Investigating key contributors to hospital appointment no-shows using explainable AI.

YIYE, V., UGBOMEH, O., EZENKWU, C.P., IBEKE, E., SHARMA, V. and ALKHAYYAT, A.

2024

Investigating Key Contributors to Hospital Appointment No-Shows Using Explainable AI

Versse Yiye
School of Creative and Cultural Business
Robert Gordon University
Aberdeen, United Kingdom
v.yiye@rgu.ac.uk

Ogochukwu Ugbomeh
School of Creative and Cultural Business
Robert Gordon University
Aberdeen, United Kingdom
o.ugbomeh@rgu.ac.uk

Chinedu Pascal Ezenkwu
School of Creative and Cultural Bus.
Robert Gordon University
Aberdeen, United Kingdom
p.ezenkwu@rgu.ac.uk

Ebuka Ibeke
School of Creative and Cultural Business
Robert Gordon University
Aberdeen, United Kingdom
e.ibeke@rgu.ac.uk

Vandana Sharma
Computer Science Department
Christ University
Delhi-NCR Campus, India
vandana.juyal@gmail.com

Ahmed Alkhayyat
College of Technical Engineering
The Islamic University
Najaf, Iraq
ahmedalkhayyat85@gmail.com

***Abstract*—The healthcare sector has suffered from wastage of resources and poor service delivery due to the significant impact of appointment no-shows. To address this issue, this paper uses explainable artificial intelligence (XAI) to identify major predictors of no-show behaviours among patients. Six machine learning models were developed and evaluated on this task using Area Under the Precision-Recall Curve (AUC-PR) and F1-score as metrics. Our experiment demonstrates that Support Vector Classifier and Multilayer Perceptron perform the best, with both scoring the same AUC-PR of 0.56, but different F1-scores of 0.91 and 0.92, respectively. We analysed the interpretability of the models using Local Interpretable Model-agnostic Explana- tion (LIME) and SHapley Additive exPlanations (SHAP). The outcome of the analyses demonstrates that predictors such as the patients' history of missed appointments, the waiting time from scheduling time to the appointments, patients' age, and existing medical conditions such as diabetes and hypertension are essential flags for no-show behaviours. Following the insights gained from the analyses, this paper recommends interventions for addressing the issue of medical appointment no-shows.**

***Index Terms*—Artificial Intelligence, AI, No-Shows, LIME, SHAP, XAI, Interpretable, Explainable, Machine Learning, Health Informatics, Health Analytics, Healthcare, Hospital**

I. Introduction

Appointment no-shows have been a major concern for the customer-facing aspects of businesses across several industries, including healthcare, education, beauty services, and banking. In healthcare, an appointment is regarded as a no-show when the patient misses the scheduled appointment time without prior cancellation or rescheduling. This can potentially lead to financial losses, inefficient resource allocation, inconveniences to the healthcare providers and significant negative impacts on other patients who could have benefited from the wasted appointment slot.

For example, in 2019, an annual loss of about $150 billion due to no-shows was recorded in the United States of America healthcare [1]. The report further highlights that every missed appointment costs a physician an average of $150,000 annually. Similarly, in the United Kingdom's National Health Service (NHS), about £1 billion of resources are wasted annually due to appointment no-shows [12]. Glauser [7] found that over 20% of patients do not attend their scheduled appointments, further highlighting the severity of the issues of no-shows. Following a more data-centric report based on a history of no-shows, an average of 42% of medical appointments scheduled in advance become no-shows [11].

The healthcare sector has benefitted from digital transformation, leveraging technologies such as Artificial Intelligence (AI), machine learning, cloud computing and the Internet of Things (IoT) to optimise its value proposition. AI and machine learning have revolutionised the healthcare industry in different ways, including medical diagnoses, drug discovery, epidemic management, and automation of routine tasks. However, being a life-critical industry, AI and machine learning applications within the healthcare sector must be transparent, explainable and responsible [13]. Interpreting advanced machine learning models has been a challenge due to their inherent black-box nature [8], making advancements in explainable AI (XAI) crucial for addressing these issues regarding interpretability.

This paper seeks to investigate the potential causes of medical appointment no-shows and, based on its findings, make recommendations for improving patients' attendance at appointments. Leveraging XAI, the research will identify the key red flags to no-shows, providing a transparent framework to enhance our understanding of factors influencing no-show behaviours. The outcome of this research will help to advance healthcare resource management and cut-down wastage.

