Machine learning approaches for prediction of major bleeding events in anticoagulated atrial fibrillation patients

Fatemeh Torabi, Arron Lacey, Ashley Akbari, Daniel Harris, Ronan A Lyons, Julian Halcox, Mike Gravenor

Direction of interest

I am research officer and data scientist at HDR UK

My background is in mathematical statistics and health data science

This work:

Ethical approval from Swansea University Research Committee: 2018-0056

Information Governance Review Panel (IGRP): SAIL0866

fatemeh.torabi@swansea.ac.uk

fatemeh.torabi@swansea.ac.uk

Background

- Atrial fibrillation (AF) increases risk of stroke by 5 to 7 folds.
- Anticoagulant (AC) drugs can prevent blood clots formation.
- Appropriate use of AC drugs can reduce stroke risk by two-thirds in AF patients.

- Bleeding is the main side effect of AC drugs that influences clinical decision of anticoagulation to some extent.
- Various bleeding risk schemes were introduced and used to evaluate risk of bleeding in AF patients.
- None of these schemes to date have been validated using machine learning classification approaches.

HAS-BLED
(Pisters et al., 2010)

ATRIA
(Fang et al., 2011)

ORBIT
(O' Brien et al., 2015)

ABC
(Hijazi et al., 2016)

fatemeh.torabi@swansea.ac.uk

HAS-BLED (Pisters et al., 2010)

Hypertension	1
Abnormal renal/liver function	1or2
Stroke	1
Bleeding history or predisposition	1
Labile INR	1
Elderly (age over 65)	1
Drug usage / alcohol concomitantly	1or2

fatemeh.torabi@swansea.ac.uk

Different approaches to risk prediction

All four studies identified a set of contributing factors and used a method to develop the prediction models

Study	Method	Data	Subject counts	Reported AUC
HAS-BLED (Pisters et al., 2010)	Multivariate logistic regression	Large population database: Euro Heart Survey	5,333	0.72
ATRIA (Fang et al., 2011)	Time-varying cox proportional hazard ratio with split-sample	Anticoagulation and Risk Factors in Atrial Fibrillation (ATRIA)	13,559	0.69
ORBIT (O'Brien et al., 2015)	Cox proportional hazard ratio	Outcome Registry for Better Informed Treatment of AF (ORBIT-AF)	7,411	0.69-0.67
ABC (Hijazi et al., 2016)	Cox proportional hazard ratio	ATIRSTOTLE and RE_LY trials	14,537 (ARISTOTLE trial) 8,468 (RE-LY trial)	0.68

fatemeh.torabi@swansea.ac.uk

Objectives

- Validate HAS-BLED in Welsh AF population
- Comparing performance of various models in predicting bleeding events

fatemeh.torabi@swansea.ac.uk

Data access and inclusion criteria

Data coverage
Data was accessed 2000-01-01 to 2018-01-01
Study period: 2017-01-01 to 2018-01-01

Patient Episode Dataset for Wales
PEDW

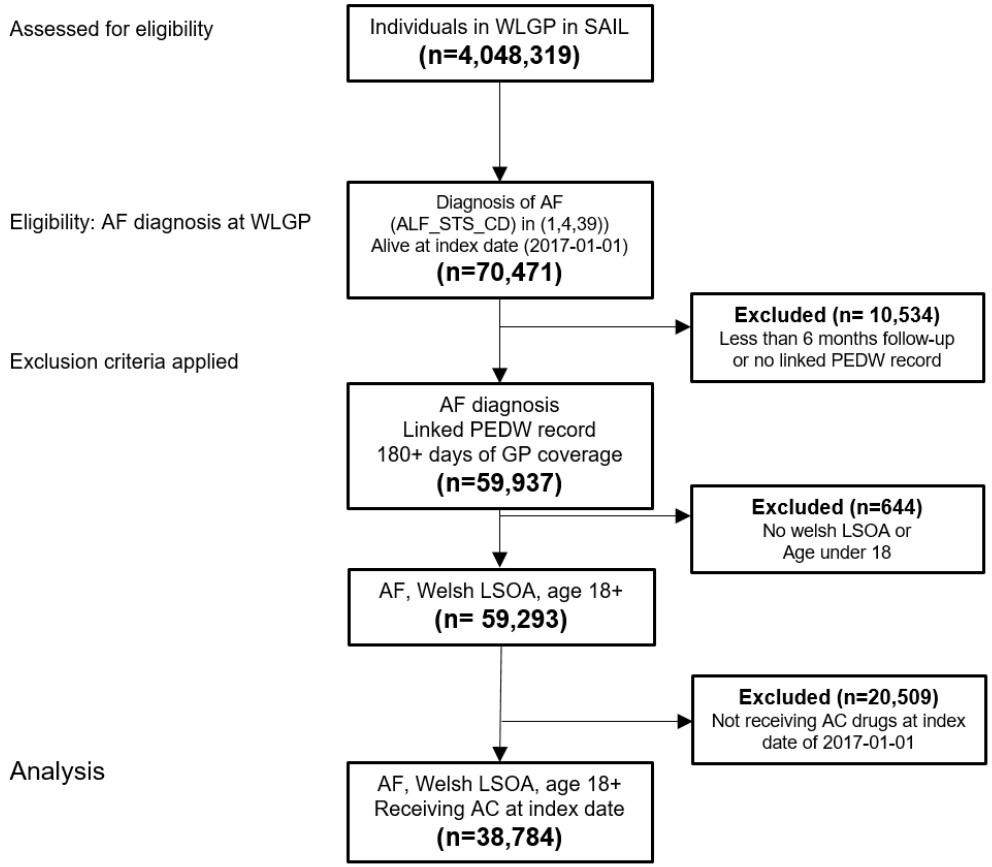
Annual District Death Extract
ADDE

Welsh Demographic Service Dataset
WDSD

WLGP
Welsh Longitudinal General Practice Dataset

Metadata of all created fields

fatemeh.torabi@swansea.ac.uk



WLGP: Welsh Longitudinal General Practice | PEDW: Patient Episode Dataset for Wales
AC: AntiCoagulant drugs
All counts in the above figure are distinct patient counts

Comorbidities

Hypertension

Stroke

Elderly

Abnormal Renal Function

Prior bleeding

Drugs or Medications

Abnormal Liver Function

Labile International Normalized Ratio

Alcohol use

BMI

Diabetes

Smoking

Vascular disease

Congestive Heart Failure

Gender

fatemeh.torabi@swansea.ac.uk

Modelling

STEP 1

Validated HAS-BLED by using same set of variables as (Pisters et el. 2010) in a logistic regression model on SAIL AF cohort.

