# Personalised Notification Scheduling in ePRO System: A Privacy-Preserving ML Framework

Ishrat Fatima Syed
ishrat.syed2@mail.dcu.ie
School Of Computing,
Dublin City University
22264608

Pooja Ballolli
pooja.ballolli2@mail.dcu.ie
School Of Computing,
Dublin City University
22261460

## ABSTRACT

**Motivation:** The motivation behind this research stems from the critical need to strike a balance between delivering personalised notifications to healthcare patients and safeguarding their sensitive medical information. Existing solutions may compromise data privacy, leading to potential misuse by third-party entities. Thus, our research aims to design an efficient and secure system that can provide personalised notifications without exposing confidential patient data.

**Data:** We utilised four diverse datasets, including Installation Data, Research Participant Data, Response Data, and Notification Data, to build a comprehensive foundation for our research. These datasets comprise various features such as app version, device type, participant details, notification responses, and questionnaire configurations. All data underwent rigorous cleaning and preprocessing to ensure data integrity and privacy.

**Approach:** To achieve privacy preservation and enhance the notification response rate, we devised a multi-step approach. First, we employed autoencoders to encode participant data into a lower-dimensional latent representation, automatically extracting essential features while preserving privacy. The encoded features were anonymised and stored securely, preventing exposure to raw patient data. Subsequently, a predictive model, leveraging XGBoost, was trained using the encoded features to determine participants likely to respond positively to notifications.

**Results:** Our experimental results demonstrated the effectiveness of the proposed approach. The XGBoost model achieved an average AUC-PR of 0.77, with a recall of 0.72 and precision of 0.74, indicating its capability to accurately identify participants likely to respond to notifications. The use of autoencoders significantly improved the model's performance by preserving data privacy while enhancing feature extraction and prediction accuracy.

**Contributions:** This research significantly contributes to the healthcare industry by offering a robust and privacy-preserving personalised notification system. By employing autoencoders, our approach effectively addresses data privacy concerns and prevents potential misuse of sensitive medical information. Moreover, the predictive model's accuracy enables timely interventions, enhancing patient engagement and healthcare outcomes. Our framework sets a precedent for applying similar privacy-preserving methodologies across various industries handling sensitive data, paving the way for improved data security in the digital age.

***Keywords***: **Personalised Notifications, Privacy Preservation, Autoencoders, XGBoost, healthcare, data security, patient engagement, data privacy, Machine learning, ePRO systems**

## I. INTRODUCTION

Electronic Patient-Reported Outcome (ePRO) systems have emerged as powerful digital platforms that enable patients to report their symptoms, side effects, and quality of life measures using electronic devices such as smartphones, tablets, or computers [1]. These systems play a crucial role in modern healthcare, providing real-time data on patient's health status and quality of life, thereby facilitating direct communication between healthcare professionals and patients participating in clinical trials. However, as patients interact with these systems, there arise significant concerns regarding the collection, storage, and use of their sensitive health information, necessitating a careful balance between patient engagement and data privacy protection. Patient engagement with ePRO systems is vital for successful clinical trials and meaningful data collection. One effective approach to encourage engagement is through the use of timely notifications. Extensive research in various health domains has demonstrated that well-timed mobile notifications can bring about positive behaviour change [2]. Intervention prompts, such as email and push notifications, have shown promise in motivating individuals to adopt positive health behaviour changes [3]. To optimise the impact of notifications, it is crucial to determine when users are most likely to attend and respond to them without causing frustration or uninstallation of associated apps due to perceived annoyance [4]. Delivering notifications at appropriate moments that do not disrupt user's daily routines is critical [5]. Context-aware notification management systems (CNMS) have been explored to address this issue and ensure that notifications add value to users without becoming a nuisance [6].

However, while enhancing patient engagement through personalised notifications, data privacy concerns must remain paramount. ePRO systems house extensive personal data,

including sensitive health information, demanding strict adherence to data privacy regulations and guidelines. Robust security measures, such as encryption, secure data storage, access controls, and regular security audits, must be implemented to protect patient data from unauthorised access and breaches [7]. Additionally, using predictive models for personalised notification send times requires careful consideration from a data privacy standpoint. Although these models can improve patient outcomes, they require access to patient data, such as usage patterns within the ePRO system. Safeguarding privacy and confidentiality is essential, and de-identification and aggregation of patient data should be practised whenever possible to prevent individual patient identification [8].

In this paper, we propose a machine learning framework that addresses the data privacy concerns in ePRO systems while enhancing patient care through personalised notifications. Our framework prioritises the protection of patient data by securely encoding any information leaving mobile devices, ensuring it remains non-human readable and safeguarded from potential misuse and unauthorised access. We illustrate the effectiveness of this framework by developing a machine learning model that predicts whether a user will respond to a notification, enabling the system to send alerts at the most appropriate time for optimised engagement. By adhering to strict data privacy regulations, ensuring transparency, and responsible data usage, ePRO systems can harness the power of technology effectively, leading to better patient outcomes, improved data quality, and efficient use of healthcare resources, all while protecting patient's sensitive information. This integration of privacy and personalised care cement ePRO systems as invaluable tools in the healthcare industry, revolutionising the way patients and healthcare professionals interact in clinical trials [9].

## II. Literature Review

In a study [10] by Mehrotra et al., they investigated the issue of intrusive and disruptive mobile app notifications. They proposed using content and context to design intelligent non-disruptive notification mechanisms and predict optimal delivery moments. Their Android app, NotifyMe, collected user data and used Naive Bayes, AdaBoost, and Random Forest algorithms to predict notification responses. The study revealed that user activity and the nature of notifications influenced response times and acceptance rates. The paper suggests enhancing prediction models with natural language processing techniques. These insights are relevant for optimising notification sent times and designing effective notification systems.