II. Background of Study

XAI has emerged as a crucial component of AI and machine learning systems, making complex model decisions interpretable. This is particularly useful given the increasing adoption of artificial intelligence in critical sectors, including healthcare, banking, aerospace, self-driving cars, and so on. Although complex machine learning models, such as deep learning, have been popular for their remarkable performances, the lack of rationale to support their decisions may hinder their adoption and confidence in the aforementioned sectors. Therefore, XAI seeks to demystify the black-box nature of machine learning algorithms, making it possible for stakeholders, industry experts, and consumers to understand why the model made a particular decision. According to Islam et al. [10], methods such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) have been used for developing XAI models due to their high accuracy rates and robustness against noise. SHAP values offer a unified approach to feature attribution, calculating how each feature contributes to a prediction. A SHAP summary plot can visually represent the no-show's feature importance for individual instances within the dataset. Each bar in the plot corresponds to a particular appointment feature, and the length of the bar signifies the SHAP value of the feature for that specific appointment. LIME provides local explanations for individual predictions, helping to bridge the gap between the inherent complexity of advanced models. A LIME plot displays the impact of individual features on the prediction of no-shows in a diverging bar chart, which helps users understand how the model arrived at a particular decision.

III. Literature Review

A prescriptive framework proposed by Srinivas and Ravindran aimed to improve the performance of appointment systems by optimising no-show appointment systems that integrate machine learning techniques [12]. This was achieved by developing innovative appointment scheduling rules that combined sequencing and overbooking policies by utilising electronic health records and various patient information variables. However, they acknowledge that this study can be improved with the use of more informative metrics to determine the best-performing model such as the AUC-PR when dealing with imbalanced datasets. Blumenthal et al. suggest that analysing a patient's history of non-adherence using Natural Language Processing (NLP) techniques can provide valuable insights for predicting adherence to colonoscopy appointments [3]. The study reported an area under the curve (AUC) value of 0.70, indicating moderate predictive performance. The sensitivity and specificity of the model were reported as 33% and 92%, respectively. However, the study highlights its limitation in the comprehensive understanding of predictors of no-shows and the generalisability of the findings to less-educated populations. Elvira et al. explored the possibility of using Big Data and machine learning models to predict medical appointment no-shows [5]. The development of their predictive model was entirely based on variables attributed to target appointment, resulting in an Area Under the Curve (AUC) of 0.74. However, they acknowledge that the predictive model's results can be enhanced, considering that the classifier presented in their study relies on patient and appointment data, which they believe may not have been adequately comprehensive for model development.

In their study, Harvey et al. focused on predicting no- shows for patients with scheduled radiology examinations using Logistic Regression analysis [9]. The findings suggested that only 16 out of 27 variables were important predictors of no-show behaviour, providing valuable insights for developing strategies to mitigate appointment non-attendance in radiology settings. A study by Alshammari, Almalki and Alshammari [2] aimed to predict no-show appointments by developing two machine learning models: Adaptive Boosting (AdaBoost) and Decision Tree. These models were trained on an open source dataset obtained from Kaggle and evaluated, resulting in the Decision Tree model outperforming AdaBoost with a precision of 0.89, recall of 0.86, an ROC of 0.88, and F-measure of 0.87. Their study suggests that younger male patients with morning appointments and no text message reminders are more likely to miss their appointments.

The extensive body of research in healthcare, machine learning, and the ever-evolving field of XAI emphasises the pressing need to understand and mitigate the challenges of no-show behaviour. However, little attention has been paid to utilising XAI to identify potential red flags of no-shows. This study investigates the key contributors of no-shows by exploring the application of explainable AI models such as LIME and SHAP. Additionally, the issue of class imbalance is very common in datasets, and this has led to the origination of class imbalance techniques such as the Synthetic Minority Oversampling Technique (SMOTE) where synthetic data is generated from the minority class to enable the models better recognise patterns [4]. As a result, the models are less biased towards the majority class, thereby providing more accurate results. The findings from this research will help hospitals strategize how best to minimise no-shows when they identify a potential red flag.

