Predicting County-Level Mobility and Cognitive Disability Prevalence Using Machine Learning: A Comparative Analysis of CDC PLACES Data

Grace Solarz, Dylan Doyle, and Julian Leff

Department of Information Systems and Business Analytics, Kent State University

Capstone in Business Analytics - Writing Intensive, BA-44099-0011

Instructor Mostafa K. Ardakani

May 5, 2026

Abstract

Disability burden in the United States varies substantially across counties. Most studies treat disability as a single outcome rather than separating its important subtypes. This study uses CDC PLACES data covering 2,293 U.S. counties to build separate binary classification models for high Mobility Disability and high Cognitive Disability, these are defined based on a 75th percentile threshold. We have 24 numeric predictors organized into small, medium, and large feature sets using minimum redundancy maximum relevance selection (mRMR). 20 machine learning models were compared across 5 runs for each small, medium and large set using accuracy, macro F1, macro recall, and training time within a weighted ranking format. The best models received a 98.83% accuracy for Mobility and 97.08% for Cognitive Disability. Cubic SVM, Wide Neural Networks, and Weighted KNN performed the strongest. We found that the two subtypes of disability showed nearly zero overlap in top predictors, and that there is a strong bias towards disability in the South region of the United States.

Keywords:

Mobility disability - Cognitive Disability - CDC PLACES - Machine Learning - Feature Selection

Introduction

Living with a disability is a significant public health concern affecting the health and health care of individuals and their ability to participate in work, school and other activities. However, not all populations live with similar levels of disability. Throughout the United States there are variations in the prevalence of disability. Understanding the prevalence of disability within a geographic area is crucial for describing disability prevalence and guiding program development and funding decisions at the local, county, and regional levels. The Centers for Disease Control and Prevention (CDC) developed the PLACES dataset, which provides reliable small-area estimates for a variety of health topics and healthcare access measures. County-level data from the PLACES dataset can develop and validate models to predict two different disability measures: Mobility Disability (serious difficulty walking or stairs) and Cognitive Disability (serious difficulty concentrating, remembering, or making decisions).

Unlike Mobility Disability, which is largely predicated on the physical health status of and physical functional limitations of individuals, Cognitive Disability is more strongly associated with mental health, behavioral risk factors and socioeconomic status (Okoro et al., 2018). Treating disability as a single measure can lead to the use of ineffective general solutions for different causes of disability. Public health entities need to be able to identify populations at risk for disability in order to implement timely interventions. However, relying solely on self-reported disability status is not sufficient (Obermeyer and Emanuel, 2016). In addition, knowing the risk of future poor health would help officials identify where disability support services will be needed to assist individuals with ongoing limitations from non-infectious chronic diseases and mental health problems.

Most studies on disability prediction use simple models for prediction or engage in descriptive analysis or reduce all disabilities to a single score (Topol, 2019). In this study, we utilize multiple machine learning approaches (Rajkomar et al., 2019) to predict Mobility Disability and Cognitive Disability, and assess the performance of the models and identify the most important county-level predictors for each disability. To demonstrate the performance of the models as the size of the feature set increases, and employ minimum redundancy maximum relevance (mRMR) (Peng et al., 2005) to select small, medium, and large predictor sets. In addition, we considered Census Region as a predictor to determine if knowing a population's geographic region improves disability prediction beyond typical health and social determinants. The most important predictors for each subtype of disability are also identified.

Literature Review

Disability can be understood as a complex outcome that is influenced by a variety of factors, including physical and mental health status, socioeconomic status, geography and environment, and access to health care (Verbrugge & Jette, 1994). Each of these factors may individually affect the development and trajectory of disability and may have a cumulative effect.

Disability rates often follow a geographic pattern. Using national health and economic data, researchers have identified patterns of clustering and mapped out the rate of disability in relation to the national average for all ages. The data is arranged by state and county, and shows a geographic variation influenced by different factors such as socioeconomic status, healthcare, environment and demographics (Lu et al., 2023). Counties with high poverty rates and limited healthcare access have the greatest percentage of citizens with disabilities. Many also have high rates of chronic disease (Centers for Disease Control and Prevention, 2025).

The geographic clustering in these data suggests that the prevalence of disability is not random, and is influenced by a variety of community-level characteristics. By viewing the data at the county level and developing a predictive model to explain the prevalence of disability, patterns that are masked by simple summary statistics can emerge.

This set of indicators tracks rates of cognitive disability for adults and adults with disabilities and describes the mental / behavioral health predictors for adults. Cognitive Disability is often more closely associated with Mental Health Risk Factors, Behavioral Risk Factors and Social Determinants than with physical impairment (Livingston et al., 2020). Mental health risk factors for cognitive disability include depression, anxiety and emotional/behavioral disorders; severe mental illness; feelings of loneliness and isolation; substance use disorders; and normal age-related cognitive decline (Gale et al., 2018; World Health Organization, 2021). Cognitive Disability can also be affected by a number of mental health and behavioral risk factors. For example, smoking is a major risk factor for cerebrovascular disease, and poor physical activity is bad for brain health (Livingston et al., 2020). Like many health issues, Cognitive and Mobility Disabilities are affected by both physical and social determinants of health (Office of Disease Prevention and Health Promotion, n.d.). Consequently, mobility and Cognitive Disability is influenced by a number of common socioeconomic status indicators such as income and education, as well as by indicators of stability in housing, food security, and access to physical and mental health care services (Berkowitz et al., 2018; Taylor et al., 2016).

Machine learning involves the use of complex algorithms to make predictions based on large datasets of information, including to uncover highly non-linear relationships (Obermeyer & Emanuel, 2016). By applying machine learning to health outcomes data, it is possible to identify counties across the country where the risk of growing disability is higher than in other areas. Common machine learning techniques used in classification problems include decision trees, logistic regression, Naive Bayes, support vector machines (SVM), K-nearest neighbors (KNN), ensemble methods, and neural networks (Rajkomar et al., 2019).

Feature selection is a crucial component of any machine learning workflow. This becomes particularly important when you have a large number of highly correlated features, where the number of predictors exceeds the number of observations, or when your data includes a lot of redundant or irrelevant information that can reduce a model's accuracy and obscure the true relationships between relevant features and target variables (Radovic et al., 2017). A common strategy to tackle the 'curse of dimensionality' in feature selection is to select a subset of the most relevant features and, at the same time, as few redundant features as possible.

While there have been several methods proposed for ranking predictive genes or features in order to assess their contribution to a predictive model, mRMR (Maximum Relevance, Minimum Redundancy) additionally organizes the predictors into distinct feature sets (Peng et al., 2005). Therefore, in addition to ranking the individual predictors, mRMR allows one to assess the effect of adding additional features and to identify the most informative set of predictors.

Despite the vast literature on disability and related determinants, important research gaps remain. For example, most existing studies treat disability as one outcome, which hampers the ability to distinguish between different types of disability (Okoro et al., 2018).

Current research on predicting Mobility Disability using demographic, sensor, and machine learning features has several limitations. Specifically, current models use insufficient features, and only report accuracy while ignoring other important metrics such as recall and F1 score. There is also a lack of comparative studies that target Mobility Disability versus Cognitive Disability, and apply different machine learning methods.

In addition to estimating the comparative risk of two disability subtypes within a single model, we investigate the utility of feature selection and predictive performance when using different predictor sets. We apply machine learning methods to estimate the two disability subtypes separately, use the mRMR feature selection method to select a subset of the available predictors, and evaluate model performance when using a number of distinct feature sets. We also compare the predictive performance of several different models and use a weighted performance measure that prioritizes both balance and accuracy.

Using machine learning methods to predict disability outcomes at the county level using data from the public health sector, we find that the predictor patterns are distinct between Mobility Disability and Cognitive Disability.

Using a county-based approach, we provide the first predictions at high accuracy for several measures of different types of disability for each county in the contiguous United States.

We discuss our findings and their implications for public health practice. We conclude that effective disability interventions will need to be type-specific and cannot focus solely on reducing general Mobility Disability.

