

手書き帳票からの情報抽出に対するローカルLLMの適用可能性評価

佐々木風太[†] 陳思楠[†] 佐伯幸郎^{††} 中村匡秀[†]

[†] 神戸大学 〒657-8501 神戸市灘区六甲台町 1-1

^{††} 高知工科大学 〒782-8502 高知県香美市土佐山田町宮ノ口 185

E-mail: [†]safu@es4.eedept.kobe-u.ac.jp, ^{††}chensinan@gold.kobe-u.ac.jp, ^{†††}saiki.sachio@kochi-tech.ac.jp,
^{†††}masa-n@cmds.kobe-u.ac.jp

あらまし 救急医療で用いられる傷病者情報申送書は、1件あたり約20分の手作業による転記を要し、出動件数の増加に伴い事務負担が増大している。この課題に対し、先行研究では、クラウドベースのマルチモーダルLLMを活用した高精度な読み取りシステムを構築し、実用レベルの精度を達成した。ところが、クラウドLLMは個人情報を外部サーバーへ送信する必要があるため、個人情報保護法や組織のセキュリティポリシーの観点から現場への導入には障壁があった。そこで本研究では、データを外部送信せずローカル環境で完結する3つのLLM(Qwen3-VL:32b、gemma3:27b、mistral-small3.2:24b)について、手書き文字読み取り精度と構造化データ出力の観点から適用可能性を評価した。その結果、Qwen3-VL:32bが他の2モデルを大きく上回る性能を示した。詳細評価では、RTX 5090環境において28項目中26項目で完全一致(正解率92.9%)を達成し、1枚あたり平均2分15秒での処理が可能であることを確認した。これにより、セキュリティ要件の厳しい組織においても、ローカルLLMによる実用可能な手書き帳票デジタル化手段が実証された。

キーワード マルチモーダルLLM, ローカルLLM, OCR, 情報抽出, オンプレミス, セキュリティ

Evaluation of Applicability of Local LLMs for Information Extraction from Handwritten Forms

Futa SASAKI[†], Sinan CHEN[†], Sachio SAIKI^{††}, and Masahide NAKAMURA[†]

[†] Kobe University Rokkodai-cho 1-1, Nada-ku, Kobe, Hyogo 657-8501 Japan

^{††} Kochi University of Technology 185 Miyakouchi, Tosayamada, Kami City, Kochi 782-8502, JAPAN

E-mail: [†]safu@es4.eedept.kobe-u.ac.jp, ^{††}chensinan@gold.kobe-u.ac.jp, ^{†††}saiki.sachio@kochi-tech.ac.jp,
^{†††}masa-n@cmds.kobe-u.ac.jp

Abstract Handwritten patient information handover forms created in emergency medical settings contain crucial medical information for continuous patient care. Currently, these documents are manually transcribed into systems, requiring approximately 20 minutes per case and consequently increasing administrative burden for organizations. To address this challenge, previous research developed a high-accuracy reading system using cloud-based multimodal Large Language Models (LLMs), achieving practical-level accuracy. However, cloud-based LLMs require transmitting confidential information to external servers, thereby creating barriers to field implementation due to personal information protection laws and organizational security policies. Therefore, this study investigates information extraction methods that function entirely within local environments without external data transmission. First, we evaluated three local LLMs (Qwen3-VL:32b, gemma3:27b, mistral-small3.2:24b) from the perspectives of handwritten character recognition accuracy and structured data output. As a result, Qwen3-VL:32b significantly outperformed other models. Subsequently, in detailed evaluation using an NVIDIA GeForce RTX 5090 environment, Qwen3-VL:32b achieved 26 out of 28 items perfectly matched (accuracy of 92.9%) with an average processing time of 2 minutes and 15 seconds per page. These results demonstrate the practical feasibility of handwritten form digitization using local LLMs in environments with strict security requirements.

