# Explainable Machine Learning and Social Determinants of Health in Stroke Prediction

*Gauri Sharma*

Department of Electrical & Computer Engineering

McGill University

Montréal, Québec, Canada

August, 2024

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Master of Science in Electrical Engineering

# Abstract

**Background:** Current stroke prediction models, relying solely on traditional medical data, overlook the role of Social Determinants of Health (SDoH) like socioeconomic status and education. This narrow focus can lead to inaccurate predictions, potentially exacerbating healthcare disparities and hindering the development of effective preventive measures. This work investigates the role of SDoH in stroke and how incorporating SDoH data into AI models can improve stroke prediction, ultimately empowering healthcare providers with a more holistic view of patient risk for better decision-making and equitable healthcare delivery.

**Research Objectives:**

1. To improve the performance of stroke prediction AI models by integrating SDoH into these models.
2. To ensure transparency and interpretability in stroke prediction through the application of explainable AI (XAI) methodologies.

**Method:** The study employs datasets from the Institut de la statistique du Québec that include both clinical indicators (e.g. diabetes, heart disease, weight) and SDoH (e.g.economic, neighbourhood conditions). We applied seven machine learning models (Random Forest), Gradient Boosting Machine (GBM), CatBoost (CB), XGBoost (XGB), Light Gradient Boosting Machine (LGBM), Neural Networks (NN), and K-Nearest Neighbors (KNN) alongside XAI techniques to investigate the role SDoH plays in the models' predictive performances. XAI methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) were implemented, shedding light on the influence of SDoH in the algorithms' predictions. Performance of models was evaluated using standard metrics such as accuracy, precision, recall, F1 score and AUC (Area under the curve).

**Results:** Our study investigated the impact of incorporating SDoH data into stroke prediction models. SDoH data variably improved performance depending on the model and specific SDoH factors incorporated, illustrating its important role alongside traditional medical data in assessing stroke risk. Our LGBM model showed maximum improvement on incorporation of SDoH features where its accuracy improved by 11.2% (from 65.9% to 77.1%). The inclusion of demographic, economic, and personal SDoH factors were the most influential. XAI methods revealed self-perceived health and stress levels as key factors for stroke prediction, emphasizing the importance of personal well-being in stroke assessment. Notably, the Light Gradient Boosting Machine (LGBM) model achieved the best performance, demonstrating an Area Under the Curve (AUC) of 81%. This translates to accuracy of 77.6%, precision of 78.6%, recall of 75.5%, and F1 score of 77.0%, showcasing LGBM's proficiency in handling the complex relationships within SDoH data. These findings suggest the importance and potential of SDoH-integrated AI models for improved stroke prediction.

**Conclusion:** Our findings highlight the role of SDoH data in building accurate and equitable healthcare models. Integrating SDoH factors improve stroke prediction accuracy by 1% to 3%, and foster fairer and more comprehensive patient risk assessments by considering the broader social and environmental influences on health. Furthermore, XAI techniques provide deeper insights into how SDoH and other factors contribute to predictions, promoting transparency and interpretability in these AI-driven solutions. This transparency is essential for building trust and ensuring ethically sound decision-making in healthcare.

# Abrégé

**Contexte:** Les modèles actuels de prédiction des AVC, qui reposent uniquement sur des données médicales traditionnelles, négligent le rôle des déterminants sociaux de la santé (DSS) tels que le statut socioéconomique et l'éducation. Cette focalisation étroite peut entraîner des prédictions inexactes, potentiellement exacerber les disparités en matière de santé et entraver le développement de mesures préventives efficaces. Ce travail examine le rôle des DSS dans les AVC et comment l'intégration des données DSS dans les modèles d'IA peut améliorer la prédiction des AVC, en donnant aux prestataires de soins de santé une vue plus holistique du risque des patients pour une meilleure prise de décision et une prestation de soins plus équitable.

**Objectifs de recherche:**

1. Améliorer la performance des modèles d'IA de prédiction des AVC en intégrant les DSS dans ces modèles.

2. Assurer la transparence et l'interprétabilité de la prédiction des AVC par l'application de méthodologies d'IA explicable (XAI).

**Méthode:** L'étude utilise des ensembles de données de l'Institut de la statistique du Québec incluant à la fois des indicateurs cliniques (par exemple, diabète, maladies cardiaques, poids) et des DSS (par exemple, conditions économiques et de voisinage). Nous avons appliqué sept modèles d'apprentissage automatique (Random Forest, Gradient Boosting Machine (GBM), CatBoost (CB), XGBoost (XGB), Light Gradient Boosting Machine (LGBM), réseaux de neurones (NN) et K-Nearest Neighbors (KNN) ainsi que des techniques XAI pour examiner le rôle des DSS dans les performances prédictives des modèles. Les méthodes XAI telles que SHAP (SHapley Additive exPlanations) et LIME (Local Interpretable Model-agnostic Explanations) ont été mises en œuvre, éclairant l'influence des DSS dans les prédictions des algorithmes. La performance des modèles a été évaluée en utilisant des métriques standard telles que la précision, la sensibilité, la spécificité, le score F1 et l'AUC (aire sous la courbe).

**Résultats:** Notre étude a examiné l'impact de l'intégration des données DSS dans les modèles de prédiction des AVC. Les données DSS ont amélioré les performances de manière variable en fonction du modèle et des facteurs DSS spécifiques incorporés, illustrant leur rôle important aux côtés des données médicales traditionnelles dans l'évaluation du risque d'AVC. Notre modèle LGBM a montré une amélioration maximale avec l'incorporation des caractéristiques DSS, où sa précision a augmenté de 11,2 % (de 65,9 % à 77,1 %). L'inclusion de facteurs DSS démographiques, économiques et personnels a été la plus influente. Les méthodes XAI ont révélé que la santé perçue et les niveaux de stress étaient des facteurs clés pour la prédiction des AVC, soulignant l'importance du bien-être personnel dans l'évaluation des AVC. Notamment, le modèle Light Gradient Boosting Machine (LGBM) a obtenu la meilleure performance, démontrant une aire sous la courbe (AUC) de 81 %. Cela se traduit par une précision de 77,6 %, une sensibilité de 75,5 %, une spécificité de 78,6 % et un score F1 de 77,0 %, montrant la capacité du LGBM à gérer les relations complexes au sein des données DSS. Ces résultats suggèrent l'importance et le potentiel des modèles d'IA intégrant les DSS pour une meilleure prédiction des AVC.

**Conclusion:** Nos résultats soulignent le rôle des données DSS dans la construction de modèles de santé précis et équitables. L'intégration des facteurs DSS améliore la précision de la prédiction des AVC de 1 % à 3 % et favorise des évaluations des risques des patients plus justes et plus complètes en tenant compte des influences sociales et environnementales sur la santé. De plus, les techniques XAI fournissent des informations plus approfondies sur la contribution des DSS et d'autres facteurs aux prédictions, favorisant la transparence et l'interprétabilité de ces solutions pilotées par l'IA. Cette transparence est essentielle pour instaurer la confiance et garantir une prise de décision éthique en matière de soins de santé.

# Acknowledgements

I am immensely grateful for the guidance and support I received throughout my Master's thesis at the Department of Electrical and Computer Engineering, McGill University. I extend my deepest thanks to my supervisor, Prof. AJung Moon, and co-supervisor, Prof. Samira A. Rahimi. Their expertise, patience, and insights were not only invaluable to my research but also instrumental in shaping me into a better researcher and individual.

My family has been my foundation and my strength. I am deeply thankful to my parents for their endless love and unwavering support. To my sister, Bubbo, whose encouragement and companionship have been my constant source of motivation, thank you. I am also profoundly grateful to my extended family—uncles, aunts, cousins—who have always been there for me. Your enthusiasm for my studies, your calls, and your pride in my accomplishments have buoyed me through this journey. I also appreciate the simple joys brought by my dogs, whose presence, even from afar, was a comforting constant during these years.

I am truly fortunate to have friends who brought laughter and light into the most stressful days. Their friendship sustained me, lightening my burdens and brightening my mood. They were always there with a joke when I needed a lift or a word of encouragement when things got tough. I'm deeply thankful for their support throughout this journey.

The RAISE Lab has been more than just a part of my academic journey—it's been like a second family. The supportive and creative energy from everyone in the lab truly transformed it into a place where I not only learned but also felt at home. Their encouragement and camaraderie extended my learning and personal growth in ways I hadn't imagined possible.

*This thesis is dedicated to my grandparents, whose values and teachings have profoundly shaped who I am today. I am forever grateful for your love, blessings, wisdom and guidance. Thank you for always looking upon me.*

# Contents

# List of Figures

# List of Tables

# List of Acronyms

**AI** Artificial Intelligence.

**CB** CatBoost.

**CVD** Cardiovascular Diseases.

**GBM** Gradient Boosting Machines.

**KNN** K-Nearest Neighbors.

**LGBM** LightGBM.

**LIME** Local Interpretable Model-agnostic Explanations.

**ML** Machine Learning.

**NN** Neural Network.

**SDoH** Social Determinants of Health.

**SHAP** SHapley Additive exPlanations.

**XAI** Explainable Artifical Intelligence.

**XGB** XGBoost.

**XML** Explainable machine learning.

# Chapter 1

# Introduction

## 1.1 Current Challenges in AI for Healthcare

Stroke, an event where blood flow to the brain is disrupted, remains a global health threat. It's the second leading cause of death worldwide [1], and each year, millions experience strokes, with many facing long-term impairments [2]. However, the risk of stroke isn't equally distributed. Social determinants of health (SDoH) – the social and economic factors where we live, work, and age – significantly influence stroke risk [3, 4]. Studies have shown that factors like low income, lack of access to healthy food and quality education, and inadequate housing are linked to a higher risk of stroke [5]. This increased risk underscores the need for healthcare interventions that address these underlying social inequities. This research explores the considerations surrounding the use of artificial intelligence (AI) in stroke prediction, with a specific focus on integrating SDoH data into these models. By incorporating this crucial information, the research aims to promote fairness and equity in healthcare decision-making related to stroke prevention and treatment.

In recent years, AI and data-driven technologies have proliferated across various sectors, including healthcare, promising to revolutionize disease diagnosis, drug development and healthcare [6, 7]. For example, DeepMind's AI has shown application for early detection of eye diseases thus improving diagnostic accuracy and patient outcomes [8]; IBM Watson's oncology support system provides evidence-based treatment options, showcasing how AI can assist in complex treatment planning [9]. These advancements underscore AI's transformative impact on healthcare, offering new avenues for enhancing patient care [7] and operational efficiency [10]. However, the rapid advancement of AI in healthcare has raised significant ethical concerns such as fairness, transparency and privacy of AI tools, informed consent, and the potential for increased healthcare disparities [11].

The healthcare sector is increasingly reliant on AI for complex tasks like early disease prediction and risk assessment. This growing dependence necessitates thorough ethical considerations throughout the development and deployment of AI-powered healthcare technologies [12]. The recent global health crises have shone a harsh light on pre-existing healthcare disparities, often linked to the SDoH [13]. These disparities highlight the need for equitable healthcare interventions that ensure everyone has a fair and just opportunity to achieve their full health potential [14].

Research suggests that an equitable healthcare system would strive to eliminate these disparities by focusing on several key areas. First, it would prioritize accessibility. Everyone, regardless of income, geographic location, race, ethnicity, or other social factors, should have affordable and timely access to high-quality preventive, diagnostic, and treatment services [15]. This might involve implementing universal health coverage or subsidized insurance plans for low-income individuals, alongside geographically accessible clinics and hospitals [16].

Second, an equitable system would ensure quality of care for all. This means delivering evidence-based, culturally competent, and patient-centered care to every individual [17]. In Canada, this includes integrating Indigenous health practices and perspectives into the healthcare system and ensuring universal access to healthcare regardless of immigration status [18]. This could involve diversifying the healthcare workforce to better reflect the communities served, and ensuring language accessibility for patients with limited English proficiency [17].

Finally, an equitable system wouldn't just treat illness, it would address the root causes of health disparities. This necessitates a focus on the SDoH, such as safe housing, quality education, and healthy food options [19]. Collaboration between healthcare providers and social service organizations, or policies promoting income security and affordable housing, could be key strategies in achieving this [3]. By prioritizing these core principles, a healthcare system can move towards ensuring everyone has the chance to live a healthy life.

This work utilizes AI to enable equitable healthcare, focusing on stroke prediction. We employ a diverse range of machine learning models – Random Forest (RF) [20], Gradient Boosting Machines (GBM) [21], K-Nearest Neighbors (KNN), Neural Network (NN) [22], LightGBM (LGBM) [23], CatBoost (CB) [24], and XGBoost (XGB) [25], each chosen for their effectiveness in binary classification tasks and their ability to capture complicated relationships within the data. The study aims to shed light on potential disparities in healthcare outcomes by integrating SDoH factors into AI models, thereby contributing to a more equitable healthcare system. It emphasizes the ethical imperative of fairness, transparency and equity by identifying systemic inequalities that contribute to disparate health outcomes among different populations.

In essence, this research is a response to the growing global need for equitable healthcare access and delivery of health-related services. The research aims to explore if the inclusion of SDoH factors into AI models improves the accuracy of stroke prediction and which SDoH factors contribute most to the prediction task by using explainable AI (XAI) techniques.

## 1.2 Research Questions & Objectives

The research is guided by the following key questions:

1. Does the inclusion of SDoH factors improve prediction of stroke?

2. Which specific SDoH factors are especially pertinent in the prediction of stroke?

This study aims to refine stroke prediction by incorporating SDoH into ML models, focusing on accuracy enhancement and leveraging XAI techniques to identify the most impactful SDoH factors. This dual approach seeks to improve model accuracy while ensuring transparency and explainability in predictions.

- **Objective 1: Enhance Stroke Prediction Accuracy by Incorporating SDoH**

  The first objective focuses on improving the accuracy of stroke risk predictions by incorporating SDoH factors—such as socioeconomic status, education, and environmental conditions—into ML models. These factors are recognized for their significant influence on health outcomes and are expected to contribute to more precise predictive analytics.

- **Objective 2: Identify Impactful SDoH Factors Using XAI Techniques**

  The second objective aims to leverage XAI techniques alongside our ML models to pinpoint which SDoH factors affect stroke the most. This approach is designed to enhance the explainability and transparency of our predictive models, providing clear insights into how specific social determinants influence stroke predictions.

By addressing these objectives, we aim to contribute to the fields of healthcare equity, AI ethics, and predictive modeling.

## 1.3 Contribution of Thesis

This thesis presents two contributions towards more equitable and ethically advanced healthcare practices.

- **Advancing Healthcare Equity:**

  A significant motivation behind this research is its prospective contribution to healthcare equity. In numerous regions globally, the quality of healthcare access is uneven, often swayed by SDoH like income, education, and environmental factors. The stroke prediction model in this work integrates these determinants and provides a clearer understanding of healthcare disparities. By identifying these disparities through AI models, there is potential to inform strategies that mitigate healthcare outcome inequities. Consequently, this could facilitate more equitable healthcare provision for individuals from underserved communities, taking a step towards reducing global health disparities.

- **Explainable AI in Healthcare:**

  In today's landscape, being able to explain meaningful patterns behind AI outcomes in healthcare is crucial. This research explores fairness and transparency considerations through post-hoc XAI techniques. By employing XAI, this research contributes to an environment where AI technologies in healthcare are grounded on a strong ethical foundation, by showcasing the importance of explainability in AI-driven healthcare decisions.

## 1.4 Thesis Organization

The thesis examines the use of AI in predicting strokes, with focus on SDoH. Chapter 1 introduces the topic, underscoring the relevance of incorporating SDoH in AI models for predicting strokes and highlighting the ethical implications. In Chapter 2, the thesis reviews existing literature on stroke and SDoH, explores their interplay, and delves into the challenges and ethical considerations associated with using ML for health predictions with a focus on stroke. This sets a foundation for the methodology described in Chapter 3, which details the preparation of datasets, development of models, and use of XAI techniques. Chapter 4 presents the findings, including the performance of AI models and insights from XAI techniques such as SHAP (SHapley Additive exPlanations) [26] and LIME (local interpretable model-agnostic explanations) [27], underscoring the significance of understanding how features influence model decisions. Chapter 5 discusses these results, particularly the ethical implications of incorporating SDoH in healthcare. This chapter also outlines the limitations of the study and potential future directions. The thesis concludes with Chapter 6, summarizing the key insights and reinforcing the importance of SDoH in the use of AI in healthcare, especially for stroke prediction.

# Chapter 2

# Background

In this chapter, we explore stroke, the role of SDoH, and how ML can help in healthcare, especially with cardiovascular diseases. We outline the significance of stroke and the multifaceted risk factors associated with it. This is followed by a discussion on SDoH, emphasizing their role in influencing stroke incidence and outcomes. We further discuss how SDoH and stroke are related, explaining how people's living conditions can change their risk of having a stroke. Next, we delve into the integration of SDoH within ML models for cardiovascular diseases, including stroke, highlighting the potential for these technologies to enhance predictive accuracy and treatment personalization. Finally, we address the incorporation of XAI in this context, underscoring the idea for equity, transparency, and fairness in deploying AI to combat health inequalities.

## 2.1 SDoH in Stroke: Definitions, Impact, and Roles

Stroke, a severe medical condition marked by the sudden interruption of blood supply to the brain, represents a substantial challenge to global healthcare systems [28]. The World Health Organization (WHO) defines stroke as "the rapid onset of clinical symptoms indicative of localized or sometimes widespread disruption of cerebral function, enduring for more than 24 hours or leading to fatality, with no discernible cause other than vascular origin" [29].

Despite healthcare improvements, stroke remains a major health issue worldwide, being the second top cause of death and the third leading cause of death and disability [30]. The economic impact of stroke is substantial, with an estimated global cost surpassing US$721 billion, equivalent to 0.66% of the total global GDP [31]. In Canada, the impact of stroke is significant, affecting approximately 878,500 adults over 20 years of age, including 438,700 men and 439,800 women (2017–18) [32, 33]. Notably, one-quarter of Canadians struggling with stroke, fall below the age of 65, highlighting the alarming trend of stroke risk escalating significantly beyond the age of 55 [34]. The prevalence of stroke in Canada underscores the urgent need for improved predictive models that can accurately identify individuals at high risk of stroke.

Recent efforts to mitigate persistent health disparities have sparked a growing interest in investigating the underlying causes of risk factors associated with stroke, often referred to as the 'causes of the causes [3].' These factors include SDoH, which are central to our research goals. Social determinants of health (SDoH), as defined by the World Health Organization

(WHO) [35], constitutes non-medical factors with a substantial influence on health outcomes. These determinants encompass the multifaceted conditions under which individuals are born, flourish, engage in livelihoods, and age [35]. At their core, they are fundamentally influenced by the distribution of financial resources, power dynamics, and access to essential assets [36]. These dynamics operate across global, national, and local scales, playing an important role in perpetuating health disparities within and between nations [35]. These social determinants, along with the healthcare system, are widely recognized as playing a significant role in maintaining health inequities [37].

The term 'SDoH' refers to the strong influence that a person's socioeconomic environment has on their health. The concept of SDoH, as described by Wilkinson and Marmot (2003) [38], highlights how the community and conditions in which people live can significantly affect their health. Historical research has demonstrated the interplay between economic development, social structures, and their profound implications for public health [39]. As the global prevalence of cardiovascular risk factors, such as obesity, hypertension, and diabetes, continues to surge [39–41], there exists an imperative to turn our attention towards the SDoH. This knowledge is crucial as we aim to narrow the gap in healthcare equality. [42].

