Differential Privacy based on Bayesian Optimisation in Deep Neural Networks

Jiling Zhou
Cyber Security Analytics
University of Exeter, UK
jz525@exeter.ac.uk

Dr Tinkle Chugh (*Supervisor*)
Department of Computer Science
University of Exeter, UK
T.Chugh@exeter.ac.uk

***Abstract*—The objective of this project is to explore the integration of local differential privacy into deep neural networks (DNNs) through the utilisation of Bayesian optimisation. The purpose of this research is to achieve a trade-off between preserving privacy and maintaining the effectiveness of the model, with a specific focus on medical applications. By implementing a three-hidden-layer feed-forward neural network (FFNN) encompassing three levels (Normal, Prediabetes, and Diabetes), we have successfully classified a dataset related to diabetes for the intent of diagnosis. The Laplace mechanism is applied in conjunction with local differential privacy to enhance the privacy assurance of FFNN. This method provides an efficient way of safeguarding the privacy of data samples, while simultaneously attaining optimal model accuracy. Furthermore, we illustrate the benefits of using the Laplace mechanism in preserving privacy compared to the Gaussian mechanism. In addition, we confirm that the integration of differential privacy in DNNs is capable of producing precise predictions while upholding privacy. Additionally, the experiment demonstrates the efficacy of Bayesian optimisation in the context of multi-objective optimisation within privacy scenarios, while also identifying the most optimal hyper-parameter configurations. This approach provides evidence that the utilisation of Bayesian optimisation leads to enhanced accuracy in the FFNN model, as well as in the Laplace and Gaussian mechanisms (privacy). We put forward an argument regarding the limited influence of Bayes's theorem on the K-anonymity mechanism. This work offers a comprehensive framework for enhancing the efficiency and structure of DNNs in order to classify medical data while ensuring enhanced privacy guarantees.**

***Index Terms*—Deep neural networks (DNNs), Differential privacy, Feed-forward neural network, Laplace mechanism, Bayesian optimisation**

I. Introduction

Machine learning is extensively employed in numerous medical applications that involve confidential data, which cannot be lawfully disclosed due to privacy considerations. The importance of privacy-preserving technologies is increasing in the effort to address this problem. Differential privacy is a widely recognised approach employed in the field of machine learning that aims to safeguard individual data points by adding random noise into the dataset. However, the procedure of adjusting hyper-parameters in machine learning models necessitates the execution of numerous data queries, which has the potential to adversely affect data privacy. This research project investigates the integration of differential privacy, and multi-objective optimisation techniques utilising Bayesian optimisation within deep neural networks (DNNs). The overarching objective is to preserve privacy while enhancing the overall accuracy of machine learning models.

Currently, machine learning experiences significant challenges due to various types of privacy attacks [1]. There are two noteworthy occurrences that merit attention: Membership inference attacks [2] refer to the malicious attempts made by adversaries to identify specific data points and ascertain whether they are part of the training dataset used to build a model. 2) Reconstruction Attacks [3] indicate instances where the opponent leverages the behaviour of the model on different inputs to reconstruct either the entire training dataset or a specific element of it. The increasing amount and importance of medical data have raised concerns regarding potential privacy risks. The aforementioned attacks possess the potential to expose sensitive data, thereby constituting a breach.

Nevertheless, there exist numerous technical challenges that need to be addressed in order to enhance the efficacy of data privacy preservation in such situations. Differential privacy is a well-recognised methodology implemented in the field of machine learning to ensure the preservation of privacy. Differential privacy has emerged as a prominent method for developing machine learning algorithms that uphold privacy. One of the methods to introduce differential privacy in machine learning algorithms is utilising a mathematical technique to safeguard privacy by introducing a suitable level of noise into the input dataset. The incorporation of differential privacy into the machine learning modelling process entails the introduction of noise with an optimisation procedure, thereby safeguarding the privacy of the data. Differential privacy encounters various challenges encompassing issues related to data quality and precision, the delicate balance between safeguarding privacy and ensuring model performance, as well as the preservation of privacy during hyper-parameter optimisation. The process of optimising the hyper-parameters of a machine learning model requires multiple data queries, which could pose a threat to the privacy of individual information.

Bayesian optimisation [7] is an approach for global optimisation of black-box functions that implements Bayesian machine learning techniques and Gaussian regression to quantify the uncertainty in the objective surrogate and determine sampling points. Bayesian optimisation determines the optimal

set of hyper-parameters for any particular machine learning model while reducing the total amount of data queries in order to safeguard data privacy. Consequently, the implementation of differential privacy in Bayesian optimisation could result in some issues, particularly when multiple conflicting objectives are involved or when an excessive amount of noise is added, resulting in unbalanced or inconsistent accuracy during the optimisation process.

In order to address these challenges, the utilisation of multi-objective optimisation [8,9] becomes increasingly significant. Mathematical optimisation involves the simultaneous satisfaction of multiple conflicting objectives in order to identify an optimal solution that encompasses all objectives. The feasibility of automating the tuning of model hyper-parameters is enhanced by the incorporation of multi-objective optimisation into Bayesian optimisation with differential privacy. The combination of features described enables the effective examination of the hyper-parameter space, taking into account both privacy assurances and model accuracy. The following offers a potentially advantageous resolution for the current undertaking.

The primary objective of this research project is to make a scholarly contribution towards the advancement of a privacy-preserving approach for hyper-parameter optimisation in DNNs applied to medical datasets. The project is designed to expand the researcher's comprehension of the privacy implications associated with hyper-parameter optimisation. Additionally, it seeks to propose a practical and efficient solution to effectively tackle these challenges. The research findings will bear substantial ramifications for a diverse array of domains, encompassing healthcare, finance, and social media, wherein the preservation of individuals' privacy within the data holds paramount importance. The project is driven by the following primary objectives:

1) Conduct a comprehensive review of the existing literature on differential privacy, Bayesian optimisation, and multi-objective optimisation.
2) Implement a classification task on a medical dataset with a deep neural network to solve a multi-class problem. The objective is to apply and train the accurate deep neural network model as a foundation for subsequent privacy-preserving experiments.
3) Apply differential privacy in combination with the deep neural network.
4) Find the optimal set of hyper-parameters for the privacy-preserving FFNN model through multi-objective Bayesian optimisation, with the aim of maximising the predictive accuracy of the model while satisfying strict differential privacy constraints.

The remaining part of this project is broken into six major chapters: In the second chapter, a comprehensive review of the essential articles is presented on the topic at hand. The details of the dataset employed within this experiment were presented in the third chapter. We explained complete experimental procedure in the fourth chapter. Moreover, the fifth and sixth chapters provided a full overview (results) and discussion of this research effort. The last chapter provides an overall summary of this project.

II. Research Context

The following part presents a comprehensive overview of the research and key academic publications pertaining to Bayesian optimisation and differential privacy in deep neural networks. The aim is to supply theoretical support and guidance for the current project. This literature review primarily examines three significant fields of research, namely differential privacy, Bayesian Optimisation, and multi-objective selection.

Differential privacy is a fundamental technique implemented to safeguard privacy during the processing of data. Dwork C presented the conceptualization of differential privacy, elucidated its fundamental characteristics, and explored its connection with additional privacy definitions [10]. Additionally, various methodologies for attaining differential privacy were examined, such as data disruption through the introduction of noise. This work provided theoretical substantiation for the aforementioned project. In addition, the survey article by Dwork C comprehensively addressed diverse applications of differential privacy, elucidating the significance of privacy loss parameters and privacy budgets in safeguarding data privacy during the process of analysis [5]. Furthermore, the approach contributed forth by McSherry F and Talwar K investigates the concept of differential privacy as a metric for quantifying the extent of privacy compromise in systems that handle private data, thereby enhancing the level of privacy assurance [6]. Shuying Qin et al. stated a noise addition method to optimise random noise by investigating discretely dispersed data in order to accomplish a trade-off between privacy and performance [40]. Dwork C provided a distributed approach to differential privacy and introduced the concept of distributed noise generation, with an emphasis on studying the (δ, ϵ) differential privacy objective in this proposal [11]. The prevailing approach involves the utilisation of differential private stochastic gradient descent (DP-SGD) as a means of achieving differential privacy. The following happens through the modification of the gradient working as the stochastic gradient descent algorithm. DP-SGD incorporates four privacy parameters, namely the maximum Euclidean norm of the gradient (L2), the quantity of noise added, the training batch size, and the learning rate [12-13].

