mathematics

MDPI

Article

Distinguishing Bladder Cancer from Cystitis Patients Using Deep Learning

Dong-Her Shih [1], Pai-Ling Shih [2], Ting-Wei Wu [1,*], Chen-Xuan Lee [1] and Ming-Hung Shih [3]

[1] Department of Information Management, National Yunlin University of Science and Technology, Douliu 64002, Taiwan; shihdh@yuntech.edu.tw (D.-H.S.); m11023018@yuntech.edu.tw (C.-X.L.)
[2] Department of Information Management, National Chung Cheng University, Chiayi 621301, Taiwan; d08530003@ccu.edu.tw
[3] Department of Electrical and Computer Engineering, Iowa State University, 2520 Osborn Drive, Ames, IA 50011, USA; mshih@iastate.edu
* Correspondence: wutingw@yuntech.edu.tw

Abstract: Urinary tract cancers are considered life-threatening conditions worldwide, and Bladder Cancer is one of the most malignant urinary tract tumors, with an estimated number of more than 1.3 million cases worldwide each year. Bladder Cancer is a heterogeneous disease; the main symptom is painless hematuria. However, patients with Bladder Cancer may initially be misdiagnosed as Cystitis or infection, and cystoscopy alone may sometimes be misdiagnosed as urolithiasis or Cystitis, thereby delaying medical attention. Early diagnosis of Bladder Cancer is the key to successful treatment. This study uses six deep learning methods through different oversampling techniques and feature selection, and then through dimensionality reduction techniques, to establish a set that can effectively distinguish between Bladder Cancer and Cystitis patient's deep learning model. The research results show that based on the laboratory clinical dataset, the deep learning model proposed in this study has an accuracy rate of 89.03% in distinguishing between Bladder Cancer and Cystitis, surpassing the results of previous studies. The research model developed in this study can be provided to clinicians as a reference to differentiate between Bladder Cancer and Cystitis.

Keywords: Cystitis; Bladder Cancer; deep learning; dimensionality reduction; data imbalance

MSC: 68T05; 68T35; 68Q32

Citation: Shih, D.-H.; Shih, P.-L.; Wu, T.-W.; Lee, C.-X.; Shih, M.-H. Distinguishing Bladder Cancer from Cystitis Patients Using Deep Learning. *Mathematics* **2023**, *11*, 4118. https://doi.org/10.3390/math11194118

Academic Editors: Xujuan Zhou, Lemai Nguyen and Guohun Zhu

Received: 10 September 2023
Revised: 24 September 2023
Accepted: 25 September 2023
Published: 28 September 2023

1. Introduction

According to the latest global cancer data for 2020 released by the International Agency for Research on Cancer (IARC) of the World Health Organization, cancer caused nearly 10 million deaths in 2020 [1]. As one of the top ten most common cancers in the world, Bladder Cancer is the tumor with the highest cost associated with lifetime treatment and one of the tumors with the most significant impact on postoperative quality of life [2]. Even in 2019, an estimated 17,600 people died from Bladder Cancer in the United States, a figure that accounts for 2.9% of all cancer deaths [3].

Bladder Cancer is one of the most common and expensive human malignancies to treat [4–6]; Bladder Cancer refers to various malignant tumors in the bladder. Possible symptoms include frequent urination, dysuria, pain in urination, low back pain, abdominal pain, and severe painless hematuria [7]. In the general adult population, invisible hematuria is found in 2–7% of men and 3–15% of women [8]; therefore, it cannot be made cost-effective for screening [9]. Bladder Cancer may initially be misdiagnosed as Cystitis or infection, which may be a complication during the early growth of the tumor before clinical diagnosis [10]. Previous studies have also suggested that Bladder Cancer may be mistaken for Interstitial Cystitis [11,12]. Because mass or thickening of the bladder wall is the dominant imaging feature of Eosinophilic Cystitis, similar to Bladder Cancer,

differentiation from Bladder Cancer can be challenging [13]. In addition, Interstitial Cystitis (IC) may be misdiagnosed as Cystitis, leading to the misdiagnosis of Bladder Cancer [14]. On imaging, it sometimes presents masses that mimic various papillary urethral tumors and are misdiagnosed as Bladder Cancer before surgery [15].

Clinical chemistry tests and urine analysis are the main diagnostic screening tests in clinical laboratories [16], and each change in the test can be interpreted as a relationship with the disease, such as complete urine analysis including color, clarity, specific gravity, and chemical analysis according to the past literature. Furthermore, urine sediment examination [17] and chemical analysis are usually performed using a strip system to check for pH, glucose, ketones, occult blood, bilirubin, and protein [18]; however, some diseases may present abnormal results of chemical analysis, for example, ketonuria can be found in the urine of diabetic patients. At the same time, hematuria indicates bleeding in the urinary tract [19]. In addition, the examination of urinary sediment may reveal crystals, red blood cells, white blood cells, bacteria, and tubules, which provide information about the urinary tract system [20]; thus, interpreting clinical tests alone can lead to misleading diagnoses [21]. Laboratory test results can be interpreted by experienced clinicians but can also be integrated and interpreted in combination with artificial intelligence (AI), such as machine learning algorithms [22].

Machine learning (ML) is a type of artificial intelligence that enables it to learn independently from data without human intervention [23]. In addition to various machine learning algorithms often used in medical research, examples such as decision trees, random forests, XGBoost, and GBM have been frequently used in medical research [24–26]. Recently, many applied machine learning studies have been introduced into clinical practice, and machine learning has become a powerful tool to improve the accuracy of cancer diagnosis and prognosis. For example, Garapati et al. [27] established an objective computer-aided system to identify the stage of Bladder Cancer through CT urography. In addition, machine learning is also applied to metabolomics to identify early and late stages of Bladder Cancer [28], and Tsai et al. [29] applied machine learning to predict Bladder Cancer based on clinical laboratory data. This study will explore the use of deep learning technology to distinguish patients with Bladder Cancer and Cystitis and hope to improve the accuracy of prediction.

This study attempts to correctly identify patients with Cystitis and Bladder Cancer by using actual numerical clinical data, and the research scenario is shown in Figure 1. This study attempted to match multiple deep learning models with different data preprocessing and dimensionality reduction techniques and attempted to surpass previous studies in accuracy on clinical datasets. The data preprocessing includes missing value (MI), feature selection (RFE/CST/MBR), and imbalanced data processing (SMOTE/ADASYN). After data preprocessing is completed, dimensionality reduction technology is adopted to avoid excessive features and overfitting of the model. Then, the reduced feature set is sent to six deep learning models (DBN, GRU, LSTM, DNN, ANN, and MLP) input feeds to obtain the deep learning technology with the best accuracy.

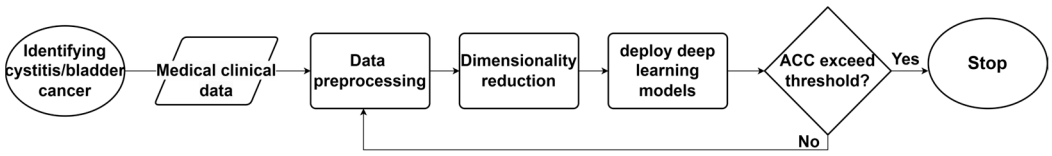

Figure 1. Research scenario.

2. Background and Related Work

2.1. Bladder Cancer and Cystitis

2.1.1. Cystitis

Cystitis is a generic term used to define any bladder inflammation, which can be acute or chronic. At the same time, the severity can range from mild discomfort in the lower abdomen to life-threatening bleeding [30]. There are several categories to describe the

various causes of Cystitis, divided into infectious, radiation, chemical, mechanical, interstitial Cystitis/chronic pelvic pain syndrome, and several conditions disguised as Cystitis. However, on a broader level, Cystitis can be divided into infectious and noninfectious. Patients with infectious Cystitis often complain of irritating emptying symptoms, difficulty urinating, frequency, urgency, and pain in the pubic hair, with severe bleeding occurring only in rare cases. Another major category of Cystitis is sterile or noninfectious Cystitis, which can be caused by radiation and chemical irritations. Unlike infection-induced Cystitis, noninfectious Cystitis is more clinically severe and can cause extreme pain, hematuria, and irritating emptying symptoms [30].

Interstitial Cystitis (IC) is a unique chronic syndrome in that it does not fit the classical distinction of infectious or noninfectious Cystitis [30]. Characterized by a range of lower urinary tract irritation symptoms and pain, the broad clinical definition of IC includes any patient who complains of urgency, frequency, and pelvic/perineal pain in the absence of bacterial infection or cancer [31]. Common symptoms include frequent urination, nocturnal urination, the urgency to urinate, bladder allergy, bladder discomfort, and bladder pain [32]. Unfortunately, IC is difficult to diagnose. To do this, clinicians must rule out urinary or vaginal infections, Bladder Cancer, bladder inflammation or infections caused by radiation therapy, eosinophilic and tuberculous Cystitis, kidney stones, endometriosis, neurological disorders, sexually transmitted diseases, low-count bacteria in the urine, and male prostatitis. Finally, Cystoscopy performed under general or regional anesthesia is used [33].

Eosinophilic Cystitis is a rare inflammatory disease of the bladder with uncertain etiology, first described by Brown [34] as eosinophilic granuloma on the bladder wall. Although many cases have been reported in adult and pediatric populations, their etiology remains elusive [35]. The most common symptoms are frequency, dysuria, urgency, pain, and hematuria, and the typical clinical findings are bladder mass, peripheral eosinophilic thickening, and bladder wall thickening [36].

2.1.2. Bladder Cancer

Bladder Cancer is the second most common malignancy of the urinary tract, accounting for approximately 3.2% of all cancers worldwide, and it is a significant cause of cancer morbidity and mortality. Most Bladder Cancers are diagnosed after the presence of macrohematuria [37], and the majority of cases (80%) occur in people over 65 years of age, with the highest incidence in people aged 85–89 years [38]. The most common symptom of Bladder Cancer is severe painless hematuria. In addition, unexplained symptoms of frequent urination, urgency, or irritating emptying should alert clinicians to the possibility of Bladder Cancer [7]. Bladder Cancer is invasive and non-invasive, with non-invasive having a reasonable risk and prognosis but muscle-aggressive having a poor prognosis [39].

The clinical spectrum of Bladder Cancer can be divided into three categories: Prognosis, management, and treatment goals. The first category includes non-muscular aggressive tumors, where treatment is designed to reduce recurrence and prevent progression to more advanced stages. The second category includes muscle-invasive lesions, where the goal of treatment is to determine whether the bladder should be removed or preserved without affecting survival and determine whether the primary lesion can be managed independently or whether the patient is at high risk of long-distance transmission, which requires a systematic approach to improve the likelihood of a cure. Finally, the critical issue is prolonging the quantity and quality of life, including metastatic diseases. Many agents with different mechanisms of action have antitumor effects against this disease [40].

Bladder Cancer does have the same symptoms as other diseases in its early stages, such as kidney cancer, prostate cancer, interstitial Cystitis, kidney stones, benign prostatic hyperplasia, and trauma [41]. Bladder Cancer may initially be misdiagnosed as Cystitis or infection, which may be a complication during early tumor growth before clinical diagnosis [42]. The symptoms of in situ Bladder Cancer and chronic Cystitis are similar, such as hematuria, frequent urination, and lower abdominal discomfort, and some patients even

have urinary incontinence. If a patient has been treated for chronic Cystitis with antibiotics, cancer cells will take the opportunity to spread. While bladder inflammatory diseases such as interstitial Cystitis are sometimes challenging, they are often misdiagnosed [43].