The influence of SDoH on health outcomes is substantial, particularly concerning health disparities—inequitable differences in health status observed both within and between countries. Health outcomes worsen as socioeconomic status decreases, affecting those in lower socioeconomic positions more severely [35]. Recognizing the crucial role of SDoH components, such as income and education, as strong predictors of adverse health outcomes is vital [3]. Research indicates that social factors often influence health more than access to healthcare or individual lifestyle choices [43, 44]. SDoH may account for anywhere from 30% to 55% of health outcomes [35]. Addressing SDoH is important for improving population health and ameliorating long-standing health disparities [35]. These determinants also exhibit an association with an elevated risk of stroke incidence [45].

Cardiovascular disease is linked with SDoH [42]. Approximately 80% of an individual's health is influenced by factors beyond clinical care, including their physical environment, social determinants, and behavioral choices such as exercise and smoking [46]. The impact of social determinants on cardiovascular disease outcomes underscores the importance of integrating these factors into ML models for more accurate and fair predictions [47, 48]. SDoH factors often exist beyond the direct control of healthcare providers, presenting unique challenges in patient care [49]. Healthcare professionals, by understanding the significant influence of social factors, can better tailor their care for patients at higher stroke risk [49]. This insight emphasizes the importance of integrating SDoH considerations into our study.

Exploring the SDoH and its impact on stroke incidents necessitate a review of prior research that highlight the influence of socioeconomic status (SES) on cardiovascular health, especially in affluent nations [50]. Schultz et al. also shed light on four key components of SES: income, educational achievement, employment status, and environmental factors [50]. They describe how each impacts cardiovascular disease risk by influencing access to healthcare, health-related behaviors, and exposure to chronic stressors.

There is a need to understand disparities in stroke-related mortality, for example, younger black adults have a high mortality rate due to stroke, which contributes to overall healthcare disparities [51, 52]. Despite the identification of well-established medical factors, hypertension, diabetes mellitus, left ventricular hypertrophy, atrial fibrillation, cigarette smoking, and a history of heart disease account only about 50% of the elevated stroke risk observed in black populations [53, 54]. The remaining 50% suggests the presence of other significant determinants, likely associated with SDoH, highlighting the necessity for deeper investigation into their impact.

Moreover, we must acknowledge the substantial influence of geographic location on an individual's health, operating through both direct and indirect mechanisms. A study by Gabb et al. [55] showed that people in remote areas experience higher cardiovascular disease rates due to less frequent doctor visits, reduced cholesterol testing, poor blood pressure management, and older adults with diabetes receiving fewer statins. Multiple obstacles hinder access to healthcare in rural areas, such as inadequate public transportation to medical facilities [56]. Australia in particular exhibits geographic differences in cardiovascular disease prevalence, showing elevated rates in rural, regional, and remote locations [57]. Elements such as access to fresh fruits and vegetables at a geographic location also contribute to stroke incidence [58].

The accessibility of healthcare services impacts the frequency of health screenings and the prescription of necessary medications. For instance, a 2017 study focusing on cardiovascular health revealed that individuals with indications for statin medications received disparate levels of care based on their SDoH status [59]. Notably, only 45% of those with four or more SDoH indicators were prescribed statins, in contrast, 65% of individuals did not face barriers related to SDoH when accessing care [59]. Incorporating SDoH highlights disparities in stroke risk assessments, but solving healthcare inequities requires tackling systemic issues, shown by differences in statin prescriptions due to SDoH barriers.

Recent progress in making neighborhood data more accessible and the ability to add this data into Electronic Health Records (EHR) have sparked new interest in studying how neighborhood factors affect health of individuals [60]. The link between neighborhoods and health outcomes is influenced by factors such as heightened stress levels [61], reduced physical activity [62], and suboptimal dietary choices [63], all of which, affect both immediate risk factors like blood pressure, diabetes management, and inflammation, as well as more distal health outcomes such as cardiovascular diseases [64]. A study by Diez Roux [65] highlighted that neighborhoods affect the well-being of their residents by influencing both physical and social aspects of their environments. This research explored how geographical areas, defined by various socioeconomic and demographic characteristics, can impact health outcomes through complex interactions between the environment and individual behaviors [65].

Interestingly, current research has predominantly focused on incorporating specific SDoH factors, with an emphasis on socioeconomic aspects [66]. However, there remains a notable research gap in comprehensively examining how the coexistence of various SDoH within the same individual contributes to stroke incidence.

## 2.2 SDoH & Explainability in Machine Learning for Cardiovascular Disease

Zhao et al [48]'s study provides crucial insights into the field of ML for stroke prediction and the inclusion of SDoH factors. They reveal that out of 48 studies reviewed, only three compared the effectiveness of models with and without SDoH factors. Remarkably, all three studies found that adding SDoH data significantly improved prediction accuracy. Additionally, two of these studies highlighted the benefits of including gender and race information in the models [67, 68].

ML promises to address ethical challenges in SDoH research, particularly with regard to ensuring fairness in algorithms [69]. This is crucial as researchers delve into whether ML inadvertently introduces bias into SDoH findings, potentially impacting health policy recommendations [70]. The study of algorithmic fairness is therefore important, aiming at both comprehending and mitigating biases within SDoH [69]. Techniques to train ML algorithms to detect biases in health records [71, 72] are part of this effort to ensure equitable analysis. Other research works explore the relationships between health policies, their outcomes, and the blend of fairness in algorithms with ML methods [70]. For instance, Daoud et al. [73] combine causal inference with fairness, illustrating how algorithmic strategies can support decisions balancing economic and health impacts equitably.

Kino et al.'s scoping review [69] shows that in ML research related to SDoH, only five studies out of 82 focused on integrating SDoH factors for prediction task. In these predictive models, SDoH features were integrated alongside various other predictors. However, these inclusions were often lacking a coherent rationale to explain how the addition of SDoH features would yield improvement in predictive performance. [69]. While a few studies [74–76] evaluated the advantages and challenges of using ML to explore SDoH, such evaluations were rare. Although SDoH features are being incorporated, a thorough analysis of their impact on improving prediction accuracy is still in its early stages [69].

The application of ML models raises legitimate concerns regarding their interpretability and the potential pitfall of overfitting to the dataset [77]. Surprisingly, a comprehensive discourse on these concerns was absent in most of the literature reviewed by us. Few studies explored methods to evaluate the significance of each variable in NN [78], yet in many instances, the selection of features for models relied on their established medical importance. This indicates that model's ability to predict was often based on factors known to be related to cardiovascular disease, even though the significance of each variable wasn't fully evaluated. This absence of understanding on model interpretability calls for a more deliberate approach to model design and analysis in healthcare applications.

The "black box" nature of many ML algorithms is a challenge, obscuring the decision-making process and thus highlighting the need for interpretability [79, 80]. This opacity is problematic, especially given the ethical and societal stakes involved in healthcare AI applications [81]. Transparency, interpretability, and explainability, although often used interchangeably, hold distinct meanings. Interpretability concerns the degree to which a model's processes can be understood [82], often overlapping with the concept of explainability [83]. Transparency, on the other hand, seeks to elucidate the model's internal mechanics [84]. In this work, we use XAI techniques to more delibrately analyze the impact of SDoH factors on stroke prediction task.

Fellous et al. [85] demonstrate how applying XAI principles can clarify complex domains and improve understanding of underlying mechanisms, particularly in areas where the decision-making process of AI systems needs to be transparent. Acknowledging that ML can inadvertently reinforce societal biases is crucial, as this contrasts with the goals of SDoH research, which focus on uncovering and addressing social inequities [81]. Research in algorithmic fairness frequently uncovers the difficult trade-offs between predictive accuracy,interpretability and fairness, underscoring the necessity for thoughtful model design in healthcare AI [86, 87].

Implementing XAI techniques into ML models for predicting cardiovascular disease can help explain the influence of various factors on health outcomes. This makes the decision-making process more transparent and informative, thereby improving the understanding of model output for healthcare professionals and patients and enabling a more integrated view of health outcomes [88–90]. This contributes to the responsible development of AI technology by enhancing interpretability and transparency, crucial for meaningful insights and actionable healthcare decisions.

# Chapter 3

# Methodology

Following the exploration of stroke, SDoH, and the potential of ML in healthcare (Chapters 1 & 2), this chapter delves into the methodological foundation of our research. We begin by outlining the data acquisition process, specifically the utilization of survey data from the Institut de la statistique du Québec (ISQ). This data underwent preprocessing to select features relevant to stroke prediction, including both traditional health factors and SDoH features. We addressed missing values and data imbalances to ensure model robustness.

Next, we describe the selection and evaluation of our machine learning models. We employed a diverse range of models, including Random Forest (RF), Gradient Boosting Machine (GBM), K-Nearest Neighbors (KNN), Neural Networks (NN), Light Gradient Boosting Machine (LGBM), CatBoost (CB), and XGBoost (XGB). These models were evaluated using a comprehensive set of metrics including accuracy, precision, recall, F1 score, and ROC-AUC, enabling a robust assessment of their performance in stroke prediction.

Finally, recognizing the importance of interpretability and fairness in AI-driven healthcare solutions, we incorporated XAI techniques into our study. LIME, SHAP, and feature importance analysis were employed to gain deeper insights into model predictions and understand the impact of SDoH factors on stroke risk assessment. This multifaceted approach strengthens the foundation of our research and paves the way for responsible and transparent AI development in healthcare.

The source code can be found here: `https://github.com/gsharma15/sdoh_stroke_predicton`

## 3.1 Dataset

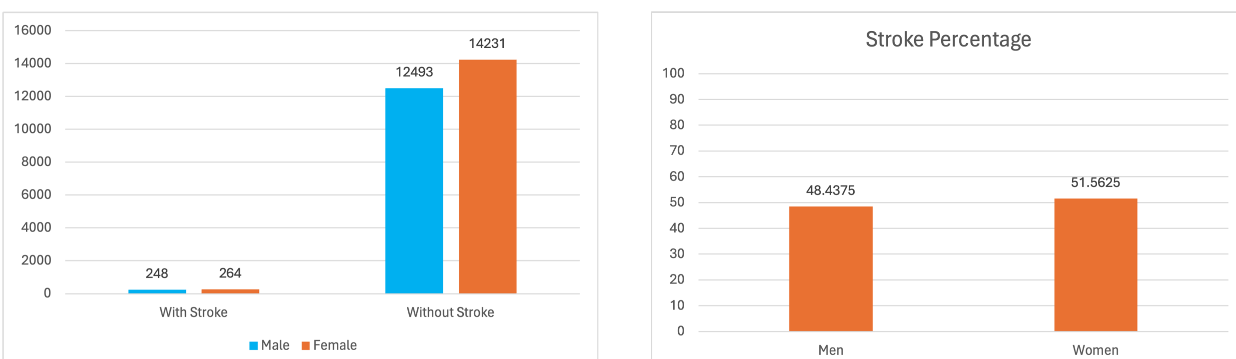
The dataset used for this research was obtained from the Institut de la statistique du Québec (ISQ) [91]. ISQ is the provincial statistical agency responsible for producing, analyzing, and disseminating official statistics in Quebec. The Québec government commissioned the ISQ to facilitate access to specific information held by public institutions for research purposes [92]. The dataset comprises designated survey data collected by various government bodies, including the Ministère de la Santé et des Services sociaux.

The survey dataset is comprised of raw, non-aggregated and self-reported responses, meaning the information is presented in its original, unprocessed state. The surveys comprising the dataset include responses to questions posed in custom-designed questionnaires, often conducted by the Institut de la statistique du Québec and Statistics Canada. For our research, we specifically requested surveys containing cardiovascular diseases. The data acquisition process resulted in the inclusion of 7 surveys previously conducted between the years 2007 and 2019 for Quebec population.

The final dataset comprised 27,236 rows and 69 columns. Each row represents a patient whereas each column highlights their corresponding SDoH and stroke-related features. The mean age is 52, whereas the median age is 56, in the dataset. The number of biological male and female participants is 12,741 and 14,495 respectively.

We examined the statistical distribution of biological sex, race, and ethnicity to identify any existing healthcare disparities or anomalies. Our analysis showed that among individuals who experienced a stroke, 48.44% were women and 51.56% were men (Figure 3.1). Additionally, we conducted a race-based analysis to investigate the relationship between race and stroke incidence, evaluating how stroke rates vary among different racial groups (Figure 3.2a). Finally, an ethnicity-focused assessment was carried out to uncover any disparities in stroke prevalence. The stroke percentage data revealed notable disparities when compared to the demographic distribution of the population. Certain ethnicities such as Canadian, English, German, Scottish, and Irish exhibited higher rate of stroke (Figure 3.3a) than their respective population representations, suggesting potential health disparities or data collection biases. Conversely, the Chinese group showed a lower stroke percentage relative to their demographic presence, indicating either a lower risk or potential underreporting. These patterns highlight the need for a detailed investigation into the socioeconomic, healthcare access, genetic, and lifestyle factors that may contribute to these observations, as well as a review of the data collection methodologies to ensure accurate representation and reporting.

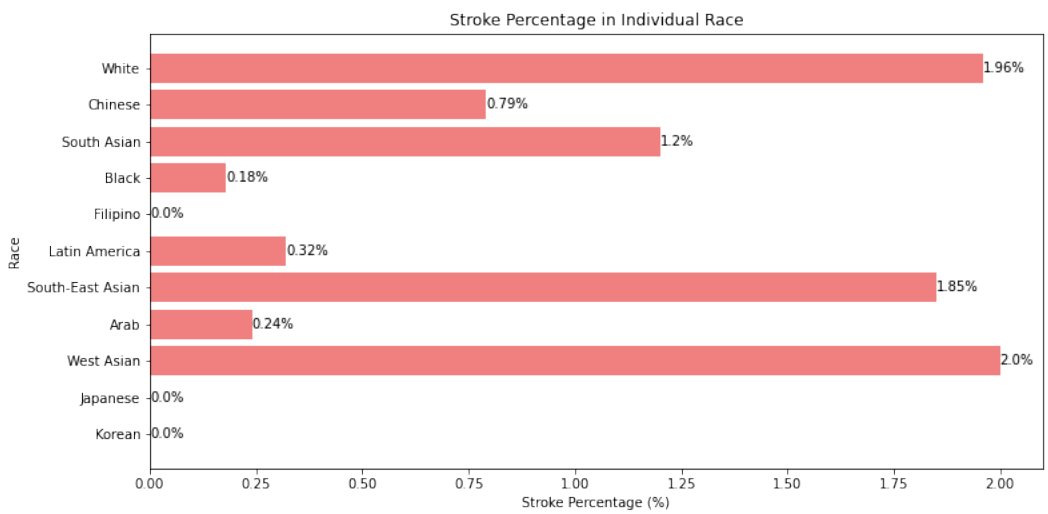
**Sex:** Sex refers to the biological and physiological characteristics that distinguish individuals as male or female [93]. We divided the dataset into individuals who suffered from stroke and those who had not. Subsequently, sex distribution and stroke incidence rates for both men and women were calculated and compared. This analysis aimed to reveal any notable sex-related patterns in stroke occurrences.




**(a)** Overall stroke distribution in the dataset **(b)** Stroke distribution in different sex

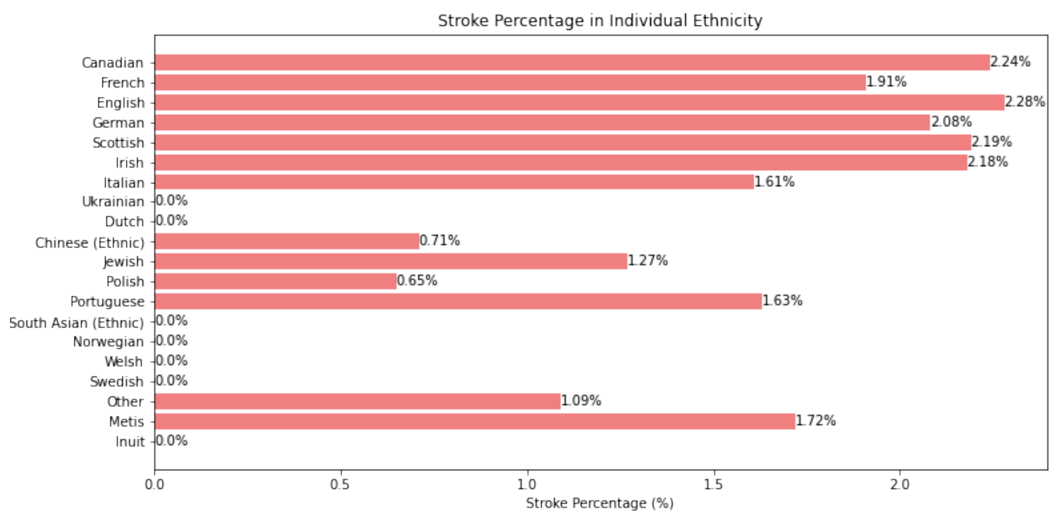
**Figure 3.1:** Sex Differences in Stroke Incidence

**Race:** Race is a social construct that categorizes individuals based on shared physical characteristics and geographical ancestry [94]. We identified specific racial categories within the dataset and examined the prevalence of strokes within each category (Figure 3.2a). Examining race reveals stroke risk disparities. While West Asians have the highest stroke prevalence (2.0%), Filipinos, Japanese, and Koreans show none (though this might be due to small sample sizes). White, South-East Asian, and South Asian populations have moderate stroke rates (1.2% - 1.96%), while Latin American, Arab, and Black populations have a lower prevalence (0.18% - 0.32%).




**(a)** Stroke Distribution in Individual Race Groups

**Ethnicity:** Ethnicity refers to an individual's cultural identity, including shared language, customs, traditions, and values [95]. Similar to the race analysis, the ethnicity-based investigation explored the potential link between ethnicity and strokes. We identified different ethnic groups within the dataset and calculated the percentage of individuals within each group who had experienced strokes (Figure 3.3a). This analysis aimed to shed light on any potential disparities in stroke prevalence among different ethnic backgrounds. Canadians (2.24%), French (1.91%), English (2.28%), Scottish (2.19%), Irish (2.18%), Italians (1.61%), Portuguese (1.63%), Metis (1.72%), and Jews (1.27%) showed moderate prevalence. Notably, Ukrainians, Dutch, South Asians, Norwegians, Swedes, and Welsh had a 0% prevalence, though this might be due to small sample sizes. The "Other" category had a prevalence of 1.09%.




**(a)** Stroke Distribution in Individual Ethnic Groups

## 3.2 Data Preprocessing

Our data preprocessing steps include data cleaning, handling missing data, and data normalization to ensure the dataset's quality and uniformity.

### 3.2.1 Feature Selection

Feature selection is a step in model development, making sure only relevant information is used for predictions. A list of features selected can be found in Appendix A in table A.1. Data preprocessing involved selecting appropriate surveys and identifying features relevant to the research question from these surveys, a crucial step in refining the dataset to include only necessary information for meeting research goals.

- **Selection of Surveys:** This involves identifying a set of surveys that housed the pertinent data for our research question. The selected surveys represented diverse data obtained over different timeframes (2007-2019), encompassing a wide range of demographics and including specific stroke-related data. A total of 5 surveys from ISQ were included out of 7.

- **Selection of Features:** Not all features in the original surveys were relevant to the research objectives, either due to missing data or because they were not related to stroke, SDoH or traditional stroke risk factors. Relevance of features was determined based on whether existing literature in cardiovascular diseases considered them.

Typically, these selected columns pertained to features such as demographic attributes (e.g. age, sex) , health indicators (e.g. presence of diabetes, stroke), and SDoH (e.g. access to food, education level).

- **Mapping Column Names:** To facilitate the mapping of features across different surveys, we adopted a strategy to maintain consistency in feature naming, prioritizing the nomenclature from the first survey in which each feature appeared. For example, the marital status feature, which appeared as 'DHH_MS' in surveys before 2015 and was shortened to 'MS' in subsequent surveys, was labeled as 'DHH_MS' in our analysis. This simplification aimed to maintain a clear and consistent framework for all features, facilitating easier cross-survey comparisons and analysis.