Key words Multimodal LLM, Local LLM, OCR, Information Extraction, On-premise, Security

1. はじめに

救急医療、行政手続き、医療機関における診療記録など、様々な業務領域において手書きで作成された帳票が依然として利用されている [1]~[3]。これらの帳票に記載された情報を電子システムへ転記する作業は、多くの場合手作業で行われており、作業時間の増大やヒューマンエラーの発生といった課題を抱えている。従来、こうした課題への対策として光学文字認識 (OCR) 技術が用いられてきたが、手書き文字の読み取り精度や非定型フォーマットへの対応において限界があった。

近年、マルチモーダル大規模言語モデル (LLM) の登場により、手書き帳票からの情報抽出技術は大きく進展している。先行研究 [4],[5] では、救急医療で用いられる傷病者情報申込書を対象に、マルチモーダル LLM を活用した高精度な読み取りシステムを構築した。

しかしながら、クラウドベースの LLM は処理の過程で画像データを外部サーバーへ送信する必要がある。手書き帳票には氏名、病歴、住所といった「要配慮個人情報」が含まれることが多く、こうした情報を外部サーバーへ送信することは個人情報保護法や組織のセキュリティポリシーの観点から制限されるケースが多い。例えば、神戸市消防局では神戸市における AI の活用等に関する条例により、神戸市情報公開条例第 10 条各号に定める非公開情報を入力することが禁止されている。そのため、どれほど精度が高いシステムであっても、クラウドサービスを利用する限り現場への導入は困難である。

そこで本研究では、セキュリティ要件により外部送信が制限される環境において、データを外部に送信することなくローカル環境 (オンプレミス) で完結する手書き帳票情報抽出手段を検討する。まず、ローカル環境で動作する 3 つのマルチモーダル LLM (Qwen3-VL:32b [8], gemma3:27b [9], mistral-small3.2:24b [10]) について、手書き文字読み取り精度と構造化データ出力の観点から適用可能性を評価した。その結果、Qwen3-VL:32b が最も優れた性能を示した。次に、Qwen3-VL:32b を対象として、NVIDIA GeForce RTX 5090 環境における項目抽出精度と処理速度について定量的な評価実験を実施した。その結果、28 項目中 26 項目で完全一致 (正解率 92.9%) を達成し、1 枚あたり平均 2 分 15 秒での処理が可能であることを確認した。これにより、ローカル LLM による手書き帳票デジタル化の実用可能性が示された。

2. 準備

2.1 手書き帳票のデジタル化の必要性

近年、デジタルトランスフォーメーション (DX) による業務効率化やデータ利活用の推進が求められており [6]、多くの組織で業務プロセスのデジタル化が進められている。しかしながら、一部では紙媒体の帳票を前提とした運用が依然として存在する。

本研究の協力機関である神戸市消防局を例にとると、この組織では傷病者情報申込書と呼ばれる帳票を扱っている。現在、この文書は手書きで作成され、そのち手作業でシステムへ転

記入力が行われている。1 件あたり約 20 分を要するこの作業は、出動件数の増加に伴い組織全体の事務負担を増大させている。類似の課題は、医療機関 [1]、行政機関 [2]、企業 [3] など手書き帳票を扱う多様な組織においても存在する。

このような背景から、手書き帳票から情報を利用しやすい形で抽出できる仕組みが求められている。

2.2 手書き帳票からの情報抽出の要件

傷病者情報申込書に限らず、帳票のデジタル化において求められる要件は、単に画像内の文字をテキストデータ化する (OCR) だけでは不十分である。業務システムに取り込むためには、帳票内の「項目 (Key)」とそこに記載された「入力値 (Value)」を正しく対応付け、構造化されたデータとして抽出する必要がある。例えば、「氏名」欄に書かれた手書き文字を認識し、それを単なる文字列としてではなく、「氏名: "神戸 太郎"」のように意味的なラベルと紐付けて出力しなければならない。特に傷病者情報申込書は、定型の枠内だけでなく、フリースペースへの記述やチェックボックス、丸囲みなど多様な記入形式を含んでいる。また、現場での活動中に記載されるため、文字の乱れや枠からはみ出しも頻繁に見られる。これらの特徴を持つ手書き帳票から、項目と値を正確に 1 対 1 で対応付けて抽出することは、従来のルールベース型 OCR では極めて困難な課題であった。

