervised learning (học không có giám sát), và reinforcement learning (học tăng cường).

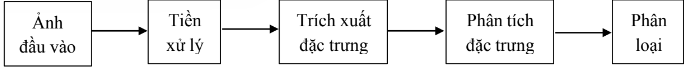
- Học có giám sát được dùng trong trường hợp một thuộc tính (nhãn) có sẵn cho một tập dữ liệu nhất định (tập huấn luyện), nhưng thiếu và cần được dự đoán cho các trường hợp khác.

- Học không có giám sát thì ngược lại, nó được sử dụng trong trường hợp khám phá các mối quan hệ tiềm ẩn trong một tập dữ liệu không được gán nhãn (các mục không được chỉ định trước).

- Học tăng cường thì nằm giữa 2 loại trên, có một số hình thức phản hồi sẵn cho môi bước tiên đoán hoặc hành động, nhưng không có nhãn chính xác hoặc thông báo lỗi [5].

Trong lĩnh vực xử lý ảnh, bài toán phân loại mẫu và nhận dạng đối tượng được hiểu là việc phân loại, xác định các mẫu thuộc lớp nào đó một cách chính xác dựa vào việc đo lường về đối tượng dựa trên các đặc trưng của đối tượng. Nói một cách khác, nhận dạng đối tượng nhằm mục đích phân loại các mẫu dựa trên kiến thức có trước về đối tượng hoặc dựa vào thông tin thống kê, đo lường được trích rút ra từ các mẫu trong tập dữ liệu huấn luyện. Ngày nay, nhận dạng được ứng dụng rộng rãi trong nhiều hệ thống như nhận dạng người trong các hệ thống tương tác người mày; nhận dạng khuôn mặt và định danh cá thể, nhận dạng phương tiện giao thông trong các hệ thống giao thông thông minh, hỗ trợ lái xe; nhận dạng hành động trong các hệ thống giám sát thông minh; nhận dạng chữ viết... Lĩnh vực nhận dạng đối tượng liên quan đến các phương pháp, kỹ thuật xử lý ảnh và học máy. Để huấn luyện mô hình và nhận dạng đối tượng, các mẫu dữ liệu thường được số hóa và biểu diễn thành các vector đặc trưng trong một không gian đa chiều tương ứng, được gọi là không gian đặc trưng. Nhờ vào các vector đặc trưng này để có thể phân tích, đánh giá nhằm xây dựng được mô hình đối tượng phục vụ phân loại, nhận dạng chúng. Trong nhận dạng, các phương pháp học máy có giám sát và không giám sát đều được nghiên cứu và ứng dụng. Các kỹ thuật học máy có giám sát thường được sử dụng như cây quyết định, mạng neural, SVM, Boosting, rừng ngẫu nhiên (random forest). Các kỹ thuật học máy không giám sát như gom cụm (Clustering Algorithms), phân tích thành phần chính (Principal Component Analysis - PCA), phân tích thành phần độc lập (Independent Component Analysis).

Quá trình nhận dạng đối tượng được mô tả bằng sơ đồ tổng quát như sau:



Hình 2.8. Quá trình nhận dạng đối tượng

2.5.2. Máy phân loại vector hỗ trợ (Support Vector Machine)

Support Vector Machine (SVM) là một mô hình phân loại hoạt động bằng việc xây dựng một siêu phẳng (hyperplane) có (n - 1) chiều trong không gian n chiều của dữ liệu sao cho siêu phẳng này phân loại các lớp một cách tối ưu nhất. Nói cách khác, cho một tập dữ liệu có nhãn (học có giám sát), thuật toán sẽ dựa trên dữ liệu học để xây dựng một siêu phẳng tối ưu được sử dụng để phân loại dữ liệu mới. Ở không gian 2 chiều thì siêu phẳng này là 1 đường thẳng phân cách chia mặt phẳng không gian thành hai phần tương ứng hai lớp với mỗi lớp nằm ở một phía của đường thẳng.