IV. Experiment

A. Data Description

The data used for this study is open-source Brazil medical appointment data that is hosted on Kaggle. It comprises numerical and categorical variables, with 110,527 rows and 14 columns. It comprises diverse patient demographics and appointment-related attributes, as described in Table I. Each row is a medical appointment, including features describing the patient and the appointment details. The target variable (no-shows) indicates whether an appointment was attended or not. The dataset has a skewed distribution, with 80% of the patients adhering to their scheduled appointments and 20% not. This indicates that the data is imbalanced. An imbalanced dataset can potentially lead to algorithmic bias and poor generalization, which will affect the performance of the models. We have used the synthetic minority over-sampling technique (SMOTE) to improve the data distribution. SMOTE has shown to improve model performances, this will be demonstrated later in the result section. SMOTE creates artificial or synthetic

TABLE I
DATA DICTIONARY

Variables	Data Description
Number_Missed	The number of missed appointments by a patient
AwaitingTime	The duration(days) between missed appointments and the actual appointment day
Age	The age of the patient
SMS_recieved_1	SMS reminder sent to patient (1 – Yes)
SMS_recieved_0	SMS reminder sent to patient (0 – No)
Diabetes_1	Does patient have diabetes (1 – Yes)
Diabetes_0	Does patient have diabetes (0 – No)
Scholarship_1	Did patient receive financial aid? (1 – Yes)
Scholarship_0	Did patient receive financial aid? (0 – No)
Gender_F	Patient gender is female
Gender_M	Patient gender is male
Alcoholism_1	Is patient an alcoholic? (1 – Yes)
Alcoholism_0	Is patient an alcoholic? (0 – No)
Hypertension_1	Does patient have high blood pressure? (1 – Yes)
Hypertension_0	Does patient have high blood pressure? (0 – No)
Handcap_1	Does patient have a physical disability? (1 – Yes)
Handcap_0	Does patient have a physical disability? (0 – No)
No-Show	True – 1 or False - 0

samples from the minority class, instead of duplicating existing sample instances of the data by generating [4].

B. RESULTS AND DISCUSSION OF FINDINGS

This study uses evaluation metrics such as precision, re- call, AUC-PR, and F1-score to evaluate the performance of six machine learning models. In this study, precision is the proportion of predicted no-shows that are actual no-shows; recall measures the percentage of actual no-shows that are correctly predicted; AUC-PR evaluates the balance between precision and recall, focusing on the accuracy of predicting the positive class (no-shows); and F1-score provides a balanced evaluation of precision and recall. The two main evaluation metrics for the study are AUC-PR and F1-score. These two metrics provide a comprehensive assessment of the models' performance, considering both the balance between precision and recall and their sensitivity to false positives.

1) EVALUATION OF THE PREDICTIVE MODELS: Before applying SMOTE, the six machine learning models responded to the class imbalance with high recall values, indicating a potential bias towards the majority class, as shown in Table II. To address this issue and enhance the models' ability to handle the imbalance, SMOTE was employed.

Following the application of SMOTE to the training dataset, this study tackled the issue of data imbalance, resulting in the Multilayer Perceptron model achieving a precision and recall of 0.96 and 0.88, respectively, as shown in Table III. This demonstrates the importance of addressing data imbalance and potential bias in models to enhance the overall perfor- mance of machine learning models. In comparison with the results achieved before the application of SMOTE technique, the models exhibit significant improvement, highlighting the impact of balancing techniques on the performance of machine learning models. The increased performance metrics, including accuracy, precision, recall, F1-score, and AUC-PR, signify the effectiveness of SMOTE in addressing the issue of class imbalance in the dataset. This enhances the models' ability to capture patterns related to appointment no-shows and shows, thereby improving the models' reliability and effectiveness in real-world applications within healthcare systems.

C. APPLICATION OF EXPLAINABLE AI

This section uses XAI techniques, namely, SHAP and LIME, to determine the variables contributing to appointment no-shows. The use of XAI helps to explain why a model predicts if a patient might miss an appointment. Table IV summarises the consistent SHAP and LIME results for all models' no-show predictors.

1) SHapley Additive exPlanation (SHAP): A SHAP plot visually represents no-show features in order of importance for individual instances within the dataset. This plot illustrates how individual features contribute to a particular model predic- tion and ranks them in descending order of importance, with the most contributing features displayed at the top. Figure 1 shows this visual representation of a SHAP plot, with class 1 representing no-shows and class 0 representing shows.