2.3 先行研究：傷病者情報申込書読み取りシステム

先行研究 [4],[5] では、これらの課題に対し、マルチモーダル LLM を活用した「傷病者情報申込書読み取りシステム」を開発した。このシステムの基本機能は、ユーザーが傷病者情報申込書の画像をアップロードするだけで、そこに記載されている多様な情報を自動的にデジタル化できる点にある。システム内部では、Google の Gemini モデル [7] が画像の視覚的特徴 (レイアウトや罫線) と言語的情報 (文字の意味) を同時に解析し、書類内の「どの欄に」「何が書かれているか」を特定する。そして、抽出した情報を、「出動番号」「傷病者氏名」「搬送先病院」といったシステム処理に適した構造化データの形式 (JSON 等) で出力する。これにより、従来は人間が傷病者情報申込書を目視しながら手作業で行っていたシステム入力業務を、画像のアップロードという単純作業のみで代替可能にしたものである。図 1 に、このシステムの画面を示す。



図 1: 傷病者情報申込書読み取りシステムの画面

2.4 システムの課題

先行研究により技術的な実現可能性は示されたものの、実運

用に向けては「セキュリティ」が決定的な課題として残されている。先行研究のシステムはクラウド上の API を利用する構成であるため、入力された傷病者情報申込書の画像データは必ず外部のサーバーへ送信される。手書き帳票には氏名、病歴、住所といった「要配慮個人情報」が含まれることがある。そして、こうした情報を外部サーバーへ送信することは、個人情報保護法や組織のセキュリティポリシーの観点から制限されるケースが多い。例えば、神戸市消防局では神戸市における AI の活用等に関する条例により、神戸市情報公開条例第 10 条各号に定める非公開情報を入力することが禁止されている。したがって、クラウドサービスを利用前提とする限り、どれほど精度が高くても現場への導入は困難である。

2.5 ローカル LLM

この課題を根本から解決する手段として、データを組織外に出さない「ローカル環境 (オンプレミス)」でのシステム運用が挙げられる。そこで本研究では、外部送信をせず、かつ手書き帳票からの高精度な情報抽出を実現しうるモデルの選定を行った。選定にあたっては、以下の基準を設定した：(1) マルチモーダル機能を有し、画像を直接入力可能であること (2) ビジョン-言語タスクにおいて高い性能を示すこと (3) RTX 5090 搭載の検証環境で実行可能なモデルサイズであること

これらの条件を満たすモデルとして、Qwen3-VL:32b [8], gemma3:27b [9], mistral-small3.2:24b [10],[11] の 3 つを選定した。

(1) **Qwen3-VL:32b**: Qwen ファミリーの中で最も強力な視覚言語モデルである。テキスト理解と生成、視覚コンテンツの認識と推論において大幅な改善が施されており、特に OCR 処理能力は 32 言語に対応し、暗い場所、ぼやけた文字、傾いた文字など実際の使用環境において信頼性の高い性能を発揮する。また、長文文書の理解能力に優れ、最大 256K トークンのコンテキストをサポートする。

(2) **gemma3:27b**: Google が Gemini テクノロジーを基盤として開発した軽量マルチモーダルモデルである。128K トークンのコンテキストウィンドウと 140 以上の言語サポートを備え、質問応答、要約、推論といったタスクに優れる。

(3) **mistral-small3.2:24b**: Mistral AI が開発したビジョン対応モデルである。ビジョン理解機能により画像を分析し、視覚的なコンテンツに基づいて洞察を提供できる。128K トークンのコンテキストウィンドウを持ち、関数呼び出しや JSON 出力といったエージェント機能に優れている。数十の言語をサポートし、高度な推論能力を有する。

