

# Note\_Review\_ApplicationOfMachineAndDeepLearningToThyroid

April 6, 2024

## 1 Chủ đề

Ứng dụng của học máy và học sâu về tuyến giáp tế bào học và mô bệnh học: Review các phương pháp

## 2 Tổng quan

Đánh giá này tổng hợp các nghiên cứu trước đó về cách học máy và học sâu có thể cải thiện quy trình xử lý tế bào (cytology) và mô (histopathology) cho chẩn đoán ung thư tuyến giáp. Phương pháp tiêu chuẩn hiện nay trước khi thực hiện phẫu thuật là việc tiêm chọc tế bào mảnh (**cytology**) - lấy sinh thiết có độ **biến thiên giữa các quan sát viên cao**, thường trả về các mẫu không xác định và không thể nhận dạng một số bệnh lý một cách đáng tin cậy; **phân tích mô histopathology** giải quyết các vấn đề này một phần, nhưng nó yêu cầu **phải phẫu thuật cắt bỏ các mảnh u nghi ngờ** nên không thể ảnh hưởng đến quyết định trước phẫu thuật (tức là không thể dự đoán trước có thực sự nên cắt bỏ mảnh u nghi ngờ này không).

Bởi những vấn đề này, cũng như bởi sự thiếu hụt lâu dài về các bác sĩ bệnh học được đào tạo, đã có nhiều nghiên cứu về cách trí tuệ nhân tạo có thể cải thiện các quy trình hiện tại và giảm áp lực lên các bác sĩ. Nhiều nghiên cứu trước đây đã chỉ ra tiềm năng đáng kể của phân tích hình ảnh tự động trong việc phân loại các u tuyến giáp, đặc biệt là đối với các u của ung thư tuyến giáp biểu mô (có vẻ họ đang tập trung vào histopathology hơn là cytology), **nhưng những nghiên cứu này thường là quan sát lại**, vì vậy vẫn còn **câu hỏi về hiệu quả thực tế của các công cụ tự động này và thực tế của việc tích hợp chúng vào các luồng làm việc lâm sàng**. Hơn nữa, bản chất của việc phân loại u tuyến giáp là đáng kể phức tạp hơn trong thực hành so với nhiều nghiên cứu hiện tại đã giải quyết, và điều này, cùng với sự đa dạng của các luồng xử lý trong các phòng thí nghiệm khác nhau, có nghĩa là không có giải pháp nào đã chứng minh được sự mạnh mẽ đủ cho việc áp dụng lâm sàng.

Do đó, có nhiều hướng nghiên cứu cho tương lai: nghiên cứu việc thực hiện thực tế của các thuật toán này như hệ thống hỗ trợ quyết định của các bác sĩ bệnh học; điều cần thiết để xây dựng niềm tin với các bác sĩ và các cơ quan quản lý; và nghiên cứu phân loại đa lớp trên các tập dữ liệu đa trung tâm đa dạng, mục tiêu là **các phương pháp mà chứng minh được có hiệu suất cao một cách không phụ thuộc vào quy trình và thiết bị**.

## 3 Giới thiệu

### 3.1 Ung thư tuyến giáp

Tuyến giáp là một tuyến nhỏ hình con bướm nằm ở cổ – khối chất lỏng nhỏ hoặc khối rắn – có thể phát triển. Những ung thư tuyến giáp có thể được phát hiện bằng cách sờ nắn trong khoảng 2–6% trường hợp và có thể phát hiện qua siêu âm trong khoảng 19–67%. Đa số khả năng là lành tính, nhưng có khoảng 5% là ung thư.

Có bốn loại ung thư tuyến giáp (ung thư biểu mô) chính – nhú (PTC), tuyến giáp dạng nang (FTC), thoái hóa (ATC) và tủy (MTC) – chiếm 70–90%, 5-10%, 2% và <2% các trường hợp tương ứng.

Các biến thể / loại phụ tồn tại, chẳng hạn như các biến thể tế bào nang (FTC) hoặc tế bào nhú (PTC). Một chẩn đoán bổ sung lưu ý đặc biệt là ‘u tuyến giáp dạng nang không xâm lấn’ với các đặc điểm hạt nhân giống như nhú’ (NIFTP). Thuật ngữ này đã được giới thiệu vào năm 2016 để thay thế phương pháp đóng gói không xâm lấn biến thể nang của PTC (EFV-PTC): **mặc dù NIFTP có tế bào giống như PTC, chúng có tiềm năng ác tính thấp và việc phân biệt hiệu quả giữa NIFTP và PTC cổ điển có thể hỗ trợ trong việc giảm thang trị liệu.**

Các nốt lành tính thường được biểu hiện bằng các nốt không tân sinh (tăng sản đa nhân / bướu cổ, u nang) và tân sinh (nang tổn thương u tuyến - FA) (non-neoplastic (multinodular hyperplasia/goitre, cysts) and neoplastic [follicular adenoma (FA)] lesions).

#### Chẩn đoán và điều trị

Thông thường việc phát hiện ung thư tuyến giáp xuất hiện từ việc khám định kỳ. Cụ thể là thông qua siêu âm → xác định được thành phần, vị trí, độ vang, vôi hóa. Tùy thuộc vào kết quả siêu âm mà bệnh nhân được đề nghị chọc hút tế bào bằng kim nhỏ (FNA) để xác định rõ hơn.

The Bethesda System (TBS) for Reporting Thyroid Cytopathology thường được sử dụng cho các mẫu được thu thập bằng FNA.

Khoảng 10% FNA được thu thập không đạt yêu cầu chẩn đoán (TBS1) do các thông số như máu bị che khuất, bảo quản tế bào kém hoặc tế bào không đủ lấy mẫu, với khuyến nghị lặp lại sinh thiết dưới sự hướng dẫn của siêu âm để xác định vùng lấy sinh thiết hiệu quả hơn. Đối với một số bệnh ác tính tế bào học chẩn đoán là một thách thức, dẫn đến tỷ lệ cao sự biến đổi giữa các kết quả chẩn đoán bởi các trung tâm xác định khác nhau. Ước tính khoảng 15 – 30% trả về kết quả không xác định (TBS3 hoặc 4). Đối với một phát hiện, nguy cơ ác tính tương ứng đã được đưa ra trong một tổng hợp phân tích lên tới 30,5% đối với TBS3 và lên tới 28,9% đối với TBS4.