2.1.3. Distinguishing Bladder Cancer and Cystitis

Bladder Cancer is the ninth most common malignancy worldwide [5], and clinicians use knowledge from different specialties to analyze histological, clinical, and demographic information [44]. Statistical methods such as Cox regression, logistic regression, and Kaplan–Meier estimators are commonly used in the analysis. For example, Kaplan–Meier methods and Cox proportional risk models were used to assess prognostic factors for recurrence, progression, and disease mortality in patients with Bladder Cancer [45]. Logistic regression based on 12 variables was used to determine predictors of 5-year overall survival in patients with Bladder Cancer who underwent radical cystectomy (Bassi et al., 2007). However, with the rapid development of health technology and informatics, the accuracy of predictions largely depends on the efficient integration of information from data obtained from various sources (clinical or pathological), which makes traditional statistical analysis relying on clinician knowledge and experience a difficult task. For example, regression modeling is a standard statistical technique that often requires some explicit assumptions about the relationship between the data that may not be valid [46]. Therefore, machine learning has been introduced into medicine to overcome the problems of statistical methods and reveal the knowledge hidden in complex clinical data [47].

In medicine and healthcare, machine learning has been applied to personalized and predictive medicine [48], cancer diagnosis and detection [49], and prevention and treatment policy research [44]. For Bladder Cancer, reliable predictions of patients undergoing cystectomy were achieved using an artificial neural network (ANN) prediction model and optimized by genetic algorithms (GA) [50]. This system has the potential for widespread use in medical decision support. In addition, Ji et al. [51] used ANN and radial basis function networks to predict the survival rate of patients with Bladder Cancer. In addition, clinic pathological and molecular markers were also used to create an ANN model to predict the one-year survival rate of patients with invasive Bladder Cancer [52]. Relevant studies on patients with Bladder Cancer and Cystitis are summarized in Table 1.

Table 1 shows the study on machine learning in the medical field and Cystitis and Bladder Cancer. Machine learning has made essential contributions to medical diagnosis, including diagnosis classification, image recognition, and disease diagnosis. It can be known from Table 1 that distinguishing Bladder Cancer from Cystitis patients can only be achieved using machine learning [29]. There are few studies on using deep learning and other data pre-processing methods to improve their work. Therefore, this study intends to fill this research gap to improve the accuracy of distinguishing Bladder Cancer from Cystitis patients.

Table 1. Machine learning in the medical field and studies on Cystitis or Bladder Cancer.

Studies on Cystitis or Bladder Cancer	Method	ACC	Author
Prediction of one-year survival in patients with muscular invasive Bladder Cancer	ANN	82%	[52]
Prediction of survival in patients with Bladder Cancer after diagnosis	RBF	85%	[51]
Accuracy of prognosis in patients with radical cystectomy	ANN, LR	LR:75.9% ANN:76.4%	[46]
Distinguish between Cystitis and Bladder Cancer	DT, RF, SVM, XGBoost GBM	ACC:87.6%	[29]
Test for Bladder Cancer	Trasfer learning (CNN)	ACC:96.9%	[53]
Related studies in the medical field	Subject	Method	Author
Medical image recognition	Quantify tumor characteristics	RF	[23]

Table 1. *Cont.*

Studies on Cystitis or Bladder Cancer	Method	ACC	Author
Complex data classification	Heart disease and hypertension data classification	Tree-based	[24]
	Hypertensive patient prediction	RF SVM XGBoost	[25]

2.2. Deep Learning

2.2.1. Artificial Neural Network (ANN)

Artificial neural networks (ANNs) and their recent advances in deep learning (DL) are booming in computer science [54]. The standard network for pattern recognition in the MATLAB® toolbox is a two-layer feedforward network. There are transmission functions in both the hidden layer and the output layer. The design idea of ANN is to imitate the working mode of the human brain. ANN consists of an input, multiple hidden, and an output layer. The units of adjacent layers are fully connected. Therefore, it has a solid ability to fit, especially for nonlinear functions. Due to the complex model structure, training artificial neural networks is very time-consuming. It is worth noting that ANN models are trained by backpropagation algorithms [55]. Figure 2 shows the basic structure of neural networks, including input, hidden, and output layers.

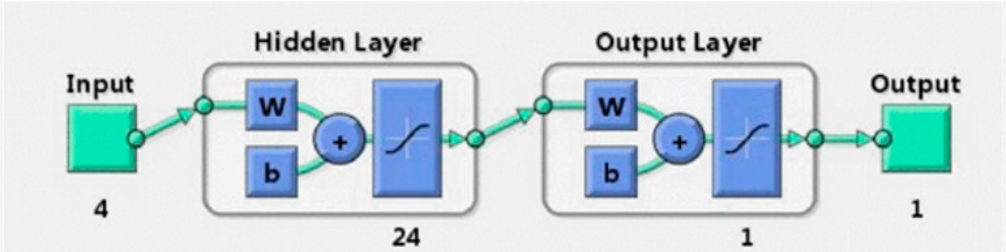

Figure 2. ANN structure.

2.2.2. Deep Neural Network (DNN)

When training Deep Neural networks (DNN), unlabeled data are first used to learn parameters in an unsupervised feature learning stage. The network is then adjusted by labeling the data to form a supervised learning phase [55]. Figure 3 shows an example of a deep neural network [56]. DNN is a collection of neurons organized in a multilayered sequence where the neuron receives neuronal activation from the previous layer as input and performs simple calculations (such as a weighted sum of the input, followed by nonlinear activation). The neurons of the network work together to realize complex nonlinear mapping from input to output. This mapping is learned from the data by adjusting each neuron's weight using error backpropagation [57].

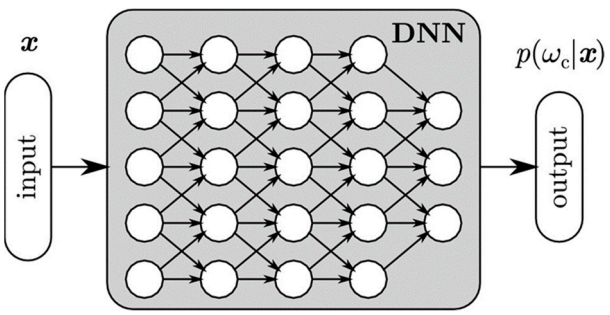

Figure 3. DNN structure.

2.2.3. Deep Belief Network (DBN)

A Deep belief network (DBN) is a probabilistic generative model consisting of multiple layers of random latent variables, often with binary values, referred to as hidden units or feature detectors. The top two layers have non-directional, symmetric connections between them and form an associative memory, the lower layer receives top-down directional connections from the upper layer, and the state of the lowest layer's units represents a data vector [58]. As shown in Figure 4, a DBN consists of an input layer (in red circle), multiple hidden layers (in white circle), and an output layer (in green circle). Each layer is a node or neuron connected, and the previous layer's output is regarded as the input layer of each hidden layer [59]. Training a DBN consists of two stages: Unsupervised pre-training and supervised fine-tuning [60]. First, each RBM is trained using the layering pre-training of greedy, and then the weights of softmax layers are learned from labeled data.

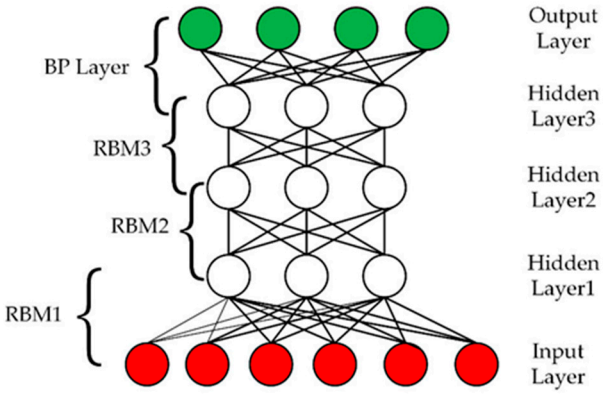

Figure 4. DBN structure.

The key idea of the deep confidence network is that the weight W learned by the restricted Boltzmann machine defines both the prior distribution p (v | h, W) and the hidden vector p (h | W), so the probability of producing a visible vector is v. After learning W, we keep p (v | h, W) and replace p (h | W) with a better model of the aggregate posterior distribution on hidden vectors; that is, the non-factor distribution is generated by averaging the factor posterior distribution generated by the various data vectors, learning this better model by taking the hidden activity vector generated by the training data as the training data for the next learning module. Hinton et al. [60] showed that if this substitution is performed in the right way, the lower limit of variation for the probability of training data under the comprehensive model can be raised.

$$p(v) = \sum_{h} p(h|W)p(v|h, W) \quad (1)$$

2.2.4. Long Short-Term Memory (LSTM)

LSTM is a neural network model that integrates implicit internal memory, and many researchers have demonstrated its high efficiency in time series prediction [61]. The LSTM unit extends memory units with additional memory units. LSTM can solve gradient disappearance and explosion problems in many sequential tasks. LSTM models can be used in deep learning machines as complex nonlinear functions that take advantage of long-term memory. LSTM extends three control gates: Input, output, and forget. These gates, which control the flow of information into or out of the unit and reset the memory unit, build a complex memory unit structure in the LSTM model. The general structure of the LSTM unit that holds the network's time unit state and hides the state is shown in Figure 5.

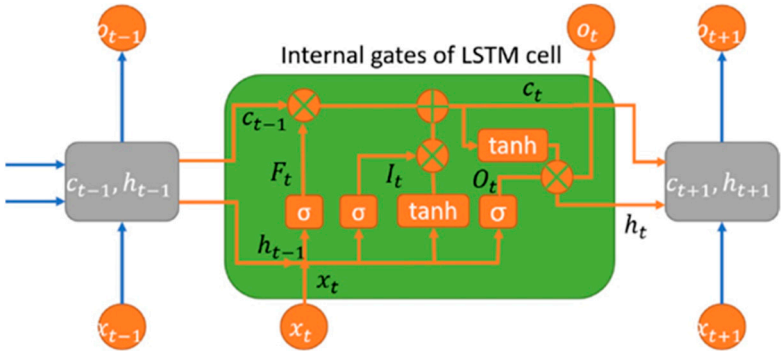

Figure 5. LSTM structure.

2.2.5. Gate Recurrent Unit (GRU)

The GRU model was proposed by Cho et al. [62], who chose a new type of hidden unit driven by an LSTM unit to read and modify the memory unit by controlling the input gate, ignore gate, and output gate. Different functions were then used to update the state of the hidden layer [61]. The performance of GRUs in speech signal modeling is similar to long-term short-term memory. In addition, its training time is shorter than that of traditional LSTMS, and its parameters are fewer than those of LSTM because it lacks output gates. Figure 6 shows the structure of GRU, and the final model is more straightforward than the standard LSTM model and is a prevalent variant [59].

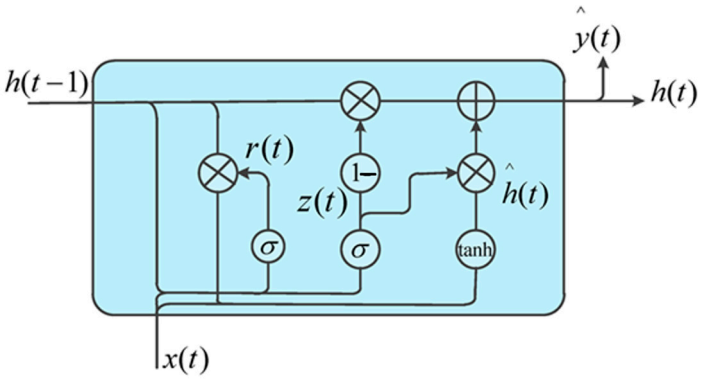

Figure 6. GRU structure.

