明治大学総合数理学部

2025 年度

卒　業　研　究

多要素認証に対してセッションIDを取得する中間者攻撃ツールEvilginxの評価

学位請求者　先端メディアサイエンス学科

関口　智大

目 次

第 1 章　はじめに **2**

第 2 章　基本定義 **3**
2.1　パスキー 3
2.2　セッション ID 3
2.3　ソーシャルログイン 3
2.4　先行研究 [3] 4

第 3 章　実験 **5**
3.1　実験方法 5
3.2　実験結果 6
3.2.1　Evilginx を用いた AiTM 攻撃の結果 6
3.2.2　セッション ID を窃取 7
3.2.3　2024 年度と 2025 年度の導入率の比較 7
3.2.4　ソーシャルログイン導入率 8
3.2.5　考察 9
3.3　本研究の限界 9
3.3.1　本研究の倫理配慮 10

第 4 章　おわりに **11**

謝辞 **17**

付 録 A　100 万人のヘルスケアデータからの心疾患の罹患率 **19**
A.1　はじめに 19
A.2　実験 20
A.2.1　ヘルスケアデータの概要 20
A.2.2　実験方法 21
A.2.3　結果 21
A.2.4　考察 25
A.3　おわりに 26

第1章　はじめに

2022 年 9 月 15 日，Uber 社はハッキンググループ LAPSUS$による MFA 疲労攻撃を受け，社内ネットワークに侵入され，システム権限を奪取された [1]．Microsoft 社の報告 [2] によれば，多要素認証 Multi-Factor Authentication(MFA) の導入が広がる一方で，Adversary-in-the-Middle（AiTM）と呼ばれる中間者攻撃型フィッシングは 2024 年に 146%増加したとされている．太田ら [3] は国内主要 102 サイト中 87 サイト（85%）が Evilginx による AiTM 攻撃に対して脆弱である可能性を指摘している．これらの事例は，MFA を導入していても AiTM 攻撃により迂回され得ることを示している．MFA 環境に残存する脆弱性の対策が求められている．

しかし，AiTM 攻撃が MFA を迂回する具体的過程や，認証方式ごとの安全性を体系的に検証した研究は十分とはいえない．特に，中間者攻撃ツール Evilginx に関する実証的分析は不足しており，認証方式が攻撃に対してどの様に脆弱性があるか不明であった．

そこで, 本研究では，AiTM 攻撃が単一の認証をバイパスするだけでなく，セッション ID を窃取することでログイン状態（セッション）を継承し得る脅威に着目する．国内外の主要 Web サービスにおける 2024 年度と 2025 年度の MFA 導入状況の変化を調査する．主要な MFA 方式（SMS, TOTP 等）を対象として，AiTM 攻撃に対する単一ログイン耐性およびセッション ID 窃取率を明らかにする．

本研究は，以下の新規性を有する．

- Evilginx がセッション ID を窃取して, ログインフェーズを経ずに侵入できるかを明らかにする．
- Evilginx を用いた AiTM 型中間者攻撃の，各認証方式の耐性を実験的に検証する．
- 2024 年と 2025 年の MFA 普及率を比較する．

第2章　基本定義

2.1　パスキー

パスキーは，FIDO2／WebAuthn[4] を基盤としており，利用者がアクセスしているドメインと，パスキーが発行されたドメインが完全に一致しているかを厳密に検証する仕組みである．この特性により，フィッシングサイトでは正しいドメイン一致が成立せず，秘密鍵による署名処理が実行されないため，パスキーはフィッシング攻撃に対して高い耐性を有するとされている．

2.2　セッションID

セッションID[5] は，Web サービスにおいてユーザのログイン状態を保存し，複数の対話メッセージを結びつけたセッションを管理するために一時的に発行される識別子である．代表的な Cookie にある方式では, ログインする時にサーバが発行したセッション ID を，ユーザのブラウザに Cookie として保存することで，以後のリクエストが同一ユーザによるものであることを確認する．このセッション ID が第三者に窃取された場合，攻撃者は認証操作を行うことなく正規ユーザになりすましてログイン状態を奪取できる．従って，セッション ID の窃取は重大なセキュリティリスクとなり，AiTM 攻撃における主要な攻撃手法の一つとなっている．

2.3　ソーシャルログイン

ソーシャルログイン [6] は，Facebook や Google などの外部 ID プロパガンダで発行されたアカウント情報を利用して，他の Web サイトへログインする仕組みである．これにより，各サイトごとにパスワードやユーザ ID を個別に登録する必要がなく，利用者に利便性上のメリットがある．図 2.1 は，Google や Apple などの単一のアカウントを用いて，複数の Web サイトにログインできる仕組みを示している．

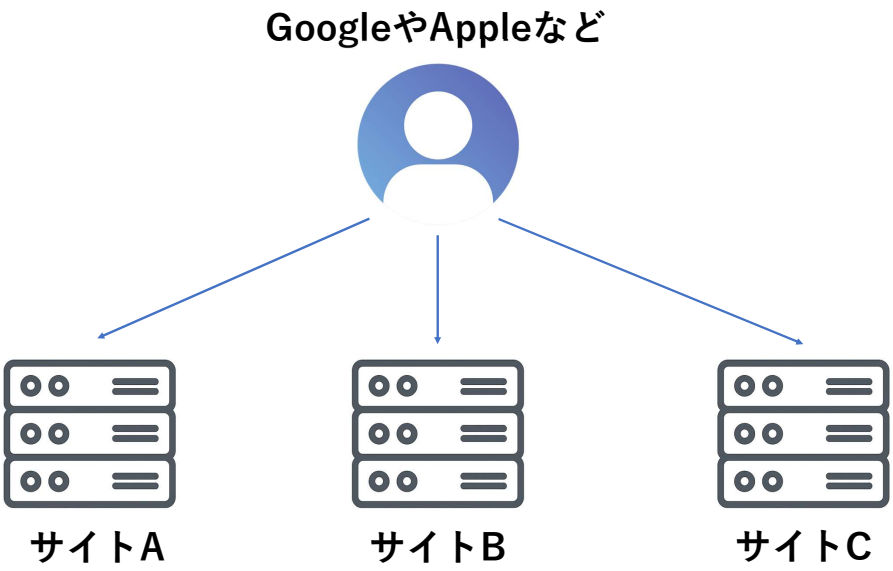

図 2.1: ソーシャルログイン

2.4 先行研究 [3]

