

## 論文内容の要旨

氏名	大塚 紀朗
Diagnosing Psychiatric Disorders from History of Present Illness Using a Large-Scale Linguistic Model	
(和訳)	
大規模言語モデルを使用した現病歴からの精神疾患の診断	

### 論文内容の要旨

精神医学における診断支援を目的とした自然言語処理モデルについて報告した。AIの中でも自然言語処理は、言葉が情報収集や治療の中心となる精神科において有望視されているが、精神科医と比較した研究はまだまだ乏しいのが現状である。これまでの研究は、主要な疾患のみを対象としたり、サンプルサイズが小さかったり、精神科医との比較がなかったりと限定的だった。そこで、私たちは、精神疾患全体をカバーする診断モデルを構築し、予測結果を精神科医と比較した。

奈良県立医科大学病院から、除外基準をクリアした2007年から2021年までの2642症例の電子カルテ上の退院サマリーから現病歴と診断を抽出した。診断はICD-10の第V章のコードに従って11クラスに分類した。東京大学病院の電子カルテに事前学習されたUTH-BERTを使用して、現病歴に基づいて退院時の主診断を予測した。

モデルの一致率は74.3%で、指定医取得前の精神科医が71.5%、研修医が69.4%だった。指定医取得前の精神科医によって正しく回答された症例に限定すると、モデルと研修医の一致率はそれぞれ84.9%、83.3%だった。作成したモデルの一致率は、入局2か月の精神科医や指定医取得前の精神科医を上回り、統計的な有意差はないものの、AIの有用性が示唆された。

疾患群別F値を確認すると、F4群では、モデルのF値が精神科医のF値を比較的大きな差で上回った。F4群には不安障害、適応障害、身体表現性障害などがあり、「めまい」「頭痛」「過呼吸」などF4に比較的多く含まれる医学用語が一致率を上昇させた可能性がある。一方、アルコールや覚せい剤などの物質使用障害を指すF1群では、精神科医のF値がモデルのF値を上回った。その理由として、UTH-BERTを生み出した東京大学医学部附属病院では依存症治療を行っておらず、奈良医大と患者層が異なり、F1診断に重要な「覚せい剤」などの用語は、不明トークンとして扱われるか不適切なサブワード化をされたことが考えられる。このように本モデルにおいては、診断を示唆する単語の出現頻度やそれを認識できるかどうかは予測精度に影響したと推測できた。

また、架空の現病歴を用いてアテンションの可視化を行い、診断予測に重要と思われるトークン(単語)をより濃い色で例示した。

そして、さまざまな本研究の限界点をあげた。具体的には、疾患の分類の仕方とラベルの誤り、現病歴の質、不明トークンと不適切なサブワード化、診断の一致率を診断精度として扱う際の問題、不均衡データ、説明可能性、汎用性をあげた。この研究を通じて、機械学習モデルを実際の精神疾患の診断に用いる研究がさらに進展することを期待する。