Bayesian optimisation has been recognised as a method for improving the parameter tuning process in deep neural networks. Rasmussen and Williams made significant contributions to the advancement of Bayesian optimisation in the field of machine learning through the introduction of Gaussian processes and probabilistic models [14]. Moreover, the researchers Gan W et al. investigated various aggregate functions within the context of Bayesian optimisation, thereby making a valuable contribution to the progress of optimisation techniques for deep neural networks [15]. In this project, we delivered a tutorial and a fundamental understanding of Bayesian optimisation. Additionally, the efficacy of Bayesian

optimisation is extensively examined in the identification of optimal parameters within machine learning models, encompassing deep neural networks [7].

Multi-objective optimisation provides theoretical foundations and guidance for implementing the use of multi-parameter optimisation in deep neural networks. A multi-objective optimisation is a computational approach that aims to identify the most effective approach that satisfies multiple conflicting objectives concurrently. Kaisa Miettinen showcased a range of solution methods for addressing the challenges of nonlinear multi-objective optimisation [16]. Konak A presented a tutorial that focuses on the application of genetic algorithms for multi-objective optimisation [17]. In addition, Zhao F contributed a new strategy that combines multi-objective reinforcement learning to address the challenges posed by distributed optimisation problems [18]. The results of the research presented in this study highlight the significance of employing multi-objective optimisation techniques to address intricate optimisation challenges encountered in DNNs.

Furthermore, the incorporation of differential privacy into Bayesian optimisation plays a vital role in safeguarding privacy within machine learning models. Kusner M introduced a method that incorporates differential privacy into the acquisition function for Bayesian optimisation [19]. This methodology involves introducing noise to the objective function using the Laplace mechanism. Nguyen T D implemented a privacy-preserving Bayesian optimisation for the purpose of designing search spaces with high dimensions. This approach aims to minimise the number of queries made to the database [20]. Tobaben conducted a study on the trade-off between accuracy and privacy in differential privacy deep learning models [21]. In their research, they put forth a method for preserving privacy during hyper-parameter tuning. Abadi M provided a method for incorporating differential privacy into DNNs in order to ensure privacy during the training process of these networks using sensitive data [4]. In their research, Fan T and Cui Z have presented a novel method for safeguarding privacy in deep neural network learning [22]. This approach, which is rooted in multi-objective optimisation, is designed to adaptively incorporate differential privacy measures. The research findings presented in this study highlight the significance of incorporating differential privacy techniques in the field of machine learning. Furthermore, these outcomes serve as a valuable resource for exploring various combinations of differential privacy techniques and Bayesian optimisation to effectively maintain privacy in the project.

III. Dataset and Resources

The experiment conducted in this dissertation utilised the Diabetes Dataset [23] sourced from Mendeley data [24], specifically the 2020 release version 1. The dataset used in this study was obtained from a sample of 1000 patients, collected by the laboratory of Medical City Hospital and the Specialised Centre for Endocrinology and Diabetes-Al-Kindy Teaching Hospital in Iraq. The Mendeley data platform is a freely accessible research database website that facilitates the sharing and citation of research data. The dataset available on this platform has undergone de-identification processes to ensure the removal of patient identification and sensitive data. Consequently, it does not include any individually identifiable data points or sensitive data associated with patients [25-26]. In contrast to the MIMIC dataset [27], the dataset under consideration exclusively comprises structured and valid numerical data pertaining to individuals diagnosed with diabetes. It is devoid of any unstructured or invalid data, thereby ensuring that the dataset is suitably prepared for subsequent model training. It plays a significant role in the achievement of a favourable result in the experiment. The dataset comprises 14 columns, including the number of patients, blood sugar levels, age, gender, and various medical and laboratory analysis data pertaining to symptoms associated with Diabetes mellitus. Each row within the dataset corresponds to a distinct patient.

IV. Experiment Design & Methods

The project is divided into four parts. Initially, The diabetes dataset is processed and cleaned in the Data preparation stage. DNN models are constructed and trained for multi-classification tasks under subsection *The implementation of classification in DNNs*. Differential privacy is presented in the subsection *The implementation of differential privacy* to increase privacy while retaining accuracy in the machine learning model. Finally, in the subsection of *The implementation of Bayesian optimisation*, Bayesian optimisation is utilised to identify the ideal hyper-parameter configuration for privacy-preserving DNN models.

A. *Data Preparation*

Data preparation is the initial and essential stage in this experiment. The primary objective of data preparation is to effectively cleanse, manipulate, and enrich data, as the quality of data has a direct impact on the precision and reliability of subsequent research endeavours. The process of data preparation encompasses six distinct steps: data cleanup, labelling and sorting, feature selection, data shuffling, normalisation, and dataset splitting. **Data cleansing** [28], also known as data cleaning or data scrubbing, is a crucial process in which inaccurate, corrupt, duplicate, or incomplete data within a dataset is corrected, removed, or otherwise addressed. This is typically achieved through various methods such as pruning or deletion. There exist multiple techniques for data cleansing and preparation, including the elimination of duplicate or extraneous observations, handling missing data, rectifying structural errors, and other similar procedures. In the current dataset, two unnecessary columns are dropped to make the dataset with 12 significant columns. **Labeling and sorting.** The process of assigning a distinct category (class) to a given dataset is referred to as labelling [29]. In supervised machine learning, the model has the ability to acquire knowledge from a dataset that contains labelled instances. This knowledge enables the model to establish a relationship between the input features and the corresponding output labels. Consequently, the trained model can make precise predictions regarding the

labels of new input data. When considering the column named "HbA1c," which serves as one of the features, no modifications were made to any metrics. Initially, the data is organised in an ascending manner based on the 'HbA1c' value. Three categories are determined based on the 'HbA1c' values: those below 5.7 are assigned label 1, those between 5.7 and 6.4 are assigned label 2, and those above 6.4 are assigned label 3 [33]. These labels are subsequently converted into binary vectors, denoted as target1, target2, and target3, respectively. **Feature selection** is a widely recognised practice in machine learning predictive modelling, wherein the number of input variables is restricted [30]. The process of feature selection is commonly employed to eliminate irrelevant features from a given dataset, thereby retaining only the relevant features. In the current dataset, all 12 significant features are incorporated after the feature selection. **Data shuffling** is the process of randomly reordering the data samples in a dataset to assure randomization. **Normalization** refers to the process of scaling and converting numerical features in order to bring them to a consistent scale. This is done with the aim of enhancing the convergence of the algorithm. In this study, we apply a normalisation technique to the dataset in order to transform it into discrete numerical values as part of the data preprocessing phase. **Splitting dataset.** During the Split dataset stage, the features (Input) and labels (Target) are segregated and subsequently stored in variables 'X' and 'y', respectively. The dataset is then divided into two subsets: 80% of the data is allocated for training the model, while the remaining 20% is reserved for testing purposes. The test set serves as a means to assess the performance of the trained model on previously unseen data.

B. The implementation of classification in DNNs

The following section will provide an analysis of the classification problem addressed in the experiments and present a comprehensive overview of the application of feed-forward neural networks for achieving multi-class classifications. The accuracy and effectiveness of the model in addressing the classification task will be assessed through the utilisation of data visualisation techniques.