2024 年，太田ら [3] は国内外 102 サイトの二要素認証の導入状況を調査した．調査の結果，全体の 68 サイトが二要素認証を導入し，ソーシャルログインを含めると 85 サイトで導入が確認された．また，国内と国外を比較すると，国外サービスのほうが二要素認証の導入率が高い傾向にあることが示された．さらに SMS, 電話,Email を用いた OTP 認証方式，認証アプリ（TOTP），専用アプリ，プッシュ通知など多くの認証方式が AiTM 攻撃に対して脆弱である可能性を指摘した．一方で，セキュリティキーやパスキーといった公開鍵暗号方式を用いる認証方式は，AiTM 攻撃に対して安全であると主張している．

第3章　実験

3.1　実験方法

本研究では，研究目的を達成するために以下の三つの方法を用いる．

(1) Evilginx を用いて AiTM 型中間者攻撃を疑似的に実施し，各認証方式の耐性を検証する．

(2) 主要 Web サービスを対象とし，2024 年度と 2025 年度の多要素認証方式（SMS・メール OTP・認証アプリ・パスキー等）の導入状況を比較する．

(3) 国内外のソーシャルログイン導入状況を調査し，利用可能なアカウント種別（ID プロバイダ）を整理する．

図 3.1 に，中間者攻撃者が認証情報およびセッション ID を窃取する一連のフローを図示する。図 3.1 では，攻撃者がユーザ A に対して偽の URL を提示し，ユーザ A と正規サーバ（meiji.com）との通信の間に介在する。この過程において，(2) および (5)・(6) の通信区間で認証情報や個人情報が盗聴される。さらに，ログイン後に発行されるセッション ID を窃取し，それを第三者であるユーザ B へ転送することで，ユーザ B は追加の認証操作を行うことなく，ユーザ A になりすまして正規サーバ（meiji.com）へログイン可能となる。

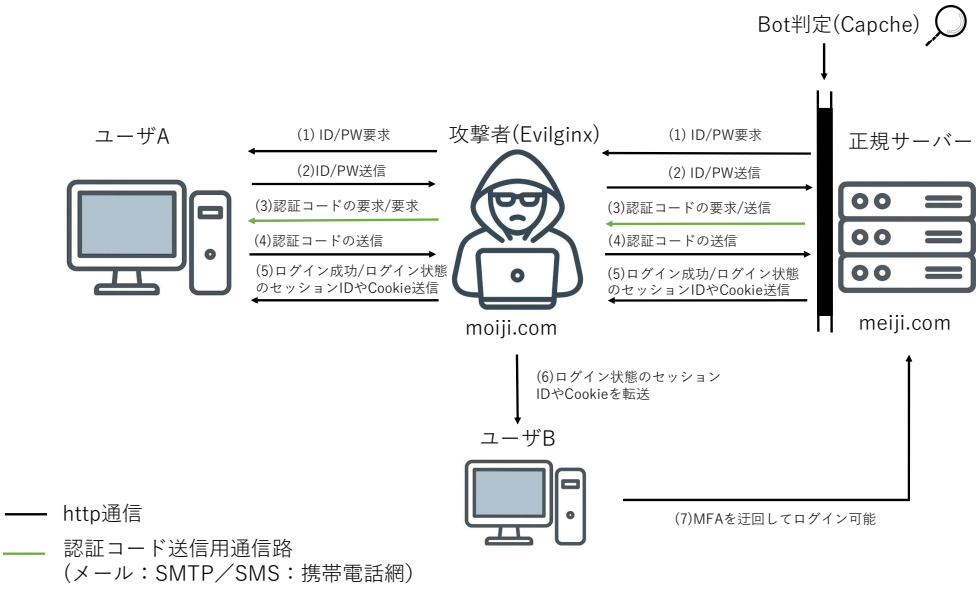

図 3.1: AiTM 型中間者攻撃の概要図

3.2 実験結果

3.2.1 Evilginx を用いた AiTM 攻撃の結果

表 3.1 に,2025 年に Web サービス 17 サイトに対して AiTM 型中間者攻撃挙動を再現した, 検証結果を示す．なお，多要素認証を導入していないサイトを「No 2FA」と表記する．

表 3.1: 認証方式別の AiTM 攻撃結果（17 サイト）

Login 確立	SessionID	個人情報	No 2FA	SMS	Email	TOTP	合計
×	✓	✓	1	2	3	4	10
×	×	✓	0	0	1	2	3
×	×	×	3	0	0	0	3

表 3.1 から，17 サイト中 13 サイト（多要素認証なし，TOTP，SMS ／メール OTP）において，何らかの情報が AiTM 攻撃によって窃取可能であったことが確認された．一方で，Bot 判定機構 (reCaptch など) を導入しているサービスでは攻撃が成立せず，認証方式に加えて，認証以外の防御機構も AiTM 耐性に大きく影響することが示唆された．