The delta features were also added to this model. (Chao et al., 2918).

STEP 2

Variables from HAS-BLED score were evaluated in:
Cox proportional hazard ratio

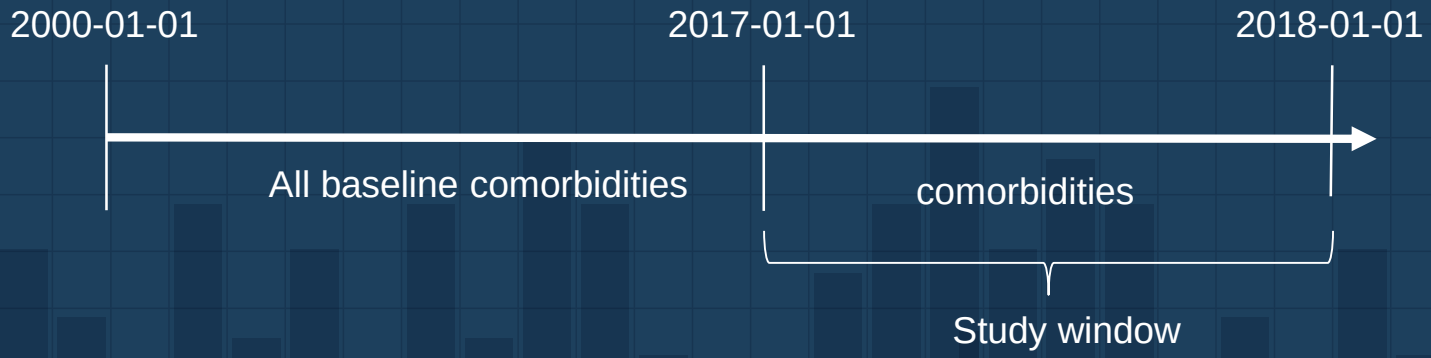
In addition we evaluated all clinically relevant variables.

STEP 3

The logistic modelling approach was compared to two different machine learning approaches for prediction:
Random Forest
Naïve Bayes

fatemeh.torabi@swansea.ac.uk

Delta feature (Chao et al.,2018): absolute changes in baseline comorbidities within study period



Delta features= |comorbidities – Baseline comorbidities|

fatemeh.torabi@swansea.ac.uk

Follow-up of 2018-01-01

Baseline of 2017-01-01

HAS-BLED
CHA_2DS_2VASc

Delta features HAS-BLED

fatemeh.torabi@swansea.ac.uk

38,784

(Mean age of 76 and 43.5% female)

Average follow-up 96% of study window: total of 37,133 person years in study

fatemeh.torabi@swansea.ac.uk

Random Forest

- Each tree is identically distributed as each one is grown using a randomization strategy: using bootstrapping and have a finite outcome => for a given set of samples, a statistic of interest will converge to a mean (WLLN).
- WLLN: independence & identically distributed:
- $P(|X-\mu| \geq k) \leq \frac{var(X)}{k^2}$

Naïve Bayes

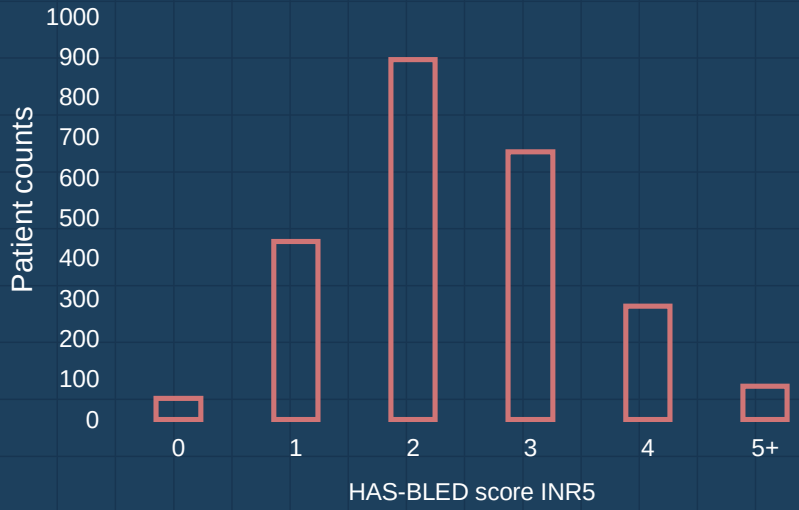
- Probabilistic classifier based on conditional probability of event of interest and Bayes theorem:
- $p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)}$
- $posterior = \frac{prior\ .likelihood}{evidence}$

