# Using machine learning tools to investigate factors associated with trends in 'no-shows' in outpatient appointments

Eduard Incze, Penny Holborn, Gary Higgs [*], Andrew Ware

*Faculty of Computing, Engineering and Science, University of South Wales, Pontypridd, CF37 1DL, United Kingdom*

ARTICLE INFO

ABSTRACT

Missed appointments are estimated to cost the UK National Health Service (NHS) approximately £1 billion annually. Research that leads to a fuller understanding of the types of factors influencing spatial and temporal patterns of these so-called "Did-Not-Attends" (DNAs) is therefore timely. This research articulates the results of a study that uses machine learning approaches to investigate whether these factors are consistent across a range of medical specialities. A predictive model was used to determine the risk-increasing and risk-mitigating factors associated with missing appointments, which were then used to assign a risk score to patients on an appointment-by-appointment basis for each speciality. Results show that the best predictors of DNAs include the patient's age, appointment history, and the deprivation rank of their area of residence. Findings have been analysed at both a geographical and medical speciality level, and the factors associated with DNAs have been shown to differ in terms of both importance and association. This research has demonstrated how machine learning techniques have real value in informing future intervention policies related to DNAs that can help reduce the burden on the NHS and improve patient care and well-being.

## 1. Introduction

Previous research of trends in missed outpatient appointments has suggested that 'Did-Not-Attends' (DNAs) are a significant problem internationally, with evidence of DNA rates varying between specialities and locations with some specialities (e.g. psychiatry related appointments) having higher DNA rates than average (Lehmann et al., 2007; Mohamed et al., 2016). Previous studies have drawn attention to the economic consequences of no-shows (Kheirkhah et al., 2016). In Wales alone, for example, there were 1.5 million missed appointments in the five years between 2015 and 2019 with an estimated cost to the National Health Service (NHS) in Wales of approximately £240 million (BBC News, 2020). Hidden within national averages are regional variations between the seven Local Health Boards (LHBs) in Wales (http://www.wales.nhs.uk/nhswalesaboutus/structure). Missed appointments which averaged 7.7% across Wales for 2018/19, were closer to 10% for those LHBs serving the areas of highest population in the South Wales valleys and the capital city of Cardiff. The direct and indirect consequences of such missed appointments are significant. These include the waste of staff time and resources (each missed appointment costing the NHS approximately £120 (NHS England, 2018)), and potential health

impacts caused by the delay in treatment and longer waiting lists (which themselves produce a chain effect because longer waits for appointments have been linked to a higher rate of non-attendance (NHS England, 2018)). A knock-on effect is the increased demand in other sectors of the health service, including accident and emergency or out-of-hours services.

While the factors associated with outpatient DNAs have been examined in previous research, most studies to date have been concerned with examining the factors that affect missed appointments for specific specialities or within individual clinic/hospital settings. There are relatively very few examples of research that has looked at variations in non-attendance across specialities (Frankel et al., 1989; McCarthy et al., 2000). It follows that more research is needed to examine those factors affecting DNA rates for a wider variety of medical specialities that investigates the importance of both contextual (area-level) and compositional (patient-specific) influences. The main aims of this paper are to draw on a national database of non-attendance records to identify the types of factors that are important for predicting DNAs using machine learning (ML) techniques and to highlight trends in missing appointments across medical specialities. This study is part of a longer-term project which aims to implement an automated reminder

---

system that targets patients at risk of not attending their outpatient appointment in such a way that frees up staff time in pursuing follow-up tasks.

The findings presented here could also be seen as part of a wider contribution to NHS plans which aim to expand the use of digital technology by developing new systems to reduce the workload on staff and improve patient outcomes (NHS England, 2018). Systems such as 'DrDoctor' which have been developed in the UK have been shown to have lowered DNA rates by almost a third in some hospitals. This has been achieved by providing patients with routes to hospitals and the ability to manage upcoming appointments and by allowing NHS staff to send custom notifications, rescheduling appointments and setting booking confirmations (NHS England, 2018). This research contributes to such systems by providing a DNA risk score for each appointment aiding those staff charged with sending reminders where they are most needed by the most appropriate means.

The structure of the rest of this paper is as follows: in Section two, the findings from previous literature on the types of factors associated with non-attendance are summarised in order to identify current gaps in knowledge and to provide a wider rationale for this research. Section three presents the data and methods used within this research and draws attention to the potential for machine learning tools that go beyond the use of traditional approaches such as those based on logistic regression models to address these types of health applications. The findings from using such tools on a dataset extracted from a national database of outpatient non-attendances are presented in Section four. In the remaining parts of the paper, the policy relevance of this research is re-iterated in relation to the potential for such tools to add to the evidence base concerning patterns on 'no-shows' and help inform policies aimed at reducing the likelihood of non-attendance across medical specialities.

## 2. Literature review

### 2.1. Review of factors affecting non-attendance

Table 1 provides a summary of previous literature concerned with examining factors associated with trends in DNA rates. Dantas et al. (2018) provide a systematic literature review of the types of patient characteristics that have been associated with no show appointments for a range of medical specialities. In summary, whilst there are differences in findings between studies, partly arising from variations in the nature of the methodological approach adopted, the majority have drawn attention to the importance of factors such as the age and socioeconomic status of patients as characteristics associated with missed appointments (Campbell et al., 1991; Catz et al., 1999; Kruse et al., 2002; Lehmann et al., 2007). History of illness, presenting complaint, severity of illness and comorbidities have also all been shown to affect patient attendance (Carpenter et al., 1981; Catz et al., 1999; Chen, 1991; Frankel et al., 1989; Nel, 2014; Nicholson, 1994; Sparr et al., 1993; Stone et al., 1999). However, there is less agreement regarding the importance of factors such as educational level, race, or the distance from the patient's home to the clinic. Gender, for example, was found to be a factor in some studies (Junod Perron et al., 2010; Stubbs et al., 2012) but not others (Bush et al., 2014; Catz et al., 1999; Lehmann et al., 2007). Similar results can be found for variables such as the day of the week, season, time of day, new or follow-up appointment, employment status, and others. This could suggest that localised factors are potentially involved and/or the specialities themselves are affected by different factors. Previous research in the NHS in the UK seems to confirm both possibilities and suggests that "local systems are needed to address local problems" (Sharp and Hamilton, 2001, p.1081). Other meta-analyses have also found inconsistencies in results (Chen, 1991; Nicholson, 1994).

An interesting finding is that missed appointments are disproportionately distributed among patients, with a small number of patients accounting for a large percentage, or even all, of the total number of missed appointments (Norris et al., 2014). One study, for example,

**Table 1**
Previous Studies of factors influencing rates of missed appointments.

| Author(s)-Year | Medical Setting/country | Methods | Influencing factors |
|---|---|---|---|
| Campbell et al. (1991) | Psychiatry | Discriminant function analysis | Age, attendance history, mental illness history, rural/urban, routine appointment, distance, source of referral, staff type |
| Catz et al. (1999) | HIV (USA) | t-tests, correlation matrix, multiple regression, ANOVA | Age, minority status, religiosity, severity of disease, lower perceived social support, hopelessness, rural/urban |
| Kruse et al. (2002) | Serious mental illness (USA) | Pearson correlation, logistic regression | Age, minority status, lower perceived social support, having health insurance, not taking medicine |
| Sparr et al. (1993) | Psychiatry (USA) | Qualitative, | Initial appointment, PTSD, history of substance abuse, major depression, intensity of treatment, qualitative: forgetting, oversleeping, getting the date wrong |
| Carpenter et al. (1983) | Psychiatry, first appointment (USA) | Chi square, follow-up calls to ask why | Age, previous psychiatric treatment, source of referral, chief complaint, and number of days they waited for an appointment |
| Nicholson (1994) | Mental Health, (Canada) | Meta review | Age, waiting time, chief complaints, history of mental health contact, referral source, anxiety about outcome, appointment no longer being needed, |
| Chen (1991) | Mental Health (USA) | Meta review | Age, waiting time, socioeconomic status/class, referral source, education level, history of hospitalisation, clinical improvement during hospitalisation, length of stay, therapist continuity, days between discharge from hospital and appointment, denial of illness, perceived need of medication, high hopelessness-helplessness, history of substance abuse, distance to clinic, divorcees, minority status |
| Sharp and Hamilton (2001) | | Multivariable analysis; review | Sex, age, waiting time, deprivation, car ownership, phone ownership, unemployment; self-reported: forgetting, family or work commitments, longer at-the-door waiting times due to overbooking |
| Hamilton and Gourley (2002) | Antenatal care (UK) | Face-to-face surveys (patients and professionals) | Transport related (parking space, late buses, difficulties traveling with children, congestion, no bus stop close enough), too long a wait, admin. errors, childcare responsibilities, language barrier, low income |