2.2.6. Multilayer Perceptron (MLP)

A multilayer perceptron is a system of neurons interconnected by weights and output signals generated by the weighting and application of a nonlinear activation function to node inputs and their respective weights [63]. Since it is a feedforward neural network, the output of one layer of neurons is fed to the next layer of neurons, and the layer between the input and output layers is called the hidden layer [64]. In addition, the output of this neural network is compared to the target value to determine the error, which is then backpropagated to update the weights, resulting in a lower error in the next period [65].

MLP is the basis for backpropagation algorithms and continuous approximation techniques [66]. A general MLP is a network consisting of three layers ($n \geq 2$), including the input, output, and hidden layers, as shown in Figure 7. The general architecture of an MLP can be described as follows:

Step 1: Input vector (x_1, x_2, …, x_p) in p-dimensional space and the output vector (y_1, y_2, …, y_q) in the q-dimensional space.
Step 2: Each neuron in one layer is linked to all the neurons in the previous layer. The output of the previous layer is the input of the next layer.

Step 3: The MLP works as follows: In the input layer (in yellow square), the neuron receives the input signal and processes it (calculates the weight and passes it to the activation function), producing the result of the activation function. The results obtained from the first hidden layer (in blue circle) are then processed and passed to the second hidden layer (in purple circle). This process continues until the neurons in the output layer (in green circle) produce results.

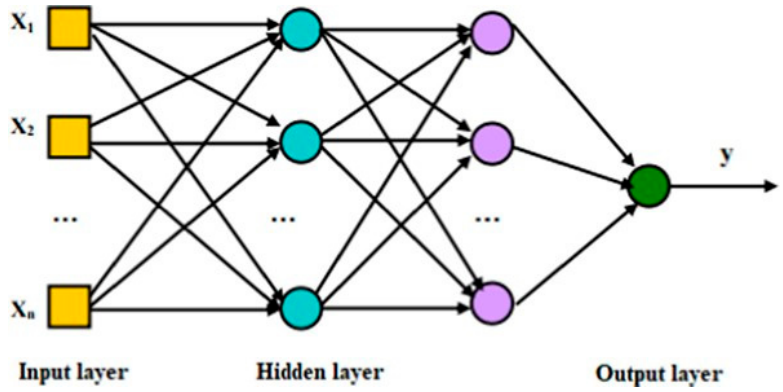

Figure 7. MLP structure.

3. Material and Methods

3.1. Dataset

Data for this study were collected from clinical laboratory data of patients diagnosed with Cystitis, Bladder Cancer, and other types of cancer at Mackay Memorial Hospital from January 2017 to February 2020 [29]. Among them were 144 patients with Cystitis (56 women and 88 men, aged 60.12 $\pm$ 11.99 years), 200 patients with kidney cancer (62 women and 138 men, age 63.41 $\pm$ 10.45 years), 201 prostate cancer patients (201 male patients, age 71.83 $\pm$ 6.42 years), 591 Bladder Cancer patients (205 female patients and 386 male patients, age 66.73 $\pm$ 9.4 years), and 200 uterine cancer patients (200 female patients, age 60.86 $\pm$ 10.26 years). The total number of data is 1336, and the description of data variables is shown in Table 2.

3.2. Research Process

The research process of this study is shown in Figure 8. Data pre-processing is divided into three parts, namely, feature selection (RFE, Chi-Squared Test, and Model-based ranking), missing value processing, and imbalanced data processing (SMOTE/ADASYN). In the data pre-processing stage, min-max normalization technology is used to normalize the dataset. In the next stage, the dimensionality reduction technique (PCA) is used, and then the dataset is divided into the training set and test set by using 5-fold cross-validation. Selecting hyperparameters is a crucial step in creating effective deep-learning solutions. DL algorithms often include specific hyperparameters that control various factors, such as memory usage and execution costs. A hyperparameter is a variable set before a learning algorithm is applied to a context-specific dataset. The optimal number depends on the characteristics of the dataset associated with each task and each situation. The reduced feature set generated by the dimensionality reduction technique is fed into six depth models (DNN, ANN, DBN, LSTM, GRU, and MLP) as input. In the training model stage, the optimizer is used to determine the best model for the model with the best results. In the model training stage, the hyperparameters of the hidden layer, such as the learning rate and activation function, are adjusted before the input is processed. Finally, after adjusting the parameters, the efficiency model can be obtained to distinguish between the patients with Bladder Cancer and Cystitis.

Table 2. The description of variables.

Variable	Variable Name	Type
X_1	Patient Number	Integer
X_2	age	Numeric
X_3	gender	Boolean
X_4	hypertension	Boolean
X_5	diabetes	Boolean
X_6	smoking	Boolean
X_7	drinking	Boolean
X_8	beetle nuts	Boolean
X_9	family history	Boolean
X_{10}	A/G Ratio	Numeric
X_{11}	Albumin	Numeric
X_{12}	ALP	Numeric
X_{13}	ALT	Numeric
X_{14}	AST	Numeric
X_{15}	BUN	Numeric
X_{16}	Calcium	Numeric
X_{17}	Chloride	Numeric
X_{18}	Creatinine	Numeric
X_{19}	Direct Bilirubin	Numeric
X_{20}	Estimated GFR	Numeric
X_{21}	Glucose AC	Numeric
X_{22}	pH	Numeric
X_{23}	Potassium	Numeric
X_{24}	Sodium	Numeric
X_{25}	Specific Gravity	Numeric
X_{26}	Total Bilirubin	Numeric
X_{27}	Total Cholesterol	Numeric
X_{28}	Total Protein	Numeric
X_{29}	Triglyceride	Numeric
X_{30}	Uric acid	Numeric
X_{31}	Urine epitheilum (UL)	Numeric
X_{32}	Urine epithelium count	Numeric
X_{33}	Nitrite	Boolean
X_{34}	Strip WBC	Numeric
X_{35}	Urine Bilirubin	Numeric
X_{36}	Urine Glucose	Numeric
X_{37}	Urine Ketone	Numeric
X_{38}	Urine Protein	Numeric
X_{39}	Urobilinogen	Numeric
X_{40}	Urine occult Blood	Numeric

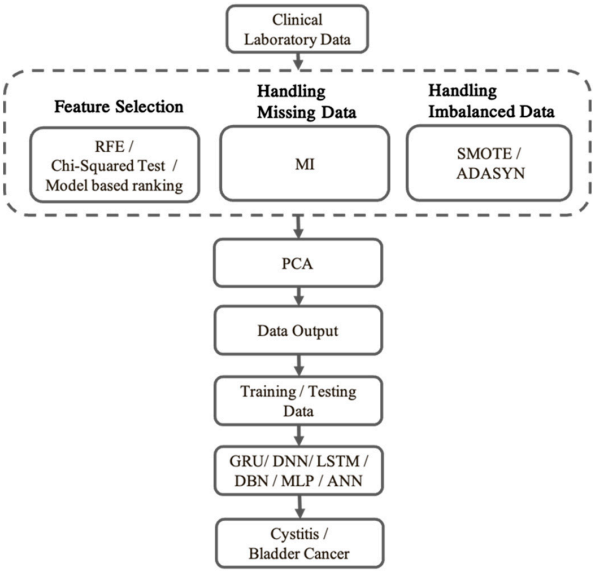

Figure 8. Research Process.

3.3. Data Preprocessing

3.3.1. Missing Value Processing

This study uses multiple imputations (MI) to process missing values. Multiple imputation (MI) is widely regarded as an excellent method to calculate missing values of existing data [67]; the purpose of multiple interpolation is to reconstruct data to preserve the original relationship between variables as much as possible [68]. Multiple interpolations solve the problem of too small or too large standard error obtained by using traditional methods of processing missing data, and multiple interpolation has become an essential and influential method in the statistical analysis of incomplete data. During this time, its application has expanded to include analyzing observational data from public health studies and clinical trials. Therefore, MI may be a better choice for medical-related clinical data [69–71]. Its principle is to provide unbiased and valid correlation estimates based on information from available data, producing estimates similar to those calculated for complete data. Missing data and multiple interpolation may affect coefficient estimates for variables with missing data and other variables without missing data [72].

3.3.2. Feature Selection

After calculating missing values, it is also necessary to identify essential features with strong positive correlations with important features in disease diagnosis, and extracting vector features eliminates useless predictive features and irrelevant features [73]. Feature selection is a crucial step in machine learning and data analysis, as it helps improve model performance by identifying and retaining the most relevant features while discarding irrelevant or redundant ones. The choice of feature selection methods should be motivated by the specific characteristics of the dataset and the problem to solve.

In order to compare whether different feature selection methods will affect the final accuracy, this study adopts three feature selection methods. Here is how Recursive Feature Elimination (RFE), the Chi-Squared Test (CST), and Learning Model-Based Feature Ranking (MBR) methods relate to the literature and their motivations:

RFE is a feature selection method that recursively fits a model (such as a classifier or regressor) to the data, ranks the features by their importance, and eliminates the least important features at each iteration. RFE is widely used and has been applied in various domains, including bioinformatics, finance, and image processing. Researchers have demonstrated its effectiveness in selecting informative features while improving model generalization. RFE is motivated by the idea that by iteratively removing less important features, the model's performance can be enhanced. This method is particularly useful when dealing with high-dimensional datasets where feature selection is essential for model interpretability and efficiency [74].

The CST is a statistical method used for feature selection in classification tasks. It assesses the independence between a feature and the target variable by measuring the difference between observed and expected frequencies of feature-target pairs. The motivation for using CST lies in its ability to capture relationships between categorical features and categorical target variables. It helps in selecting features that are most likely to have a significant impact on the target variable's class distribution [75]. Chi-square (χ^2) statistics are used to test the independence of two variables by calculating scores to measure the degree of independence of the two variables. In feature selection, χ^2 measures the independence of features relative to categories. The initial assumption of χ^2 is that features and categories are independent before scores are calculated [76], and scores with larger values indicate a high dependency. A Chi-square [77] score with c class and r values is defined as (2) and (3):

$$x^2 = -\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - u_{ij})}{u_{ij}} \quad (2)$$

n_{ij} is the number of samples value with the ith value of the feature.

$$u_{ij} = \frac{(n*_j n_{i*})}{n} \tag{3}$$

n_i is the number of samples and i is the feature value. n_i is the number of samples in class j. n is the number of samples.

MBR methods involve training a machine learning model (e.g., decision trees, random forests, and gradient boosting) and ranking features based on their contribution to the model's performance. Many machine learning algorithms inherently rank features based on their importance during the training process. For example, decision trees and random forests assign feature importance scores. Researchers have leveraged these methods for feature selection, and their effectiveness is well-documented. MBR methods are motivated by the Idea that features contributing the most to the model's performance should be retained, while irrelevant features can be discarded. This approach is data-driven and adapts to the specific problem at hand.

3.3.3. Imbalanced Data Processing

Imbalanced data mean that the number of specific categories in the data is particularly large or small, which quickly leads to the prediction of a large number of categories during model training or prediction, resulting in the prediction results of the model being worse than expected. However, the sampling method transforms the imbalanced dataset into a balanced dataset by processing the training set, which will improve the final result in most cases. Two imbalanced data processing methods, the Synthetic minority oversampling technique (SMOTE) and the Adaptive Synthetic Sampling Approach (ADASYN), will be adopted in this study. Using SMOTE and ADASYN in imbalanced data processing offers several benefits for improving the performance of machine learning models in such scenarios. The benefits of SMOTE are addressing class imbalance, improved model performance, the preservation of information, and a reduced risk of overfitting [78].

