

Nationwide AI Models for Depression Prediction: Regional Generalizability and Culturally Sensitive Variable Selection

Alissa Valentine¹, Helen Coupland², Eike Petersen³, Emily Beaman¹, Merete Osler^{1,4},
Aasa Feragen⁵, Melanie Ganz^{1,2}

¹University of Copenhagen, Denmark

²Neurobiology Research Unit at Copenhagen University Hospital (Rigshospitalet), Denmark

³Fraunhofer Institute for Digital Medicine, Germany

⁴Center for Clinical Research and Prevention at Copenhagen University Hospital (Bispebjerg/Frederiksberg Hospital)

⁵Technical University of Denmark, Denmark

alissa.valentine@di.ku.dk, ganz@di.ku.dk

Abstract

The prevalence of psychiatric disorders is increasing globally. Major depressive disorder (MDD) is the most burdensome, contributing to growing strains on healthcare systems. Artificial intelligence (AI) for healthcare is deemed a potential solution but requires large-scale, high-quality data; therefore, we use Denmark as a case study with comprehensive register data and long psychiatric wait times. Although Denmark is often considered a homogeneous society, its sociodemographic and socioeconomic factors differ across the country, including immigration rates. There are also regional differences in the prevalence of psychiatric diagnoses. However, we have yet to understand how AI models account for such differences. We examine how regional and sociodemographic inequities affect MDD diagnosis and AI-based prediction. First, we conduct an observational analysis showing disparities in MDD diagnosis odds across regions of Denmark, as well as patient sociodemographic and socioeconomic factors. Second, we develop two MDD prediction models using tabular transformer architecture that differ in how they represent heritage and immigration-related information. We assess the performance of the two predictive models across the five administrative regions in Denmark, and their generalizability when trained on data from only a single region. We observe that model performance varies across Denmark's five administrative regions, along with immigration status and heritage. Incorporating more nuanced sociodemographic and socioeconomic variables improves predictive performance across all regions and most sociodemographic groups. These findings illustrate the importance of fairness assessments in psychiatric AI and the need to account for regional, sociodemographic, and socioeconomic inequities when developing and deploying AI in high-risk healthcare settings.

Code — <https://github.com/valentinealissa/MDD-Denmark>

Introduction

Rates of psychiatric disorders are increasing worldwide. Major depressive disorder (MDD) is one of the most prevalent, with over 330 million people living with depression (Institute for Health Metrics and Evaluation 2025). Nevertheless, preventing and treating psychiatric diseases is far

from an easy task. For example, studying MDD necessitates access to high-quality and large-scale healthcare data, due to the diagnosis rate often being below its prevalence rate (Weyer et al. 2023). Hence, we selected Denmark as a case study because it offers nationwide, high-quality registry data. In Denmark, approximately 6% of the population is currently affected by depression (Dalsager et al. 2025). Compared to other mental health disorders, MDD represents the largest contribution to both Denmark's nationwide annual healthcare costs and income loss (Christensen et al. 2022b). Yet, the average wait time for primary and secondary care continues to increase (Danish Broadcasting Corporation 2024). For instance, a psychiatric appointment in Denmark has increased from five months to over two years since 2017 (Jyllands-Posten 2024). These delays in obtaining psychiatric care for MDD contribute to the burden of disease for the patient, their family, and the healthcare system (Armbrecht et al. 2021; Oluboka et al. 2017; Ghio et al. 2014).

Artificial intelligence (AI) is often framed as a tool to use clinical resources more efficiently by targeting high-risk patient groups. Within psychiatry, this includes implementing AI tools to assist with diagnosis, triage patients, or predict adverse events such as self-harm. However, existing research on these topics is often concentrated in specific regions, healthcare systems, or institutions, largely due to limitations in data access. For instance, recent work in eastern Denmark has demonstrated that AI methods have the potential to predict hospital readmission (Sasha Thorn Jakobsen et al. 2025) and suicide in psychiatric patients (Nielsen et al. 2023). Meanwhile, psychiatric research groups in western Denmark have leveraged AI to predict treatment needs (Hansen et al. 2025b), diagnostic progression (Hansen et al. 2025a), and involuntary admission (Perfalk et al. 2024). In an ideal scenario, models such as these would be deployed throughout the country to improve psychiatric outcomes in the national healthcare system. But due to restrictions on data access, most of the above models were developed locally. Given the good performance of local models as well as national healthcare records, we want to investigate how feasible it is to implement AI models for MDD risk prediction nationwide.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Whilst Denmark is often touted as a homogeneous society, its five administrative regions differ greatly in sociodemographic diversity and socioeconomic status (SES). The regions – North Jutland, Middle Jutland, South Denmark, Zealand, and the Capital Region (Figure 1A) – function similarly to counties in other countries. Immigrants make up over 20% of the population ($N = 137,000$) in Copenhagen, the capital and largest city in Denmark, compared to 11% ($N = 42,000$) in Aarhus, the second largest city (Statistics Denmark 2025d). These regional differences lead us to question the generalizability of prediction models developed in one region only. Moreover, ongoing work fails to evaluate the fairness of these models across subgroups such as region of care (i.e., the Capital Region versus North Jutland), immigration status (Danish origin versus immigrant), or heritage (e.g., West Asia versus North Europe) (Hansen et al. 2025a,b; Perfalk et al. 2024; Sasha Thorn Jakobsen et al. 2025; Nielsen et al. 2023). Fairness evaluations such as these are critical for recognizing the structural inequities within the psychiatric system that shape who receives a diagnosis, thus access to care. More specifically, MDD diagnoses are a human conceptualization and do not reflect the ground truth of who experiences depressive-like symptoms. In general, psychiatric diagnoses represent the outcomes of complex interactions between patient presentation (which is often shaped by one's heritage (Kirmayer et al. 2001)), clinician judgment, and systemic biases. We define heritage as a cultural or ethnic identity that is inherited from previous generations, and is distinct from immigration status but may be associated with nationality or region of origin.

When gathering demographic or more general data across time and data sources, variables are often not consistently defined and recorded. Beyond issues with the inhomogeneity of the existing data, current variable specification approaches mean that measures of individual heritage are often coarse, conflated, or inconsistently coded, and models inevitably inherit these limitations (Movva et al. 2023). For instance, in Denmark, ethnicity is often only characterized as either “Western” or “non-Western” (Wallach-Kildemoes et al. 2014; Knudsen et al. 2022; Stevnsborg et al. 2016), a harmful trope that perpetuates colonialist attitudes towards cultural differences (Bhopal and Donaldson 1998). In reality, research demands more granular measures of immigration status and heritage to assess how equitable the distribution of psychiatric care is (Benthall and Haynes 2019). Failing to do so leads to the development of models that learn to attribute regional diversity to poorly designed variables that categorize people based on “Western-ness,” which in turn hides differing impacts of cultural heritage versus immigration status, leading to underestimation of disparities in model performance (Movva et al. 2023).

This study aims to explore the feasibility of using models for MDD prediction across a nation and to understand how such models may perpetuate regional disparities. We first investigate the sociodemographic, socioeconomic, and regional factors that influence MDD diagnosis using a matched cohort containing patients with MDD and healthy controls from the Danish registries. Deep learning models are then used in an MDD prediction task wherein we evalu-

ate model performance and fairness across different regions of Denmark, as well as across individual immigration status and heritage. In doing so, we explore the importance of variable selection by comparing the performance of two models: one with an extended set of variables to account for immigration status versus heritage and SES, the latter with variables typically included in existing research. To assess the generalizability of registry data pertaining to MDD across the five administrative regions, models are trained on data from one region and evaluated on the remaining regions.

Our contributions revolve around the following question: **If we were to implement an MDD prediction tool nationwide in Denmark, would it perpetuate existing disparities in mental healthcare?** Specifically, we:

1. Provide an association analysis of regional differences in MDD diagnoses, controlling for the sociodemographic and socioeconomic composition of the administrative regions. This confirms earlier work suggesting lower MDD prevalence in North Jutland compared to the four other regions (Raft et al. 2026). Given this finding, we expect models to learn regional biases in MDD prediction due to prevalence differences.
2. Examine the importance of variable selection in model development. We construct two variables to account for heritage, one following Statistics Denmark's standard classification, and the other using internationally recognized groupings to provide a more nuanced global characterization. We find that using more culturally sensitive variables improves model performance.
3. Assess the generalizability of ‘regionalized’ models by training MDD prediction models within each administrative region and evaluating performance both within and across regions. We find that some regions generalize better than others, and that generalizability is not reciprocal.