In a paper [11] titled "Nurture: Notifying Users at the Right Time Using Reinforcement Learning," Ho et al., introduce Nurture, a personalised notification system that utilises reinforcement learning to identify the most opportune moments for sending notifications. The primary objective of this system is to improve user engagement and retention rates. Traditional notification systems often lack personalisation, relying on fixed schedules or basic heuristics, which can lead to users ignoring or perceiving notifications as spam. To overcome these challenges, Nurture leverages machine learning and user data, including contextual features like time of day and user activity level, along with historical behaviour, to predict the best notification timing for each individual user. Through reinforcement learning, the system continuously improves its notification strategy, maximising user engagement and retention rates. Evaluating Nurture on real-world mobile app usage data, the authors demonstrate its superiority over traditional methods, indicating its potential to revolutionise notification systems beyond mobile apps. The paper presents a valuable contribution to personalised notification systems, emphasising the advantages of machine learning-driven optimisation, and highlights Nurture's versatility for application in various domains such as email and social media notifications.

Araújo et al., introduced a novel service [12] aimed at determining the best time interval for sending marketing messages to individual subscribers based on their profiles. The study analysed a dataset of email campaign observations over two years, revealing that Thursdays and Fridays during morning hours were optimal for sending email campaigns. Their findings underscored the significance of timing in the effectiveness of email campaigns, emphasising the need to consider subscriber behaviour for successful marketing communication. They evaluated three regression models for predicting the optimal time intervals to send marketing messages: Random Forest Regressor, Multiple Linear Regression, and K-Neighbors Regressor. They employed a parallel ensemble strategy known as stacking to enhance prediction performance. The stacking strategy involves combining the predictions of multiple models to create a meta-model that yields improved results, making it the most successful approach among the models tested in determining the optimal timing for marketing communication. The stacking strategy demonstrated the best performance for predicting optimal communication timing among the models tested.

Additionally, Deligiannis et al., in their research [13] employed a similar regression-based approach to determine the optimal timing of product repurchases through SMS messages. Using regression algorithms, their models calculated the time between a subscriber's last and potential next purchase, considering factors like open message rates, purchase transactions, participation frequency, and click-through rates. The second proposed model utilised the output of the first to establish approximate dates for automatic reminders. Although their regression-based models provided valuable insights into optimising product repurchase timing, the study acknowledged the limitations of a small dataset, which could impact the generalisability of their results [13].

Rieke et al., in their study [14] delve into the transformative impact of artificial intelligence (AI), machine learning (ML), and deep learning (DL) in the fields of radiology, pathology, and genomics. They highlight the critical need to overcome data centralisation and privacy concerns prevalent in healthcare applications. To address these challenges, the researchers propose the adoption of Federated Learning (FL), a collaborative learning approach that eliminates the necessity

of sharing raw data, and which has already demonstrated successful applications in digital health.

One of the key features of FL is that the ML process occurs locally at each participating institution, and only model characteristics such as parameters and gradients are transferred among the institutions. Notably, recent research findings showcase that models trained through FL achieve performance levels comparable to those trained on centrally hosted datasets, and even outperform models that rely solely on isolated single-institutional data [15] [16].

This study sheds light on the diverse stakeholders in the healthcare sector and highlights the advantages FL offers to them. Large-scale initiatives like the Trustworthy Federated Data Analytics (TFDA) project and the German Cancer Consortium's Joint Imaging Platform are specifically acknowledged for fostering secure and innovative collaboration in healthcare research. Additionally, the study acknowledges international collaborations, such as those focusing on the development of AI models for mammogram assessment, which serves as demonstrative examples of FL's potential in enhancing diagnostic consistency and sensitivity to rare cases.

Despite the implementation challenges, the authors emphasise the crucial efforts required to ensure data compliance and uniformity across collaborating institutions. Furthermore, they underscore how FL can positively impact manufacturers of healthcare software and hardware, enabling continuous validation and improvement of ML-based systems without compromising patient-specific information.

Cha et al., in a recent study [17] focussed on implementing Federated Learning (FL) on vertically partitioned data to achieve comparable performance with centralised models while maintaining data privacy. The researchers employed overcomplete autoencoder-based models to transform the data into latent representations, enabling data aggregation without compromising the raw data's privacy. Subsequently, they trained a tabular neural network model with categorical embedding using the aggregated latent data. To assess performance, the study also used a centrally based model as a baseline, evaluating the accuracy and the area under the receiver operating characteristic curve (AUC-ROC).

The study's findings demonstrated the successful transformation of data into latent representations using the autoencoder-based network, eliminating the need for domain-specific knowledge at individual sites and ensuring data security. The performance loss was minimal with the overcomplete autoencoder, resulting in accuracy and AUC-ROC reductions ranging from 0% to 8.89% across the various datasets.

The authors utilised autoencoders with non-negativity constraints as a key component of their privacy-preserving intrusion detection pipeline. Autoencoders are neural network architectures commonly used for feature extraction tasks. By imposing non-negativity constraints on the autoencoder's weights and activations, the model is encouraged to learn sparse and non-negative representations of the input data. This characteristic is crucial for privacy preservation as it helps prevent the generation of sensitive information or patterns that could potentially reveal individual or sensitive data. By extracting less redundant features through the non-negative autoencoders, the system reduces the risk of exposing private information while maintaining the necessary information to perform effective intrusion detection. This way, the proposed approach enhances the privacy protection of IoT health systems during the incremental learning process, making it suitable for handling sensitive data without compromising individual privacy.

The proposed approach of autoencoder-based Federated Learning proved to be an effective method for privacy-preserving machine learning on vertically partitioned data. It allowed for the development of robust models without the necessity of centralising sensitive data, thus addressing the privacy and data governance challenges associated with traditional ML algorithms.