Data and Methods

3.1 Dataset Description

The CDC PLACES dataset provides county health estimates that allow researchers to compare risk factors and health outcomes across the US (Centers for Disease Control and Prevention, 2024). As for the dataset, each row represents one county in a state. The final cleaned dataset contained 2,293 U.S. counties, with twenty-four numeric predictors, one categorical predictor, and two binary target variables. The project's purpose regarding preparing a dataset is to determine if patterns in chronic disease, behavioral health, distress, access-to-care, and social-needs indicators can classify counties as high or low disability-burden communities.

The numeric predictor variables were grouped into multiple categories. The Chronic Disease / Health Outcome Predictors were Stroke, Diabetes, Obesity, COPD, Coronary Heart Disease, Arthritis, High Blood Pressure, High Cholesterol, Current Asthma, Cancer, and Poor General Health. The Behavioral Predictors were Smoking, Physical Inactivity, and Binge Drinking. The Mental / Physical Distress Predictors were Frequent Mental Distress, Frequent Physical Distress, and Depression. The Social Needs and Access-to-Care Predictors were No Health Insurance, Annual Checkup, Cholesterol Screening, Food Insecurity, Housing Insecurity, Loneliness, and No Social Support. As for the categorical predictor variable, the dataset featured Census Region, split up into the Northeast, the Midwest, the South, and the West. These categories were included because disability burden is usually connected to multiple health, behavioral, distress, access-to-care, and social factors rather than one factor alone.

3.2 Target Variables

Two binary target variables were created for this study: mobility class and cognitive class. This decision was made because Mobility Disability and Cognitive Disability represent different types of functional limitation and may be influenced by different upstream health factors (Okoro et al., 2018; Verbrugge & Jette, 1994). Rather than predicting one broad “Any Disability” outcome, this project separately shows Mobility Disability and Cognitive Disability, so the results could show whether each disability type had a different predictor pattern.

Both target variables were created using a seventy-fifth percentile threshold. Counties in the top 25% for each disability measure were labeled as high-disability counties, while counties below that threshold were labeled as low-disability counties. This created two separate binary classification questions: one for Mobility Disability and one for Cognitive Disability. The seventy-fifth percentile threshold was used because it identifies counties with high disability prevalence while still leaving enough counties in both classes for our models and analysis.

Mobility Disability is intended to measure serious difficulty walking or using stairs, which makes it relevant to physical and cardiovascular health, as well as to muscular or skeletal limitations and chronic disease burden (Centers for Disease Control and Prevention, 2024). In the final dataset, approximately 25.3% of counties were classified as high Mobility Disability, and 74.7% were classified as low Mobility Disability. Cognitive Disability refers to serious difficulty focusing, remembering, or making decisions. This outcome was modeled separately because cognitive limitations may be linked to different upstream factors than mobility limitations. These include mental distress, depression, social isolation, behavioral risks, and vascular health (Livingston et al., 2020; Gale et al., 2018). In the final dataset, approximately 25.1% of counties were classified as high Cognitive Disability, and 74.9% were classified as low Cognitive Disability.

3.3 Variables Excluded from Modeling

Some variables were useful for identifying and mapping counties, but they were not used as direct predictors in the main machine learning models. These variables were County Name, State Name, State Abbreviation, Location ID, FIPS code, and other administrative identifiers. These variables were excluded because they do not describe health, behavior, distress, access to care, or social needs. They are mostly there to identify where a county is. If they were included in the model, the model could just use the locations, rather than health patterns. These identification variables were still useful during the data preparation process, but they were not used as regular modeling predictors.

Census Region was handled differently because it is a broader category, not a direct county identifier. The region was tested separately to see whether broad geographic grouping added predictive value beyond just the project's health feature panel. This allowed the project to test geography in a controlled way without letting the model memorize individual county or state labels.

3.4 Data Cleaning

Data cleaning was an important part of this project because the final models depended on having a consistent dataset. The main goal was to ensure that each county appeared only once in the final dataset. It was also important to make sure that there was one clear value for each predictor and target variable. The cleaning process focused on checking for duplicate records and confirming that the target variables were coded correctly as zero or one. The dataset was also checked to make sure each county had one Mobility Disability label and one Cognitive Disability label. An earlier version of the project contained duplicate or conflicting county records. This created problems because some counties could appear in ways that led to inconsistent class labels. If the same county or the same type of county appears with conflicting labels, the model was not able to learn a clean pattern. The final revision corrected this issue before model training, and the cleaned dataset was implemented. The cleaning step improved the integrity of the project because it made sure the model was learning from a consistent county-level dataset.

3.5 Feature Sets

Feature sets were created to test whether model performance changed as more predictors were added. Having a smaller set of strong predictors can perform almost as well as a larger set. There are good reasons to try a small set, primarily because it can reduce run time, costs incurred by computing, and feature a smaller dataset. The predictors were organized into three main feature sets: Small, Medium, and Large. The Small feature set included 8 predictors. The Medium feature set included 15 predictors. The Large feature set included all 24 numeric predictors. A fourth condition, L + Region, was also tested by adding Census Region to the Large feature set.

The Small set included 8 core predictors: Stroke, Diabetes, Obesity, COPD, Coronary Heart Disease, Arthritis, Smoking, and Physical Inactivity. This group focused mostly on major chronic diseases and basic behavioral risk factors. It was used to see whether a simpler model could still classify high-disability counties well.

The Medium set included everything in the Small set, plus High Blood Pressure, High Cholesterol, Depression, Current Asthma, Cancer, Mental Distress, and Physical Distress. This added more cardiovascular, mental health, and distress-related predictors. The purpose of this group was to see whether these extra health measures improved prediction, especially for Cognitive Disability.

The Large set included all 24 numeric predictors. It added Binge Drinking, Poor General Health, No Health Insurance, Annual Checkup, Cholesterol Screening, Food Insecurity, Housing Insecurity, Loneliness, and No Social Support. This was the full feature set from the CDC PLACES; it included broader access-to-care and social-needs variables. A final version, called L + Region, added Census Region to the Large set. This was used to test whether geography added useful information after the health and social variables were already included.

Research and Design

4.1 Research Design

This project was designed as a comparative predictive analytics study. The main goal was to see if machine learning models could classify counties as either high or low disability burden. This was done separately for both Cognitive Disability and Mobility Disability. This means that the project is not just about building one model, but rather about comparing two related but different disability outcomes (Obermeyer & Emanuel, 2016). The study focused on using binary classification. For Mobility Disability, the model is meant to predict whether a county was in the high Mobility Disability group or the low Mobility Disability group. For Cognitive Disability, the same idea. Each target was modeled separately so that the results would indicate if the same predictors and models worked for both disability types.

The main comparisons in this study were Mobility Disability versus Cognitive Disability, Small versus Medium versus Large feature sets, and different machine learning model families and models. A separate test also compared the Large feature set with and without the Census Region. This was done to see whether Region added useful information after the health and social variables were already included. Multiple tools were implemented during the project. Data cleaning and preparation were done before the machine learning models were trained via RStudio. This step involved viewing and reorganizing the CDC PLACES data, checking the target variables, and removing duplicates. MATLAB Classification Learner was used for model training, and made it possible to test a plethora of machine learning models under the same system. Google Sheets was also used to organize model results and compare performance scores.

After the models were trained, the results were placed into spreadsheets so accuracy, Macro F1, Macro Recall, training time, and weighted rankings could be compared. Visualizations were created in Python from these results to help show model performance, feature set differences, and target differences.

4.2 Machine Learning Methods and Feature Selection

The study tested a broad group of machine learning methods. For the main paper, we focused on 20 model types to keep the methodology clear and readable. These models included decision trees, logistic models, Naive Bayes models, support vector machines, k-nearest neighbors models, ensemble models, and neural network models (Rajkomar et al., 2019). The 20 machine learning methods discussed in the paper were: Fine Tree, Medium Tree, Coarse Tree, Binary GLM Logistic Regression, Efficient Logistic Regression, Gaussian Naive Bayes, Kernel Naive Bayes, Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM, Coarse Gaussian SVM, SVM Kernel, Fine KNN, Weighted KNN, Bagged Trees, Boosted Trees, RUSBoosted Trees, and Wide Neural Network.