**Table 1** (*continued*)

| Author(s)-Year | Medical Setting/country | Methods | Influencing factors |
|---|---|---|---|
| Bush et al. (2014) | Paediatric Urology (USA) | Chi-squared, ANOVA, multivariate logistic regression | Having government insurance, staff type, initial appointment, waiting time, clinic type, |
| Lehmann et al. (2007) | General medicine (Switzerland) | Chi-squared univariate analysis, Spearman rank-order correlations | Age, minority status, waiting time, date of birth earlier in the year, non-Europeans, common language with physician, follow-up appointment |
| Coodin et al. (2004) | Severe mental illness (USA) | Chi-square, independent groups *t*-test, Bonferroni correction | Age, drug use history, low level of community functioning |
| Stone et al. (1999) | Plastic surgery (UK) | Doctor questionnaire, patient questionnaire, chi-square | Appointment history, new or follow-up, hand trauma, excision of benign skin lesions, chronic disorders with little change, time since surgery greater than three months (after one missed appointment); patient self-reported: forgot, work-related, illness, lack of transport, poor weather, failure of being informed of appointment, appointment being cancelled |
| McCarthy et al. (2000) | Multi-speciality (UK) | Patient questionnaire, Pearson correlation, | Patient self-reported: being unable to attend due to illness or work-related reasons, and transport, being unwilling to attend due to long expected waiting time or improved health, waiting time, previous negative experience and fear, not receiving a clinic appointment, |
| Frankel et al. (1989) | Multi-speciality | Patient questionnaire | Short notice of appointment, seriousness of condition, age, sex (insignificant), relationship status, retiree status, difficulty attending (mainly due to work), not having enough information about the appointment, being on holiday, treatment unnecessary, illness, appointment being changed, history of non-attendance |
| Chua and Chow (2019) | Multi-speciality | Multiple logistic regression | age, race, nationality, speciality, appointment lead time, appointment month, appointment day of the week, referral source, previous visit type |

**Table 1** (*continued*)

| Author(s)-Year | Medical Setting/country | Methods | Influencing factors |
|---|---|---|---|
| | | | and previous visit status |

found that all the missed appointments in their sample were caused by only 19.4% of patients (Sparr et al., 1993). A history of missed appointments for the patient is a factor in future non-attendance (Campbell et al., 1991; Stone et al., 1999). A more recent study found that "less than 10% of patients who failed to attend twice, turned up for the third appointment" (Mohamed et al., 2016, p.3). Deprivation may also be a factor with some suggestions that nonattendance at NHS outpatient clinics is more common in deprived populations (Sharp and Hamilton, 2001).

One potential drawback in those studies conducted to date concerns the variations in the number of appointments and participants considered, with sample sizes varying from a few dozen to tens of thousands (Bush et al., 2014). Another limitation relates to the focus on individual specialities or institutions which limits the transferability of findings to other specialities or contexts. Most use data from a single clinic, but others involve a wider variety of settings; such as the study by Krishna and Amarjothi (2012) that includes 114 clinics drawing on a cohort that consists of 843 patients in mental health outpatient clinics. Furthermore, previous research has focused almost exclusively on a select few types of clinics, most commonly mental health, HIV, cancer, and paediatrics. To address such concerns, this paper builds on previous findings by drawing on a more extensive database of patients across a wider range of specialities to consider the types of factors associated with trends in DNA rates.

The approaches used within previous studies also vary but have tended to be based on what could be termed 'traditional statistical methods' (such as t-tests, correlation analysis, logistic regression, chi-squared, analysis of variance, Markovian models and others). These methods, while powerful and commonly used, may not be ideal for an in-depth analysis of a dataset numbering hundreds of thousands of records and more than 100 variables. With so many variables, complex non-linear interrelationships arise, which are both difficult to find and explain using classical analysis. The large number of nominal variables that are part of the dataset increase the complexity of both understanding the data and the predictive task in hand. For these reasons, our study promotes the use of machine learning approaches to investigate the causes of DNAs. In the next section, we place the use of such techniques in context by briefly describing how such tools have been used in health studies to date.

### 2.2. Overview of machine learning use in healthcare

The use of machine learning techniques has divided opinion among both clinicians and researchers due to perceived issues such as non-interpretability, security concerns and susceptibility to directed attacks (Mozaffari-Kermani et al., 2015). Despite these misgivings, machine learning is increasingly being used in a range of health applications (see, for example, Wiemken and Kelley, 2020). In medical imaging, for example, deep learning ML techniques are increasingly used for automated diagnosis thanks to "excellent accuracy in the interpretation and identification of patterns of disease in medical imaging" and lower costs for the patient (dos Santos and Carvalho, 2015, p. 15). Text mining is also being used for transcribing handwritten patient files into an electronic format, thus improving the use of electronic health records (Wang et al., 2012). The recent 'Artificial Intelligence in Healthcare' report commissioned by NHS Digital presents a cautious, but optimistic view on the continued development of ML in healthcare suggesting that "while AI in healthcare promises great benefits to

patients, it equally presents risks to patient safety, health equity and data security" (Academy of Royal Medical Colleges, 2019, p. 6).

Decision tree-based algorithms have a long history of use in healthcare (Koh and Tan, 2005), probably because they are easy to understand, are easily explainable with transparent decision making, and they are an established and accepted technology (Loh, 2014). Some examples include those studies where decision trees were used in drug discovery and outperformed neural networks and logistic regression (Muhammad et al., 2008; Obenshain, 2004). Decision tree-based methods have also been used in health geography research, such as a recent study of West Nile Virus where decision trees were used to identify complex relationships between data sets (Young et al., 2013). Their model's predictive ability was geographically inconsistent, possibly owing to unmodelled factors in the areas of poor prediction. Gradient boosted tree methods (defined in section 2.3) have gained popularity in healthcare due to their interpretability and increased performance compared to decision trees (Che et al., 2015). However, to our knowledge, few studies in health geography have, to date, used LightGBM, a type of gradient boosted model, utilised in this study. This is in stark contrast to its application in other disciplines/study areas such as bioinformatics (Chen et al., 2019), malware detection (Wang and Liu, 2020), acoustic scene classification (Stowell et al., 2015), cryptocurrency price forecasting (Sun et al., 2020) and others. LightGBM has several advantages over other techniques, including performance, interpretability, non-linearity, robustness to nominal and ordinal data, and ability to handle a large amount of data.

There has been some ML research into the causes of outpatient appointment DNAs; studies in Spain and the UK, for example, have found gradient boosting to be the best predictive model, outperforming both classical methods (such as logistic regression) and other ML techniques (Elvira et al., 2018; Nelson et al., 2019). The gradient boosting methods used were H2O.ai and XGBoost, respectively. Dashtban and Li (2019) also found ML methods to be superior to logistic regression, although their method of choice was a deep neural network. However, such studies have tended not to include sociodemographic data; for example, Nelson et al. (2019) did not include any age, sex, or employment variables in their study. In addition, unlike the approach taken in this study, differences between specialities were not sought or reported.

### 2.3. Review of machine learning methods used in this study

For the analysis described in this paper, Python 3 was used with data manipulation and machine learning libraries. The predictive algorithm chosen was Microsoft's LightGBM, drawing on consistently good results in a variety of research studies and competitions especially in binary classification tasks (Anghel et al., 2018; Ma et al., 2018). This is a 'gradient boosting decision tree' method introduced in 2017 (Ke et al., 2017). Linear models were also considered, but ultimately rejected because, as it is often the case with such large data, when "there is a highly non-linear and complex relationship between the features and the response,. . . Decision trees may outperform classical approaches" (James et al., 2013, p. 314). Decision trees on their own, however, are often prone to overfitting and do not generalise well. Due to this and other shortcomings, they are rarely used individually (Bramer, 2007). Since our dataset features dozens of variables, resistance to overfitting was a major consideration, providing another reason for choosing a gradient boosting algorithm. LightGBM specifically compares favourably to other gradient boosting algorithms, such as XGBoost (Anghel et al., 2018; Ke et al., 2017; Ma et al., 2018).