1) Three class problem of HbA1c

Glycated hemoglobin, referred to as HbA1c, is a form of hemoglobin that undergoes a chemical bonding process with sugar [31]. The HbA1c metric is employed as a means of quantifying elevated blood glucose levels and is frequently utilised as a diagnostic tool for diabetes [32]. In the current research, HbA1c was employed as a significant evaluation metric for training the dataset. Based on the diagnostic criteria for diabetic HbA1c established by the American Diabetes Association, a normal level is diagnosed when HbA1c is less than 5.7%. Prediabetes is diagnosed when HbA1c falls within the range of 5.7% to 6.4%. Diabetes is diagnosed when HbA1c exceeds 6.4% [33]. Therefore, we classify HbA1c into three categories in this analysis: Class 1 contains values within the normal range of diabetes and corresponds to Target 1; Class 2 contains values within the range of diabetes and corresponds to Target 2; Class 3 contains values within the range of diabetes and corresponds to Target 3. We correspond to Target 3, and the detailed list is as follows:

TABLE I: The classification of HbA1c

Number	Target	HbA1c Value	Diagnosis	Class
1	Target 1	$< 5.7\%$	Normal	Class 1
2	Target 2	5.7% ~ 6.4%	Prediabetes	Class 2
3	Target 3	$> 6.4\%$	Diabetes	Class 3

2) Definition of the Feed-forward Neural Network

In this experiment, the diabetes dataset utilised is a standard medical dataset presented in tabular format. Each row represents a sample, while each column represents an attribute. The dataset possesses appropriate numerical features for classification, and its data structure is deemed satisfactory. Therefore, considering the comprehensive selection criteria encompassing nature and complexity, the feed-forward neural network model (a subset of DNNs) is defined and implemented in this section. One significant benefit of utilising feed-forward neural networks for classification tasks is their strong capacity to recognise patterns in tabular and numerical data, thereby enabling their application to more intricate datasets.

The Keras and TensorFlow libraries are used for the purpose of defining and training feed-forward neural networks in order to address multi-classification tasks. The feed-forward neural network is designed to process 12 input feature vectors in the input layer and classify them into three classes: Normal, Prediabetes, and Diabetes, in the output layer. The model architecture encompasses various components, such as hidden layers, activation functions, and Dropout. The number of hidden units chosen in the three hidden layers is 100, 50, and 15, respectively (DNN). This choice is motivated by the relatively simple and small size of the dataset. The dataset consists of a collective of 12 input features and is divided into three distinct categories, as depicted in Figure 1. The ReLU activation function, denoted as f(x) = max(0, x), guarantees that any negative values in the layer's output are transformed to zero, while positive values remain unaltered. The ReLU activation function is employed in the three hidden layers of the feed-forward neural network to facilitate the acquisition of the nonlinear mapping between the data and the hidden representation, thereby enabling expedited convergence. Dropout is a regularisation technique that incorporates stochasticity. During the training stage, a random selection of neurons is deliberately discarded in order to mitigate the risk of overfitting and enhance the model's capacity for generalisation. During the model compilation phase, the loss function 'categorical_crossentropy' is employed to address the classification problem of encoding the target label. This loss function quantifies the disparity between the predicted probability and the actual category label. The optimiser is responsible for accelerating the model's convergence by modifying the learning rate. In this section, the 'Adam' algorithm is used for optimisation. The important indication of this experiment is accuracy, which is used to measure the classification accuracy of the model throughout training and testing. The training data

is used to update the model's parameters during the model training stage, while the testing data is used to assess the model's performance. The epoch parameter sets the number of times the model iterates through the full training dataset.

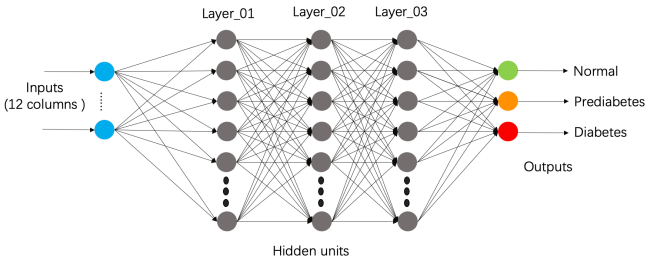


Fig. 1: FFNN structure with hidden layers

*3) **Evaluation of model***

The classification accuracy of the model is assessed at the model testing step by comparing the predicted class labels with the real class labels in the test dataset by inference using a trained DNN. The proportion of properly categorised samples in the overall test dataset is represented by the accuracy value. In general, the higher the accuracy, the better the performance of categorising test data into the proper class is considered. In terms of classification, we discovered that the accuracy of the three class-FFNN classification is about 97.5%.

- **The confusion matrix of Diabetes dataset.** According to the confusion matrix, class 3 (Diabetes) contributes more to overall accuracy than the other two classes. The confusion matrix's diagonal represents the total percentage of correctness with regard to the classifications.
- **The ROC curve of Diabetes dataset.** Judging by the area of the ROC curve, in the diabetic dataset, class 1 and class 3 data are more accurate than class 2. The rate of false positives in class 2 is greater than in classes 1 and 3.

Figure 2 illustrates the results of classification (without privacy) with the help of (a) The confusion matrix and (b) The ROC curve:

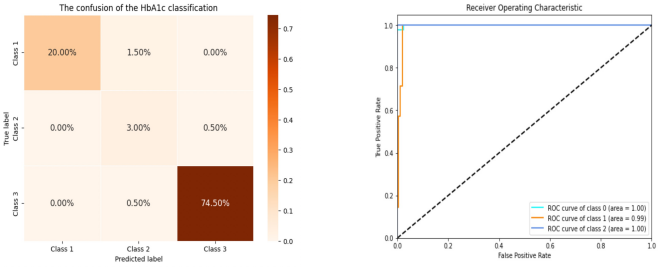


Fig. 2: (a) The confusion matrix (b) The ROC curve

In terms of a classification problem, the achieved accuracy is 97.5%. The utilisation of a uniform dataset for the three classes may assist in increasing the accuracy of the model. Moreover, the selected features contribute to the classification process by enhancing the accuracy of the results obtained. Acquiring a standardised dataset for the three classes poses a significant challenge. The entirety of the diabetes dataset was employed by the author in order to ensure an adequate number of observations for each class. The confusion matrix and receiver operating characteristic (ROC) curves provide an assessment of the classification model's accuracy in relation to all three classes. Due to the limited size and exceptional quality of the dataset, the accuracy rate is notably elevated. However, it is still possible to examine the influence on accuracy by investigating the quantity of hidden layers and the number of hidden units within each respective hidden layer.

C. The implementation of differential privacy

This section mainly focuses on the application of the Laplace mechanism within the context of local differential privacy, with the aim of augmenting the privacy assurance of the feed-forward neural network. The primary objective is to conduct an analysis of the epsilon (ϵ) value in order to examine the impact of the privacy budget on the dispersion of the dataset. Furthermore, the accuracy of the feed-forward neural network is computed and assessed using various techniques for introducing noise. The purpose of investigating these factors is to acquire a deeper understanding of how privacy-preserving technical mechanisms affect the performance of the model.

*1) **Selecting the Laplace mechanism of local differential privacy***

In the following subsection, we rely on the Laplace mechanism of local differential privacy to offer privacy assurances for the FFNN model. Additionally, we evaluate the privacy budget of differential privacy by examining the model's accuracy.