Tree models are often easier to understand because they split the data based on predictor values. Logistic models are simpler and can work well when the relationship between predictors and the target is more direct. Naive Bayes models are usually fast and simple. SVM models are useful when the classes are separated by more complex boundaries. KNN models classify observations based on nearby counties with similar predictor values. Ensemble models combine multiple models to improve performance. Neural networks can capture more complex patterns, although they are usually harder to explain (Obermeyer & Emanuel, 2016; Topol, 2019). Using different model families was important because the project did not assume one model type would be best. The goal was to test many models and then compare them using the same performance criteria.

Feature selection helped organize the predictors before modeling. The method used was mRMR, which stands for Minimum Redundancy Maximum Relevance (Peng et al., 2005). This method tries to select variables that are strongly related to the target variable, while also avoiding variables that repeat the same information as other predictors (Radovic et al., 2017). This was useful for this project because many county-level health variables are related to each other. The mRMR ranking helped organize the predictors into the Small, Medium, and Large feature sets described in the data preparation section. We hoped to test whether a smaller set of strong predictors could perform close to the full set of 24 numeric predictors. It also helped the study look at whether adding more variables actually improved model performance.

4.3 Design of Experiments and Validation

The Design of Experiments setup was used to test how model performance changed under different conditions. The project compared two disability targets, three feature set sizes, repeated model sessions, and one Region-based condition. The main experimental conditions were Mobility S, Mobility M, Mobility L, Cognitive S, Cognitive M, Cognitive L, and Mobility L + Region. The S condition used the Small feature set, the M condition used the Medium feature set, and the L condition used the full Large feature set. The L + Region condition used all twenty-four numeric predictors plus Census Region.

Each condition was repeated 5 times. This was done because one model run can sometimes produce results that are because of specific train-test splits or random variation. Repeating each condition made the results more reliable and made it less likely that the final findings were based on one lucky/good run. The same general validation and testing structure was used across the sessions. Without this setup, it would be harder to tell whether one model performed better because it was actually stronger or if it happened to be tested under easier conditions.

Each model used 5-fold cross-validation during model training and a test set for final performance reporting. Cross-validation was utilized because it gives the model several chances to train and validate on different parts of the data; again, this is meant to help compare models in a more reliable way (Obermeyer & Emanuel, 2016). The held-out test set was used after model training to measure how well the model performed on data it had not already used for training or validation. The analysis included this, as a model can sometimes perform well during training but fail on new data. The held-out test result gives a more honest view of how the model might perform on unseen counties. Using both cross-validation and held-out testing helped reduce overfitting. It also made the model comparison stronger because the models were judged using a similar process across the different targets and feature sets.

4.4 Performance Metrics and Weighted Framework

Several performance metrics were used to evaluate the models. Accuracy was included because it shows the overall percentage of correct predictions. While accuracy is useful, it was not enough by itself because the dataset had more low-disability counties than high-disability counties. Macro F1 Score was also used. Macro F1 balances precision and recall across both classes. This matters because the model should not only perform well on the larger low-disability class. It also needs to perform well on the smaller high-disability class.

Macro Recall was one of the most important metrics in this study. Recall is meant to measure how well the model identifies the true cases in a class. In this project, missing a high-disability county would be a major issue because the model would fail to identify a county that may need more public health attention. For that reason, Macro Recall was given more importance than accuracy alone (Rajkomar et al., 2019). Training Time was also included, and it measured how long the model took to train. Training Time was not the main priority, but it was useful to track regardless. A model that performs slightly better, but takes significantly longer to train, may not always be the best choice, especially when compute costs could be high. At the same time, because this is a public health-focused project, prediction quality was considered more important than speed.

Models were not selected based on just accuracy, because a model can have high accuracy, but still miss many high-disability counties. Since the high-disability class is the group the projects care most about identifying, weighting the scores was recommended. The final weights were Accuracy at 15%, Macro F1 at 35%, Macro Recall at 45%, and Training Time at 5%. Macro Recall received the highest weight because false negatives are costly in regard to public health(Obermeyer & Emanuel, 2016). If the model predicts a false negative that a county does not have high disability when it actually does, it could affect their funding, hospital access, and general statistics. Macro F1 received the second-highest weight because it gives a more balanced view of model performance across classes. Accuracy was still included, but it was not the main weight. Training Time received the lowest weight because speed matters, but it is less important than correctly identifying high-risk counties. This weighted system helped the project choose models that were not only accurate but also useful for the study.

4.5 Z-Score Normalization and Feature Importance

Z-score normalization was used to compare model results across different performance metrics. This was needed because the metrics were not all on the same scale. Accuracy, Macro F1, Macro Recall, and Training Time cannot be combined directly without first standardizing them. For Accuracy, Macro F1, and Macro Recall, higher values were better, while lower values are better for Training Time. After the metrics were standardized, the weighted scores were combined into one overall score for each model. This made it easier to rank the models across the different feature sets and targets. The final ranking helped show which models had the best overall balance of correctness, class-balanced performance, recall, and efficiency. Feature importance was used to understand which predictors mattered most in the final models. This was important because the project was not only about getting a high accuracy score. It was also about understanding whether Mobility Disability and Cognitive Disability were driven by different health, social, or other predictors within the dataset.

Results

5.1 Model Performance

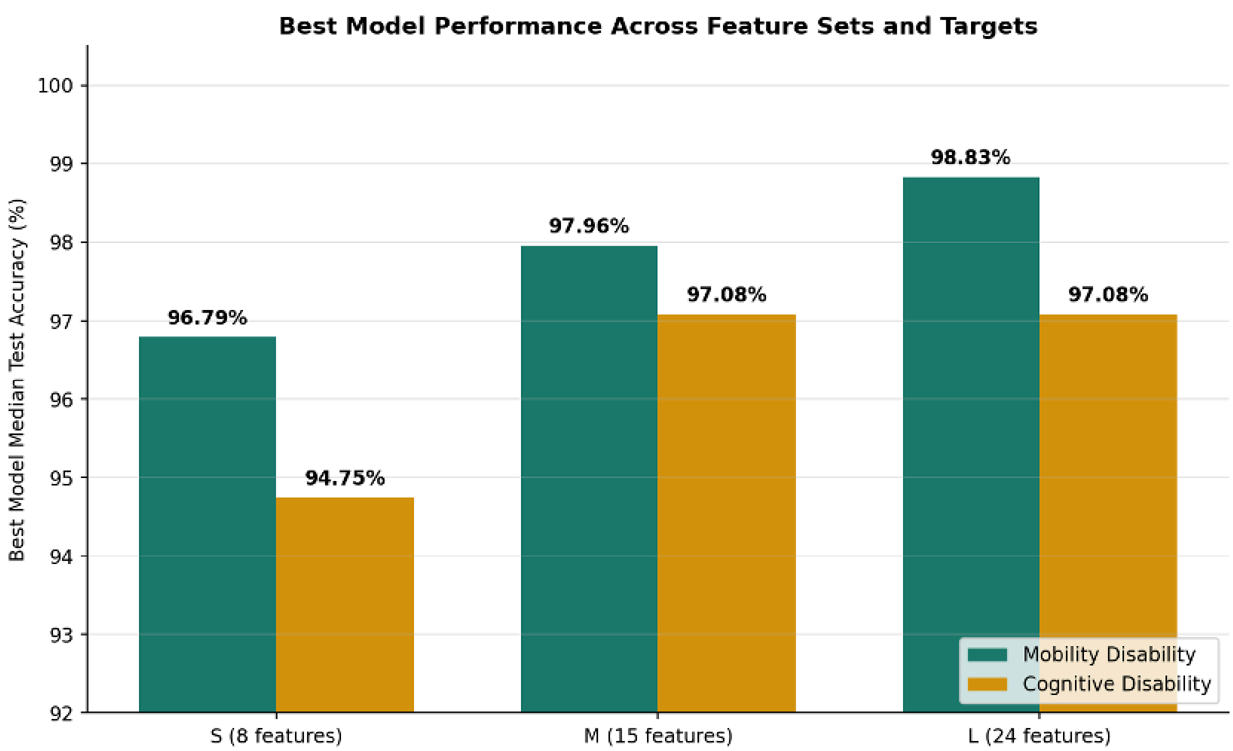

Figure 1. Best model performance across feature tiers

Our best model median test accuracy ranged from 94.75% (Cognitive S, Cubic SVM) to 98.83% (Mobility L, Cubic SVM). Mobility was more predictive across every single feature tier. Because the target classes aren't balanced, with approximately 75% of counties in the low disability class and 25% in the high disability class, accuracy alone was not sufficient. A majority Class classifier could have around 75% accuracy without identifying any high disability counties. Due to this, Macro F1 and Macro Recall were heavily emphasized in model comparison. The Mobility to Cognitive accuracy gap ranged from 0.88 percentage points in the medium feature set to 2.04 percentage points in the small feature set. The Large feature set showed a 1.75 point gap.

