

Development of Visual Diagnostic Expertise in Pathology: An Information-processing Study

Journal of the American Medical Informatics Association 10 (1) 39--51
Crowley, R, S., Naus, G, J., Stewart III, J., and Ridman, C, P.

1. Introduction

- 微細病理学診断とは
 - 生検や手術により細胞を採取し，マイクロスコープにより組織を同定
- 微細病理学診断における熟達化
 - 最低 5 年のトレーニング（研修医の制度）
- 本研究の目的
 - 微細病理学診断における熟達化の背景にある認知プロセスの理解
 - Novice – Intermediate - Expert の比較
 - Intelligent tutoring system (ITS) へのインサイトを得ること
 - 課題分析
 - スキャフォールドの同定
 - Expert の保持する認知ルールの同定
 - Novice が起こしがちなエラーの同定

2. Background

2.1. 微細病理学診断に関する実証研究

- 皆無

2.2. 臨床診断（カルテからの診断）に関する実証研究

- Schmidt, Norman, and Boshuizen (1992)
 - 熟達化は多数の症例の経験によって“疾患に関するスクリプト”を獲得するプロセス
- Patel et al. (1986; 2001)
 - 熟達化によって推論の方向が変化
 - Novice はトップダウン（後ろ向き）に診断を行うため，間違える
 - Expert は所見からのボトムアップ（前向き）に診断を行い正確

2.3. 放射線診断

- Lesgold et al (1988)
 - Expert の特徴
 1. Expert は多くの所見を報告
 2. 因果関係に関する多くの言及
 3. 多く，かつ長い推論の連鎖
 - Expert の典型的な診断プロセス
 1. 画像提示直後にスキーマのトリガリング
 2. スキーマのチューニング
 - 熟達化は知覚と認知のバランス変化のプロセス
 - 1 初期の知覚学習
 - 2 推論ルールの獲得
- Azevedo and Lajoie (1997)
 - マンモグラフィ診断における熟達化の検討 ITS の開発
 - 正確さ，診断所要時間，所見数，オペレータ数，エラー数に差がない
- Kundel et al (1969; 1983; 1987; 1996)
 - 信号検出理論に基づく眼球運動の測定

- 熟達化の効果
 1. 診断時間, 診断効率, 診断能力の向上
 2. トップダウン, ないし後ろ向き推論の増加 (という仮説を提唱)
Patel et al と矛盾
 - 診断のプロセス
 1. グローバルインプレッション
 2. 探索と発見
 3. 内省的探索
 4. 探索後の想起
- 2.4. 皮膚科診断 (マイクロスコープ)
- Norman et al (1990)
 - 2つの熟達化モデルを検証
 - モデル 1: “Independent cues”の学習
疾患間を効率的に区別できる特徴の学習
 - モデル 2: 事例ベースのカテゴリ化
以前見た症例とのパターンマッチングにより診断能力が向上
 - 実験結果から, モデル 1 よりもモデル 2 が妥当
- 2.5. 微細病理学に固有の事情
- 一度に見ることのできるエリアが限定
 - スライド内のエリアを動き回る必要 系列的な探索
- 2.6. 著者たちの考える熟達化
- 3つの側面: 認知的な熟達化, 社会文化的な熟達化, 団体組織的な熟達化
 - 熟達化の理解には認知科学的手法が有用
 - 解像度の高い課題分析が可能
- 2.7. 著者たちの考えるモデル = Dreyfus & Dreyfus のモデル
- 5段階の系列的な発展
 - Novice Beginner Competent Proficient Expert
 - 熟達化による変化
 - 抽象的知識, ルールから過去経験(インスタンス)へ
 - 状況付けられた知覚の発達
 - 観察者から実践者への変化の移行
3. Research questions regarding microscopic diagnostic pathology
- 熟達化によって診断のパフォーマンスは変化するか
 - 熟達化によって自身の診断に対する自信や診断の困難さに関する評価は変化するか
 - 熟達化によって認知プロセスは変化するか
 - 熟達化によってエラーの質は変化するか
 - 熟達化によって診断プロセスは時間的に変化するか
4. Methods
- 4.1. Research design
- 独立変数: 熟達化 (Novice, Intermediate, Expert)
 - 従属変数: 正確さ・評価値・プロセス・エラー・時間
 - 実験材料: 確定診断付き胸部 8 症例
 - 被験者: 28 人, 1 人 4 症例ずつ診断