Definition 1.1 (Local differential privacy [10, 34]). Based on the analysis of the dataset in this research, the selection of (ϵ, δ) differential privacy [11] was made due to its appropriateness for situations where a slight likelihood of privacy breach is deemed acceptable, while simultaneously offering a less stringent level of privacy safeguarding:

$$P_r[\kappa(D_1 \in S)] \leq exp(\epsilon) \times P_r[\kappa(D_2 \in S)] + \delta \qquad (1)$$

The privacy budget, denoted as epsilon (ϵ), serves as a metric for quantifying the level of privacy protection. It accomplishes this by constraining the likelihood of a specific model output. Generally, smaller values of ϵ offer stronger guarantees of privacy. Conversely, larger values of epsilon indicate weaker privacy protection. The delta (δ) is utilised to quantify the likelihood of privacy protection mechanism failure. Typically, a smaller value of δ corresponds to a reduced risk of privacy breach. The experimental range for the value of epsilon (ϵ) in this study spans from 0.1 to 10.

Definition 1.2 The theory of sensitivity is employed in evaluating the level of noise necessary for ensuring privacy protection. It serves as a metric for quantifying the extent of fluctuation observed in the outcomes of queries. The sensitivity of the query function $f : D \longrightarrow \mathbb{R}^K$, is defined as:

$$Sensitivity = max_{D_1,D_2} ||f(D_1 - D_2||_1 \qquad (2)$$

In general, the sensitivity of a counting query is 1 for all neighbouring datasets D_1 and D_2. The sensitivity value for this experiment was recorded as 1.

Definition 1.3 (Laplace mechanism [36]). The Laplace distribution is a probability distribution commonly employed for the purpose of generating random noise. The Laplace Mechanism introduces noise to the output of the query for a function $f : D \longrightarrow \mathbb{R}^K$:

$$f(D) + Lap(0, \frac{Sensitivity}{\epsilon}) \tag{3}$$

The notation $Lap(0, \frac{Sensitivity}{\epsilon})$ represents a stochastic variable that follows the Laplace distribution with a mean of 0 and a scale parameter of $Sensitivity/\epsilon$. The parameters associated with privacy loss encompass epsilon (ϵ), delta (δ), sensitivity, dataset size, and query complexity. It can be inferred that the assessment of the privacy budget in the context of local differential privacy holds significant importance for the Laplace mechanism, as it directly influences the level of noise introduced into the data and handles the balance between preserving privacy and maintaining data accuracy.

*2) **Adding Laplace nose to the diabetes dataset***

This section presents a comprehensive analysis of safeguarding data privacy during the training of feed-forward neural network models on the diabetes dataset. During the experimental session, the initial step involves the establishment of two fundamental parameters, namely epsilon (ϵ) and sensitivity. Furthermore, the Laplace mechanism is applied to introduce Laplace noise to each component of the entire dataset. This is accomplished by applying the add_noise() function to each data point in the dataset, thereby ensuring the desired objective of achieving differential privacy. The diabetes training dataset is augmented with noise in order to safeguard data privacy before the process of training the model. Four distinct samples of epsilon (ϵ) values were utilised in our experimental study, specifically 0.1, 1, 5, and 10. The primary objective was to examine the impact of varying epsilon (ϵ) on the generation of noise. In order to support the analysis, the present data visualisation pertains to the *AGE* variable within the diabetes dataset, serving as the observed data. The outcomes of noise addition are depicted in Figure 3 and Figure 4.

- **The violin plot of adding Laplace nose to the diabetes dataset.**
  As shown in Figure 3, it is evident that the violin plot exhibits a shortage of consistency with the original data sample when ϵ = 0.1, whereas when ϵ = 10, the violin plot demonstrates an adequate amount of consistency with the original data sample. Therefore, it can be inferred that as the value of epsilon (ϵ) increases, the proximity to the original data sample also increases.
- **The scatter plot of adding Laplace nose to the diabetes dataset.**
  Based on the scatter plot presented in Figure 4, it can be observed that variations in the value of epsilon (ϵ) have an impact on the intensity of the introduced noise.

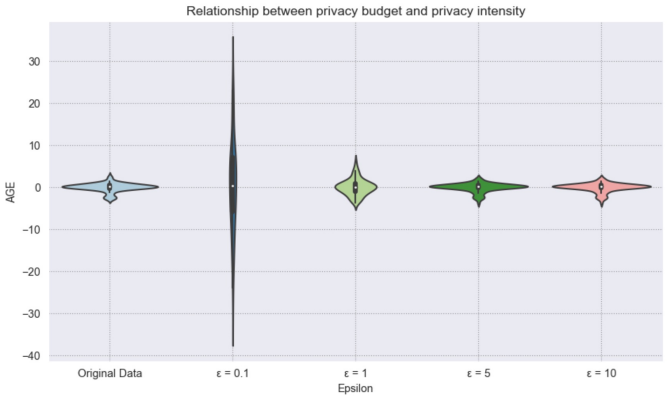


Fig. 3: The violin plot of adding Laplace noise

When the value of ϵ decreases, the intensity of the added noise grows, whereas when the value of ϵ increases, the intensity of the added noise drops. According to terms of

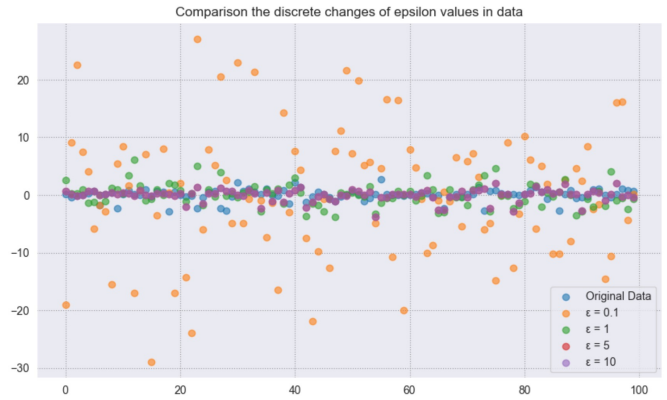


Fig. 4: The scatter plot of adding Laplace noise

incorporating Laplace noise, the approach employed is that of local differential privacy, wherein the sensitivity parameter is maintained at a value of 1. The epsilon (ϵ) value was manipulated in order to examine the impact on noise variation. The findings of our study suggest that a decrease in the epsilon (ϵ) value is associated with a greater enhancement of privacy preservation. Conversely, as the value of epsilon (ϵ) increases, the level of privacy protection diminishes. This experiment demonstrates the significance of the parameter epsilon (ϵ) in differential privacy. A smaller value of ϵ can offer a stronger guarantee of privacy. However, it also provides additional noise, which directly impacts the accuracy of the dataset. Consequently, this can influence the outcomes of subsequent machine learning training. Hence, it is imperative to undertake a more comprehensive examination of the methods to successfully achieve a harmonious equilibrium between privacy preservation and data accuracy.

*3) **Comparison of different kinds of noise***

Within this particular section, we posit that the category of noise holds significant influence over the magnitude of

the noise present, and further assert that distinct types of noise yield varying impacts on the dataset that has undergone training. Hence, in the conducted experiment, Gaussian noise and K-anonymity are incorporated into the dataset for the feed-forward neural network model. The objective is to assess their impact on the accuracy of the model, in comparison to the utilisation of a single Laplace noise.

Definition 1.4 (Gaussian mechanism). The Gaussian mechanism [37] is a commonly employed method for introducing noise. This method follows the Gaussian distribution and applies random noise to it. Specifically, the Gaussian mechanism adds noise to the output of the query in the following manner:

$$f(D) + N(0, \sigma^2) \tag{4}$$

Where δ is the deviation, the privacy amplification parameter is $Sensitivity/\delta$.

K-anonymity: The process of K-anonymity [38] is applied as a data anonymization method. This study exploits the approach of K-anonymity grouping to construct a K-anonymity dataset. This involves duplicating each row of the original dataset K times and generating a new dataset from these replicated rows. The objective is to ensure that each row in the resulting dataset cannot be distinguished from at least K-1 other rows, thereby rendering the identification of individual records difficult.