Feature Set	Mobility (best model)	Cognitive (best model)	Δ
S (8 feat)	96.79% - Cubic SVM	94.75% - Cubic SVM	+2.04
M (15 feat)	97.96% - Wide NN / Quad SVM	97.08% - Weighted KNN	+0.88
L (24 feat)	98.83% - Cubic SVM	97.08% - Wide NN	+1.75

Figure 2. Best model performance across feature tiers shown with numerical differences

5.2 Model Comparison Heatmap

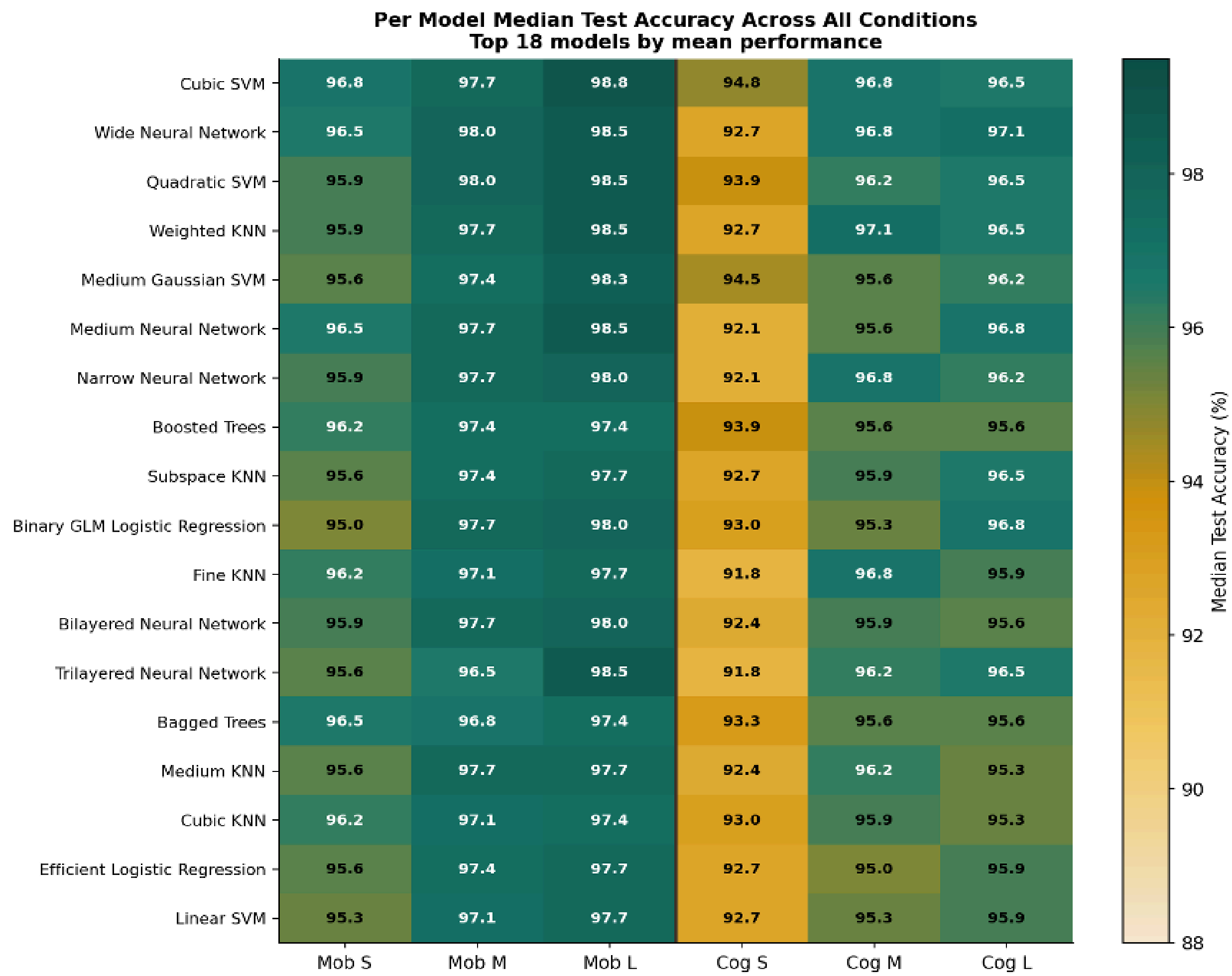

Figure 3. Per-Model test accuracy across all 6 main conditions. Top 18 models (accuracy) are shown

This Heat map visualizes how the top 18 models performed across both Mobility, and Cognitive Disability and all three feature tiers. Darker color means a higher accuracy, while lighter color represents less accuracy. The black line down the middle separates Mobility and Cognitive. More Complex models like SVMs, Neural Networks, and KNN, cluster at the top with very consistent high accuracy. More simple models such as Naive Bayes and Trees show a decrease in accuracy as feature count decreases. You can observe that the same model families tend to perform well for both Mobility and Cognitive.