Ví dụ, ta có các điểm dữ liệu như hình 2.9 với mỗi điểm thuộc 1 trong 2 lớp cho trước:

A couple of red objects

Description automatically generated with medium confidence

Hình 2.9. Đường thẳng phân chia 2 lớp trong không gian 2 chiều

Bản chất của phương pháp SVM là chuyển không gian dữ liệu ban đầu thành một không gian mới hữu hạn chiều mà ở đó cho khả năng phân lớp dễ dàng hơn. Một quả bất kì nằm trên mặt bàn sẽ được gắn với một tọa độ cụ thể. Ví dụ, quả táo nằm cách mép trái 2cm và cách mép dưới 5cm được thể hiện trên trục tọa độ (x, y) tương ứng là (2, 5), x và y chính là tọa độ trong không gian hai chiều của quả táo. Khi đưa lên chiều thứ 3 là z(x, y), ta có thể tính được tọa độ của z trong không gian 3 chiều dựa vào tọa độ x,y ban đầu. Điểm làm SVM hiệu quả hơn các phương pháp khác chính là việc sử dụng Kernel Method giúp cho SVM không còn bị giới hạn bởi việc phân lớp một cách tuyến tính, hay nói cách khác các siêu phẳng có thể được hình thành từ các hàm phi tuyến.

Margin là khoảng cách giữa siêu phẳng đến 2 điểm dữ liệu gần nhất tương ứng với các phân lớp. Trong ví dụ quả táo quả lê đặt trên mặt bàn, margin chính là khoảng cách giữa cây que và hai quả táo và lê gần nó nhất. Điều quan trọng ở đây đó là phương pháp SVM luôn cô gắng cực đại hóa margin này, từ đó thu được một siêu phẳng tạo khoảng cách xa nhất so với 2 quả táo và lê. Nhờ vậy, SVM có thể giảm thiểu việc phân lớp sai (misclassification) đối với điểm dữ liệu mới đưa vào.

A diagram of a graph

Description automatically generated

Hình 2.10. Margin trong SVM

Là một kỹ thuật phân lớp khá phổ biến, SVM thể hiện được nhiều ưu điểm trong số đó có việc tính toán hiệu quả trên các tập dữ liệu lớn. Có thể kể thêm một số ưu điểm của phương pháp này như:

- Xử lý trên không gian 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 nơi chiều có thể cực kỳ 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 ra quyết dịnh.

- Tính linh hoạt - phân lớp thường là phi tuyến tính. Khả năng áp dụng Kernel 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.

Tuy nhiên SVM cho kết quả không tốt khi số lượng thuộc tính (p) của tập dữ liệu lớn hơn rất nhiều so với số lượng dữ liệu (n). Đồng 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 thành viên trong một nhóm là như thế nào.

SVM là một phương pháp hiệu quả cho bài toán phân lớp dữ liệu. Nó là một công cụ đắc lực cho các bài toán về xử lý ảnh, phân loại văn bản, phân tích quan điểm. Một yếu tố làm nên hiệu quả của SVM đó là việc sử dụng Kernel function khiến cho các phương pháp chuyển không gian trở nên linh hoạt hơn.

2.6. Mạng nơ-ron tích chập (CNN) trong chẩn đoán bệnh ngoài da qua hình ảnh

2.6.1. Giới thiệu

Mạng nơ-ron tích chập (Convolutional Neural Network - CNN) là một loại mạng nơ-ron nhân tạo (Artificial Neural Network - ANN) được thiết kế đặc biệt để xử lý dữ liệu có cấu trúc dạng lưới, ví dụ như hình ảnh. CNN được lấy cảm hứng từ cấu trúc của vỏ não thị giác của động vật, nơi các tế bào thần kinh được sắp xếp theo một cấu trúc phân cấp để xử lý thông tin thị giác.