Nguy cơ tương đối cao thường thúc đẩy phẫu thuật chẩn đoán dưới hình thức cắt thùy hoặc cắt tuyến giáp, với phân tích các mô (histopathology) được cắt bỏ là tiêu chuẩn vàng để chẩn đoán. Phẫu thuật cắt bỏ là quá trình hành động được đề nghị cho các nốt ác tính, nhưng một lượng lớn tỷ lệ các lần phẫu thuật chẩn đoán rõ ràng là không cần thiết. Ca phẫu thuật mang theo rủi ro, có thể gây ra đau khổ và lo lắng đáng kể, và trong nhiều trường hợp cần phải điều trị thay thế thyroxine suốt đời. Nhiều bệnh nhân trải nghiệm điều này một cách không cần thiết.

Như vậy việc xác định các phân loại ung thư với tỷ lệ chính xác cực cao là cực kỳ quan trọng để tránh đau khổ và phiền hà cho bệnh nhân.

Tuy nhiên, việc **phân tích thủ công các mẫu sinh thiết và mô tổn rất nhiều công sức**, với áp lực thời gian đặt lên các nhà nghiên cứu bệnh học. Do khối lượng công việc tăng lên và sự thiếu

hạt kinh niên của nhân viên được đào tạo. **Các phương pháp dự đoán bệnh ác tính tốt hơn cần thiết trong suốt quá trình chẩn đoán để giảm bớt áp lực, giảm số ca phẫu thuật không cần thiết, và cải thiện sức khỏe chung của bệnh nhân.**

**Thử nghiệm phân tử** đã được đề xuất để tăng cường bệnh ác tính dự đoán về các nốt không xác định về mặt tế bào học, với nhiều nghiên cứu báo cáo thành công. Mặc dù những phương pháp như vậy chắc chắn có một vị trí trong chẩn đoán ung thư tuyến giáp, có mối lo ngại rằng các xét nghiệm này **thường giảm nguy cơ xuất hiện ung thư thay vì đảm bảo sự vắng mặt của nó**, bởi bị giới hạn ở một số ít chuyên môn hóa và tập trung cao độ phòng thí nghiệm và tăng tổng chi phí chăm sóc sức khỏe, điều này ngăn cản việc áp dụng lâm sàng ở những nơi có nguồn lực hạn chế\*\*.

Được thúc đẩy bởi những điều trên, chúng tôi tiến hành vào các kỹ thuật sử dụng trí tuệ nhân tạo (AI) có thể **cải thiện quy trình làm việc lâm sàng hiện có.**

## 3.2 Trí tuệ nhân tạo

### 3.2.1 Kiến thức nền tảng

AI là một lĩnh vực liên quan đến việc **dạy máy tính và máy móc cách đưa ra quyết định và giải quyết vấn đề một cách thông minh** (33). Về mặt lịch sử, nó liên quan đến việc tái tạo bằng máy tính khả năng của bộ não con người, mặc dù AI hiện đại ít tập trung hơn về việc bắt chước các quá trình sinh học và hơn thế nữa cho cả việc giải quyết các vấn đề phức tạp bất kể cảm hứng sinh học (34).

**Học máy (ML)** là một nhánh của AI được định nghĩa là nghiên cứu về các **thuật toán máy tính tự động cải thiện thông qua kinh nghiệm - quá trình học** (35). Dữ liệu được sử dụng để huấn luyện các thuật toán này thực hiện một nhiệm vụ – ví dụ, hồi quy, phân loại hoặc phân cụm – trong một cách tối ưu hóa một số số liệu hiệu suất mà không cần lập trình rõ ràng.

**Deep learning (DL)** là một nhánh của ML có **khả năng tự động hóa khám phá các mẫu trực tiếp từ dữ liệu thô** (36). Nó liên quan đến ứng dụng mạng lưới thần kinh – việc học tập ‘sâu’ như thế này. Mạng bao gồm nhiều lớp, lần lượt mỗi lớp bao gồm nhiều tế bào thần kinh tính toán – và đã chứng minh được hiệu quả cao thực hiện các nhiệm vụ liên quan đến **dữ liệu phi cấu trúc**, chẳng hạn như hình ảnh phân tích (37) và nhận dạng giọng nói (38).

Nghiên cứu AI đã tăng lên trong những năm gần đây do sự phát triển lớn hơn sự sẵn có của **các bộ dữ liệu lớn, sức mạnh xử lý được cải thiện** – đặc biệt với sự ra đời của các bộ xử lý đồ họa (GPU) để tính toán song song (39) – và tăng tính sẵn sàng của **các thư viện phần mềm nguồn mở** giúp dễ dàng thực hiện thuật toán.

### 3.2.2 Ứng dụng trong y sinh

Thường sử dụng ML & DL cụ thể:

ML và DL đặc biệt có thể áp dụng cho y sinh, vì chúng có thể được sử dụng để **khám phá các mẫu mà con người không nhìn thấy được** (chẳng hạn như trong phát hiện và phân tích di truyền), hỗ trợ **phân tích hình ảnh y sinh để giảm áp lực cho bác sĩ lâm sàng** và dự đoán kết quả từ dữ liệu lâm sàng (40). Ví dụ, họ đã được tuyển dụng để ước tính các tương tác sinh học chưa biết giữa các hợp chất thuốc và mục tiêu sinh học (41), dự đoán các tác dụng phụ khi phát hiện thuốc (42), dự đoán tính đặc hiệu của trình tự trong các protein gắn DNA và RNA (43, 44), tự động hóa việc giải thích siêu âm tim (45), tự động hóa việc phân loại các hình ảnh cụ thể

của cơ quan hoặc bộ phận cơ thể (46), và cả hai đều sàng lọc (47, 48), đồng thời dự đoán tỷ lệ tử vong và nhập viện, suy tim (49).

ML và DL cũng đã được áp dụng thành công cho nhiều lĩnh vực phân loại ung thư khác: để tự động phân loại các nốt từ **hình ảnh CT** (50) và dự đoán sự có mặt của đột biến từ **hình ảnh mô bệnh học** (51) trong ung thư phổi, để dự đoán tình trạng hạch nách từ **hình ảnh siêu âm** (52) và liên kết hình thái khối u và biểu hiện gen định vị theo không gian từ **hình ảnh mô bệnh học** (53) trong ung thư vú, và để phân loại và phân đoạn nghi ngờ tổn thương từ **MRI** (54) và tự động phân loại **sinh thiết Glory** (55) trong bệnh ung thư tuyến tiền liệt.