5.3 Value of Adding Features

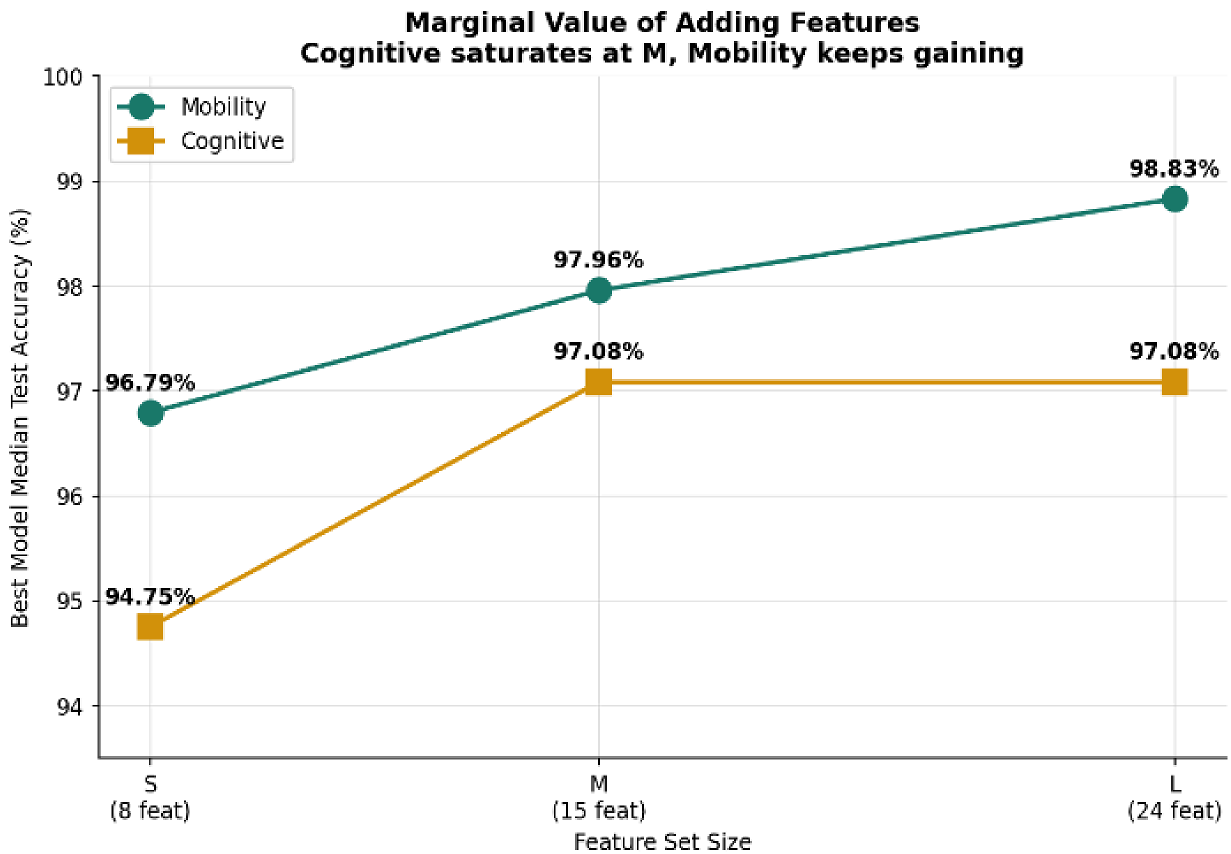

Figure 4. Best model accuracy as a function of feature set size

The accuracy gains for adding additional features were relatively consistent. For Mobility, accuracy gains increased +1.17 points from S to M, +.087 points from M to L. Cognitive showed a +2.33 point gain from S to M but then showed no gain from M to L. This shows that the 9 additional features in L (binge drinking, poor general health, no health insurance, annual checkup, cholesterol screening, food insecurity, housing insecurity, loneliness, lack of social support) collectively added no predictive value for Cognitive Disability, even though they did add value for mobility. This suggests that the key predictors of Cognitive Disability were all in the small and medium feature set, so the additional access to care and social factors don't provide any new information needed for Cognitive Disability.

5.4 Region as a Feature Has no Benefit

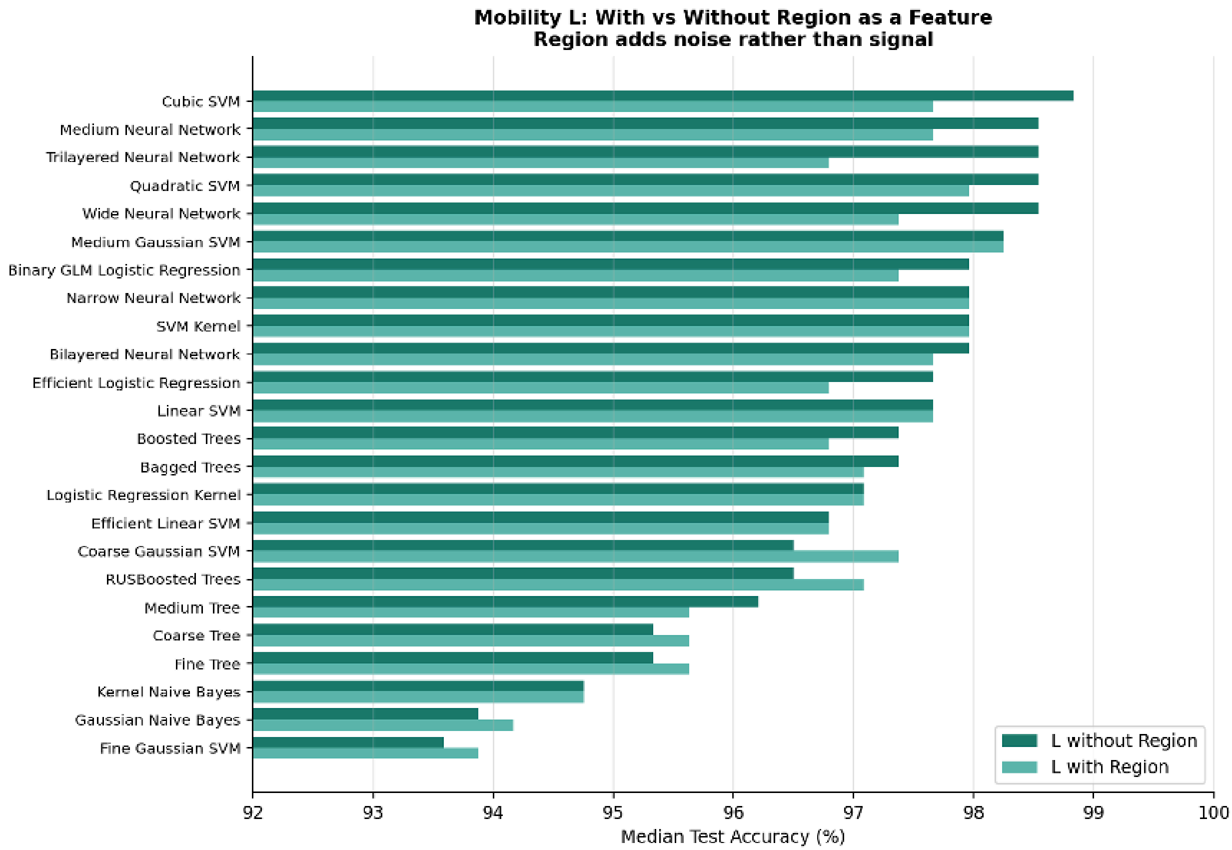

Figure 5. Mobility runs with Region vs. without

On the 20 models that ran in both conditions, including Region as a categorical feature decreased the best model median test accuracy by 0.58 percentage points (98.83 to 98.25) and decreased the mean of the top 5 model margins by the same margin. Not one model increased by more than 0.5 from including Region. The reason for this is because things like smoking rates, diabetes, and physical inactivity, already gives a clear enough picture of where a county is. These health factors already capture the differences between regions. Adding “Region” is repetitive and repeats information that the model already knows. However, Region may still be useful for interpreting findings, such as identifying variation in disability rates across regions.

5.5 Run to Run Stability

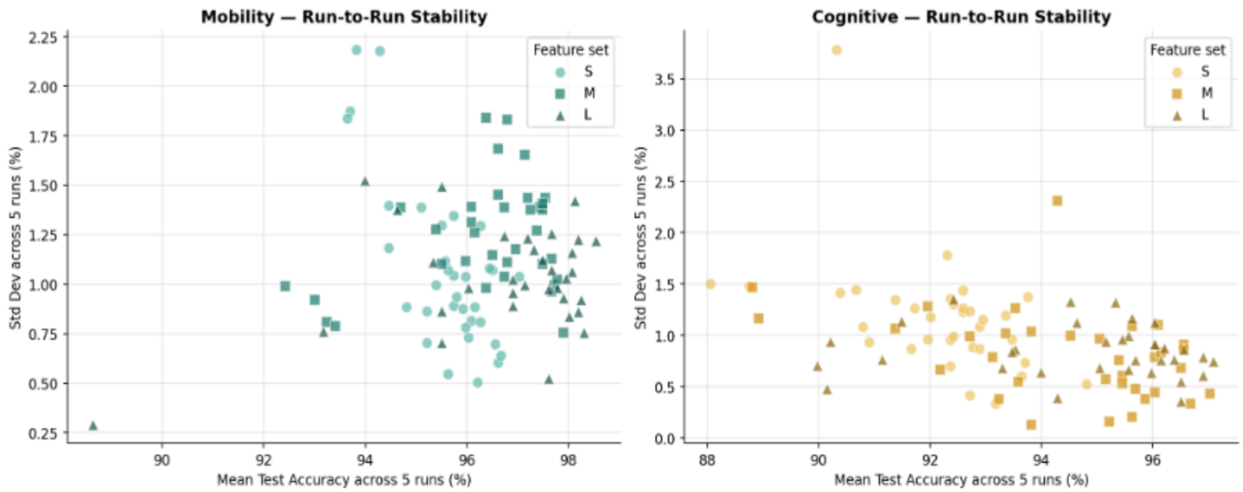

Figure 6. Standard deviation of test accuracy across 5 runs, ideally located bottom right (high accuracy, low variance)

Across all setups , most of the models were consistent. The results only changed about 1 percentage point. The best models (Cubic SVM, Wide Neural Network, Linear SVM, and Weighted KNN) were very stable. They performed about the same every time they were run. Some models (Efficient Linear SVM, Logistic Regression Kernel, and Naive Bayes) had more variation. The Cognitive models were more consistent, but still not accurate overall.