- 実験材料×被験者 118 ケース
 - 分析対象: 108 ケースのみ (テクニカルな問題から 10 ケースを除外)
- 4.2. Case Materials (Table 1)
 - 単一の大学病院のファイルから選択 (確定診断付き)
 - 新たに病理学者 1 名が診断を加え, オリジナルの診断と一致したもののみ採用
 - バリエティに考慮 - 難しさ, 病巣の大きさ, 典型性, 複数疾患
 - 診断の怪しい症例 (ボーダーラインの症例) は除外
 - 症例の構成
 - 単一のスライドから構成
 - 1 文, もしくは 2 文の履歴つき
- 4.3. Participants
 - Novice: 10 人 ピッツバーグ大学医学部学部 3 年
 - Intermediate: 10 人 2 年目, 3 年目の研修医 (ほとんどがピッツバーグ大)
 - Expert: 8 人 学会役員, 26.5 年の経験 (ピッツバーグ大を含む複数大から採用)
Novice, Intermediate のみ謝礼
ピッツバーグ大以外の被験者はメールや学会におけるポスターで公募
- 4.4. Task
 - Lesgold et al. を参考に臨床情報进行操作
 1. 臨床情報なしで診断
 2. 診断が終わった後に臨床情報の提示, 臨床情報ありで診断
 3. 「診断の確実さ」・「症例の困難さ」を評定(視覚的アナログ尺度)
- 4.5. Data collection
 - Think aloud 法 (認知科学における標準的技法) によりプロトコルデータの記録
 - 被験者が見ているスライドを記録 (小型カメラ)
 - CD-ROM に DV データの記録 発話の起こし
- 4.6. Coding schemes
 - 分析対象となった 108 事例のうち 24 事例をコーディングスキーマ作成用データに
 - 2 種類のコーディングスキーマ プロセスとエラー
 - 4.6.1. プロセス
 - Newel & Simon (1972) を理論的背景
 - 発話プロトコルから知識状態の推移プロセスを同定
 - 各プロトコルセグメント オペレータと引数
 - PAW (Protocol Analyst's Workbench) の利用
 - オペレータに関するテンプレートの定義, 参照機能
 - 24 事例を繰り返しコーディングし, コーディングルールの決定
 - 4.6.2. Error coding
 - 同様に 24 事例を繰り返しコーディングし, コーディングルールの決定
 - エラーとみなせるものは全て挙げつくした
 - ケースレベルのエラー, プロトコルレベルのエラーに分けた
- 4.7. Protocol coding
 - 第一著者による

- PAW を使用してコーディングの揺れを最小限
- コーディングされたファイルは Excel, SPSS, Statview によって解析

4.8. Measurements

4.8.1. Coding Diagnostic accuracy

- カルテ提示前, 提示後, 各々の診断について正解/不正解を判断
- 2 種類の判断基準
 - Specific: Gold standard に記載された病名, もしくは同義語のみ正解
 - General: Gold standard に記載された病名, もしくは親カテゴリが同じもののみ正解

4.8.2. Measures of certainty and difficulty

- 「確実さ」「困難さ」の評定 (視覚的アナログ尺度) を 0 から 10 に数値化

4.8.3. Process and error measures

- オペレータの頻度を症例ごとにカウント
- オペレータ間の推移関係を検討
- 独立した仮説の数を検討
- 診断中に正解となる疾患を想起したかを検討

4.8.4. コーディングの一致率

- 材料: 無作為に選択された 12 ケース (11%)
- オペレータに関する一致率のみ検討
 - 同一評定者内での一致率 3 ヶ月後 89%
 - 異なる評定者での一致率 79%
 - 評定者: JS 病理学関係の学会役員, プロトコル分析の経験なし
 - 2 時間のトレーニング
- 信頼性はギリギリのところ