これら 3 つのモデルについて、手書き帳票からの情報抽出と JSON 形式での構造化出力の観点から適用可能性を評価し、ローカル環境において最も適切なモデルを選定する。

3. ローカル LLM の適用可能性評価

3.1 評価目的

本節では、前述した 3 つのローカル LLM (Qwen3-VL:32b, gemma3:27b, mistral-small3.2:24b) の適用可能性を評価する。

本評価実験では、傷病者情報申込書のみを対象とした。傷病

者情報申込書は、定型の枠内記入、フリースペースへの記述、チェックボックス、丸囲みなど多様な記入形式を含む複雑な構造を持つ書類である。このような複雑な特性を持つ書類でローカル LLM の性能を検証することで、他の手書き帳票への適用可能性も含めた評価が可能となる。なお、本評価で使用される傷病者情報申込書は、消防局で使用されている様式に準拠して作成したサンプルであり、架空の傷病者情報を記載したものである。実際に使用された文書ではない。評価の観点は以下の 2 つである。

(1) **手書き文字読み取り精度**: 手書き文字を正確に認識できるか。

(2) **構造化データ出力**: プロンプトで指示した JSON 形式で出力され、かつ書類の内容が正しく抽出されるか。

3.2 検証環境

評価実験は、以下の検証環境で行った。

- **GPU**: NVIDIA GeForce RTX 5090
- **推論フレームワーク**: Ollama
- **評価対象モデル**:
 - Qwen3-VL:32b
 - gemma3:27b
 - mistral-small-3.2:24b
- **量子化**: なし

3.3 評価 1: 手書き文字読み取り精度

3.3.1 評価方法

3 つのローカル LLM の基礎的な文字認識能力を測るため、傷病者情報申込書内の文章領域を切り出し、テキスト化させた。評価で使用した切り出し画像を図 2 に示す。画像の左から順に画像 1、画像 2、画像 3 とする。

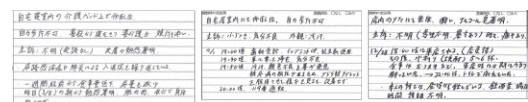


図 2: 文字読み取り精度評価に使用した切り出し画像 (左から画像 1, 画像 2, 画像 3)

各モデルに対して、以下のプロンプトを用いて文字の読み取りを指示した。

あなたは高精度な OCR (光学文字認識) エンジンとして振る舞ってください。アップロードされた画像のテキストを、一字一句正確に書き起こしてください。

評価指標として、文字誤り率 (Character Error Rate: CER) を使用した。CER は、正解テキストに対する置換・削除・挿入のエラー数の割合を示す指標であり、値が小さいほど認識精度が高いことを意味する。

3.3.2 評価結果

3 つの画像それぞれに対する各モデルの CER を表 1 に示す。

表 1: 3つのローカル LLM の手書き文字読み取り精度の比較 (CER)

モデル	画像 1	画像 2	画像 3
Qwen3-VL:32b	0.1212	0.0641	0.2125
gemma3:27b	0.6288	0.2756	0.5375
mistral-small3.2:24b	0.5000	0.2949	0.7188

結果より、Qwen3-VL:32b は全ての画像において他の 2 つのモデルを大きく上回る認識精度を示した。特に画像 2 においては、CER **0.0641** を記録し、100 文字中誤読が 6 文字程度に抑えられており、十分な実用性がある。

一方、gemma3:27b と mistral-small3.2:24b は手書きの日本語文字の認識にかなり苦戦していることが確認された。ただし、数字については比較的良い精度を示しており、数字の記載が多い画像 2 においては他の画像と比較して CER が低めになっていた。これは、これらのモデルが英数字には一定の認識能力を持つものの、複雑な日本語の手書き文字に対しては十分な学習がなされていないことを示唆している。