3.2.2 セッション ID を窃取

Evilginx を用いて，代表的な Web サービスのセッション ID を窃取し，ログイン操作を行うことなくセッションを乗っ取ることが可能かを調査した.

```
[10:44:10] [dbg] POST: /user/top/
[10:44:10] [dbg] POST body =
[10:44:10] [dbg] .syosetu.com: ks2 = z90fe5ta39f3
[10:44:10] [dbg] .syosetu.com: autologin =
2700594%3C%3Ecae41c11615f29e87e3a4ad7246fa75e40df622d
957bd30fcaf12be9b79c7e3a%3C%3E1709057219
```

図 3.2: セッション ID 窃取

図 3.2 に, 小説投稿サイト「小説家になろう（syosetu.com）」に対して Evilginx を用いた中間者型攻撃を実施し，cookie 挙動を観測した結果を示す．POST の中身は無く, その下に ks2 と autologin という 2 種類の cookie が確認された．次に，別の PC 上のブラウザにおいて Google Developer Tools を用い，取得した autologin Cookie を手動で追加したところ，2FA による追加認証を行うことなくログインが成立することを確認した.

この結果から，autologin Cookie はログイン状態を管理するセッション ID として機能しており，Evilginx によって当該セッション ID が窃取されることで，認証操作を経ずにセッションが再利用可能となることが示された.

3.2.3 2024 年度と 2025 年度の導入率の比較

本調査は 2025 年 7 月から 11 月にかけて実施し，調査対象は太田ら [3] が 2024 年度に調査した国内 53 サイトおよび国外 49 サイトの計 102 サイトである.

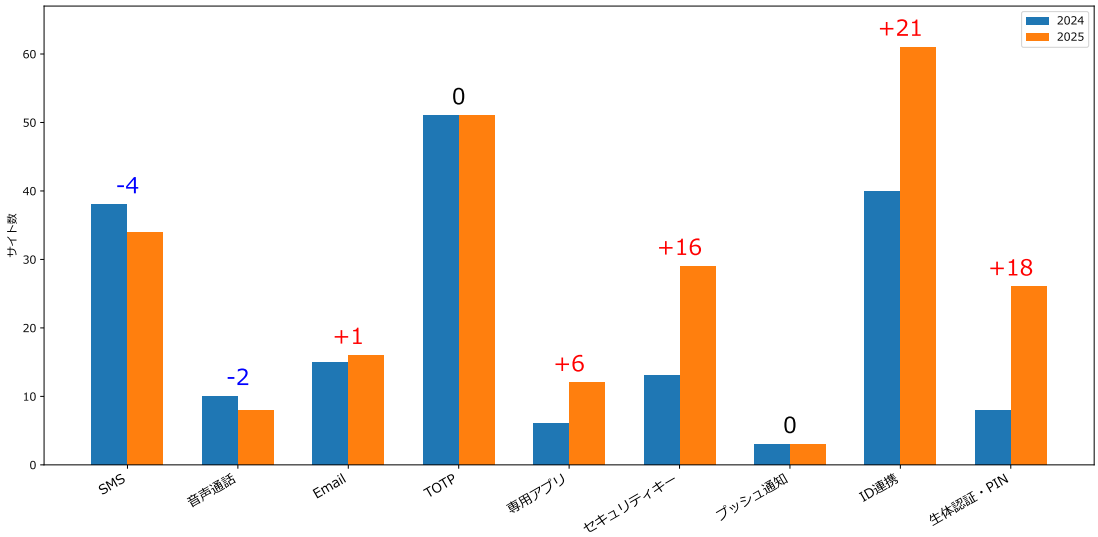

図 3.3: 2024 年度と 2025 年度における認証方式導入率の比較

図 3.3 から，2025 年は 2024 年と比較して，SMS ベース OTP の導入件数が 38 件から 34 件へと 4 件減少し，電話認証も 10 件から 8 件へと 2 件減少していることが分かる．一方で，ID 連携は 40 件から 61 件へと 21 件増加しており，セキュリティキーも 13 件から 29 件へと増加していた．