SMOTE is a nearest-neighbor-based technique used by Euclid to determine the distance between data points in a feature space, and SMOTE is a popular and effective method for solving class imbalances in many fields [78]. SMOTE's central idea was to synthesize more samples based on the feature-space similarities between the few existing instances. Specifically, given the imbalanced data T, for each minority class instance $x_i \in T$, SMOTE first uses the Euclidean distance for x_i, then randomly selects one of the K nearest neighbors, computes the difference between the eigenvectors x_i and its corresponding nearest neighbor, and finally, multiplies the eigenvector difference by a random number and adds the new vector to x_i. The mathematical formula used to synthesize the new minority sample is shown in Equation (4) [79]:

$$x_{new} = x_i + \left(x_i^k - x_i\right) \times \delta \tag{4}$$

In x_i^k, one of the nearest neighbors is x_i, and δ is a random value belonging to (0, 1), so the resultant few instances x_{new} are a point x_i connected along the segment and its nearest neighbor x_i^k.

The ADASYN algorithm is an adaptive synthetic oversampling algorithm proposed by He et al. [80]. The benefits of ADASYN are adaptability, handling class overlap, robustness to noise, and balanced distribution with focused effort. The ADASYN method aims to reduce the imbalanced data between classes and adaptively adjust classification restrictions using complex samples. To synthesize different amounts of new sample data for a smaller number of samples, the key is to find a probability distribution r_i and apply n to determine the number of samples to be synthesized for each small class of samples.

For each class of samples, we obtain the K nearest neighbor of x_i in the n-dimensional space, the ratio of $r_i = \frac{\Delta i}{k}, i = 1,2, \ldots, m$. Δi is the class number near the K value of x_i.

Therefore, $r_i \in (0,1]$. Given the probability distribution of ($\sum r_i = 1$) of $\hat{r} = \frac{r_i}{\sum_{i=1}^{ms} r_i}$, the case for most classes around each minority sample is calculated. We calculate the number of synthetic samples based on the xi of each minority sample: $g_i : g_i = \hat{r_i} + G$ [81].

SMOTE and ADASYN are valuable tools for addressing class imbalances in machine learning datasets. They can improve model performance, generalize better, and reduce bias against most classes. The choice between SMOTE and ADASYN depends on the nature of the dataset. SMOTE is usually a good choice when the occupational imbalance does not change much, and the occupational overlap is not significant. ADASYN, on the other hand, is better suited to datasets with different densities, overlapping categories, and noisy datasets because it can adapt to the specific challenges these datasets pose.

3.3.4. Dimensionality Reduction

Principal Component Analysis (PCA) is a multivariable dimensionality reduction technique that analyzes a data table in which several interrelated quantitatively dependent variables describe observations. Its goal is to extract meaningful information from a table and represent it as a new set of orthogonal variables called principal components [82]. The quality of PCA models can be evaluated using cross-validation techniques. PCA models can be summarized as correspondence analysis (CA) to deal with qualitative variables and multi-factor analysis (MFA) to deal with heterogeneous variable sets [82]. Historically, PCA was first proposed in the context of statistics for estimating the principal component of the multivariable random variable x [83,84]. Specifically, given a zero-mean multivariable $x \in \mathbb{R}^D$ and an integer $d < D$, d is the principal component of x, and $y \in \mathbb{R}^D$ is defined as d uncorrelated linear components of x [85].

$$y_i = u_i \top x \in \mathbb{R}, u_i \in \mathbb{R}^D, i = 1, 2, \ldots, d, \tag{5}$$

Maximizing the variance of y_i

$$u_i \intercal u_i = 1 \; and \; Var(y1) \geq Var(y2) \geq \ldots \geq Var(yd) > 0. \tag{6}$$

For example, to find the first principal component y_1, let us find a vector $u_1^* \in \mathbb{R}^D$ such that

$$u_1^* = \underset{u_i \in \mathbb{R}^D}{\operatorname{argmax}} Var(u_1 \intercal x) \;\; s.t. \;\; u_1 \intercal u_1 = 1. \tag{7}$$

3.4. Evaluation Metrics

Through the following evaluation metrics, the effectiveness obtained after classification can be evaluated, and the classification results can be divided into True Positive (TF), that is, correct acceptance, True Negative (TN) or correct rejection, False Positive (FP) or wrong acceptance, and False Negative (FN) or wrong rejection. Accuracy, Sensitivity, Specificity, and other relevant metrics can be calculated from these four metrics [86].

(1) Accuracy

An accuracy assessment is defined as the ratio of correctly predicted samples to the total number of predicted samples. TP is true positive, TN is true negative, and total represents the total number of predictions.

$$Accuracy = \left(\frac{(TP + TN)}{Total} \right) \tag{8}$$

(2) Precision

Precision calculates the correct classification number that is penalized for an incorrect classification number.

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

(3) Recall

Recall counts the number of correct categories punished for missing entries

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

(4) Specificity

Specificity is the proportion of people without an original negative test, commonly referred to as the proportion of "true negative"

$$Specificity = \frac{TN}{FP + TN} \tag{11}$$

(5) F-Score

F-Score is the harmonic average of measurement accuracy and recall rate

$$F - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{12}$$

(6) ROC

The larger the area under the ROC curve (AUC), the higher the accuracy of the detection

$$ROC = \frac{sensitivity + specificity}{2} \tag{13}$$

4. Experimental Results and Discussion

4.1. Software Environment Settings

The operating environment of this research experiment was developed by MacOS Monterey v12.2.1 (Apple Inc., Cupertino, CA, USA), and the software used was Visual Studio Code v1.71.2 (Microsoft, Redmond, WA, USA). The following table shows the method and software name of the pre-processing stage, including missing value processing, imbalanced data processing, and feature selection. The software summary of the selection method is shown in Table 3.

Table 3. Pre-processing software setup.

Preprocessing Stage	Method	Software
Missing value processing	MI	IBM SPSS v29
Unbalanced data processing	SMOTE	Imbalanced-learn v0.10.1
	ADASYN	
Feature selection	RFE	scikit-learn v1.3.0
	CST	
	MBR	

The six deep learning frameworks that distinguish Cystitis and Bladder Cancer are DNN, ANN, GRU, LSTM, MLP, and DBN. Except for DBN, which uses Tensorflow v2.11.0 (Python, open source), they all use Keras v2.11.0 in Tensorflow. The dimensionality reduction technology Principal Component Analysis (PCA) uses scikit-learn v1.3.0 (Python, open source), and the optimization algorithm is developed using Python 2.7.18 (open source).

4.2. Hyperparameter Setting

The hyperparameter settings of the deep learning model will affect the classification results. The parameter setting table of this study is shown in Table 4. The parameter setting tables in the respective deep models are all referenced from different literature. For

example, the optimal parameter setting of the GRU model comes from Lee et al. [87], the optimal parameter settings of the LSTM model are from Heikal et al. [88], the optimal parameter settings of the DNN model are from Gunawan et al. [89], the optimal parameter settings of the MLP model are from Arabzad et al. [90], the optimal parameter settings of the ANN model are from Helwan and Tantua [91], and the optimal parameter settings of the DBN model are from Suparwito et al. [92].

Table 4. Parameter setting of each deep learning model.

Method	Attributes				
DNN	Class	Number of character	Number of hidden layer	Number of neurons in 1st hidden layer	Number of neurons in 2nd hidden layer
	Digits	10	3	300	50
	Letters	26	3	300	50
GRU	Epoch	Batch size	Learning rate	Gradient threshold	Dropout rate
	1000	200	0.005	1	10
LSTM	Hidden state dimension	Dropout rate	Learning rate	Number of epochs	Batch size
	200	0.5	0.001	10	50
ANN	Number of neurons in output layer	Number of neurons in hidden layer	Iterations number	Learning rate	Momentum rate
	2	40	5000	0.001	0.4
MLP	Number of neurons in input layer	Number of neurons in output layer	Number of neurons in each hidden layer	Learn rate	Number of hidden layers
	10	2	20	0.05	2
DBN	Hidden layers	Activation function	Learning rate	Learning rate scale	Momentum rate
	8-17-9	Thnh	0.1	5000	0.4

4.3. Feature Selection

The original data are processed using multiple imputation (MI) to handle missing values, and after LabelEncoder and normalization, it is then applied to three feature selection methods. Table 5 shows the features selected by the three different feature selection methods. It can be seen that there are some differences in the features selected by different feature selection methods.

In addition, this dataset has the problem of data imbalance, so this study uses the SMOTE algorithm to deal with the problem of data imbalance and then conducts training and testing of six deep learning models. Table 6 shows the accuracy of the six deep learning models. As can be seen from the table, the preliminary research results of the six deep learning models have not yet reached expectations despite the extraction of important features, as shown in Table 7. The research results in Table 6 show that the MLP deep learning model has better results, and the accuracy reached 78.48%. Therefore, this study will conduct further model optimization research on the MLP model.

Table 5. Features selected by three different feature selection methods.

Feature Name	Feature Content	Feature Selection		
		CST	RFE	MBR
Nitrite	Whether it contains nitrite	√	√	√
Age	Patient age	√	√	√

Table 5. *Cont.*

Feature Name	Feature Content	Feature Selection		
		CST	RFE	MBR
Diabetes	Whether the patient has diabetes	√	√	√
Smoking	Whether the patient smokes	√	√	√
Beetle nuts	Whether the patient ate betel nut	√	√	√
Family history	Whether the patient has a genetic history	√	√	√
Drinking	Whether the patient has consumed alcohol	√		√
Urine Bilirubin	Urinary bilirubin values	√		√
Urine Protein	Urine Protein values	√		√
Hypertension	Whether the patient has high blood pressure	√		√
Chloride	Chloride values		√	√
Estimated GFR	Renal bulb filtration rate		√	√
Potassium	Potassium values		√	√
Specific Gravity	Specific gravity of urine		√	√
Strip WBC	white blood cell values		√	√
Total Cholesterol	Total cholesterol values		√	√
Uric acid	Uric acid values		√	√
Urine Ketone	Urine Ketone values		√	√

Note: √ indicates used.

Table 6. Model accuracy evaluation after initial feature selection.

Method / Model	RFE	CST	MBR
	Accuracy		
DBN	74.36%	63.29%	52.32%
GRU	62.87%	60.76%	51.53%
LSTM	67.51%	70.04%	51.48%
DNN	70.04%	70.01%	72.15%
ANN	70.16%	67.08%	68.35%
MLP	78.48%	74.39%	75.94%

Table 7. Evaluation metrics for other deep learning models.

Method	Accuracy	Precision	F1 Score	Sensitivity	Specificity	AUC
DNN	68.8%	65.12%	69.71%	75%	63.11%	69.06%
ANN	67.52%	65.25%	67%	68.75%	66.39%	67.57%
LSTM	63.68%	64.84%	58.13%	52.68%	73.77%	63.22%
GRU	60.68%	68.52%	44.58%	33.04%	86.07%	59.55%

4.4. Deep Learning Optimizer

The main purpose of adding optimizers to deep learning is to train neural networks so that they can learn from data and improve performance. Optimization algorithms play a key role in the neural network training process because they help the network adjust model

parameters. There are many optimization algorithm choices for deep learning training. Different algorithms may affect the quality of the training effect. Different optimization algorithms are used in the process of deep learning model training and optimization, hoping to improve the performance of existing models.