- **The scatter plot of different ways of adding noise**
  The scatter plot in Figure 5(a) demonstrates that manipulating the privacy parameter to its maximum value has a notable impact on the feed-forward neural network model in this experiment, as evidenced by the various types of noise addition. Regrettably, our experiment did not yield any significant correlation in the ordering of these noise methods when the privacy parameter was set to the same scale. There is no evidence of substantial differentiation observed between the application of Laplace noise and Gaussian noise with equivalent scales. K-anonymity does not exclusively rely on the usage of additive noise as a privacy preservation technique. Instead, it allows for data manipulation to enable comparisons while maintaining privacy. However, it should be noted that the comparisons made under the framework of K-anonymity may lack substantial significance.

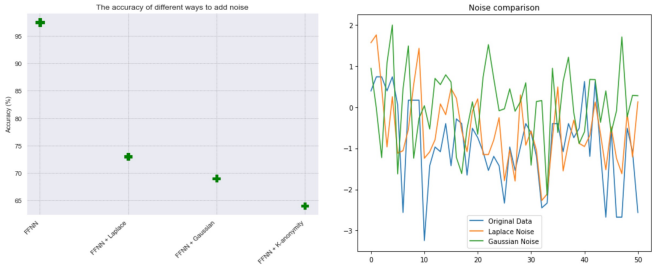


Fig. 5: (a)The scatter plot (b) The noise comparison

D. The implementation of Bayesian optimisation

This section illustrates a description of the implementation of multi-objective Bayesian optimisation. The aim of this section is to identify the most suitable set of parameters that can maximise the predictive accuracy of the model, while also ensuring compliance with stringent differential privacy constraints, through an iterative process of refining the hyper-parameter configuration. Simultaneously, the influence of the parameters on the model is examined through the execution of several iterations of experiments.

1) Explanation of dataset selection and objective

The current research utilises multi-objective Bayesian optimisation [39] as the foundation for the experiment, aiming to identify the optimal hyper-parameter set by maximising the objective function. In this section, the dataset that has undergone normalisation is utilised to train the Bayesian optimisation approach in order to achieve the optimal evaluation of the model's accuracy and its ability to preserve privacy. The primary objective of Bayesian optimisation is to identify the hyper-parameters of the FFNN model that yield the highest accuracy in prediction while adhering to stringent differential privacy limitations.

2) Training of the Neural Network model

During the phase of model training, the process involves iterative training of the model with input and target datasets, and optimising the model using categorical cross-entropy loss, the 'Adam' optimiser, and accuracy as the evaluation metrics. The number of training rounds on the target labels is determined by specifying the value of epochs. Upon completion of the training process, the performance of the model is assessed by evaluating it on a separate test dataset. This evaluation involves the computation of loss and accuracy metrics.

3) Define the DNN modelling

A neural network model with adjustable hyper-parameters is established using the create_model() function. This function takes as input the number of hidden layers and hidden units. A deep neural network is defined as FFNN with multiple hidden layers. The number of hidden units in the feed-forward neural network model, and the dropout rate to prevent overfitting are the hyper-parameters chosen for Bayesian optimisation. Additionally, the function requires training features in the data. The purpose of this model is to predict the most probable class for each input sample. The create_model() function returns the compiled neural network model.

4) Define the objective function of hyper-parameters

The objective () function is used to specify the aim in Bayesian optimisation for hyper-parameter tweaking of privacy-preserving FFNN. In our experiment, we employed one noise-affecting and four DNN hyper-parameters as inputs: epsilon, units_1, units_2, units_3, and dropout. We trained the model with these parameters, created model predictions for the test set, and translated the predicted labels to class labels. The model prediction accuracy is measured above by comparing the actual labels to the predicted labels. The accuracy is used to assess the target value.

Furthermore, the search space for the chosen hyper-parameters is defined using pbounds, which delineates the permissible values for each hyper-parameter within the search space. Through the establishment of a search space for the hyper-parameters, Bayesian optimisation demonstrates efficacy in the exploration of various hyper-parameter configurations, ultimately leading to the identification of the optimal combination of hyper-parameters. The experiment specifies that the epsilon (ϵ) value falls within the interval [0.1, 10], while the number of units in each hidden layer ranges from 10 to 200. The dropout parameter is constrained within the interval (0, 0.5), thereby denoting that the dropout rate has the capacity to range from 0% to 50%.

*5) **Performance of Bayesian optimisation***

By conducting the Bayesian optimisation procedure and the optimal combination of parameters are provided in the table below. The optimal parameters are obtained based on the experimental results.

TABLE II: The best noise parameters

Number	Noise Parameters	Value
1	Epsilon	4.57134858567302
2	Units 1	103
3	Units 2	164
4	Units 3	16
5	Dropout	0.36093748856273283

- **The histogram of various combinations.** The parameters that have been determined as optimal through Bayesian optimisation are incorporated into the noise-processed feed-forward neural network, and subsequently, the differences in accuracy are compared. Based on the data presented in Figure 6, four distinct outcomes can be derived.

(i) **Bayesian optimisation improves the accuracy of FFNN model.** Based on the comparison between FFNN and FFNN + Bayesian, it can be inferred that the inclusion of Bayesian optimisation in the latter results in an additional 0.5% in accuracy. This finding suggests that Bayesian optimisation improves the accuracy of the model. The valid parameters for the set of experiments include the quantities of hidden units within hidden layers 1 to 3, as well as the implementation of dropout.

(ii) **Bayesian optimisation significantly enhances the accuracy of Laplace mechanism.** Based on the comparison between FFNN + Laplace and FFNN + Laplace + Bayesian, it can be inferred that the inclusion of Bayesian optimisation in the selection process has resulted in a 13% increase in the accuracy of FFNN with Laplace noise. Within this set of experiments, the key variables under consideration are epsilon, the quantities of hidden units within the 1 to 3 hidden layers, and the implementation of dropout.

(iii) **Bayesian optimization considerably increases the accuracy of Gaussian mechanism.** The influential variables in this set of experiments include the quantities of hidden units in hidden layers 1-3 and the implementation of dropout. However, it is valuable to mention that the privacy budget of Gaussian is a factor that should be considered. The improvement in accuracy can be attributed to the optimisation of the parameters of the feed-forward neural network as it is influenced by the deviation of the dataset.

(iv) **Bayesian optimisation has little effect on K-anonymity.** The main argument is that the operational mechanism of K-anonymity does not aim to amplify noise, but rather to attain privacy protection by modifying the dataset itself. Consequently, even if the parameters of FFNN are optimised, the accuracy cannot be enhanced due to the substantial alterations in the initial dataset.

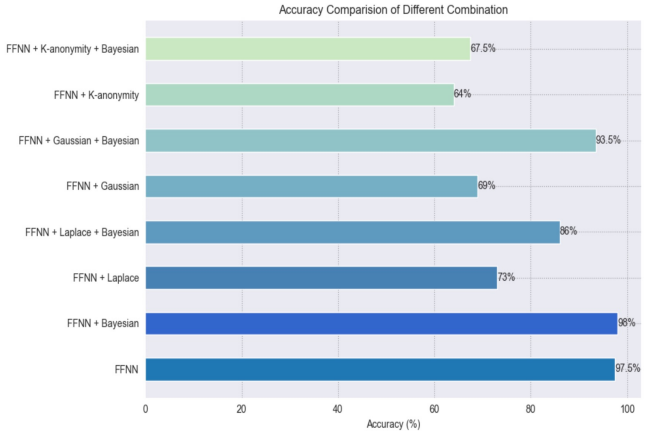


Fig. 6: Comparison of Accuracy of Different Combinations

V. Results

This section provides a comprehensive analysis and synthesis of the four parts presented in the previous section. These sections are examined critically and summarised, encompassing the following 4 components:

A. *Data preparation*

The task of data preparation serves as a fundamental prerequisite for the successful execution of the experiment. There are six essential steps involved in the task of data preparation. These steps encompass data cleaning, labelling and sorting, feature selection, data shuffling, normalisation, and splitting the dataset. The dataset utilised in this experiment has undergone a series of processing steps, resulting in the acquisition of a high-quality dataset that is well-suited for diabetes classification. This achievement has established a solid groundwork for the subsequent experiments to be conducted.