3.2.4 ソーシャルログイン導入率

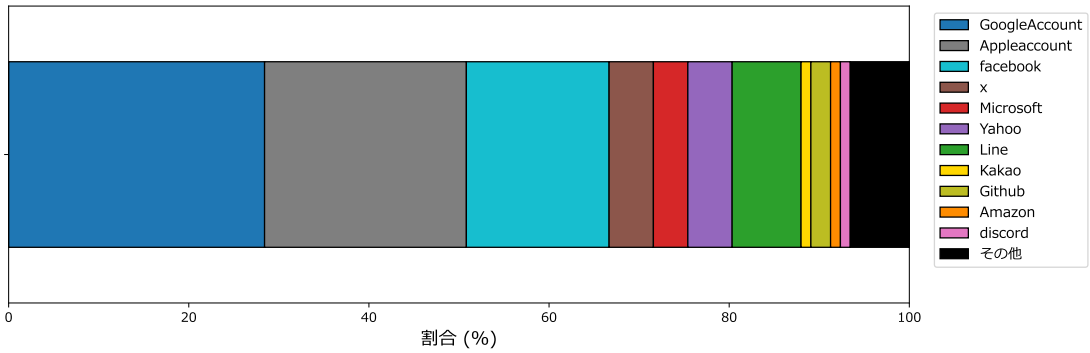

図 3.4: 全体のソーシャルログイン導入率

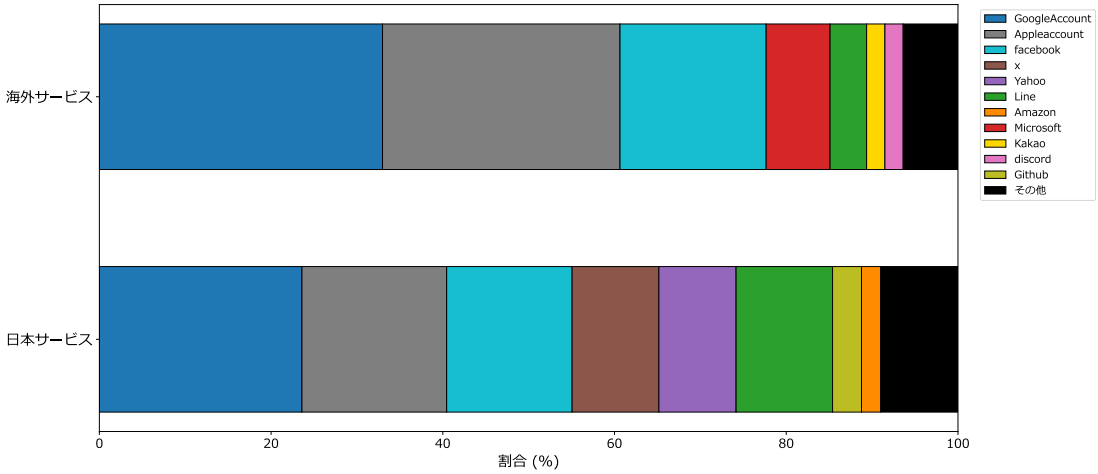

図 3.5: 日本と海外のソーシャルログイン導入率比較

図 3.5 にソーシャルログインにおける PP の割合を示す．国内サービスでは LINE や Yahoo! JAPAN が多く，海外サービスでは Google および Apple によるログインが RP の主流であることが確認された．

3.2.5 考察

AiTM 攻撃の実験結果表 3.1 から，16 サイト中 13 サイト（81.3%) において, セッション ID の窃取または個人情報の取得が可能であった特に，多要素認証を用いないパスワード単独認証や OTP 認証方式では，ワンタイムコードを盗み見ずともセッション ID の窃取のみでログインが成立するケースが確認された．一方，Bot 判定サービスを導入しているサイトでは，認証方式とは無関係に AiTM 攻撃が阻止される例が確認された．これは，認証方式以外の仕組みが AiTM 耐性向上に寄与し得ることを示唆する．

2024 年度と 2025 年度の比較より，SMS ベースの OTP 認証方式の導入率は減少し，セキュリティキーや ID 連携方式の導入が増加した．特にパスキーの導入増加は注目すべき点であり，これはフィッシング攻撃対策として企業の危機意識が強まっていることの表れといえる．

ソーシャルログインでは，国内は LINE・X が中心であり，海外は Google・Apple が主流であった．国内でも海外と同様に Google・Apple が主流であるが，これに加えて LINE・X の採用数が多い点が特徴である．利便性の向上という利点がある一方で，連携元アカウントの侵害時に複数サービスへ被害が波及するリスクも存在する．

3.3 本研究の限界

(1) 認証方式の導入率は調査時点に依存しており，仕様変更の追跡は行っていない.

(2) ソーシャルログインの評価は導入率の分析に留まり，認証プロトコル自体の脆弱性評価は行っていない.

3.3.1 本研究の倫理配慮

(1) 本研究では, アカウント本人の同意の元で行う動作検証を除いて, 実在するシステムへの検証は行っていない.

第4章　おわりに

本研究では，多要素認証環境が AiTM 攻撃に対してどの程度耐性を有するかを明らかにするため，Evilginx を用いた攻撃実験，認証方式の導入率比較，およびソーシャルログイン導入状況の調査を行った．OTP ベースの認証方式は AiTM 攻撃に脆弱である一方，Bot 判定機構など認証方式以外の追加的防御策が AiTM 耐性向上に有効であることが確認された．また，cookie を窃取して別の端末でログインが可能であることを確認した．2024 年度と 2025 年度の比較より，企業がパスキー・ID 連携などの強固な認証方式へ移行する傾向が明らかになった．さらに，ソーシャルログインの利用傾向から，国内外で異なる ID プロバイダが採用されている実態が確認された．今後は，ソーシャルログインや OAuth 2.0 ／ OIDC フローに対する AiTM 攻撃の影響評価や，サービス横断での認証強度比較など，より詳細な分析を進めたい．

カテゴリ	サイト名	ドメイン	国	多要素認証	SMS	phone call	Email	TOTP(認証アプリ)	専用アプリ	セキュリティキー(USB)	生体認証	PIN コード	プッシュ通知 (PC スマホ)
Backup and Sync	Dropbox	https://www.dropbox.com/	アメリカ	○	○	-	-	○	-	○	○	○	△
	icloud(AppleID)	https://www.icloud.com/	アメリカ	○	○	○	-	-	-	○	○	-	△
	Files.com	https://www.files.com/	アメリカ	○	○	-	-	◎	-	○	-	○	-
	Evernote	https://evernote.com/	アメリカ	○	-	-	-	◎	-	-	-	-	-
	firestroage	https://firestorage.jp/	日本	-	-	-	-	-	-	-	-	-	-
	infinicloud	https://infini-cloud.net/	日本	-	-	-	-	-	-	-	-	-	-
Communication	LINE	https://line.me/ja/	日本	-	-	-	-	-	○	-	-	-	-
	Chatwork	https://go.chatwork.com/	日本	○	-	-	-	◎	-	-	-	-	-
	Discord	https://discord.com/	アメリカ	○	○	-	-	◎	△	○	-	○	-
	slack	https://slack.com/	アメリカ	○	○	-	◎	-	-	-	-	-	-
	zoom	https://zoom.us/	アメリカ	○	○	-	-	○	-	○	-	-	-
	miro	https://miro.com/	アメリカ	○	-	-	-	○	-	-	-	-	-
	pixiv	https://www.pixiv.net/	日本	○	-	-	-	○	-	◎	-	○	-
	小説家になろう	https://syosetu.com/	日本	○	-	-	◎	○	-	-	-	-	-
	studio	https://studio.design/	日本	-	-	-	-	-	-	-	-	-	-
	Canva	https://www.canva.com/	オーストラリア	○	-	-	-	△	-	○	○	○	-
	Adobe	https://www.adobe.com/	アメリカ	○	○	-	○	○	○	○	○	○	-
Developer	unity	https://unity.com/	アメリカ	○	○	-	-	-	-	-	-	-	-
	readme	https://readme.com/	外国	○	-	-	-	○	-	-	-	-	-
	openAI	https://openai.com/	アメリカ	○	-	-	-	○	-	-	-	-	-
	Github	https://github.com/	アメリカ	○	○	-	-	○	○	○	○	○	-
	Monaca	https://monaca.mobi/	日本	-	-	-	-	-	-	-	-	-	-