The idea of boosting (using multiple weak models to approximate a robust model) was first introduced in the eighties (Kearns and Valiant, 1989) and confirmed in the early nineties in a landmark paper for machine learning and statistics (Schapire, 1990). Gradient boosting, hypothesised by Breiman and first constructed by Friedman, builds upon the idea of boosting, by introducing adaptive reweighing of the parameters (Breiman, 1998; Friedman, 1999a, 1999b). Bagging (bootstrap

aggregation) was also used in conjunction with deep (unpruned) trees, to reduce the variance, leading to better generalisation. Through bagging, each decision tree is trained on a subset of the dataset, which leads to higher individual variance but lower bias (James et al., 2013). In summary, this overview points to some successful applications of machine learning within healthcare research and more specifically draws attention to the potential use of LightGBM for this study.

## 3. Data and methods

### 3.1. Data sources

The principal source of the data for this research is the NHS Wales Informatics Service (NWIS) National Outpatient Appointment Database. For this analysis, every patient present within the data between January 1, 2018 and December 31, 2018 was selected. To account for repeated measures in the data, only the latest appointment per patient within the time-period was kept. This timeframe was selected because it is the most recently available data and includes every outpatient appointment scheduled in Wales during the selected timeframe, for every available medical speciality. As such, this data is not a sample, but the entire population. The data includes information on 1,011,897 unique individuals, 561,393 females and 450,504 males. The mean age is 49 years old while the median is 53 years old, providing evidence of negative skewness. The DNA rate averaged by individual is 8.6%, although this greatly varies by Health Board and speciality.

Deprivation measures have also been included because they have previously been found to have an impact on non-attendance rates (Sharp and Hamilton, 2001). The Welsh Index of Deprivation (WIMD) Rank (for 2014) and the Townsend Index (for 2011) are both measures of deprivation, but they are calculated differently. WIMD is both an area-based measure and a measure of relative deprivation, since it ranks all Lower Layer Super Output Areas (LSOAs, approx. 1500 people) based on eight deprivation domains: income, employment, health, education, access to services, community safety, physical environment and housing. These domains are themselves calculated from combining underlying variables (Statistics For Wales, 2014). The Townsend Index, on the other hand, uses four census variables to calculate a score for each area (Output Area (OA, approx. 300) or LSOA) namely unemployment, non-car ownership, non-home ownership, and overcrowding (Townsend, 1988; Yousaf and Bonsall, 2017). More deprived areas have higher Townsend scores and are ranked higher on the WIMD, with respect to their method of measuring deprivation. Both WIMD and Townsend were included in the dataset in order to compare their use as contextual variables in an analysis of DNA rates. Both the WIMD rank and Townsend Index were joined to each patient record by LSOA code.

### 3.2. Data pre-processing

As expected for such a dataset, a degree of pre-processing was required. Some records had invalid or impossible values; these were removed entirely. There were approximately 3.2% of records where the waiting time (measured in days between the date of referral and date of appointment) was zero or negative. Of these, the majority were same-day appointments, which are not relevant to this research, for two reasons: firstly, the motives for non-attendance are likely to be different to the usual (forgetting, no longer needing the appointment, not considering it a priority, etc.) and secondly, reminders are unlikely to be sent in time in these cases. As such, all these records were also removed. Another 1% of records were removed due to missing the referral date entirely. As a result of these pre-processing steps, the final count of data numbered 976,562 records.

Where possible, new variables were derived from existing data; an example being 'appointment history scores' which were calculated individually using each patient's history of attendance. This measure of attendance is designed to give more recent appointments a higher

impact, by including a time factor in the calculation of the formula which scores patients on their previous attendance (detailed in Equation (1)). The Charlson Comorbidity Index was also calculated for each patient; this is a "weighted index that takes into account the number and the seriousness of comorbid disease" (Charlson et al., 1987, p. 373). Other computed variables include road distance/drive time to hospital based on the patient's and hospital's postcodes, attendance outcome of previous appointment (attended, non-attended, cancellation), hospital attendance rates (by type), days since last appointment, days since last DNA, and 'waiting time' calculated as the difference between the appointment and referral dates.

The appointment history used in this research is unlike other similar measures, both because it considers all three attendance outcomes and includes an adaptive weighting based on date. This is an original equation, obtained through experimentation and based on the principle of recent appointment attendance having a higher impact on future attendance than older appointment attendance. Each of the three factors (attendances, DNAs, cancellations) were computed individually and stored as separate variables. The equation used to calculate each appointment's score can be seen in Equation (1). It is calculated for all appointments, five years prior to each appointment in the dataset. To calculate all the appointment scores, over thirty million records were queried.

$$A_{score} = \sum \frac{1_A}{\sqrt{(Currentdate - Pastdate)}} \tag{1}$$

### 3.3. Exploratory data analysis

Initial exploratory analysis highlighted that some numerical variables were highly correlated, such as the 'percentage of the population with long term health problems and disabilities' (LTHPD) and 'health level' (self-reported). Due to correlations with other variables being higher for LTHPD (thus lower overall information gain), that variable was removed from the dataset. Similarly, date-related variables such as

'day of the year', 'week of the year' and 'month of the year' were highly correlated; thus, all but the day of the year were removed. Priority type (routine or urgent) was also correlated with attendance type (new or follow-up). Some variables had a high percentage of missing values; 'days since last DNA' had 56.7% of its values missing, because no DNAs were found in the lookback period of five years. All those appointments had a DNA score of 0. Driving distance also could not be computed for 22% of appointments; however, it is present for a large proportion of the data. The majority of patients had a Charlson Comorbidity index of 0, indicating no comorbidities; however, it still proved to be an important variable, especially for some specialities.

The distribution of the Townsend scores was compared to the distributions of DNA rates. For visualisation, we have chosen to include local-level analysis for one Local Health Board; the Cwm Taf Morgannwg University Health Board (CTM) area in South Wales which has the highest DNA rate in Wales. CTM also includes a diverse spread of socioeconomic circumstances, from some of the most deprived LSOAs to some of most affluent in Wales. As can be seen in Fig. 1, there is a large overlap between areas with high Townsend scores (A) and high DNA rates in general (B); however, the opposite can be seen as well, and DNAs for some specialities show an inverse relationship with Townsend scores, (such as 'Old Age Psychiatry', Fig. 1(D)). Of the 1,015,165 patients, 175,247 (17%) had at least one non-attendance within the timeframe, adding up to 248,905 non-attendances in total. Based on this, while most patients attend all their appointments, patients who missed at least one appointment within this timeframe, missed 1.42 appointments on average. This figure varies by individual LHB within the region. In Cwm Taf Morgannwg University Health Board, this rate rises to 1.54, suggesting a larger repeat non-attendance problem whereas Powys Teaching LHB had the lowest rate of 1.28. This finding concurs with the work of others that suggests that missed appointments lead to more missed appointments (Campbell et al., 1991; Stone et al., 1999).
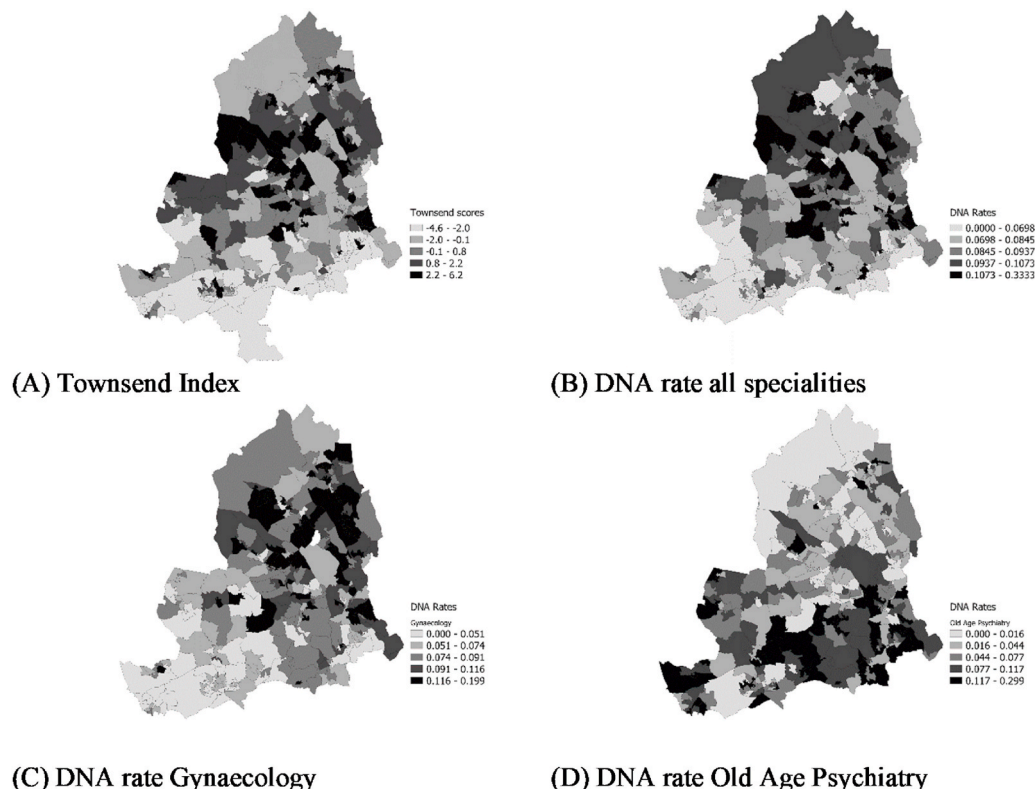