fatemeh.torabi@swansea.ac.uk

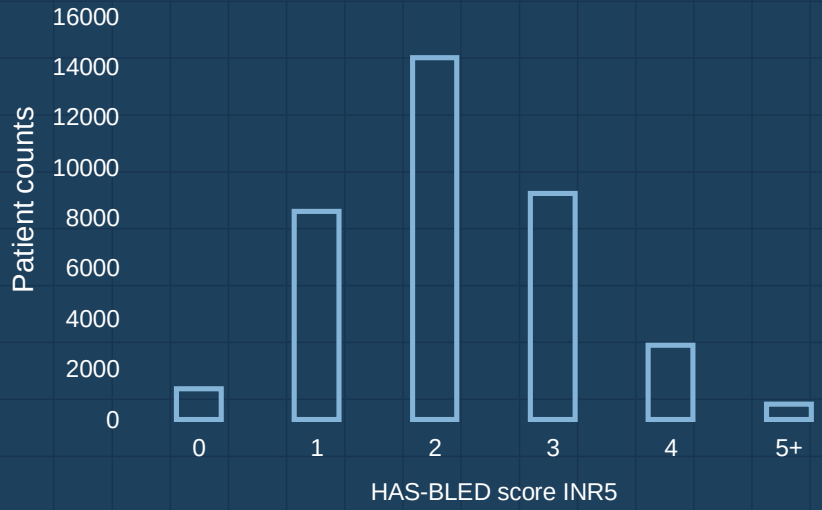
Results

Baseline Characteristics	No bleed	Bleed	P-value
n	36368	2416	
Age mean(sd)	75.93 (9.85)	77.43 (9.51)	<0.001
18-65	4268 (11.7)	217 (9.0)	
65-74	11880 (32.7)	699 (28.9)	
75+	20220 (55.6)	1500 (62.1)	
Female	15875 (43.7)	981 (40.6)	0.147
CHA2DS2Vasc score	3.43 (1.46)	3.76 (1.52)	<0.001
0	738 (2.0)	26 (1.1)	
1	2595 (7.1)	117 (4.8)	
2	6187 (17.0)	345 (14.3)	
3	9343 (25.7)	590 (24.4)	
4	9156 (25.2)	609 (25.2)	
5	5537 (15.2)	417 (17.3)	
6	2217 (6.1)	231 (9.6)	
7	521 (1.4)	60 (2.5)	
8	68 (0.2)	15 (0.6)	
9	6 (0.0)	6 (0.2)	
HAS-BLED score INR5	2.17 (1.03)	2.39 (1.09)	<0.001
0	1221 (3.4)	52 (2.2)	
1	8260 (22.7)	442 (18.3)	
2	14362 (39.5)	893 (37.0)	
3	8968 (24.7)	665 (27.5)	
4	2953 (8.1)	282 (11.7)	
5+	604 (1.6)	82 (3.4)	
HAS-BLED score INR8	2.01 (0.97)	2.21 (1.02)	<0.001
0	1401 (3.9)	58 (2.4)	
1	9554 (26.3)	538 (22.3)	
2	15315 (42.1)	960 (39.7)	
3	7746 (21.3)	612 (25.3)	
4	2021 (5.6)	206 (8.5)	
5+	333 (0.9)	44 (1.8)	
History of bleed (hospital)	8160 (22.4)	842 (34.9)	0.226
History of bleed (GP)	2672 (7.3)	344 (14.2)	0.313
Stroke	5558 (15.3)	369 (15.3)	0.267
Congestive Heart failure	10871 (29.9)	827 (34.2)	0.19
Hypertension	20490 (56.3)	1395 (57.7)	0.147
Liver disease	1085 (3.0)	80 (3.3)	0.339
Renal disease	650 (1.8)	99 (4.1)	0.346
Labile INR (over5)	9837 (27.0)	736 (30.5)	0.203
Labile INR (over8)	4248 (11.7)	301 (12.5)	0.288
Alcohol consumption	893 (2.5)	45 (1.9)	0.342
Vascular disease	12646 (34.8)	960 (39.7)	0.17
Diabetes	9837 (27.0)	716 (29.6)	0.203
Antiplatelet drugs	23315 (64.1)	1586 (65.6)	0.167

BLEED



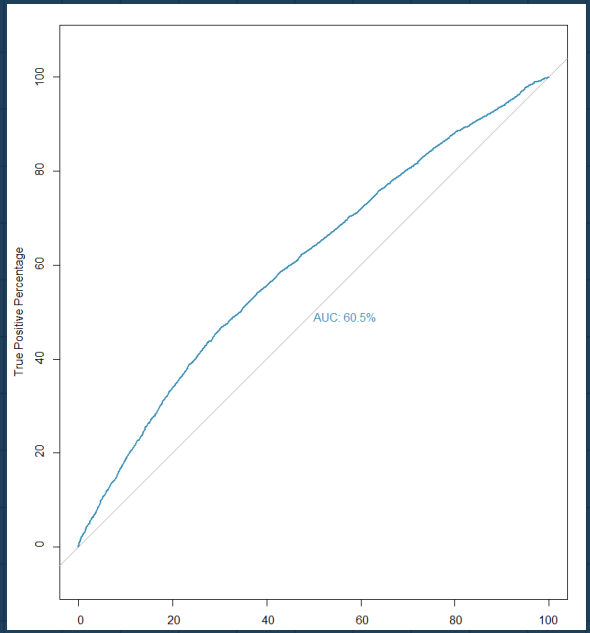
NO-BLEED



fatemeh.torabi@swansea.ac.uk

Validation of HAS-BLED

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.19231	0.08112	-39.354	< 2e-16
Hypertension-BL	0.01861	0.04288	0.434	0.6643
Abnormal Liver function-BL	0.09775	0.11961	0.817	0.4138
Abnormal renal function-BL	0.75325	0.11117	6.776	1.24E-11
Stroke-BL	-0.07819	0.05896	-1.326	0.1848
Any major bleeding-BL	0.58798	0.04488	13.101	< 2e-16
INR 5 or above-BL	0.10162	0.04631	2.195	0.0282
Age 65-74	0.10803	0.08017	1.348	0.1778
Age 75+	0.30986	0.07567	4.095	4.22E-05
Drug and medication use	0.07873	0.04454	1.768	0.0771
Alcohol-BL	-0.21876	0.15719	-1.392	0.164