Novoa et al., in their paper [18] present a novel, fast, and privacy-preserving implementation of deep autoencoders known as Fast Deep Autoencoder for Federated Learning (DAEF). This approach ensures that the exchange of information during federated learning does not compromise the privacy of the underlying training data. The study compares DAEF with other methods, such as OS-ELM (Online Sequential Extreme Learning Machine), SHL-AE (Stacked Hybrid Learning with Autoencoders), and MHL-AE (Multimodal Hybrid Learning with Autoencoders), in terms of anomaly detection in time-critical environments, specifically edge computing and federated learning scenarios.

The research methodology comprehensively evaluates DAEF's performance, highlighting its strengths in parallel and incremental learning, low computational overhead, and ability to protect data privacy. Seven diverse datasets from reputable repositories were utilised for evaluation, including the UCI Machine Learning Repository, Kaggle, and ODDS.

The research concludes that DAEF offers distinct advantages in edge computing and federated learning contexts due to its decentralised architecture and utilisation of MQTT as the communication protocol. Compared to OS-ELM, DAEF demonstrates superior performance, especially in scenarios with a limited number of instances per node. Additionally, its capability to handle deep architectures further positions DAEF as an excellent alternative for building complex models in time-critical environments. The study establishes DAEF as a highly promising and practical approach for effective anomaly detection in edge computing and federated learning applications while safeguarding data privacy.

## III. Methodology

In this section, we present an overview of the techniques utilised in formulating our solution, accompanied by a rationale for the choices made. The underlying principles governing our model design were centred around two key aspects: privacy-preserving algorithms and personalised models to enhance the accuracy of notification time prediction.

Considering the sensitive nature of ePRO systems and the confidential patient data involved, there exists a significant risk

of potential misuse by companies seeking to develop models that selectively target individuals based on their medical history. Therefore, the methodology outlined in this paper endeavours to establish a robust framework that prioritises privacy, safeguarding personal data from exposure to models or third-party entities.

Our primary objective is to optimise the efficiency of ePRO systems by guaranteeing timely notifications to our users. To achieve this, our solution employs training autoencoders for each participant, which effectively encode their data and automatically extract essential features from it. This approach enables us to build models that learn from the unique characteristics of individual participants, thereby enhancing the accuracy and relevance of our predictions. Additionally, by utilising this data encoding framework, we aim to demonstrate the effectiveness of our approach in optimising notifications and at the same time preserving privacy.

By striking a balance between preserving privacy and enhancing efficiency, our proposed approach aims to address the challenges posed by ePRO systems, offering a secure and effective system for timely notifications without compromising on data protection. It is worth noting that this framework is not limited to the ePRO systems or healthcare industry alone but can be effectively applied across various industries that handle sensitive data. By demonstrating the versatility and reliability of our data encoding techniques, we seek to provide a valuable and applicable solution for data privacy and security concerns in a wide range of domains beyond healthcare.

*Understanding the data*

In this section, we present an overview of four datasets utilised in our longitudinal study, which was conducted in collaboration with In The Wild Research. The study involved 350 participants who received notifications five times a day for a total of 24 days over the course of 12 weeks. The data was collected in three cycles of 8 days each, spaced 5 weeks apart, resulting in a grand total of 832,064 data points.

The first two datasets, namely "Installation Data "and "ResearchParticipant Data," contained device-related information collected from each of the 350 participants at the beginning of the study. These datasets helped us understand the diversity of devices used by participants and how it may influence their responses to notifications.

The third dataset, "Response Data," provides valuable insights into how participants reacted to the notifications they received throughout the study period. Lastly, the fourth dataset, "Notification Data," encompasses the configurations of the questionnaires used during the study.

Collectively, these four datasets formed the basis of our research, enabling us to optimise system efficiency, prioritise privacy through personalised models, and offer valuable solutions for data privacy and security concerns across various industries.

*Data Cleaning*

Data cleaning is a critical step in ensuring the reliability and coherence of datasets for research studies. In this study, we undertook a comprehensive data cleaning process to prepare four datasets for analysis. We examined each dataset to identify unique values and check for null values. We then calculated two essential columns, "HowLongToEngage" and "HowLongToSubmit," for the "Response" dataset. To better understand participant behaviour, we extracted significant timestamp information, such as "submitHour" and "probeHour." We handled duplicate records in the "Installation" dataset and merged all four datasets into a unified dataset. We converted timestamp values to a consistent datetime format and dropped columns with mostly null values. We addressed null values in the "Variance" column by filling them with the mean value. We excluded test participants from the merged dataset and eliminated duplicate records for "timeBasedSurveyQuestionId" and "researchParticipantId." We removed undelivered records and deleted erroneous records for a specific probe day. Finally, we introduced a "Class" column, categorising responses as '1' for responded and '0' for not responded.

As a result of this comprehensive data cleaning process, we successfully reduced the dataset from 832,064 data points to 26,796 data points. This resulted in a reliable and coherent dataset, devoid of duplicates, null values, and errors. The thorough cleaning allowed us to derive valuable insights from the study.

*Feature Engineering*

Feature engineering plays a crucial role in any machine learning model, as the quality of inputs determines the quality of outputs. Our dataset contained limited participant-related features, such as app version and device type, while having numerous notification-related features, including probetimestamp, engagetimestamp, submittimestamp and timebased notificationId, among others. To enhance the dataset and provide meaningful context, we calculated additional fields like Notification numbers, day notification numbers, and cycles, which served as features representing the notification context for the model.

Moreover, we derived time-based features from the probetimestamp, such as probehour, day, month, year, and weekend, which helped capture the timing of notifications.

Due to the scarcity of participant-related features for our model, we employed clustering techniques such as K-means. Specifically, we clustered participants based on the number of notifications they responded to across different hours of the day. This approach resulted in three distinct clusters that we utilised as features, thereby enriching the representation of participant information in the model.

By combining feature engineering and clustering, we effectively addressed the challenge of limited participant-related features, improving the model's performance and enabling meaningful analysis in our study.