(A) Townsend Index

(B) DNA rate all specialities

(C) DNA rate Gynaecology

(D) DNA rate Old Age Psychiatry

**Fig. 1.** Relationship between the Townsend Index and DNA rates by speciality for CTM.

### 3.4. Analytical methods

#### 3.4.1. Logistic regression

Logistic regression is one of the simplest and most used methods for two-class classification. It is easy to implement and often used, as is the case here, as a baseline for a binary classification problem. Logistic regression was performed using the scikit-learn API in Python. The numerical data was min-max normalised by subtracting the minimum value and dividing by the range for each numeric column. The result is that each numerical factor is rescaled to the range [0,1]. [0, 1]Normalization is necessary for logistic regression to perform well but comes with a disadvantage in interpretability of the model which is discussed in the Results section. Scikit-learn uses L2 penalty by default with the 'lbfgs' solver, however using 'saga' solver, an elastic net penalty (L1+L2) was applied (Zou and Hastie, 2005).

A significant disadvantage of logistic regression is that it is greatly affected by colinear variables, being a linear method. Another disadvantage is that categorical variables must be either one-hot-encoded, a process by which categorical variables of k distinct values are split into k-1 binary variables (resulting in increased complexity) or factorised, which means they are treated as if they were linear variables. Logistic regression also does not handle null values, leading to information loss. To address the limitation of logistic regression and drawing on literature that shows the potential for applying machine learning techniques in health geography, the next section outlines the use of LightGBM in the analysis of DNAs.

#### 3.4.2. LightGBM

The machine learning method used in this study is the gradient boosted trees method 'LightGBM'. Being a tree-based method, collinear variables do not have a significant effect on the results. 'Gradient boosted' means that the predictive model is formed iteratively from an ensemble of weaker models (in this case, decision trees), optimised by a gradient descent function. Boosting is applied in the following way: after each weak model is trained, their performance is evaluated, with inputs which previous models failed to correctly classify being weighed higher and successfully predicted inputs being weighted lower. This process changes the sample distribution at each iteration by focusing subsequent models on misclassified data. Finally, the method for making predictions using boosting in classification is to have each weak model assign a class, and their choices are aggregated in some form (majority class, average class, weighted average, etc.). The weighted average is commonly used, with weaker models being typically assigned a lower weight.

Gradient boosting works by successively fitting each model to the gradient of the previous model's loss function, with the aim of maximising the error reduction at each step. Effectively, this is a type of functional gradient descent. Importantly, it does not change the sample distribution, unlike traditional boosting, but still emphasises learning from mistakes. Thus, LightGBM trains decision trees iteratively, one tree at a time, with each tree taking some of the previous output into account.

#### 3.4.3. Implementation

Machine learning models require splitting the data into training and testing sets, to avoid data leakage. The portion of the data used for setting the parameters of a model ('training the model') is called the training set, and usually contains 80% of the data, while the other 20% is used as a hold-out set for testing. The hold-out set contains data unseen by the model, and it is used in order to ensure that the results are reproducible. A validation set was also utilised, serving as an intermediary between the training and test sets. When the dataset was split into training, testing and validation sets, a multi-stage sampling process formed of stratified splits was performed based on DNA rates and patient IDs, to ensure that the datasets are representative and that there is no information leakage. Due to the large amount of data available, a model was trained on 80% of records, while validation was performed on 10%

of records. Finally, testing was carried out on a hold-out set of 10% of records. A model with the same parameters was then trained and tested on subsets of the data containing each speciality, and results were compared to the original.

Models were fine-tuned using a dual approach of random search, to find promising values or ranges for hyperparameters, followed by grid search with cross-validation, to home in on the best performing hyperparameters. Parameter ranges for both logistic regression and LightGBM can be found in Table 2. The metrics used to score the model's performance were precision (PPV), sensitivity and F1 score (harmonic mean of precision and sensitivity). Due to the class imbalance present in the data, attendances outnumber DNAs at a ratio of 10:1; two often-used metrics, accuracy and ROC-AUC, were not used for model selection because they are not appropriate in an imbalanced class situation (Davis and Goadrich, 2006; Saito and Rehmsmeier, 2015). The model outputs the probabilities of each appointment belonging to one of the two classes (attendance or non-attendance). While by default, most models set 50% as the threshold for separating the predictions into classes, in imbalanced datasets this is often not the correct choice (Saito and Rehmsmeier, 2015). The precision-sensitivity plot was used to determine the threshold that maximises the F1 score.

While the results are discussed for a scoring threshold that maximises the resulting F1 score, this threshold can be modified to favour either precision or sensitivity, depending on which metric is deemed more important. A cost-benefit analysis can reveal what is costlier: mislabelling a missed appointment or an attended one. The SHAP (SHapley Additive exPlanations) library was used to calculate the relative importance of each variable. Each variable is assigned a 'shap value' which equates to the average change in magnitude of the model's output when that variable is hidden, or in other words, the change in log odds. SHAP is a unified model, combining some previously existing methods in a novel way, using game theory principles. It is more accurate and consistent than traditional log loss or Gini Impurity (Lundberg and Lee, 2017). Dot plots, violin plots and individual importance plots drawn using SHAP are analysed in Section 4. Violin plots provide much more information than traditional box plots as they can highlight both the distribution of each variable and the importance of it within the model, respective to its distribution. These plots will be compared to visually show the differences between the use of logistic regression and LightGBM in our analysis.

## 4. Results

### 4.1. Using classical methods (logistic regression)

The results of logistic regression were poor, both in terms of metrics and information gain. On the test set, an F1 score of 0.31 was achieved.

**Table 2**
Parameter tuning for logistic regression and LightGBM.

| Logistic regression | | LightGBM | |
|---|---|---|---|
| Parameter | Value/Range | Parameter | Value/Range |
| Solver | Saga | bagging_freq | 1 |
| Penalty | Elasticnet | learning_rate | 0.05 |
| class_weight | Balanced | objective | Binary |
| max_iter | ∞ | metrics | f1_score |
| Tol | {0.001, 0.003, 0.01, 0.03} | num_iterations | ∞ |
| C | [0,001, 100] | early_stopping_rounds | 50 |
| l1_ratio | [0, 1] | scale_pos_weight | [1,10] |
| | | num_leaves | [7,4095] |
| | | max_depth | [-1, 63] |
| | | bagging_fraction | [0.4, 1] |
| | | colsample_bytree | [0.4, 1] |
| | | lambda_l1 | [0,1] |
| | | lambda_l2 | [0,1] |

This F1 score was calculated from a recall of 0.37 and precision of 0.27. Therefore only 37% of DNAs were successfully identified, and 27% of predicted DNAs were correct. Factor importance calculated using SHAP can be seen in Fig. 2(A) for the top 20 variables in terms of model importance. In the case of logistic regression, the results are somewhat misleading, for example looking at 'DaysSinceLastDNA', the logistic regression model finds a linear relationship with DNAs, but LightGBM shows that the true relationship is more complex (Fig. 2(B)). The inadequacy of logistic regression for predicting DNAs becomes even more evident in Fig. 3 which shows a linear relationship between the patient's age and DNA rate, which exploratory data analysis has shown to be overly simplified and misleading. More in-depth results, such as factors by speciality and LHB, were not computed using logistic regression due to the poor performance and as discussed, the potential for misleading results.