  Moreover, we undertook the task of renaming certain features to enhance clarity and comprehension. For example, the variable 'SDC_020B', which represented the South Asian race across all surveys, was renamed to 'South_Asian'. This renaming strategy was applied judiciously to various features, with the aim of making the dataset more accessible and intuitive for researchers, thereby facilitating a more efficient data analysis process.

This process helped refine our dataset, including only essential features for analysis and reducing interpretation errors during our analysis. The distribution of features across categories is as follows: 18 in traditional, 3 in demographic, 3 in economic, 2 in neighbourhood, 5 in personal, 4 in social, 3 in mental health, 12 in race, and 20 in ethnicity. A list of these features can be found in Appendix A in table A.1.

### 3.2.2 Data Linkage

We integrated data from various survey years by standardizing the names for similar features across datasets. This process ensured that the features were consistent, making it easier to analyze the impact of SDoH on health outcomes by ensuring that all data points were comparable. In the previous sub-section 3.2.1, we discussed how features were extracted and mapped for individual surveys selected for this research. This section explains how we combined different surveys into a single dataset. This merged dataset was key in analyzing how SDoH affect cardiovascular diseases.

Given the substantial volume of data at our disposal and the constraints of the computational resources available at ISQ, we adopted a strategic **chunked processing approach** [96] for merging the data from surveys. This approach offered several advantages, including enhanced efficiency, minimized memory usage, and preservation of data.

(a) Determining Chunk Size: The process starts by deciding on a chunk (batch) size of 1000 rows. This means that for every survey file processed, up to 1000 rows are handled at a time. We decided a batch size of 1000 to strike a balance between efficiency and system resource utilization. It allowed for processing substantial data volumes without overwhelming memory, ensuring smooth and manageable data handling.

(b) Setting Up for Data Aggregation: An empty space, or "container", is prepared for collecting the data. As surveys are processed in batches, this container will hold the accumulated data.

(c) Processing Surveys in Batches: The approach involves going through the survey files in batches, each time taking a set that corresponds to the batch size of 1000 rows.

(d) Extracting and Merging Data: Within each batch, the surveys are opened, and data is extracted. This data is then merged into a single collection, making it a unified structure of information. This step consolidates the data from multiple surveys, simplifying further analysis.

(e) Compiling the Data: After processing a batch of surveys and merging their data, this combined information is added to the prepared container. This process is repeated for each batch until all surveys have been processed.

(f) Completing the Dataset: Through repeating this batch process for all survey files, a final, combined dataset is formed. This dataset comprises merged data from all the processed surveys, ready for analysis or further processing.

The merged dataset prior to further preprocessing had 57,131 individuals and 69 features. The data merging process created a unified and detailed dataset by combining features from various surveys. This ensured that each variable was consistently represented, setting a solid foundation for our further analysis.

Using this dataset, we proceeded to address missing data, conduct exploratory data analysis, build ML models, and examine aspects related to our research goals.

### 3.2.3 Missing Data Handling & Data Transformation

The process of preparing the dataset for analysis involved two crucial aspects: handling missing data and data transformation.

#### Handling Missing Data

Missing data can arise due to various reasons, including survey non-responses or data recording errors. In this research, the following strategies were employed to address missing data:

- **Replacement of Missing Values:** Initially, missing values, represented as NaN (Not a Number), in the dataset were systematically replaced with zero (0). This replacement was necessary to enable mathematical operations and ensure the dataset's suitability for modeling and analysis.
- **Imputation with Median Values:** For remaining missing values, a median imputation strategy was employed. Median imputation involves calculating the median value for each respective column. This approach ensured that the imputed

values retained the statistical characteristics of the original data. Furthermore, the imputed median values were rounded to the nearest integer to maintain data uniformity.

### Data Transformation

Data transformation steps were executed to enhance the quality and relevance of the dataset:

- **Removal of Unwanted Survey Responses:** Many survey datasets contain responses that are not applicable to the research objectives or contain responses like "didn't report" or "don't know." 62 features had such responses. To focus the analysis on meaningful survey data, these unwanted responses were identified based on their specific response codes. Corresponding rows containing such responses were systematically removed from the dataset.

- **Data Type Conversion:** The survey responses, originally integers, were represented as decimals in the dataset, for instance, 1 was shown as 1.0. To streamline subsequent analyses and maintain consistency within the dataset, all column data types were converted to integers (int). This transformation ensured that all values in the dataset were whole numbers, simplifying the modeling process.

- **Verification of Removal:** To ensure the effective removal of unwanted responses, a validation step was implemented. A count of unique values in the specific columns was performed to confirm that rows containing unwanted responses had been successfully excluded from the dataset.

After handling missing data, the cleaned dataset was saved to a new file. This cleaned dataset was used for further analysis and modeling. The cleaned dataset had 27,236 individuals and 69 features.

### 3.2.4 Handling Class Imbalance

Class imbalance is a common challenge in healthcare datasets, as certain health conditions, such as strokes, are relatively rare compared to non-stroke cases. In our analysis, we encountered a substantial class imbalance with 512 stroke cases and 26,724 non-stroke cases. This significant disproportion can lead to models that are biased towards predicting the majority class (non-stroke), as they tend to optimize overall accuracy by favoring the more prevalent class. This results in poor predictive performance for the minority class (stroke). To address this issue and ensure that our predictive models are robust and effective, we employed Under-Sampling [97] technique for our problem. This approach is justified for the following reasons:

1. **Enhancing Model Sensitivity to Stroke Cases:** Given the nature of accurately identifying stroke cases, undersampling the non-stroke cases to balance the dataset can increase the model's sensitivity to the stroke class.

2. **Mitigating Model Bias:** The dataset's imbalance heavily biases prediction towards the majority class (non-stroke). By undersampling, we reduce this bias, enabling the model to learn a more balanced decision boundary that improves its ability to distinguish between stroke and non-stroke cases.

3. **Computational Efficiency:** Owing to the limitation of computational resources, undersampling was the optimal method to enhance model efficiency and focus on accurately detecting stroke cases.

Undersampling was chosen over oversampling or SMOTE [98] as it avoids the computational burden and potential overfitting associated with increasing the dataset size through synthetic data generation or duplication.

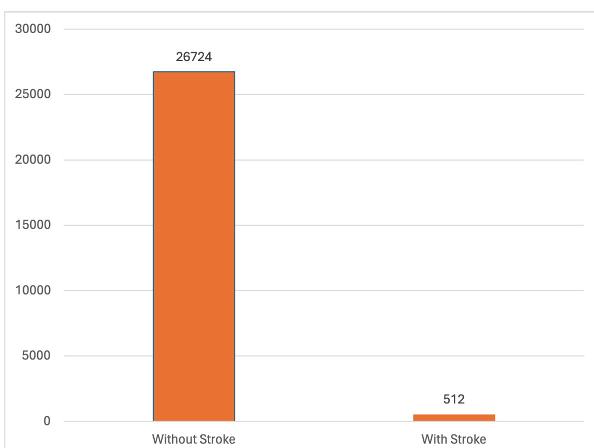



**Figure 3.4:** Dataset Imbalance - Overall stroke distribution in the dataset

### Under-Sampling

Under-sampling, seeks to balance the class distribution by reducing the number of majority class instances. Specifically, we randomly selected a subset of majority class instances to match the number of minority class instances. The goal of random under-sampling is to create a balanced dataset, but it comes with the trade-off of reducing the overall dataset size. Mathematically, we ensured that the number of majority class instances ($M$) equaled the number of minority class instances ($N$):

$$M = N$$

While random under-sampling can balance the dataset, it may lead to a loss of valuable information from the majority class.

We applied NearMiss version 3 [99] for undersampling to consider equity by ensuring balanced representation between the majority and minority classes. This method selects majority class instances based on their proximity to the minority class, aiming to preserve the most informative and challenging examples to maintain a meaningful decision boundary. By matching the number of instances between classes, it seeks to reduce bias and ensure that both stroke and non-stroke cases are equally represented and learned by the model.

## 3.3 Model Development

#### Data Splitting

To evaluate the predictive capabilities of various ML models, we divided the dataset into training (80%) and test sets (20%).

### 3.3.1 Model Selection

We employed a diverse set of ML algorithms, each with its own strengths and characteristics, to predict stroke occurrences. The selected models included Random Forest (RF) [20], Gradient Boosting Machines (GBM) [21], K-Nearest Neighbors (KNN), Neural Network (NN) [22], LightGBM (LGBM) [23], CatBoost (CB ) [24], and XGBoost (XGB) [25].

### 3.3.2 Model Evaluation

In the evaluation process, each model was assessed on the test dataset to generate predictions, based on performance metrics like accuracy, precision, recall, and F1 score. Additionally, model performance was assessed across various feature subsets (demographic, economic, ethnicity, race, social factors, neighbourhood, personal factors and mental health) to identify the impact of different feature combinations on model accuracy and other key metrics. This evaluation aimed to uncover potential variations in model's accuracy and effectiveness using different SDoH.

### 3.3.3 Model Tuning

In our study, we utilized GridSearchCV [100] for thorough hyperparameter tuning across various ML algorithms to develop predictive models for stroke. This method enabled us to systematically explore and optimize hyperparameters, focusing on performance enhancement, complexity management, and overfitting prevention, all while maintaining computational efficiency. Such adjustment of parameters like the number of estimators and learning rates was important for building models capable of accurately predicting stroke occurrences. The details on the specific hyperparameters can be found in table 3.1.

### 3.3.4 Explainable Artificial Intelligence (XAI)

In our study, we integrated different XAI techniques, particularly feature importance, LIME [27], and SHAP [26], to interpret the predictive models and understand the impact of SDoH on stroke prediction.

**Feature Importance** The feature importance method, which ranks and visualizes the top most influential features, was used to highlight the key features driving predictions across different models. We identified top 15 features, to allow for a focused analysis on the most impactful features, balancing detail with manageability in model interpretation and visualization. This approach leveraged the `feature_importances_` attribute of models to rank features based on their impact on the model's predictive accuracy. A bar chart ( 4.2.1)

| Model | Parameters |
|---|---|
| Random Forest | • n_estimators: [100, 200]<br>• max_depth: [10, 20]<br>• min_samples_split: [2, 5]<br>• min_samples_leaf: [1, 2] |
| Gradient Boosting Machine | • n_estimators: [100, 200]<br>• learning_rate: [0.01, 0.05]<br>• max_depth: [3, 4] |
| KNN | • n_neighbors: [3, 5]<br>• weights: [‘uniform’, ‘distance’]<br>• p: [1, 2] |
| Neural Network | • hidden_layer_sizes: [(50, 50)]<br>• activation: [‘relu’]<br>• alpha: [0.0001, 0.001]<br>• learning_rate: [‘constant’] |
| LGBM | • n_estimators: [100, 200]<br>• learning_rate: [0.01, 0.05]<br>• max_depth: [3, 4] |
| CatBoost | • iterations: [500]<br>• learning_rate: [0.01, 0.05] |
| XGBoost | • n_estimators: [100, 200]<br>• learning_rate: [0.01, 0.05]<br>• max_depth: [3, 4] |

**Table 3.1:** Hyperparameters used for each machine learning model

visualization highlights the relative importance of these features. This process aids in model interpretation and guides feature selection and model refinement, ensuring that the most relevant features are considered in stroke prediction.

**LIME** LIME was applied for local interpretability of individual predictions. By generating perturbations around a specific instance and observing the impact on the prediction, LIME identifies relevant features. This instance-level explanation is useful for understanding model behavior on complex or borderline cases, offering detailed insights into the reasons behind a model's prediction. Through this method, we can communicate the rationale behind specific predictions, enhancing trust and transparency in the model's predictive output.

**SHAP** SHAP values were used to break down how each feature contributed to predictions. This method offers a global and local understanding, meaning, it gives a broad overview of which features are important overall (global), and also shows how each feature affects specific predictions (local). By employing SHAP, we were able to generate summary plots illustrating this information. This dual perspective of global and local interpretability makes SHAP an invaluable tool for comprehensively understanding model predictions and the underlying feature interactions.

# Chapter 4

# Results

Building upon the foundations established in Chapter 3, this chapter presents our findings on the impact of incorporating SDoH into stroke prediction models. We explore model performance (4.1), different subsets of SDoH (4.1.2), and the role of SDoH factors in stroke prediction (4.2). We utilize different evaluation metrics and the aforementioned XAI techniques (Chapter 3) to understand the role of SDoH in stroke prediction.

## 4.1 Model Performance

As our research objective aims to examine how consideration for SDoH improves ML prediction of stroke risk, we first examined ML models on the stroke risk prediction task without any SDoH. We sought to verify that our models perform comparatively to existing models in the literature. We examine various metrics to assess performance of our prediction models. This evaluation shows how well the models work and highlights the effects of SDoH factors on stroke prediction. To evaluate the performance of all models, we calculated key metrics including accuracy, precision, recall, and F1 score, in addition to generating ROC-AUC curves (Figure 4.1). These measures provided an assessment of each model's ability to predict stroke occurrences. The ROC-AUC curve further offered insights into the trade-off between true positive rates and false positive rates across various threshold settings, improving our understanding of model effectiveness for different stroke events.

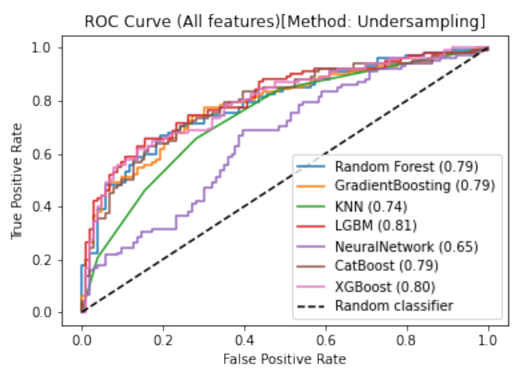



**Figure 4.1:** ROC-AUC Curve for Stroke Prediction

Figure 4.1 displays an ROC curve for models that incorporated all features, including both traditional and all SDOH categories.. The diagonal dashed line represents a random classifier that serves as a baseline. An effective model's ROC curve should significantly deviate from this line. Ideally, the curve should rise sharply towards the top left corner, indicating the model accurately identifies individuals with stroke (true positives) with minimal false alarms (incorrectly classifying healthy people as having a stroke). By analyzing the location and shape of a model's ROC curve in Figure 4.1 compared to the baseline, we can assess its effectiveness in differentiating between those at high and low risk of stroke.

### 4.1.1 Insights from Models

1. **LGBM:** Demonstrating the highest AUC of 81%, the LGBM model is the top performer among all the models. This implies that LGBM offers the most significant discriminative power in identifying potential stroke cases in our dataset.

2. **XGB:** XGB demonstrated an AUC of 80%, highlighting its effectiveness in predicting stroke with high accuracy.

3. **Ensemble Methods - RF , GBM, and CatBoost:** Interestingly each of the ensemble methods displayed an AUC of 79%. Their similar performance necessitates further investigation to understand differences in their predictions.

4. **KNN:** With an AUC of 74%, KNN, doesn't match the predictive capabilities of the aforementioned ensemble methods in this context.

5. **NN:** Surprisingly, the NN model registered the lowest AUC of 65% among the models. The model's lower AUC, compared to other models, suggests it might not have captured the complex relationships in the data as effectively, possibly due to insufficient training data, or a need for more sophisticated architecture adjustments. This underlines the

notion that complex models like NN are not always superior.

**Table 4.1:** Performance metrics for all models (in percentage)

| Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| RF | 77.10 | 77.80 | 75.50 | 76.60 |
| GBM | 73.20 | 72.80 | 73.50 | 73.20 |
| KNN | 69.80 | 71.30 | 65.70 | 68.40 |
| NN | 62.90 | 62.00 | 65.70 | 63.80 |
| **LGBM** | **77.60** | **78.60** | **75.50** | **77.00** |
| CB | 75.10 | 76.80 | 71.60 | 74.10 |
| XGB | 74.10 | 75.30 | 71.60 | 73.40 |

The results from Table 4.1 highlight the performance of various ML models in stroke prediction. These metrics collectively provide insight into the models' predictive capabilities, balancing the identification of true stroke cases against the minimization of false positives and false negatives. Detailed results can be found in Appendix B.

- **RF** has shown a balanced performance with an accuracy of 77.10%, precision of 77.80%, recall of 75.50%, and an F1 score of 76.60%. High recall indicates the model's strength in identifying a high proportion of actual stroke cases, making it particularly valuable in medical settings where missing a stroke case could have dire consequences. The precision score reflects its relative success in minimizing false positives, while the F1 score suggests a balanced trade-off between precision and recall.

- **GBM** achieves an accuracy of 73.20%, precision of 72.80%, recall of 73.50%, and an F1 score of 73.20%. Its superior recall suggests it's good at detecting stroke cases, crucial for early intervention and treatment. The high precision indicates fewer false positives, but the F1 score is slightly lower than LGBM.

- **KNN** exhibits a higher precision of 71.30% but a lower recall of 65.7%, resulting in an F1 score of 68.40%. The higher precision means the model is effective at correctly identifying stroke cases when it predicts one, though the lower recall points to a potential limitation in capturing all true stroke cases.

- **NN** shows the lowest performance, with an accuracy of 62.90%, precision of 62.00%, recall of 65.70%, and an F1 score of 63.80%. The lower values across these metrics indicate challenges in both identifying true stroke cases and in avoiding false stroke predictions, suggesting that this model might be less suitable for applications where both accurate detection and minimization of false alerts are important.

- **LGBM** outperforms other models with an accuracy of 77.60%, precision of 78.60%, recall of 75.50%, and an F1 score of 77.00%. The high recall rates are indicative of its capability to identify most true stroke cases, the precision scores reflect its effectiveness in minimizing false positives and the F1 scores indicate a strong balance between precision and recall, making them highly suitable for stroke prediction in healthcare contexts.

- **CB** achieves similar results to LGBM with an accuracy of 75.10%, precision of 76.80%, recall of 71.60%, and an F1 score of 74.10%. The high recall rates makes it a valuable tool in stroke prediction. Their high recall rates are indicative of their capability to identify most true stroke cases, which is essential for preventing missed diagnoses.