3.4 評価 2: 構造化データ出力の評価

3.4.1 評価方法

次に、3 つのローカル LLM に対して傷病者情報申込書の全体画像を入力し、構造化データとして情報を抽出させた。評価に使用した傷病者情報申込書の画像を図 3 に示す。

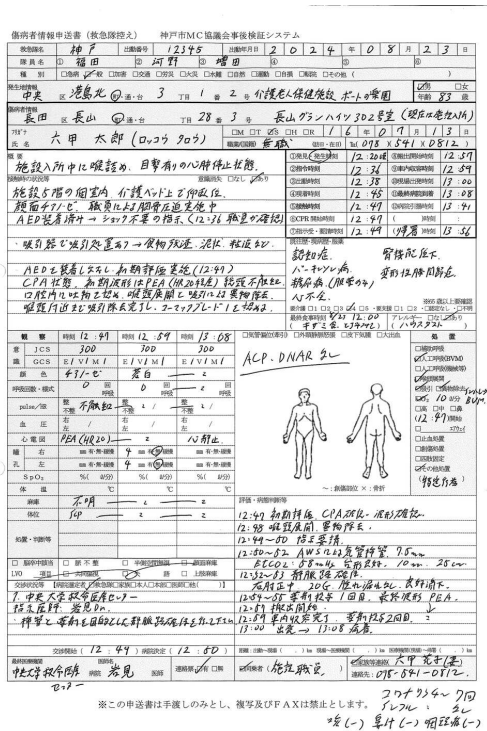


図 3: 評価に使用した傷病者情報申込書

各モデルに対して、先行研究で構築したシステムと同様のプロンプトを用いて項目抽出を指示した。評価基準は以下の 2 点である。

- プロンプトで指示した JSON 形式で出力されているか
- 書類の内容が正しく抽出されているか

3.4.2 評価結果

各モデルの出力結果を分析したところ、以下のような特徴が確認された。

Qwen3-VL:32b の出力は、指定した JSON 形式を正確に遵守し、かつ書類の内容を適切に抽出することができた。以下に出力例の一部を示す。

```
[
  {
    "救急隊名": "神戸",
    "出勤番号": "12345",
    "出勤年月日": "2025年03月09日",
    "隊員名": "高田, 上山, 藤井",
    "種別": "01_急病",
    "発生地情報": "中央区 港島上町 7丁目 1番 6号
      港島ニューホートレジデンス 2302号
      室",
    "傷病者情報": {
      "氏名": "長田 鉄太郎",
      "生年月日": {
        "和暦": "04_H",
        "年月日": "1990.1.1"
      },
      "年齢": "83",
      "性別": "01_男性",
      "住所": "中央区 港島上町 7丁目 1番 6号
        港島ニューホートレジデンス 2302号室",
      "職業": "無職",
      "TEL": "(090)(8827)(4094)",
    },
    "概要": "自宅にて意識レベル低下、発熱、SpO2値低下",
    (以下省略)
  }
]
```

この出力例からわかるように、Qwen3-VL:32b は項目名と値の対応関係が正しく認識され、実用的な構造化データが得られた。

gemma3:27b の出力は、JSON 形式での出力は可能であったものの、ほとんどの項目で値を抽出することができなかった。以下に出力例を示す。

```
[
  {
    "救急隊名": null,
    "出勤番号": "12345",
    "出勤年月日": null,
    "隊員名": null,
    "種別": "null",
    "発生地情報": "中央区 港島上町 7丁目 1番 6号
      港島ニューホートレジデンス 2302号室",
    "傷病者情報": {
      "氏名": null,
      "生年月日": {
        "和暦": null,
      }
    }
  }
]
```

```

    "年月日": null
  },
  "年齢": null,
  "性別": null,
  "住所": "中央区 港島上町 7丁目 1番 6号
           港島ニュートレジデンス 2302 号室",
  "職業": null,
  "TEL": "(090)(8827)(4094)"
},
"概要": null,
"接触時の状況等": null,
(以下省略)