### 4.2. Using machine learning

The results of the LightGBM model tested on the test set can be seen in Table 3, both for Wales overall and for the 7 Local Health Boards. The importance of the most important twenty factors, as calculated using SHAP values, can be found in Table 4. To illustrate the principles only the general case and three selected specialities are presented in this table but it is possible to extend this analysis to consider more specialities and this is returned to in the Discussion section of the paper as an avenue for further research. Nationally, the most important individual factors for a model trained and tested on all specialities were found to be the patient's age, appointment history (attendance score and DNA score), treatment site, speciality, waiting time and Townsend Index. Age is a variable very commonly found amongst the other studies, and it was expected to be an important factor. The results for age are as expected: younger people miss more appointments. The other variables are more difficult to compare because they are either rare (appointment history, waiting time) or completely missing from the literature (Charlson Comorbidity Index).

The subsequent models were trained and tested on distinct specialities and yielded better results for some specialities and worse for others,

as shown for the 20 most populated specialities in Table 5. The fact that the ability to predict DNAs changes when the model is trained exclusively on them, indicates that the specialities differ in the factors affecting their attendance. Specifically, for those specialities where the performance dropped, it suggests that attendance is affected by factors that are not present in the data. Furthermore, the factor importance charts in Fig. 4 and results in Table 4 reveal that the same factors can affect specialities differently, i.e. although the Townsend Index is an important factor in most specialities, the way that its variation affects attendance differs between specialities.

For Fig. 4, in the general case (a), the SHAP value forms a positive linear relationship with the Townsend Index, thus implying that more deprived areas correlate with a higher DNA rate. Neither the linearity nor the strictly positive relationship holds on a speciality-by-speciality basis, with some having a drastically different shape. For example, Adult Mental Illness (b) presents a relatively flat curve with lower values being more impactful. While both Dermatology (c) and Neurology (d) show a positive relationship, Dermatology appears to approximate logistic growth, while Neurology looks logarithmic. For 'Adult Mental Illness' (AMI) and Cardiology, the best factor in predicting DNAs is the hospital where the appointment is scheduled. Differences in the other factors can also easily be observed. For the Clinical Haematology speciality, the most important factors were previous attendance history and age, followed by waiting time. In contrast, for the 'Geriatric Medicine' speciality, the top three factors were: previous attendance history, treatment site, and ambulance score. Geriatric Medicine has the highest importance for ambulance score of any speciality, and one of the lowest for age, tied with Paediatrics.

## 5. Discussion

A key strength of this study is the availability of a rich source of national data, allowing a large-scale approach to the problem. This data was used to derive new statistics, such as the appointment history scores, which have proven to be very useful. While most studies focus on a single speciality, or a single hospital, this study has included records from every medical speciality for an entire country (Wales). This,
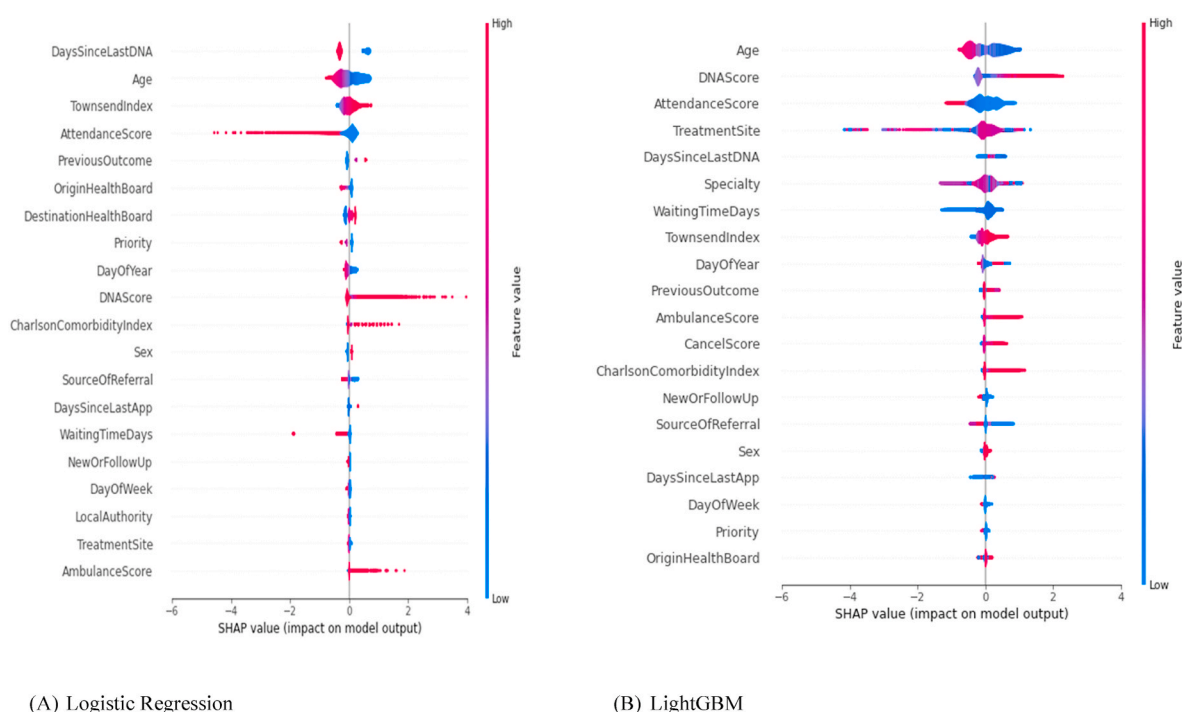


(A) Logistic Regression

(B) LightGBM

**Fig. 2.** A comparison of factor importance for Wales.

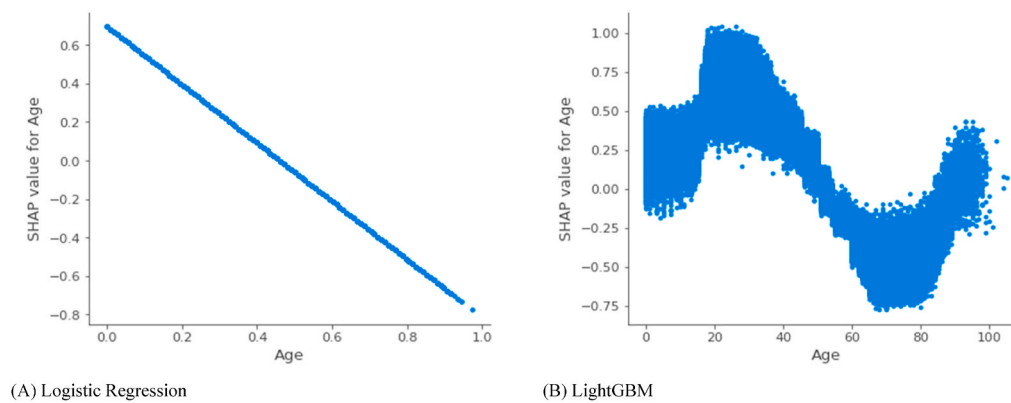(A) Logistic Regression                                      (B) LightGBM

**Fig. 3.** A comparison of the factor importance of Age for Wales.

**Table 3**
Performance metrics for LightGBM by Local Health Board.

|  | Overall (Wales) | Betsi Cadwaladr University | Cardiff and Vale University | Cwm Taf Morgannwg University | Hywel Dda University | Powys Teaching | Swansea Bay University | Aneurin Bevan University |
|---|---|---|---|---|---|---|---|---|
| F1 score | 0.37 | 0.36 | 0.4 | 0.37 | 0.35 | 0.35 | 0.37 | 0.35 |
| Precision | 0.33 | 0.28 | 0.29 | 0.3 | 0.29 | 0.31 | 0.31 | 0.31 |
| Recall | 0.42 | 0.5 | 0.61 | 0.48 | 0.42 | 0.4 | 0.44 | 0.4 |