Trong bệnh ung thư tuyến giáp, ML và DL đã được áp dụng cho các trường hợp: - Phương pháp chẩn đoán hình ảnh: để **chẩn đoán di căn hạch cổ ở hình ảnh CT** (56) và **chẩn đoán các nốt tuyến giáp trong siêu âm** (57–59). - Sử dụng các **mô hình mô phỏng từ dữ liệu giải trình tự DNA và RNA** cũng tồn tại để chẩn đoán (26–29), rủi ro phân tầng (57) và dự đoán di căn hạch (58).

Đánh giá này sẽ tập trung vào việc áp dụng ML và DL để **tế bào học và mô bệnh học của ung thư tuyến giáp**. Các nghiên cứu nghiên cứu đã được xác định bằng cách tìm kiếm PubMed với các thuật ngữ tìm kiếm bao gồm **tuyến giáp ung thư, học máy, học sâu và trí tuệ nhân tạo**.

Các nghiên cứu sâu hơn được xác định từ các tài liệu tham khảo trong phạm vi phù hợp giấy tờ và đề xuất của người đánh giá. Để nắm bắt được các nghiên cứu chỉ có trong tài liệu gần đây được xuất bản từ năm 2017 đã được tóm tắt, với các tài liệu tham khảo đến các giấy tờ cũ hơn được cung cấp.

## 4 Các ứng dụng / phương pháp hiện tại

Nghiên cứu hiện tại tập trung vào **phân loại nhân tuyến giáp**. Điều này có thể rộng - ví dụ, phân loại các nốt là lành tính hoặc ác tính – hoặc dạng hạt, với các chẩn đoán phân biệt được đưa ra cho loại nốt cụ thể.

Nói chung, **các phương pháp tiếp cận cho cả tế bào và mô bệnh học** có thể được phân chia dựa trên việc họ có sử dụng phương pháp truyền thống hay không **ML hoặc DL**.

Hình 1 cho thấy các quy trình xử lý điển hình cho số hóa hình ảnh toàn bộ slide (WSI) trong cả hai trường hợp. WSI được vát lần đầu tiên thành các vùng ô nhỏ hơn; điều này có thể được thực hiện bởi một nhà nghiên cứu bệnh học, người có thể làm nổi bật các vùng thông tin và chú thích chúng (chẳng hạn như lành tính hoặc nghi ngờ ác tính), hoặc tự động, trong đó thông thường, **chẩn đoán ở cấp độ trượt (slide-level) sẽ được chuyển xuống các bản vá (patches-level)**.

Với các phương pháp **ML**, các **hạt nhân sau đó được phân đoạn** (có thể sử dụng DL hoặc các kỹ thuật khác), các **tính năng / đặc trưng** của chúng **được trích xuất và được tổng hợp vào thuật toán ML** được sử dụng để **phân loại bản vá** dựa trên về các giá trị đặc tính.

Đối với các quy trình dựa trên **DL**, các **bộ dữ liệu thường được tăng cường trước khi đưa vào mạng lưới thần kinh tích chập (CNN)**.

**Phân loại cấp độ bản vá cục bộ** sau đó có thể được tổng hợp thành **chẩn đoán cấp độ WSI toàn cầu - toàn bộ ảnh**.

Các lĩnh vực nghiên cứu khác bao gồm việc phân khúc hiệu quả tế bào nang, đánh giá cách phân mềm sàng lọc có thể cải thiện quy trình làm việc của nhà nghiên cứu bệnh học, dự đoán cả tình

trạng đột biến gen và di căn hạch.

## 4.1 Tế bào học (Cytopathology)

Bảng 2 tóm tắt nghiên cứu được công bố từ năm 2017 áp dụng AI với tế bào học ung thư tuyến giáp. **Mục tiêu khác nhau giữa các nghiên cứu** – ví dụ, một số người cố gắng phân biệt giữa lành tính và ác tính, trong khi những loại khác tập trung vào việc phân biệt FTC với FA – cũng như các kỹ thuật được sử dụng.

**Bản chất và kích thước của các tập dữ liệu và liệu các số liệu đánh giá có phù hợp hay không** được tính toán trên một slide / bệnh nhân hoặc mức độ hình ảnh được trích xuất. Kết quả là do đó không thể so sánh trực tiếp giữa các kết quả nghiên cứu.

Các bài báo cũ hơn từ năm 1996–2014, kém tiến bộ hơn và phức tạp.

### 4.1.1 Các phương pháp dựa trên học máy (ML)

**Phân loại ung thư biểu mô (Classification of carcinomas)** Margari và cộng sự. (61) đã sử dụng **cây phân loại và hồi quy (CART)** để đánh giá các tổn thương tuyến giáp và trích xuất kiến thức dễ hiểu về quá trình chẩn đoán. Nghiên cứu gồm 521 trường hợp lành tính (261), ác tính (256) và các nốt không xác định (4) được xác nhận bằng mô học. Phân loại đặc điểm hình thái tế bào được trích xuất và sử dụng để phát triển hai mô hình: CART-C để dự đoán danh mục TBS và CART-H để dự đoán chẩn đoán mô học lành tính hoặc ác tính.

- CART-C đạt được độ chính xác 91,0% khi dự đoán loại TBS; nếu TBS3 được sử dụng như một điểm cắt để phân loại các nốt là lành tính hay ác tính, CART-C đạt được độ nhạy và độ đặc hiệu tương ứng là 88,5% và 79,7%, và nếu sử dụng TBS4 thì các giá trị tương tự là 78,1% và 94,4%. Những cái này kết quả không khác biệt về mặt thống kê so với kết quả của các nhà giải phẫu bệnh.

- CART-H đạt độ chính xác 93,0%, độ nhạy 92,4% và 93,6% độ đặc hiệu khi dự đoán chẩn đoán mô học.