- **XGB** displays a balance with an accuracy of 74.10%, precision of 75.30%, recall of 71.60%, and an F1 score of 73.40%. Like CB and LGBM, XGB's high recall rate makes it a valuable tool in stroke prediction for its ability to catch a high number of stroke cases, with its F1 score affirming a favorable balance between identifying true cases and minimizing false diagnoses

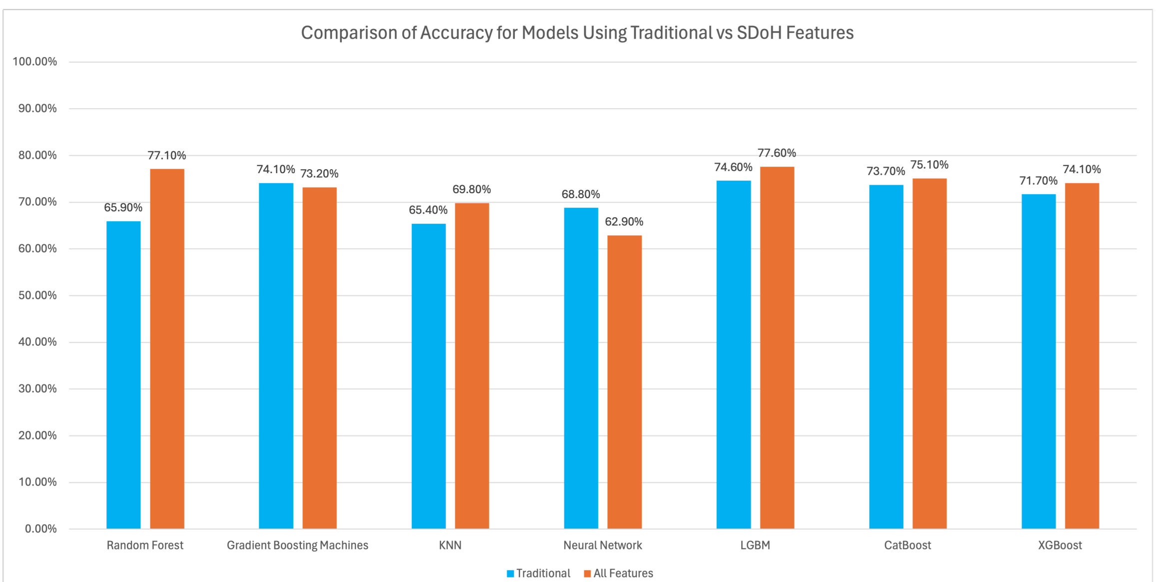



**Figure 4.2:** Accuracy Comparison Based on Traditional vs Traditional & SDoH features

When SDoH data is included alongside traditional features (Figure 4.2), we see an improvement in the accuracy of many prediction models. This emphasizes the importance of using a wider range of data to improve predictions. Models like RF and LGBM jump in accuracy (from 65.90% to 77.10% and 74.60% to 77.60%, respectively) when SDoH features are included. These improvements highlight that SDoH data, which includes economic factors, social aspects, demographics, and personal information, captures crucial details missed by medical indicators alone. However, it is important to note that not all models demonstrate improvement upon the inclusion of SDoH data, underscoring the complexity and variability in how these factors interact with health outcomes. The results partially support the research hypothesis, indicating that incorporating SDoH data can enhance prediction models, though the impact varies, reflecting the complex relationship of these factors in determining health outcomes.

### 4.1.2 SDoH

In this section, we evaluate how different subsets of SDoH factors such as demographic, socioeconomic, and personal characteristics affect the performance of our prediction models. By examining these subsets individually, we gain insights into how specific groups of SDoH contribute to the accuracy and reliability of stroke predictions. A detailed list of results for all the metrics evaluated for each SDoH subset can be found in Appendix B.

**Demographic Factors:** Including demographic data like age and marital status improved the performance of several stroke risk prediction models (Figure 4.3), particularly GBM (77.10% accuracy with demographics vs 74.10% without) and XGB (76.10% vs 71.70%). Similar results were obtained for KNN models (69.30% vs 65.40%).While NN initially suffered a drop in accuracy with all features (62.90% all features vs 68.80% with traditional features only), incorporating demographics still yielded a modest improvement (65.90%). These findings suggest that demographic data can enhance the accuracy of various predictive models, though the impact may vary depending on the specific model type.

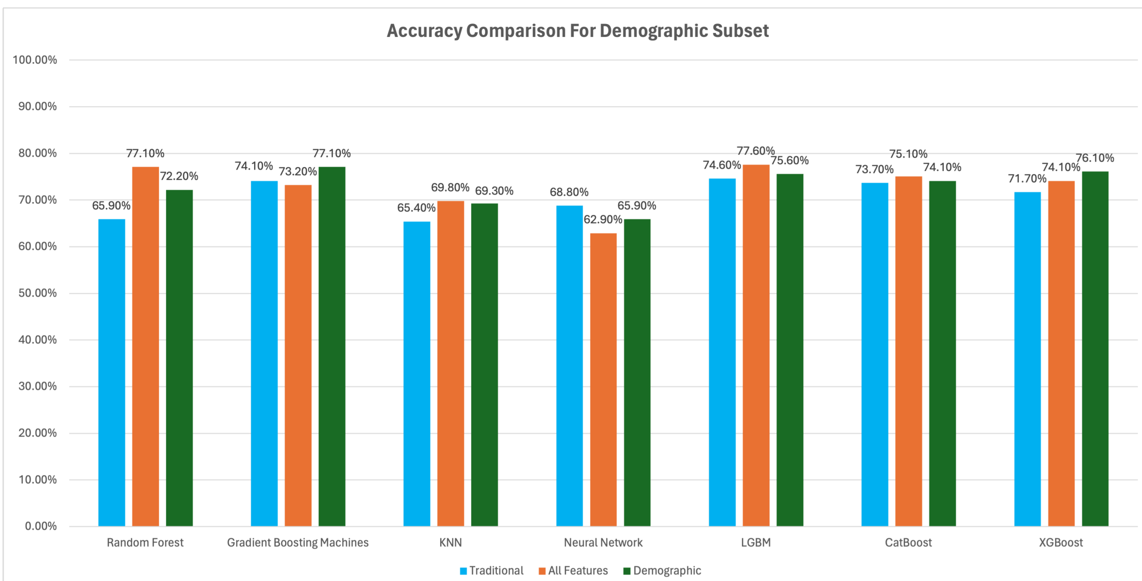



**Figure 4.3:** Impact of Demographic Features on Model Performance

**Economic Factors:** Economic features' (Figure 4.4) influence on stroke risk prediction models varies by algorithm. GBM and LGBM see accuracy gains (77.10% and 77.60% respectively) with economic features, suggesting these variables like employment and housing are strong predictors. KNN also benefits greatly, reaching its highest accuracy (70.70%) with economic data. Conversely, (69.80%) and XGB (75.60%) show more modest improvements, while NN see a slight recovery in accuracy (66.80%) compared to the full feature set. CB is the only model with a slight decrease (73.20%) in accuracy, indicating it might not utilize economic variables as effectively as others. Economic features enhance the performance of specific models, highlighting their relevance, but the impact varies depending on the model type.

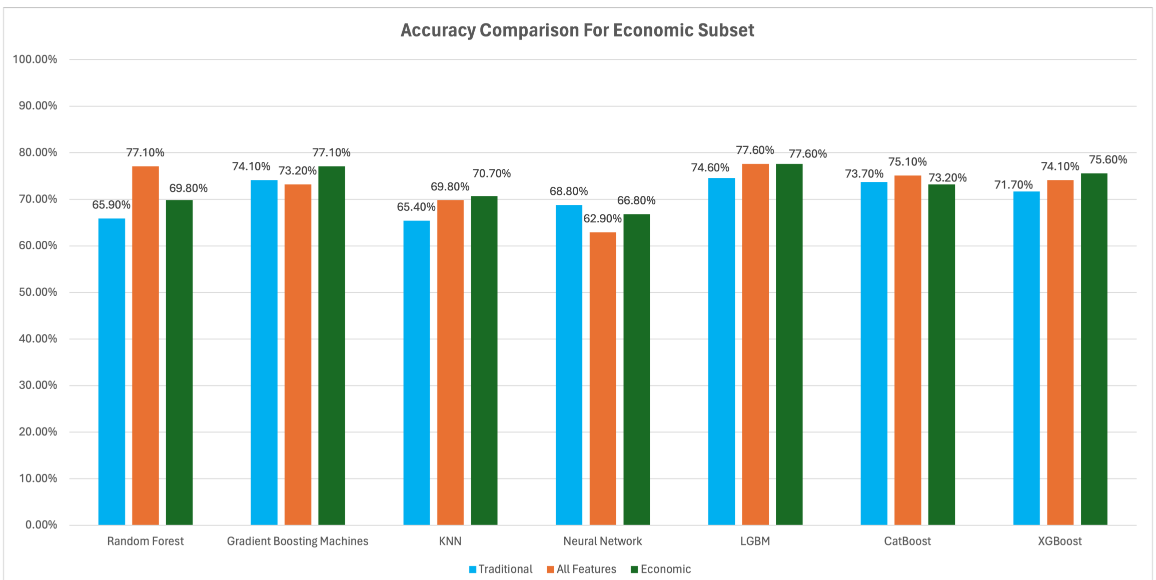



**Figure 4.4:** Impact of Economic Features on Model Performance

**Ethnicity:** Examining the impact of ethnicity (Figure 4.5) on stroke risk prediction models reveals different outcomes across algorithms. RF (65.90%) shows no change in accuracy, suggesting ethnicity offers no additional predictive power for this model. GBM (73.70%) and XGB (73.70%) experience slight decreases, implying ethnicity has some value, but not as much as traditional features. LGBM (73.70%) and CB (72.70%) see moderate reductions, indicating a mild impact of ethnicity on their performance. KNN

(64.40%) and NN (63.40%) exhibit the most significant drops, suggesting the ethnicity feature may not align well with their underlying mechanisms or may even introduce noise. The influence of ethnicity on model accuracy varies considerably, with some models potentially benefiting little and others struggling to integrate this feature effectively.

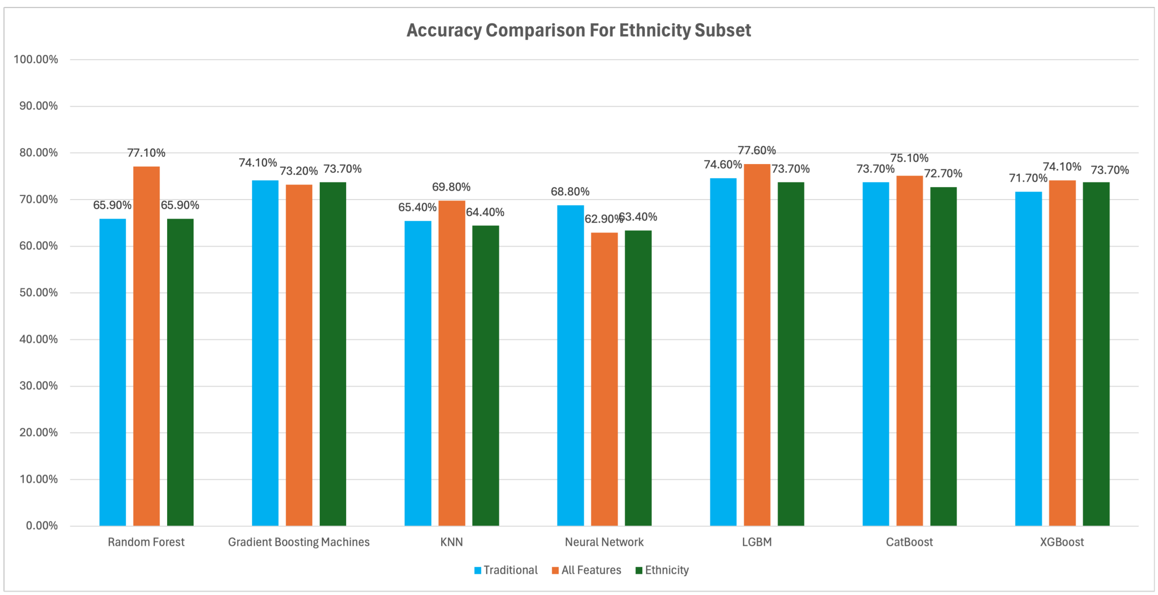



**Figure 4.5:** Impact of Ethnicity on Model Performance

**Mental Health Factors:** Integrating mental health data (Figure 4.6) into stroke risk prediction models yielded interesting insights but generally decreased accuracy across all algorithms. NN exhibited the maximum drop (62.00% accuracy), suggesting mental health features alone offer less predictive power for this model compared to the complete dataset. GBM and XGB also showed accuracy declines, indicating some value in mental health data but less than the full feature set. LGBM and CB displayed moderate decreases, suggesting a milder influence of mental health on their performance. The most significant drops occurred with KNN (69.30%) and NN (61.50%), potentially due to these models struggling to integrate mental health data effectively. Mental health features may offer valuable insights when combined with others, their inclusion highlights the complexity of mental health as a predictor and the need for further analysis to capture its full potential.

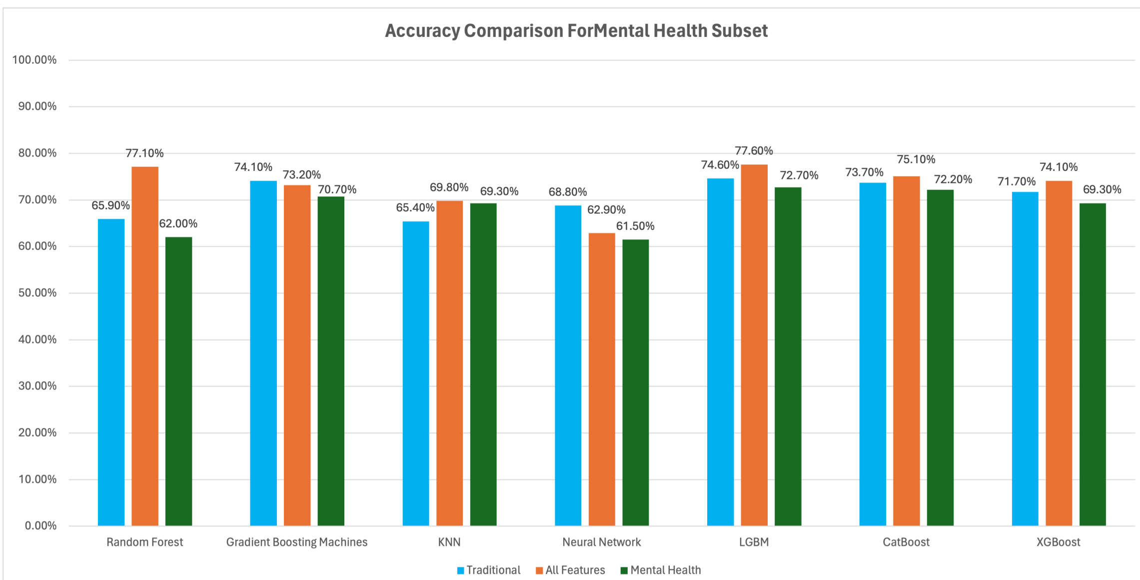



**Figure 4.6:** Impact of Mental Health Features on Model Performance

**Neighbourhood Factors:** Analyzing the impact of neighborhood features (Figure 4.7) reveals a trend where most models either see no improvement or experience a decrease in accuracy compared to using all features. (65.90%) shows no change, suggesting neighborhood data offers no additional benefit for this model. GBM (71.20%), XGB (72.20%), CB (72.70%), and LGBM (73.70%) all experience minor to moderate accuracy drops, indicating neighborhood features hold some value but less than the full dataset. The most significant decrease occurs with KNN (59.50%), suggesting this model struggles to utilize neighborhood data. Similarly, the NN (63.90%) shows a decrease, further highlighting potential challenges in leveraging this type of information. While neighborhood features might hold some predictive power, they appear less influential on their own and may require integration with other data types for more robust prediction in these models.

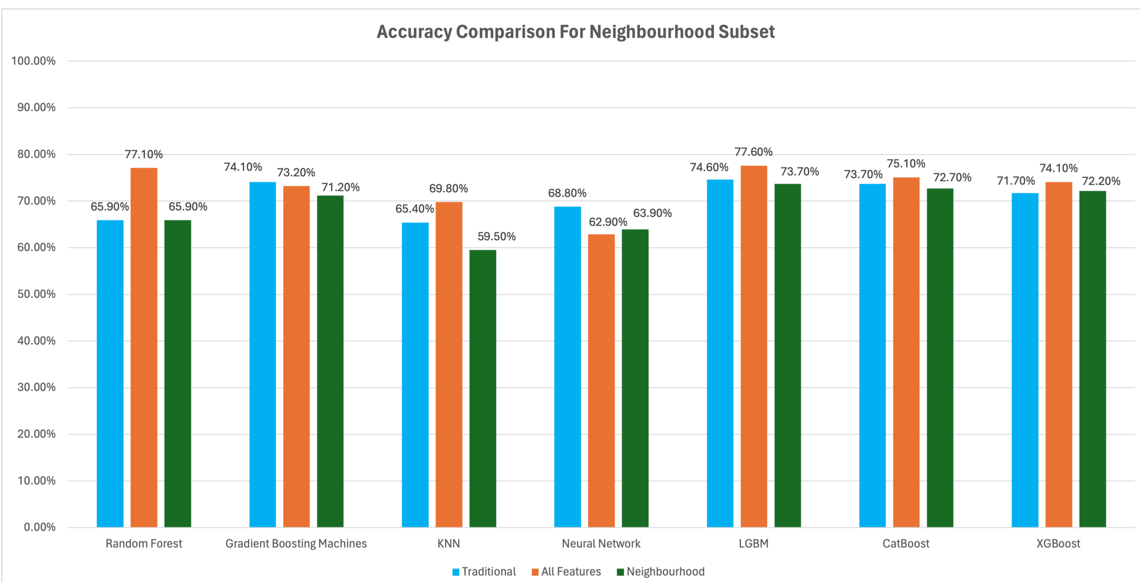



**Figure 4.7:** Impact of Neighbourhood Features on Model Performance

**Personal Factors:** Incorporating personal data (Figure 4.8) impacts model performance, revealing its value for stroke risk prediction. XGB sees a substantial accuracy gain (72.20%) from personal features, highlighting their positive influence. Notably, GBM achieve their peak accuracy (77.10%) with personal features, suggesting these features hold the strongest predictive power within this model. KNN shows a moderate benefit (68.80%) from personal data, while LGBM (75.60%) and XGB (75.60%) experience slight decreases compared to the full feature set but remain improved over traditional features alone. Interestingly, the NN struggles with personal features (61.50%), indicating a potential mismatch between this data type and the model architecture. CB also shows a minor decrease (74.10%) from the full set but performs better than with only traditional features (73.70%). Personal features generally enhance model performance compared to traditional features, with some models even achieving peak accuracy with them.

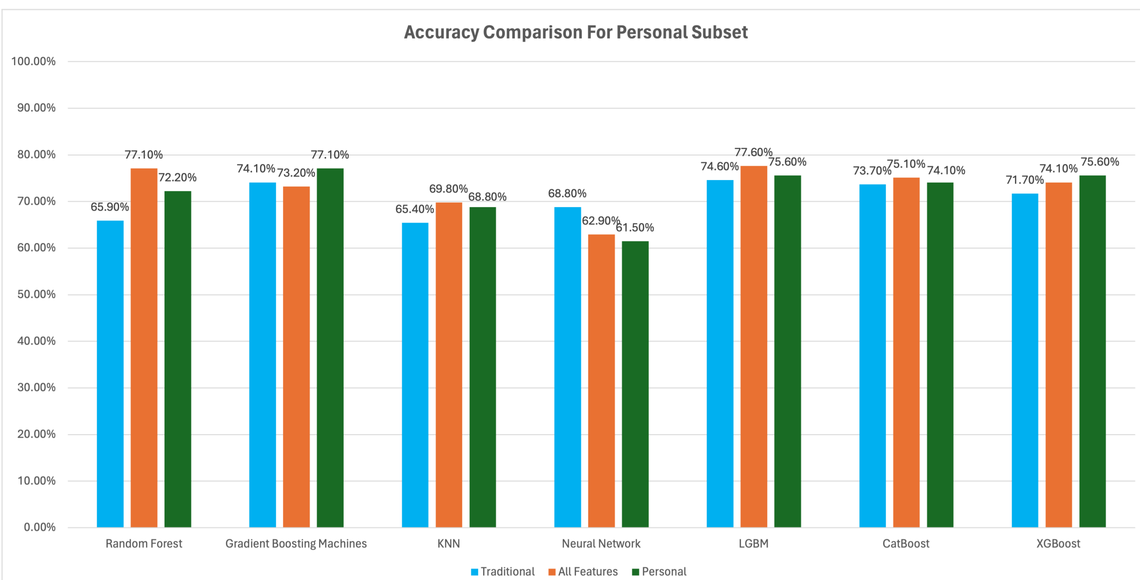



**Figure 4.8:** Impact of Personal Features on Model Performance

**Race:** Examining the influence of race (Figure 4.9) on stroke prediction models reveals different effects. KNN (65.40%) experience minimal changes, suggesting race offers little additional predictive power beyond traditional features. GBM (74.10%) and LGBM (74.10%) show slight accuracy dips, indicating race has some value but isn't a major factor. CB (75.10%) remains stable, demonstrating its ability to utilize race without compromising performance. Interestingly, the NN shows a slight improvement (66.80%), suggesting some benefit from race features. Overall, the impact of race on model accuracy is varied, with some models indifferent and others experiencing minor changes. This highlights that race may hold some predictive power, but its influence is often subtle compared to other features.