*Data Encoding and Privacy preservation*

Once we identified the pertinent features for our framework, the pivotal aspect of our solution unfolds. In a typical classification task, the chosen features are directly used by

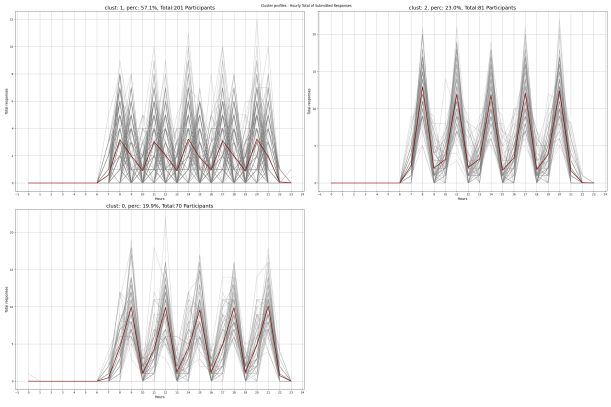

Figure 1: Participant Clusters

the model. However, in our proposed approach, we prioritise participant privacy and retain their data on their devices. Instead of directly using the selected features, we employ a simple autoencoder for each participant's chosen features.

Figure 2 illustrates the architecture of an autoencoder, which comprises two main components: an encoder and a decoder. The encoder compresses the input data into a lower-dimensional latent representation, capturing essential features and eliminating noise. This transformation is achieved through neural networks. The decoder then reverses this process, reconstructing the original data from the compressed representation. These components enable unsupervised learning, data compression, and various data analysis tasks.

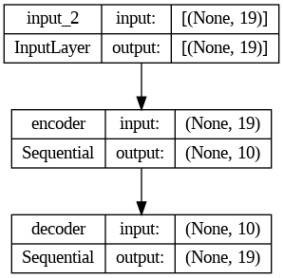


Figure 2: AutoEncoder Architecture

We trained the autoencoder model described above on individual participant data with the objective of learning the latent representation of their unique behavioural features in a lower-dimensional space. This latent representation captured personalised interaction patterns with notifications, and the encoded data's complexity made it practically impossible for humans to interpret or utilise the information for any purpose beyond sending timely notifications.

During training, the autoencoder employed an unsupervised approach, where the input data served as its own target. The model iteratively optimised its parameters to minimise the reconstruction error between input and output data. Figure 3 illustrates the learning curves, which provide crucial insights into the model's performance and convergence. The x-axis represents the number of training iterations or epochs, while the y-axis shows the value of the reconstruction loss.

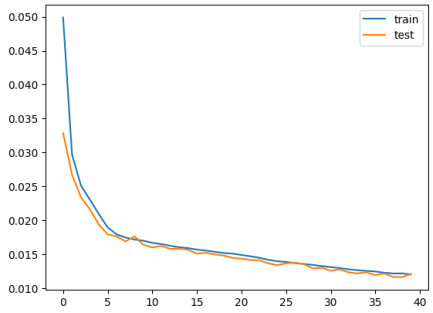


Figure 3: AutoEncoder Learning curves

As training progresses, the reconstruction loss gradually decreases, indicating that the autoencoder is becoming more proficient at capturing the essential features of the input data and generating accurate reconstructions. In the initial stages, the loss reduction may be more rapid as the model quickly learns basic patterns. However, as it approaches convergence, the loss reduction rate may slow down, indicating that the model is refining its representation of the data's nuances. Achieving minimal loss indicated an effective encoding and decoding process, signifying meaningful and compact representations of the input data—an indispensable aspect of our specific use case. These representations can subsequently be leveraged for training a classification model to optimise the delivery of notifications.

In summary, the trained autoencoder takes all the selected features as input and compresses them down to just 10 unrecognisable features. These encoded features serve three main purposes:

a) Privacy Preservation: The encoding makes the features unrecognisable, ensuring that anyone accessing them cannot decipher their original composition, thus safeguarding the user's privacy.

b) Automatic Feature Extraction: The encoding process is an automatic feature extraction mechanism [19], capturing the interplay and relationships between the selected features within the participant's data.

c) Dimensionality Reduction: The encoded features effectively reduce the dimensionality of the data [20], simplifying the representation while retaining essential information.

By encoding the features in this manner, we strike a balance between preserving participant privacy and maintaining data utility. This approach ensures that sensitive information remains secure, making it a crucial component of our framework for privacy-preserving data analysis.

### *Building Predictive Model*

Building upon the previous steps, our next step was constructing a predictive model. The core research questions we aimed to address were as follows:

***Research Question 1: Can we predict a participant's response to a notification based on the set of features provided?***

By utilising the raw engineered features as inputs to our predictive model, we sought to determine a baseline for accurately forecasting whether a participant would respond positively or not to a given notification. The predictive model's output would provide valuable insights for optimising notification delivery and enhancing user engagement.

***Research Question 2: Can the model's performance using encoded features match or surpass that achieved without encoded features?***

To assess the effectiveness of the autoencoder-derived features, we aimed to compare the predictive model's performance using encoded features against a model constructed solely with the original, unencoded features. A successful outcome would demonstrate the utility and efficiency of the autoencoder's latent representation in improving prediction accuracy.

***Research Question 3: Will this approach preserve privacy and enable dataset augmentation with sensitive participant information, such as age, medical condition, and gender?***

An essential aspect of our approach was to ensure data privacy while still harnessing sensitive participant information to enrich the predictive model. We aimed to augment the dataset with features like age, gender, and medical condition without compromising confidentiality. By using the encoded features, which inherently mask the original data, we strived to strike a balance between data utility and privacy preservation, thus allowing us to augment the dataset with valuable information.