After the MLP model undergoes the RFE feature selection method, this study will use different optimization algorithms, such as the Gray Wolf Optimizer (Gwo), Salp Swarm Algorithm (Ssa), Cuckoo Search (Cs), Sine Cosine Algorithm (Sca), Firefly Algorithm (Fa), Particle Swarm Optimization (Pso), Whale Optimization Algorithm (Woa), Gene Algorithm (Ga), Flower Pollination Algorithm (Fpa), and Bat Algorithm (Ba). The MLP model parameters are adjusted in the hope of improving the accuracy. The experimental results are shown in Table 8. As a result, the accuracy has not been significantly improved. This may be because the optimization algorithm itself also has feature selection steps. Using the same feature selection may lead to the loss of important features. It may be because these optimization algorithms are still unable to find the global optimal solution. There are certain limitations. Although the optimization algorithm can effectively find the local optimal solution, it is difficult to find the global optimal solution sometimes.

Table 8. Classification accuracy results of different optimizers.

Model	Original Acc	Feature Selection	Optimizer/Acc
MLP	78.48%	RFE	Pso:68.77% Woa:68.78% Ga:70.04% Sca:71.31% Ssa:74.26%
		CST	Ssa:66.24% Sca:67.09% Ga:70.04% Ba:70.89% Gwo:71.73%
		MBR	Ba:71.73% Ssa:79.32% Gwo:75.95% Cs:76.79% Ga:78.48%

Mirjalilietal [93] developed the GWOA intelligent optimization algorithm in 2014. This approach attempts to mimic gray wolves and their leadership hierarchy. The Gray Wolf is one of nature's predators. They usually come in groups of 5 to 12. In their social ruling class, these wolves follow strict rules. The internal leadership hierarchy is split into α, β, δ, and ω wolves [94].

Alpha (α) Wolf: It makes decisions and leads hunting and other activities.

Beta (β) wolves: They support the alpha in decision-making and are subordinate to the alpha.

Type delta (δ) wolves: These wolves (δ) obey the alpha and beta wolves and are responsible for protecting the pack and assisting the dominant wolves in hunting.

Omega (ω) Wolves: Affiliated with alpha, beta, and delta Wolves, responsible for monitoring territorial boundaries.

The Salp swarm algorithm (SSA) is based on Salp predation behavior that can solve complex daily life optimization problems in nature. Marine life is diverse, and most of these species share the same behaviors and characteristics, such as communication styles, motor performance, and finding food. The Salp is a species of Marine organism in the Salpidae family. Its shape is very similar to that of a jellyfish (cylindrical). Marine organisms share certain behaviors, such as group behavior. Salp is known as the salp chain, and

biological researchers believe this behavior helps sea squirts achieve better movement and foraging [95].

The Sine-cosine algorithm (SCA) is a population-based optimization algorithm introduced by [96], motivated by the trigonometric sine and cosine functions. SCA generates various initial random solutions and uses mathematical models based on sine and cosine functions to require them to transition to the best solution. The algorithm also combines various random and adaptive variables to maintain exploration and utilization of the search space at various optimization milestones.

Cuckoo search (CS) is a relatively new algorithm proposed by [97], and CS is highly efficient at solving global optimization problems. Cuckoo search (CS) is a new nature-inspired meta-heuristic algorithm based on larval parasitism in some cuckoo species. The steps are as follows:

Each cuckoo lays one egg at a time and then drops it into a randomly selected nest.

The best nests and high-quality eggs will be passed on to the next generation.

The number of available host nests is fixed, and cuckoo eggs are likely to be found by host birds. In this case, the host bird either throws away the eggs or abandons the nest and builds a new one.

4.5. Dimensionality Reduction

Since the experiment of adding optimizers to deep learning did not improve the performance of the deep learning model, this study also used dimensionality reduction technology (PCA) and different oversampling methods, plus feature selection methods. The experimental results are shown in Table 8, and it can be seen from Table 9 that the classification accuracy of the MLP deep learning model composed of the SMOTE method for oversampling in data preprocessing, the MBR method for feature selection, and the dimensionality reduction technology PCA significantly increased to 89.03%, showing that the oversampling method can indeed effectively deal with the problem of data imbalance. The MBR feature selection method can effectively select the best feature subset. Coupled with dimensionality reduction technology, it can indeed improve the performance of the model.

Table 9. Feature selection of different oversampling and classification accuracy results with dimensionality reduction.

Method	Oversampling	Feature Extraction	ACC	With PCA
MLP	ADASYN	CST	66.09%	68.94%
		RFE	77.77%	79.20%
		MBR	79.48%	85.47%
	SMOTE	CST	71.83%	73.52%
		RFE	77.77%	79.20%
		MBR	78.9%	89.03%

Finally, this study summarizes all experimental results in Table 10. This study uses six deep learning methods (ANN, DNN, DBN, LSTM, GRU, and MLP), combined with different feature selection methods (MBR, CST, and RFE). In addition to methods dealing with data imbalance (ADASYN and SMOTE), optimization algorithms were employed during the process. The final results show that preprocessing uses MBR for preliminary feature selection, then uses oversampling (SMOTE) technology to process the data, and finally uses principal component analysis (PCA) to select the final features with the highest accuracy. Excellent results were achieved with an accuracy of 89.03% in classifying Cystitis and Bladder Cancer.

Table 10. Comparison of the accuracy in distinguishing Bladder Cancer from Cystitis.

Method	Imbalanced	Feature Extraction	Dimensionality Reduction	Optimizer	ACC
ANN	ADASYN	RFE	PCA	-	71.77%
DNN					72.57%
DBN					65.82%
LSTM					47.26%
GRU					52.74%
MLP	ADASYN	MBR	PCA	Bat	79.75%
				Genetic	80.17%
				Salp Swarm	81.43%
				Cuckoo Search	82.28%
				Grey Wolf Optimizer	84.81%
		CST		Salp Swarm	64.4%
				Genetic	66.67%
				Sine Cosine	68.78%
				Bat	68.78%
				Grey Wolf Optimizer	69.2%
				Genetic	73.41%
		RFE		Whale Optimization	75.11%
				Particle Swarm Optimization	75.94%
				Sine Cosine	77.21%
				Salp Swarm	78.06%
				Sine Cosine	78.9%
		-		Firefly	79.75%
				Cuckoo Search	82.28%
				Salp Swarm	83.12%
				Grey Wolf Optimizer	86.5%
		MBR			89.03%
	SMOTE	CST	PCA	-	73.52%
		RFE			77.75%

4.6. Comparison with Previous Research

This study implements a deep learning approach that can effectively distinguish patients with Bladder Cancer and Cystitis, thereby contributing to the medical field. Table 11 is a comparison table of related research. In previous studies, Tong et al. [98] used machine learning methods (ML) to classify the urine metabolome and its potential application in the diagnosis of interstitial Cystitis in the study of Cystitis. Freitas [53] used feature fusion, transfer learning, and CapsNets to detect Bladder Cancer, but they did not make the dataset public. In other studies, Yu et al. [99] and Tsai et al. [29] both used machine learning technology to distinguish Bladder Cancer from Cystitis, while this study used a deep learning model to distinguish patients with Bladder Cancer and Cystitis and the accuracy rate reached 89.6%, which is the highest accuracy rate among all studies so far.

In addition, the ROC curve of this study is shown in Figure 9. The ROC curve (Receiver Operating Characteristic curve) is a chart commonly used to evaluate the performance of binary classification models. It plots the true positive rate (also called the recall or sensitivity) of the classifier as the y-axis and the false positive rate as the x-axis. The AUC in the ROC curve in Figure 10 is 0.92 (ROC curve area), which shows the good performance of the model in classification problems, with high discrimination ability and prediction accuracy.

Table 11. Comparison with previous research.

Disease		Dimensionality Reduction	Method	Dataset	Author	Result
Cystitis	Bladder Cancer					
√			(ML)SVM, LR	Cystitis data set	[98]	Acc: 86%
	√		(DL)CNN	Experimental data set (unpublished)	[53]	-
√	√		(ML)Decision tree	Shandong Provincial Hospital Dataset (unpublished)	[99]	-
√	√		(ML)lightGBM	Clinical trial data set	[29]	Acc: 87.6%
√	√	√	(DL) MLP	Clinical trial data set	Our Study	Acc: 89.03%

Note: √ indicates used.

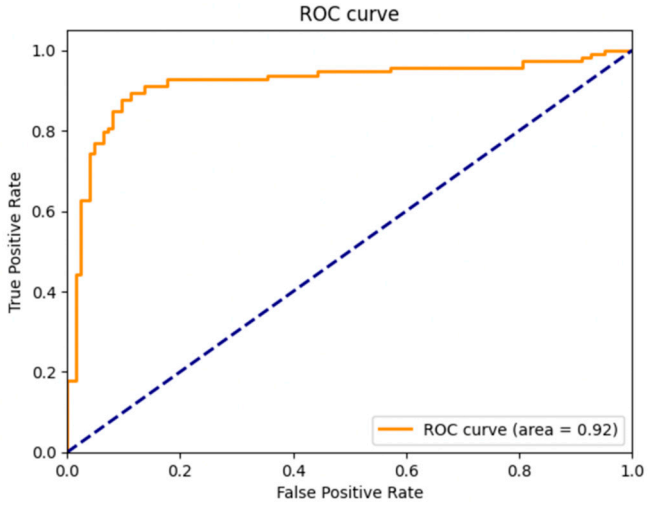

Figure 9. ROC curve of our study.

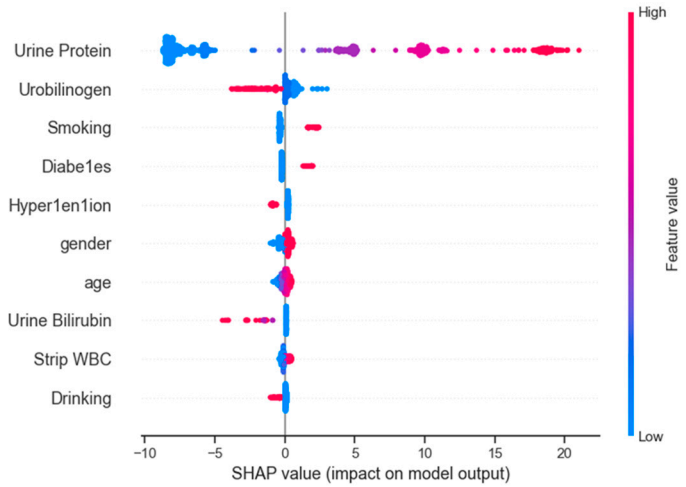

Figure 10. SHAP feature importance analysis.

In addition to accuracy, this study also included other evaluation metrics for comparison with relevant studies. As shown in Table 12, the deep learning method used in

this study also achieved good performance in terms of precision and specificity. Precision refers to the accuracy of the model in predicting positive samples, that is, the proportion of samples predicted as positive samples by the model that are actually positive samples. In this study, our research precision is as high as 94.85%, which means that the model can accurately identify Bladder Cancer, which is very important for application scenarios that require high precision. At the same time, the specificity of a clinical trial refers to the test's ability to correctly identify patients who do not have the disease, so a test with 100% specificity can correctly identify all patients who do not have the disease [100], and the specificity in our study was as high as 95.97%. This means that the model can better distinguish between true negative samples and false positive samples, further improving the reliability and effectiveness of the proposed model. The F1 score is a harmonic average of accuracy and recall and is particularly useful when dealing with unbalanced datasets. Lopsided datasets are standard in the medical field, and the occurrence of a particular disease may be rare compared to non-disease cases. Sensitivity measures the proportion of actual positive cases (patients with a disease) correctly identified by the model. In a medical context, sensitivity is critical because it indicates the ability of a diagnostic test to correctly identify an individual who truly has a disease. The AUC represents the probability that the model will rank a randomly selected positive instance higher than a randomly selected negative instance. The higher the AUC value (close to 1), the better the model can distinguish between diseased and non-diseased individuals.