4.8.5. Measuring latencies

1. 課題にかかった総時間
2. 解剖学的位置の同定にかかった時間
3. 初期仮説を述べるまでの時間
4. 病巣のある症例では病巣を同定するまでの時間

4.9. Analysis

- SPSS, SAS を使用
- 評定値・プロトコルの頻度・診断時間 一元配置分散分析
被験者をユニット
多重比較に Tukey-HSD 法を利用
- 正解率 2 検定
ケースをユニット
- 有意水準は 5% に設定

5. Results

5.1. Diagnostic accuracy (Table 4)

- カルテ提示前・提示後/Specific・General
 - 全ての指標で Expert>Intermediate>Novice
- Novice のみがカルテの提示によって意見を変化
- Novice のみがカルテ提示後で正確さを上昇

5.2. Ratings of certainty and difficulty (Table 5)

- 確実さ: Expert > Intermediate > Novice
- 困難さ: Expert < Intermediate < Novice
- 正確な診断をおこなったケースと不正確な診断をおこなったケースでの評定値の差
 - Intermediate の「確実さ」のみ有意 ($t = -2.45, p = 0.019$)
 - Intermediate は自信のないときに不正確な診断

5.3. Process measures (Table 6)

5.3.1. Data examination

- データの調査, 所見を発見すること
- 量が多かったため 3 カテゴリに分割
 - Visual identification (全データの 35%)
 - 知覚的な所見を同定すること
 - Expert < Novice = Intermediate
 - より詳細な分析
 - Novice は正常所見に関する言及が多い
 - Intermediate は組織に関する言及が多い
 - Visual comparison (全データの 5%)
 - 複数所見の比較や関係性についての言及
 - Intermediate > Expert = Novice
 - Examination of history (全データの 4%)
 - 臨床履歴の参照
 - 統計的差なし

5.3.2. Data exploration and explanation (全データの 7%)

- データの評価, 説明
- Intermediate > Expert = Novice

5.3.3. Data interpretation (全データの 20%)

- データの解釈
- Intermediate > Expert = Novice

5.3.4. Control processes (全データの 2%)

- 自身の診断プロセスに関する言及
- 統計的差なし

5.3.5. Operational processes (全データの 6.5%)

- マイクロスコープの操作に関する言及
- Intermediate > Expert = Novice

5.3.6. Goal-setting (全データの 5% 以下)

- 考慮に入れている診断の限定, 仮説の統制, まだ見つけていない所見を口に出すこと
- 統計的な差なし

5.3.7. Unique hypothesis

- 診断プロセス全体の中で独立した仮説の数
- Expert = Intermediate > Novice

5.3.8. 最終的に誤った診断を下したケースの途中で正しい診断を考慮した率

- Novice 15%

- Expert 45%

5.4. Latencies (Figure 1)

- 全体的に Expert は素早い診断
 - 何らかの仮説を言及 (H1) : Expert = Intermediate < Novice
 - 最終仮説の言及 (FH) : Expert < Intermediate = Novice
 - 解剖学的部位の言及 (AL) : Expert < Intermediate = Novice
 - 病巣の言及 (DFL) : Expert < Intermediate = Novice
- Expert の H1・FH の差が短い理由
 - Expert のほとんどの H1 は FH と同じ
 - Novice , Intermediate は試行錯誤を繰り返す
- Expert , Intermediate の FH・END が長い理由
 - 仮説検証 , 対立仮説の棄却が長い