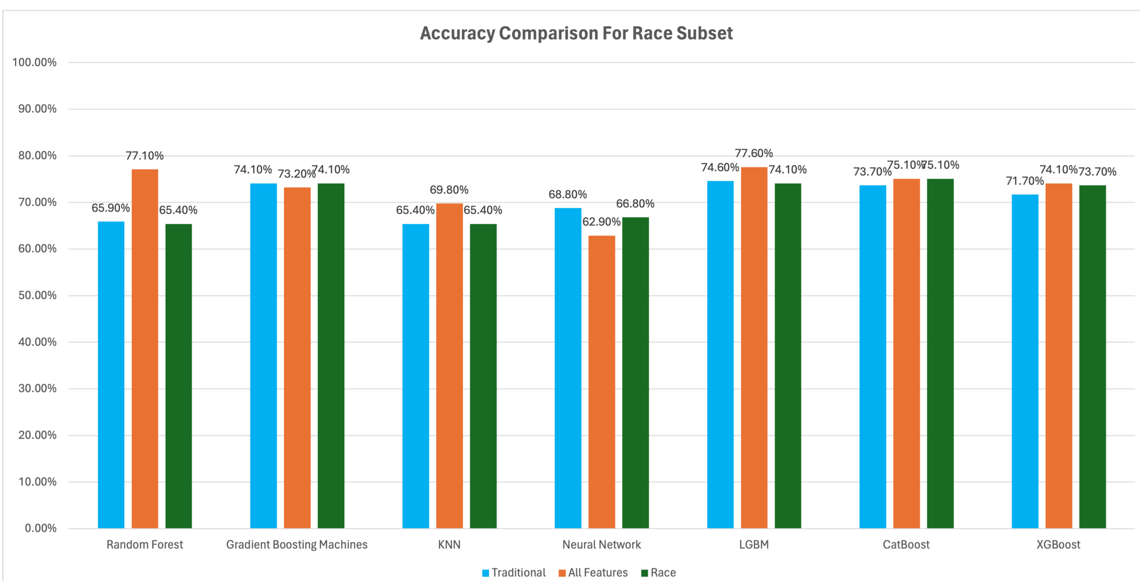



**Figure 4.9:** Impact of Race on Model Performance:

**Social Factors:** Incorporating social features (Figure 4.10) into the models yields mixed results, impacting each model's accuracy differently compared to both traditional and all feature sets. RF (68.30%) and LGBM (74.10%) experience accuracy gains from social data, but these increases are smaller than those observed with the full feature set. Conversely, GBM (71.20%) and KNN (65.40%) see their performance decline when using social features alone compared to both traditional and all features. The NN (63.40%) shows a slight improvement with social data, but it remains lower than its peak accuracy. Interestingly, CB (74.10%) performs equally well with social features as with traditional features, and XGB (74.10%) maintains its top accuracy from the full feature set even when using social features alone. Social features tend to provide some benefit or maintain performance compared to traditional features.

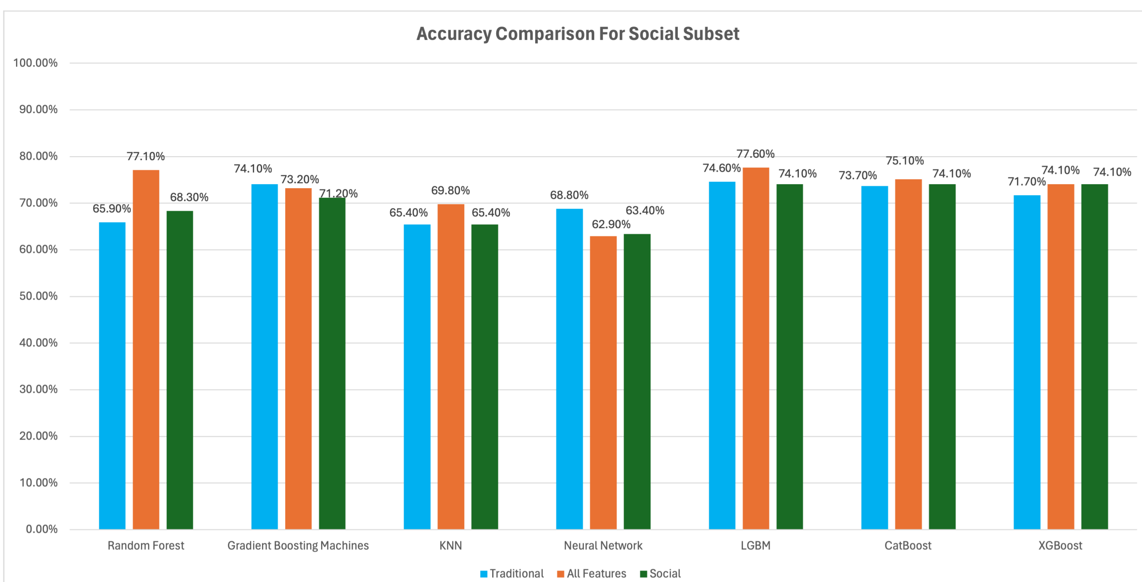



**Figure 4.10:** Impact of Social Features on Model Performance

Results show that adding SDoH subsets improves model accuracy in some cases, giving a better picture of what affects stroke risk and boosting prediction strength. Demographic, economic, personal and social subset had the most positive impact on performance. The models GBM, LGBM, CB, and XGB performed best in all scenarios.

## 4.2 XAI

We employed XAI methods such as feature importance, SHAP, and LIME which allowed us to pinpoint the specific SDoH factors that play important roles and offer insights into their relative importance.

### 4.2.1 Feature Importance

We analyse the feature importance plot based on feature ranking and the relative importance of each feature. The order of the features from top to bottom indicates the rank of importance as assessed by the model. Features higher on the y-axis contribute more to the model's decisions. Whereas, the length of the bars represents the relative importance of each feature. A longer bar means a higher score and, therefore, a greater impact on the model's predictions. Such plots serve as tools for interpretability, helping to explain the model's behavior in a human-understandable way.

**CatBoost:**

In the CB model (Figure 4.11), the type of smoker (Smoker_Type) and the presence of heart disease (Heart_Disease) stand out as the foremost influential factors, highlighting the significance of lifestyle habits and medical conditions on predicting stroke. Demographic aspects such as age (DHH_AGE) and marital status (DHH_MS) also hold substantial weight, showing the model's attention to personal and social circumstances. Delving further, the model considers both traditional risk factors, such as high cholesterol (High_Cholesterol) and blood pressure medication usage (High_BP_took_meds), and social factors like stress perception (Perceived_Life_Stress), employment status (Working), alcohol consumption (Had_Alcohol_12mo), and health perception (Perceived_Health). This mix highlights the influence of SDoH factors along with clinical indicators on stroke using CB model.

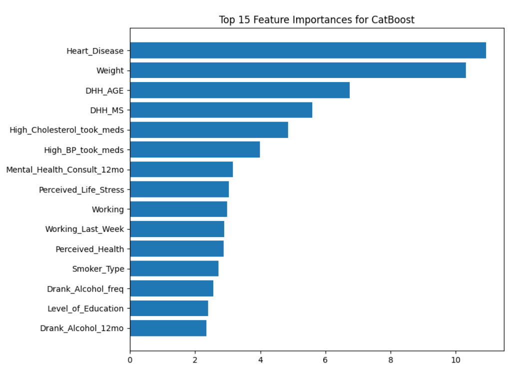



**Figure 4.11:** Feature Importance - CatBoost

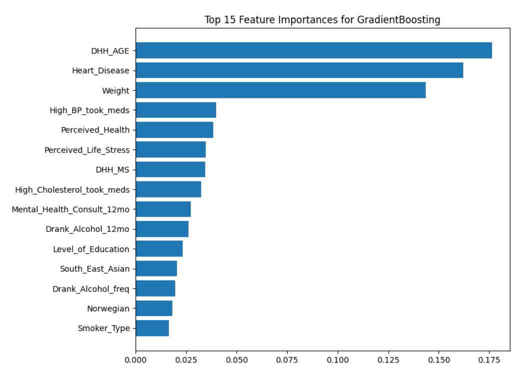



**Figure 4.12:** Feature Importance - GBM

**Gradient Boosting Machine:**

In the GBM model (Figure 4.12), age (DHH_AGE) is identified as the most crucial factor, emphasizing its role in stroke predictions. The importance of smoking type (Smoker_Type) is also highlighted, maintaining its significance across different models. Following these are medical conditions and social factors such as heart disease (Heart_Disease) and marital status (DHH_MS), showing their consistent impact. Features representing health perceptions and lifestyle choices, including perceived health (Perceived_Health), blood pressure medication usage (High_BP_took_meds), and alcohol consumption (Had_Alcohol_12mo), form the middle tier of importance. Furthermore, the model considers the effects of stress perception (Perceived_Life_Stress) and education level (Level_of_Education), albeit with less weight than the leading factors. This illustrates that SDoH had an impact on GBM's assessment of stroke along with traditional factors.

**LGBM:**

In the LGBM model (Figure 4.13), smoking type (Smoker_Type) and age (DHH_AGE) emerge as top predictors, with the model also placing significant emphasis on stress perception (Perceived_Life_Stress) and health perception (Perceived_Health), sometimes even more than on traditional medical factors such as heart disease (Heart_Disease). Following these, factors like marital status (DHH_MS), blood pressure medication usage (High_BP_took_meds), and doctor consultations in the past year (Dr_Consult_12mo) further contribute to the accuracy of predictions. Notably, this model distinguishes itself by valuing unique features such as living environment (Urban_Rural) and ethnicity (Norwegian), showcasing its consideration of SDoH influence on stroke.

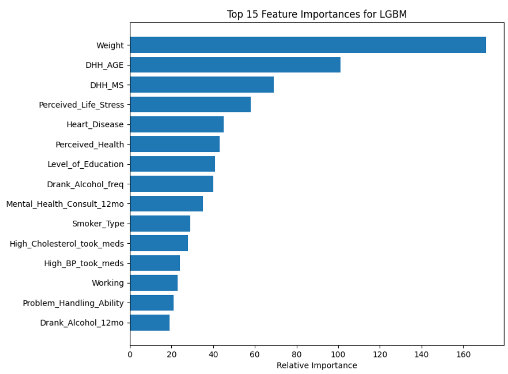



Figure 4.13: Feature Importance - LGBM

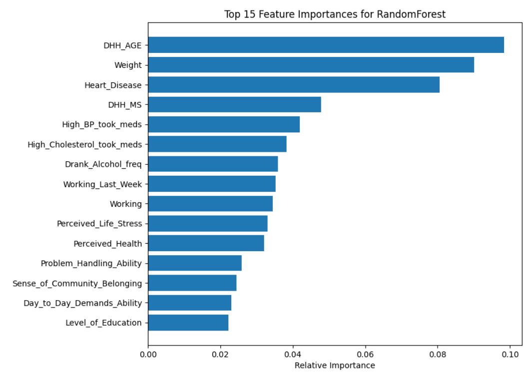



Figure 4.14: Feature Importance - RF

**Random Forest:**

In the RF model (Figure 4.14), the significance of smoking type (Smoker_Type) once again tops the list, followed by age (DHH_AGE) and heart disease (Heart_Disease), with marital status (DHH_MS) also highlighted as a key factor. Further down the importance scale, the model benefits from information provided by blood pressure medication usage (High_BP_took_meds), alcohol drinking frequency (Drink_Alcohol_freq), and high cholesterol (High_Cholesterol), among other features. Additionally, employment status (Working), education level (Level_of_Education), and sense of community (Community_Belonging) contribute to the model's decision-making, although with a more subtle impact, demonstrating the role of SDoH in RF's assessment of stroke.

**XGBoost:**

In the XGB model (Figure 4.15), heart disease (Heart_Disease) emerges as the most influential factor, underscoring its role in predictions. A closely ranked group of features, including blood pressure medication usage (High_BP_took_meds), ethnic backgrounds (Chinese, Norwegian), high cholesterol (High_Cholesterol), and age (DHH_AGE), illustrates the model's consideration of a blend of medical and demographic factors. Ethnicities such as Chinese and Norwegian highlight the model's attention to demographic diversity. Lower in the importance order, ethnicities like South East Asian, alongside lifestyle choices and medical consultations (Had_Alcohol_12mo, Dr_Consult_12mo), and other ethnic backgrounds (German, Metis), contribute to varying degrees. Closing the list of top features are type of smoker (Smoker_Type), marital status (DHH_MS), English, and health perception (Perceived_Health), which, despite being at the end, still influence the model. This showcases the role SDoH plays in XGB's prediction for stroke.

We weren't able to generate feature importance plots for KNN and NN in this analysis. Unlike other models, KNN doesn't assign individual weights to features, relying on data point proximity for predictions. NN, with their complex layers and transformations, make it difficult to isolate the contribution of each feature.

The analysis of feature importance across different models consistently highlights a few

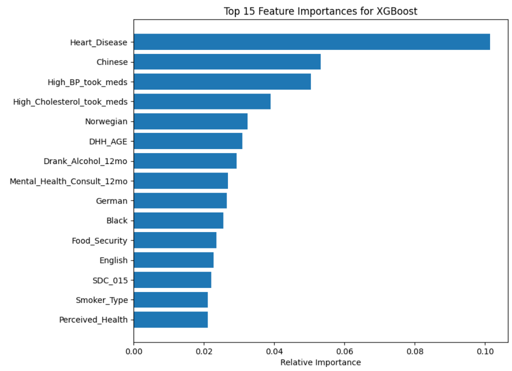



**Figure 4.15:** Feature Importance - XGBoost

key predictors for stroke, notably the type of smoker (Smoker_Type), age (DHH_AGE), and the presence of heart disease (Heart_Disease). These factors are frequently top-ranked across models, indicating their influence on stroke prediction. In addition, demographic aspects such as marital status (DHH_MS) and diverse SDoH like employment status (Working), perceived health (Perceived_Health), and stress levels (Perceived_Life_Stress) also play important roles, although their impact varies among the models.

### 4.2.2 SHAP

SHAP, offers insights into the model's decision-making on a comprehensive level while still detailing individual predictions. In stroke prediction, a red bar in the SHAP plot signifies that a feature is positively influencing the prediction of a stroke occurring. In contrast, a blue bar would indicate a positive influence toward predicting no stroke. The length of the bars in SHAP plots denotes the relative importance of each feature, with longer bars indicating a stronger impact on the model's output.

**CatBoost:**

For the CB model (Figure 4.16), smoking behavior (Smoker_Type) and age (DHH_AGE) are important in the model's predictions, highlighting their role in health outcomes. Interestingly, the model gives considerable importance to personal health perception (Perceived_Health), mental well-being (Perceived_Life_Stress), and social connections (Sense_of_Community_Belonging). These findings reveal the model's sensitivity to both health-related behaviors and broader social factors. The presence of medical condition management features like high blood pressure medication (High_BP_took_meds), and cholesterol levels (High_Cholesterol) further indicates the model's integration of traditional health metrics, reinforcing the influence of SDoH on stroke prediction.

**Gradient Boosting Machine:**

In the GBM model (Figure 4.17), weight (Weight) emerges again as a dominant feature, consistent with its significance in other models. The continued prominence of non-traditional factors such as stress perception (Perceived_Life_Stress) and community ties (Sense_of_Community_Belonging) across models suggests a universal impact of these factors on health predictions. The recurrence of features related to heart disease (Heart_Disease), high blood pressure medication (High_BP_took_meds), and cholesterol

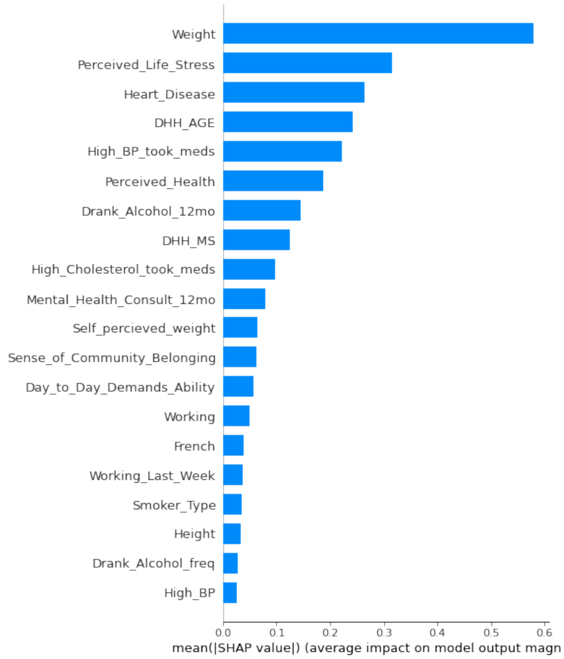



**Figure 4.16:** SHAP Plot - CatBoost

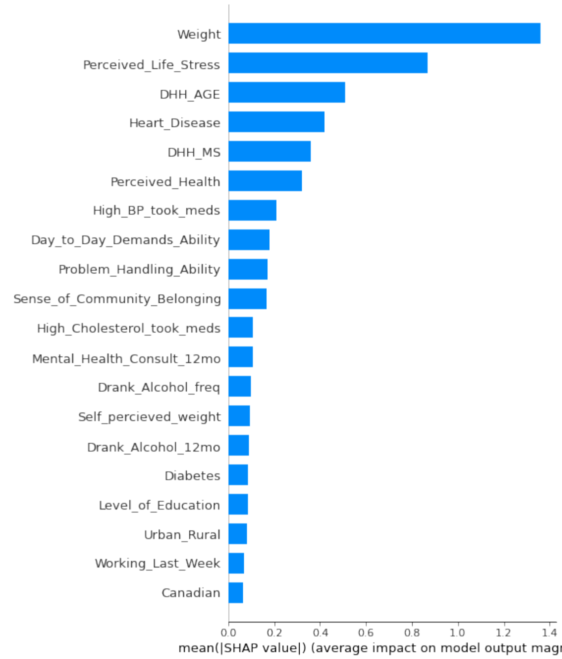



**Figure 4.17:** SHAP Plot - GBM

levels (High_Cholesterol) across different models underscores their relevance in predicting stroke along with the important role of SDoH.

**LGBM:**

In the LGBM model (Figure 4.18), the consistency in highlighting weight (Smoker_Type) as an influencer aligns with findings from other models, pointing to its integral role in health outcomes. The model places significant weight on an individual's health perception (Perceived_Health) and stress levels (Perceived_Life_Stress), indicating that how a person views their health and handles stress is crucial for prediction. Lifestyle habits like blood pressure management (High_BP_took_meds) and alcohol consumption frequency

(Drank_Alcohol_freq) are also key, emphasizing the role of SDoH in influencing health.