The Logistic Regression model revealed that features such as *"Number Missed"*, *"AwaitingTime"*, and *"SMS received"* were among the most influential in predicting no-show behaviours. From the interpretation of the patient's predicted outcome, patients with a longer wait time between scheduling and the actual appointment day exhibited higher SHAP values, indicating an increased likelihood of a no-show. Notably, patients with a history of missed appointments are more likely to have higher SHAP values, suggesting a correlation between number of appointments missed and no-show probability. Additionally, *"SMS received"* emphasises the potential impact of communication through text messages on patient attendance in the Logistic Regression model, as patients who didn't receive an SMS reminder of their appointment are more likely to convert to no-shows. In the Decision Tree model, SHAP highlighted *"Number Missed"*, *"AwaitingTime"*, and *"Age"* as the most contributing predictors of no-shows' behaviour. Patients with a history of missed hospital appointments and longer wait

TABLE II
SUMMARY OF MODELS PERFORMANCE BEFORE THE APPLICATION OF SMOTE

Results	Logistic Regression	Decision Tree	Support Vector Classifier	Random Forest	XGBoost	Multilayer Perceptron
Accuracy	80%	73%	80%	75%	80%	79%
Precision	0.80	0.81	0.80	0.81	0.80	0.80
Recall	1.0	0.86	1.0	0.89	0.99	0.99
F1-Score	0.89	0.84	0.89	0.85	0.89	0.88

TABLE III
SUMMARY OF MODELS PERFORMANCE AFTER THE APPLICATION OF SMOTE

Results	Logistic Regression	Decision Tree	Support Vector Classifier	Random Forest	XGBoost	Multilayer Perceptron
Accuracy	83%	84%	84%	85%	85%	87%
Precision	0.97	0.92	0.86	0.94	0.98	0.96
Recall	0.81	0.88	0.96	0.87	0.83	0.88
F1-Score	0.91	0.90	0.91	0.91	0.90	0.92
AUC-PR	0.51	0.48	0.56	0.51	0.55	0.56

TABLE IV
SUMMARY OF KEY CONTRIBUTORS TO HOSPITAL APPOINTMENT NO-SHOWS USING XAI

Machine Learning Models	SHAP Features	LIME Features
Logistic Regression	Number_Missed, AwaitingTime, SMS_recieved	Number_Missed, Hypertension, SMS_recieved
Decision Tree	Number_Missed, AwaitingTime, Age	Number_Missed, SMS_recieved
Support Vector Classifier	Number_Missed, AwaitingTime, Age	Number_Missed, Age
Random Forest	Number_Missed, AwaitingTime, Age	Number_Missed, Age
XGBoost	Number_Missed, AwaitingTime, SMS_recieved	Number_Missed, Diabetes, Age
Multilayer Perceptron	Number_Missed, AwaitingTime, Gender_M	Number_Missed, Age, SMS_recieved

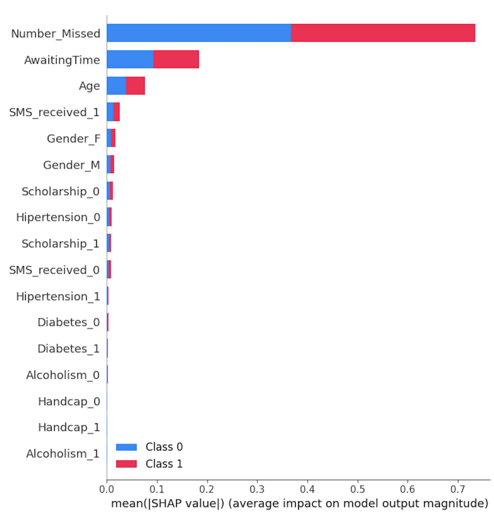

Fig. 1. Visual Representation of a SHAP Plot

times between their scheduled day of appointment and the actual day of appointment resulted in higher SHAP values, suggesting an increased probability of a no-show. Similarly, the *"Age"* variable also has higher SHAP values compared to the existing variables. The Support Vector Classifier model highlights *"Number Missed"*, *"AwaitingTime"* and *"Age"* as significant contributors to no-shows, indicating that a patient's history of missed appointments, age and a longer waiting time increases the chance of a patient missing a scheduled appointment. Random Forest revealed that features such as *"Number Missed"*, *"AwaitingTime"* and *"Age"* were the most influential features in predicting no-shows.