Table 12. Comparison of evaluation metrics.

Method	Accuracy	Precision	F1 Score	Sensitivity	Specificity	AUC	Author
MLP	89.03%	94.85%	87.62%	82.30%	95.97%	92%	Our Study
lightGBM	87.6%	86.3%	87.7%	89.5%	85.5%	93.2%	[29]

5. Discussion

5.1. Feature Selection

This study uses three feature selection methods, MBR, CST, and RFE, to screen out the main features. From Table 6, it is found that the characteristics of nitrite test, age, diabetes, smoking, betel nut consumption, and family history are duplicated among the three feature selection methods, and the recurring characteristics are discussed as follows:

1. The nitrite test is a commonly used detection method. It can measure the nitrite content (Nitrite) in urine. High nitrite content may be related to an increased risk of Bladder Cancer. When we consume foods containing nitrite, nitrites can be converted into carcinogens in the body, especially in acidic environments, such as urine in the bladder. These carcinogens can cause damage to the cells lining the bladder and increase the risk of Bladder Cancer. Using nitrites alone to differentiate between Bladder Cancer and Cystitis (inflammation of the bladder or urinary tract infection) is not a reliable diagnostic method because, in both cases, nitrites may be present in the urine. However, nitrite testing can be part of a broader diagnostic approach.
2. Age is one of the known risk factors for Bladder Cancer. As age increases, the risk of Bladder Cancer also increases. Age can be a factor in distinguishing between Bladder Cancer and Cystitis patients, but it is not a definitive diagnostic criterion on its own. Instead, age is one of several factors that healthcare professionals consider when evaluating a patient's condition. Bladder Cancer is more commonly diagnosed in older individuals, typically over the age of 60. Older individuals may be more likely to develop symptoms of Bladder Cancer, such as hematuria (blood in the urine), lower back pain, or weight loss. It is important to emphasize that while age can provide some context, it is not a standalone diagnostic tool for distinguishing between Bladder Cancer and Cystitis. The final diagnosis and treatment plan should be determined by a healthcare professional based on a comprehensive evaluation that takes into account multiple factors.

3. Diabetes, a chronic medical condition characterized by high blood sugar levels, can impact the risk and presentation of various health issues, including Bladder Cancer and Cystitis. However, diabetes alone is not a definitive diagnostic tool for distinguishing between these two conditions. Diabetes is associated with certain risk factors that can affect the development and progression of Bladder Cancer. These risk factors include obesity, a sedentary lifestyle, and a history of smoking, which are more common in individuals with diabetes. Diabetes can sometimes cause urinary symptoms, such as frequent urination, increased thirst, and recurrent urinary tract infections (UTIs). It is important to note that while diabetes may influence the risk and presentation of bladder conditions, including Bladder Cancer and Cystitis, it is not a standalone diagnostic tool. Patients with diabetes who experience urinary symptoms or have concerns about their bladder health should promptly seek medical attention. Early detection and appropriate evaluation are essential for the effective management of these conditions, and healthcare providers will use a multifaceted approach to reach a diagnosis.
4. Smoking and eating betel nuts have also been confirmed to be risk factors for Bladder Cancer and are related to an increased incidence of Bladder Cancer. Smoking is a recognized risk factor for Bladder Cancer. Smokers are more likely to develop Bladder Cancer than non-smokers. Therefore, smoking history is an important consideration when evaluating patients with urinary symptoms. Smoking-related Bladder Cancer often presents with symptoms such as hematuria, frequent urination, and low back pain. However, these symptoms can also be seen with Cystitis. Healthcare providers may use diagnostic tests including urinalysis, urine culture, imaging studies (such as CT scan or ultrasound), cystoscopy (visual examination of the bladder), and tissue biopsies. These tests provide a more definite diagnosis. Betel nut consumption has been linked to an increased risk of oral and esophageal cancer, but not directly to Bladder Cancer or Cystitis. However, it is still necessary for healthcare providers to evaluate patients for betel nut use for an overall health assessment. In conclusion, while smoking and betel quid consumption are important factors to consider when assessing bladder condition, they are not diagnostic tools. The final diagnosis of Bladder Cancer or Cystitis relies on a comprehensive evaluation, including the patient's medical history, symptoms, and diagnostic test results.
5. Genetic history (Family history) is also considered to be one of the risk factors for Bladder Cancer, and individuals with Bladder Cancer cases in their families may be at increased risk. Distinguishing between patients with Bladder Cancer and Cystitis based solely on genetic family history is challenging because both diseases can occur independently of genetic predisposition. However, family history can be a valuable piece of information when assessing a patient's risk factors and should be considered as part of a comprehensive evaluation. Healthcare providers will ask patients about their family history of any relevant medical conditions, including Bladder Cancer, urinary tract disease, and other genetic conditions that may be associated with an increased risk of Bladder Cancer. Certain genetic syndromes are associated with an increased risk of Bladder Cancer, such as Lynch syndrome (hereditary nonpolyposis colorectal cancer) and familial adenomatous polyposis (FAP).

These results provide further evidence to support the association of these features with Bladder Cancer. The consistency of the feature selection method in selecting these features may mean that they have important information in the diagnosis and prediction of Bladder Cancer. Further research can explore the role of these features. Biological mechanisms and clinical significance can help to better understand their role in the development of Bladder Cancer. In addition, these characteristics can also serve as the basis for Bladder Cancer risk assessment models and contribute to the early detection and prevention of Bladder Cancer occurrence.

5.2. Feature Importance Ranking

In terms of model explanation, some studies have begun to use SHAP (Shapley Additive explanation) [101]. SHAP is a method used in machine learning and artificial intelligence to explain the output of complex predictive models. It is particularly useful in situations where you want to understand why a model made a specific prediction or decision. Many machine learning models, such as deep neural networks or gradient-boosted trees, are often considered "black boxes" because they lack transparency. SHAP helps make these models more interpretable by providing insights into the contribution of each input feature to the model's predictions. SHAP values allow one to rank input features based on their contribution to model predictions. This can be helpful for feature selection, understanding which features are most influential, and simplifying complex models by focusing on the most important features [102].

In the SHAP feature importance map, each feature has a SHAP value, which indicates the degree of influence of the feature on the model's prediction results. The positive or negative SHAP value indicates the positive or negative impact of the feature on the prediction results. The size of the numerical value represents the degree of impact. This study uses SHAP to select the top ten features that are important for Bladder Cancer and Cystitis from the 28 features of the MBR method, displays them in Table 13 and Figure 10, and further discusses the relationship between these features and Bladder Cancer:

1. The average SHAP values of Urine Protein and Urobilinogen are 8.85 and 1.13, respectively, which means that they have a large positive impact on predicting the outcome of Bladder Cancer, indicating that in this dataset, high urinary levels of protein and urobilinogen may be associated with increased risk of Bladder Cancer.
2. The average SHAP values of smoking and diabetes were 0.83 and 0.62, respectively, showing a positive correlation between them and Bladder Cancer, which means that smoking and diabetes may increase the risk of Bladder Cancer.
3. The average SHAP values for Hypertension, patient gender (Gender), and patient age (Age) were 0.57, 0.33, and 0.29, respectively. These characteristics were positively correlated with Bladder Cancer, indicating that patients with hypertension, male sex, and older age may be associated with a slightly increased risk of Bladder Cancer.
4. The average SHAP values of Urine Bilirubin, Strip WBC, and Drinking were 0.24, 0.21, and 0.14, respectively. These features showed a weak positive correlation with Bladder Cancer, which means increased levels of urinary bilirubin and stripped white blood cells, as well as alcohol-drinking behavior, may be associated with a slightly increased risk of Bladder Cancer.

Table 13. SHAP values of the first ten features.

Rank	Feature Name	Content	SHAP Value (Average)
1	Urine Protein	Urine Protein values	8.85
2	Urobilinogen	Urobilinogen values	1.13
3	Smoking	Whether the patient smokes	0.83
4	Diabetes	Whether the patient has diabetes	0.62
5	Hypertension	Whether the patient has high blood pressure	0.57
6	Gerder	Patient gender	0.33
7	Age	Patient age	0.29
8	Urine Bilirubin	Urine Bilirubin values	0.24
9	Strip WBC	white blood cell values	0.21
10	Drinking	Whether the patient has consumed alcohol	0.14

In summary, these discussion results provide preliminary insights into the characteristics associated with Bladder Cancer and Cystitis. Our findings have important implications for the prediction and diagnosis of Bladder Cancer and lay the foundation for further research and clinical practice. However, further research still needs to consider other factors

and interactions to fully understand the complexity of Bladder Cancer and Cystitis and its relationship with these characteristics.

5.3. Correlation Heatmap

Correlation heatmap is a graphical representation method used to visualize the correlation between features. It shows the degree of correlation between features and uses color to represent the strength of the correlation. Dark colors indicate low correlation or no correlation, bright colors indicate high correlation, and the Correlation Heatmap represents the correlation matrix between the numerical variables to be considered in the data model [103]. Figure 11 is the correlation matrix heatmap of this study, which uses a matrix The form displays the correlation coefficient or degree of correlation between features and uses different colors to indicate the strength of the correlation. If there is a high degree of correlation between multiple features, there may be a multicollinearity problem.

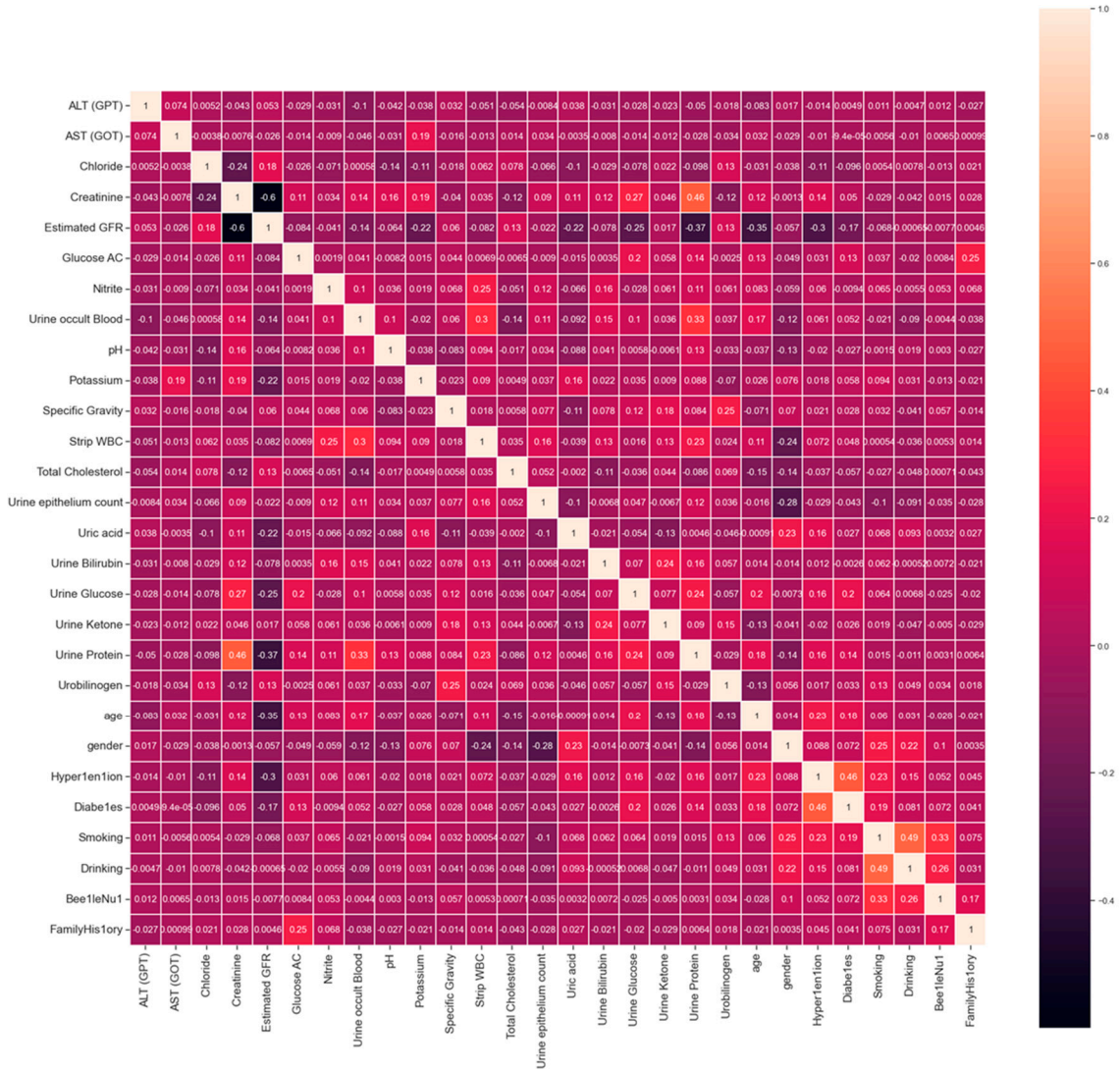

Figure 11. Correlation heatmap of our study.