	Qiita	https://qiita.com/	日本	○	-	-	-	○	-	-	-	-	-
	zenn	https://zenn.dev/	日本	-	-	-	-	-	-	-	-	-	-
Domains	Crazy Domains	https://www.crazydomains.com	オーストラリア	○	○	○	-	-	-	-	-	-	-
	101domains	https://www.101domain.com/	アメリカ	○	-	-	-	○	-	-	-	-	-
	namecheap	https://www.namecheap.com/	アメリカ	○	△	-	-	○	-	○	○	○	-
	ライブドアブログ	https://blog.livedoor.com/	日本	-	-	-	-	-	-	-	-	-	-
Entertainment	Netflix	https://www.netflix.com/	アメリカ	-	-	-	-	-	-	-	-	-	-
	DAZN	https://www.dazn.com/	イギリス	-	-	-	-	-	-	-	-	-	-
	Twitch	https://www.twitch.tv/	アメリカ	○	○	-	-	○	-	-	-	-	-
	Spotify	https://www.spotify.com/	スウェーデン	-	-	-	-	-	-	-	-	-	-
	DMM	https://www.dmm.com/	日本	○	-	-	-	○	-	-	-	-	-
	コミックシーモア	https://www.cmoa.jp/	日本	-	-	-	-	-	-	-	-	-	-
	SHOWROOM	https://www.showroom-live.com/	日本	-	△	-	-	-	-	-	-	-	-
	ブックライブ	https://booklive.jp/	日本	-	-	-	-	-	-	-	-	-	-
Food	wolt	https://wolt.com/	フィンランド	-	-	-	-	-	-	-	-	-	-
	UberEats	https://www.ubereats.com/	アメリカ	○	○	-	-	○	-	-	○	-	-
	Starbucks	https://www.starbucks.com/	アメリカ	-	-	-	-	-	-	-	-	-	-
	食べログ	https://tabelog.com/	日本	-	-	-	-	-	-	-	-	-	-
	cookpad	https://cookpad.com/	日本	-	-	-	-	-	-	-	-	-	-
Gaming	VRChat	https://hello.vrchat.com/	アメリカ	○	-	-	-	○	-	-	-	-	-
	Nvidia	https://www.nvidia.com/	アメリカ	△	-	-	-	△	-	△	-	-	-
	Electronic Arts	https://www.ea.com/	アメリカ	-	-	-	-	-	-	-	-	-	-

	Nintendo	https://www.nintendo.com/	日本	○	-	-	-	○	-	○	○	○	-
	KONAMI	https://www.konamistyle.jp/mypage/login.php	日本	○	-	-	○	-	-	-	○	○	-
	CAPCOM	https://www.capcom.co.jp/	日本	○	-	-	-	○	-	-	-	-	-
	gamewith	https://gamewith.jp/	日本	-	-	-	-	-	-	-	-	-	-
Hosting/VPS	KnownHost	https://www.knownhost.com/	アメリカ	○	-	-	-	○	-	-	-	-	-
	PlanetHoster	https://www.planethoster.com/en	カナダ	○	-	-	-	○	-	-	-	-	-
	さくらインターネット	https://www.sakura.ad.jp/	日本	○	○	○	○	○	-	-	-	-	-
	ConoHa	https://www.conoha.jp/	日本	○	-	-	-	○	-	-	-	-	-
Hotels and Accommodations	NH Hotels	https://www.nh-hotels.com/	スペイン	-	-	-	-	-	-	-	-	-	-
	トリバゴ	https://www.trivago.com/	ドイツ	-	-	-	-	-	-	-	-	-	-
	Hotels.com	https://www.hotels.com/	アメリカ	-	-	-	-	-	-	-	-	-	-
	一休	https://www.ikyu.com/	日本	○	○	-	-	-	-	-	-	-	-
	エアトリ	https://www.airtrip.jp/	日本	-	-	-	-	-	-	-	-	-	-
	HIS	https://www.his-j.com/Default.aspx	日本	-	-	-	-	-	-	○	○	○	-
IoT	Xiaomi	https://www.mi.com/	中国	-	-	-	-	-	-	-	-	-	-
	IFTTT	https://ifttt.com/	アメリカ	○	-	-	-	○	-	-	-	-	-
	Nature	https://www.nature.com/	日本	-	-	-	-	-	-	-	-	-	-
	TOLIGO	https://www.toligo.jp/	日本	-	-	-	-	-	-	-	-	-	-
Payments	paypal	https://www.paypal.com/	アメリカ	○	△	△	△	○	△	○	○	○	-
	Stripe	https://stripe.com/en-jp	アメリカ	○	○	-	-	○	-	○	○	○	-
	paypay アプリ	https://paypay.ne.jp/	日本	-	-	-	-	-	-	-	○	-	-