Maleki và cộng sự. (64) đã làm việc để phân biệt 3 loại PTC cổ điển với NIFTP và EFV-PTC mà không xâm lấn sử dụng **máy vectơ hỗ trợ (SVM)** được đào tạo về **mô tả văn bản**. Các trường hợp bệnh lý phẫu thuật với một trong các chẩn đoán, FNA phù hợp với khối u và mô tả vi mô có sẵn đã được chọn, dẫn đến 59 trường hợp (29 PTC, 30 NIFTP/EFV-PTC). Tổng cộng có 59 cụm từ tế bào học khác nhau được chiết xuất từ kính hiển vi mô tả; chúng được cô đọng thành 32 loại (với, dành cho ví dụ, tế bào ung thư và tế bào ung thư được giảm xuống thành một loại). Các bộ phân loại sau đó được huấn luyện bằng cách sử dụng tất cả các trường hợp ngoại trừ một trường hợp ngẫu nhiên trường hợp được chọn của PTC và NIFTP, được sử dụng cho sự đánh giá; điều này đã được thực hiện cho tất cả các lần lặp lại có thể bị loại trừ cặp. Bộ phân loại đạt độ chính xác 76,1%, độ nhạy 72,6% và độ đặc hiệu 81,6%. Các cụm từ chính liên quan đến NIFTP là ‘rất ít’ keo’, ‘mô hình vi nang’, ‘tế bào ung thư’, ‘đông đúc’ và ‘nhỏ tế bào nang’; những thứ liên quan đến PTC là ‘nhú’, ‘nhạt nhiễm sắc’, ‘tiêu điểm’, ‘hạt nhân nổi bật’ và ‘rãnh hạt nhân’.

**Các lĩnh vực nghiên cứu khác** Yao và cộng sự. (70) đã đánh giá **ThinPrep**, một chế phẩm hiện có kỹ thuật được tối ưu hóa cho bệnh lý kỹ thuật số và **thuật toán ML**, như một phương tiện **cải thiện độ chính xác chẩn đoán và khả năng tái tạo cho trường hợp TBS3**. Nghiên cứu của họ sử dụng 40 FNA (20 lành tính, 20 FA) xác nhận với mô học; sự khác biệt về hình thái giữa hai trường hợp này được biết là chủ quan hơn, ít chẩn đoán có thể tái tạo. Tổng cộng có 800 hình ảnh được trích xuất – **20 từ mỗi trường hợp**, 10 công suất trung bình (100x) và 10 công suất cao (400x) – với mỗi hình ảnh được nhà nghiên cứu tế bào học xem xét để đánh giá đầy đủ

tính di động và chỉ định một loại TBS. Thị giác máy tính truyền thống - các kỹ thuật dựa trên đã được sử dụng để **trích xuất 86 đặc điểm hạt nhân cấp thấp**, và chúng được nhóm lại để tạo thành 3 tính năng cấp trung và 3 tính năng cấp cao **dựa trên kiến thức tế bào học** của tác giả. Dốc bộ phân loại cây tăng cường và cây bổ sung đã được đào tạo riêng trên độ phóng đại công suất trung bình và cao. Mô hình có hiệu suất cao nhất đã được huấn luyện về bộ công suất trung bình và đạt độ chính xác 71%, 72% độ chính xác và khả năng thu hồi 71% trên bộ thử nghiệm, với FA được coi là lớp tích cực. Để so sánh, nhà tế bào học đạt được 63% chính xác, độ chính xác 57% và khả năng thu hồi 95%, cho thấy sự thận trọng hơn phương pháp tiếp cận với chi phí của độ chính xác rộng hơn.

Kể từ **sự ra đời của CNN và sự phát triển tiếp theo của DL** - dựa trên các phương pháp phân loại tế bào học, các tác giả sử dụng ML truyền thống đã áp dụng nó trong các trường hợp ít thông thường hơn, chẳng hạn như thay vào đó bằng cách sử dụng CART cho các tính năng dựa trên văn bản. Các thành công vừa phải của Maleki et al. (64) và Yao và cộng sự. (70) đề nghị rằng ML có một số tiềm năng để giải quyết những khó khăn trước phẫu thuật những thách thức, mặc dù vẫn cần phải cải thiện.

#### 4.1.2 Các phương pháp dựa trên học sâu (DL)

**Phân loại ung thư biểu mô** Savala và cộng sự. (60) đã sử dụng **mạng lưới thần kinh** để phân biệt các trường hợp của FA (26) và FTC (31). Hình ảnh được chuẩn bị với hai WSI nhiễu khác nhau - 'May Grunwald-Giemsa' (MGG) và 'Haematoxylin và eosin' (H&E) - đã được đưa vào và mô học được sử dụng làm tiêu chuẩn vàng, trong đó 90% trường hợp FTC được phát hiện là xâm lấn tối thiểu. Đơn hình ảnh ô được trích xuất thủ công - khoảng 100 cho mỗi trường hợp - các đặc điểm hạt nhân riêng lẻ sau đó được tính toán bằng hình ảnh phần mềm xử lý và các giá trị trung bình cho các phép đo hình thái khác nhau và các tính năng đo mật độ đã được đưa vào cho mỗi tập thể vật mẫu. Bộ xác nhận và bộ kiểm tra mỗi bộ chứa chín mẫu; cái trước được phân loại với độ chính xác 78%, độ nhạy 75% và độ đặc hiệu là 80%, và sau đó được phân loại hoàn hảo. Cho dù phân loại thử nghiệm hoàn hảo này, cần phải nghiên cứu thêm: thử nghiệm tập hợp nhỏ và có sự khác biệt lớn giữa kết quả đạt được trên các bộ kiểm tra và xác nhận có cùng kích thước làm nổi bật tính chất tự nhiên phương sai dự kiến ở quy mô như vậy. (Tức kết quả dự đoán trên tập test và tập valid quá khác biệt dù set có cùng size)

Sanyal và cộng sự. (62) đã **sử dụng CNN để phân loại PTC và không PTC**. Chỉ những trường hợp đã đạt được chẩn đoán tế bào học mới được bao gồm, và chẩn đoán đã được xác nhận bằng mô học; ranh giới trường hợp đã bị loại trừ. Để đào tạo người mẫu, 20 slide tế bào học (hoặc nhuộm Romanowsky- hoặc Papanicolaou) được chọn từ hai trung tâm khác nhau và hai kính hiển vi kỹ thuật số khác nhau được sử dụng để trích xuất 370 hình ảnh 512x512 pixel (184 PTC, 186 không PTC; 209 tại 10x và 161 ở độ phóng đại 40x) tập trung vào các khu vực chẩn đoán quan tâm. Một bộ thử nghiệm riêng biệt được phát triển bằng cách chọn 87 vùng từ mười vết bôi có hiển thị PTC (21) hoặc không PTC (78) hành vi, với tất cả các vùng được chụp ở mức 10x và 40x sự phóng đại. CNN đã phân loại những hình ảnh này một cách riêng biệt và khi sử dụng các tiêu chí quyết định dựa trên hoặc dựa trên (trong đó, nếu một mẫu được phân loại là PTC ở độ phóng đại, mức độ mẫu phân loại là PTC), CNN đã đạt được mức độ hình ảnh độ chính xác 85,1%, độ nhạy 90,5% và độ đặc hiệu 83,3%.