B. *The implementation of classification in DNNs*

In this section, the experiment effectively employs a feed-forward neural network with three hidden layers to successfully accomplish a three-class classification task on the

diabetes dataset. For the purpose of determining the issue pertaining to descriptive data, it is recommended to partition the chosen dataset into a classification problem that involves three distinct levels of severity: Normal, Prediabetes, and Diabetes. The selection of features is based on their ability to aid in the classification of the issue at hand. In addition, the feed-forward neural network (FNNN) was selected as the chosen method for implementing the classification task. This approach proved successful in addressing the multi-classification problem. The effectiveness of the classification was evaluated using the confusion matrix and ROC curve. Furthermore, the impact of the activation function, loss function, and regularisation technology function on the classification task was taken into consideration. Through these considerations, a deep neural network model was trained, demonstrating both generalisation and convergence capabilities that are well-suited for diabetes classification. In this field experiment, we investigate the variables that influence the precision of the model, specifically the quantity of hidden layers and hidden units within each hidden layer, and dropout. However, given the current magnitude of the dataset, the utilisation of hidden layers as a subsequent Bayesian experiment is not employed. In subsequent problems of increasing complexity, the effective use of the number of hidden layers can be considered as one of the hyper-parameters. Additionally, it is worth examining the impact of altering the number of hidden layers and the number of hidden units in each hidden layer on the accuracy of the model.

C. *The implementation of differential privacy*

This section is dedicated to the application of the Laplace mechanism of local differential privacy in order to augment the privacy assurance of FFNN. For this experiment, the (ϵ, δ) differential privacy framework was selected due to its provision of a more lenient privacy guarantee and its exploration of the applicability within medical scenarios. In this study, we assess the manner in which the privacy budget, denoted as epsilon (ϵ), effectively handles the delicate equilibrium between privacy preservation and the associated trade-offs with model accuracy. Furthermore, we conduct a comparative analysis of three distinct noise-addition mechanisms to ascertain their respective influences on the accuracy of the model. In summary, the experiment demonstrates that the utilisation of Bayesian optimisation leads to enhanced accuracy in the FFNN model. Moreover, it yields substantial improvements in the accuracy of both the Laplace mechanism and the Gaussian mechanism. The impact of the K-anonymity mechanism is constrained as it primarily relies on the characteristics of the dataset rather than the introduction of additional noise. Additionally, we have also derived the subsequent conclusions:

1) In terms of ensuring privacy, the Laplace mechanism outperforms the Gaussian mechanism. Figure 7:(a) illustrates the discrepancy between the data with added noise and the original data under the condition of the highest privacy parameter. Based on the provided diagram, it is evident that the significance of the Laplace mechanism's performance in noise intensity is evident.

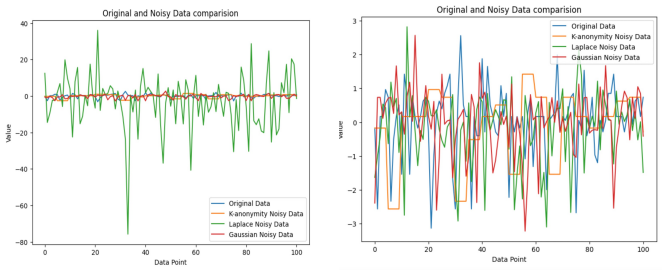


Fig. 7: (a) The privacy budget is maximized (b) The privacy budget is best

2) The application of Gaussian noise and Laplacian noise, both having equal scales, does not yield substantial distinctions in the data. The distinction between the two types of noise may be less discernible, particularly when the magnitude is limited. However, the utilisation of various scales may result in significant disparities. The noise distribution comparison depicted in Figure 8 is conducted under identical scaling conditions.

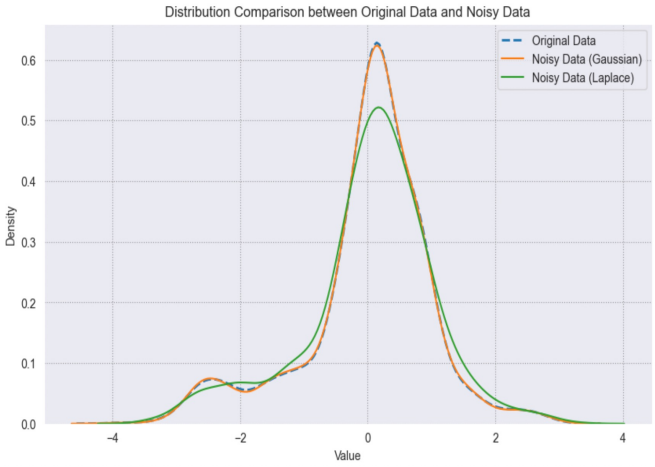


Fig. 8: Noise distribution on the same scale

D. *The implementation of Bayesian optimisation*

In this particular phase of the experiment, we successfully accomplished two objectives. One of the key contributions of our study is the utilisation of multi-objective Bayesian optimisation to identify the optimal hyper-parameter configuration for the privacy-preserving deep neural network model. This was achieved through a systematic approach involving modelling, model training, and the definition of hyper-parameter objectives and other essential steps. The second objective is to validate the optimal parameter set identified in the Experiment of the implementation of Bayesian optimisation by conducting eight additional rounds of experiments. This aims to maximise the predictive accuracy of the model while adhering to stringent differential privacy constraints. The findings indicate that the utilisation of Bayesian optimisation during the training process enhances the accuracy of the

feed-forward neural network model (see Figure 6). Regarding the protection of privacy, Bayesian optimisation has demonstrated notable enhancements in the precision of the Laplace mechanism and Gaussian mechanism. However, it is worth noting that Bayesian optimisation does not exert a substantial influence on the concept of K-anonymity. Based on what is known, it can be inferred that the Laplace mechanism is more appropriate for Bayesian optimisation in comparison to the Gaussian mechanism, primarily due to the privacy parameters involved. Figure 7 (b) illustrates the discrepancy between the data with added noise and the original data when the optimal parameters are determined through Bayesian optimisation. The figure illustrates that all three methods exhibit a certain level of privacy protection, with the Gaussian mechanism and Laplace mechanism demonstrating superior performance in this regard.

VI. Discussion

This project provides a modest contribution to the investigation of privacy-preserving methodologies for DNN within the domain of medical data classification. Nevertheless, it is imperative to acknowledge certain constraints that necessitate further investigation in subsequent research endeavours. The dataset utilised in this study is of a relatively limited scale, potentially resulting in an overestimation of the accuracy of the feed-forward neural network model prior to the introduction of noise. In subsequent experimental endeavours, it is imperative to employ larger datasets in order to validate the model's capacity for generalisation and to investigate the efficacy of more intricate deep neural network models. Furthermore, the comprehensive evaluation of the Adam optimiser for training feed-forward neural network has been limited. As a subsequent measure, it would be advisable to conduct a more extensive survey that delves into various optimisers, such as DP-SGD, in order to gain a deeper understanding of their influence on model performance within the confines of differential privacy constraints. Another limitation arises from the utilisation of identical datasets for both testing and validation purposes, thereby impeding our ability to investigate the rationale for exclusively applying differential privacy measures solely to the test dataset. Subsequent investigations may explore the rationales and ramifications of this configuration in order to enhance comprehension of privacy mechanisms within deep learning models.