5.5. Error measures (Table 3)

- Case レベルのエラー
 1. 診断すべき場所を見なかった
 2. 診断すべき場所を見たのにも関わらず認識しなかった
 3. 解剖学的部位を同定しなかった
 - A) 解剖学的部位を同定しようとしなかった
 - B) 同定しようとしたができなかった
 - C) 間違えて同定した
- プロトコルレベルのエラー
 4. 正常構造を誤って同定
 5. 異常構造を誤って同定
 6. データを誤って解釈
 - A) 解剖学的部位を同定しようとしなかった
 - B) 同定しようとしたができなかった

5.5.1. 被験者群間でのエラーの差

- おおまかな傾向として
 - Expert = Intermediate < Novice

5.6. 正しい病巣を見つけた後に正しい診断を下した割合

- Intermediate 36%
- Expert 90%
- $\chi^2 = 14.27, n = 54, df = 1, p < 0.0001$
- Intermediate は病巣をすぐに見つけるのにも関わらず , 正確に解釈できない

6. Discussion

6.1. Significance

- 微細病理学診断における熟達化の効果を確かめた初の研究
 - 6.1.1. 微細病理学診断の特徴 1
 - マイクロスコープの扱いに対する意識的な注意の必要性
 - Intermediate: 操作に関する発話の多さ
 - Expert が見せた特徴: 操作に関する発話ほとんどなし
- Dreyfus & Dreyfus のモデルと一貫
 - 熟達化によってマイクロスコープが身体化

6.1.2. 微細病理学診断の特徴 2

- Intermediate が病巣を見つけたのに診断できなかった
 - 探索プロセスと診断プロセスのカップリングができていないということ
 - 物理的探索・知覚・注意と診断は別物
- Lesgold et al のモデルと一貫
 - 熟達化の初期で知覚学習がおきる
 - その後に推論と関係する認知プロセスの発達がおきる

6.1.3. 微細病理学診断の特徴 3

- Intermediate は視覚的特徴の同定において Novice よりも優れる
- 経験による知覚的シンボル・ボキャブラリーの獲得

6.1.4. 微細病理学診断の特徴 4

- Expert は Intermediate よりも所見の言及が少ない
- コンパイルの理論 (ACT-R) と一貫
- Deyfus のモデルと一貫 (部分の知覚から全体のパターン認知)

6.1.5. 前向き推論 vs. 後ろ向き推論について

- 本研究では直接的に検討できなかった
- 微細病理学診断という課題はこのことを検討するのに利点があるようにも見える
スライスの切り替え作業により、プロセスの系列的データを採取可能
- 今後の課題
 - 2つの推論をどのように統合するのか？
 - いつ推論方略をスイッチするのか？
 - どういった要因が二つの推論のバランスを変化させるのか？
 - 知識ベースの量や構造はそれらのバランスに影響するのか？
 - どのようにして探索、知覚、推論がインタラクトするのか？
 - どのようにして、ゴールドリブンの探索が起きるのか？
 - どのようにして知識ベースがそれらのプロセスに影響するのか？
 - それらのストラテジーとエラーにどういった関係があるのか？
 - どのようにしてそれらの理解を支援につなげられるか？

6.2. Limitations of the present study

- 被験者数の少なさ
 - プロトコル分析に必然的におきる問題
 - 実験結果を他集団・他症例へ適用することは困難
- ホーソン効果
 - 実験に参加するという意識がパフォーマンスを向上させたのではないが
 - 被験者群間でホーソン効果の影響に差があるとは思えない
 - 相対的な比較は妥当
- think-aloud 法による弊害
 - フィルタリングしないこと、自己説明をするわけではないことを教示
 - 思考が変化することはない (Ericsson, & Simon, 1993)
 - それにしても課題時間が伸びた
 - 相対的比較に限定 (絶対的な時間に意味なし)