```

この出力例からわかるように、gemma3:27b は JSON 構造自体は正しく生成できているものの、ほとんどの項目が 'null' になっており、実際の書類から情報を読み取ることができていない。抽出された項目は「出勤番号」「発生地情報」「TEL」などごく一部に限られており、これらはプロンプト内に記述例として明示的に示していた項目であった。

しかしながら、本モデルの不調が単なる認識能力の限界によるものか、あるいは複雑なスキーマ定義を含むプロンプトへの追従負荷に起因するものかは検討の余地がある。そこで、構造化の制約を緩和し、単に「書類から情報を抽出せよ」という平易な指示を与えた場合の挙動を検証した。得られた出力結果を以下に示す。

```

{
  "書類の種類": "傷害補償請求書(救急医療控)",
  "受付番号": "12345",
  "提出年月日": "2023年3月9日",
  "被災者情報": {
    "氏名": "長田 鉄郎",
    "フリガナ": "オダ テツロウ",
    "性別": "男性",
    "年齢": "83",
    "電話番号": "090-8287-9094",
    "住所": "滋賀県蒲生郡竜王町伊勢 570-2",
    "続柄": "本人"
  },
  "事故状況": {
    "事故の状況": "自転車転倒",
    "負傷部位": "左足首骨折、打撲、擦過",
    "治療の状況": "入院",
    "緊急性": "あり"
  },
  "事故発生日時": {
    "日時": "2023年2月16日 14時30分",
    "場所": "滋賀県蒲生郡竜王町伊勢",
    "状況": "近隣の畑の道で自転車走行中、
             前方不明の段差に引っかかり転倒"
  },
  (以下省略)
}

```

この結果には依然として多くの誤りが散見されるものの、手書き文字からの情報抽出そのものは試みられていることが確認できる。この事実は、先の実験における抽出失敗が、モデルの認識能力の欠如のみならず、複雑なプロンプト指示（特に厳格な JSON スキーマへの準拠）への追従負荷に起因していた可能性を強く示唆している。

mistral-small3.2:24b の出力も同様に、JSON 形式での出力は可能であったものの、書類の内容をほぼ抽出できなかった。gemma3:27b と同様に、ほとんどの項目が 'null' または空文字列となっており、手書き文字認識能力の不足により実用的な情報抽出ができないことが確認された。

3.4.3 考察

以上の結果から、傷病者情報申込書の読み取りにおいて、ローカル環境で実用可能なモデルは **Qwen3-VL:32b のみ** であると結論付けた。gemma3:27b と mistral-small3.2:24b は、プロンプトで指示した構造化出力の形式は理解できるものの、本研究で要求される複雑なプロンプト指示に従った高精度な情報抽出を実現できず、傷病者情報申込書の読み取りにおいて実用的な性能を示さなかった。

次節以降では、この Qwen3-VL:32b を用いた際の項目抽出精度と処理速度について詳細な検証を行う。

4. Qwen3-VL:32b の詳細評価

4.1 実験設定

前節で選定された Qwen3-VL:32b について、その情報抽出能力を定量的に検証する。実験は、前節と同じ検証環境 (RTX 5090, Ollama) で実施した。

4.1.1 使用データ

評価用データとして、傷病者情報申込書のサンプル画像 3 枚 (画像 1~画像 3) を用意した。これらは消防局で使用されている様式に準拠した模擬データであり、全て手書き文字によって記述されている。各画像には、架空の傷病者情報や活動内容が記録されている。使用した 3 枚のサンプル画像を図 4 に示す。各画像について、人手による正確な書き起こしを行い、これを正解データ (Ground Truth) とした。