Quan và cộng sự. (63) đã **sử dụng CNN dựa trên VGG16** thông qua transfer học cách **phân biệt PTC với các nốt tuyến giáp lành tính**. Các nghiên cứu bao gồm 279 tiêu bản tế bào được nhuộm H&E (159 PTC, 120 lành tính). Tất cả các trường hợp PTC được phân loại là TBS5

hoặc 6 và có các đặc điểm PTC điển hình và chẩn đoán được xác nhận về mặt mô học; tất cả hình ảnh lành tính được phân loại là TBS2, nhưng bệnh nhân thì không phải phẫu thuật nên không có chẩn đoán mô học (pathology). Mỗi hình ảnh lớn hơn được phân chia thủ công thành nhiều hình ảnh nhỏ hơn 224x224 các mảnh pixel chứa các ô, mang lại 887 hình ảnh trong tổng cộng (476 PTC, 411 lành tính) và tập dữ liệu này đã được tăng thêm hệ số tám thông qua các phép lật và phép quay. CNN đã đạt được độ chính xác 97,7%, độ nhạy 100% và độ đặc hiệu 94,9% trên images-level và độ chính xác 95% ở cấp độ patient-level. Đặc điểm hạt nhân cũng tự động trích xuất và so sánh với t-test; số lượng đường viền, chu vi và diện tích ô và cường độ điểm ảnh trung bình là tất cả đều lớn hơn về mặt thống kê đối với các tế bào ác tính. Một hạn chế của nghiên cứu là rằng nó chỉ bao gồm các slide được phân loại là TBS2, 5 hoặc 6 (bài báo phân loại chỉ có 2 nhãn) – những hình ảnh FNA nói chung sẽ nắm bắt được – và các tác giả khuyến rằng trong tương lai nghiên cứu điều tra hiệu suất CNN trên các trường hợp không xác định (TBS3 hoặc 4) và các loại ung thư tuyến giáp khác.

Fragopoulos và cộng sự. (65) đã triển khai **mạng lưới thần kinh để phân loại WSI tế bào học chất lỏng** là lành tính hoặc ác tính. Các nghiên cứu bao gồm 447 mẫu (288 lành tính, 159 ác tính), tất cả đều có chẩn đoán mô học tiêu chuẩn vàng. Đường viền hạt nhân được thực hiện thủ công được đánh dấu từ mỗi slide và các đặc điểm hình thái (hình học và đo mật độ) sau đó được tách ra cho 41.324 hạt nhân.

Mô hình được đào tạo để phân loại những thứ này một cách riêng lẻ và được sử dụng các lớp chức năng cơ sở xuyên tâm thay vì kích hoạt điển hình hơn chức năng. Chẩn đoán mức độ trượt được xác định thông qua đa số phiếu, được dựa trên số lượng hoặc tỷ lệ phần trăm hạt nhân được phân loại là ác tính với tỷ lệ phần trăm hiệu quả hơn. Mô hình tốt nhất đạt độ chính xác 95,0%, độ nhạy 95,0% và độ đặc hiệu 95,1%.

Phạm vi Elliott và cộng sự. (66) đã phát triển một hệ thống bao gồm **hai CNN để dự đoán bệnh ác tính từ WSI tế bào học**, loại bỏ nhu cầu xác định thủ công các vùng thông tin của các ô như đã thấy trong các triển khai khác. Nghiên cứu bao gồm 908 FNA được nhuộm màu Papanicolaou với kết quả mô bệnh học đã được xác nhận chẩn đoán. Các trường hợp FNA và mô bệnh học không chẩn đoán được không lành tính hay ác tính đã được loại trừ. **CNN đầu tiên được sử dụng để xác định vị trí các vùng thông tin của các nhóm nang trứng; thứ hai phân tích các nhóm nang này và đưa ra phân loại TBS và chẩn đoán liên quan là lành tính hoặc ác tính.**

Để phát triển đào tạo được thiết lập cho CNN đầu tiên, một nhà nghiên cứu bệnh học được dán nhãn thông tin theo cách thủ công vùng. Các vùng không có thông tin là các vùng được chọn ngẫu nhiên quá trình quét; hầu hết quá trình quét không mang lại nhiều thông tin, vì vậy kết quả này có tỷ lệ cao xác suất cung cấp các vùng không chứa tế bào nang. CNN này đã được đào tạo, áp dụng cho từng WSI và được sử dụng để trích xuất 1.000 khu vực được xác định được xác định có xác suất xảy ra cao nhất mang tính thông tin; 1.000 khu vực này sau đó được sử dụng để đào tạo CNN thứ hai, phân loại các khu vực địa phương là lành tính hoặc ác tính và tổng hợp chúng thành một dự đoán cấp độ toàn cầu cho bệnh lý cuối cùng.

Cả hai CNN đều được đào tạo bằng phương pháp học chuyển giao và dựa trên VGG11. Hệ thống đạt được độ chính xác 90,8%, độ nhạy 92,0%, độ đặc hiệu 90,5% và AUROC là 0,932, giá trị cuối cùng ở mức của nhà nghiên cứu bệnh học ban đầu chẩn đoán (0,931). Các tác giả nhận thấy rằng việc kết hợp hệ thống của họ với chẩn đoán ban đầu đã cải thiện AUROC lên 0,962, nêu bật tiềm năng của **AI như một thử nghiệm phụ trợ**.