5.6 Predictor Correlation Matrix

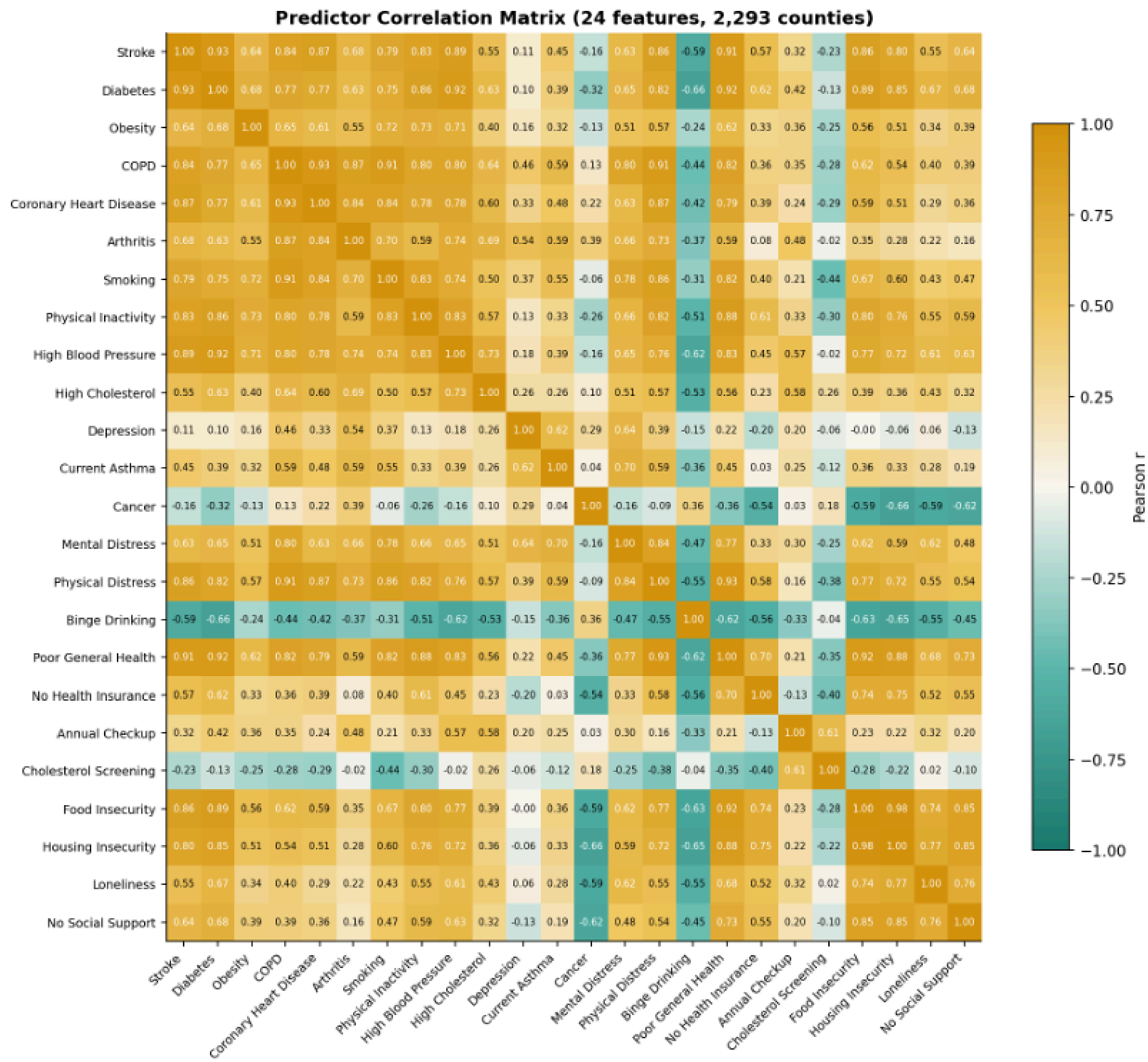

Figure 7. Predictor correlation across the 24 predictors

This heatmap shows the correlation structure of our predictor panel. A large physical health feature group (Stroke, Diabetes, COPD, Coronary Heart Disease, High Blood Pressure) show correlations of over .7, this indicates substantial multicollinearity. If one of these is high, the others usually are too. Another group is behaviors and mental health (smoking, inactivity, stress, depression) these tend to go together. Social issues (housing problems, loneliness, lack of support) are another group with strong correlation. Cancer is the most distinctive feature in this panel, showing weak and sometimes even negative correlations with several of the physical measures. This can be because cancer prevalence increases with age, this means it's higher in healthy aging populations.

5.7 Divergent Drivers

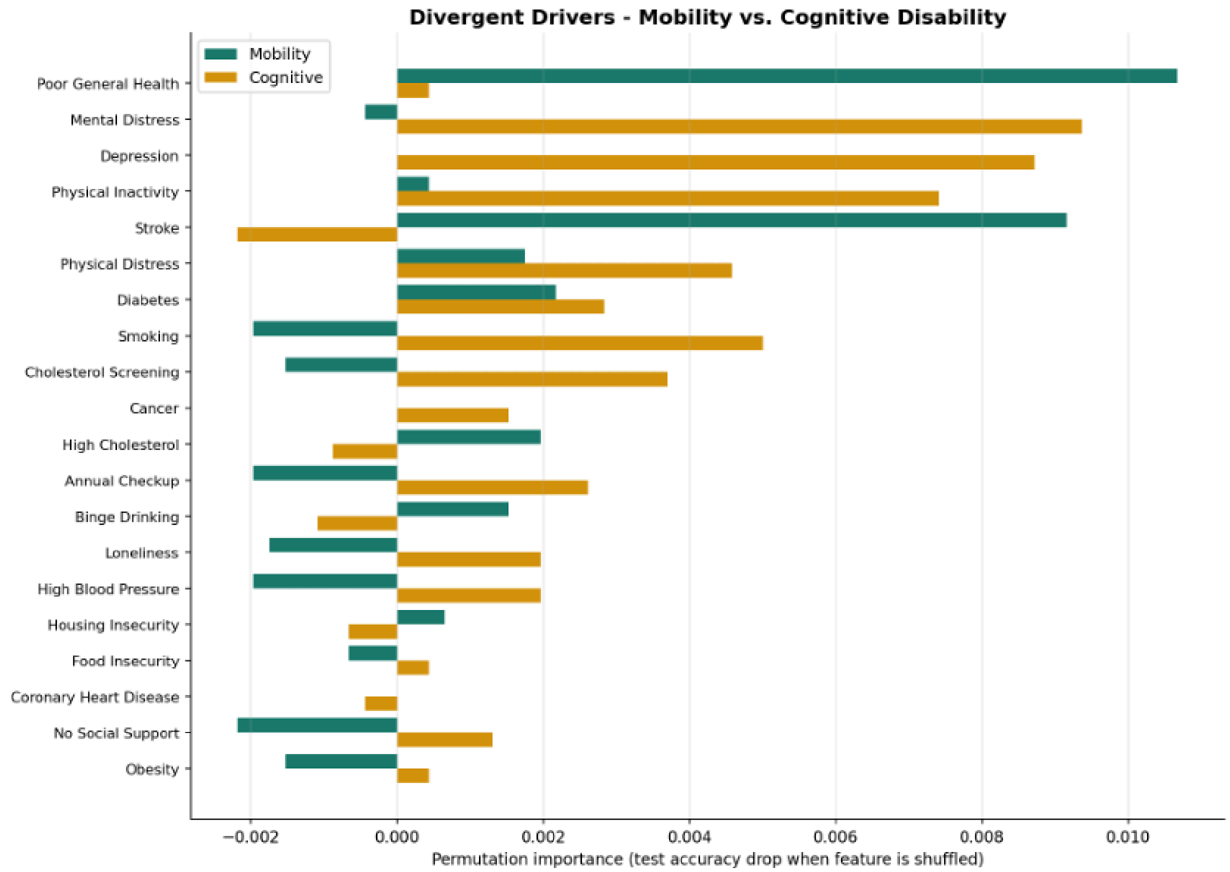

Figure 8. Side by side comparison of permutation importances - Top features for Mobility are mainly irrelevant for Cognitive, same the other way around