	Vpass	https://www.smbc-card.com/	日本	-	-	-	-	-	-	-	-	-	-
	MIXI M	https://m.mixi.com/	日本	-	-	-	-	-	-	○	○	○	-
Retail	Amazon	https://www.amazon.com/	アメリカ	○	-	-	-	○	-	○	○	○	-
	ZARA	https://www.zara.com/	スペイン	-	-	-	-	-	-	-	-	-	-
	楽天市場	https://www.rakuten.co.jp/	日本	○	-	-	○	-	-	○	○	○	-
	zozotown	https://zozo.jp/	日本	-	-	-	-	-	△	-	-	-	-
	minne	https://minne.com/	日本	○	○	○	-	-	△	-	-	-	-
	Yahoo! JAPAN	https://www.yahoo.co.jp/	日本	○	○	-	○	-	○	○	○	○	-
Security	Norton	https://norton.com/	アメリカ	○	△	△	-	○	-	○	-	-	-
	DNSFilter	https://www.dnsfilter.com/	アメリカ	○	-	-	-	○	-	-	-	-	-
	Malwarebytes	https://my.malwarebytes.com/	アメリカ	○	-	-	○	-	-	-	-	-	-
	Trend Micro	https://www.trendmicro.com/ja_jp/forHome.html	日本	○	○	-	○	○	-	-	-	-	-
	ソースネクスト	https://www.sourcenext.com/	日本	△	-	-	-	△	-	-	-	-	-
Social	X	https://x.com	アメリカ	○	-	-	-	○	-	○	-	-	-
	instagram	https://www.instagram.com/	アメリカ	○	○	-	-	○	-	-	-	-	-
	facebook	https://www.facebook.com/	アメリカ	○	○	-	-	○	-	○	-	-	-
	TikTok	https://www.tiktok.com/	中国	○	○	-	-	○	-	△	△	△	-
	LinkedIn	https://www.linkedin.com/	アメリカ	○	○	-	-	○	-	○	○	○	-
	Reddit	https://www.reddit.com/	アメリカ	○	-	-	-	○	-	-	-	-	-
	yay	https://yay.space/	日本	○	-	-	-	○	-	-	-	-	-
	Wantedly	https://www.wantedly.com/	日本	-	-	-	-	-	○	-	-	-	-

	mixi	https://mixi.jp/	日本	-	-	-	-	-	-	-	-	-	-
Cloud Computing	DigitalOcean	https://www.digitalocean.com/	アメリカ	-	○	-	-	○	-	-	-	-	-
	Heroku	https://www.heroku.com/	アメリカ	○	-	-	-	○	○	○	○	○	○
Education	udemy	https://www.udemy.com/	アメリカ	○	-	-	○	-	-	-	-	-	-
Task Management	Nifty	https://niftypm.com/	アメリカ	○	-	-	-	○	-	-	-	-	-
Universities	Oh-o! Meiji	https://oh-o2.meiji.ac.jp/	日本	○	○	○	-	○	-	○	-	-	-
VPN Providers	Cyberghost	https://www.cyberghostvpn.com/	ルーマニア	-	-	-	-	-	-	-	-	-	-
other	リクルート	https://www.recruit.co.jp	日本	○	○	-	○	○	-	-	-	-	-
	マイナビ	https://job.mynavi.jp/	日本	○	○	-	○	-	-	-	-	-	-
	one career	https://www.onecareer.jp/	日本	-	-	-	-	-	-	-	-	-	-
	GMO	https://point.gmo.jp/	日本	○	-	-	○	○	-	-	-	-	-
	Mercari	https://jp.mercari.com/	日本	-	-	-	-	-	-	○	○	○	-
	au(au ID)	https://id.auone.jp/	日本	○	○	-	○	-	○	-	-	-	-
	docomo(d アカウント)	https://id.smt.docomo.ne.jp/	日本	○	○	-	-	-	○	-	-	-	-
	Google	https://www.google.co.jp/	アメリカ	○	○	○	-	○	-	○	○	○	-
	Microsoft	https://www.microsoft.com/	アメリカ	○	○	-	○	◎	-	○	○	○	-

謝辞

本研究を行うにあたり，多くの方より御指導いただきました．特に，多大なる御指導を受け賜りました，明治大学総合数理学部先端メディアサイエンス学科，菊池浩明教授に深く感謝申し上げます．また，研究にご協力，ご助言頂いた太田和希氏をはじめとした菊池研究室の皆様に深く感謝の意を表すとともに，謝辞とさせていただきます．

参考文献

[1] NTT セキュリティジャパン，“ Cyber Security Report 2022 ”，2022. https://jp.security.ntt/resources/cyber_security_report/CSR_202209.pdf

[2] Microsoft，“ Microsoft Digital Defense Report 2024 ”，2024. https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Microsoft%20Digital%20Defense%20Report%202024%20(1).pdf

[3] 太田和希，菊池浩明，“ 2FA をバイパスする中間者攻撃ツール Evilginx によるフィッシング攻撃の脅威分析 ”，情報処理学会第 87 回全国大会，講演番号 1ZE-04，2025.