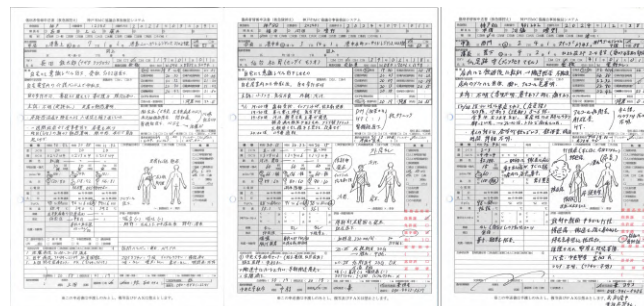


図 4: 実験に使用した傷病者情報申込書のサンプル (左から画像 1, 画像 2, 画像 3)

4.2 項目抽出精度の評価

実際のシステム運用を想定し、傷病者情報申込書の画像全体を入力として、システムに必要な重要項目 (28 項目) を JSON

形式で抽出させた。対象項目には、「救急隊名」「出動番号」「傷病者氏名」「生年月日」「年齢」「性別」「発生場所」「主訴」「各種バイタルサイン」などが含まれる。図5に、評価対象項目の具体的な位置を示す。評価指標には、**項目正解率 (Field Accuracy)** を用いた。これは、抽出された項目の値が、正解データと完全に一致 (完全一致) している項目の割合である。些細な文字等の違いも誤りとして扱う厳格な評価である。

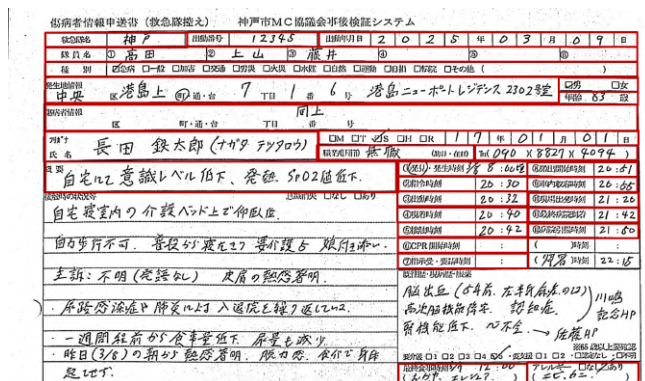


図 5: 評価対象項目の位置

結果を表2に示す。

表 2: 項目抽出精度の結果

画像 ID	正解数 / 全 28 項目	正解率
画像 1	23	82.1%
画像 2	26	92.9%
画像 3	20	71.4%

文字が整っていた「画像2」では、28項目中26項目で完全一致し、正解率は**92.9%**に達した。抽出に失敗した2項目についても、人間が見ても判読が難しい、あるいは枠線をはみ出して記述されていた箇所であった。乱雑な文字を含んでいた「画像3」においても71.4%の精度を維持しており、最低限の情報抽出は可能であることが示された。

4.3 処理速度の評価

実運用におけるスループットを確認するため、RTX 5090 環境において、画像の読み込みから JSON データが出力されるまでの処理時間 (推論時間) を計測した。計測結果を表3に示す。

表 3: 処理速度の評価結果

画像 ID	処理時間 (秒)
画像 1	176.21
画像 2	88.00
画像 3	139.91
平均	134.71

3枚の画像の平均処理時間は**134.71秒 (約2分15秒)**であった。事後的なデータ入力業務においては、1件あたり十数秒といったリアルタイム性は必ずしも求められない。隊員が帰署した後にまとめてバッチ処理を行う運用を想定すれば、この処理時間は十分に実用範囲内であるといえる。

5. 結 論

本研究では、3つのローカル LLM (Qwen3-VL:32b, gemma3:27b, mistral-small3.2:24b) の適用可能性を評価した結果、Qwen3-VL:32b がローカル環境において傷病者情報申送書から高精度な情報抽出を実現できることが明らかとなった。