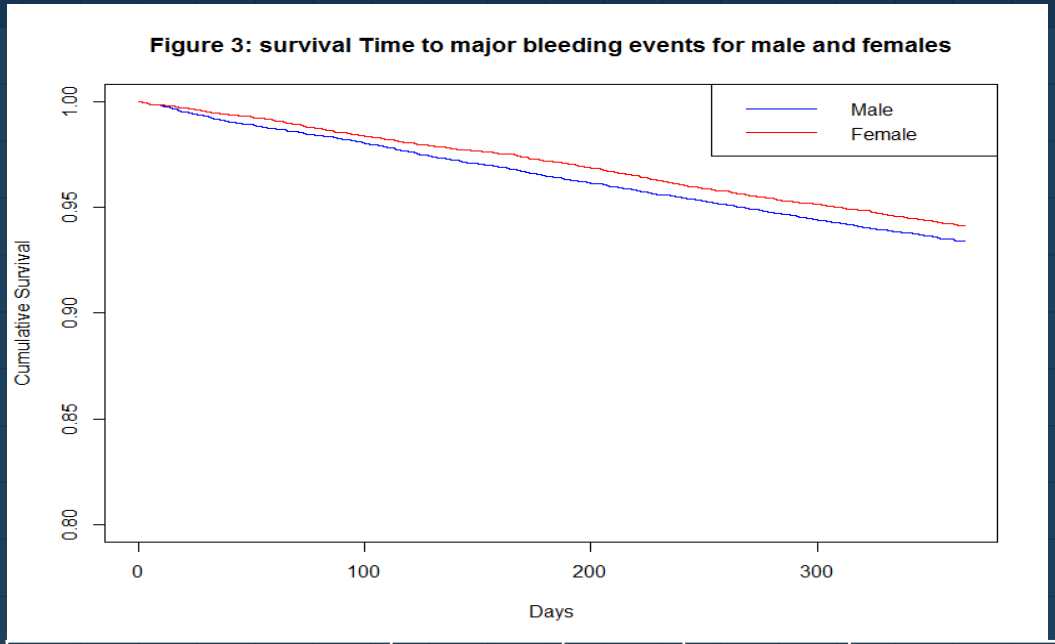
fatemeh.torabi@swansea.ac.uk

Hazard ratio of HAS-BLED variables

Covariates Entire sample(n=35,784) Number of events (n=2,416)					
	Coefficient	HR	Lower 95% CI	Upper 95% CI	Pr(>|z|)
Hypertension-BL	0.01721	1.01736	0.9381	1.103	0.6776
Abnormal Liver function-BL	0.09907	1.10414	0.8815	1.383	0.3887
Abnormal renal function-BL	0.71825	2.05084	1.6751	2.511	3.52E-12
Stroke-BL	-0.07029	0.93213	0.8338	1.042	0.2163
Any major bleeding-BL	0.56805	1.76482	1.6222	1.92	< 2e-16
INR 5 or above-BL	0.10004	1.10521	1.0128	1.206	0.0247
Age 65-74	0.10625	1.11209	0.9545	1.296	0.173
Age 75+	0.30842	1.36128	1.1785	1.572	2.75E-05
Drug and medication use	0.07643	1.07942	0.9922	1.174	0.0755
Alcohol-BL	-0.20786	0.81232	0.602	1.096	0.1738
Concordance= 0.592 (se = 0.006) Likelihood ratio test= 270.9 on 10 df, p=<2e-16 Wald test= 295.9 on 10 df, p=<2e-16 Score (logrank) test = 305.9 on 10 df, p=<2e-16					

fatemeh.torabi@swansea.ac.uk

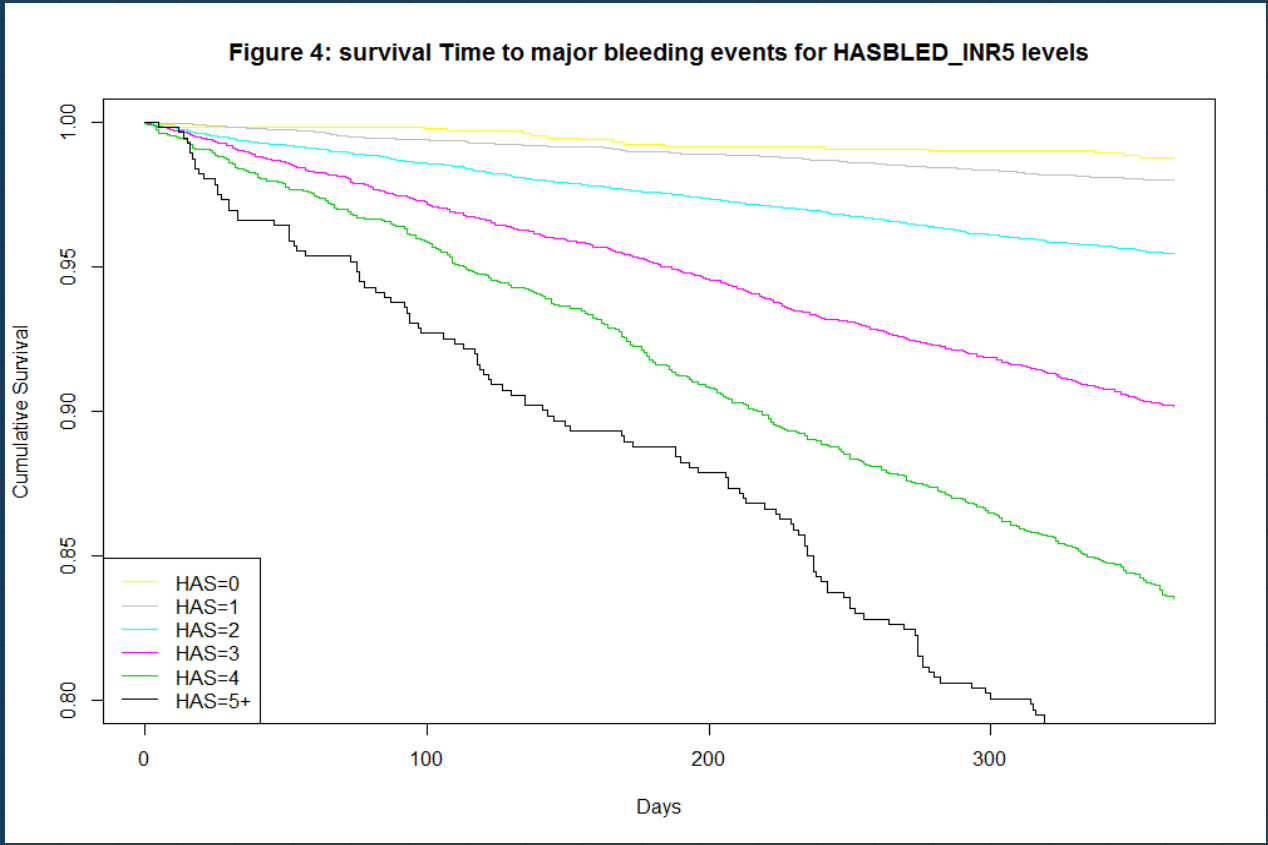
Hazard ratio of Bleeds stratified by Gender