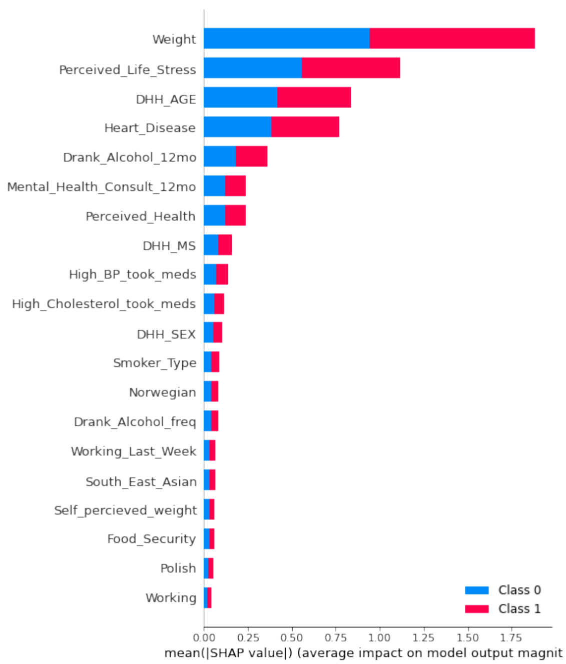



**Figure 4.18:** SHAP Plot - LGBM

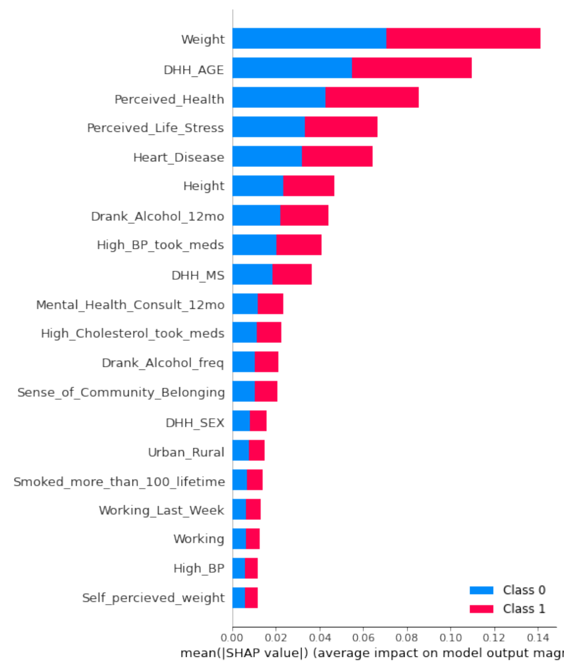



**Figure 4.19:** SHAP Plot - Random Forest

**Random Forest:**

For the RF model (Figure 4.19), weight, type of smoker (Smoker_Type) again ranks as a primary influencer, underscoring its consistent importance across models in predicting health outcomes. The model gives considerable importance to mental well-being (Perceived_Life_Stress) and social connections (Sense_of_Community_Belonging), echoing the sentiment of other models about the impact of psychological and social factors. Additionally, it highlights the relevance of managing health conditions like high cholesterol (High_Cholesterol) and lifestyle choices (Drank_Alcohol_freq), showcasing the model's capability to integrate both SDoH factors and traditional health metrics in its assessment.

**XGBoost:**

The XGB model (Figure 4.20) aligns with the trend seen in other models, where weight is the most dominant feature, reinforcing the influence of lifestyle habits on health. The model places considerable emphasis on non-traditional features such as stress perception (Perceived_Life_Stress) and community belonging (Sense_of_Community_Belonging), signifying its attentiveness to social and psychological aspects of health. Alongside these, traditional health indicators like heart disease (Heart_Disease), blood pressure medication usage (High_BP_took_meds), and cholesterol levels (High_Cholesterol) maintain their established importance, reflecting the model's comprehensive approach to include both medical conditions and SDoH factors in stroke prediction.

For SHAP analysis, KNN and NN require specialized versions of SHAP explainers due to their unique model architectures. Unfortunately, these specific SHAP explainers were incompatible with the existing software environment on ISQ's office machines, leading to technical installation issues. This limitation prevented the effective application of SHAP-based interpretability methods for these models within the ISQ infrastructure.

SHAP analysis consistently identifies smoking behavior, age, sense of community belonging, perceived health, and stress levels as key predictors across models, along with traditional health metrics like medication use and cholesterol levels. This underscores the

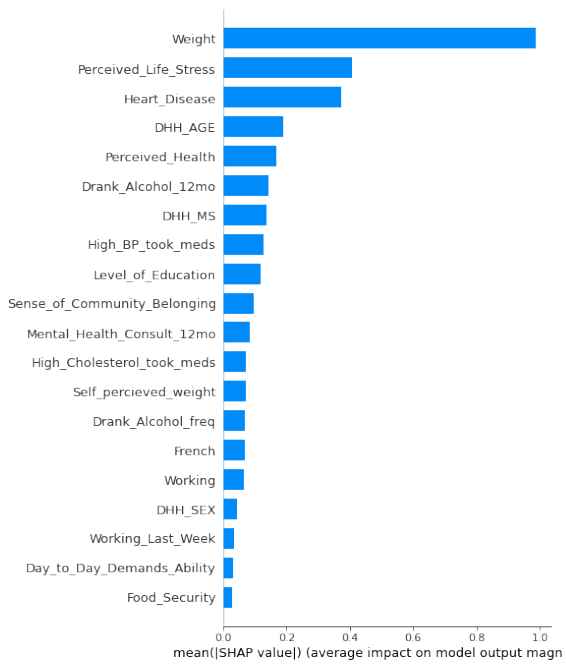



**Figure 4.20:** SHAP Plot for XGBoost

importance of integrating SDoH factors along with traditional medical factors in stroke risk assessment.

### 4.2.3 Lime

LIME provides a detailed look at how each feature affects an individual prediction in a model. In the context of stroke prediction, a red bar in the LIME plot means that the feature is contributing to the model predicting the absence of stroke, while a green bar suggests a contribution toward a stroke prediction. The length of these bars indicates the strength of each feature's impact, with longer bars signifying greater influence on the prediction for that specific patient's case.

**Catboost:**

For the CB model (Figure 4.21), the prediction for a stroke incident is heavily influenced by the presence of heart disease (Heart_Disease) and the age (DHH_AGE) of the individual, with both of them being the top contributing features. The model interprets these factors as strong indicators, where existing heart conditions and increased age raise the model's prediction probability for a stroke. Other features such as blood pressure medication (High_BP_took_meds), cholesterol management (High_Cholesterol_took_meds), and recent alcohol consumption (Drank_Alcohol_12mo) also play roles in the predictive process, but to a lesser extent.

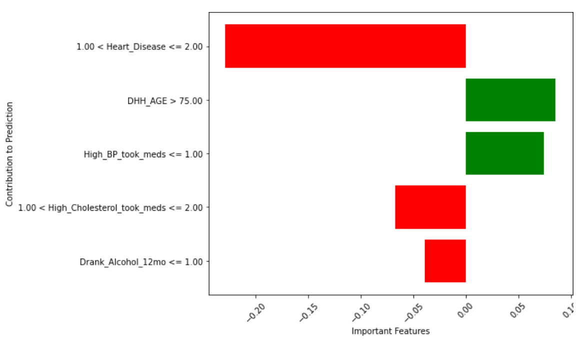



**Figure 4.21:** LIME plot - CatBoost

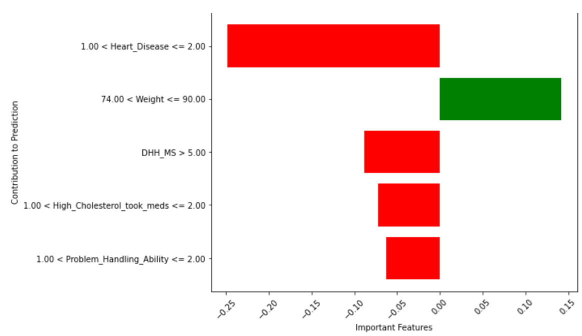



**Figure 4.22:** LIME plot - GBM

**Gradient Boosting Machine:**

In the GBM model's analysis (Figure 4.22), features like body weight (Weight) and an individual's problem-handling ability (Problem_Handling_Ability) appear to have a protective effect, reducing the likelihood of a stroke prediction. Conversely, the presence of heart disease (Heart_Disease) is interpreted as increasing stroke risk. The model also considers marital status (DHH_MS) and cholesterol medication (High_Cholesterol_took_meds) when making its prediction, reflecting the way the model weighs various health-related factors and SDoH attributes in assessing stroke risk.

**KNN:**

For the KNN model (Figure 4.23), the prediction leans towards 'Stroke', with heart disease and cholesterol levels (Heart Disease, High_Cholesterol_took_meds) marked as significant contributors. The model's prediction also takes into account blood pressure medication usage (High_BP_took_meds), blood pressure levels (High_BP), and smoking type (Smoker_Type), each playing a role in influencing the likelihood of a stroke prediction, though their impact is secondary to that of heart-related factors.

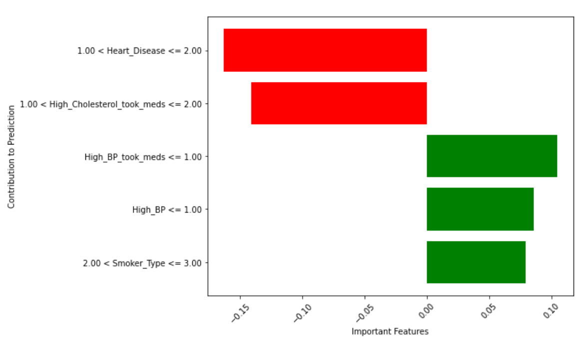



**Figure 4.23:** LIME plot - KNN

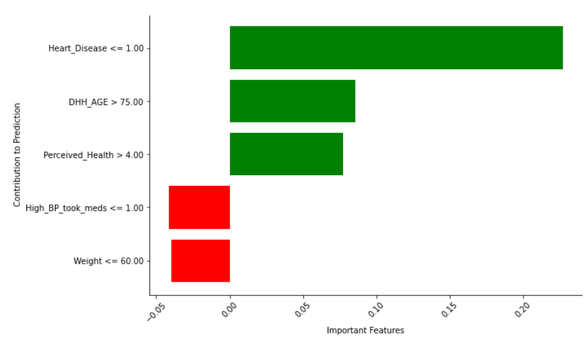



**Figure 4.24:** LIME plot - LGBM

**LGBM:**

In the case of LGBM (Figure 4.24), an individual's perceived health (Perceived_Health) and age (DHH_AGE) emerge as the main influences. Interestingly, this suggests that while advanced age is generally considered a risk factor, a positive health perception might mitigate this risk according to the model. Other factors the model considers include heart disease (Heart_Disease), medication for high blood pressure (High_BP_took_meds), and weight (Weight), all contributing to the model's assessment of stroke risk, highlighting the interplay between medical conditions and SDoH.

**XGBoost:**

In the XGB analysis (Figure 4.25), the prediction leans towards 'Stroke', with heart disease (Heart_Disease) and specific weight parameters (Weight in the range 61.00 to 74.00) identified as important contributors. Additionally, marital status (DHH_MS), and the usage of medication for high cholesterol and high blood pressure (High_Cholesterol_took_meds and High_BP_took_meds), also play significant roles. The visualization suggests that the presence of heart disease and being within the specified weight range greatly increase the likelihood of a stroke prediction, highlighting the model's attention to these specific health indicators and treatments as significant risk factors along with few SDoH factors.

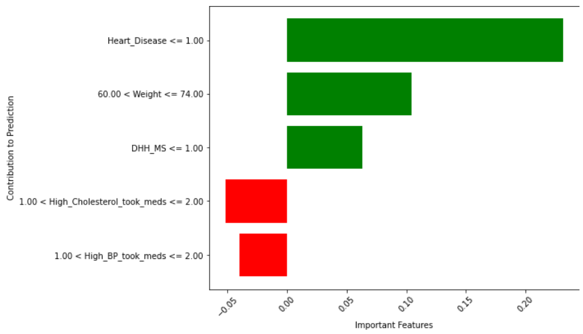



**Figure 4.25:** LIME plot - XGBoost

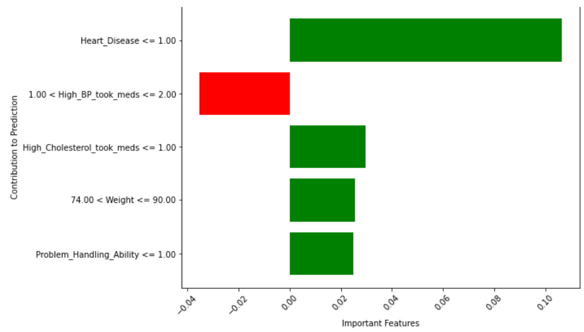



**Figure 4.26:** LIME plot - RF

**Random Forest:**

For the RF model (Figure 4.26), the prediction is mainly driven by presence of heart disease (Heart_Disease) and a different weight range (Weight in the range 74.00 to 80.00). This model also considers the impact of high blood pressure and cholesterol medication (High_BP_took_meds and High_Cholesterol_took_meds), as well as the individual's problem-handling ability (Problem_Handling_Ability). The orientation of green bars for these features suggests that despite the presence of heart disease, the specific weight range and how an individual manages problems play crucial roles in predicting stroke. This demonstrates the influence of SDoH factors in combination with medical conditions in assessing stroke.

**Neural Network:**

For the NN model (Figure 4.27), the prediction indicates biological sex (DHH_SEX) and the presence of heart disease (Heart_Disease) emerging as the dominant factors influencing the presence of stroke. The inclusion of features such as access to a regular healthcare provider (Regular_Healthcare_Provider), usage of cholesterol medication (High_Cholesterol_took_meds), and smoking habits (Smoker_Type) further shapes the prediction landscape. This analysis underlines the NN's capacity to integrate both SDoH metrics and broader health-related behaviors in formulating its predictions, illustrating the interplay of various factors in stroke prediction.

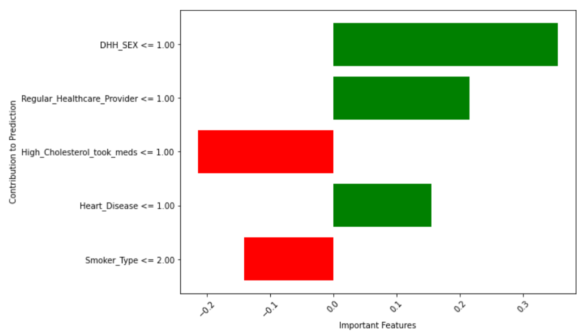



**Figure 4.27:** LIME plot for NN

LIME analysis across multiple models emphasizes the strong impact of heart disease and age on stroke predictions. Additional important contributors include medication usage for blood pressure and cholesterol, as well as lifestyle factors like recent alcohol consumption and weight. This multifactorial influence highlights the importance of integration of both traditional medical indicators and broader SDoH in assessing stroke risk.

# Chapter 5

# Discussion

Our discussion delves into the implications of incorporating SDoH data into ML models for stroke prediction. We utilized ML models ( such as LGBM, XGB, RF, GBM, CB, and NN) alongside interpretability techniques (such as feature importance, LIME and SHAP) to understand how SDoH data influences model performance and decision-making. Our results align with previous research suggesting that at least for specific ML models, including SDoH factors improves stroke prediction accuracy [101, 102].

## 5.1 Impact of SDoH on Model Performance

The AUC scores were calculated considering all features, encompassing both traditional and all SDoH categories. Among these models, the LGBM achieved the highest AUC score (81%), highlighting its effectiveness in utilizing complex data for accurate predictions. This aligns with studies showing the strengths of gradient boosting methods in other medical prediction tasks [103]. However, our NN model achieved a lower AUC score (65%). This suggests potential limitations in the NN model's suitability for this specific data or an extensive need for further hyperparameter tuning. Similar findings have been reported in other studies, where the effectiveness of all models, in our case NN, can be data-dependent [104]. XGB also performed well with an AUC of 80%, further establishing the efficacy of gradient boosting methods in this context. Ensemble methods like RF, GBM, and CB also demonstrated better performance each with AUC scores of 79%. These findings suggest their capability of handling features effectively, which aligns with research highlighting the strengths of ensemble methods for complex data analysis [105].

These observations emphasize several key takeaways. First, model choice plays an important role when predicting stroke. Based on the results, models like RF, XGB, LGBM, and CB seem well-suited for handling clinical and SDoH data due to their ability to leverage complex features. Second, identifying relevant featues is important for stroke prediction. As observed from our analysis, GBM's performance demonstrated sensitivity to the type of features used. Finally, the lower AUC score achieved by the NN model suggests potential challenges in utilizing SDoH data effectively. This may necessitate more sophisticated model design or optimization strategies for NN approaches. Thus, our analysis underscores the importance of selecting the appropriate model based on available data and the specific prediction task. It also shows that the inclusion of SDoH data enhances model performance for stroke prediction.

## 5.2 Role of SDoH in Stroke Prediction

Addressing concerns about the interpretability of AI models in healthcare, our research focused on integrating XAI techniques and SDoH data into stroke prediction models. LIME and SHAP, as XAI methods, proved crucial in enhancing the interpretability of our models [106]. They clarified how both traditional health metrics and SDoH factors influence predictions. This transparency reinforces the importance of including medical data, alongwith demographic, economic, social, and personal factors. This aligns with the growing emphasis on a “whole patient" approach in healthcare [107]. “Whole patient” care involves assessing lifestyle, comorbidities, communication, mental health, socioeconomic status, and medication issues – all areas impacted by SDoH [107]. By addressing these broader factors, our approach aims to provide more holistic and patient-centred solutions, potentially fostering wider adoption of AI in clinical settings.

XAI techniques provided valuable insights into the role of SDoH within our stroke prediction models. Studies suggest integrating demographic and economic features enhances model accuracy [101] and our findings corroborate this. However, the degree of improvement in predictions varied across different models, reflecting the diverse mechanisms employed by machine learning models to handle these features. This aligns with existing research highlighting the challenges of integrating SDoH into ML models [108]. Ethnicity and mental health features showed less consistent effects on model performance, indicating a complicated relationship between these factors and health outcomes. This suggests a need for more refined feature engineering to optimize their integration. Similar findings have been reported in other studies, where the effectiveness of incorporating complex data can be data-dependent and require careful feature engineering for optimal impact [109].

Interestingly, incorporating social and personal features alongside clinical predictors such as heart disease and weight positively impacted model performance. These additional features provided valuable context, enhancing the models' understanding of the interplay between SDoH and stroke risk. This aligns with the growing recognition of the importance of a holistic approach in healthcare AI, where considering social determinants alongside traditional medical data leads to more comprehensive and informative models [48]. Notably, race emerged as a significant predictor in certain cases, highlighting its potential importance in stroke prediction. However, it's crucial to acknowledge the complexities surrounding the use of race in medical AI, ensuring its responsible integration to avoid potential biases [72].

Research has established the link between mental health and physical health [110]. Consistent with this, perceived health and life stress emerged as prominent factors in our study, underlining the crucial role of an individual's mental well-being. The sense of community belonging also emerged as a notable factor, indicating the impact of social support on health outcomes. Employment status and the ability to handle daily activities further reinforced the connection between economic stability and daily functional capacity and health outcomes, as suggested in existing literature [111]. Additionally, urban or rural residency influenced health likely through factors like access to healthcare and environmental conditions. Finally, the level of education played a role, suggesting its impact on overall health and well-being, consistent with findings reported in the literature [112]. These insights highlight the importance of a comprehensive healthcare approach that integrates SDoH alongside traditional health metrics. This holistic approach fosters a deeper understanding of individual health outcomes, ultimately facilitating improved model performance and providing actionable insights for targeted health interventions.

XAI techniques proved crucial in pinpointing the most influential factors across different models, aligning with research highlighting the importance of interpretability for understanding model behavior [89]. For example, in CB and GBM models, XAI methods revealed smoking habits and heart disease as top features. This aligns with established knowledge regarding the strong influence of lifestyle choices and existing health conditions on stroke risk [113]. LGBM's focus on psychosocial factors, as uncovered by XAI, further aligns with studies demonstrating the link between mental well-being and stroke risk [114]. Additionally, XGB's emphasis on demographic factors like "Chinese" and "Norwegian" underscored the model's sensitivity to ethnic backgrounds, a complex interaction of medical and demographic factors that XAI helps to elucidate [101].