Notwithstanding these constraints, the project's contribution is noteworthy. The proposed framework effectively integrates a deep neural network architecture that synergistically incorporates Bayesian optimisation and differential privacy techniques for the purpose of classifying medical data, with a specific focus on severity classification within the diabetes dataset. The framework under consideration has the potential to be applied to various medical scenarios, thereby ensuring the preservation of data privacy without compromising the accuracy of the model. Additionally, providing a comprehensive elucidation of the correlation between differential privacy and objective optimisation, as well as emphasising the significance of hyperparameter optimisation, contributes to a more profound comprehension of privacy-preserving deep learning. Besides, the efficacy of the project's utilisation of Bayesian optimisation in tackling the intricacies associated with parameter tuning has been substantiated through the consideration of diverse factors such as problem complexity, dataset size, and network architecture. This framework offers a pragmatic approach for the implementation of privacy-preserving deep learning models in practical scenarios by effectively managing the trade-off between privacy and accuracy.

VII. Conclusion

The present research endeavour explores the utilisation of DNNs in the scenario of multi-objective optimisation for the purpose of safeguarding differential privacy. The primary objective is the utilisation of Bayesian optimisation approaches to attain differential privacy in DNNs. The main purpose is to achieve a suitable equilibrium between preserving privacy and guaranteeing the efficacy of models, with a specific emphasis on conducting focused validation within the field of medicine. The significant findings can be briefly summarised as follows: The incorporation of differential privacy into DNNs enables the achievement of precise predictions while concurrently ensuring the protection of privacy, thereby providing significant insights into the intricate trade-off between accuracy and privacy preservation. Furthermore, an empirical investigation is carried out to validate the practicality and efficacy of Bayesian optimisation in the context of multi-objective optimisation. This investigation aims to determine the most optimal hyperparameter configuration for the aforementioned method.

Along with this, the project successfully implemented the DNNs for the purpose of classifying medical datasets based on their severity. More specifically, the network was able to classify diabetes datasets into three distinct levels: normal, prediabetes, and diabetes. Further, the integration of the Laplace mechanism, a method of local differential privacy, enhances the privacy guarantee of the feed-forward neural network model. The proposed methodology presents a feasible approach to ensuring privacy preservation within the medical dataset sample, while simultaneously ensuring the highest level of accuracy in the model's implementation. The present research aims to investigate the influence of privacy parameters on the accuracy of models and conducts a comparative analysis of different techniques for adding noise. The results underscore the advantages of the Laplace mechanism in terms of preserving privacy, as compared to the Gaussian mechanism.

The project eventually showcases the relationship and significance of differential privacy within the framework of DNNs and multi-objective optimisation. The provided example effectively demonstrates the efficacy of Bayesian methodologies in preserving data privacy while simultaneously achieving a suitable equilibrium between model accuracy and hyperparameter optimisation. In brief, this research facilitates the incorporation of privacy-preserving methodologies into deep learning models, providing valuable knowledge regarding the optimal setup of DNNs for the categorization of medical data, all the while ensuring heightened privacy assurances.

VIII. DECLARATIONS

A. *Declaration of Originality*

I am aware of and understand the University of Exeter's policy on plagiarism and I certify that this assignment is my own work, except where indicated by referencing, and that I have followed the good academic practices.

B. *Declaration of Ethical Concerns*

This work does not raise any ethical issues. No human or animal subjects are involved neither has personal data of human subjects been processed. Also no security or safety critical activities have been carried out.

REFERENCES

[1] Rigaki M, Garcia S. A survey of privacy attacks in machine learning[J]. arXiv preprint arXiv:2007.07646, 2020.

[2] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2017. Membership inference attacks against machine learning models. In 2017 IEEE Symposium on Security and Privacy (SP). IEEE, San Francisco, CA, USA, 3–18.

[3] Neil Zhenqiang Gong and Bin Liu. 2016. You are who you know and how you behave: Attribute inference attacks via users' social friends and behaviors. In 25th USENIX Security Symposium (USENIX Security 16). Usenix, Austin, TX, USA, 979–995.

[4] Abadi M, Chu A, Goodfellow I, McMahan H B, Mironov I, Talwar K, Zhang L. Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. 2016, 308–318

[5] Dwork C. Differential privacy: a survey of results. In: Agrawal M, Du D, Duan Z, Li A, editors. Theory and applications of models of computation. Berlin: Springer; 2008. p. 1–19.

[6] McSherry F, Talwar K. Mechanism design via differential privacy. In: Proceedings of the 48th IEEE Symposium on Foundations of Computer Science; 2007 Oct 21–23; Providence, RI, USA. New York: IEEE; 2007. p. 94– 103.

[7] Frazier P I. A tutorial on Bayesian optimization[J]. arXiv preprint arXiv:1807.02811, 2018.

[8] Kaisa Miettinen (1999). Nonlinear Multiobjective Optimization. Springer. ISBN 978-0-7923-8278-2. Retrieved 29 May 2012.

[9] Konak A, Coit D W, Smith A E. Multi-objective optimization using genetic algorithms: A tutorial[J]. Reliability engineering & system safety, 2006, 91(9): 992-1007.

[10] Dwork C. Differential privacy[C]//Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II 33. Springer Berlin Heidelberg, 2006: 1-12.

[11] Dwork C, Kenthapadi K, McSherry F, et al. Our data, ourselves: Privacy via distributed noise generation[C]//Advances in Cryptology-EUROCRYPT 2006: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28-June 1, 2006. Proceedings 25. Springer Berlin Heidelberg, 2006: 486-503.

[12] Nicolas Papernot, Martín Abadi, Ulfar Erlingsson, Ian Goodfellow, and Kunal Talwar. Semi-supervised knowledge transfer for deep learning from private training data. In Proceedings of the 5th International Conference on Learning Representations, 2017.

[13] www.tensorflow.org/responsible_ai/privacy/tutorials/classification_privacy

[14] Rasmussen C E, Williams C K I. Gaussian processes in machine learning[J]. Lecture notes in computer science, 2004, 3176: 63-71.

[15] Gan W, Ji Z, Liang Y. Acquisition Functions in Bayesian Optimization[C]//2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE). IEEE, 2021: 129-135.

[16] Kaisa Miettinen (1999). Nonlinear Multiobjective Optimization. Springer. ISBN 978-0-7923-8278-2. Retrieved 29 May 2012.

[17] Konak A, Coit D W, Smith A E. Multi-objective optimization using genetic algorithms: A tutorial[J]. Reliability engineering & system safety, 2006, 91(9): 992-1007.

[18] Zhao F, Ren X, Yang S, et al. Federated Multi-objective Reinforcement Learning[J]. Information Sciences, 2023.

[19] Kusner M, Gardner J, Garnett R, et al. Differentially private Bayesian optimisation[C]//International conference on machine learning. PMLR, 2015: 918-927.

[20] Nguyen T D, Gupta S, Rana S, et al. A privacy preserving Bayesian optimisation with high efficiency[C]//Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conference, PAKDD 2018, Melbourne, VIC, Australia, June 3-6, 2018, Proceedings, Part III 22. Springer International Publishing, 2018: 543-555.

[21] Tobaben M. Hyperparameters and neural architectures in differentially private deep learning[J]. 2022.

[22] Fan T, Cui Z. Adaptive differential privacy preserving based on multi-objective optimization in deep neural networks[J]. Concurrency and Computation: Practice and Experience, 2021, 33(20): e6367.

[23] Rashid A. Diabetes dataset[J]. Mendeley Data, 2020, 1.

[24] Swab M. Mendeley data[J]. Journal of the Canadian Health Libraries Association/Journal de l'Association des bibliothèques de la santé du Canada, 2016, 37(3).