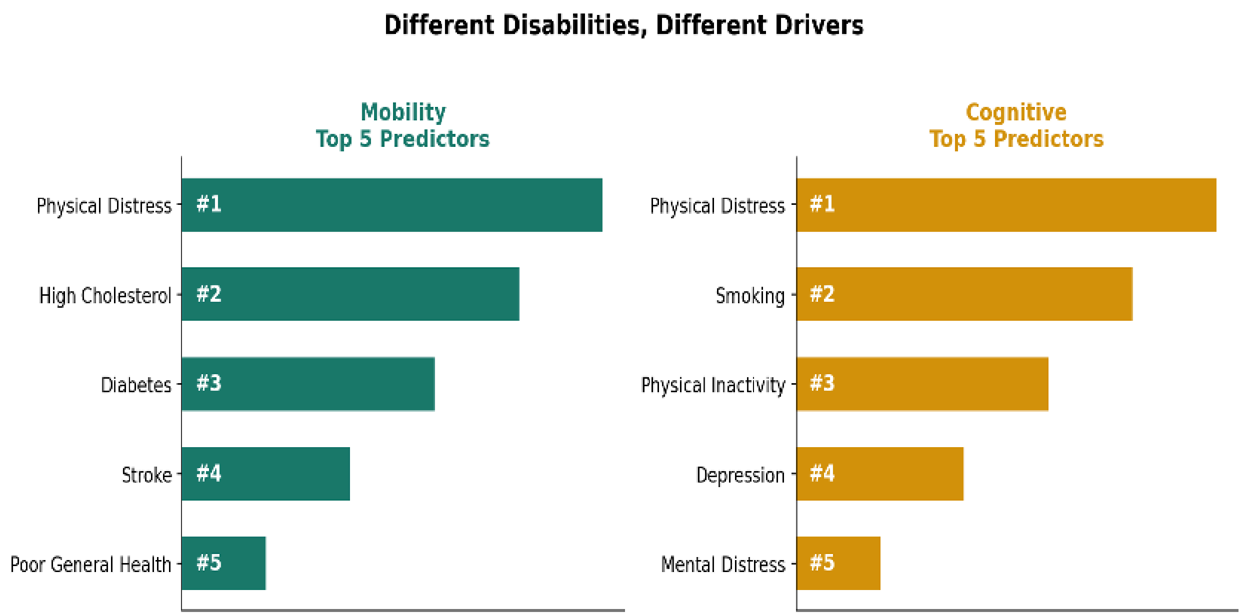

Figure 9. A clear comparison of the top 5 predictors of Mobility and Cognitive Disability

The two targets Mobility and Cognitive have almost no overlap in their top predictors.

Mobility's top features are: Poor General Health, Stroke, Diabetes, High Cholesterol, Physical Distress, Binge Drinking. These are physical and cardio-metabolic indicators that we can expect to cause walking and stair climbing difficulty. Cognitive's Top features are: Mental Distress, Depression, Physical Inactivity, Smoking, Physical Distress, Cholesterol Screening, Diabetes, Annual Checkup, High Blood Pressure, and Loneliness. These are mental health, behavior, and social isolation indicators. Of the top 5 predictors of each target binary variable (cognitive and mobility) the only feature that appears in both is Physical Distress, and that ranks higher for Cognitive than for Mobility. Stroke is the second most important predictor feature for Mobility, but ranks near 0 for Cognitive. This is the central finding of the project. Even though the data is the same, the important factors change based on which type of disability you are predicting.

5.8 Geographic Distribution

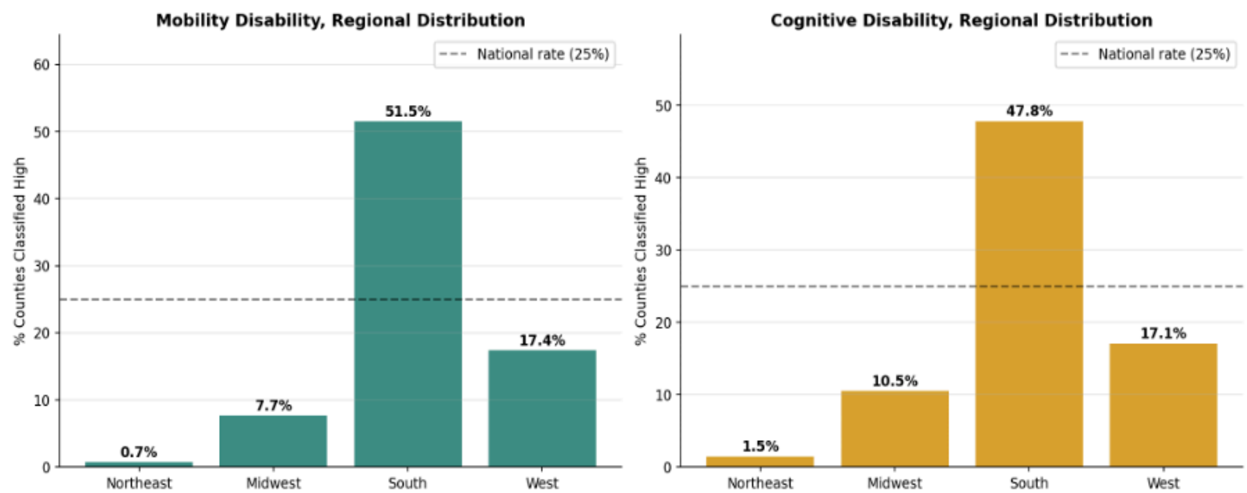

Figure 10. Percentage of counties classified as "high disability" by Region

Both of our targets show a strong regional concentration. For Mobility, the South has 51% high disability counties. This is more than seven times the Midwest rate (7.7%) and almost seventy times the Northeast rate (0.7%). The West is in the middle at 17.4%. The Cognitive distribution is similar but not as extreme. The South sits at 47.8%, West 17.1%, Midwest 10.5%, Northeast 1.5%. This is more concentrated than what would be expected from a normal regional distribution. The South is producing more than half of all of the top quartile Mobility Disability counties, despite having only 38% of the datasets.

5.9 State Level Patterns

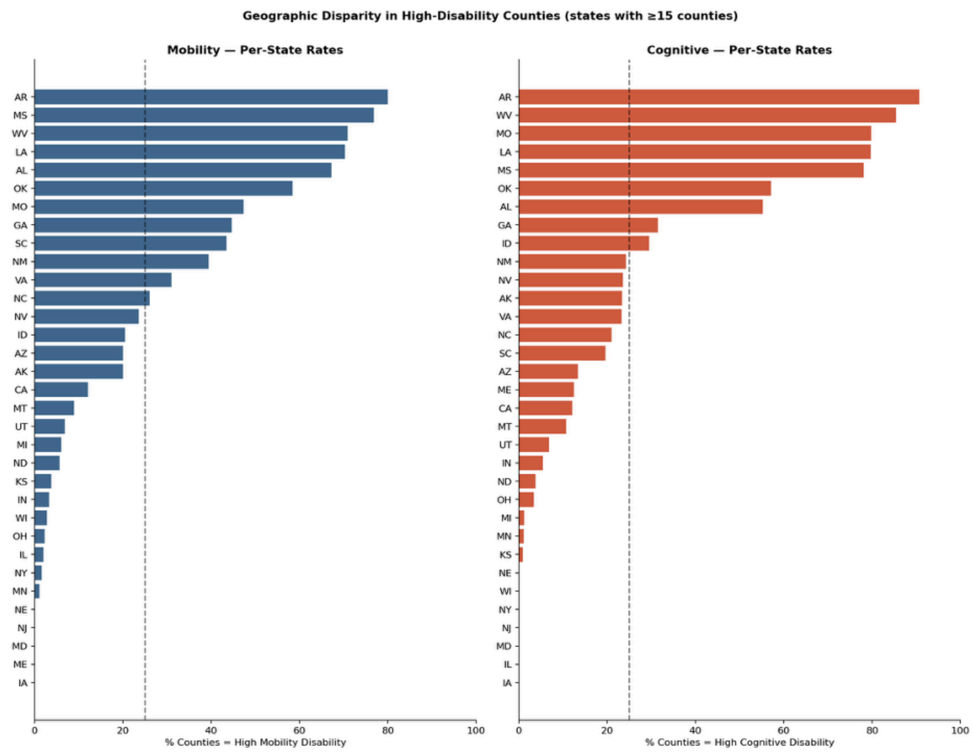

Figure 11. Percentage of counties classified as high disability by state, only states with >15 counties shown