According to the results of the correlation heatmap in Figure 11, this study summarizes several features with higher correlations in Table 14 and discusses them as follows:

1. There is a negative correlation between the glomerular filtration rate (Estimated GFR) and creatinine (Creatinine), which means that when the glomerular filtration rate

increases, the creatinine level decreases, which may be related to healthy kidney function.

2. There is also a negative correlation between the glomerular filtration rate (Estimated GFR), urinary protein (Urine Protein), and age (Age), which may mean that when the glomerular filtration rate decreases, the concentration of urinary protein and age may increase. These factors may be associated with the risk of kidney disease.
3. There is a positive correlation between smoking and drinking, which means that these two behaviors often occur together, which may help understand the risk factors and preventive measures for Bladder Cancer.
4. There is also a positive correlation between urinary protein (Urine Protein) and urinary occult blood (Urine occult Blood), which may imply that these two characteristics are related to the diagnosis or monitoring of Bladder Cancer. Changes in these characteristics may reflect the development and progression of Bladder Cancer.

Table 14. Feature correlation coefficient ranking.

Feature (1)	Feature (2)	Correlation Coefficient
Smoking	Drinking	0.49
Urine Protein	Creatinine	0.46
Estimated GFR	Urine Protein	−0.37
Estimated GFR	Age	−0.35
Urine Protein	Urine occult Blood	0.33
Estimated GFR	Creatinine	−0.6

In summary, the results of these correlation matrix heat maps provide valuable information for our understanding of the relevant characteristics and possible risk factors that distinguish Bladder Cancer from Cystitis and help assess the risk of Bladder Cancer and formulate corresponding preventive measures. Further studies can build on these findings to gain insight into these characteristics and correlates of distinguishing between Bladder Cancer and Cystitis. According to the empirical study of yuan [104], SHAP interpretation will fluctuate when small background datasets are used, and these fluctuations will decrease when the sample size of background datasets increases. However, evaluations like the BLEU score and the Jaccard index indicate that SHAP is more reliable in ranking the most and least important than the moderately important variables.

6. Conclusions

Bladder Cancer is a heterogeneous disease. Patients with Bladder Cancer may initially be misdiagnosed as Cystitis or infection. Accurate diagnosis of early Bladder Cancer is the key to successful treatment. This study proposes a deep learning model to distinguish Cystitis from Bladder Cancer. This model can classify these two diseases more accurately. However, due to the imbalance of the dataset, we found difficulties in classification. Therefore, this study used data imbalanced technology (SMOTE) to balance the dataset. In addition, we compared six different classification models based on deep learning algorithms, namely ANN, DNN, GRU, DBN, LSTM, and MLP. By comparing these models, we found that the MLP model performed better than the other five deep learning models. Therefore, we further improved the MLP deep learning model by using three different feature selection methods and multiple optimization algorithms to achieve better accuracy. In order to further verify the effect of the deep learning model proposed in this study, this study compared it with the lightGBM model proposed by Tsai et al. [29]. The experimental results show that using the feature ranking technology based on the learning model as the feature selection method is effective. Using principal component analysis (PCA) for dimensionality reduction can achieve the best accuracy, with an accuracy rate of 89.03%.

This result is better than past research results and can effectively reduce the possibility of misdiagnosis and provide a better clinical diagnosis tool.

Although the MLP model proposed in this study has good classification performance for Cystitis and Bladder Cancer, it still has limitations. First, all training data are from Mackay Memorial Hospital, excluding multi-center data. The dataset for training and validating the MLP model should include more data, and further incorporating multi-center data to improve the generalization performance of the model will be an important consideration before the model is widely used in clinical practice. In the future, the efficiency of evaluation indicators can also be improved by adopting different data preprocessing methods, such as the Kalman Filter technology [105]. Moreover, we will continue to collect data from other hospitals and include these data in training to strengthen the model's generalization ability, further improve the performance and reliability of the model, and provide better support for clinical diagnosis and treatment.

Author Contributions: Conceptualization, D.-H.S.; formal analysis, C.-X.L.; funding acquisition, D.-H.S. and P.-L.S.; investigation, T.-W.W.; methodology, D.-H.S., P.-L.S. and C.-X.L.; project administration, D.-H.S.; resources, D.-H.S. and P.-L.S.; software, T.-W.W. and C.-X.L.; supervision, M.-H.S.; validation, P.-L.S., T.-W.W. and M.-H.S.; visualization, M.-H.S.; writing—original draft, C.-X.L.; writing—review and editing, T.-W.W. and M.-H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the National Science and Technology Council, Taiwan (grant number NSTC 112-2410-H-224-007). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability Statement: The dataset can be obtained from Tsai et al. [29].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ferlay, J.; Colombet, M.; Soerjomataram, I.; Parkin, D.M.; Piñeros, M.; Znaor, A.; Bray, F. Cancer statistics for the year 2020: An overview. *Int. J. Cancer* **2021**, *149*, 778–789. [CrossRef] [PubMed]
2. Panebianco, V.; Pecoraro, M.; Del Giudice, F.; Takeuchi, M.; Muglia, V.F.; Messina, E.; Cipollari, S.; Giannarini, G.; Catalano, C.; Narumi, Y. VI-RADS for Bladder Cancer: Current applications and future developments. *J. Magn. Reson. Imaging* **2022**, *55*, 23–36. [CrossRef] [PubMed]
3. Saginala, K.; Barsouk, A.; Aluru, J.S.; Rawla, P.; Padala, S.A.; Barsouk, A. Epidemiology of Bladder Cancer. *Med. Sci.* **2020**, *8*, 15. [CrossRef] [PubMed]
4. Svatek, R.S.; Hollenbeck, B.K.; Holmäng, S.; Lee, R.; Kim, S.P.; Stenzl, A.; Lotan, Y. The economics of Bladder Cancer: Costs and considerations of caring for this disease. *Eur. Urol.* **2014**, *66*, 253–262. [CrossRef]
5. Antoni, S.; Ferlay, J.; Soerjomataram, I.; Znaor, A.; Jemal, A.; Bray, F. Bladder Cancer incidence and mortality: A global overview and recent trends. *Eur. Urol.* **2017**, *71*, 96–108. [CrossRef] [PubMed]
6. Leow, J.J.; Cole, A.P.; Seisen, T.; Bellmunt, J.; Mossanen, M.; Menon, M.; Preston, M.A.; Choueiri, T.K.; Kibel, A.S.; Chung, B.I. Variations in the costs of radical cystectomy for Bladder Cancer in the USA. *Eur. Urol.* **2018**, *73*, 374–382. [CrossRef]
7. Kaufman, D.S.; Shipley, W.U.; Feldman, A.S. Bladder Cancer. *Lancet* **2009**, *374*, 239–249. [CrossRef]
8. Kamat, A.M.; Hahn, N.M.; Efstathiou, J.A.; Lerner, S.P.; Malmström, P.-U.; Choi, W.; Guo, C.C.; Lotan, Y.; Kassouf, W. Bladder Cancer. *Lancet* **2016**, *388*, 2796–2810. [CrossRef]
9. Carel, R.; Silverberg, D.; Kaminsky, R.; Aviram, A. Routine urinalysis (dipstick) findings in mass screening of healthy adults. *Clin. Chem.* **1987**, *33*, 2106–2108. [CrossRef]
10. Jhamb, M.; Lin, J.; Ballow, R.; Kamat, A.M.; Grossman, H.B.; Wu, X. Urinary tract diseases and Bladder Cancer risk: A case–control study. *Cancer Causes Control.* **2007**, *18*, 839–845. [CrossRef]
11. Duldulao, K.E.; Diokno, A.C.; Mitchell, B. Value of urinary cytology in women presenting with urge incontinence and/or irritative voiding symptoms. *J. Urol.* **1997**, *157*, 113–116. [CrossRef]
12. Tissot, W.D.; Diokno, A.C.; Peters, K.M. A referral center's experience with transitional cell carcinoma misdiagnosed as interstitial Cystitis. *J. Urol.* **2004**, *172*, 478–480. [CrossRef] [PubMed]
13. Wong-You–Cheong, J.J.; Woodward, P.J.; Manning, M.A.; Davis, C.J. Inflammatory and nonneoplastic bladder masses: Radiologic-pathologic correlation. *Radiographics* **2006**, *26*, 1847–1868. [CrossRef]
14. Wu, E.Q.; Birnbaum, H.; Mareva, M.; Parece, A.; Huang, Z.; Mallett, D.; Taitel, H. Interstitial Cystitis: Cost, treatment and co-morbidities in an employed population. *Pharmacoeconomics* **2006**, *24*, 55–65. [CrossRef] [PubMed]