While we perform our case study on Danish registry data, our findings are relevant for any healthcare system serving a population with regional differences in urbanization, immigration rates, or mental health resources.

The analysis in this paper has two main parts: an observational association analysis and an MDD prediction task using deep learning. We first outline the background literature, explain our choice of predictive models and fairness evaluation toolbox, and describe the data and cohort design. Next, we present the association analysis, followed by the experimental design and results of the deep learning approach. Finally, we conclude with a discussion on the takeaways of this work and summary of the limitations surrounding registry data and predicting MDD.

Background and Related Work

Our work builds upon the high-quality personal data available via Statistics Denmark. Upon birth or arrival in Denmark, all residents are given a unique personal identification number (CPR), which is used by the country's administrative sectors, including the Danish National Patient Registry. Denmark provides free universal healthcare, including psychiatric care, to all its residents, and the CPR number allows linkage between details of a person's interactions

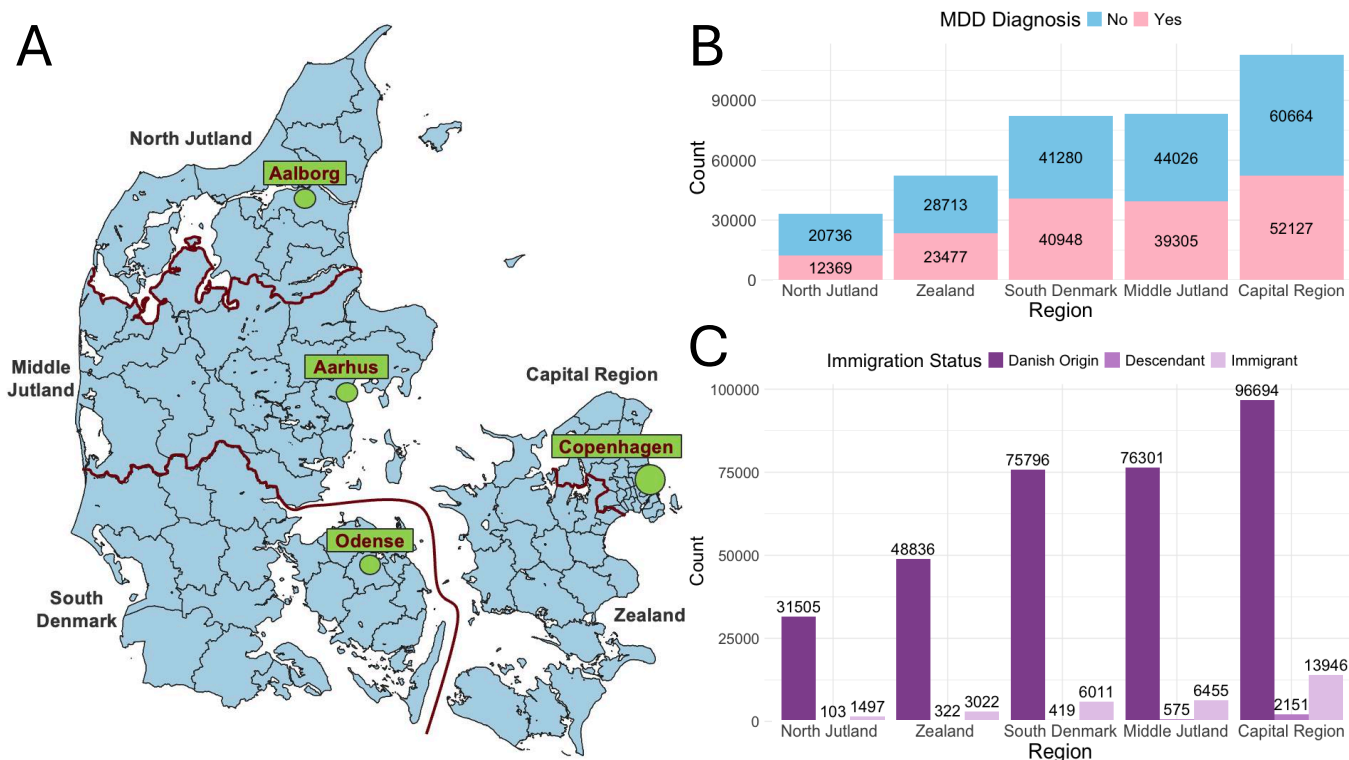


Figure 1: A. Map of the 5 Administrative Regions of Denmark (North Jutland, Middle Jutland, South Denmark, Zealand, and the Capital Region), their municipalities, and largest cities as green dots. B. MDD Prevalence per administrative region in our dataset. C. Immigration Status Counts (equaling unique cohort participants) per administrative region in our dataset.

with the hospital system and information on employment, education, immigration, and more. In this way, the Danish National Patient register is essentially electronic health records, containing routinely collected data at the hospital level, wherein data related to diagnosis is recorded and input by clinicians starting at triage and ending in discharge. All data from hospital records are governed centrally by the Danish Health Data Authority (DHDA), ensuring uniform data quality and consistency. It is worth noting that whilst the hospital records are part of the Danish National Patient Register, personal data from general practitioner (GP) practices is not yet available. Data access is strictly guarded, and all information is anonymized for research purposes. These registers contain decades of carefully curated data on almost all Danish residents, enabling the construction of comprehensive individual-level profiles for the entire population and thereby facilitating data-driven healthcare in psychiatry (Schmidt et al. 2015; Plana-Ripoll et al. 2025).

Regional Differences

Each of the five administrative regions are responsible for its own hospitals, specialized care, and psychiatric facilities. Despite this, there is very little variability in data collection within and between regions due to DHDA regulations. Each administrative region is further split into several municipalities (Figure 1A). This is similar to, for example, how the United States has states made up of counties, or China has

provinces made up of prefectures. Municipalities are responsible for local level care and facilities, such as care workers and social services, which does not impact hospital records.

Although there are similar data collection practices across the five regions, there is limited existing literature examining regional variation in general health and psychiatric care within Denmark. Previous research has described the five regions as broadly similar in medication use, healthcare utilization, and patient demographic composition, reinforcing the assumption of national homogeneity (Henriksen et al. 2015). However, studies on the treatment of depression, such as investigations on the use of antidepressants or electroconvulsive therapy (ECT), have shown a difference in prescription and treatment rates across regions and municipalities (Raft et al. 2026; Hundrup, Osler, and Jørgensen 2017). Notably, municipalities with higher levels of socioeconomic disadvantage had higher rates of antidepressant prescriptions (Raft et al. 2026). In addition, substantial and currently unexplained regional differences exist in both ADHD prevalence and psychiatric service capacity (Sonne et al. 2025). These findings underscore a growing need for research that examines regional differences in disease prevalence, treatment practices, and their underlying causes.

Variable Selection

Statistics Denmark sets standards for variable definitions and usage within the registries, including for ethnic-

ity (Hjorth et al. 2025; Statistics Denmark 2025c). These standards recommend using the immigration status and the country of origin variables to categorize ethnicity as Danish, “Western”, or “non-Western.” Western countries include: the European Union, Andorra, Iceland, Liechtenstein, Monaco, Norway, San Marino, Switzerland, the United Kingdom, the Vatican State, the United States, Canada, Australia, and New Zealand. Non-western countries include: Albania, Bosnia and Herzegovina, Belarus, Yugoslavia, Kosovo, Macedonia, Moldova, Montenegro, Russia, Serbia, Soviet Union, Turkey, Ukraine, all countries in Africa, all countries in South and Central America, all countries in Asia, all countries in Oceania (except Australia and New Zealand), and those who are Stateless. The immigration status variable classifies individuals into three groups: immigrants, descendants, and people of Danish origin. Descendants are defined as people born in Denmark whose parents were not born in Denmark nor have Danish citizenship, whereas people of Danish origin have at least one parent born in Denmark with Danish citizenship, regardless of whether they were born in Denmark themselves. Accordingly, the country of origin is recorded as Denmark for all people of Danish origin, the country of citizenship for descendants, and the country of birth for immigrants. Statistics Denmark also recommends using the immigration status variable to aggregate those with “immigrant” and “descendant” status into one group. These standards reverberate through existing research, such that even when research groups claim to account for ethnicity in their models (Naesager et al. 2024; Rasmussen et al. 2025), the reality is that they categorize people in ways that over-generalize ethnicity and heritage.