6.3. Implications for future microscopic pathology education systems

- 物理的探索を支援するシステムの重要さ

- Novice のブアさ
- Novice には物理的探索の支援をしたほうが良い
- 伝統的に教育システムは知識偏重，物理的探索を支援してこなかった
 - コンパイルの前に視覚的特徴の同定が必要
 - 視覚的特徴に対する明示的なフィードバックを与えるシステム

7. Conclusion

- 微細病理学における熟達化に関する初の検討
- 異なる認知スキルが熟達化に関与していることがわかった
- たくさんのエラータイプを発見することができた
- エラーを列挙してリストを作ることで，教育システムを作れる

Table 1
Case Materials

Case	Gold Standard Diagnosis	Description
1	Infiltrating Ductal Carcinoma	Focal lesion of poorly differentiated cancer adjacent to biopsy site and scar.
2	Ductal Carcinoma in Situ (DCIS)	Widespread solid and cribriform in-situ carcinoma present throughout majority of sample.
3	Infiltrating Lobular Carcinoma	Widespread classical type infiltrating lobular carcinoma. Scant adjacent normal tissue.
4	Lobular Carcinoma in Situ (LCIS)	Small focus of LCIS with retrograde extension in otherwise normal breast.
5	Fibroadenomas, Sclerosing Adenosis and Intraductal Papilloma	Multiple focal lesions, including sclerosing adenosis - a benign lesion that shares some visual features with cancer.
6	Paget's Disease	Nipple with focal area of intra-epidermal Paget's disease. No underlying carcinoma.
7	Adenomyoepithelioma	Small circumscribed lesion with uniform features.
8	Atypical Papilloma	Large lesion with numerous atypical features

Table 2
Fragment of protocol showing process and content coding

Line	Protocol segment	Process codes (Operators)	Content codes (Arguments)	
			Descriptor	Value
137	"I have to figure out areas of obvious necrosis"	Set-goal-identify	Finding	areas of obvious necrosis
138	"No hemorrhage."	Note-absent-finding	Absent-Finding	hemorrhage
139	"I still think it is malignant."	Statement-of-hypothesis	Hypothesis #	4
			Hypothesis Category	benign/malignant
			Hypothesis	malignant
140	"I still think it is breast."	Identify-anatomic-location	Location	breast
141	"I think..."	Not-coded	Reason	incomplete thought
142	"Could it be a lymph node that is replaced by something?"	Statement-of-hypothesis	Hypothesis #	5
			Hypothesis Category	general
			Hypothesis	replaced lymph node
143	"Except for the fat, I don't see anything else."	Confirm-with-present-finding	Hypothesis #	5
			Hypothesis category	General
			Hypothesis	Replaced lymph node
			Present-finding	Fat

コーディングスキーマ

メインカテゴリ	オペレータ (総数 48)
Data examination	<i>Identify-normal-structure, Identify-histopathological-cue, Compare-findings-from-multiple-location</i>
Data exploration and explanation	<i>Evaluate-certainty-findings, Evaluate-salience, Associate-location</i>
Data interpretation	<i>Statement-of-hypothesis, Confirm-hypothesis-with-present-findings, Recall-evidence-hypothesis-relationship, Evaluate-certainty-hypothesis</i>
Control processes	メタ推論, プランニング
Operational processes	マイクロスコブの操作