Diagram of a diagram of a network

Description automatically generated with medium confidence

Hình 2.11. Mô hình mạng nơ-ron tích chập

2.6.2. Cấu trúc

Mạng nơ-ron tích chập thường bao gồm 3 lớp chính:

1. *Lớp tích chập (Convolutional layer):* Thực hiện phép tích chập giữa bộ lọc (kernel) và dữ liệu đầu vào để trích xuất các đặc trưng cục bộ.
2. *Lớp gộp (Pooling layer):* Giảm kích thước dữ liệu bằng cách lấy giá trị lớn nhất (max pooling) hoặc giá trị trung bình (average pooling) trong một vùng nhỏ.
3. *Lớp kết nối đầy đủ (Fully connected layer):* Kết nối tất cả các nơ-ron từ lớp trước đó và thực hiện phép phân loại dựa trên các đặc trưng đã trích xuất.

A screenshot of a computer

Description automatically generated

Hình 2.12. Mô tả cấu trúc mô hình CNN AlexNet

* **Lớp tích chập (Convolutional Layer)**

Lớp này là nơi thể hiện ý tưởng ban đầu của CNN. Thay vì kết nối toàn bộ điểm ảnh, layer này sẽ sử dụng một tập các bộ lọc (filters) có kích thước nhỏ so sánh với ảnh (thường là 5x5 hoặc 3x3) áp  vào một vùng trong ảnh và tiến hành tính tích chập giữa bộ lọc và giá trị điểm ảnh trong vùng cục bộ đó. Bộ lọc sẽ lần lượt được dịch chuyển theo một giá trị bước trượt (stride) chạy dọc theo ảnh và quét toàn bộ ảnh.

A close-up of a diagram

Description automatically generated

Hình 2.13. Hình ảnh mô tả lớp tích chập

Như vậy, trong quá trình tích chập, filter sẽ quét qua toàn bộ ảnh, từng phần một và tạo ra một feature map mới, trích xuất các đặc trưng quan trọng từ ảnh gốc. Số lượng feature maps phụ thuộc vào số lượng filter trong lớp tích chập. Ví dụ, nếu có 10 filter thì sẽ tạo ra 10 feature maps. Những ảnh này sau đó được truyền vào các lớp tiếp theo của mô hình để tiếp tục xử lý. Ban đầu, các trọng số trong filter được khởi tạo ngẫu nhiên và sau đó được điều chỉnh trong quá trình huấn luyện, giúp mô hình học cách nhận diện các đặc trưng chính xác hơn từ dữ liệu. Một hình minh họa của quá trình tích chập với filter 5×5 sẽ thể hiện rõ cách mà từng phần của bức ảnh gốc tương tác với filter để tạo ra feature map tương ứng.

* **Lớp gộp (Polling Layer)**

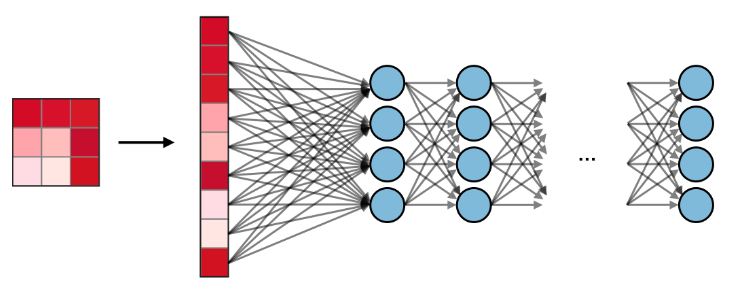
Layer này sử dụng một cửa sổ trượt để quét qua toán bộ ảnh dữ liệu, mỗi lần trượt theo một bước trượt (stride) cho trước. Khác với layer Convolution, layer Pooling không tính tích chập mà tiến hành lấy mẫu (subsampling). Khi cửa sổ trượt trên ảnh, chỉ có một giá trị được xem là giá trị đại diện cho thông tin ảnh tại vùng đó (giá trị mẫu) được giữ lại. Các phương thức lấy phổ biến trong layer Pooling là MaxPooling (lấy giá trị lớn nhất) và AveragePooling (lấy giá trị trung bình).