Predictive Models with Tabular Data

In public health research, predictive models are most often based on traditional statistical approaches, such as logistic regression and Cox proportional hazards models. In contrast, in health data science, predictive models tend to be machine learning models. Examples include tree-based models, e.g., Random Forest (Breiman 2001) or XG-Boost (Chen and Guestrin 2016); as well as deep learning based models, e.g., CNNs, RNNs, or transformers. Whilst most predictive AI models for images or text currently utilize deep learning architectures, earlier work has argued that tree-based models still outperform deep learning architectures and remain state-of-the-art for medium-sized tabular datasets (Grinsztajn, Oyallon, and Varoquaux 2022). In our case, even though the registries inside Statistics Denmark contain purely tabular data, they consist of hundreds of thousands of individuals’ data, making deep learning architectures preferable. By now, several deep learning models specifically for tabular data have been developed. In our work, we have therefore chosen to utilize a standard statistical model, namely logistic regression, for association analysis as well as a transformer model for tabular data (Huang et al. 2020; Gorishniy et al. 2021). One additional advantage of using foundational transformer models for prediction tasks is the ability to assess feature importance using the attention weights, thus we chose this approach (Gorishniy et al. 2021).

Fairness Evaluations of Medical Data

Fairness evaluations offer a framework for distinguishing differential model performance across subgroups. For example, clinical prediction models often fail to perform consistently across all patient subgroups, leading to some patients receiving better care than others if said models were deployed. Hence, performance difference evaluation is the first step towards fair models (Petersen et al. 2023a). Fairness evaluations can also help to detect underlying systemic biases, e.g., differences in error rates across subpopulations. This is especially important in medical applications, since type I error is worse than a type II error in many medical applications. For example, the consequences of leaving depression untreated could include a serious decline in physical health or even suicide. This makes underdiagnosis far more dangerous than sending too many patients to psychotherapy, where the cost and harm of unnecessary treatment are relatively low. Hence, understanding the nature of the prediction errors helps address and mitigate them. Furthermore, bias detection can be highly valuable to ensure ethical development of medical AI, as well as compliance with regulatory requirements and health law. Policies such as the EU AI Act have led to new regulatory requirements for AI models that emphasize fairness and model explainability (European Parliament 2023). This aims to improve the transparency, interpretability, and accountability of AI models by leveraging methods like feature importance to offer users insight into how the model makes decisions.

Data

We leveraged Denmark’s nationwide registry data, using CPR numbers to link health, demographic, and socioeconomic registry data drawn from the Danish National Patient Register, the Danish Civil Registration System, the Danish Psychiatric Central Research Register, the Income Register, and the Education Register. All experiments and analyses were conducted on Statistics Denmark servers in compliance with data security rules and contingent on project approval. There are many other studies that use the same data, providing detailed descriptions about the variables available in the registries and how such data was obtained (Plana-Ripoll et al. 2025; Weye et al. 2023; Christensen et al. 2022b).

We created a matched cohort composed of two groups: an MDD group and a healthy control group. For each year between 2000 and 2018, we included all individuals residing in Denmark who received an MDD diagnosis in the MDD group, provided they had at least five years of prior medical records from age 15 or older showing no previous MDD diagnosis and had not been included in the MDD group at an earlier time point. Due to the five-year look-back, the youngest eligible age at the index year is effectively 19. MDD diagnoses were defined as primary or secondary diagnoses of ICD-10 codes F32 or F33. The reference index year was the year in which a patient entered the MDD group. Each individual in the MDD group was matched to an individual from the general population for the same index year. To be eligible for the healthy control group, individuals were required to have no MDD diagnosis in the 5 years prior to

the index year, be at least 15 years old (including those who turn 15 during the index year) at the start of the look-back period, and remain alive through the end of that year. We excluded individuals from both groups if they were over 65 at the index year, died on the date of receiving an MDD diagnosis, died before the index date, or were not continuously residing in Denmark during the five-year look-back period. Additionally, individuals in the healthy control group who later joined the MDD group were excluded.

Patient-level features were derived from Danish register data cross-sectionally at the index year. This included the following sociodemographic factors: age in index year, sex, LGBT identity (individuals in a registered same-sex marriage or civil partnership, or those with a registered sex change), civil status (married, not married, divorced, or widowed) (Statistics Denmark 2025a), immigration status (Danish origin, descendant or immigrant) (Statistics Denmark 2025b), country of origin¹, administrative region of residence (North Jutland, Middle Jutland, South Denmark, Zealand, or Capital Region), and municipality (51 "kommune groups"). We also derived the following socioeconomic features from the Danish register: highest education level (unknown, short, medium, or long) (Hjorth et al. 2025), income rank (equivalised household income quartile) (Hjorth et al. 2025), cohabitation status (living alone or with others), and address changes (number of address changes in the index year).

We obtained additional sociodemographic variables by mapping country of origin to: 1) Statistics Denmark's definition of Western and non-Western countries (Statistics Denmark 2025c) and 2) the United Nations Intermediate Region categories (United Nations Statistics Division 2006). In this way, we were able to design two variables to reflect heritage as a sociodemographic factor, as described below.

1. Heritage 1: This variable definition was based on previous health studies using Danish registry data, where Statistics Denmark recommends that immigrants and descendants are grouped according to whether their country of origin is defined as "Western" or "non-Western".
2. Heritage 2: This variable definition was based on the geographic region of the world they or their family originated from, using the UN's Intermediate Regions.

The second heritage variable aims to group individuals using a broader, global representation of heritage, with the acknowledgment that some regions may share cultural characteristics. By using both the immigration status and the heritage variable, we are able to characterize individuals with more cultural sensitivity.

Lastly, besides delineating heritage more sharply and accounting for the SES associated with one's country of origin, we chose to create and add another socioeconomic variable, Country of Origin Economic Level (COEL). COEL maps country of origin to the 2019 World Bank Levels (World

¹Country of origin for immigrants is their country of birth. For descendants, it is their country of citizenship despite having born in Denmark. For people of Danish origin (i.e., at least one parent with Danish citizenship/born in Denmark), this is Denmark regardless of where they were born.

Bank Data Team 2018) provided by the UN. This variable sorts countries into four groups according to their gross national income per capita, ranging from low-income to high-income, allowing us to control for the SES associated with one's country of origin. By including this variable, we aimed to provide a more granular metric of SES that may be related to the healthcare access and utilization of an individual or their parents before immigrating to Denmark. An overview of the cohort composition detailing population statistics is shown in Figure 1 B and C, while detailed sociodemographic and socioeconomic variables can be seen in Tables 5 and 6 in the Appendix (located in the *MDD-Denmark* Github Repository).

Observational Association Analysis

Our first aim is to examine differences in MDD diagnosis rates across Denmark's five administrative regions. We use binary logistic regression to estimate the association between the likelihood of receiving an MDD diagnosis, region of care, and our sociodemographic and socioeconomic variables. While doing so, we aim to control for individual sociodemographic and socioeconomic characteristics using variables designed to provide greater cultural sensitivity than those typically seen in Danish public health studies. Odds ratios (ORs) and 95% confidence intervals (CIs) are calculated using the exponents of the coefficients, and statistical significance denoted if $p < 0.05$. We use the more informative variables to represent heritage, and consequently, the formula used for the analysis is given below:

$$\begin{aligned} \text{logit}(P(\text{MDD})) = & \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Sex} + \beta_3 \text{Civil Status} \\ & + \beta_4 \text{Region} + \beta_5 \text{Municipality} \\ & + \beta_6 \text{Education} + \beta_7 \text{Income} \\ & + \beta_8 \text{Heritage2} + \beta_9 \text{COEL} \\ & + \beta_{10} \text{Immigration Status} \\ & + \beta_{11} \text{Cohabitation} \\ & + \beta_{12} \text{Address Changes} + \beta_{13} \text{LGBT} \end{aligned}$$

Association Analysis Results

Regional Differences The logistic regression-based analysis demonstrated significant regional differences in the odds of receiving an MDD diagnosis. Among the five administrative regions, residents of Middle Jutland had the highest odds of receiving an MDD diagnosis (OR=1.03 [95% CI=1.02-1.03]) compared to the Capital Region. In contrast, living in Zealand was associated with lower odds of diagnosis (0.98 [0.95-1.00]) relative to the Capital Region. At the municipal level, the lowest odds of MDD diagnosis were observed in North Jutland's municipalities (OR=0.55-0.93). The six municipalities with the highest odds (OR=1.04-1.08) were located in Zealand and South Denmark. Notably, South Denmark and the Capital Region showed the greatest variation in odds across municipalities (OR=0.95-1.03). Across all regions, 22 municipalities had significantly lower odds, and 9 had significantly higher odds of receiving an MDD diagnosis compared to Copenhagen municipality.