The states with the most disability are mostly in the South and Appalachian (Arkansas, Mississippi, and West Virginia). Some states (Iowa, New York) have almost none. This is not just based on location. These high disability health areas have worse health overall. They have more chronic disease, unhealthy behaviors, and social challenges. This suggests lifestyle and mental health interventions may be valuable targets.

Discussion and Public Health Implications

6.1 Public Health Implications

Models like ours can be used to help health agencies identify counties that need earlier intervention. Counties projected to have high Mobility Disability prevalence are likely to benefit from programs targeting cardiovascular, and metabolic health, including stroke prevention, physical activity programs, diabetes management, COPD support, cardiovascular screening, and arthritis support. Counties projected to have high Cognitive Disability prevalence point toward interventions focused on mental health and social connection, including depression screening, social support, smoking limitations, physical activity, and mental health services to reduce social isolation. Treating disability as a single category would combine these differences and risk putting resources in interventions that do not match the drivers of disability in a community. The models can also be a practical tool for closing data gaps when complete PLACES estimates aren't available for every county.

6.2 Limitations

Approximately one quarter of the U.S. counties were left out due to missing data, some states had more missing than others. The results may not apply the same way in these areas. The analysis also relies on a single year of the PLACES data from 2023, which constrains the claims that can be drawn from findings and the certainty that models will apply to future years. Defining High Disability as the top 25% of counties simplifies a complex, continuous problem into a binary classification. Finally, many of the predictors are correlated with one another, which complicates the interpretation of the feature importance scores. A high ranking predictor may be capturing variance shared with several related variables rather than acting as an independent driver of disability.

Conclusion

7.1 Summary of What Was Done

Using county-level data from the CDC PLACES project, models were developed to predict the proportion of the population with Mobility Disability and Cognitive Disability. Two binary target variables were created that identify counties above the 75th percentile as having a high disability burden. The models incorporate a number of potential predictors such as the prevalence of chronic diseases, behavioral risk factors, mental and physical distress, and access to medical care for diabetes, treatment of depression, and other health needs.

To improve the accuracy and interpretability of the models, mRMR, 20 machine learning models were trained and tested on these feature sets. A 5-fold cross-validation was performed on each model as well as testing on a held-out test set. Model performance was evaluated using Accuracy, Macro F1 Score, Macro Recall, and Training Time. A weighted ranking system was implemented to rank and select the top performing models based on a combination of predictive performance and training efficiency.

7.2 Summary of Key Findings

Using machine learning methods, projecting disability burden at the county level is a feasible goal. The best performing model for Mobility Disability had an accuracy of approximately 98.83%, and the best Cognitive Disability model had an accuracy of approximately 97.08%.

Results showed that Mobility and Cognitive Disability were predicted by different sets of factors. Mobility Disability was most strongly associated with the physical health and chronic disease variables whereas Cognitive Disability was most strongly associated with the mental health, behavioral risk factors, and social determinants. The variable Census Region did not add to the model beyond the health and social variables. Future increments in the number of variables may be associated with decreasing gains, consistent with the performance of the model in the past.

References

Arthritis Foundation. (2022). *Arthritis and disability statistics*. https://www.arthritis.org

Benjamin, E. J., et al. (2019). Heart disease and stroke statistics—2019 update. *Circulation, 139*(10), e56–e528. https://doi.org/10.1161/CIR.0000000000000659

Berkowitz, S. A., et al. (2018). Food insecurity and health care expenditures. *Health Services Research, 53*(3), 1600–1620. https://doi.org/10.1111/1475-6773.12730

Booth, F. W., Roberts, C. K., & Laye, M. J. (2012). Lack of exercise is a major cause of chronic diseases. *Comprehensive Physiology, 2*(2), 1143–1211. https://doi.org/10.1002/cphy.c110025

Centers for Disease Control and Prevention. (2024). *PLACES: Local data for better health*. https://www.cdc.gov/places

Centers for Disease Control and Prevention. (2024). *PLACES: Measure definitions—Disability*. https://www.cdc.gov/places/measure-definitions/disability.html

Centers for Disease Control and Prevention. (2025). *About chronic diseases*. https://www.cdc.gov/chronic-disease/about/index.html

Centers for Disease Control and Prevention. (2025). *Fast facts: Health and economic costs of chronic conditions*. https://www.cdc.gov/chronic-disease/data-research/facts-stats/index.html

Centers for Disease Control and Prevention. (2023). *Why is addressing social determinants of health important?* https://www.cdc.gov/about/priorities

Gale, C. R., et al. (2018). Loneliness, social isolation, and cognitive function. *Journal of Alzheimer's Disease, 63*(2), 669–681. https://doi.org/10.3233/JAD-170655

Krahn, G. L., Walker, D. K., & Correa-De-Araujo, R. (2015). Persons with disabilities as an unrecognized health disparity population. *American Journal of Public Health, 105*(S2), S198–S206. https://doi.org/10.2105/AJPH.2014.302182

Livingston, G., et al. (2020). Dementia prevention, intervention, and care: 2020 report. *The Lancet, 396*(10248), 413–446. https://doi.org/10.1016/S0140-6736(20)30367-6

Lu, H., Greenlund, K. J., Lu, Y., & Croft, J. B. (2023). County-level geographic disparities in disabilities among U.S. adults, 2019. *Preventing Chronic Disease, 20*, E04. https://doi.org/10.5888/pcd20.230004

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data and machine learning in health care. *New England Journal of Medicine, 375*(13), 1216–1219. https://doi.org/10.1056/NEJMp1606181

Office of Disease Prevention and Health Promotion. (n.d.). *Social determinants of health*. https://odphp.health.gov/healthypeople

Okoro, C. A., et al. (2018). Prevalence of disabilities and health care access by disability status and type among adults—United States. *Morbidity and Mortality Weekly Report, 67*(32), 882–887. https://doi.org/10.15585/mmwr.mm6732a3

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(8), 1226–1238. https://doi.org/10.1109/TPAMI.2005.159

Radovic, M., et al. (2017). Minimum redundancy maximum relevance feature selection approach. *BMC Bioinformatics, 18*(1), 1–14. https://doi.org/10.1186/s12859-017-1825-1

Rajkomar, A., et al. (2019). Machine learning in medicine. *New England Journal of Medicine, 380*(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

Richardson, T., et al. (2017). The impact of loneliness on mental health. *Clinical Psychology Review, 48*, 1–15. https://doi.org/10.1016/j.cpr.2016.05.002

Taylor, L. A., et al. (2016). Leveraging the social determinants of health. *American Journal of Preventive Medicine, 51*(6), 1022–1033. https://doi.org/10.1016/j.amepre.2016.08.036

Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books.

Verbrugge, L. M., & Jette, A. M. (1994). The disablement process. *Social Science & Medicine, 38*(1), 1–14. https://doi.org/10.1016/0277-9536(94)90294-1

World Health Organization. (2021). *Depression*. https://www.who.int