LIME, as a vital XAI tool, provided in-depth insights into individual predictions of various models, revealing how specific features influence stroke risk predictions. For example, in the CB model, LIME highlighted heart diseases and age as key factors, illustrating how older age and existing heart conditions substantially increase stroke risk. Similarly, in LGBM, age and how an individual perceives their health were significant, indicating a relationship where an individual's self-perceived health status could mitigate or amplify the risk posed by age. These insights from LIME, similar to findings in other studies [115, 116], help understand the model's reasoning at an individual prediction level, offering a transparent view of how different health-related factors are weighted and interpreted.

SHAP, another XAI method, provided a broader, more aggregated view of feature importance across multiple models. This aligns with the strengths of SHAP for offering global explanations of model behavior [26]. It identified and quantified the impact of both traditional health metrics and non-traditional SDoH factors. For instance, weight was consistently a dominant factor in models like CB, GBM, and LGBM, as revealed by SHAP, highlighting its significant role in health predictions, which aligns with existing research [117]. Furthermore, SHAP brought attention to the importance of psychosocial factors such as self-perceived health, self-perceived life stress, and sense of community belonging, emphasizing the need to consider both physical and mental well-being in stroke risk assessment, as supported by prior studies [114].

The distinct importance placed on SDoH across various models points to the potential for personalized healthcare interventions. This aligns with the growing interest in utilizing AI for personalized medicine [118]. By understanding the specific SDoH that influence stroke risk for individuals or groups, healthcare providers can tailor interventions more effectively, potentially mitigating risk through targeted strategies. These targeted interventions could address factors like access to healthy food, stress management techniques, or social support networks, potentially leading to improved health outcomes.

## 5.3 Limitations

Our study, while extensive, has limitations inherent to the scope of data and methodologies applied. Our study's scope is bounded by the dataset provided by the ISQ and the chosen methodologies, limiting our ability to capture the full larger population and SDoH that influence stroke. The use of machine learning models and XAI techniques like LIME and SHAP, while beneficial for interpretability, cannot completely untangle the complex relationship between clinical and SDoH. Notably, the NN model exhibited poorer performance compared to other models, suggesting potential mismatches between model complexity and dataset characteristics. The correlational nature of our analysis limits our capacity to infer causality, underscoring the need for methodologies capable of establishing direct cause-and-effect relationships. Additionally, the inconsistent impact of features such as ethnicity and mental health on model performance underscores the necessity for refined feature engineering and broader data sources to improve predictive precision and model robustness.

## 5.4 Future Work

To overcome the limitations mentioned above, future directions of research will focus on broadening the dataset to include a wider range of SDoH, potentially from diverse geographic and socio-economic contexts. Enhancing the dataset will facilitate an in-depth analysis of underrepresented determinants. Moreover, advancing machine learning methodologies and exploring newer XAI approaches could elevate model accuracy and interpretability. Investigating the longitudinal effects of SDoH on stroke risk is another important area, requiring different models capable of adapting to temporal changes in risk factors. Furthermore, building on the correlations identified in this study, future research should focus on delineating causality to more accurately guide intervention strategies. Addressing these specific challenges will deepen our understanding of stroke risk factors and pave the way for more targeted and fair healthcare solutions.

# Chapter 6

# Conclusion

This study investigated the role of SDoH in enhancing model performance to predict strokes. This research work addresses the limitations identified in Chapter 1, where traditional models solely reliant on clinical data can lead to biased and inaccurate results. We address this gap by integrating SDoH data along with clinical data while prioritizing fairness and explanability in healthcare.

Building upon the foundations established in Chapters 2 and 3, our study confirmed the importance of SDoH in stroke incidence. Chapter 2 explored the existing literature on the relationship between SDoH and stroke risk in individuals, while Chapter 3 detailed the methodological approaches for incorporating various SDoH factors into AI models. Our method involved rigorous data pre-processing, model development, and the application of XAI techniques. XAI, as discussed in Chapter 3, plays a crucial role in understanding how the model arrives at its predictions, ensuring transparency and addressing ethical implications.

The findings in Chapter 4, demonstrate that incorporating SDoH data improves model accuracy. This chapter also delves into how different social factors influence stroke risk. Notably XAI analysis revealed that SDoH factors like demographics, economic status, and personal factors influenced stroke risk predictions. This underscores the importance of considering SDoH alongside traditional medical data for a more comprehensive understanding of stroke risk, as emphasized throughout this study.

Chapter 5 elucidates how our research contribute to healthcare AI in several ways. Firstly, we demonstrate the role of SDoH in stroke prediction, advocating for a more holistic approach. Secondly, we highlight the potential for AI to promote healthcare equity by integrating SDoH data. Finally, the research emphasizes the importance of developing ethically sound and transparent models that address the complexities of SDoH, as outlined in Chapter 5. This work paves the way for an inclusive, equitable AI-powered healthcare system capable of delivering personalized care and ultimately reducing health disparities.

In conclusion, this study establishes the crucial role of SDoH in stroke prediction, advocating for a holistic approach in healthcare AI. By integrating SDoH data and employing XAI techniques, we can build fairer, more accurate, and interpretable models.

# Bibliography

[1] “The top 10 causes of death.” https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death. [Accessed: 11-March-2024].

[2] “WHO EMRO | Stroke, Cerebrovascular accident | Health topics.” https://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/index.html. [Accessed: 11-March-2024].

[3] P. Braveman and L. Gottlieb, “The social determinants of health: it’s time to consider the causes of the causes,” *Public Health Rep*, vol. 129 Suppl 2, pp. 19–31, Jan. 2014.

[4] “Overview of all 12 Social Determinants of Health (SDOH) &x2013; Post-Secondary Peer Support Training Curriculum.” https://opentextbc.ca/peersupport/chapter/overview-of-all-12-social-determinants-of-health-sdoh/. [Accessed: 23-February-2024].

[5] J. M. D. Raphael, “The canadian facts.” https://thecanadianfacts.org/the_canadian_facts.pdf. [Accessed: 11-March-2024].

[6] A. Bohr and K. Memarzadeh, “The rise of artificial intelligence in healthcare applications,” in *Artificial Intelligence in Healthcare*, pp. 25–60, Elsevier, 2020.

[7] S. A. Alowais, S. S. Alghamdi, N. Alsuhebany, T. Alqahtani, A. I. Alshaya, S. N. Almohareb, A. Aldairem, M. Alrashed, K. Bin Saleh, H. A. Badreldin, M. S. Al Yami, S. Al Harbi, and A. M. Albekairy, “Revolutionizing healthcare: the role of artificial intelligence in clinical practice,” *BMC Med. Educ.*, vol. 23, p. 689, Sept. 2023.

[8] J. Yim, R. Chopra, T. Spitz, J. Winkens, A. Obika, C. Kelly, H. Askham, M. Lukic, J. Huemer, K. Fasler, G. Moraes, C. Meyer, M. Wilson, J. Dixon, C. Hughes, G. Rees, P. T. Khaw, A. Karthikesalingam, D. King, D. Hassabis, M. Suleyman, T. Back, J. R. Ledsam, P. A. Keane, and J. De Fauw, “Predicting conversion to wet age-related macular degeneration using deep learning,” *Nature Medicine*, vol. 26, p. 892–899, May 2020.

[9] S. Somashekhar, M.-J. Sepúlveda, S. Puglielli, A. Norden, E. Shortliffe, C. Rohit Kumar, A. Rauthan, N. Arun Kumar, P. Patil, K. Rhee, and Y. Ramya, “Watson for oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board,” *Annals of Oncology*, vol. 29, p. 418–423, Feb. 2018.

[10] S. Joshi, M. Sharma, R. P. Das, J. Rosak-Szyrocka, J. Żywiołek, K. Muduli, and M. Prasad, “Modeling conceptual framework for implementing barriers of AI in

public healthcare for improving operational excellence: Experiences from developing countries," *Sustainability*, vol. 14, p. 11698, Sept. 2022.

[11] S. Gerke, T. Minssen, and G. Cohen, "Chapter 12 - ethical and legal challenges of artificial intelligence-driven healthcare," in *Artificial Intelligence in Healthcare*, pp. 295–336, Academic Press, 2020.

[12] "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI." `https://ssrn.com/abstract=3248829`. [Accessed: 11-March-2024].

[13] CDC, "Social Determinants of Health." `https://www.cdc.gov/about/sdoh/index.html`. [Accessed: 11-March-2024].

[14] "Health Care Access and Quality." `https://www.ncbi.nlm.nih.gov/books/NBK596397/`. [Accessed: 11-March-2024].

[15] "Mirror, Mirror 2021: Reflecting Poorly." `https://www.commonwealthfund.org/publications/fund-reports/2021/aug/mirror-mirror-2021-reflecting-poorly`. [Accessed: 11-March-2024].

[16] "Universal Health Coverage Data." `https://datatopics.worldbank.org/universal-health-coverage/`. [Accessed: 11-March-2024].

[17] "nationalacademies.org." `https://www.nationalacademies.org/our-work/the-quality-of-health-care-in-america`. [Accessed: 11-March-2024].

[18] M. MacPherson, "Immigrant, refugee, and indigenous canadians' experiences with virtual health care services: Rapid review," *JMIR Hum. Factors*, vol. 10, p. e47288, Oct. 2023.

[19] "Health Equity | Public Health Ontario." `https://www.publichealthontario.ca/en/Health-Topics/Health-Equity`. [Accessed: 11-March-2024].

[20] C. Fernandez-Lozano, P. Hervella, V. Mato-Abad, M. Rodríguez-Yáñez, S. Suárez-Garaboa, I. López-Dequidt, A. Estany-Gestal, T. Sobrino, F. Campos, J. Castillo, S. Rodríguez-Yáñez, and R. Iglesias-Rey, "Random forest-based prediction of stroke outcome," *Sci. Rep.*, vol. 11, p. 10071, May 2021.

[21] Y. Xie, B. Jiang, E. Gong, Y. Li, G. Zhu, P. Michel, M. Wintermark, and G. Zaharchuk, "Use of gradient boosting machine learning to predict patient outcome in acute ischemic stroke on the basis of imaging, demographic, and clinical information," *AJR Am. J. Roentgenol.*, vol. 212, pp. 44–51, Jan. 2019.

[22] S. Dev, H. Wang, C. S. Nwosu, N. Jain, B. Veeravalli, and D. John, "A predictive analytics approach for stroke prediction using machine learning and neural networks," *Healthcare Analytics*, vol. 2, p. 100032, Nov. 2022.

[23] P.-Y. Su, Y.-C. Wei, H. Luo, C.-H. Liu, W.-Y. Huang, K.-F. Chen, C.-P. Lin, H.-Y. Wei, and T.-H. Lee, "Machine learning models for predicting influential factors of early

outcomes in acute ischemic stroke: Registry-based study," *JMIR Med. Inform.*, vol. 10, p. e32508, Mar. 2022.

[24] P. B. Dash, J. Nayak, C. R. Kishore, M. Mishra, and B. Naik, "Efficient ensemble learning based CatBoost approach for early-stage stroke risk prediction," in *Smart Innovation, Systems and Technologies*, Smart innovation, systems and technologies, pp. 475–483, Singapore: Springer Nature Singapore, 2023.

[25] R. Chen, S. Zhang, J. Li, D. Guo, W. Zhang, X. Wang, D. Tian, Z. Qu, and X. Wang, "A study on predicting the length of hospital stay for chinese patients with ischemic stroke based on the XGBoost algorithm," *BMC Med. Inform. Decis. Mak.*, vol. 23, p. 49, Mar. 2023.

[26] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," 2017.

[27] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," 2016.

[28] eClinicalMedicine, "The rising global burden of stroke," *eClinicalMedicine*, vol. 59, p. 102028, May 2023.

[29] S. Hatano, "Experience from a multicentre stroke register: a preliminary report," *Bulletin of the World Health Organization*, vol. 54, no. 5, p. 541, 1976.

[30] V. L. Feigin and S. et al, "Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the global burden of disease study 2019," *The Lancet Neurology*, vol. 20, p. 795–820, Oct. 2021.

[31] V. L. Feigin, M. Brainin, B. Norrving, S. Martins, R. L. Sacco, W. Hacke, M. Fisher, J. Pandian, and P. Lindsay, "World stroke organization (WSO): Global stroke fact sheet 2022," *Int J Stroke*, vol. 17, pp. 18–29, Jan. 2022.

[32] P. H. A. of Canada, "Stroke in Canada - Canada.ca." `https://www.canada.ca/en/public-health/services/publications/diseases-conditions/stroke-in-canada.html`. [Accessed: 20-February-2024].

[33] Public Health Agency of Canada, "Canadian chronic disease surveillance system data tool," 2023.

[34] Government of Canada, 2023.

[35] World Health Organization, "Social determinants of health." `https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1`, 2023. Accessed: 25-September-2023.

[36] D. C. Dover and A. P. Belon, "The health equity measurement framework: a comprehensive model to measure social inequities in health," *International Journal for Equity in Health*, vol. 18, Feb. 2019.

[37] "Putting health inequities on the map." https://www.hsph.harvard.edu/wp-content/uploads/sites/2068/2016/10/09_krieger_putting_health_inequities_on_the_map_geojournal.pdf. [Accessed: 23-February-2024].

[38] R. G. Wilkinson and M. Marmot, *Social determinants of health: the solid facts*. World Health Organization, 2003.

[39] S. Yusuf, S. Reddy, S. Ôunpuu, and S. Anand, "Global burden of cardiovascular diseases: part i: general considerations, the epidemiologic transition, risk factors, and impact of urbanization," *Circulation*, vol. 104, no. 22, pp. 2746–2753, 2001.

[40] A. D. Lopez, C. D. Mathers, M. Ezzati, D. T. Jamison, and C. J. Murray, "Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data," *The lancet*, vol. 367, no. 9524, pp. 1747–1757, 2006.

[41] C. D. Mathers and D. Loncar, "Projections of global mortality and burden of disease from 2002 to 2030," *PLoS Medicine*, vol. 3, no. 11, p. e442, 2006.

[42] C. Kreatsoulas and S. S. Anand, "The impact of social determinants on cardiovascular disease," *Can J Cardiol*, vol. 26 Suppl C, pp. 8C–13C, Aug. 2010.

[43] "Sdoh evidence review." https://aspe.hhs.gov/reports/sdoh-evidence-review. [Accessed: 20-February-2024].

[44] S. Magnan, "Social determinants of health 101 for health care: Five plus five," *NAM Perspectives*, vol. 7, Oct. 2017.

[45] Centers for Disease Control and Prevention, "CDC quick maps," 2023.

[46] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, *et al.*, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," *Information fusion*, vol. 58, pp. 82–115, 2020.

[47] M. W. Segar, J. L. Hall, P. S. Jhund, T. M. Powell-Wiley, A. A. Morris, D. Kao, G. C. Fonarow, R. Hernandez, N. E. Ibrahim, C. Rutan, A. M. Navar, L. M. Stevens, and A. Pandey, "Machine learning–based models incorporating social determinants of health vs traditional models for predicting in-hospital mortality in patients with heart failure," *JAMA Cardiology*, vol. 7, p. 844, Aug. 2022.

[48] Y. Zhao, E. P. Wood, N. Mirin, S. H. Cook, and R. Chunara, "Social determinants in machine learning cardiovascular disease prediction models: A systematic review," *American Journal of Preventive Medicine*, vol. 61, no. 4, pp. 596–605, 2021.

[49] J. Lutz, "Social determinants & stroke."

[50] W. M. Schultz, H. M. Kelli, J. C. Lisko, T. Varghese, J. Shen, P. Sandesara, A. A. Quyyumi, H. A. Taylor, M. Gulati, J. G. Harold, J. H. Mieres, K. C. Ferdinand,

G. A. Mensah, and L. S. Sperling, "Socioeconomic status and cardiovascular outcomes: Challenges and interventions," *Circulation*, vol. 137, pp. 2166–2178, May 2018.

[51] B. Kissela, A. Schneider, D. Kleindorfer, J. Khoury, R. Miller, K. Alwell, D. Woo, J. Szaflarski, J. Gebel, C. Moomaw, *et al.*, "Stroke in a biracial population: the excess burden of stroke among blacks," *Stroke*, vol. 35, no. 2, pp. 426–431, 2004.

[52] V. J. Howard, D. O. Kleindorfer, S. E. Judd, L. A. McClure, M. M. Safford, J. D. Rhodes, M. Cushman, C. S. Moy, E. Z. Soliman, B. M. Kissela, *et al.*, "Disparities in stroke incidence contributing to disparities in stroke mortality," *Annals of neurology*, vol. 69, no. 4, pp. 619–627, 2011.

[53] G. Howard, M. Cushman, B. M. Kissela, D. O. Kleindorfer, L. A. McClure, M. M. Safford, J. D. Rhodes, E. Z. Soliman, C. S. Moy, S. E. Judd, *et al.*, "Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass," *Stroke*, vol. 42, no. 12, pp. 3369–3375, 2011.

[54] J. Y. Chong and R. L. Sacco, "Epidemiology of stroke in young adults: race/ethnic differences," *Journal of thrombosis and thrombolysis*, vol. 20, pp. 77–83, 2005.

[55] G. Gabb and L. Arnolda, "Geographic location as a modifiable cardiac risk factor," *Canadian Medical Association Journal*, vol. 189, p. E482–E483, Apr. 2017.

[56] N. Douthit, S. Kiv, T. Dwolatzky, and S. Biswas, “Exposing some important barriers to health care access in the rural usa,” *Public health*, vol. 129, no. 6, pp. 611–620, 2015.

[57] A. Phillips, *Rural, regional and remote health: a study on mortality.* Australian Institute of Health and Welfare, 2007.

[58] D. Hu, J. Huang, Y. Wang, D. Zhang, and Y. Qu, “Fruits and vegetables consumption and risk of stroke: a meta-analysis of prospective cohort studies,” *Stroke*, vol. 45, no. 6, pp. 1613–1619, 2014.

[59] P. Schroff, C. M. Gamboa, R. W. Durant, A. Oikeh, J. S. Richman, and M. M. Safford, “Vulnerabilities to health disparities and statin use in the regards (reasons for geographic and racial differences in stroke) study,” *Journal of the American Heart Association*, vol. 6, no. 9, p. e005449, 2017.

[60] N. A. Bhavsar, A. Gao, M. Phelan, N. J. Pagidipati, and B. A. Goldstein, “Value of neighborhood socioeconomic status in predicting risk of outcomes in studies that use electronic health record data,” *JAMA Netw Open*, vol. 1, p. e182716, Sept. 2018.

[61] C. E. Cutrona, G. Wallace, and K. A. Wesner, “Neighborhood characteristics and depression: An examination of stress processes,” *Current Directions in Psychological Science*, vol. 15, p. 188–192, Aug. 2006.

[62] N. Stappers, D. Van Kann, N. De Vries, and S. Kremers, "Do physical activity friendly neighborhoods affect community members equally? a cross-sectional study," *International Journal of Environmental Research and Public Health*, vol. 15, p. 1062, May 2018.

[63] E. H. Belarmino, J. Malacarne, A. C. McCarthy, S. Bliss, J. Laurent, S. C. Merrill, M. T. Niles, S. Nowak, R. E. Schattman, and K. Yerxa, "Suboptimal diets identified among adults in two rural states during the covid-19 pandemic," *Journal of Hunger and Environmental Nutrition*, p. 1–16, Feb 2024.

[64] G. K. Ghodeshwar, A. Dube, and D. Khobragade, "Impact of lifestyle modifications on cardiovascular health: A narrative review," *Cureus*, July 2023.

[65] A. V. Diez Roux, "Investigating neighborhood and area effects on health," vol. 91, pp. 1783–1789, Nov. 2001.

[66] E. Reshetnyak, M. Ntamatungiro, L. C. Pinheiro, V. J. Howard, A. P. Carson, K. D. Martin, and M. M. Safford, "Impact of multiple social determinants of health on incident stroke," *Stroke*, vol. 51, no. 8, pp. 2445–2453, 2020.