After setting out to tackle the research questions one by one, our initial focus was on the first question. To establish a baseline, we aimed to predict whether a participant responds to a notification using the raw features. For this binary classification task, we opted for the well-established and widely used Logistic Regression model.

As mentioned earlier, our raw features were all numerical and scaled, and the class was clearly defined, representing whether a participant responded to the notification or not. However, during the exploratory data analysis (EDA), we observed a class imbalance in the dataset, with approximately three times more instances of responded notifications than non-responded notifications. To address this imbalance, instead of employing oversampling, undersampling, or SMOTE techniques, we used the class weights parameter for the logistic regression model in the baseline. Our aim was to keep the baseline model simple and straightforward.

To evaluate the model's performance, we implemented 5-fold cross-validation across participants. This approach ensured robustness in the evaluation process, as it accounted for potential variations in participant behaviour across different folds. Additional details and specifications were incorporated into the logistic regression model to fine-tune its performance and capture the nuances of participant responses effectively.

In this study, given the imbalanced nature of our dataset and the critical importance of correctly identifying both classes (participant responding and not responding to notifications), we adopted a comprehensive evaluation approach to assess our model's performance. Our primary focus was on achieving high precision and recall for each class.

To address the class imbalance, we incorporated class-specific evaluation metrics into our analysis. For the "not responded" class, precision was calculated as the ratio of correctly predicted "not responded" instances to all instances labelled as "not responded." Recall was determined as the ratio of correctly predicted "not responded" instances to the total actual "not responded" instances in the dataset. Similarly, for the "responded" class, precision and recall were computed.

In addition to class-specific metrics, we also evaluated our model's performance using macro F1, AUC-ROC (Area Under the Receiver Operating Characteristic), and AUC-PR (Area Under the Precision-Recall curve) scores. All three metrics showed promising results, but AUC-PR was the best evaluation criterion for our case.

Our goal was to identify participants who would respond positively, and we placed a higher emphasis on precision over recall. It was crucial to predict with high precision to ensure that we sent notifications to those participants who would respond correctly.

AUC-PR proved to be the most suitable metric for this scenario. Unlike AUC-ROC, which considers the true positive and false positive rates, AUC-PR focuses on precision and recall, making it ideal for imbalanced datasets where the positive class is of greater importance [21].

By using logistic regression as our baseline model and considering the class weights parameter to tackle the class imbalance, we set the stage for the subsequent research questions and the comparison with the performance achieved using encoded features from the autoencoder.

Table I: Logistic Regression performance on Raw (unencoded) data

| Fold | AUC-PR Score | True Negative | False Positive | False Negative | True Positive |
|---|---|---|---|---|---|
| 1 | 0.766229 | 933 | 725 | 1488 | 2214 |
| 2 | 0.772774 | 960 | 697 | 1332 | 2370 |
| 3 | 0.82094 | 1058 | 599 | 1174 | 2528 |
| 4 | 0.756699 | 929 | 729 | 1512 | 2189 |
| 5 | 0.763864 | 981 | 677 | 1604 | 2097 |
| **Best Threshold** | | | 0.334 | | |
| **Mean AUC-PR** | | | 0.7761 | | |

Once the baseline model was established using the appropriate metrics, the subsequent step was to verify whether the model's performance remained consistent when replacing the raw features with the encoded features. As previously mentioned, the encoding process was conducted at the participant level, necessitating the gathering and aggregation of individually encoded participant data into a single dataset for ML model training. Figure 4 illustrates this process.

In the initial iteration, we retained the logistic regression model without modification and simply substituted the raw features with the encoded features, which were only ten in number. We also handled the class imbalance using the same technique of class weights and kept all model parameters consistent. The model was trained using 5-fold cross-validation to assess its performance across participants. For testing, the

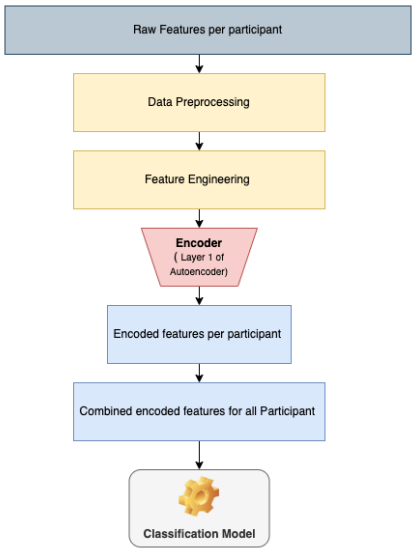


Figure 4: Modeling with encoded features

same test set that was reserved for the non-encoded baseline model was used to evaluate the model's performance.

Remarkably, we observed minimal degradation in the classification metrics and the overall model performance. Precision and recall for both classes remained almost unchanged. This empirical evidence suggested that there was no significant loss in model performance even with the utilisation of the encoded features.

Table II: Logistic Regression performance on Encoded data

| Fold | AUC-PR Score | True Negative | False Positive | False Negative | True Positive |
|---|---|---|---|---|---|
| 1 | 0.718602 | 641 | 684 | 1231 | 1710 |
| 2 | 0.731832 | 670 | 655 | 1185 | 1756 |
| 3 | 0.731067 | 650 | 675 | 1180 | 1761 |
| 4 | 0.732535 | 657 | 668 | 1174 | 1767 |
| 5 | 0.723744 | 638 | 686 | 1172 | 1769 |
| **Best Threshold** | | | 0.3289 | | |
| **Mean AUC-PR** | | | 0.7276 | | |

As a result, we confidently concluded that the encoding of features was highly effective, and the model could predict participant responses with the same level of performance as with the non-encoded features. This finding substantiated the value and quality of the encoded features, affirming their suitability for accurately predicting participant responses to notifications.