Qwen3-VL:32b は、RTX 5090 環境において28項目中26項目で完全一致 (正解率92.9%) を達成し、1枚あたり平均2分15秒での処理が可能であることが確認された。これにより、データを外部サーバーへ送信することなく、個人情報保護法や組織のセキュリティポリシーに準拠した形で、手書き帳票のデジタル化が可能であることが示された。

6. ま と め

本研究では、セキュリティ要件により外部送信が制限される環境を対象に、ローカル環境で完結する手書き帳票情報抽出手段を検討した。先行研究で構築したクラウドベースのシステムは高精度を実現したものの、個人情報保護法や組織のセキュリティポリシーの観点から、現場への導入には障壁があった。

3つのローカル LLM (Qwen3-VL:32b, gemma3:27b, mistral-small3.2:24b) の適用可能性を評価した結果、手書き文字読み取り精度と構造化データ出力の両面で Qwen3-VL:32b が最も優れた性能を示した。詳細評価では、RTX 5090 環境において正解率92.9%、処理時間平均2分15秒を達成し、バッチ処理を前提とする運用においては十分な実用性を有することが示された。

本研究により、ローカル LLM (Qwen3-VL:32b) を用いることで、データの外部送信を行わない安全な環境下においても、傷病者情報申送書の高精度な情報抽出が実現可能であることが実証された。これにより、消防局をはじめとする厳格なセキュリティポリシーを持つ組織においても、マルチモーダル LLM 技術を活用した帳票デジタル化の道が開かれた。今後は、より多様な帳票への適用や、実運用に向けた検証を進めていく必要がある。

謝辞 本研究の一部は JSPS 科研費 JP25H01167, JP25K02946, JP25K24389, JP24K02765, JP24K02774, JP23K17006, JP23K28091, JP23K28383 の研究助成を受けて行われている。本研究は神戸市消防局との共同研究の一環で行われている。

文 献

- [1] “紙カルテ利用の診療所の電子化対応可能性に関する調査 (設問は全3問)にご協力を!” <https://www.med.or.jp/nichii/online/article/012198.html>, accessed January 28, 2026.
- [2] “行政手続のオンライン化,” https://www.digital.go.jp/policies/administrative_procedures_online, accessed January 28, 2026.
- [3] “アドビ、「バックオフィス業務のデジタル化に関する調査」の結果を発表,” https://www.adobe.com/jp/news-room/news/202308/20230807_back-office-survey.html, accessed January 28, 2026.
- [4] 佐々木風太, 陳思楠, 中村匡秀, “生成 AI を活用した救急出動報告書読み取りシステムの研究開発,” 電子情報通信学会技術

研究報告, vol.124, no.334, LOIS2024-66, pp.84-89, January 2025.

- [5] 佐々木風太, 佐伯幸郎, 中村匡秀, “生成 AI を用いた手書きフォームの高精度なデータ認識・抽出手法の検討,” 電子情報通信学会技術研究報告, vol.125, no.115, SS2025-20, pp.116-121, July 2025.
- [6] “DX 動向 2024,” <https://www.ipa.go.jp/digital/chousa/dx-trend/eid2eo0000002cs5-att/dx-trend-2024.pdf>, accessed February 3, 2026.
- [7] “Gemini 2.5: Our most intelligent AI model,” <https://blog.google/technology/google-deepmind/gemini-model-thinking-updates-march-2025/>, accessed June 12, 2025.
- [8] “qwen3-v1,” <https://ollama.com/library/qwen3-v1>, accessed January 28, 2026.
- [9] “gemma3,” <https://ollama.com/library/gemma3>, accessed January 28, 2026.
- [10] “Mistral-Small-3.2-24B-Instruct-2506,” <https://huggingface.co/mistralai/Mistral-Small-3.2-24B-Instruct-2506>, accessed January 28, 2026.
- [11] “Mistral-Small-3.1-24B-Instruct-2503,” <https://huggingface.co/mistralai/Mistral-Small-3.1-24B-Instruct-2503>, accessed January 28, 2026.