Covariates Entire sample(n=35,784) Number of events (n=2,416)					
	Coefficient	HR	Lower 95% CI	Upper 95% CI	Pr(>|z|)
Gender	-0.12175	0.88537	0.8163	0.9603	0.00329

Concordance= 0.515 (se = 0.005)
Likelihood ratio test= 8.7 on 1 df, p=<0.003
Wald test= 8.64 on 1 df, p=<0.003
Score (logrank) test = 8.65 on 1 df, p=<0.003

fatemeh.torabi@swansea.ac.uk



fatemeh.torabi@swansea.ac.uk

Hazard ratio of bleed for levels of CHA_2DS_2VASc score

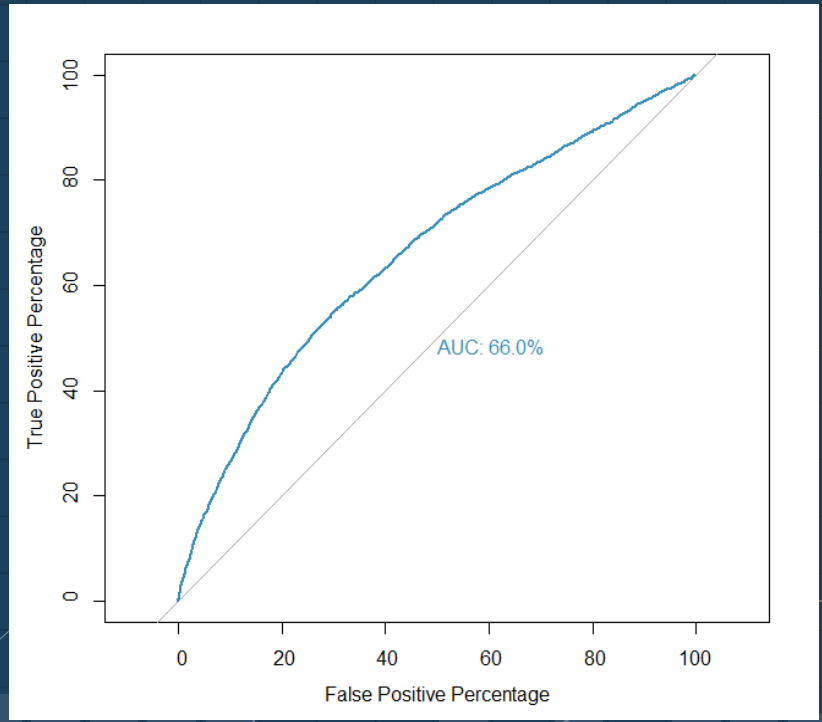
Covariates Entire sample(n=35,784) Number of events (n=2,416)					
	Coefficient	HR	Lower 95% CI	Upper 95% CI	Pr(>|z|)
Hypertension-BL	-0.11595	0.89052	0.8125	0.976	1.32E-02
Abnormal Liver function-BL	0.05324	1.05468	0.8416	1.3217	0.64379
Abnormal renal function-BL	0.62206	1.86275	1.5182	2.285	2.49E-09
Stroke-BL	-0.11332	0.89287	0.798	0.9991	0.04812
Any major bleeding-BL	0.54696	1.72798	1.5879	1.8804	< 2e-16
INR 5 or above-BL	0.07044	1.07298	0.9828	1.1715	0.11594
Age 65-74	-0.07152	0.93098	0.7825	1.1076	0.41983
Age 75+	0.01182	1.01189	0.8421	1.2159	0.89966
Drug and medication use	0.07537	1.07828	0.9911	1.1731	0.07963
Alcohol-BL	-0.20036	0.81843	0.6066	1.1043	0.18986
CHADSVASC_BL2	0.26483	1.30321	0.8454	2.0088	0.23034
CHADSVASC_BL3	0.45974	1.58367	1.0366	2.4194	0.03348
CHADSVASC_BL4	0.56999	1.76824	1.1498	2.7193	0.00944
CHADSVASC_BL5	0.60841	1.83751	1.1857	2.8478	0.00649
CHADSVASC_BL6	0.70665	2.0272	1.2973	3.1678	0.00192
CHADSVASC_BL7	0.97367	2.64764	1.6726	4.1912	3.26E-05
CHADSVASC_BL8	1.10446	3.01759	1.8122	5.0247	2.18E-05
CHADSVASC_BL9	1.71544	5.55913	2.8332	10.9077	6.09E-07
CHADSVASC_BL10	3.08568	21.88232	8.7431	54.7673	4.33E-11

Concordance= 0.602 (se = 0.006)
Likelihood ratio test= 345.1 on 19 df, p=<2e-16
Wald test= 394.5 on 19 df, p=<2e-16
Score (logrank) test = 432.1 on 19 df, p=<2e-16