Having successfully addressed research questions 1 and 2 with positive outcomes, we were well-positioned to tackle the final and critical research question: *"Will this approach preserve privacy and enable dataset augmentation with sensitive participant information, such as age, medical condition, and gender?"* To assess this, rigorous end-to-end testing of the system was imperative to ensure that user privacy was effectively preserved at all stages of data handling, including data collection, encoding, model training, and the prediction pipeline.

Our paramount concern was to design a system that not only achieved accurate predictions but also upheld the highest standards of data privacy. With this objective, we meticulously mapped out the architecture of the system, carefully integrating privacy-preserving methodologies at every step.

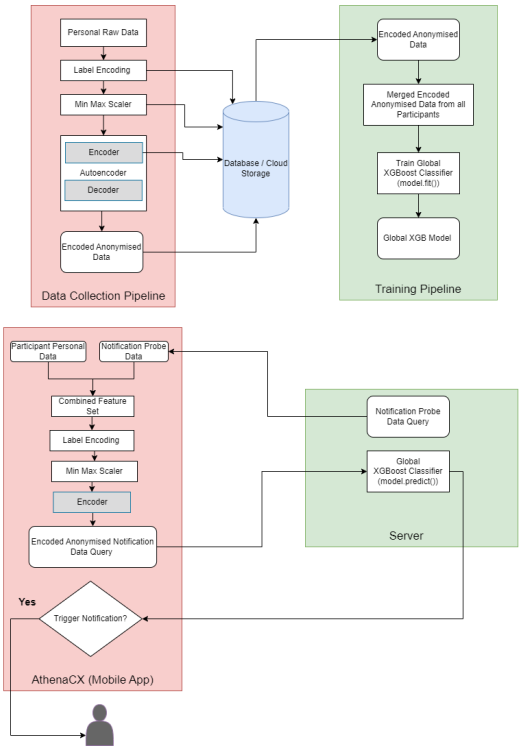


Figure 5: System Architecture

To ensure end-to-end privacy and anonymity, we divided our system into four distinct units: Data Collection Pipeline, Model Training Pipeline, Server, and Edge Device (Mobile App). The entire system was thoughtfully designed with privacy as the central focus.

The initial step in building a predictive model for notification optimisation was data collection. Our system's core principle was to ensure that raw data, potentially containing sensitive information about participants, never left the edge device and was never utilised for modelling. The Data Collection Pipeline diagram illustrates the process.

The participant's personal data underwent label encoding to convert categorical features into numerical representations. Next, we applied Min-Max scaling to standardise the features. Subsequently, an autoencoder was employed to encode the raw data into anonymous and non-human-readable features. The encoded data was labelled as "feature_1", "feature_2", and so on, making it impossible to discern the original raw features used to generate them. As a result, these encoded features carried no information about the participant's personal data, rendering them useless for building any other model or purposes beyond the intended prediction task. This safeguarded the data from any potential breach, as even in the event of a breach, the raw data did not exist within our system, and the stored encoded data was useless to any human.

We securely stored this encoded data in a database, ensuring

that individual privacy was fully protected. This process was repeated for each participant, generating a substantial collection of encoded data. Additionally, we stored the encoder part of the autoencoder model that was responsible for the feature encoding, but we consciously excluded the decoder from the system. The decoder component solely resided on the edge device.

This approach maintained data privacy and confidentiality by not exposing the raw data and relying solely on the encoded features for model training and prediction. The combination of robust anonymisation, feature encoding, and secure data storage assured end-to-end privacy, instilling confidence in the system's ability to preserve participant privacy In the data collection phase.

In the next phase, after accumulating an ample amount of data, we embarked on the Model Training Pipeline, which entailed the training of our predictive model. The participant-specific encoded data, securely stored in the database, was concatenated to form a comprehensive training dataset. Subsequently, a global classification model was trained to utilise this composite dataset. The model's performance was assessed using a robust 5-fold cross-validation technique, employing prominent evaluation metrics such as AUC-ROC, AUC-PR, precision, and recall for each class.

To ensure stringent privacy preservation, we adhered to the principle of not utilising the raw data at any model training stage. Instead, we solely relied on the encoded features derived through the earlier anonymisation process. This strategic approach safeguarded participant privacy while allowing us to develop an efficient and effective global model capable of generating predictions for all participants. The Server and the Mobile Device together form the pivotal components of our system. The Server maintains close communication with the Edge Device, which, in our case, is the Athena CX mobile application. It assumes multiple responsibilities, including hosting the trained classification model, transmitting Notification Probe data to the Athena CX application, and promptly responding to notification query requests from the application.

The Mobile Application holds a central and critical role in safeguarding user data privacy. It serves as the primary entity within the system, meticulously preserving privacy throughout the process. Upon receiving the Notification Probe data, the Mobile Application securely concatenates it with participant data stored on the user's device to form a comprehensive feature set. Subsequently, it executes the data preprocessing pipeline, encompassing label encoding and min-max scaling, followed by the encoder application. This encoder is also locally stored on the user's device.

Following the encoding process, the Mobile Application generates an anonymised notification query comprising all encoded features used during model training. This data is securely transmitted to the server. Upon receiving the notification query, the server initiates the classification model, generating a prediction on whether it deems it appropriate to send a notification to the user. Thereafter, the server sends the response back to the mobile app.

Upon receiving the response from the server, the Mobile Application evaluates whether the prediction is affirmative or negative. In the case of a positive prediction, the application creates a notification pop-up for the user. However, if the response from the server is negative, indicating an inappropriate timing for the notification, the application defers the notification and retries at a more suitable time.

This tightly coordinated interaction between the Server and the Mobile Application exemplifies our commitment to preserving user data privacy while ensuring effective and timely notification delivery. The Mobile Application serves as a secure gateway, protecting user information and supporting seamless communication with the server to generate timely notifications while upholding privacy principles throughout the entire process.