Sociodemographic Group Older age was associated with significantly lower odds of receiving an MDD diagnosis. For example, individuals at the minimum age of 19 had an OR of 0.96 (95% CI=0.96-0.96), compared to 0.88 (0.88-0.88) for those at the maximum age of 65. Males had significantly lower odds of being diagnosed with MDD compared to females (0.89 [0.89-0.89]). There was a significant association between those identified as LGBT in the registers and greater odds of MDD diagnosis (1.10 [1.08-1.13]). Compared to those who had never married, married individuals had significantly lower odds of receiving a diagnosis (0.99 [0.99-1.00]), while those registered as divorced had significantly higher odds (1.07 [1.06-1.08]). Our findings show that descendants and immigrants have distinctly different odds of an MDD diagnosis; those registered as descendants had significantly lower odds of being given an MDD diagnosis (0.91 [0.89-0.93]) compared to individuals of Danish origin. In contrast, individuals recorded as immigrants had significantly higher odds of diagnosis (1.04 [1.03-1.06]) relative to persons of Danish origin. Those with a country of origin in Northern Africa (1.10 [1.05-1.15]), Southern Asia (1.08 [1.04-1.11]), Southern Europe (1.07 [1.04-1.11]), or Western Asia (1.05 [1.02-1.08]) had significantly higher odds of receiving an MDD diagnosis compared to Northern Europe. In contrast, heritage from Eastern Africa (0.86 [0.82-0.90]), Eastern Asia (0.84 [0.79-0.88]), Eastern Europe (0.97 [0.95-0.99]), South-eastern Asia (0.89 [0.85-0.92]), Western Africa (0.88 [0.83-0.94]), and Western Europe (0.96 [0.93-0.98]) were associated with significantly lower odds of diagnosis relative to Northern Europe.

Socioeconomic Factors Longer education was associated with significantly lower odds of receiving an MDD diagnosis (0.93 [0.93-0.94]) compared to individuals with short education. Similarly, individuals in the top income rank had lower odds (0.81 [0.80-0.81]) of diagnosis relative to those with the lowest income rank. Individuals had higher odds of obtaining an MDD diagnosis if they lived alone (1.12 [1.12-1.12]) or moved in their index year such that at least one address change in their index year increased their odds (1.08 [1.08-1.08]). Additionally, having a country of origin with an upper middle-income level (1.04 [1.02-1.07]), low-income level (1.05 [1.02-1.09]), or an unknown country of origin (1.11 [1.08-1.14]) was related to significantly higher odds of obtaining an MDD diagnosis compared to having a high income country of origin. On the contrary, a lower middle-income level of one's country of origin resulted in significantly lower odds of obtaining a diagnosis (0.93 [0.90-0.96]).

Summary

Our findings suggest that the likelihood of receiving an MDD diagnosis varies across Denmark's five Danish administrative regions, confirming regional heterogeneity. Moreover, we observe a large variation at the municipality level within each region, supporting recent work (Raft et al. 2026). Sociodemographic and socioeconomic variables also showed significant associations with MDD diagnoses. Notably, odds of diagnosis are more similar between immi-

grants and individuals of Danish origin than between immigrants and their descendants, despite the latter two often being grouped together in other modeling studies (Hjorth et al. 2025). Both heritage, defined as the region of their country of origin, and the economic status of an individual's country of origin significantly affects diagnosis likelihood. This suggests models should include more robust variables beyond simplistic broad categorizations of background to better reflect how cultural variation impacts one's odds of diagnosis. Lastly, our results highlight the importance of incorporating multiple SES measures to fully understand and address health disparities.

MDD Prediction Model

In the second part of the analysis, we aim to assess the feasibility of developing an MDD prediction model using Danish registry data. In doing so, we seek to address the following questions:

1. If we were to deploy an MDD prediction tool nationwide, would it perpetuate or exacerbate existing disparities in mental healthcare?
2. Does model performance improve if we include culturally sensitive variables instead of the typical variable based upon the "Westernness" of the country of origin? *If so...* which variables most greatly influence model predictions?
3. Can models trained on data from only one region generalize well to another region?

Experimental Methodology

We used the dataset described in the Data Section to construct six subsets, each corresponding to distinct regional groupings: 1) All Regions; 2) the Capital Region; 3) North Jutland; 4) Middle Jutland; 5) South Denmark; 6) Zealand.

For the analyses, each dataset was divided into training, validation, and test sets (with percentages 50%, 20%, and 30%, respectively). A tabular transformer model, TF-Transformer (Gorishniy, Rubachev, and Babenko 2022; Gorishniy et al. 2021), was trained twice on each training set using two distinct sets of variables. The Baseline Model included variables commonly used in psychiatric prediction tasks based on Danish registry data and recommended by Statistics Denmark (Hjorth et al. 2025; Naesager et al. 2024; Rasmussen et al. 2025; Perfalk et al. 2024; Nielsen et al. 2023). In contrast, the Heritage-Informed Model incorporated variables designed to better reflect socioeconomic factors and distinguish immigration from heritage, as described in the Data Section. The feature sets used by the two prediction models are detailed below:

1. Baseline Model: Age, Sex, Civil Status, Region, Municipality, Education, Income, Heritage 1 (3-level: Danish / Western Immigrant / Non-Western Immigrant)
2. Heritage-Informed Model: Age, Sex, Civil Status, Region, Municipality, Education, Income, Heritage 2 (UN Intermediate Regions), COEL (World Bank Economic Level), Immigration Status, Cohabitation, Address Changes, LGBT

	Baseline Model				Heritage-Informed Model			
	Acc	Brier	AUROC	DRMSCE	Acc	Brier	AUROC	DRMSCE
All	0.649	0.220	0.701	0.006	0.662	0.216	0.715	0.006
Capital Region	0.642	0.224	0.691	0.000	0.652	0.220	0.703	0.000
Middle Jutland	0.657	0.217	0.710	0.006	0.668	0.212	0.725	0.008
North Jutland	0.668	0.236	0.682	0.000	0.683	0.230	0.704	0.006
Zealand	0.643	0.225	0.690	0.018	0.661	0.219	0.707	0.000
South Denmark	0.648	0.217	0.708	0.011	0.661	0.213	0.721	0.009

Table 1: Model performance by model and region for the two MDD prediction models trained on data from all five administrative regions.

Training the model with two distinct variable sets across six regional dataset subsets resulted in 12 fitted models. Each model was evaluated on the test set corresponding to its own region (the region on which they were trained) as well as on the test sets from all other regions.

Model Evaluation We assessed several model performance metrics that are recommended for assessing fairness in machine learning (Maier-Hein et al. 2024) and can be meaningfully compared between subgroups of different size and prevalence (Sutariya and Petersen 2025; Petersen et al. 2023a), including: Area Under the Receiver Operating Characteristic (AUROC), Debaised Root Mean Square Calibration Error (DRMSCE), Accuracy (Acc), and balanced Brier Score (BS). When calculating the accuracy, we utilized a threshold equal to the average predictive probability. We evaluated the models using the following subgroups: administrative region, immigration status, Heritage 1, and Heritage 2. Subgroups of Heritage 1 and Heritage 2 that had sufficient sample sizes for interpreting model performance ($N > 1,000$) were selected for further analysis. For Heritage 1, this included all but those with unknown heritage. For Heritage 2, this included nine subgroups (Table 2). We used the meval package for computing all metrics and associated uncertainties (i.e., confidence intervals) within subgroups and for generating the final figure (Figure 2 in the Appendix) (Sutariya and Petersen 2025).