**Table 4**
Factor importance for Wales and selected specialities.

| All specialities | | Adult mental illness | | Dermatology | | Neurology | |
|---|---|---|---|---|---|---|---|
| Factor | SHAP importance | Factor | SHAP importance | Factor | SHAP importance | Factor | SHAP importance |
| Age | 0.38 | Treatment site | 0.35 | Age | 0.51 | Age | 0.35 |
| DNA score | 0.31 | DNA score | 0.27 | Waiting time | 0.29 | DNA score | 0.30 |
| Attendance score | 0.26 | Attendance score | 0.20 | Treatment site | 0.27 | Attendance score | 0.29 |
| Treatment site | 0.20 | Age | 0.16 | Attendance score | 0.24 | Days since last DNA | 0.23 |
| Days since last DNA | 0.19 | Days since last DNA | 0.12 | DNA score | 0.21 | Townsend Index at LSOA | 0.19 |
| Speciality | 0.17 | Waiting time | 0.11 | Days since last DNA | 0.18 | Waiting time | 0.13 |
| Waiting time | 0.16 | Destination Health Board | 0.07 | Townsend Index at LSOA | 0.14 | Destination Health Board | 0.13 |
| Townsend Index at LSOA | 0.14 | Day of year | 0.06 | Day of year | 0.14 | Day of year | 0.10 |
| Day of year | 0.09 | Previous outcome | 0.03 | Days since last app | 0.10 | Treatment site | 0.09 |
| Previous outcome | 0.08 | Days since last app | 0.03 | Previous outcome | 0.09 | Sex | 0.08 |
| Ambulance score | 0.07 | Townsend Index at LSOA | 0.03 | Cancel score | 0.05 | Drive distance (miles) | 0.08 |
| Cancel score | 0.07 | Source of referral | 0.02 | Day of week | 0.04 | Days since last app | 0.07 |
| Charlson Comorbidity Index | 0.05 | Local authority | 0.02 | Charlson Comorbidity Index | 0.04 | Previous outcome | 0.07 |
| New or follow-up | 0.05 | Cancel score | 0.01 | Ambulance score | 0.04 | Source of referral | 0.06 |
| Source of referral | 0.04 | Origin Health Board | 0.01 | Drive distance (miles) | 0.03 | Day of week | 0.06 |
| Sex | 0.04 | Priority | 0.01 | Local authority | 0.02 | Ambulance score | 0.06 |
| Days since last app | 0.03 | Day of week | 0.01 | Priority | 0.02 | Local authority | 0.05 |
| Day of week | 0.02 | Drive distance (miles) | 0.00 | Destination Health Board | 0.01 | New or follow-up | 0.05 |
| Priority | 0.02 | Ambulance score | 0.00 | Origin Health Board | 0.01 | Priority | 0.03 |
| Origin Health Board | 0.02 | Charlson Comorbidity Index | 0.00 | Sex | 0.01 | Origin Health Board | 0.03 |

combined with a sophisticated machine learning approach, has allowed key insights to be drawn on the differences between locations and specialities in terms of non-attendance. Results obtained through LightGBM show that the impact of contextual factors such as area-level deprivation on DNA rates is dependent on the speciality under consideration. Such an approach has real value in informing future targeted mechanisms, as it is shown that results vary by area and speciality. For example, for specialities where deprivation was found to be an important factor, incentives such as compensated bus tickets could be trialled. Special care

should be afforded both to understanding why deprivation measures are more important for some specialities than others, and why in specialities where it is important, the relationship between DNA rates and deprivation is different.

One very commonly used method to alleviate the effects of DNAs is overbooking, that is booking more slots than are available to capitalise on missed appointments and keep staff utilisation high. Arguments against overbooking have been made in several studies, with concerns mainly relating to patient satisfaction and the additional stress placed on

**Table 5**
F1 scores by speciality for the 20 most populated specialities.

| Speciality | F1 score |
|---|---|
| Adult mental illness | 0.46 |
| Cardiology | 0.34 |
| Clinical Haematology | 0.36 |
| Dermatology | 0.38 |
| Endocrinology | 0.46 |
| ENT | 0.32 |
| Gastroenterology | 0.37 |
| General medicine | 0.51 |
| General surgery | 0.33 |
| Geriatric medicine | 0.44 |
| Gynaecology | 0.32 |
| Neurology | 0.42 |
| Obstetrics | 0.33 |
| Ophtalmology | 0.37 |
| Oral surgery | 0.46 |
| Paediatrics | 0.40 |
| Respiratory medicine | 0.39 |
| Rheumatology | 0.31 |
| Trauma and Orthopaedics | 0.37 |
| Urology | 0.33 |

staff by this practice (Bech, 2005; Moore et al., 2001; Hamilton et al., 1999). This research could be used to directly assess the proportion of appointments to be overbooked in each clinic, each day, and even the most appropriate times of day for overbooking. The impact of this system, if used correctly, could vastly improve both the waiting times, increasing patient satisfaction, and minimise the occurrence of staff overtime due to excessive overbooking.

The factors found to be relevant in predicting non-attendances match existing literature in some respects, with some of the most important ones overall being patient attendance history, waiting time, and age. Other factors that were found influential include: Townsend Index, day of the year, drive time and the Charlson Comorbidity Index of the patient. These factors can be grouped into three major categories: compositional (household and individual level), contextual (community, OA level) and clinical (hospital, appointment). By summing the SHAP importance measure for these groups, it has been found that compositional factors account for most of the SHAP importance. The fact that the performance of the model dropped for some specialities when analysed on their own, shows that the current data does not model those specialities adequately. There must be other factors that influence the DNA rates for those specialities, which are not included in the data. Rheumatology is an example of such a speciality, where an F1 score of 0.31 was reached.

Performance varied by area as well, F1 score ranging from 0.4 to 0.61. The results might be affected by the diversity of areas included in the dataset as reflected in their deprivation ranking. Another possible factor is data quality. For example, almost half (47%) of the cases where a waiting period could not be calculated due to the absence of a referral date, were in one LHB (CTM). The most import area factor, in general, was the Townsend Index. Other demographics failed to make a significant impact. No difference was found between patients living in urban or rural areas. In the data collected, compositional factors were found to be more important than contextual factors, but deprivation was consistently found to be in the top 10 most important factors, for every speciality. When looking at clinic-related variables, the most important factor is the medical speciality. The next most important is the source of referral, followed by days between when the appointment was requested, and the appointment was scheduled to take place. Hospital type was
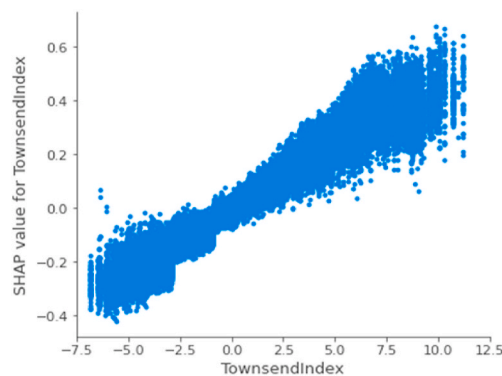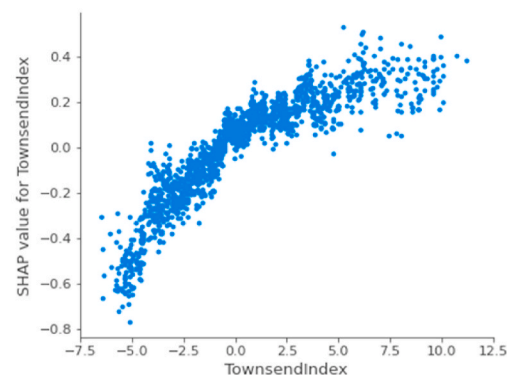


(A) All specialities
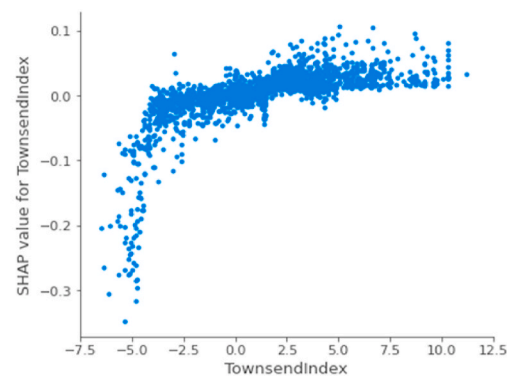
(B) Adult Mental Illness
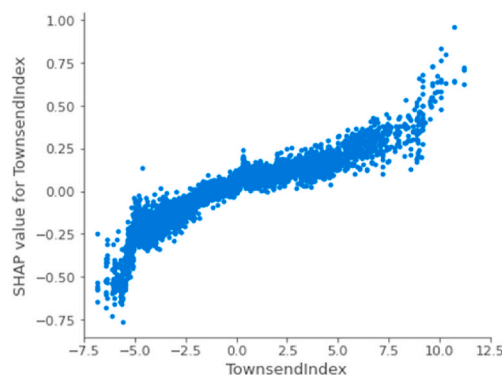
(C) Dermatology

(D) Neurology

**Fig. 4.** A comparison of the factor importance of the Townsend Index by speciality for Wales.

found to be much more important than the exact hospital where the appointment was scheduled.