Table3. Errors

Errors	Description	Novice		Intermediate		Expert		Statistics			
Case Level	Errors coded as present or absent in each case	Number of cases / Total (%)		Number of cases / Total (%)		Number of cases / Total (%)		Chi-Square	P Value	Pairwise Comparison	P Value
1	Lesion never brought under objective	8/30 (26.7 %)		1/28 (3.6 %)		0/23 (0 %)		13.25	.0021	N, I N, E E, I	.0059 .0006 .3189
2	Lesion traversed without recognition	7/30 (23.3 %)		0/28 (0 %)		0/23 (0 %)		8.11	.0209	N, I N, E E, I	.0056 .0058 -
3	Error in identifying anatomic location	14/40 (35.0 %)		2/38 (5.3 %)		0/32 (0 %)		22.76	<.0001	N, I N, E E, I	.0008 <.0001 .1612
Segment Level	Errors counted for each case	Mean/ case	SD	Mean/ case	SD	Mean/ case	SD	F Value	P Value	Pairwise Comparison (Tukey HSD)	P Value
4	Incorrectly names normal structure	0.35	0.74	0.11	0.31	0	0	5.28	.012	N, I N, E E, I	.068 .013 .656
5	Incorrectly names histopathologic cue	0.93	1.4	0.76	1.08	.003	.18	7.05	.004	N, I N, E E, I	.669 .024 .004
6	Error in assigning significance, declarative knowledge failures	0.48	0.82	0.32	0.57	.003	.18	3.17	.059	N, I N, E E, I	.552 .048 .298

Table 4. Accuracy

Measure	Novice	Intermediate	Expert	Chi-Square	P Value
Accuracy before clinical history					
Correct specific diagnoses/ Total (%)	1 / 40 (2.5 %)	16 / 40 (40%)	25 / 32 (78.1%)	26.32	<.0001
Correct categoric diagnoses/ Total (%)	9 / 40 (22.5%)	21 / 40 (52.5%)	31 / 32 (96.9%)	23.32	<.0001
Number diagnoses changed by history/ Total (%)	20 / 40 (50%)	9 / 40 (22.5 %)	2 / 32 (6.3 %)	13.33	.0013
Accuracy after clinical history					
Correct Specific Diagnoses/ Total (%)	4 / 40 (10 %)	18 / 40 (45 %)	26 / 32 (81.3 %)	40.39	<.0001
Correct categoric diagnoses/ Total (%)	14 / 40 (35 %)	23 / 40 (57.5 %)	31 / 32 (96.9 %)	17.93	.0001

Table 5. Difficulty and Certainty Ratings

Measure	Novice		Intermediate		Expert		F Value	P Value
	Mean	SD	Mean	SD	Mean	SD		
Difficulty Score	5.8	1.1	4.7	2.1	1.9	1.8	50.08	<.001
Certainty Score	4.2	1.3	6.7	1.2	9.5	0.5	11.95	<.001

Table6

Protocol measures										
Measure	Novice		Intermediate		Expert		F Value	P Value	ANOVA	
	Mean	SD	Mean	SD	Mean	SD			Pairwise Comparison (Tukey HSD)	P Value
Protocol lines	63.6	16.5	78.0	27.9	44.6	14.2	5.7	.009	N, I N, E E, I	.285 .154 .006
Data examination										
Identification	14.1	5.4	20.5	8.2	8.0	2.8	9.4	.001	N, I N, E E, I	.065 .110 .001
Comparison	2.2	1.2	3.3	1.5	0.93	0.66	8.7	.001	N, I N, E E, I	.132 .077 .001
History	1.4	0.5	1.7	0.7	1.2	0.8	1.7	.210	N, I N, E E, I	.511 .741 .191
Data exploration and explanation	2.6	1.6	4.5	2.4	2.0	1.6	4.4	.023	N, I N, E E, I	.081 .794 .027
Data interpretation	4.9	1.6	11.4	4.1	8.4	3.6	9.8	.001	N, I N, E E, I	.000 .084 .153
Control Processes	0.4	0.5	0.9	1.3	0.4	0.4	0.8	.454	N, I N, E E, I	.500 1.00 .557
Operational processes	2.6	2.1	4.4	1.8	1.5	1.6	5.4	.011	N, I N, E E, I	.119 .408 .009
Goal-setting	0.4	0.4	1.7	1.3	1.8	2.0	3.3	.052	N, I N, E E, I	.094 .083 .981
Unique hypotheses	2.2	0.6	4.0	0.9	3.3	1.2	11.3	<.001	N, I N, E E, I	<.001 .031 .210