[25] www.elsevier.com/authors/tools-and-resources/research-data/mendeley-data-for-journals

[26] www.mendeley.com/terms/

[27] [Clore,John, Cios,Krzysztof, DeShazo,Jon, and Strack,Beata. (2014). Diabetes 130-US hospitals for years 1999-2008. UCI Machine Learning Repository. doi.org/10.24432/C5230J.

[28] James G, Witten D, Hastie T, et al. An introduction to statistical learning[M]. New York: springer, 2013.

[29] Hastie T, Tibshirani R, Friedman J H, et al. The elements of statistical learning: data mining, inference, and prediction[M]. New York: springer, 2009.

[30] Encyclopedia of machine learning[M]. Springer Science & Business Media, 2011.

[31] Bunn H F, Higgins P J. Reaction of monosaccharides with proteins: possible evolutionary significance[J]. Science, 1981, 213(4504): 222-224.

[32] World Health Organization. Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. Diabetes Res Clin Pract 2011;93:299–9.

[33] Executive Summary: Standards of Medical Care in Diabetes—2010. Diabetes Care 1 January 2010; 33 (Supplement$_1$): S4–S10.

[34] Bebensee B. Local differential privacy: a tutorial[J]. arXiv preprint arXiv:1907.11908, 2019.

[35] Nissim K, Raskhodnikova S, Smith A. Smooth sensitivity and sampling in private data analysis[C] Proceedings of the thirty-ninth annual ACM symposium on Theory of computing. 2007: 75-84.

[36] Dwork C, McSherry F, Nissim K, et al. Calibrating noise to sensitivity in private data analysis[C]//Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3. Springer Berlin Heidelberg, 2006: 265-284.

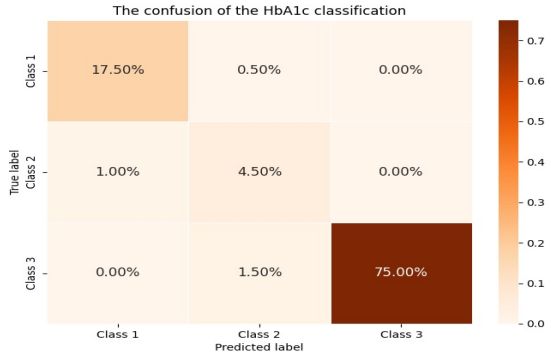
[37] Dwork C, Roth A. The algorithmic foundations of differential privacy[J]. Foundations and Trends® in Theoretical Computer Science, 2014, 9(3–4): 211-407.

[38] Sweeney L. k-anonymity: A model for protecting privacy[J]. International journal of uncertainty, fuzziness and knowledge-based systems, 2002, 10(05): 557-570.

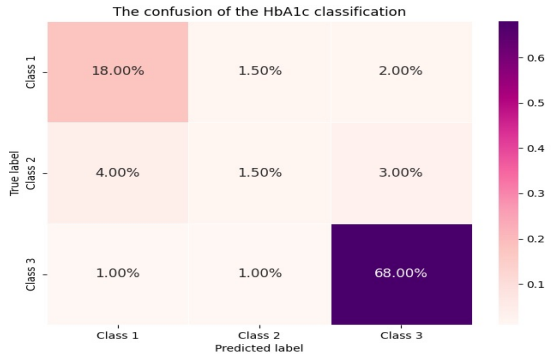
[39] Deb K, Pratap A, Agarwal S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II[J]. IEEE transactions on evolutionary computation, 2002, 6(2): 182-197.

[40] Qin S, He J, Fang C, et al. Differential private discrete noise adding mechanism: Conditions, properties and optimization[J]. arXiv preprint arXiv:2203.10323, 2022.

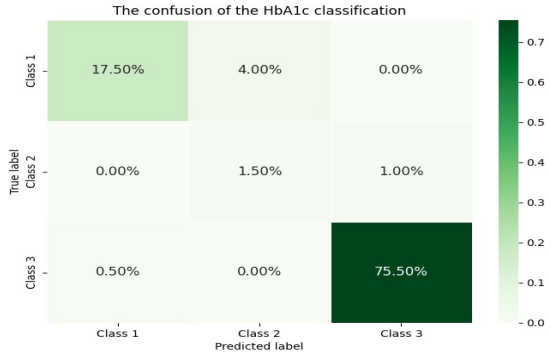
APPENDIX



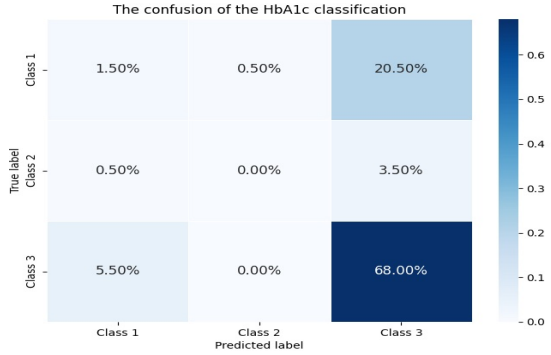
(a)



(b)

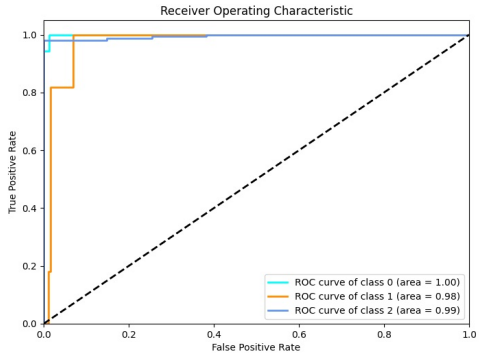


(c)

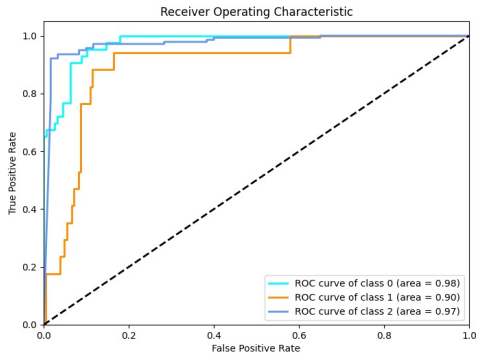


(d)

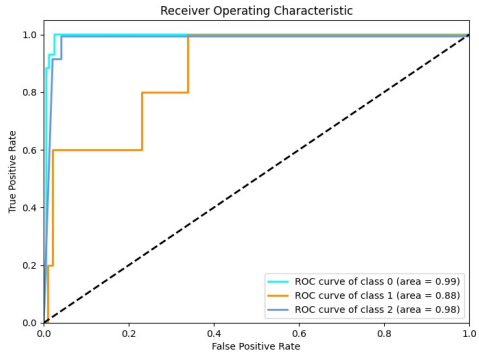
Fig. 9: Confusion matrixes after Bayesian optimisation: (a) FFNN + Bayesian (b) FFNN + Laplace + Bayesian (c) FFNN + Gaussian + Bayesian (d) FFNN + K-anonymity + Bayesian



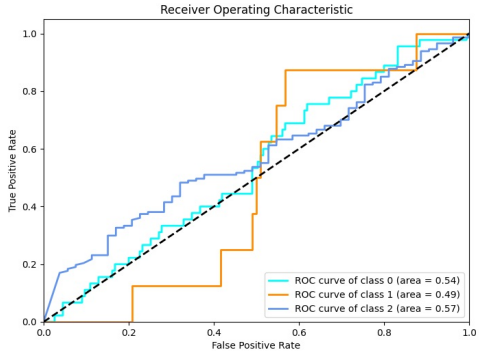
(a)



(b)



(c)



(d)

Fig. 10: ROC curves after Bayesian optimisation: (a) FFNN + Bayesian (b) FFNN + Laplace + Bayesian (c) FFNN + Gaussian + Bayesian (d) FFNN + K-anonymity + Bayesian