Dov và cộng sự. (69) đã **mở rộng công việc trên trong Elliot Range et al. (66)** bằng cách

sử dụng phương pháp học tập có giám sát yếu để không cần can thiệp dự đoán bệnh ác tính tuyến giáp từ cùng một bộ dữ liệu WSI. Thông thường, các slide tế bào học có cấu trúc phụ độc đáo với các trường hợp thông tin được phân bố thưa thớt trong suốt slide và vị trí và đánh giá các trường hợp này đặt ra một thách thức. Họ đã sử dụng một kỹ thuật dựa trên ước tính khả năng tối đa để truyền bá các nhãn cấp độ trượt đến các vùng cục bộ, sử dụng nhãn là nhãn ‘ồn ào’, dẫn đến chiến lược đào tạo được cải thiện.

Họ phát hiện ra rằng thuật toán hai giai đoạn của họ – tương tự như được sử dụng trong công việc trước (66) – đã đạt được trình độ con người ở cấp độ chuyên gia hiệu suất với AUROC là  $0,870 \pm 0,017$  để có hiệu suất tốt nhất thực hiện triển khai.

**Các lĩnh vực nghiên cứu khác** Zhu và cộng sự. (67) [và Tao và cộng sự. (87)] làm việc hướng tới hiệu quả **phân chia tế bào nang từ WSI**. Nghiên cứu bao gồm 43 WSI (17 PTC, 26 lành tính) và 6.900 bản vá đã bị cắt từ 13 trong số những thứ này (tất cả PTC) và được sử dụng để huấn luyện ngữ nghĩa dựa trên DeepLabv3 mô hình phân đoạn. Nghiên cứu **không áp dụng phương pháp học chuyển tiếp; thay vào đó, mô hình được đào tạo trực tiếp với tập dữ liệu**. Các tác giả đã **thêm một nhánh phân loại có thể chỉ định các bản vá là vùng tế bào nang, dạng keo hoặc nền**; khu vực được xác định là chứa các tế bào nang sau đó được đưa vào ngữ nghĩa cấu trúc phân khúc, giúp cải thiện hiệu quả – lên tới 93% thời gian phân đoạn được giảm đi bằng cách bỏ qua các vùng keo hoặc nền. Khi áp dụng cho 30 WSI thử nghiệm, hỗn hợp mô hình phân đoạn đạt được độ chính xác pixel là 99,3% trong 49,5 giây; nó hoạt động tốt hơn mạng tích chập hoàn toàn, U-Net và DeepLabv3, đạt độ chính xác pixel là 96,3%, 96,3% và 97,7% trong 370,8, 146,4 và 712,6 giây tương ứng.

Lin và cộng sự. (68) đã phát triển **phương pháp sàng lọc nhanh PTC phân đoạn từ WSI**. Nghiên cứu bao gồm 131 Papanicolaou- nhuộm các lam kính tế bào PTC FNA (120) và ThinPrep (11), và chú thích thực tế về các khu vực PTC được cung cấp bởi hai chuyên gia các nhà nghiên cứu bệnh học. Mỗi WSI lần đầu tiên được xử lý trước để loại bỏ các vùng nền, được vá và cuối cùng được phân đoạn bởi CNN với cốt lõi là dùng VGG16. Hệ thống của họ đạt được độ chính xác 99%, 86% độ chính xác và độ nhạy 94%, vượt trội so với U-Net và điểm chuẩn SegNet và có thể xử lý WSI nhanh hơn 7,8 lần và 9,1 lần tương ứng hơn các phương pháp đó.

Dov và cộng sự. (71) tiếp tục mở rộng công việc trước đây của họ bằng cách **kiểm tra tác động lâm sàng của phần mềm sàng lọc dựa trên AI sử dụng nghiên cứu trước của họ so sánh với đánh giá của nhà nghiên cứu bệnh học trên dữ liệu khác** cụ thể là trên 109 Sinh thiết FNA nhuộm Papanicolaou (84 lành tính, 25 ác tính) với và không sử dụng công cụ này. Nhãn được xác định dựa trên phẫu thuật kết quả giải phẫu bệnh. Ban đầu, nhà nghiên cứu bệnh học đánh giá tập dữ liệu một cách độc lập và sau khoảng thời gian 117 ngày, kết quả tương tự nhà nghiên cứu bệnh học đã kiểm tra cùng một tập dữ liệu với sự hỗ trợ của phần mềm. Hệ thống này bao gồm sàng lọc CNN dựa trên VGG11 thuật toán được đào tạo như trong tác phẩm trước đó của tác giả (66), với phần mềm trình bày lựa chọn 100 khu vực quan tâm cho nhà nghiên cứu bệnh học thông qua giao diện người dùng đồ họa cũng như gợi ý tiên lượng cho bệnh ác tính. Sự phù hợp của kết quả được đo với số liệu thống kê kappa có trọng số theo cặp; đối với TBS được chỉ định danh mục khi nhà nghiên cứu bệnh học đã sử dụng và không sử dụng phần mềm 0,924, cho thấy sự phù hợp gần như hoàn hảo. Khi sử dụng phần mềm, thời gian trung bình dành cho mỗi FNA là 81,6 giây và mặc dù tương tự số liệu thống kê cho đánh giá độc lập của nhà nghiên cứu bệnh học không với điều kiện là thời gian xử lý tình huống thấp này nêu bật mức độ hiệu quả của một quyết định hệ thống hỗ trợ có thể cải thiện quy trình làm việc của nhà nghiên cứu bệnh học. (Ngắn gọn là ứng dụng AI đã giúp giảm thời gian xuống).



Những điều trên đưa ra bằng chứng thuyết phục rằng **CNN nói riêng có thể cải thiện quy trình chẩn đoán ung thư tuyến giáp**. Hầu hết các nghiên cứu đều có tập trung vào việc xác định PTC hoặc dự đoán bệnh ác tính nói chung, nhưng có bằng chứng cho thấy FTC và FA có thể được xác định bằng cách sử dụng riêng các đặc điểm hạt nhân (60, 88, 89), một nhiệm vụ thường đặt ra nhiều thách thức trong quá trình phân tích slide thủ công, do đó thực hiện các phương pháp tiếp cận AI có khả năng hỗ trợ chẩn đoán tế bào học hiện có thể. Các thuật toán khó hoặc không thể thực hiện được làm loại bỏ yêu cầu về vấ thủ công và trích xuất tính năng được đặc biệt quan tâm (66): các giải pháp đòi hỏi ít sự can thiệp của con người hơn có tiềm năng lớn hơn để giảm bớt khối lượng công việc lâm sàng. Bài báo gần đây nhất của Dov et al. (71) cũng là đáng chú ý, vì nó điều tra tác động thực tế mà các phương pháp AI như vậy có thể gây ra một khi được dịch sang phòng khám, một bước tiến đáng hoan nghênh từ hơn phân tích hồi cứu thường được quan sát.