Bảng 2.1. Bảng minh họa phương thức Max Pooling và Average Pooling

|  |  |  |
| --- | --- | --- |
| **Kiểu** | Max pooling | Average Pooling |
| **Chức năng** | Từng phép pooling chọn giá trị lớn nhất trong khu vực mà nó đang được áp dụng | Từng phép pooing tính trung bình các giá trị trong khu vực mà nó đang được áp dụng |
| **Minh họa** |  |  |
| **Nhận xét** | * Bảo toàn các đặc trưng đã phát hiện * Được sử dụng thường xuyên | * Giảm kích thước feature map * Được sử dụng trong mạng LeNet |

* **Lớp kết nối đầy đủ (Fully Connected)**

Tầng kết nối đầy đủ (FC) nhận đầu vào là các dữ liễu đã được làm phẳng, mà mỗi đầu vào đó được kết nối đến tất cả neuron. Trong mô hình mạng CNN, các tầng kết nối đầy đủ thường được tìm thấy ở cuối mạng và được dùng để tối ưu hóa mục tiêu của mạng ví dụ như độ chính xác của lớp.



Hình 2.14. Hình ảnh mô phỏng lớp kết nối đầy đủ

2.6.3. Ưu điểm

Mạng nơ-ron tích chập (CNN) là một mô hình học sâu phổ biến, hiệu quả, đặc biệt trong xử lý hình ảnh và thị giác máy tính. CNN có khả năng tự động trích xuất các đặc trưng quan trọng từ dữ liệu thông qua các lớp tích chập, giúp nhận diện và phân loại hình ảnh với độ chính xác cao. Một ưu điểm đáng chú ý của CNN là khả năng giảm thiểu số lượng tham số nhờ các lớp tích chập và pooling, giúp giảm nguy cơ overfitting và cải thiện hiệu suất.

Bên cạnh đó, CNN có tính chuyển vị bất biến, cho phép nhận diện đối tượng trong ảnh bất kể vị trí, xoay, hoặc thay đổi kích thước. Mô hình này cũng rất linh hoạt và có thể được điều chỉnh để áp dụng vào nhiều lĩnh vực khác nhau như phân tích video, xử lý văn bản, và nhận diện giọng nói. Thêm vào đó, khả năng song song hóa trên GPU giúp tăng tốc quá trình huấn luyện và suy luận, làm cho CNN trở thành một công cụ quan trọng trong nhiều ứng dụng trí tuệ nhân tạo hiện đại như nhận diện khuôn mặt, xe tự lái, và chăm sóc sức khỏe.

2.6.4. Các kiến trúc CNN phổ biến

Một số kiến trúc CNN phổ biến được sử dụng trong bài toán chẩn đoán bệnh ngoài da:

* + *AlexNet:* Một trong những CNN đầu tiên đạt được kết quả vượt trội trong bài toán phân loại hình ảnh.[1][2]
  + *VGGNet:* Kiến trúc sâu hơn AlexNet, sử dụng nhiều lớp tích chập nhỏ để trích xuất đặc trưng.
  + *GoogLeNet (Inception):* Kiến trúc sử dụng các module inception để học các đặc trưng ở nhiều kích thước khác nhau.
  + *ResNet:* Kiến trúc rất sâu, sử dụng để kết nối tắt (skip connection) để tránh vấn đề gradient biến mất.

CHƯƠNG 3. KẾT QUẢ THỰC NGHIỆM

* 1. Dữ liệu thực nghiệm
     1. Nguồn gốc và đặc điểm của bộ dữ liệu
* **Nguồn gốc của bộ dữ liệu**

Các hỉnh ảnh tổn thương đến từ Bộ dữ liệu HAM10000[6] và được thu thập bằng nhiều loại da liễu, từ tất cả các vị trí giải phẫu (trừ niêm mạc và móng tay), từ lịch sử một số mẫu bệnh nhân được sàng lọc bệnh ngoài da của một số tổ chức khác nhau. Hình ảnh được thu thập với sự chấp thuận của Ủy ban đánh giá đạo đức tại Đại học Queensland (Nghị định thư số 2017/001223) và Đại học Y khoa Vienna (Nghị định thư số 1804/2017).