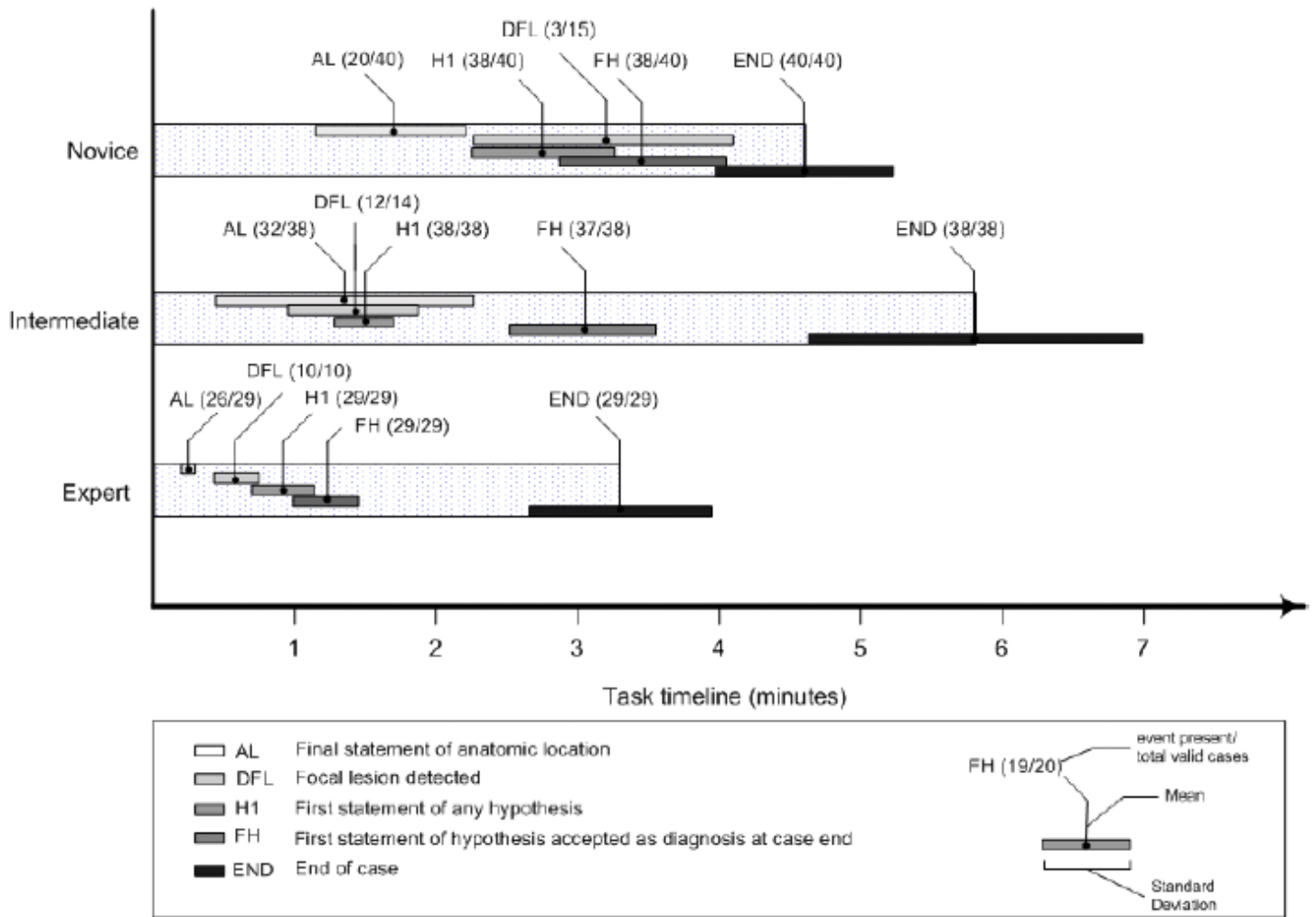


Figure 1. Task timeline depicting latencies and event counts