MDD Prediction Results

Below, we first present the performance of the Baseline and Heritage-Informed models after training on the entire dataset comprising all five administrative regions. We examine the feature importance of both models to better understand their performance differences (Figure 3 in the Appendix). Following this, we compare model performance for the Baseline and Heritage-Informed Models trained on the entire dataset for each of the five regions separately (Table 1). Then we examine disparities in model performance across the different immigration and Heritage groups (Table 2). Secondly, we present a fairness analysis of the Heritage-Informed Model trained on all regions in Figure 2 (Appendix). Finally, in Table 3 we report the performance of models trained on the region-specific subsets of the dataset and evaluated across the other regions to assess their generalizability.

Model Performance The Heritage-Informed Model, which included more granular representations of heritage,

immigration, and socioeconomic factors, had slightly improved model performance (AUROC = 0.72) compared to the Baseline model that uses the simpler characterization of heritage recommended by Statistics Denmark (AUROC = 0.70). The Heritage-Informed Model also demonstrated higher accuracy, and lower BS as seen in Table 1. Including more culturally-sensitive variables also benefited model performance across all five administrative regions, despite differences in their demographic makeup, with increases in all accuracy, Brier scores, and AUROC values, along with some improvements in calibration as seen by the DRMSCE scores (Table 1). We also saw increased AUROC values for population subgroups delimited according to their heritage and immigration status (Table 2). Due to small differences in performance, we also evaluated both models using multiple seeds. This is discussed in detail in the Appendix.

Feature Importance In addition, we compared the feature importance of the Baseline and the Heritage-Informed model to better understand what may be driving the performance improvements. The Baseline Model MDD predictions are most influenced by municipality (Figure 3 in Appendix). In contrast, the Heritage-Informed model depends most on the Heritage 2 and COEL features, and did not show the strong influence of one’s municipality on model predictions (Figure 3 in Appendix). These results suggest differences in the demographic makeup of different municipalities, which may be related to cultural background and heritage, thus our model learned to attribute importance to municipality level values if the more robust variables are not included. It’s important to note that researchers typically use socioeconomic variables that are fixed to someone’s life in Denmark instead of incorporating socioeconomic variables related to life before migrating to Denmark (i.e., COEL). Quantifying socioeconomic status in this way allows us to account for economic resources an individual may have had access to or possible hardships experienced before immigrating to Denmark, which may be culturally influenced but otherwise impossible to fully account for with other variables.

Fairness Evaluation When trained on the full dataset comprising all five Danish administrative regions, the greatest performance disparities are seen in North Jutland. We report regional fairness metrics for the Heritage-Informed Model because it demonstrated the best overall performance. Although North Jutland demonstrated the highest accuracy

	Baseline Model				Heritage-Informed Model			
	Acc	Brier	AUROC	DRMSCE	Acc	Brier	AUROC	DRMSCE
Immigration Status								
Danish origin	0.650	0.221	0.700	0.006	0.662	0.217	0.712	0.004
Descendant	0.634	0.227	0.700	0.000	0.632	0.227	0.678	0.000
Immigrant	0.640	0.228	0.678	0.009	0.657	0.219	0.710	0.003
Heritage 1								
Danish	0.650	0.221	0.700	0.006	0.663	0.217	0.713	0.004
Non-Western Immigrant	0.641	0.233	0.671	0.004	0.658	0.223	0.706	0.009
Western Immigrant	0.632	0.228	0.678	0.000	0.642	0.222	0.700	0.000
Heritage 2								
Eastern African heritage	0.580	0.238	0.683	0.123	0.639	0.225	0.683	0.000
Eastern European heritage	0.632	0.227	0.673	0.000	0.632	0.221	0.693	0.000
Northern African heritage	0.653	0.223	0.724	0.030	0.643	0.222	0.702	0.000
Northern European heritage	0.650	0.221	0.700	0.006	0.663	0.217	0.713	0.005
Southeastern Asian heritage	0.534	0.245	0.613	0.163	0.648	0.251	0.627	0.000
Southern Asian heritage	0.636	0.232	0.665	0.000	0.651	0.221	0.709	0.000
Southern European heritage	0.656	0.226	0.685	0.051	0.658	0.234	0.695	0.000
Western Asian heritage	0.683	0.232	0.683	0.043	0.677	0.234	0.697	0.000
Western European heritage	0.635	0.229	0.683	0.000	0.653	0.227	0.704	0.000

Table 2: Model performance by model, immigration status, and heritage for the two MDD prediction models trained on data from all five administrative regions. *Note: Metrics are reported for Heritage 1 and Heritage 2 subgroups with sufficient sample size ($N > 1,000$).*

(0.68), it had the poorest reliability ($BS = 0.23$) and the lowest AUROC (0.70). Figure 2E in Appendix further suggests that the Heritage-Informed model frequently predicts a lower risk of MDD diagnosis in North Jutland than other regions, although the model appears to be well calibrated (Figure 2D in Appendix). Otherwise, the model appeared to perform best on Middle Jutland ($AUROC = 0.73$; $BS = 0.21$, $Acc = 0.67$) and South Denmark ($AUROC = 0.72$; $BS = 0.21$, $Acc = 0.66$). The model performed similarly well on Zealand ($AUROC = 0.71$; $BS = 0.22$, $Acc = 0.66$) and the Capital Region ($AUROC = 0.70$; $BS = 0.22$, $Acc = 0.65$). The model is well calibrated in all regions ($DRMSCE < 0.01$).

When trained on data from all five regions, the Baseline and Heritage-Informed models exhibit performance differences according to heritage and immigration status. When comparing performance across immigration status, the Heritage-Informed model performed best for immigrants ($AUROC = 0.71$; $BS = 0.22$, $Acc = 0.66$) and those of Danish origin ($AUROC = 0.71$; $BS = 0.22$, $Acc = 0.66$), but worst for descendants ($AUROC = 0.68$; $BS = 0.23$, $Acc = 0.63$). In comparison, the Baseline model has similar performance for descendants ($AUROC = 0.70$; $BS = 0.22$, $Acc = 0.65$) and persons of Danish origin ($AUROC = 0.70$; $BS = 0.23$, $Acc = 0.63$), with the lowest scores for immigrants ($AUROC = 0.68$; $BS = 0.23$, $Acc = 0.64$). When comparing the performance between the different subgroups of the Heritage 1 variable, both the Baseline model and the Heritage-Informed model showed the highest performance for Danes, and the lowest performance for immigrants (Table 2). Exploring model performance across the subgroups of the Heritage 2 variable allows us to have a more granular view of which global regional subgroups may drive the model's differential

performance across the Heritage 1 subgroups. Of individuals that would be characterized as "non-Western" immigrants, the highest performance is for people with Northern African heritage for both the Baseline and Heritage-Informed model (Table 2), with similarly good performance for those with Southern Asian heritage ($AUROC = 0.71$; $BS = 0.25$, $Acc = 0.65$) if using the Heritage-Informed model. Both models have the lowest performance metric values for individuals with Southeast Asian heritage (Table 2).

Regional Generalizability After training models with region-specific data, we can see that some models are more generalizable than others. This trend holds true for both the Baseline Model and the Heritage-Informed Model. For instance, the Heritage-Informed model improves the generalizability of models trained only on data from the Capital Region, Zealand, or South Denmark (Table 3). However, when trained only on data from North Jutland, the Baseline model is preferred for enhanced regional generalizability. There are similar trends for Middle Jutland. Overall, the western regions generalize well with each other via the Baseline or Heritage-Informed model; models trained on data from Middle Jutland or South Denmark demonstrate fair performance when evaluated on each other's data ($AUROC > 0.70$). Models trained on data from Zealand also perform fair when evaluated on data from Middle Jutland and South Denmark ($AUROC > 0.70$). However, when evaluated on Zealand data, models trained on data from Middle Jutland ($AUROC = 0.66$) or South Denmark ($AUROC = 0.67$) perform poorly. North Jutland and the Capital region were the two regions that performed the worst, both performing poorly on data from the other ($AUROC < 0.67$). This suggests that certain regions are more generalizable than others, and generaliz-