Patients with a longer wait time between scheduling and the actual appointment day exhibited higher SHAP values, indicating an increased likelihood of a no-show. Similarly, the *"Age"* tended to have higher SHAP values, suggesting a correlation between the age of patients and no-show probability. The XGBoost result highlights *"Number Missed"*, *"AwaitingTime"* and *"Age"* as the key factors for predicting no-show behaviour. The SHAP values indicate that patients with a longer wait between scheduling and appointment days are more likely to have a no-show behaviour than those who have a shorter wait. Similarly to other models, Multilayer Perceptron identifies *"Number Missed"* and *"AwaitingTime"* as important factors. The model also considers *"Gender M"* in the top three most dominant features in predicting no-shows. Moreover, an increase in SHAP values was observed for patients who had a prolonged wait between scheduling and appointment day. The higher SHAP values for *"Gender M"* suggest that the male gender is more likely to miss their scheduled appointment. This could introduce and reinforce the existing gender bias, thereby leading to unfair or discriminatory predictions. The research did not seek to explore the preceding issue further.

Across all models, the SHAP plot identifies *"Number Missed"*, *"AwaitingTime"* and *"Age"* as the dominant flags for identifying no-show behaviour. *"Number Missed"* strongly correlates with no-show behaviour, implying that patients with a history of missing appointments are more likely to continue this behaviour. Similarly, *"AwaitingTime"* points to a higher probability of no-shows, meaning that prolonged lead times to scheduled appointments might contribute to no-shows. The

SHAP values for the *"Age"* variable demonstrate that children and older patients are more likely to keep to their scheduled appointments, while younger adults are more likely to miss their scheduled appointments. Identifying these common indicators gives healthcare facilities actionable insights for designing interventions and strategies to reduce no-show rates.

2) Local Interpretable Model-agonistic Explanation (LIME): A LIME plot provides insight into the contribution of individual features in a machine learning prediction. Figure 2 demonstrates the feature attribute for a patient, showing diverging bars where the most contributing features of no-shows are displayed on the right, and the most important shows are displayed on the left. The longer bars with higher values indicate a greater importance of the feature. The LIME plot for the Logistic Regression model predicts that this patient will miss their appointment with a 100% probability.

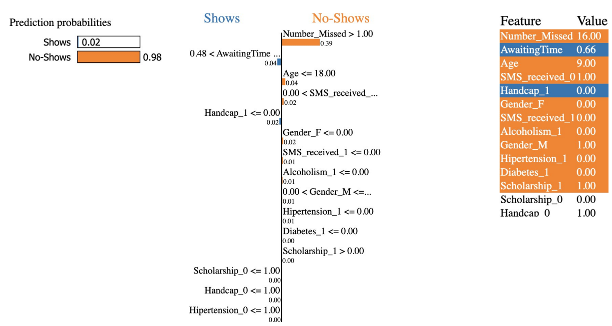

Fig. 2. Visual Representation of a LIME Plot

For the Logistic Regression model, *"Number_Missed"*, *"SMS_received"* and *"Hypertension"* are identified as the most contributing factors to no-show behaviours, meaning that a patient's appointment history and medical conditions can underscore if the patient will attend a scheduled appointment. In the case of the Decision Tree model, LIME highlights *"Number_Missed"* and *"SMS_recieved"* as key flags for no-show behaviours. Similarly, LIME analysis of the Support Vector Classifier demonstrates that *"Number_Missed"* as well as *"Age"* and *"Alcoholism"* are critical in predicting no-shows. While the history of a patient's missed appointments and age indicates a high probability of no-shows, the extent of the relationship between a patient's alcoholic status and no-shows would necessitate further investigation, maybe leveraging qualitative research. Based on LIME's interpretation of the Random Forest model, *"Number_Missed"*, *"Age"*, and *"Alcoholism"* are the dominant features for no-show behaviours. The XGBoost LIME result identifies *"Number_Missed"* and *"Diabetes_1"* as influential factors to no-show behaviours, indicating that a patient's history of appointments and medical conditions can help to inform if the patient will likely not attend a scheduled appointment. MLP, in line with other models, highlights *"Number_Missed"*, *"Age"* and *"SMS_recieved"* the key predictors of no-show.

The LIME plot across all models highlights *"Number_Missed"* as a strong contributor of no-shows. This indicates that the probability of a patient missing scheduled appointments increases with the number of missed appointments in the past. Additionally, several other features emerged as important factors, although their influence varied among models. These features include *"Age"*, *"SMS–received"*, and specific medical conditions like *"Hypertension"*, *"Alcoholism"* and *"Diabetes–1"*. The diversity of patient attributes illustrates the complexities of predicting medical appointment no-shows. However, it is important to note that *"AwaitingTime"* appeared to strongly contribute to a patient showing up for their scheduled appointment.