## 4.2 Mô bệnh học (Histopathology)

### 4.2.1 Các phương pháp dựa trên học máy

### 4.2.2 Các phương pháp dựa trên học sâu

Phân loại ung thư biểu mô

Dự báo đột biến gen (Prediction of genetic mutational status)

Các lĩnh vực nghiên cứu khác

## 5 Thảo luận

Đánh giá này đã xác định được rất nhiều bằng chứng thuyết phục gợi ý rằng AI có thể cải thiện tế bào học và mô bệnh học để chẩn đoán nguy cơ ung thư tuyến giáp.

Các vấn đề hiện tại với sinh thiết FNA tuyến giáp bao gồm tính biến thiên giữa các máy chủ cao (21, 22), một ý nghĩa quan trọng tỷ lệ trả về các mẫu không xác định (11, 16) và thực tế rằng **một số bệnh lý không thể được phân loại một cách đáng tin cậy bằng phương pháp tế bào học (76); mô bệnh học phân tích giải quyết những vấn đề này ở một mức độ nào đó nhưng đòi hỏi phải phẫu thuật cắt bỏ nên không thể hướng dẫn các quyết định trước phẫu thuật**. Nó quan trọng cần lưu ý rằng có sự khác biệt đáng kể giữa người quan sát không chỉ ở tế bào học tuyến giáp mà còn trong mô bệnh học tuyến giáp (104–107).

Ngoài ra, do tình trạng thiếu hụt trầm trọng các nhà nghiên cứu bệnh học được đào tạo (25), các công nghệ có thể làm giảm nhu cầu về thời gian của bác sĩ lâm sàng nên được hoan nghênh và sẵn sàng chấp nhận. Vấn đề có liên quan trước áp lực chưa từng có mà các dịch vụ y tế phải hiện đang hoạt động; trở nên trầm trọng hơn do đại dịch COVID-19, số bệnh nhân NHS đang chờ chăm sóc tự chọn đang ở mức cao kỷ lục sáu triệu tính đến tháng 11 năm 2021 (108). Công nghệ AI không được phép tất nhiên, được vội vã thực hiện mà không có sự đánh giá thích hợp; thực sự, chỉ cách đánh giá các thiết bị y tế dựa trên AI đã có hiệu quả câu hỏi gần đây (109). Phải hết sức thận trọng để đảm bảo các giải pháp hiệu quả, có thể chuyển nhượng, mạnh mẽ và không có thành kiến (110); ví dụ, cái sau đã gây ra những vấn đề đáng chú ý với các thuật toán được thiết kế để phát hiện ung thư da (111).

Mặc dù các phương pháp dựa trên AI chắc chắn cho thấy nhiều hứa hẹn trong lĩnh vực tuyến giáp

ung thư, còn thiếu nghiên cứu về tính khả thi lâm sàng của phương pháp này, ngoại trừ nghiên cứu của Dov et al. (71). Các nghiên cứu trong tương lai có thể nhằm mục đích làm sáng tỏ hơn nữa mức độ mà các nhà nghiên cứu bệnh học có thể dựa vào các hệ thống này để ra quyết định phụ trợ và điều tra có hướng dẫn.

Kiểm tra hiệu suất trên một bộ kiểm tra mang lại ý tưởng về độ chính xác và khả năng chuyển giao, nhưng **những mô hình này chắc chắn không thể được tích hợp vào thay thế ngay việc phân tích bệnh lý dựa vào con người**; hơn là, **sự tự tin phải được phát triển dần dần thông qua việc triển khai các quy trình làm việc được tăng cường bởi AI để xây dựng niềm tin**. Một phần nào đó hướng tới mục tiêu đó, và chắc chắn phù hợp với sự nhấn mạnh cần thiết gần đây vào việc phát triển AI có thể giải thích được, một số nghiên cứu gần đây đã điều tra các lĩnh vực hướng dẫn CNN phân loại (69, 97, 98); các việc tạo ra các bản đồ nhiệt mang tính thông tin có thể hướng dẫn các bác sĩ lâm sàng khu vực phân biệt đối xử và đẩy nhanh việc điều tra của họ. Tuyển giáp nghiên cứu bệnh lý có thể lấy cảm hứng từ những tiến bộ gần đây trong các lĩnh vực bệnh lý khác (112, 113) để phát triển hơn nữa những ý tưởng này. Kiến trúc đa tác vụ cũng có thể giúp ích về mặt này: việc triển khai có thể cung cấp cả sự phân loại toàn cầu và những bản đồ nhiệt và người ta có thể tưởng tượng một cánh tay bổ sung sử dụng xử lý ngôn ngữ tự nhiên, trong đó AI được sử dụng để tương tác với ngôn ngữ của con người, để cung cấp sự biện minh bằng văn bản, tất cả đều có thể củng cố niềm tin lâm sàng vào kết quả chẩn đoán. Nghiên cứu sâu hơn là chắc chắn được hoan nghênh trong lĩnh vực này: khả năng diễn giải là quan trọng đối với thiết lập niềm tin cần thiết với các bác sĩ lâm sàng và cơ quan quản lý, và khám phá tính thực tiễn về cách các công nghệ này có thể cải thiện việc kiểm tra định kỳ sẽ hỗ trợ việc áp dụng lâm sàng (40).

Đáng chú ý, **phần lớn nghiên cứu trước đây tập trung vào hệ nhị phân là phân loại no PTC và PTC do PTC chiếm ưu thế hơn các loại khác khối u ác tính khác**, nhưng việc kiểm tra FNA tuyển giáp và phẫu thuật mẫu vật có nhiều sắc thái hơn đáng kể so với nhiều nghiên cứu trước đây đã giải quyết. Nghiên cứu được thực hiện bởi Wang et al. (91) là một ngoại lệ với việc tác giả đạt được độ chính xác cao khi **cố gắng phân loại đa dạng, mặc dù có một hạn chế là loại trừ các trường hợp mà các nhà nghiên cứu bệnh học không đồng ý**.