Train	Regions Test	Baseline Model			Heritage-Informed Model		
		Acc	Brier	AUROC	Acc	Brier	AUROC
Capital Region	Capital Region	0.638	0.362	0.683	0.647	0.353	0.694
	Middle Jutland	0.626	0.374	0.686	0.643	0.357	0.692
	North Jutland	0.661	0.339	0.662	0.663	0.337	0.671
	Zealand	0.589	0.411	0.659	0.632	0.368	0.674
	South Denmark	0.607	0.393	0.675	0.626	0.374	0.684
Middle Jutland	Capital Region	0.621	0.379	0.672	0.618	0.382	0.676
	Middle Jutland	0.659	0.341	0.712	0.668	0.332	0.723
	North Jutland	0.659	0.341	0.657	0.604	0.396	0.680
	Zealand	0.608	0.392	0.662	0.593	0.407	0.657
	South Denmark	0.610	0.390	0.682	0.645	0.355	0.705
North Jutland	Capital Region	0.619	0.381	0.654	0.598	0.402	0.675
	Middle Jutland	0.629	0.371	0.676	0.623	0.377	0.701
	North Jutland	0.661	0.339	0.669	0.674	0.326	0.694
	Zealand	0.630	0.370	0.661	0.618	0.382	0.683
	South Denmark	0.625	0.375	0.676	0.612	0.388	0.699
Zealand	Capital Region	0.608	0.392	0.641	0.635	0.365	0.681
	Middle Jutland	0.613	0.387	0.638	0.647	0.353	0.701
	North Jutland	0.644	0.356	0.649	0.665	0.335	0.682
	Zealand	0.639	0.361	0.686	0.657	0.343	0.701
	South Denmark	0.605	0.395	0.641	0.634	0.366	0.697
South Denmark	Capital Region	0.626	0.374	0.671	0.629	0.371	0.678
	Middle Jutland	0.643	0.357	0.689	0.651	0.349	0.705
	North Jutland	0.635	0.365	0.667	0.625	0.375	0.682
	Zealand	0.619	0.381	0.653	0.626	0.374	0.671
	South Denmark	0.648	0.352	0.705	0.652	0.348	0.713

Table 3: Model performance by region for training and evaluation for the two MDD prediction models trained on data from individual regions.

ability is bi-directional or not always reciprocal. Lastly, the model trained on data from all five regions has greater performance on each region than models only trained on a single region. Therefore, there is something to gain from using nationwide data to build predictive models.

Discussion

We investigated whether implementing a nationwide MDD prediction tool in Denmark could exacerbate existing disparities in mental healthcare. In our analysis, we first analyzed the sociodemographic, socioeconomic, and regional factors that are associated with MDD diagnostic patterns in Denmark. Then, we examined whether the disparities and potential biases in the input datasets are learned by AI models designed for MDD prediction tasks. We observed that the **AI models reproduced the variance present in the input data, leading to inequities** in model performance across the different administrative regions, immigration status, and heritage. However, our findings show that **model performance can be improved for almost all subgroups by incorporating features with cultural sensitivity**. It should be noted that these benefits of variable selection were not seen for everyone, specifically the descendants of immigrants, raising public health questions pertaining to who is prioritized in contexts of deploying such models. Lastly, our findings show that the generalizability of regionalized models is bi-directional, meaning that a model trained on a single region

may perform well on another, but reciprocation is not guaranteed. In contrast, training on nationwide data yielded the strongest overall performance for each region.

Our work sheds light on the benefits and risks of using nationwide health models, especially for people living in rural regions. While the models trained on all regions performed the best overall in terms of generalization to other regions, they still have lower performance for specific regions like North Jutland. North Jutland is a rural region of Denmark with the largest patient-provider distances in the country, and where patients experience the longest waiting times for psychiatric care compared to the other regions (Jyllands-Posten 2024; Danish Broadcasting Corporation 2024). This is also reflected in our observational association analysis, which shows a lower prevalence of MDD in North Jutland compared to the other regions of Denmark. Whilst the reason for the lower prevalence in North Jutland remains to be determined, we see that without applying bias mitigation methods to AI models, nationwide and region-based models will exhibit lower performance and potentially perpetuate existing inequities in North Jutland due to unbalanced data that may reflect access to care. **Although these findings are focused on Denmark, they are broadly applicable to any other healthcare system with regional differences**—for example, a model being applied across states with varied insurance or healthcare access in the United States.

Another of our main takeaways is **the importance of**

thoughtful variable selection when developing AI models with real world data, especially when defining sensitive groups for fairness evaluation. We here pursue a “fairness through awareness” approach (Dwork et al. 2012), explicitly including sensitive attributes in our models. Some argue that the inclusion of sensitive variables in predictive algorithms in medicine is controversial and the subject of intense debate (Diao et al. 2024; Benthall and Haynes 2019). In this regard, our findings mirror the observations of Zink et al. that explicit inclusion of sensitive variables is often beneficial for the affected patient groups (Zink, Obermeyer, and Pierson 2024). For instance, we found that the inclusion of more culturally-sensitive features improved model predictive performance not only for immigrants and several minority groups but also for individuals of Danish origin. Moreover, we found variation in model performance across the Heritage 2 variable subgroups, which may otherwise have been masked if using the “Western”/“non-Western” categorization. Existing research has established that “non-Western” immigrants in Denmark have different mental healthcare needs than persons of Danish origin or “Western” immigrants, which can be influenced by refugee status or lived experiences of racism and xenophobia (Tegunimataka 2022; Nielsen et al. 2014; Bruckner et al. 2024; Norredam et al. 2009). Given this understanding, the cultural diversity in urban regions like Copenhagen may contribute to the differing healthcare needs between the five regions.

Limitations and Future Work

This study has several limitations, primarily because our models rely solely on registry data. Most importantly, our analysis is based on MDD diagnoses in hospital records, excluding cases diagnosed in primary care or outside hospital settings. This exclusion criteria is a limitation of all research based on national registers (e.g., the Danish Psychiatric Central Research Register). This makes it challenging to quantify the true burden of depression due to the more severe cases of MDD being recognized in registry data. Second, while useful for machine learning, the matched cohort design limits generalizability beyond the selected sample. Third, our heritage definition is constrained by the absence of race or ethnicity variables; proxies like country of birth or citizenship obscure mixed heritage and nuanced identity, and the UN geographic categories still perpetuate occidental perspectives (Coronil 1996). Similarly, there is no existing registry variable to denote LGBTQ+ identity, and our method of constructing it likely omits many LGBTQ+ folks. We also lack environmental and psychosocial factors such as family dynamics or adverse childhood experiences. Finally, although we used a state-of-the-art transformer model, other approaches might perform differently depending on regional variations or variable quality, affecting robustness.

A broader limitation in this work is the challenge of predicting MDD. In our analysis, we treat diagnostic labels as ‘ground truth,’ ignoring potential under-, mis-, or overdiagnosis. This may affect bias assessments (Pfohl et al. 2025; Dutta et al. 2020; Wick, Panda, and Tristan 2019; Petersen et al. 2023b), especially for certain subgroups. More specifically, a patient’s cultural background influences their symp-

tom presentation (Kirmayer et al. 2001), exposure to clinician bias (Beach et al. 2025; Valentine et al. 2026), and lived experiences navigating a medical system. Therefore, an MDD diagnosis is not the same thing as experiencing mental health illness and diagnoses are in themselves human conceptualizations. Diagnoses represent the points at which individuals encounter, and are categorized by, a healthcare system embedded within broader societal inequities. Nevertheless, with the increasing use of AI in healthcare, projects such as ours are critical for improving transparency about model limitations. Our work shows that if we don’t take action, deploying these systems as-is risks perpetuating the disparities seen in this paper. In short, we do *not* advocate for using models such as ours to diagnose people.

In the future, we aim to address current study limitations by expanding our analysis to additional Danish datasets with more granular MDD risk information. One such dataset is the National Health Survey cohort, which includes self-reported depression data for 183,646 adults (Christensen et al. 2022a), of whom 13% reported depression. These data will complement register-based information by capturing milder cases and adding lifestyle and behavioral variables. We also plan to focus on under-served communities using the Social Medical Database (Button 2025), which contains characteristics and treatment courses for patients referred for social medical examination since 2014 ($n \approx 7,000$). Notably, 60% have psychiatric diagnoses, and over 30% lack prior treatment records; about half of these undiagnosed cases involve depression ($n \approx 2,500$). This will enable development and testing of AI models for MDD risk prediction using fine-grained data on socially vulnerable groups.

