

令和6年度厚生労働科学研究費補助金（障害者政策総合研究事業）
発達障害への地域支援に資する継続的な情報収集・活用方法・体制整備に向けた研究

分担研究報告書

臨床データ等を活用した情報システム構築に関する研究

研究分担者 浅尾 高行（信州大学学術研究・産学官連携推進機構 特任教授）

研究要旨

「目的」多忙な業務の中で無理なく入力でき安価で安全なデータベースを目指して、発達障害関連の診療、相談支援、教育等のリアルワールドデータをデータベース化し活用するための情報システムを検討した。

「方法」専門外来における診療情報調査を行い、データベース入力負担を少なくする登録方法として手書き問診票のOCR読み取りによるデータベース化を検討した。

「結果」手書きチェックボックスの正解認識率は100%を達成した。また手書き数字の認識でも画像認識AIを用いた場合、正解率100%が認められた。電子カルテ端末のPCのスペックでも実用レベルの速度で稼働した。

「結語」安価な汎用スキャナーでOCR問診票の自動集計・スコア計算とデータ登録を同時に達成できるデータ集積方法は、レジストリーの入力の手間を省く登録方法として期待できる。今後、研究参加施設での臨床検証を進める計画である。

A、研究目的

これまでの登録中止となった疾患データベースプロジェクトの報告から、データの入力が業務負担になるような登録システムでは、継続的運用が困難なことが示唆された。本プロジェクトでは多忙な業務の中で無理なく入力でき、コスト面でも安価で安全なデータベースを目指して、発達障害関連疾患の診療、相談支援、教育等のリアルワールドデータをデータベース化し活用するための情報システム構築について検討した。

B、研究方法

1) 専門外来における診療情報調査

信州大学附属病院の発達障害の専門外来において診断・診療に利用されている情報を調査し、データベース化すべき項目を調査した。

2) 手書き問診票のOCR読み取りによるデータベース化の検討

専門外来の初診時間問診票を参考に104項目のOCR用のチェックボックス形式の問診票を作成し、汎用スキャナーによるデータベース化の可能性を検証した。プログラミング言語Pythonによる画像認識とAI手書き数字認識システムを組み合わせ、OCRによる自動データ登録システムを試作し検証した。市販のドキュメントスキャナー（スキャンSnap）

iX1600、RICOH 社製)で認識判別精度を確保するために以下の工夫を盛り込んだプログラムを試験実装した。

- ・問診票の定位置に四角の枠を配置し Python の物体認識ライブラリーを用いて位置座標を取得
- ・読み取った画像上の座標を用いて左右方向のスキンのずれを補正するアルゴリズム
- ・患者 ID 情報を入れたバーコードから患者IDの取得方法
- ・問診票自動作成プログラムからのチェックボックス位置情報によりすべてのチェックボックス部分の画像を取りだす手法
- ・手書きチェックの有無の判別アルゴリズム
- ・問診票に印刷されたページ番号と日付を機械学習の文字認識プログラムでテキスト取得
- ・手書きチェックの有無をメタデータとしてデータベースに保存する機能
- ・問診票の項目、質問テキスト、選択肢テキストの表データから半自動で問診票を作成するプログラム
- ・手書き数字を認識するAI(ViT)判別モジュール
- ・VAS のスケールのチェック位置の画像処理による数値化

3) 電子カルテへの入力支援システムの検討

OCR で取得した問診情報と各種スコアの計算結果をデータベース化と同時に電子カルテへの記載を支援するプログラムを作成し、電子カルテ操作の省力化効果を検討した。また、患者ごとの問診票を印刷する操作の自動化を行った。

- ・14種類のスコアの自動計算アルゴリズム
- ・カルテに自動で記載する Format の作成

・問診票のスコアをクリップボード経由で電子カルテに転記する支援システム

・RPA(PC 端末を自動操作)による問診票などの印刷出力の自動化

4) 模擬患者データを用いた読みとり精度試験

RPA の自動化システムを用いて模擬患者の問診票などの初診時に必要な書類を印刷し、サンプル問診票をスキャンし読みとり精度を検証した。

C、研究結果

1)臨床現場での診療情報の特性

- ・発達障害の診断には、問診を中心とした情報が重要であり、その量は初診時の問診票だけでも100項目以上となり登録データ入力の負担が大きいことが想定された。
- ・病理結果や採血の結果で診断が確定する疾患群と異なり、発達障害の診療は文章を中心とした非構造化データに依存していた。
- ・問診結果から10種類以上のスコアが利用されているが、計算機を用いた手計算により集計されており、診療開始時間前の医師の作業負担となっていた。
- ・施設ごとに使用している問診票のバリエーションがあることが示唆された。

2)OCR 登録の試行結果

- ・項目と選択肢のリストから自動で手書きチェックリストの印刷原稿を作成するシステムで各種の問診票を短時間で作成できた。
- ・チェックボックスのみからなるテストデータを用いたスキャンから集計までの速度は電子カルテ端末(CPU i5, 4Mb Memory)でも1例あたり20数秒で実用範囲内であった。

・記載鉛筆の濃さとチェック認識率の検討では、B以上の濃さの鉛筆であれば100%の認識結果が得られた

・手書き数値の認識率と認識速度の検討では、機械学習(Light_GBM)とAI(ViT)との比較では、95% : 100%でAIの手書き数字認識率は満足できるものであった。一方電子カルテのスペックのPCでの文字認識に要する時間は、機械学習で0.1秒/字に対してAIでは1秒/字かかっていた。

D、考察

本研究領域の診療、相談支援、教育等で発生する多数の項目からなる診療情報は、構造化データとしての登録がむづかしいという特徴が明らかとなった。多くの先行する同じタイプの疾患データベースにおいて、特定の病院からの登録に偏り、参加施設の拡張ができずに登録中止になった例が報告されている。継続ができない理由として、入力の手間、維持コストの高さ、有用な情報を得るのに時間とコストがかかるなどが考察されている。発達障害を対象としたレジストリー登録は、疾患の種類や情報となる項目が多く入力に相当の労力が想定されるため、新たな無理のない登録手法が求められる。

今回の検討でOCR問診票の自動集計・スコア計算とデータ登録を同時に達成できるデータ集積方法は、レジストリー構築の問題を解決する方法として期待できる。今後、研究参加施設での臨床検証を進める計画である。