[4] Passkey Central, “Introduction to Passkeys — Phishing Resistance,”https://www.passkeycentral.org/introduction-to-passkeys/passkey-security#phishing-resistance

[5] e-Words, “セッション ID とは”，https://e-words.jp/w/%E3%82%BB%E3%83%83%E3%82%B7%E3%83%A7%E3%83%B3ID.html

[6] NEC ソリューションイノベータ，“ ソーシャルログインとは？仕組みやメリット・デメリットを解説 ”，https://www.nec-solutioninnovators.co.jp/sp/contents/column/20220114.html

付 録 A　100 万人のヘルスケアデータからの心疾患の罹患率

A.1　はじめに

日本国民の死因第 2 位 [1] である心疾患には、高血圧や糖尿病が要因とされるものが多い. Saito らによって、HDL-コレステロール値が心疾患の罹患率に関与していることが判明している [5].
そこで本研究では、ヘルスケア企業によって匿名化されたビックデータを活用する. どの因子が心疾患の罹患率に影響を及ぼすのか重回帰を行う、更に、匿名性を理論的に保証するように、ヘルスケアデータに対してプライバシー費用 ε に応じて連続値にはラプラスノイズを加算し、離散値にはランダマイズレスポンス [2] を適用したデータを作成し、ロジスティクス回帰モデルに学習させ、オリジナルのヘルスケアデータに対する予測精度を評価する.

A.2 実験

A.2.1 ヘルスケアデータの概要

ヘルスケアデータには 2014 年 4 月〜2021 年 9 月までの健康診断のデータが記載されている. 本実験では 34 種類 (離散値 24 種類・連続値 10 種類) の変数を抽出し、1048576 名のデータを使用するそのうち 780 名が心疾患に罹患している. 中性脂肪、HDL コレステロール、LDL コレステロール、γ-GT は可視吸光光度法で測定され、GOT(AST)、GPT(ALT) は紫外吸光光度法で測定されている. 目的変数 'icd10-medium-code' は心疾患を罹患している人は 1 、それ以外の人は 0 に設定した. 表 A.1 に心疾患とみなした icd-10 中分類コードの一覧を示す. 本データの説明変数 (連続値) の統計量を表 A.2 に示す.

表 A.1: ICD-10 コードに基づく心疾患の分類

コード名	病名
A30-A49	その他の細菌性疾患
I00-I02	急性リウマチ熱
I10-I15	高血圧性疾患
I20-I25	虚血性心疾患
I26-I28	肺性心疾患及び肺循環疾患
I30-I52	その他の型の心疾患
M30-M36	全身性結合組織障害
O10-O16	妊娠，分娩及び産じょくにおける浮腫，たんぱく尿及び高血圧性障害
O30-O48	胎児及び羊膜腔に関連する母体ケア並びに予想される分娩の諸問題
Q20-Q28	循環器系の先天奇形

表 A.2: 主要変数の統計量

属性名	平均	標準偏差	最小値	中央値	最大値
height	158.9	9.01	123.9	158.1	191.7
weight	58.4	11.58	27.2	57.1	137.9
fukui	83.4	9.70	55.0	83.0	121.0
chusei-shibou-kashi	116.0	74.90	21.0	97.0	1514.0
hdl-kashi	64.1	16.98	24.0	62.0	141.0
ldl-kashi	122.0	30.19	30.0	121.0	252.0
got-shigai	24.0	10.39	9.0	22.0	197.0
gpt-shigai	21.4	13.67	4.0	18.0	232.0
gamma-gt-kashi	34.1	43.82	5.0	22.0	845.0
age	65.4	8.86	18.0	68.0	85.0

A.2.2 実験方法

ヘルスケアデータ内の、33 変数を説明変数、'icd10-medium-code' を目的変数としたロジスティック回帰モデルを構築する. またデータを学習させる際に LDP(連続値：ラプラスノイズ、離散値：ランダマイズレスポンス) あり、なしの 2 種類のモデルを作成し、適合率 ・ 再現率 ・ F1 スコアの 3 種類を評価する. 匿名性を保証する前のデータの年齢・体重・身長に関して心疾患罹患者の分布を図 A.1〜図 A.3 に示す. 図 A.1〜図 A.3 は黒色の部分が罹患者が少なく、白の部分が罹患者が多い.

A.2.3 結果

回帰変数を表 A.4 に示す. 表 A.4 は心疾患の罹患率が性別、コレステロール値を下げる薬を飲んでいるか、週に 2 以上日軽く汗をかく運動をしているかの影響を受ける事を示す. 図 A.1、図 A.2 は心疾患の罹患者は一定のグループを形成している事を示す. 表 A.3 は ϵ による精度の変化を示す.

表 A.3: 精度

ϵ	適合率	再現度	F
元データ	0.998	0.999	0.999
2	0.869	0.932	0.900
5	0.990	0.995	0.993
10	0.998	0.999	0.999

表 A.4: 回帰変数

特徴量	回帰変数	P	オッズ比
const	-13.612	0.000	0.000
身長 (cm)	0.053	0.000	1.055
体重 (kg)	-0.021	0.167	0.979
BMI (kg/m^2)	0.116	0.003	1.122
腹囲 (cm)	-0.025	0.000	0.975
中性脂肪 (mg/dl)	-0.002	0.000	0.998
HDL コレステロール (mg/dl)	-0.014	0.000	0.986
LDL コレステロール (mg/dl)	-0.004	0.000	0.996
GOT(AST)(U/I)	0.014	0.000	1.014
GPT(ALT)(U/I)	-0.004	0.043	0.996
γ-GT(γ-GTP)(U/I)	-0.003	0.000	0.997
保健指導レベル	-0.228	0.000	0.796
血圧を下げる薬の有無	0.005	0.907	1.005
インスリン注射又は血糖を下げる薬の有無	0.170	0.012	1.185
コレステロールを下げる薬の有無	0.534	0.000	1.706
脳卒中の罹患の有無	-0.809	0.000	0.445
心臓病の罹患の有無	-0.286	0.000	0.751
慢性の腎不全又は人工透析の有無	0.295	0.136	1.343
たばこの使用歴	-0.186	0.001	0.830
体重変化	-0.099	0.042	0.906
運動習慣	0.373	0.000	1.453
歩行習慣	-0.103	0.009	0.902
性別	0.619	0.000	1.858