Conclusions

This study explored whether deploying an AI-based MDD prediction tool could perpetuate existing disparities in mental healthcare. We used Denmark as a case study, with its universal healthcare system, nationwide register data, and long psychiatric wait times. In an observational analysis, we found that sociodemographic, socioeconomic, and regional factors influence odds of MDD diagnosis itself. By developing predictive AI models on the same data, we observed that models trained on this data learn and reproduce these biases. This could in turn lead to inequities across regions, immigration status, and heritage. While nationwide models generally performed best overall, they underperformed in rural regions such as North Jutland, highlighting the risks of utilizing unbalanced data for model development, which is very common in real medical settings. Incorporating culturally sensitive variables and more measures of socioeconomic factors improved model fairness, underscoring the importance of thoughtful feature selection. However, our analysis is limited by the reliance on assumptions regarding ground truth labels of hospital-based diagnoses. Future work will integrate further datasets, such as the Danish National Health Survey and Social Medical Database, to capture milder MDD cases and socially vulnerable populations, ultimately aiming to test predictive algorithms on more data sources and address pitfalls, thereby improving equity in predictive modeling for MDD.

Ethical Statement

This study uses Denmark as a case study and hence utilizes nationwide Danish registry data collected and maintained by Statistics Denmark, which upholds strict data access rules to ensure the highest standards of data protection. We acknowledge that registry data may reflect historical and systemic biases; these limitations are discussed extensively in the Limitations and Future Work section. The findings are intended to inform equitable policy and practice, and care was taken to avoid conclusions that could stigmatize individuals or groups.

Acknowledgments

The authors would like to thank Vibe Frøkjær, Neurobiology Research Unit, Copenhagen University Hospital, Rigshospitalet for initial help with the MDD study conception. The authors would also like to thank Marie Raft for contributing to the map of Denmark in Figure 1. Work on this project was partially funded by the Independent Research Fund Denmark (DFF, grant number 9131-00097B), the Lundbeck foundation (Lundbeck Experiment, grant number R436-2023-857 and Lundbeck Collaborative TRUSTMIND, grant number R453-2024-356). The funding agencies had no influence on the writing, submission, or publication of this manuscript.

References

- Armbrrecht, E.; Shah, R.; Poorman, G. W.; Luo, L.; Stephens, J. M.; Li, B.; Pappadopulos, E.; Haider, S.; and McIntyre, R. S. 2021. Economic and Humanistic Burden Associated with Depression and Anxiety Among Adults with Non-Communicable Chronic Diseases (NCCDs) in the United States. *Journal of Multidisciplinary Healthcare*, Volume 14: 887–896.
- Beach, M. C.; Harrigan, K.; Chee, B.; Ahmad, A.; Links, A. R.; Zirikly, A.; Han, D.; Boss, E.; Lawson, S.; Saheed, M.; et al. 2025. Racial bias in clinician assessment of patient credibility: Evidence from electronic health records. *PloS one*, 20(8): e0328134.
- Benthall, S.; and Haynes, B. D. 2019. Racial categories in machine learning. *FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency*.
- Bhopal, R.; and Donaldson, L. 1998. White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health. *American Journal of Public Health*.
- Breiman, L. 2001. Random Forests. *Machine Learning*, 45(1): 5–32.
- Bruckner, T. A.; Singh, P.; Hvidtfeldt, C.; and Andersen, L. 2024. Involuntary Psychiatric Commitments Among Non-Western Immigrants During the Muhammad Cartoon Controversy in Denmark. *Journal of Immigrant and Minority Health*, 27(2): 234–241.
- Button, M. S. F. 2025. Borgere med relation til de socialmedicinske afdelinger i Danmark og prædiktorer for deres arbejdsmarkedstilknytning. *Socialmedicinsk klinik, Frederiksberg Hospital*, 50.
- Chen, T.; and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, 785–794. ACM.
- Christensen, A. I.; Juel Lau, C.; Lund Kristensen, P.; Buch Johnsen, S.; Wingstrand, A.; Friis, K.; Davidsen, M.; and Helms Andreasen, A. 2022a. The Danish National Health Survey : Study design, response rate and respondent characteristics in 2010, 2013 and 2017. *Scandinavian Journal of Public Health*, 50.
- Christensen, M. K.; McGrath, J. J.; Momen, N. C.; Whiteford, H. A.; Weye, N.; Agerbo, E.; Pedersen, C. B.; Mortensen, P. B.; Plana-Ripoll, O.; and Iburg, K. M. 2022b. The cost of mental disorders in Denmark: a register-based study. *npj Mental Health Research*, 1(1).
- Coronil, F. 1996. Beyond Occidentalism: Toward Nonimperial Geohistorical Categories. *Cultural Anthropology*.
- Dalsager, L.; Sørensen, K.; Sørensen, J. K.; Burr, H.; Dragano, N.; Magnusson Hanson, L.; Dahl Nielsen, M. B.; Hulvej Rod, N.; Wahrendorf, I.; Morten E.H. Madsen; and Rugulies, R. 2025. Time trends in prevalence of depressive disorder in the Danish working population from 2012 to 2018. *Journal of Affective Disorders*, 385.
- Danish Broadcasting Corporation. 2024. The wait to see a psychiatrist is getting longer and longer – and longer.
- Diao, J. A.; He, Y.; Khazanchi, R.; Nguemeni Tiako, M. J.; Witonsky, J. I.; Pierson, E.; Rajpurkar, P.; Elhawary, J. R.; Melas-Kyriazi, L.; Yen, A.; Martin, A. R.; Levy, S.; Patel, C. J.; Farhat, M.; Borrell, L. N.; Cho, M. H.; Silverman, E. K.; Burchard, E. G.; and Manrai, A. K. 2024. Implications of Race Adjustment in Lung-Function Equations. *New England Journal of Medicine*, 390(22): 2083–2097.
- Dutta, S.; Wei, D.; Yueksel, H.; Chen, P.-Y.; Liu, S.; and Varshney, K. 2020. Is There a Trade-Off Between Fairness and Accuracy? A Perspective Using Mismatched Hypothesis Testing. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, 2803–2813. PMLR.
- Dwork, C.; Hardt, M.; Pitassi, T.; Reingold, O.; and Zemel, R. 2012. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ACM.
- European Parliament. 2023. EU AI Act: first regulation on artificial intelligence.
- Ghio, L.; Gotelli, S.; Marcenaro, M.; Amore, M.; and Natta, W. 2014. Duration of untreated illness and outcomes in unipolar depression: A systematic review and meta-analysis. *Journal of Affective Disorders*, 152–154: 45–51.
- Gorishniy, Y.; Rubachev, I.; and Babenko, A. 2022. On Embeddings for Numerical Features in Tabular Deep Learning. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 24991–25004. Curran Associates, Inc.
- Gorishniy, Y.; Rubachev, I.; Khrulkov, V.; and Babenko, A. 2021. Revisiting deep learning models for tabular data. In

Proceedings of the 35th International Conference on Neural Information Processing Systems, NIPS '21. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781713845393.

Grinsztajn, L.; Oyallon, E.; and Varoquaux, G. 2022. Why do tree-based models still outperform deep learning on typical tabular data? In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 507–520. Curran Associates, Inc.

Hansen, L.; Bernstorff, M.; Enevoldsen, K.; Kolding, S.; Damgaard, J. G.; Perfalk, E.; Nielbo, K. L.; Danielsen, A. A.; and Østergaard, S. D. 2025a. Predicting Diagnostic Progression to Schizophrenia or Bipolar Disorder via Machine Learning. *JAMA Psychiatry*, 82(5): 459.

Hansen, L.; Damgaard, J. G.; Lundin, R. M.; Danielsen, A. A.; and Østergaard, S. D. 2025b. Predicting the need for electroconvulsive therapy via machine learning trained on electronic health record data. *medRxiv*.

Henriksen, D. P.; Rasmussen, L.; Hansen, M. R.; Hallas, J.; and Pottegård, A. 2015. Comparison of the Five Danish Regions Regarding Demographic Characteristics, Healthcare Utilization, and Medication Use—A Descriptive Cross-Sectional Study. *PLOS ONE*, 10(10): e0140197.

Hjorth, C.; Kjærulff, T.; Thomsen, M.; Cronin-Fenton, D.; Dalton, S.; and Olsen, M. 2025. SEPLINE: Socioeconomic Position in Epidemiological Research—A National Guideline on Danish Registry Data. *Clinical Epidemiology*, Volume 17: 593–624.

Huang, X.; Khetan, A.; Cvitkovic, M.; and Karnin, Z. 2020. TabTransformer: Tabular Data Modeling Using Contextual Embeddings.