[67] M. McGeachie, R. L. B. Ramoni, J. C. Mychaleckyj, K. L. Furie, J. M. Dreyfuss, Y. Liu, D. Herrington, X. Guo, J. A. Lima, W. Post, *et al.*, "Integrative predictive model of coronary artery calcification in atherosclerosis," *Circulation*, vol. 120, no. 24, pp. 2448–2454, 2009.

[68] L. Rasmy, Y. Wu, N. Wang, X. Geng, W. J. Zheng, F. Wang, H. Wu, H. Xu, and D. Zhi, "A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous ehr data set," *Journal of biomedical informatics*, vol. 84, pp. 11–16, 2018.

[69] S. Kino, Y.-T. Hsu, K. Shiba, Y.-S. Chien, C. Mita, I. Kawachi, and A. Daoud, "A scoping review on the use of machine learning in research on social determinants of health: Trends and research prospects," *SSM - Population Health*, vol. 15, p. 100836, 2021.

[70] C. M. Clancy, S. A. Glied, and N. Lurie, "From research to health policy impact," *Health Serv. Res.*, vol. 47, pp. 337–343, Feb. 2012.

[71] M. Conway, S. Keyhani, L. Christensen, B. R. South, M. Vali, L. C. Walter, D. L. Mowery, S. Abdelrahman, and W. W. Chapman, "Moonstone: a novel natural language processing system for inferring social risk from clinical narratives," *Journal of biomedical semantics*, vol. 10, no. 1, pp. 1–10, 2019.

[72] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, vol. 366, no. 6464, pp. 447–453, 2019.

[73] A. Daoud, A. Herlitz, and S. Subramanian, "Combining distributive ethics and causal inference to make trade-offs between austerity and population health," *arXiv preprint arXiv:2007.15550*, 2020.

[74] H. Boerstler and J. M. de Figueiredo, "Prediction of use of psychiatric services: Application of the cart algorithm," *The journal of mental health administration*, vol. 18, pp. 27–34, 1991.

[75] E. Darvishi, H. Khotanlou, J. Khoubi, O. Giahi, and N. Mahdavi, "Prediction effects of personal, psychosocial, and occupational risk factors on low back pain severity using artificial neural networks approach in industrial workers," *Journal of manipulative and physiological therapeutics*, vol. 40, no. 7, pp. 486–493, 2017.

[76] S.-Y. K. Fu, D. Anderson, M. Courtney, and W. Hu, "The relationship between culture, attitude, social networks and quality of life in midlife australian and taiwanese citizens," *Maturitas*, vol. 58, no. 3, pp. 285–295, 2007.

[77] M. A. Ahmad, C. Eckert, and A. Teredesai, "Interpretable machine learning in healthcare," in *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*, pp. 559–560, 2018.

[78] *Bayesian methods for neural networks.*

[79] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009.

[80] T. L. Wiemken and R. R. Kelley, “Machine learning in epidemiology and health outcomes research,” *Annu Rev Public Health*, vol. 41, no. 1, pp. 21–36, 2020.

[81] L. F. Berkman, I. Kawachi, and M. M. Glymour, *Social epidemiology*. Oxford University Press, 2014.

[82] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, “Explaining explanations: An overview of interpretability of machine learning,” in *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pp. 80–89, IEEE, 2018.

[83] Z. C. Lipton, “The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery.,” *Queue*, vol. 16, no. 3, pp. 31–57, 2018.

[84] U. Bhatt, A. Xiang, S. Sharma, A. Weller, A. Taly, Y. Jia, J. Ghosh, R. Puri, J. M. Moura, and P. Eckersley, “Explainable machine learning in deployment,” in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 648–657, 2020.

[85] J.-M. Fellous, G. Sapiro, A. Rossi, H. Mayberg, and M. Ferrante, "Explainable artificial intelligence for neuroscience: behavioral neurostimulation," *Frontiers in neuroscience*, vol. 13, p. 1346, 2019.

[86] J. Kleinberg, J. Ludwig, S. Mullainathan, and A. Rambachan, "Algorithmic fairness," in *Aea papers and proceedings*, vol. 108, pp. 22–27, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2018.

[87] J. R. Loftus, C. Russell, M. J. Kusner, and R. Silva, "Causal reasoning for algorithmic fairness," *arXiv preprint arXiv:1805.05859*, 2018.

[88] R. Porto, J. M. Molina, A. Berlanga, and M. A. Patricio, "Minimum relevant features to obtain explainable systems for predicting cardiovascular disease using the statlog data set," *Applied Sciences*, vol. 11, no. 3, p. 1285, 2021.

[89] R. Aggarwal, P. Podder, and A. Khamparia, "Ecg classification and analysis for heart disease prediction using xai-driven machine learning algorithms," in *Biomedical Data Analysis and Processing Using Explainable (XAI) and Responsive Artificial Intelligence (RAI)*, pp. 91–103, Springer, 2022.

[90] A. M. Westerlund, J. S. Hawe, M. Heinig, and H. Schunkert, "Risk prediction of cardiovascular events by exploration of molecular data with explainable artificial intelligence," *International Journal of Molecular Sciences*, vol. 22, no. 19, p. 10291, 2021.

[91] I. de la statistique du Québec, “Statistics at the service of society.” `https://statistique.quebec.ca/en`. [Accessed: 16-February-2024].

[92] I. de la statistique du Québec, “Services for researchers.” `https://statistique.quebec.ca/en/institut/services-for-researchers`. [Accessed: 16-February-2024].

[93] “Introduction to Sociology; 2nd Canadian Edition.” `https://opentextbc.ca/introductiontosociology2ndedition/chapter/chapter-12-gender-sex-and-sexuality/`. [Accessed: 20-February-2024].

[94] “sulc.edu.” `https://www.sulc.edu/assets/sulc/officeequityinclusiontitleix/RACE.pdf`. [Accessed: 20-February-2024].

[95] “Introduction to Sociology ; 1st Canadian Edition.” `https://shorturl.at/lEP25`, 2013. [Accessed: 20-February-2024].

[96] P. John Loewen, “Optimizing Python Pandas for Large Datasets: 4 Practical Examples of Chunking.” `https://python.plainenglish.io/optimizing-python-pandas-for-large-datasets-4-practical-examples-of-\chunking-e94f1ec584a3`. [Accessed: 20-February-2024].

[97] R. Mohammed, J. Rawashdeh, and M. Abdullah, “Machine learning with oversampling and undersampling techniques: Overview study and experimental results,” in *2020*

*11th International Conference on Information and Communication Systems (ICICS)*, pp. 243–248, 2020.

[98] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, p. 321–357, June 2002.

[99] "NearMiss." `https://shorturl.at/uwEIW`. [Accessed: 11-March-2024].

[100] "GridSearchCV." https://shorturl.at/kDKL2.

[101] G. Hammond, K. Johnston, K. Huang, and K. E. Joynt Maddox, "Social determinants of health improve predictive accuracy of clinical risk models for cardiovascular hospitalization, annual cost, and death," *Circ. Cardiovasc. Qual. Outcomes*, vol. 13, p. e006752, June 2020.

[102] N. Stabellini, J. Cullen, J. X. Moore, S. Dent, A. L. Sutton, J. Shanahan, A. J. Montero, and A. Guha, "Social determinants of health data improve the prediction of cardiac outcomes in females with breast cancer," *Cancers (Basel)*, vol. 15, Sept. 2023.

[103] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, pp. 1937–1967, 2021.

[104] "Deep neural networks with data dependent implicit activation function." `https://ww3.math.ucla.edu/camreport/cam18-07.pdf`. [Accessed: 11-March-2024].

[105] G. Seni and J. Elder, *Ensemble methods in data mining: improving accuracy through combining predictions*. Morgan & Claypool Publishers, 2010.

[106] S. Ali, F. Akhlaq, A. S. Imran, Z. Kastrati, S. M. Daudpota, and M. Moosa, "The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review," *Comput. Biol. Med.*, vol. 166, p. 107555, Oct. 2023.

[107] L. Takamine, J. Forman, L. J. Damschroder, B. Youles, and J. Sussman, "Understanding providers' attitudes and key concerns toward incorporating CVD risk prediction into clinical practice: a qualitative study," *BMC Health Serv. Res.*, vol. 21, p. 561, June 2021.

[108] "Social determinants of health." `https://www.healthcarefinancenews.com/news/social-determinants-health-data-can-be-difficult-collect-and-share-its\-imperative-success`. [Accessed: 11-March-2024].

[109] M. Martínez-García and E. Hernández-Lemus, "Data integration challenges for machine learning in precision medicine," *Front. Med. (Lausanne)*, vol. 8, p. 784455, 2021.

[110] G. A. Aarons, A. R. Monn, L. K. Leslie, A. F. Garland, L. Lugo, R. L. Hough, and S. A. Brown, "Association between mental and physical health problems in high-risk adolescents: a longitudinal study," *J. Adolesc. Health*, vol. 43, pp. 260–267, Sept. 2008.

[111] E. B. Weida, P. Phojanakong, F. Patel, and M. Chilton, “Financial health as a measurable social determinant of health,” *PLoS One*, vol. 15, p. e0233359, May 2020.

[112] V. Raghupathi and W. Raghupathi, “The influence of education on health: an empirical assessment of OECD countries for the period 1995-2015,” *Arch. Public Health*, vol. 78, p. 20, Apr. 2020.

[113] M. Kelly-Hayes, “Influence of age and health behaviors on stroke risk: lessons from longitudinal studies,” *Journal of the American Geriatrics Society*, vol. 58, pp. S325–S328, 2010.

[114] L. S. Williams, S. S. Ghose, and R. W. Swindle, “Depression and other mental health diagnoses increase mortality risk after ischemic stroke,” *American Journal of Psychiatry*, vol. 161, no. 6, pp. 1090–1095, 2004.

[115] H. Ren, C. Liu, J. Li, R. Yang, F. Ma, M. Zhang, R. Wang, and L. Gan, “Self-perceived burden in the young and middle-aged inpatients with stroke: a cross-sectional survey,” *Rehabilitation Nursing Journal*, vol. 41, no. 2, pp. 101–111, 2016.

[116] H. Nakayama, H. Jørgensen, H. Raaschou, and T. S. Olsen, “The influence of age on stroke outcome. the copenhagen stroke study.,” *Stroke*, vol. 25, no. 4, pp. 808–813, 1994.

[117] L. Kalichman, B. Rodrigues, D. Gurvich, Z. Israelov, and E. Spivak, “Impact of patient’s weight on stroke rehabilitation results,” *American journal of physical medicine & rehabilitation*, vol. 86, no. 8, pp. 650–655, 2007.

[118] I.-W. G. Kwon, S.-H. Kim, and D. Martin, “Integrating social determinants of health to precision medicine through digital transformation: an exploratory roadmap,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 9, p. 5018, 2021.

# Appendix A

# List of SDoH & Traditional Features

Table A.1 provides a breakdown of the variables used in our analysis, categorized by both SDoH and traditional clinical risk factors. The table details the specific variables included under each SDoH category, offering a clear picture of the factors considered in our models.

**Table A.1:** SDoH and Traditional Categories

| **Subset** | **Feature Name** |
|---|---|
| Traditional | Drank alcohol in last 12 months |
| | Frequency of drinking alcohol in last 12 months |
| | Drank more than 5 drinks on one occasion in last 12 months |
| | Has asthma |
| | High blood pressure |
| | Takes medication for high blood pressure |
| | Has diabetes |
| | Has heart disease |
| | Stroke incidence |
| | Takes medications for high cholestrol |
| | Smoked more than 100 cigarettes in lifetime |

Continued on next page

Table A.1 – continued from previous page

| Subset | Feature Name |
|---|---|
| | Type of smoker - Daily/ Occasionally/ Not at all<br>Weight<br>Height<br>Exposed to second hand smoke (private vehicle) in last 1 month<br>Exposed to second hand smoke (public places) in last 1 month<br>Self perceived weight - underweight/ overweight/ just about right |
| Demographic | Marital Status<br>Sex<br>Age |
| Economic | Worked at a job/ business<br>Worked at a job/ business past week<br>Has a regular healthcare provider |
| Neighbourhood | Rural or Urban Area<br>Dwelling - owned or rent |
| Personal | Self-perceived ability to handle unexpected problems<br>Self-perceived ability to handle day to day demands<br>Self-perceived health<br>Self-perceived mental health<br>Self-perceived life stress |
| Social | Food Security<br>Could not afford to eat<br>Sense of belonging to local community<br>Highest level of education |
| Mental Health | Mood Disorder (Depression, Bipolar, Mania, Dysthymia)<br>Anxiety disorder<br>consulted mental health professional in last 12 months |
| Race | Aboriginal Identity - First Nations/ Metis/ Inuk (Inuit)<br>White<br>Chinese<br>South Asian<br>Black<br>Filipino<br>Latin America<br>Southeast Asian<br>Arab |

Continued on next page

Table A.1 – continued from previous page

| Subset | Feature Name |
|---|---|
| | West Asian |
| | Japanese |
| | Korean |
| Ethnicity | Canadian |
| | French |
| | English |
| | German |
| | Scottish |
| | Irish |
| | Italian |
| | Ukranian |
| | Dutch (Netherlands) |
| | Chinese |
| | Jewish |
| | Polish |
| | Portugese |
| | South Asian |
| | Norwegian |
| | Welsh |
| | Swedish |
| | Other |
| | Metis |
| | Inuit |

# Appendix B

# Evaluation Metrics for SDoH Subsets

Table B.1 dives deep into the performance of our stroke prediction models across various SDoH subsets. To evaluate how well the models function and how SDoH factors influence these predictions, we calculated key metrics for each model within each SDoH subset. These metrics include accuracy, precision, recall, and F1 score. By analyzing these measures, we gain valuable insights into the models' ability to accurately predict stroke occurrences when considering different SDoH factors.

**Table B.1:** Model Evaluation Metrics for Specific SDoH (in percentage)

| **Model** | **Feature Subset** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
|---|---|---|---|---|---|
| Random Forest | Traditional | 65.90 | 66.00 | 64.70 | 65.30 |
| | All SDoH | 77.10 | 77.80 | 75.50 | 76.60 |
| | Economic | 69.80 | 69.60 | 69.60 | 69.60 |
| | Personal | 72.20 | 72.30 | 71.60 | 71.90 |

Continued on next page

Table B.1 – continued from previous page

| Model | Feature Subset | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|---|
| | Neighbourhood | 65.90 | 65.70 | 65.70 | 65.70 |
| | Demographic | 72.20 | 72.30 | 71.60 | 71.90 |
| | Race | 65.40 | 65.70 | 63.70 | 64.70 |
| | Ethnicity | 65.90 | 67.40 | 60.80 | 63.90 |
| | Social | 68.30 | 70.30 | 62.70 | 66.30 |
| | Mental Health | 62.00 | 61.50 | 62.70 | 62.10 |
| GBM | Traditional | 74.10 | 74.70 | 72.50 | 73.60 |
| | All SDoH | 73.20 | 72.80 | 73.50 | 73.20 |
| | Economic | 77.10 | 77.20 | 76.50 | 76.80 |
| | Personal | 77.10 | 78.40 | 74.50 | 76.40 |
| | Neighbourhood | 71.20 | 72.20 | 68.60 | 70.40 |
| | Demographic | 77.10 | 77.80 | 75.50 | 76.60 |
| | Race | 74.10 | 75.80 | 70.60 | 73.10 |
| | Ethnicity | 73.70 | 75.00 | 70.60 | 72.70 |
| | Social | 71.20 | 71.70 | 69.60 | 70.60 |
| | Mental Health | 70.70 | 70.60 | 70.60 | 70.60 |
| KNN | Traditional | 65.40 | 66.00 | 62.70 | 64.30 |
| | All SDoH | 69.80 | 71.30 | 65.70 | 68.40 |
| | Economic | 70.70 | 70.60 | 70.60 | 70.60 |
| | Personal | 68.80 | 68.60 | 68.60 | 68.60 |
| | Neighbourhood | 59.50 | 60.00 | 55.90 | 57.90 |
| | Demographic | 69.30 | 70.10 | 66.70 | 68.30 |
| | Race | 65.40 | 67.00 | 59.80 | 63.20 |
| | Ethnicity | 64.40 | 66.70 | 56.90 | 61.40 |
| | Social | 65.40 | 65.70 | 63.70 | 64.70 |
| | Mental Health | 69.30 | 68.60 | 70.60 | 69.60 |
| Neural Network | Traditional | 68.80 | 69.80 | 65.70 | 67.70 |
| | All SDoH | 62.90 | 62.00 | 65.70 | 63.80 |
| | Economic | 66.80 | 66.30 | 67.60 | 67.00 |

Continued on next page

Table B.1 – continued from previous page

| Model | Feature Subset | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|---|
| | Personal | 61.50 | 60.70 | 63.70 | 62.20 |
| | Neighbourhood | 63.90 | 67.50 | 52.90 | 59.30 |
| | Demographic | 65.90 | 65.70 | 65.70 | 65.70 |
| | Race | 66.80 | 68.10 | 62.70 | 65.30 |
| | Ethnicity | 63.40 | 64.20 | 59.80 | 61.90 |
| | Social | 63.40 | 65.20 | 56.90 | 60.70 |
| | Mental Health | 61.50 | 61.40 | 60.80 | 61.10 |
| LGBM | Traditional | 74.60 | 75.50 | 72.50 | 74.00 |
| | All SDoH | 77.60 | 78.60 | 75.50 | 77.00 |
| | Economic | 77.60 | 76.40 | 79.40 | 77.90 |
| | Personal | 75.60 | 77.70 | 71.60 | 74.50 |
| | Neighbourhood | 73.70 | 75.50 | 69.60 | 72.40 |
| | Demographic | 75.60 | 75.50 | 75.50 | 75.50 |
| | Race | 74.10 | 74.30 | 73.50 | 73.90 |
| | Ethnicity | 73.70 | 74.00 | 72.50 | 73.30 |
| | Social | 74.10 | 75.30 | 71.60 | 73.40 |
| | Mental Health | 72.70 | 73.50 | 70.60 | 72.00 |
| CatBoost | Traditional | 73.70 | 73.50 | 73.50 | 73.50 |
| | All SDoH | 75.10 | 76.80 | 71.60 | 74.10 |
| | Economic | 73.20 | 73.70 | 71.60 | 72.60 |
| | Personal | 74.10 | 74.70 | 72.50 | 73.60 |
| | Neighbourhood | 72.70 | 72.50 | 72.50 | 72.50 |
| | Demographic | 74.10 | 74.70 | 72.50 | 73.60 |
| | Race | 75.10 | 75.80 | 73.50 | 74.60 |
| | Ethnicity | 72.70 | 71.70 | 74.50 | 73.10 |
| | Social | 74.10 | 75.30 | 71.60 | 73.40 |
| | Mental Health | 72.20 | 71.40 | 73.50 | 72.50 |
| XGBoost | Traditional | 71.70 | 72.90 | 68.60 | 70.70 |
| | All SDoH | 74.10 | 75.30 | 71.60 | 73.40 |

Continued on next page

Table B.1 – continued from previous page

| Model | Feature Subset | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|---|
| | Economic | 75.60 | 76.00 | 74.50 | 75.20 |
| | Personal | 75.60 | 77.10 | 72.50 | 74.70 |
| | Neighbourhood | 72.20 | 73.20 | 69.60 | 71.40 |
| | Demographic | 76.10 | 75.70 | 76.50 | 76.00 |
| | Race | 73.70 | 74.50 | 71.60 | 73.00 |
| | Ethnicity | 73.70 | 75.00 | 70.60 | 72.70 |
| | Social | 74.10 | 75.30 | 71.60 | 73.40 |
| | Mental Health | 69.30 | 69.30 | 68.60 | 69.00 |