Ngoài ra, do các bộ dữ liệu trong phòng thí nghiệm thường không đồng nhất – các phòng thí nghiệm khác nhau thường sử dụng các phương pháp tế bào khác nhau và chuẩn bị mô, cố định, nhuộm màu và chụp ảnh – các thuật toán thường phải được điều chỉnh cho phù hợp với nhu cầu lâm sàng của từng phòng thí nghiệm (99).

Để thực sự có tính khái quát và mạnh mẽ, một phương pháp phải có khả năng giải quyết những khác biệt về mặt thể chế này. Một số nghiên cứu đã bắt đầu giải quyết vấn đề này - ví dụ: bằng cách kiểm tra các tập dữ liệu có nhiều vết bản và từ các tổ chức khác nhau (60, 62, 99), hoặc bởi thực hiện chuyển đổi màu cho các cường độ nhuộm màu khác nhau (94) – nhưng các nghiên cứu trong tương lai bao gồm các bộ dữ liệu đa trung tâm đa dạng và thể hiện hiệu suất cao trong quy trình và thiết bị 1 cách bất khả tri được khuyến khích. Học tập liên kết và miền sự thích ứng là hai con đường cần được điều tra thêm trong vấn đề này tôn trọng: chúng đã được áp dụng trong các lĩnh vực bệnh lý khác để cải thiện khả năng chuyển giao liên phòng thí nghiệm (114, 115).

Ngoài ra, mặc dù phần lớn các nghiên cứu trước đây **tập trung vào phân loại chẩn đoán các mẫu tế bào và mô bệnh học, những con đường khác điều tra các lĩnh vực khác nhau của bệnh ung thư tuyển giáp liệu pháp chắc chắn tồn tại**. Dự đoán di căn hạch ở công việc của Esce et al. (95) là một ví dụ có ý nghĩa lâm sàng thực sự: chính các tác giả thừa nhận ý tưởng về một thuật toán thời gian thực có thể hướng dẫn quyết định có thực hiện mổ khối u ngay

hay không. Hơn nữa, mặc dù việc phân loại NIFTP có được nghiên cứu ở mức độ nào đó (64, 93, 99), nó đã chứng tỏ là một lĩnh vực đầy thách thức của nghiên cứu và, với tiềm năng đáng kể về việc nhận dạng NIFTP đáng tin cậy để hỗ trợ việc xuống thang trị liệu, điều này đòi hỏi phải tiếp tục nghiên cứu về cách công nghệ AI có thể cải thiện các quy trình hiện tại.

Nhiều phương pháp được xem xét ở đây yêu cầu tính đại diện nhân được gán cho các bản vá được trích xuất. Các mô hình bản thân chúng thường không lấy toàn bộ WSI làm đầu vào; thay vào đó họ đưa ra dự đoán về những hình ảnh được phân đoạn nhỏ hơn này trước tổng hợp những dự đoán riêng lẻ này thành một dự đoán bệnh nhân hoặc trượt dự báo mức độ. Để tạo ra một tập huấn luyện gồm các bản vá có nhãn – trong đó WSI thường được kiểm tra thủ công để tìm thông tin các nhà nghiên cứu bệnh học – là công việc tốn nhiều công sức và tốn kém đòi hỏi kiến thức chuyên môn. Trong một số nghiên cứu, các nhà nghiên cứu đã tự động vá các khu vực của WSI và đã phá vỡ yêu cầu chú thích thủ công các hình ảnh được phân đoạn bằng chỉ định chẩn đoán WSI ban đầu (91) – hậu quả của việc đó một cách tiếp cận là các tính năng đại diện của cấp độ trượt chẩn đoán sẽ không tồn tại ở tất cả các miếng vá, vì nó thường gặp ở tế bào hoặc WSI mô bệnh học để chứa các vùng của cả mô bình thường và mô bệnh lý nên có một. Hơn nữa, để thực hiện một số các mô hình hiện tại trong thực hành lâm sàng, một nhà nghiên cứu bệnh học có thể phải xác định thủ công các vùng đại diện để sử dụng làm đầu vào. BẢNG đã khám phá ở trên, một số nghiên cứu đã giải quyết vấn đề này bằng các phương pháp mới các phương pháp tiếp cận, bao gồm nhận dạng vùng thông tin tự động, học tập tích cực và học tập đa trường được giám sát yếu (66, 67, 69, 87, 97, 98); các nghiên cứu trong tương lai có thể nghiên cứu sâu hơn như các kỹ thuật đưa quy trình chẩn đoán đến gần hơn với tự động hóa thực sự.

**Tóm lại việc lấy mọi patch thu được từ 1 ảnh nhân A rồi gán mọi patch đó là nhãn A có thể khiến mô hình học sai nghiêm trọng.**

Tóm lại, mặc dù AI đã cho thấy tiềm năng to lớn trong việc cải thiện đường ống chẩn đoán ung thư tuyến giáp, nghiên cứu hiện tại bị ảnh hưởng một số hạn chế: **thiếu tập trung vào việc tích hợp lâm sàng các ứng dụng dựa trên AI, các phương pháp và cách thức chúng có thể cải thiện quy trình làm việc trong thực tế; việc sử dụng các nhãn cấp độ bản vá để thu được tập huấn luyện có thể tốn nhiều công sức và tốn kém; và tập trung vào phân loại nhị phân và PTC, trái ngược với sự phân loại đa dạng của tất cả các kiểu.** Những cái này những hạn chế nêu bật con đường cho nghiên cứu trong tương lai: đánh giá tiềm năng thực tế của một thuật toán để hỗ trợ quyết định lâm sàng làm; mở rộng nghiên cứu hiện tại về khả năng giải thích, có thể giúp xây dựng niềm tin với các bác sĩ lâm sàng và cơ quan quản lý; điều tra thêm các kỹ thuật chỉ yêu cầu nhãn cấp độ trượt và có thể hướng dẫn các nhà nghiên cứu bệnh học đến các vùng quan tâm; và kiểm tra các bộ dữ liệu đa trung tâm lớn để phát triển các kỹ thuật mạnh mẽ không phụ thuộc vào quy trình và thiết bị của từng phòng thí nghiệm.