Hundrup, E.; Osler, M.; and Jørgensen, M. B. 2017. Time Trends and Variations in Electroconvulsive Treatment in Denmark 2008 to 2014: A Nationwide Register-Based Study. *The Journal of ECT*, 33(4): 243–248.

Institute for Health Metrics and Evaluation. 2025. GBD Results.

Jyllands-Posten. 2024. In some places, patients wait up to three years to see a psychiatrist – and nine months to get an appointment with an ophthalmologist.

Kirmayer, L. J.; et al. 2001. Cultural variations in the clinical presentation of depression and anxiety: implications for diagnosis and treatment. *Journal of clinical psychiatry*, 62: 22–30.

Knudsen, S. V.; Valentin, J. B.; Norredam, M.; Videbech, P.; Mainz, J.; and Johnsen, S. P. 2022. Differences in quality of care, mortality, suicidal behavior, and readmissions among migrants and Danish-born inpatients with major depressive disorder. *European Psychiatry*, 65(1).

Maier-Hein, L.; Reinke, A.; Godau, P.; Tizabi, M. D.; Buettner, F.; Christodoulou, E.; Glocker, B.; Isensee, F.; Kleesiek, J.; Kozubek, M.; et al. 2024. Metrics reloaded: recommendations for image analysis validation. *Nature methods*, 21(2): 195–212.

Movva, R.; Shanmugam, D.; Hou, K.; Pathak, P.; Gutttag, J.; Garg, N.; and Pierson, E. 2023. Coarse race data conceals disparities in clinical risk score performance. *Machine Learning for Healthcare*.

Naesager, A. H. D.; Damgaard, S. N.; Rozing, M. P.; Siersma, V.; Møller, A.; and Tranberg, K. 2024. Developing a prediction model to identify people with severe mental illness without regular contact to their GP - a study based on data from the Danish national registers. *BMC Psychiatry*, 24(1).

Nielsen, S. D.; Christensen, R. H. B.; Madsen, T.; Karstoft, K.; Clemmensen, L.; and Benros, M. E. 2023. Prediction models of suicide and non-fatal suicide attempt after discharge from a psychiatric inpatient stay: A machine learning approach on nationwide Danish registers. *Acta Psychiatrica Scandinavica*, 148(6): 525–537.

Nielsen, S. S.; Jensen, N. K.; Kreiner, S.; Norredam, M.; and Krasnik, A. 2014. Utilisation of psychiatrists and psychologists in private practice among non-Western labour immigrants, immigrants from refugee-generating countries and ethnic Danes: the role of mental health status. *Social Psychiatry and Psychiatric Epidemiology*, 50(1): 67–76.

Norredam, M.; Garcia-Lopez, A.; Keiding, N.; and Krasnik, A. 2009. Risk of mental disorders in refugees and native Danes: a register-based retrospective cohort study. *Social Psychiatry and Psychiatric Epidemiology*, 44(12): 1023–1029.

Oluboka, O. J.; Katzman, M. A.; Habert, J.; McIntosh, D.; MacQueen, G. M.; Milev, R. V.; McIntyre, R. S.; and Blier, P. 2017. Functional Recovery in Major Depressive Disorder: Providing Early Optimal Treatment for the Individual Patient. *International Journal of Neuropsychopharmacology*, 21(2): 128–144.

Perfalk, E.; Damgaard, J. G.; Bernstorff, M.; Hansen, L.; Danielsen, A. A.; and Østergaard, S. D. 2024. Predicting involuntary admission following inpatient psychiatric treatment using machine learning trained on electronic health record data. *Psychological Medicine*, 54(15): 4348–4361.

Petersen, E.; Ganz, M.; Holm, S. H.; and Feragen, A. 2023a. On (assessing) the fairness of risk score models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, 817–829. ACM.

Petersen, E.; Holm, S.; Ganz, M.; and Feragen, A. 2023b. The path toward equal performance in medical machine learning. *Patterns*, 4(7).

Pfohl, S. R.; Harris, N.; Nagpal, C.; Madras, D.; Mhasawade, V.; Salaudeen, O. E.; Dieng, A.; Sequeira, S.; Arciniegas, S. E.; Sung, L.; Ezeanochie, N. P. O.; Cole-Lewis, H.; Heller, K. A.; Koyejo, S.; and D'Amour, A. N. 2025. Understanding challenges to the interpretation of disaggregated evaluations of algorithmic fairness. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.

Plana-Ripoll, O.; Liu, X.; Köhler-Forsberg, O.; Sørensen, H.; and Momen, N. 2025. Mental Disorders in Danish Hospital Registers: A Review of Content and Possibilities for Epidemiological Research. *Clinical Epidemiology*, Volume 17: 387–407.

- Raft, M.; Jørgensen, T.; Renneberg, J.; Lassen, C.; Jacobsen, R.; Jørgensen, A.; Jørgensen, M.; and Osler, M. 2026. Geographical variation and area-level socioeconomic disadvantage in recognized depression: A nationwide register-based Danish cohort study. *Social Psychiatry and Psychiatric Epidemiology*, 1–13.
- Rasmussen, M.; Santini, Z. I.; Joshanloo, M.; Petersen, C. B.; and Christensen, A. I. 2025. Determinants of Life Satisfaction and Mental Wellbeing in the Danish General Population: Shared and Distinct Associations. *International Journal of Public Health*, 70.
- Sasha Thorn Jakobsen, T.; Cristòbal Còppulo, E.; Rasmussen, S.; and Benros, M. E. 2025. Evaluating large language models for predicting psychiatric acute readmissions from clinical notes of population-based EHR. *medRxiv*.
- Schmidt, M.; Schmidt, S. A. J.; Sandegaard, J. L.; Ehrenstein, V.; Pedersen, L.; and Sørensen, H. T. 2015. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clinical Epidemiology*, 449.
- Sonne, H.; Pottgård, A.; Kjærgaard Hansen, A. S.; Bech, M.; and Kildegaard, H. 2025. Regional variation in ADHD treatment and diagnosis in Denmark. *medRxiv*.
- Statistics Denmark. 2025a. Definition of CIVST.
- Statistics Denmark. 2025b. Definition of IE TYPE.
- Statistics Denmark. 2025c. Definition of western countries and non-western countries.
- Statistics Denmark. 2025d. Population at first day of the quarter.
- Stevnsborg, L.; Jensen-Dahm, C.; Nielsen, T. R.; Gasse, C.; and Waldemar, G. 2016. Inequalities in Access to Treatment and Care for Patients with Dementia and Immigrant Background: A Danish Nationwide Study. *Journal of Alzheimer's Disease*, 54(2): 505–514.
- Sutariya, D.; and Petersen, E. 2025. meval: A Statistical Toolbox for Fine-Grained Model Performance Analysis. In *MICCAI Fairness of AI in Medical Imaging (FAIMI) Workshop*, 187–197.
- Tegunimataka, A. 2022. The Health of Immigrant Youth in Denmark: Examining Immigrant Generations and Origin. *Journal of International Migration and Integration*, 24(2): 659–694.
- United Nations Statistics Division. 2006. Standard country or area codes for statistical use (M49).
- Valentine, A. A.; Lepow, L. A.; Apakama, D.; Chan, L.; Charney, A. W.; and Landi, I. 2026. Bias Detection in Emergency Psychiatry: Linking Negative Language to Diagnostic Disparities. *ArXiv*, arXiv–2509.
- Wallach-Kildemoes, H.; Thomsen, L. T.; Kriegbaum, M.; Petersen, J. H.; and Norredam, M. 2014. Antidepressant utilization after hospitalization with depression: a comparison between non-Western immigrants and Danish-born residents. *BMC Psychiatry*, 14(1).
- Weye, N.; McGrath, J. J.; Lasgaard, M.; Momen, N. C.; Knudsen, A. K.; Musliner, K.; and Plana-Ripoll, O. 2023. Agreement between survey- and register-based measures of depression in Denmark. *Acta Psychiatr Scand*.
- Wick, M.; Panda, S.; and Tristan, J.-B. 2019. Unlocking Fairness: a Trade-off Revisited. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.
- World Bank Data Team. 2018. New country classifications by income level: 2018-2019.
- Zink, A.; Obermeyer, Z.; and Pierson, E. 2024. Race adjustments in clinical algorithms can help correct for racial disparities in data quality. *Proceedings of the National Academy of Sciences*, 121(34): e2402267121.