fatemeh.torabi@swansea.ac.uk

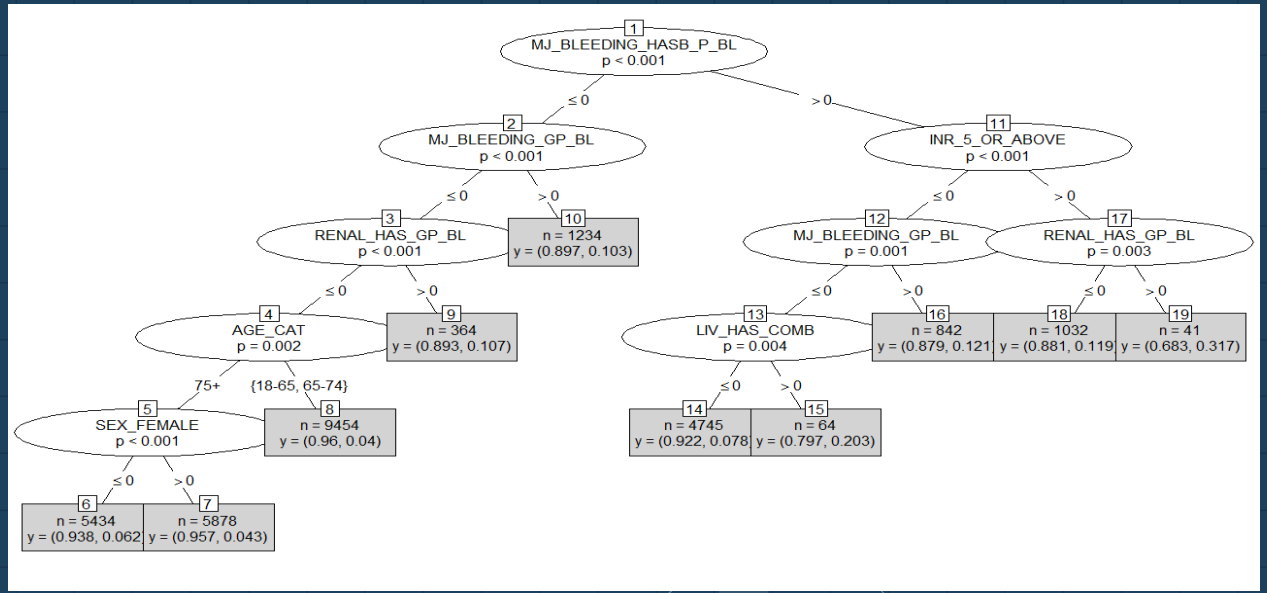
Logistic regression model with Delta features

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.26863	0.083548	-39.123	< 2e-16
HYPT_GP_BL	0.006104	0.043383	0.141	0.888113
LIV_HAS_COMB_BL	-0.17641	0.127115	-1.388	0.165196
RENAL_HAS_GP_BL	0.629284	0.117768	5.343	9.12E-08
SEX_FEMALE	-0.19863	0.044009	-4.513	6.38E-06
MJ_BLEEDING_HASB_P_BL	0.484852	0.046263	10.48	< 2e-16
MJ_BLEEDING_GP_BL	0.623708	0.063883	9.763	< 2e-16
INR_5_OR_ABOVE_BL	0.06435	0.05818	1.106	0.268704
INR_8_OR_ABOVE_BL	-0.12799	0.080596	-1.588	0.112273
AGE_CAT65-74	0.108432	0.080862	1.341	0.179938
AGE_CAT75+	0.266853	0.076781	3.475	0.00051
MEDICATION_HASB_BL	0.082058	0.045055	1.821	0.068562
ALCOHOL_BL	-0.37285	0.162709	-2.292	0.021933
CHF_CHAD_COMB_BL	-0.14946	0.048965	-3.052	0.00227
CHF_CHAD_COMB	1.053622	0.055432	19.008	< 2e-16
HYPT_GP	0.04011	0.169044	0.237	0.812444
LIV_HAS_COMB	1.093149	0.144699	7.555	4.20E-14
RENAL_HAS_GP	0.25888	0.229182	1.13	0.258653
INR_5_OR_ABOVE	0.302235	0.083901	3.602	0.000315
INR_8_OR_ABOVE	-0.07188	0.103114	-0.697	0.485766
ALCOHOL	-0.06323	0.477405	-0.132	0.89464