## IV. Results and discussion

In this section, we discuss the results of the modelling approaches employed for encoding and predictive modelling. We utilised a vanilla encoder architecture with a bottleneck layer for the autoencoder model, resulting in 10 encoded features. Figure 3 depicts the learning curves, demonstrating the model's excellent performance as it proficiently learned to encode and reconstruct the data. However, during experimentation, we observed that autoencoders tend to lose signal and the essence of classes when trained on imbalanced datasets. The majority class dominated the encoded features, prompting us to address this issue by training the encoders after downsampling the majority class. This led to improved results, with better preservation of class information.

Moving on to the modelling phase, we explored three different models: logistic regression, XGBoost, and an ensemble model comprising seven classifiers: Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier, SVC, K Neighbors Classifier (neighbors=6), logistic regression, and Naive Bayes. The ensemble model was created by combining these classifiers to leverage their collective predictive capabilities.

For evaluating the models, we adopted a stratified 5-fold cross-validation approach, considering the skewness in the dataset and the primary objective of correctly identifying participants who would respond to the notification (positive class). The Area Under the Precision-Recall Curve (AUC-PR) was chosen as our evaluation metric due to its suitability for imbalanced datasets.

Upon thorough experimentation and hyperparameter tuning, XGBoost emerged as the top-performing model, outperforming logistic regression and our ensemble model. We optimised hyperparameters such as appropriate "scale_pos weight", regularisation, and decision tree depth to achieve an average AUC-PR of 0.77, with a recall of 0.72 and precision of 0.74.

In Figure 6 the chart graphically illustrates the performance comparison among the models, clearly demonstrating XGBoost's superiority. Additionally, Table III provides a de-

Table III: Model Comparison with Metrics

| Model | Metrics | | |
|---|---|---|---|
| | AUC-PR | Precision | Recall |
| Logistic | 0.7276 | 0.72 | 0.59 |
| **XGBoost** | **0.7768** | **0.74** | **0.72** |
| Ensemble | 0.7391 | 0.70 | 0.78 |

tailed breakdown of precision and recall metrics for our best-performing XGBoost model.


Figure 6: Graphical Comparison of Model Performance

The results highlight the effectiveness of the chosen modelling approaches and the importance of addressing the class imbalance in the autoencoder phase. The predictive model's performance, particularly with the XGBoost algorithm, demonstrates its potential in accurately identifying participants likely to respond to notifications, making it a valuable tool for the targeted intervention strategy.

*Robustness Testing and Performance on Cold Start Scenarios*

In addition to the 5-fold cross-validation approach, the model's robustness was tested by evaluating its performance on "cold start" scenarios. This involved testing the model on new participants who were not part of the training data, and simulating real-world scenarios where the model encounters unseen participants for the first time.

Surprisingly, the model showed relatively good performance on these new participants achieving an AUC-PR of 0.728, indicating that it could generalise well to unseen data. However, one notable observation was a decrease in precision for the negative class, which refers to the accuracy of correctly identifying non-responders. This drop in precision suggests that the model might have difficulty distinguishing non-responders among the new participants, which is an area requiring further investigation.

To address the decrease in precision for the negative class, potential strategies include error analysis to identify common misclassifications and feature engineering to better capture the distinguishing characteristics of non-responders. Moreover, exploring more general feature extraction techniques might lead to improved overall performance, as these methods can help extract more relevant information from the data.

## V. Future Scope

The future prospects of this research on privacy-preserving autoencoders for predicting participant responses hold tremendous potential for progress and innovation. While the current global model relies on XGBoost, exploring more sophisticated models like deep learning architectures and advanced ensemble methods could lead to significant improvements in predictive performance while upholding data privacy. A critical element in understanding participants in this study is RFM (Recency, Frequency, Monetary) analysis, a customer segmentation technique commonly used in marketing and analytics [22]. RFM provides valuable insights into participants' behaviours and preferences by evaluating the recency of their interactions (R), the frequency of their responses (F), and the monetary value associated with their engagement (M). By conducting RFM analysis on the raw data before encoding it with the autoencoder, researchers can gain a comprehensive understanding of participant's characteristics without compromising their data privacy.

To ensure the approach's scalability and robustness, it can be extended to encompass a longer longitudinal study, incorporating local autoencoder models for each participant. Implementing Federated Learning with an SGD classifier offers collaborative model training opportunities across multiple participants while preserving privacy [23]. The SGD classifier enables the secure transmission of local model weights to the global model, harnessing the collective knowledge of all participants for model training without exposing individual raw data. Additionally, performing error analysis on participant classes will provide invaluable insights into model limitations and potential biases, guiding targeted improvements to enhance overall performance. Building a regression model to predict participant response times and prioritising feature engineering will further optimise study engagement strategies and refine the representation of participant data. By addressing these future scope points, this research will drive the advancement of privacy-preserving machine learning, offering accurate and personalised predictions while prioritising the paramount importance of safeguarding participant's privacy and data security.

## VI. Conclusion

This paper emphasises the utmost importance of safeguarding data privacy in Electronic Patient-Reported Outcome (ePRO) systems and personalised notifications. The driving force behind this project is the delicate balance required between promoting patient engagement and protecting their sensitive health information. These digital platforms have emerged as indispensable tools in modern healthcare, empowering patients to actively report their health status and quality of life [1]. However, to fully harness the potential of ePRO systems, it is imperative to address concerns surrounding the collection and utilisation of sensitive health data. The growing interest of insurance companies in accessing personal data and the persistent occurrence of data breaches raise serious concerns about patient's trust in technology-driven healthcare solutions.