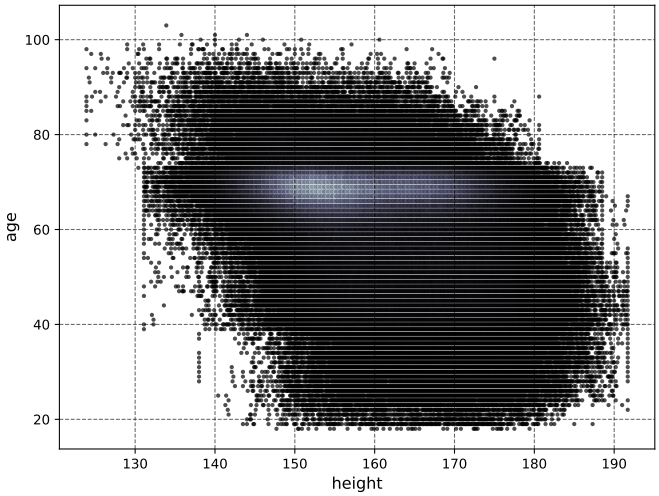

図 A.1: 身長と年齢の散布図

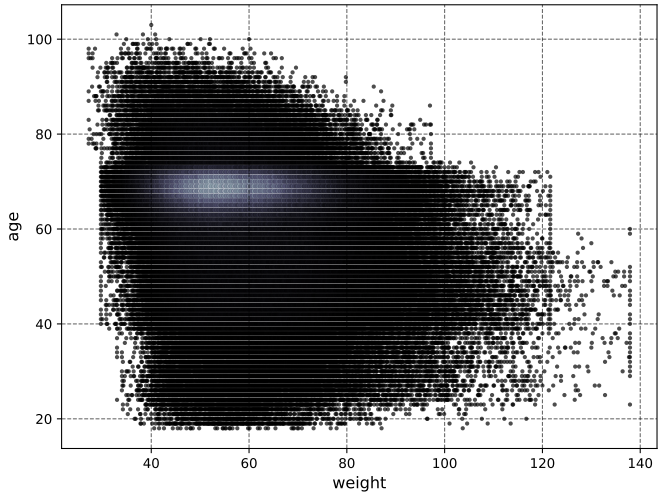

図 A.2: 体重と年齢の散布図

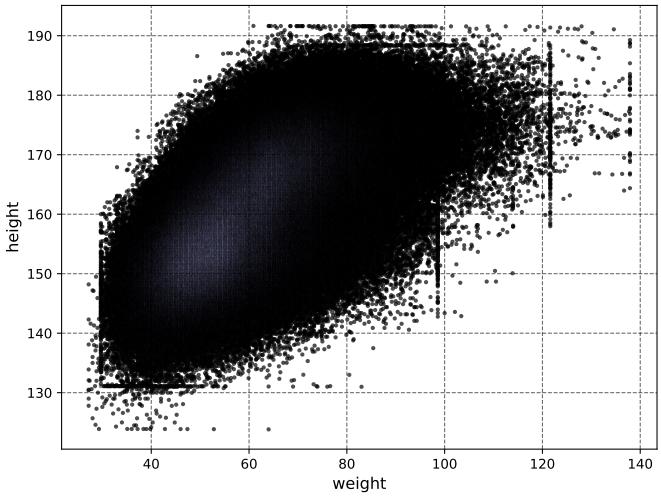

図 A.3: 体重と身長の散布図

A.2.4　考察

本研究において、心臓病の罹患歴がある人は心疾患の罹患率が低くなることが分かる. 表 A.4 は身長が罹患率に大きな影響を及ぼしていないことが示している. 一方で、図 A.3 より身長と体重が一定の組み合わせの人が心疾患を罹患している事が分かる. この矛盾のように見える現象は、回帰分析と散布図における解釈の違いによるものであると考えられる. すなわち、回帰分析では各変数を独立した単変量として評価するのに対し、散布図は複数の変数間の関係性を視覚化するため、同じ変数であっても異なる視点で情報を捉える事が出来る.

A.3　おわりに

本研究の結果は、心疾患罹患率に影響を及ぼす要因として生活習慣の重要性を示唆するものであった. しかし、罹患率の低さが解析結果に影響を及ぼした可能性を否定できない. 今後の研究では、より適切なデータセットを用いることで、この問題を克服し、心疾患予防に向けた具体的なエビデンスを提供することを目指す. 複合変数の生成やその効果の検証を通じて、より高精度なモデルの構築が期待される.

参考文献

[1] 厚生労働省, ‘令和 5 年 人口動態統計の概況’, (https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/geppo/nengai23/dl/gaikyouR5.pdf) (参照 2024 年 12 月).

[2] 伊藤聡志, 池上和輝, 菊池浩明, ‘健康診断データとレセプトデータの匿名加工情報を用いた疾病リスク分析’.　情報処理学会論文誌 Vol.62 No.9 1560–1574 (Sep. 2021)

[3] 野田博之 他, ‘住民健診の結果に基づいた脳卒中・虚血性心疾患の予測’, 日本公衛誌, Vol.53, No.4, pp.265–276 (2006).

[4] 菊池浩明 他, ‘DPC データベースからのプライバシーを保護した線形回帰による入院日数モデルの学習’, DICOMO2014 シンポジウム, pp.219–223 (2014).

[5] Saito, I., et al., ‘Association of high-density lipoprotein cholesterol concentration with different types of stroke and coronary heart disease’, Atherosclerosis, Vol.265, pp.147–154, (https://doi.org/10.1016/j.atherosclerosis.2017.08.032) (2017).