Sự phân bố của các trạng thái bệnh đại diện cho một thế giới thực tế đã được sửa đổi, trong đó có nhiều tổn thương lành tính hơn các tổn thương ác tính, nhưng đại diện quá mức cho các khối u ác tính.

Tất cả dữ liệu đầu ra được mã hóa trong một tệp CSV (giá trị được phân tách bằng dấu phẩy), với mỗi phản hồi phân loại liên tiếp. Các cột tệp phải là:

- IMAGE: Mẫu hình ảnh đầu vào.

- MEL: Chẩn đoán U ác tính.

- NV: Chẩn đoán Nốt ruồi.

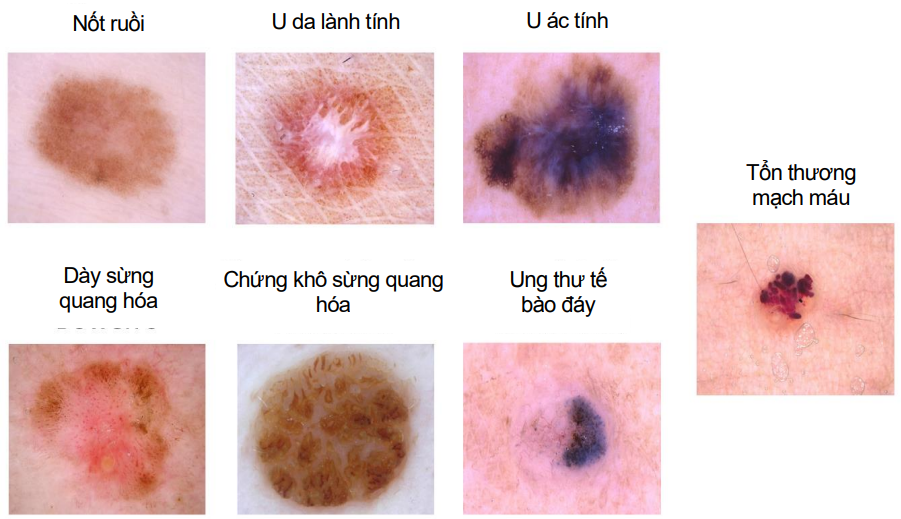
- BCC: Chẩn đoán Ung thư tế bào đáy.

- AKIEC: Chẩn đoán Dày sừng quang hóa.

- BKL: Chẩn đoán Chứng khô sừng quang hóa.

- DF: Chẩn đoán U da lành tính.

- VASC: Chẩn đoán Tổn thương mạch máu.



Hình 3.1. Hình ảnh một số bệnh ngoài da

* **Đặc điểm của bộ dữ liệu**

Tất cả các hình ảnh tổn thương được đặt tên bằng lược đồ ISIC\_.jpg, trong đó là một định danh duy nhất gồm 7 chữ số. Thẻ EXIF trong ảnh đã bị xóa; mọi thẻ EXIF còn lại sẽ không được dựa vào để cung cấp siêu dữ liệu chính xác.

Bộ dữ liệu gồm 39507 hình ảnh tổn thương da được chia thành hai tập train và test. Trong đó, mỗi tập được chia thành 7 folder gắn nhãn tương ứng với các loại bệnh. Tập train gồm 38569 files và tập test gồm 938 files.

TÀI LIỆU THAM KHẢO

[1]: Gershgorn, Dave (26 July 2017). ["The data that transformed AI research—and possibly the world"](https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/). *Quartz*.

[2]: Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E. (2017-05-4). ["ImageNet classification with deep convolutional neural networks"](https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf) (PDF). *Communications of the ACM*. **60** (6): 84–90.