A comparison of different deprivation measures was also performed, with the result that the Townsend Index generally outperforms the WIMD, with some exceptions such as Paediatrics and Child and Adolescent Psychiatry. These differences can be a result of the different domains that form the two deprivation measures. WIMD is a more complex measure formed from indices grouped into domains. A limitation of the study is the absence of other variables which could be influencing DNA rates; particularly the compositional variables such as minority status, religion, etc. Future research will aim to improve on the ability to predict DNAs by including new data and improve the model by including factors such as weather conditions, unusual events or the physical location of hospitals. This should help inform different types of interventions aimed at increasing patient attendance in different regional contexts and for a range of medical specialities.

For brevity, only the results of a few key specialities are illustrated here; future research could involve a wider investigation that includes a consideration of more specialities as part of a fuller exploration of factors leading to DNAs. Further research will also include a study on the most efficient reminders for the patient groups identified in this paper, focusing on those considered at higher risk of non-attendance. Initial hypotheses are that electronic reminders would achieve a better response with groups belonging to the younger demographic, or those living in urban areas, while traditional forms of contact might work best for the older individuals or those living in rural areas. However, more research is needed to explore the potential for alternative intervention policies in these contexts. The dataset includes no individual-level information on the patients' ethnic background, religiosity, socioeconomic status, employment status, disability status, education level, car ownership, and history of mental illness, all of which have been shown in some previous studies to have a measurable impact on non-attendance rates (Dantas et al., 2018). While there is data on patients' ethnic origin in the National Database at NWIS, it is very sparse with two-thirds of the records being 'Not stated'. As such, it was not included in the analysis. Although census data was used as a proxy for many of these factors, the deprivation rank is based on 2011 Census data (Office for National Statistics, 2011) complimented by more up-to-date 2014 data that includes a wider range of data sources; thus, it may not be completely representative of the current situation. While the newest WIMD edition had already been released in late 2019, it was not considered for this study because of a change in methodology whose impact has not yet been evaluated on our dataset. Finally, in this study we have used advanced quantitative methods involving the analysis of a national database; deeper insights could be gained by collecting primary data and by incorporating qualitative approaches to investigate the perceived reasons for non-attendance, both among staff and patients, in contrasting geographical or socioeconomic contexts.

## 6. Conclusions

Missed appointments continue to cost health organisations in many countries a considerable amount of money and may have detrimental social and health implications for individual patients. The main contribution of the paper is that it provides a greater understanding of the underlying reasons and influences on patterns of DNAs that have often been beyond the scope of previous studies. The results from this analysis can help inform the use of various policy interventions aimed at increasing the attendance rate. In particular, the study builds on an expanding literature on the advantages of machine learning approaches in healthcare (discussed in Section 2.2) and draws on the increasing amount of data becoming available through data-sharing initiatives in the health sector. To the best of our knowledge, this is the first study to use these particular machine learning techniques in an outpatient appointment context. The choice of model allows for highly non-linear relationships to be discovered between variables and highlights the

limitations of common classical approaches.

It has been shown that there are differences in those factors associated with missing outpatient appointments for different medical specialities which will have important implications for informing interventions developed on a speciality-by-speciality basis. An important addition of this study has been the analysis of both contextual and compositional factors potentially associated with DNA rates. Wider contextual variables are often omitted from studies concerned with establishing those factors associated with non-attendance. Our findings have shown that whilst for some specialities, such rates are influenced predominantly by compositional factors, others are affected more by contextual or clinical factors worthy of further investigation in follow-up research. In particular, geographical variations highlighted in the analysis suggest that further research could look to enhance the dataset by examining the influence of variables such as weather incidence, local transport conditions or the impact of other significant events at the scheduled appointment time and place.

## Declarations of interest

The authors declare no competing conflicts of interest.

## Data availability

The secondary dataset used in this study was anonymised and analysed under strict ethical and data confidentiality protocols. Appropriate ethical procedures were applied to source the data from its providers.

## Code availability

The code used in this study is not made publicly available but may be provided by the authors on request.

## Acknowledgements

## References

Academy of Royal Medical Colleges, 2019. Artificial Intelligence in Healthcare, 14 November 2019. https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/.

Anghel, A., Papandreou, N., Parnell, T., De Palma, A., Pozidis, H., 2018. Benchmarking And Optimization Of Gradient Boosting Decision Tree Algorithms, Workshop on Systems for ML and Open. Source Software at NeurIPS 2018, Montreal, Canada, 3 February 2020. https://arxiv.org/abs/1809.04559.

Bech, M., 2005. The economics of non-attendance and the expected effect of charging a fine on non-attendees. Health Pol. 74 (2), 181–191.

Bramer, M., 2007. Principles of Data Mining. Springer-Verlag London Limited, London.

Breiman, L., 1998. Arcing the edge. Ann. Probab. 26, 1683–1702, 10 January 2019. https://statistics.berkeley.edu/tech-reports/486.

Bush, R.A., Vemulakonda, V.M., Corbett, S.T., Chiang, G.J., 2014. Can we predict a national profile of non-attendance paediatric urology patients: a multi-institutional electronic health record study. J. Innovat. Health Inf. 21 (3), 132–138.

Campbell, B., Staley, D., Matas, M., 1991. Who misses appointments? An empirical analysis. Can. J. Psychiatr. 36 (3), 223–225.

Carpenter, P.J., Morrow, G.R., del Gaudio, A.C., Ritzler, B.A., 1981. Who keeps the first outpatient appointment? Am. J. Psychiatr. 138 (1), 102–105.

Catz, S.L., Mcclure, J.B., Jones, G.N., Brantley, P.J., 1999. Predictors of outpatient medical appointment attendance among persons with HIV. AIDS Care 11 (3), 361–373.

Charlson, M.E., Pompei, P., Ales, K.L., MacKenzie, C.R., 1987. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J. Chron. Dis. 40 (5), 373–383.

Che, Z., Purushotham, S., Khemani, R., Liu, Y., 2015. Distilling knowledge from deep networks with applications to healthcare domain. ArXiv151203542 Cs Stat.

Chen, A., 1991. Noncompliance in community psychiatry: a review of clinical interventions. Psychiatr. Serv. 42 (3), 282–287.

Chen, C., Zhang, Q., Ma, Q., Yu, B., 2019. LightGBM-PPI: predicting protein-protein interactions through LightGBM with multi-information fusion. Chemometr. Intell. Lab. Syst. 191, 54–64.

Chua, S.L., Chow, W.L., 2019. Development of predictive scoring model for risk stratification of no-show at a public hospital specialist outpatient clinic. Proceedings of Singapore Healthcare 28 (2), 96–104.

Coodin, S., Staley, D., Cortens, B., Desrochers, R., McLandress, S., 2004. Patient factors associated with missed appointments in persons with schizophrenia. Can. J. Psychiatr. 49 (2), 145–148.

Dantas, L.F., Fleck, J.L., Oliveira, F.L.C., Hamacher, S., 2018. No-shows in appointment scheduling – a systematic literature review. Health Pol. 122, 412–421.

Dashtban, M., Li, W., 2019. Deep learning for predicting non-attendance in hospital outpatient Appointments. In: Presented at the Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2019.451, 3 February 2020.

dos Santos, A.B.V., Carvalho, C.R., 2015. Deep learning for healthcare management and diagnosis. Iberoamerican Journal of Applied Computing 5 (2), 15–25.