V. RECOMMENDATIONS

Based on the top three contributors to no-shows, "Number _Missed", "AwaitingTime", and "Age", we propose the following recommendations to mitigate hospital no-shows.

1) Implement automated reminder systems and educate patients on the repercussions of missed appointments on their health and the healthcare system.
2) Optimisation of appointment scheduling and real-time updates on changes made to patients' appointments to enable healthcare providers to manage the prolonged wait time and its potential negative impact on patients' expectations.
3) Tailor communication strategies based on age groups by utilising digital reminders for younger patients while employing phone calls or physical mail to ensure they receive appointment notifications.
4) Adopt an integrated approach combining these recommendations and leveraging predictive machine learning models. Patient's features and historical behaviours can be incorporated into these models to provide highly tailored interventions.
5) To mitigate appointment no-shows requires continuous monitoring and continuous improvement of strategies, following feedback and further analysis.

VI. CONCLUSION

The issue of no-show appointments in healthcare has been reviewed in this paper. Based on the existing reports, appointment no-shows with the healthcare sector result in significant financial implications and inefficient resource allocation. This paper uses machine learning to predict no-shows, leveraging model interpretability techniques such as SHAP and LIME to highlight red flags indicating potential appointment no-shows. Although this paper has made important findings, based on which strategies for managing appointment no- shows have been recommended, one here is that the paper is geographically and culturally restricted to Brazil, which may mean that the results of this paper cannot be generalised to other regions. Based on the preceding reason, we are looking to expand this work to include more diverse datasets that cut across various regions and healthcare settings. We will look to explore other more advanced machine learning techniques, especially in the context of unstructured datasets. Features such as *"Number_Missed"*, *"Age"*, *"SMS_received"*, *"AwaitingTime"*, and medical conditions have been identified as red flags for predicting no-shows. These insights are crucial for effectively improving hospital appointment attendance and consequently addressing the issues of resource wastage within healthcare.

References

[1] ACS, "The $150B repercussions of patient no-shows — 5 statistics," 2019. [Online]. Available: https://www.beckersasc.com/asc-news/the-150b-repercussions-of-patient-no-shows-5-statistics.html. [Accessed: June 11, 2023].

[2] A. Alshammari, R. Almalki, and R. Alshammari, "Developing a Predictive Model of Predicting Appointment No-Show by Using Machine Learning Algorithms," *Journal of Advances in Information Technology*, vol. 12, no. 3, 2021.

[3] D. M. Blumenthal et al., "Predicting Non-Adherence with Outpatient Colonoscopy Using a Novel Electronic Tool that Measures Prior Non-Adherence," *Journal of General Internal Medicine*, vol. 30, no. 6, pp. 724–731, 2015.

[4] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Machine Learning*, 2023.

[5] C. M. Elvira et al., "Machine-Learning-Based No Show Prediction in Outpatient Visits," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 7, pp. 29, 2018.

[6] T. Gebhart, "No-show management in primary care: a quality improvement project," *unav*, Date not provided.

[7] W. Glauser, "When patients miss appointments, everyone pays," *Canadian Medical Association Journal (CMAJ)*, vol. 192, no. 6, pp. E149–E150, 2020.

[8] R. Guidotti et al., "A Survey of Methods for Explaining Black Box Models," *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–42, 2019.

[9] H. B. Harvey et al., "Predicting No-Shows in Radiology Using Regression Modeling of Data Available in the Electronic Medical Record," *Journal of the American College of Radiology: JACR*, vol. 14, no. 10, pp. 1303–1309, 2017.

[10] M. Islam et al., "A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks," *Applied Sciences*, vol. 12, pp. 1353, 2022.

[11] N. L. Lacy et al., "Why We Don't Come: Patient Perceptions on No-Shows," *Annals of Family Medicine*, vol. 2, no. 6, pp. 541–545, 2004.

[12] S. Srinivas and A. R. Ravindran, "Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework," *Expert Systems with Applications*, vol. 102, pp. 245–261, 2018.

[13] G. Stiglic et al., "Interpretability of machine learning-based prediction models in healthcare," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 5, e1379, 2020.