To address these challenges, our machine learning framework is designed to primarily focus on preserving patient data privacy. We employ secure encoding and de-identification

techniques to ensure that sensitive data leaving mobile devices remains non-human readable and shielded from any potential misuse or unauthorised access. By harnessing our model's predictive capabilities to deliver notifications at optimal times for enhanced patient engagement [6], we aim to create an environment where patients can confidently adopt AI-driven solutions in their healthcare journey [24]. Emphasising a privacy-first approach, ePRO systems can evolve into indispensable assets within the healthcare industry. Patients can confidently participate in clinical trials, knowing that their personal information is rigorously safeguarded. Furthermore, healthcare professionals can access real-time patient data, facilitating informed decision-making without compromising patient privacy. As we embrace a future driven by AI and technology, this seamless integration of privacy and personalised care becomes paramount to achieving better patient outcomes, improved data quality, and optimal utilisation of healthcare resources. Through our framework, we strive to usher in a new era of responsible data usage, transparency, and patient-centricity in healthcare, paving the way for a revolutionary transformation in patient and healthcare professional interactions within clinical trials and beyond [9].

*Acknowledgement:*

We are deeply grateful to Professor Tomás Ward and Mr. Himanshu Vashisht for their guidance, support, and valuable feedback throughout this research project. Their expert insights and thoughts have significantly enriched our work, and their suggestions have made our paper stronger and more rigorous.

We would also like to thank In the Wild Research for providing us with the data that made this research possible. Their data was invaluable to our study, and we are grateful for their willingness to share it with us. [25]

## REFERENCES

[1] C. J. van den Hurk, F. Mols, M. Eicher, R. J. Chan, A. Becker, G. Geleijnse, I. Walraven, A. Coolbrandt, M. Lustberg, G. Velikova, *et al.*, "A narrative review on the collection and use of electronic patient-reported outcomes in cancer survivorship care with emphasis on symptom monitoring," *Current Oncology*, vol. 29, no. 6, pp. 4370–4385, 2022.

[2] M. L. A. Lustria, S. M. Noar, J. Cortese, S. K. Van Stee, R. L. Glueckauf, and J. Lee, "A meta-analysis of web-delivered tailored health behavior change interventions," *Journal of health communication*, vol. 18, no. 9, pp. 1039–1069, 2013.

[3] M. Bardus, H. Blake, S. Lloyd, and L. Suzanne Suggs, "Reasons for participating and not participating in a e-health workplace physical activity intervention: A qualitative analysis," *International Journal of Workplace Health Management*, vol. 7, no. 4, pp. 229–246, 2014.

[4] L. Dennison, L. Morrison, G. Conway, L. Yardley, *et al.*, "Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study," *Journal of medical Internet research*, vol. 15, no. 4, p. e2583, 2013.

[5] A. L. Marshall, "Challenges and opportunities for promoting physical activity in the workplace," *Journal of Science and Medicine in Sport*, vol. 7, no. 1, pp. 60–66, 2004.

[6] F. Künzler, J.-N. Kramer, and T. Kowatsch, "Efficacy of mobile context-aware notification management systems: A systematic literature review and meta-analysis," in *2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 131–138, IEEE, 2017.

[7] S. T. Iqbal and E. Horvitz, "Disruption and recovery of computing tasks: field study, analysis, and directions," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 677–686, 2007.

[8] D. C. McFarlane, "Comparison of four primary methods for coordinating the interruption of people in human-computer interaction," *Human-computer interaction*, vol. 17, no. 1, pp. 63–139, 2002.

[9] P. D. Adamczyk and B. P. Bailey, "If not now, when? the effects of interruption at different moments within task execution," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 271–278, 2004.

[10] A. Mehrotra, M. Musolesi, R. Hendley, and V. Pejovic, "Designing content-driven intelligent notification mechanisms for mobile applications," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 813–824, 2015.

[11] B.-J. Ho, B. Balaji, M. Koseoglu, and M. Srivastava, "Nurture: notifying users at the right time using reinforcement learning," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pp. 1194–1201, 2018.

[12] C. Araújo, C. Soares, I. Pereira, D. Coelho, M. Â. Rebelo, and A. Madureira, "A novel approach for send time prediction on email marketing," *Applied Sciences*, vol. 12, no. 16, p. 8310, 2022.

[13] A. Deligiannis, C. Argyriou, and D. Kourtesis, "Building a cloud-based regression model to predict click-through rate in business messaging campaigns," *International Journal of Modeling and Optimization*, vol. 10, no. 1, pp. 26–31, 2020.

[14] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, *et al.*, "The future of digital health with federated learning," *NPJ digital medicine*, vol. 3, no. 1, p. 119, 2020.

[15] W. Li, F. Milletarì, D. Xu, N. Rieke, J. Hancox, W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M. J. Cardoso, *et al.*, "Privacy-preserving federated brain tumour segmentation," in *Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10*, pp. 133–141, Springer, 2019.

[16] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*, pp. 92–104, Springer, 2019.

[17] D. Cha, M. Sung, Y.-R. Park, *et al.*, "Implementing vertical federated learning using autoencoders: Practical application, generalizability, and utility study," *JMIR medical informatics*, vol. 9, no. 6, p. e26598, 2021.

[18] D. Novoa-Paradela, O. Fontenla-Romero, and B. Guijarro-Berdiñas, "Fast deep autoencoder for federated learning," *Pattern Recognition*, p. 109805, 2023.

[19] Q. Meng, D. Catchpoole, D. Skillicom, and P. J. Kennedy, "Relational autoencoder for feature extraction," in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 364–371, 2017.

[20] Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," *Neurocomputing*, vol. 184, pp. 232–242, 2016.

[21] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd international conference on Machine learning*, pp. 233–240, 2006.

[22] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "Rfm ranking–an effective approach to customer segmentation," *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 10, pp. 1251–1257, 2021.

[23] R. Rakhmiddin and K. Lee, "Federated learning for clinical event classification using vital signs data," *Multimodal Technologies and Interaction*, vol. 7, no. 7, p. 67, 2023.

[24] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nature medicine*, vol. 25, no. 1, pp. 44–56, 2019.

[25] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," *CoRR*, vol. abs/1807.00459, 2018.