Elvira, C., Ochoa, A., Gonzalvez, J.C., Mochon, F., 2018. Machine-learning-based No show prediction in outpatient visits. International Journal of Interactive Multimedia and Artificial Intelligence 4 (7), 29–34.

Frankel, S., Farrow, A., West, R., 1989. Non-attendance or non-invitation? A case-control study of failed outpatient appointments. Br. Med. J. 298, 1343–1345.

Friedman, J., 1999a. Stochastic gradient boosting, 3 February 2020. https://statweb.stanford.edu/~jhf/ftp/stobst.pdf.

Friedman, J., 1999b. Greedy function approximation: a gradient boosting machine, 3 February 2020. https://statweb.stanford.edu/~jhf/ftp/trebst.pdf.

Hamilton, P.K., Gourlay, M., 2002. Missed Hospital Appointments and Transport. Kings Fund, 38, 3 February 2020. https://www.kingsfund.org.uk/sites/default/files/UEL-missed-hospital-appointments-transport-Kerry-Hamilton-Marion-Gourlay-Kings-Fund-December-2002.pdf.

Hamilton, W., Round, A., Sharp, D., 1999. Effect on hospital attendance rates of giving patients a copy of their referral letter. Br. Med. J. 318, 1392–1395.

James, G., Witten, D., Hastie, T., Tibshirani, R. (Eds.), 2013. An Introduction to Statistical Learning: with Applications in R. Springer texts in Statistics. Springer, New York.

Junod Perron, N., Dominicé Dao, M., Kossovsky, M.P., Miserez, V., Chuard, C., Calmy, A., Gaspoz, J.-M., 2010. Reduction of missed appointments at an urban primary care clinic: a randomised controlled study. BMC Fam. Pract. 11, 79.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. LightGBM: a highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems 30. NIPS 2017, 3 February 2020. https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.

Kearns, M., Valiant, L.G., 1989. Cryptographic limitations on learning boolean formulae and finite automata. In: Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing, STOC '89. ACM, pp. 433–444. https://doi.org/10.1145/73007.73049. New York, NY, USA, 3 February 2020.

Kheirkhah, P., Feng, Q., Travis, L.M., Tavakoli-Tabasi, S., Sharafkhaneh, A., 2016. Prevalence, predictors and economic consequences of no-shows. BMC Health Serv. Res. 16, 13.

Koh, H.C., Tan, G., 2005. Data mining applications in healthcare. J. Healthc. Inf. Manag. 19 (2), 64–72.

Kruse, G.R., Rohland, B.M., Wu, X., 2002. Factors associated with missed first appointments at a psychiatric clinic. Psychiatr. Serv. 53 (9), 1173–1176.

Lehmann, T.N.O., Aebi, A., Lehmann, D., Balandraux Olivet, M., Stalder, H., 2007. Missed appointments at a Swiss university outpatient clinic. Publ. Health 121 (10), 790–799.

Loh, W.-Y., 2014. Fifty years of classification and regression trees: fifty years of classification and regression trees. Int. Stat. Rev. 82 (3), 329–348.

Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., Niu, X., 2018. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. Electron. Commer. Res. Appl. 31, 24–39.

McCarthy, K., McGee, H.M., O'Boyle, C.A., 2000. Outpatient clinic waiting times and non-attendance as indicators of quality. Psychol. Health Med. 5 (3), 287–293.

Mohamed, K., Mustafa, A., Tahtamouni, S., Taha, E., Hassan, R., 2016. A quality improvement project to reduce the 'No show' rate in a paediatric Neurology clinic. BMJ Open 5, u209266 w3789.

Moore, C.G., Wilson-Witherspoon, P., Probst, J.C., 2001. Time and money: effects of no-shows at a family practice residency clinic. *Family Medicine*-Kansas City 33 (7), 522–527.

Mozaffari-Kermani, M., Sur-Kolay, S., Raghunathan, A., Jha, N.K., 2015. Systematic poisoning attacks on and defenses for machine learning in healthcare. IEEE Journal of Biomedical and Health Informatics 19 (6), 1893–1905.

Muhammad, U.K., Choi, J.P., Shin, H., Kin, M., 2008. Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE Engineering in Medicine and Biology Society. IEEE, Vancouver, BC, pp. 5148–5151.

Nel, D.Y.M., 2014. Factors Associated with Attendance at First Clinic Appointment in HIV Positive Psychiatric Patients Initiated on Antiretroviral Therapy (ART) as In-Patients. Thesis M.Med. Psychiatry)-University of the Witwatersrand, Faculty of Health Sciences, 2014.

Nelson, A., Herron, D., Rees, G., Nachev, P., 2019. Predicting scheduled hospital attendance with artificial intelligence. Digital Medicine 2, 1–7.

NHS England, 2018. NHS to trial tech to cut missed appointments and save up to £20 million, 3 February 2020. https://www.england.nhs.uk/2018/10/nhs-to-trial-tech-to-cut-missed-appointments-and-save-up-to-20-million/.

Nicholson, I.R., 1994. Factors involved in failure to keep initial appointments with mental health professionals. Psychiatr. Serv. 45 (3), 276–278.

Norris, J.B., Kumar, C., Chand, S., Moskowitz, H., Shade, S.A., Willis, D.R., 2014. An empirical investigation into factors affecting patient cancellations and no-shows at outpatient clinics. Decis. Support Syst. 57, 428–443.

Office for National Statistics, 2011. 2011 Census: Population and Household Estimates for Small Areas in England and Wales, March 2011, 3 February 2020. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforsmallareasinenglandandwales/2012-11-23.

Obenshain, M.K., 2004. Application of data mining techniques to healthcare data. Infect. Contr. Hosp. Epidemiol. 25 (8), 690–695.

Schapire, R.E., 1990. The strength of weak learnability. Mach. Learn. 5, 197–227.

Sharp, D.J., Hamilton, W., 2001. Non-attendance at general practices and outpatient clinics: local systems are needed to address local problems. Br. Med. J. 323, 1081–1082.

Sparr, L.F., Moffitt, M.C., Ward, M.F., 1993. Missed psychiatric appointments: who returns and who stays away. Am. J. Psychiatr. 150, 801–805.

Stone, C.A., Palmer, J.H., Saxby, P.J., Devaraj, V.S., 1999. Reducing non-attendance at outpatient clinics. J. R. Soc. Med. 92 (3), 114–118.

Stowell, D., Giannoulis, D., Benetos, E., Lagrange, M., Plumbley, M.D., 2015. Detection and classification of acoustic scenes and events. IEEE Trans. Multimed. 17 (10), 1733–1746.

Stubbs, N.D., Sanders, S., Jones, D.B., Geraci, S.A., Stephenson, P.L., 2012. Methods to reduce outpatient non-attendance. Am. J. Med. Sci. 344 (3), 211–219.

Sun, X., Liu, M., Sima, Z., 2020. A novel cryptocurrency price trend forecasting model based on LightGBM. Finance Res. Lett. 32, 101084.

Townsend, P., 1988. Health and deprivation. In: Townsend, P., Phillimore, P., Beattie, A. (Eds.), Inequality and the North. Croom Helm Ltd, London, p. 221.

Wang, G., Liu, Z., 2020. Android malware detection model based on LightGBM. In: Jain, V., Patnaik, S., Popențiu Vlădicescu, F., Sethi, I.K. (Eds.), Recent Trends in Intelligent Computing, Communication and Devices. Springer Singapore, Singapore, pp. 237–243.

Wang, Z., Shah, A.D., Tate, A.R., Denaxas, S., Shawe-Taylor, J., Hemingway, H., 2012. Extracting diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning. PloS One 7, e30412.

Welsh Government, 2014. Welsh Index of Multiple Deprivation.

Wiemken, T.L., Kelley, R.R., 2020. Machine learning in epidemiology and health outcomes research. Annu. Rev. Publ. Health 41, 21–36.

Young, S.G., Tullis, J.A., Cothren, J., 2013. A remote sensing and GIS-assisted landscape epidemiology approach to West Nile virus. Appl. Geogr. 45, 241–249.

Yousaf, S., Bonsall, A., 2017. UK Townsend Deprivation Scores from 2011 Census Data. UK Data Service.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. *J. Roy. Stat. Soc.* B 67, 301–320.