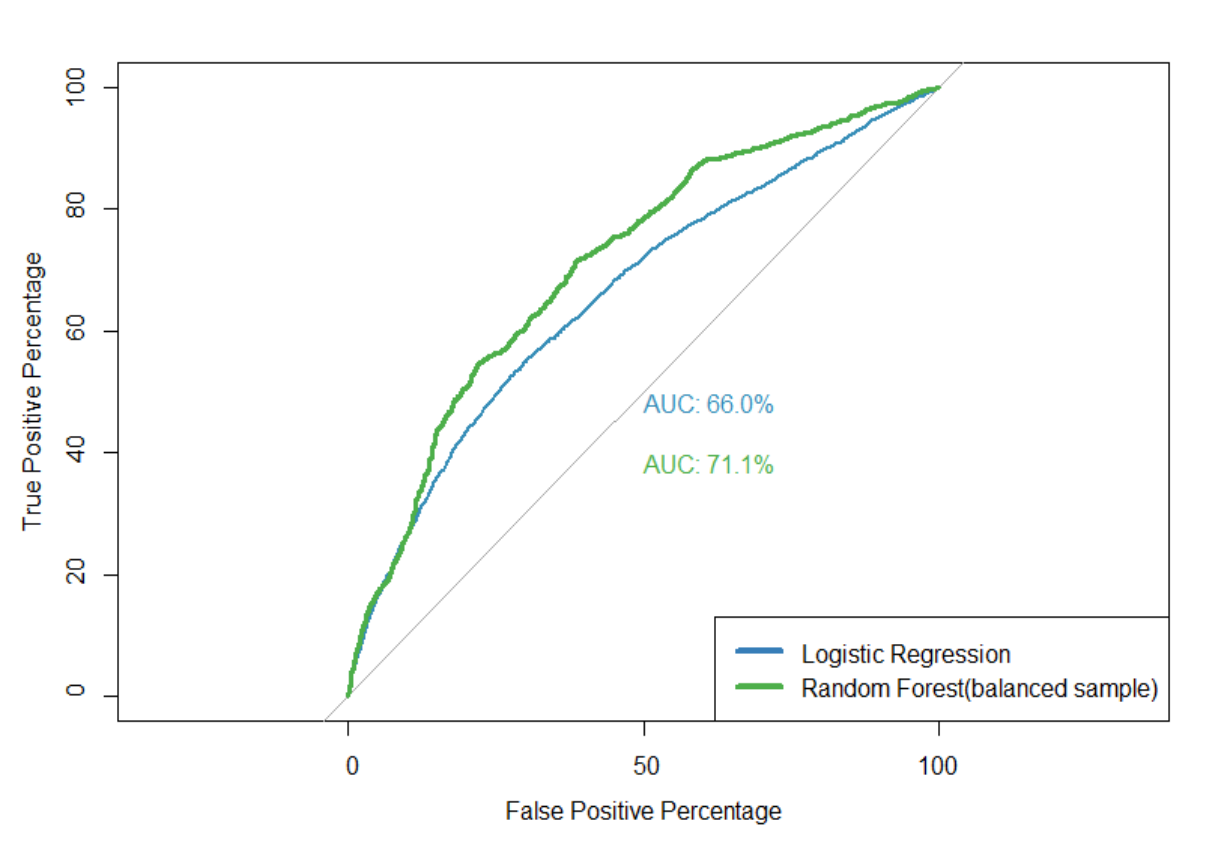


fatemeh.torabi@swansea.ac.uk

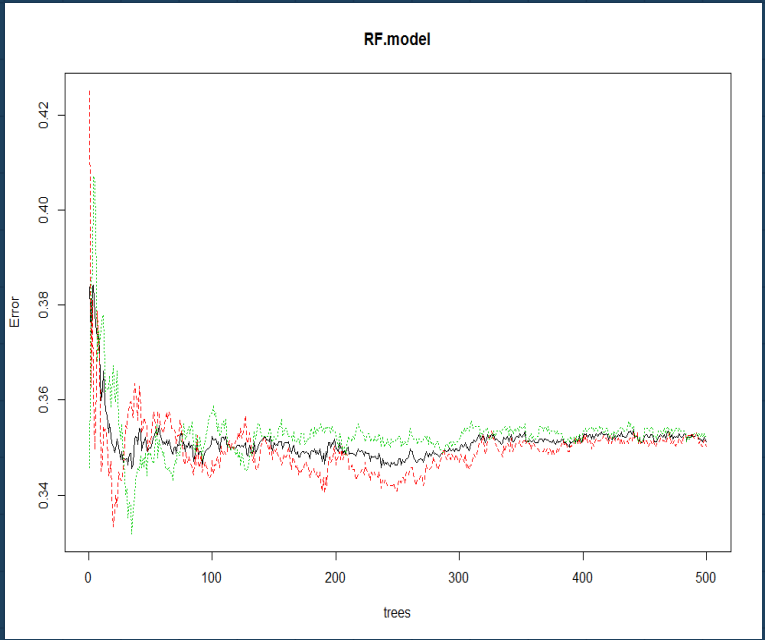
Random Forest

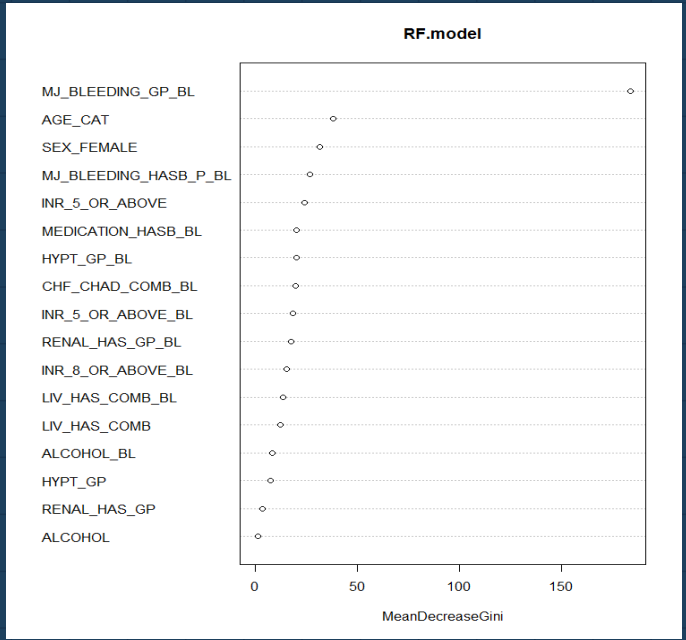


fatemeh.torabi@swansea.ac.uk



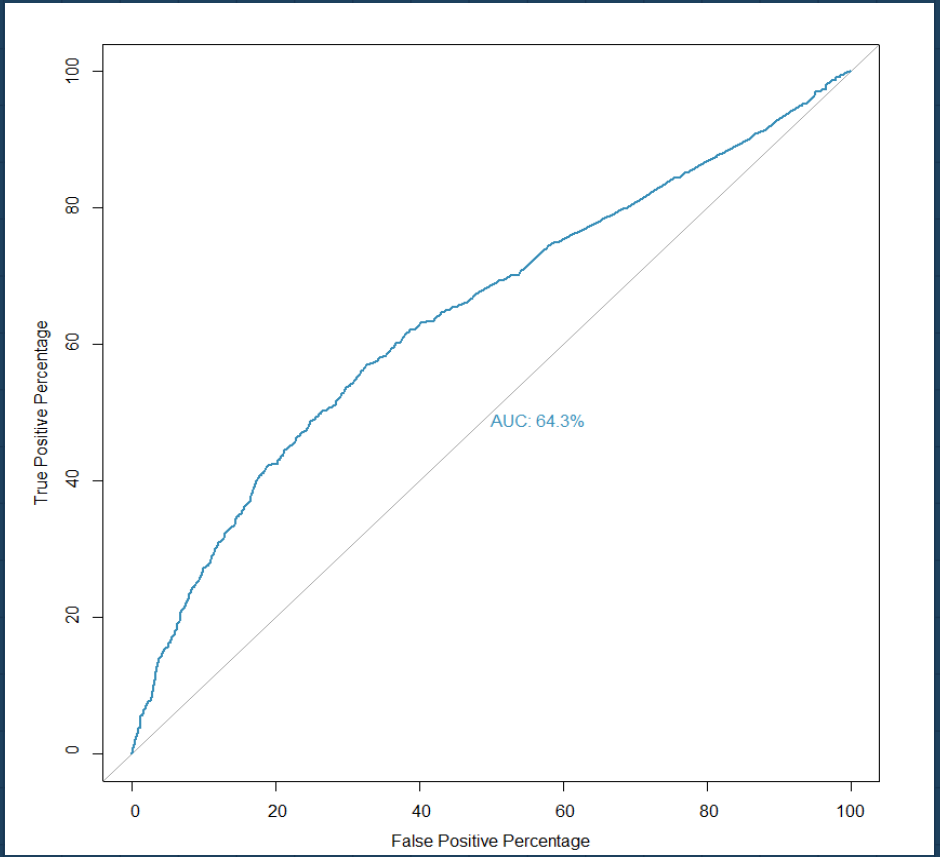
fatemeh.torabi@swansea.ac.uk





fatemeh.torabi@swansea.ac.uk

Naïve Bayes classifier

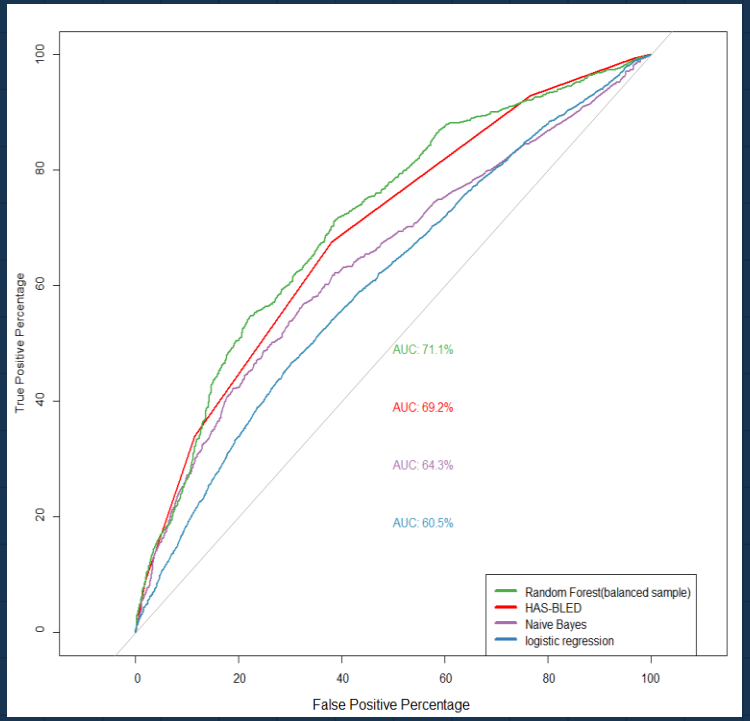


fatemeh.torabi@swansea.ac.uk

Conclusion and Discussion

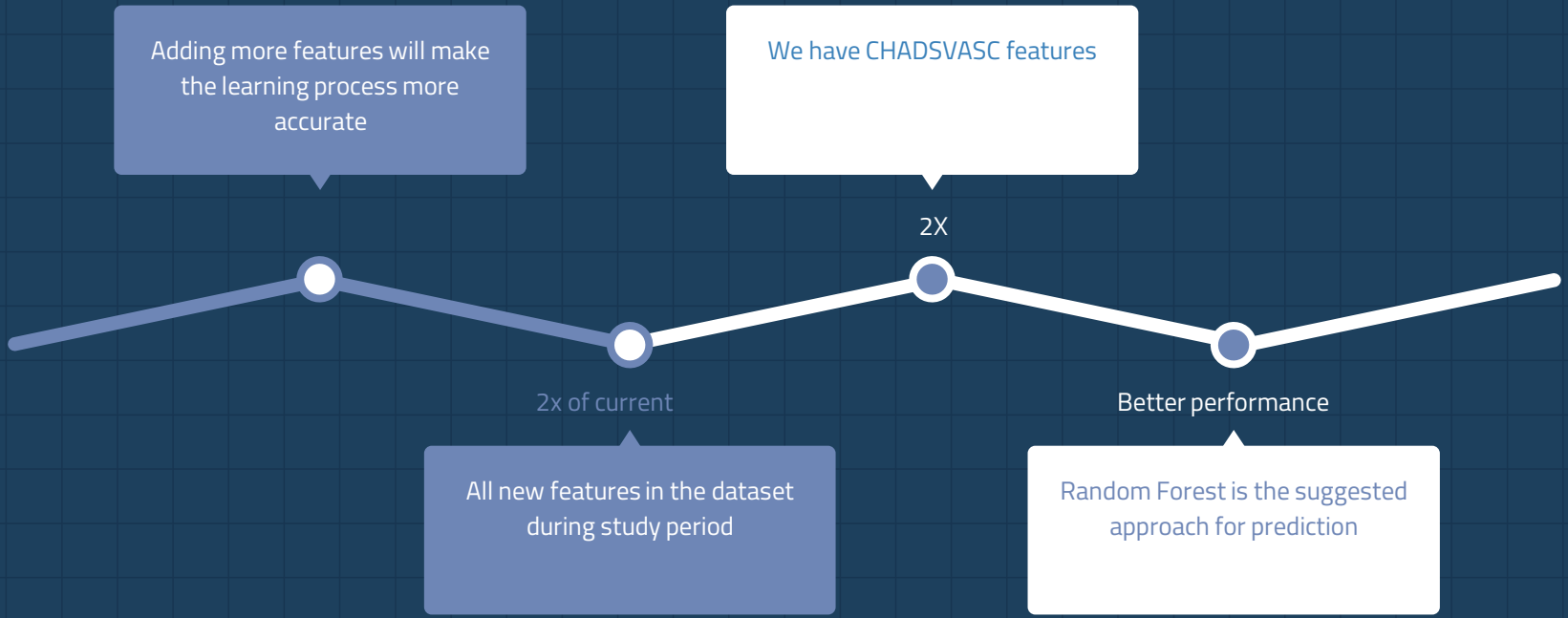
Model comparison

With same set of baseline and delta features, Random Forest performance is better than all other methods.



fatemeh.torabi@swansea.ac.uk

Model optimization



fatemeh.torabi@swansea.ac.uk

Strengths

- This study validated HASBLED on a very large cohort and showed there needs to be room for improvement in BLEED prediction.
- This work is transferable to other areas such as stroke prediction.

Limitations

- Use of limited number of features in models with regards to machine learning as well as only using the default parameters.

fatemeh.torabi@swansea.ac.uk

Future work

Improve machine learning methods: more models, additional features, feature selection, parameter tuning.

Extended follow-up on adverse outcomes: outcomes for those HASBLED gets wrong beyond a bleed. More deaths for those not identified? Ineffective use of resources for those falsely identified. Gaps in appropriate therapeutic options.

Applying these analysis to other areas.

fatemeh.torabi@swansea.ac.uk

Thank you!

Find the resources:
https://github.com/FatemehTorabi
Get in touch:
email: fatemeh.torabi@swansea.ac.uk
twitter: @Fatemetrb