15. Kiliç, S.; Erguvan, R.; İpek, D.; Gökçe, H.; Güneş, A.; Aydin, N.; Baydinç, C. Polypoid Cystitis unrelated to indwelling catheters. *Int. Urol. Nephrol.* **2002**, *34*, 293–297. [CrossRef] [PubMed]
16. Hindmarsh, J.T.; Lyon, A.W. Strategies to promote rational clinical chemistry test utilization. *Clin. Biochem.* **1996**, *29*, 291–299. [CrossRef]
17. Echeverry, G.; Hortin, G.L.; Rai, A.J. Introduction to Urinalysis: Historical Perspectives and Clinical Application. In *The Urinary Proteome*; Humana Press: Totowa, NJ, USA, 2010; pp. 1–12.
18. Simerville, J.A.; Maxted, W.C.; Pahira, J.J. Urinalysis: A comprehensive review. *Am. Fam. Physician* **2005**, *71*, 1153–1162.
19. Lillian, A.; Mundt, K. Chemical analysis of urine. In *Graff's Textbook of Routine Urinalysis and Body Fluids*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2010; Volume 1, pp. 35–53.
20. Cavanaugh, C.; Perazella, M.A. Urine sediment examination in the diagnosis and management of kidney disease: Core curriculum 2019. *Am. J. Kidney Dis.* **2019**, *73*, 258–272. [CrossRef]
21. Ismail, A.A. When laboratory tests can mislead even when they appear plausible. *Clin. Med.* **2017**, *17*, 329. [CrossRef]
22. Haymond, S.; McCudden, C. Rise of the machines: Artificial intelligence and the clinical laboratory. *J. Appl. Lab. Med.* **2021**, *6*, 1640–1654. [CrossRef]
23. Parmar, C.; Grossmann, P.; Bussink, J.; Lambin, P.; Aerts, H.J. Machine learning methods for quantitative radiomic biomarkers. *Sci. Rep.* **2015**, *5*, 13087. [CrossRef] [PubMed]
24. Banerjee, M.; Reynolds, E.; Andersson, H.B.; Nallamothu, B.K. Tree-based analysis: A practical approach to create clinical decision-making tools. *Circ. Cardiovasc. Qual. Outcomes* **2019**, *12*, e004879. [CrossRef] [PubMed]
25. Chang, W.; Liu, Y.; Xiao, Y.; Yuan, X.; Xu, X.; Zhang, S.; Zhou, S. A machine-learning-based prediction method for hypertension outcomes based on medical data. *Diagnostics* **2019**, *9*, 178. [CrossRef]
26. Zhang, J.; Mucs, D.; Norinder, U.; Svensson, F. LightGBM: An effective and scalable algorithm for prediction of chemical toxicity–application to the Tox21 and mutagenicity data sets. *J. Chem. Inf. Model.* **2019**, *59*, 4150–4158. [CrossRef] [PubMed]
27. Garapati, S.S.; Hadjiiski, L.; Cha, K.H.; Chan, H.P.; Caoili, E.M.; Cohan, R.H.; Weizer, A.; Alva, A.; Paramagul, C.; Wei, J. Urinary Bladder Cancer staging in CT urography using machine learning. *Med. Phys.* **2017**, *44*, 5814–5823. [CrossRef] [PubMed]
28. Kouznetsova, V.L.; Kim, E.; Romm, E.L.; Zhu, A.; Tsigelny, I.F. Recognition of early and late stages of Bladder Cancer using metabolites and machine learning. *Metabolomics* **2019**, *15*, 94. [CrossRef] [PubMed]
29. Tsai, I.-J.; Shen, W.-C.; Lee, C.-L.; Wang, H.-D.; Lin, C.-Y. Machine learning in prediction of Bladder Cancer on clinical laboratory data. *Diagnostics* **2022**, *12*, 203. [CrossRef]
30. Haldar, S.; Dru, C.; Bhowmick, N.A. Mechanisms of hemorrhagic Cystitis. *Am. J. Clin. Exp. Urol.* **2014**, *2*, 199.
31. Rovner, E.; Propert, K.J.; Brensinger, C.; Wein, A.J.; Foy, M.; Kirkemo, A.; Landis, J.R.; Kusek, J.W.; Nyberg, L.M.; Interstitial Cystitis Data Base Study Group. Treatments used in women with interstitial Cystitis: The interstitial Cystitis data base (ICDB) study experience. *Urology* **2000**, *56*, 940–945. [CrossRef]
32. Homma, Y.; Ueda, T.; Tomoe, H.; Lin, A.T.; Kuo, H.C.; Lee, M.H.; Lee, J.G.; Kim, D.Y.; Lee, K.S.; Committee IC, GInterstitial Cystitis Guideline Committee. Clinical guidelines for interstitial Cystitis and hypersensitive bladder syndrome. *Int. J. Urol.* **2009**, *16*, 597–615. [CrossRef]
33. Chancellor, M.B.; Yoshimura, N. Treatment of interstitial Cystitis. *Urology* **2004**, *63*, 85–92. [CrossRef] [PubMed]
34. Brown, E.W. Eosinophilic granuloma of the bladder. *J. Urol.* **1960**, *83*, 665–668. [CrossRef] [PubMed]
35. Sparks, S.; Kaplan, A.; DeCambre, M.; Kaplan, G.; Holmes, N. Eosinophilic Cystitis in the pediatric population: A case series and review of the literature. *J. Pediatr. Urol.* **2013**, *9*, 738–744. [CrossRef] [PubMed]
36. Mosholt KS, S.; Dahl, C.; Azawi, N.H. Eosinophilic Cystitis: Three cases, and a review over 10 years. *Case Rep.* **2014**, *2014*, bcr2014205708. [CrossRef]
37. Sanli, O.; Dobruch, J.; Knowles, M.A.; Burger, M.; Alemozaffar, M.; Nielsen, M.E.; Lotan, Y. Bladder Cancer. *Nat. Rev. Dis. Prim.* **2017**, *3*, 17022. [CrossRef]
38. Stroman, L.; Issa, R. Bladder Cancer. *Surgery* **2022**, *40*, 674–682. [CrossRef]
39. Hashemi, M.; Arani, H.Z.; Orouei, S.; Rostamnejad, E.; Ghorbani, A.; Khaledabadi, M.; Kakavand, A.; Tavakolpournegari, A.; Saebfar, H.; Heidari, H. Crosstalk of miRNAs with signaling networks in Bladder Cancer progression: Therapeutic, diagnostic and prognostic functions. *Pharmacol. Res.* **2022**, *185*, 106475. [CrossRef]
40. Clark, P.E.; Agarwal, N.; Biagioli, M.C.; Eisenberger, M.A.; Greenberg, R.E.; Herr, H.W.; Inman, B.A.; Kuban, D.A.; Kuzel, T.M.; Lele, S.M. Bladder Cancer. *J. Natl. Compr. Cancer Netw.* **2013**, *11*, 446–475. [CrossRef]
41. Sharp, V.J.; Barnes, K.T.; Erickson, B.A. Assessment of asymptomatic microscopic hematuria in adults. *Am. Fam. Physician* **2013**, *88*, 747–754.
42. Kantor, A.F.; Hartge, P.; Hoover, R.N.; Narayana, A.S.; Sullivan, J.; Fraumeni, J.R.J. Urinary tract infection and risk of Bladder Cancer. *Am. J. Epidemiol.* **1984**, *119*, 510–515. [CrossRef]
43. Thorsteinsson, K.; Jensen, J.B. Misinterpretation resulting in a diagnosis of Bladder Cancer–A case emphasising the value of diagnostic reconsideration. *Urol. Case Rep.* **2022**, *40*, 101928. [CrossRef]
44. Cruz, J.A.; Wishart, D.S. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform.* **2006**, *2*, 117693510600200030. [CrossRef]
45. Millan-Rodriguez, F.; Chechile-Toniolo, G.; Salvador-Bayarri, J.; Palou, J.; Vicente-Rodriguez, J. Multivariate analysis of the prognostic factors of primary superficial Bladder Cancer. *J. Urol.* **2000**, *163*, 73–78. [CrossRef]

46. Bassi, P.; Sacco, E.; De Marco, V.; Aragona, M.; Volpe, A. Prognostic accuracy of an artificial neural network in patients undergoing radical cystectomy for Bladder Cancer: A comparison with logistic regression analysis. *BJU Int.* **2007**, *99*, 1007–1012. [CrossRef] [PubMed]
47. Wang, G.; Lam, K.-M.; Deng, Z.; Choi, K.-S. Prediction of mortality after radical cystectomy for Bladder Cancer by machine learning techniques. *Comput. Biol. Med.* **2015**, *63*, 124–132. [CrossRef] [PubMed]
48. Weston, A.D.; Hood, L. Systems biology, proteomics, and the future of health care: Toward predictive, preventative, and personalized medicine. *J. Proteome Res.* **2004**, *3*, 179–196. [CrossRef] [PubMed]
49. Mccarthy, J.F.; Marx, K.A.; Hoffman, P.E.; Gee, A.G.; O'neil, P.; Ujwal, M.L.; Hotchkiss, J. Applications of machine learning and high-dimensional visualization in cancer detection, diagnosis, and management. *Ann. N. Y. Acad. Sci.* **2004**, *1020*, 239–262. [CrossRef]
50. Vukicevic, A.M.; Jovicic, G.R.; Stojadinovic, M.M.; Prelevic, R.I.; Filipovic, N.D. Evolutionary assembled neural networks for making medical decisions with minimal regret: Application for predicting advanced Bladder Cancer outcome. *Expert Syst. Appl.* **2014**, *41*, 8092–8100. [CrossRef]
51. Ji, W.; Naguib RN, G.; Ghoneim, M.A. Neural network-based assessment of prognostic markers and outcome prediction in bilharziasis-associated Bladder Cancer. *IEEE Trans. Inf. Technol. Biomed.* **2003**, *7*, 218–224. [CrossRef]
52. Qureshi, K.N.; Naguib, R.N.; Hamdy, F.C.; Neal, D.E.; Mellon, J. Neural network analysis of clinicopathological and molecular markers in Bladder Cancer. *J. Urol.* **2000**, *163*, 630–633. [CrossRef]
53. Freitas, N.R.; Vieira, P.M.; Cordeiro, A.; Tinoco, C.; Morais, N.; Torres, J.; Anacleto, S.; Laguna, M.P.; Lima, E.; Lima, C.S. Detection of Bladder Cancer with feature fusion, transfer learning and CapsNets. *Artif. Intell. Med.* **2022**, *126*, 102275. [CrossRef] [PubMed]
54. Grekousis, G. Artificial neural networks and deep learning in urban geography: A systematic review and meta-analysis. *Comput. Environ. Urban Syst.* **2019**, *74*, 244–256. [CrossRef]
55. Liu, H.; Lang, B. Machine learning and deep learning methods for intrusion detection systems: A survey. *Appl. Sci.* **2019**, *9*, 4396. [CrossRef]
56. Montavon, G.; Samek, W.; Müller, K.-R. Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.* **2018**, *73*, 1–15. [CrossRef]
57. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
58. Hinton, G.E. Deep belief networks. *Scholarpedia* **2009**, *4*, 5947. [CrossRef]
59. Chen, X.-M.; Wu, C.-X.; Wu, Y.; Xiong, N.-X.; Han, R.; Ju, B.-B.; Zhang, S. Design and analysis for early warning of rotor UAV based on data-driven DBN. *Electronics* **2019**, *8*, 1350. [CrossRef]
60. Hinton, G.E.; Osindero, S.; Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [CrossRef]
61. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
62. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
63. Gardner, M.W.; Dorling, S. Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [CrossRef]
64. Hippert, H.S.; Pedreira, C.E.; Souza, R.C. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Trans. Power Syst.* **2001**, *16*, 44–55. [CrossRef]
65. Edmond, C.; Girsang, A.S. Classification performance for credit scoring using neural network. *Int. J.* **2020**, *8*, 1592–1599. [CrossRef]
66. Bousquet, O.; von Luxburg, U.; Rätsch, G. *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, 2–14 February 2003, Tübingen, Germany, 4–16 August 2003, Revised Lectures*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 3176.
67. Patrician, P.A. Multiple imputation for missing data. *Res. Nurs. Health* **2002**, *25*, 76–84. [CrossRef]
68. Choi, J.; Dekkers, O.M.; le Cessie, S. A comparison of different methods to handle missing data in the context of propensity score analysis. *Eur. J. Epidemiol.* **2019**, *34*, 23–36. [CrossRef]
69. Klebanoff, M.A.; Cole, S.R. Use of multiple imputation in the epidemiologic literature. *Am. J. Epidemiol.* **2008**, *168*, 355–357. [CrossRef]
70. Sterne, J.A.; White, I.R.; Carlin, J.B.; Spratt, M.; Royston, P.; Kenward, M.G.; Wood, A.M.; Carpenter, J.R. Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *BMJ* **2009**, *338*, b2393. [CrossRef]
71. Graham, J.W. Missing data analysis: Making it work in the real world. *Annu. Rev. Psychol.* **2009**, *60*, 549–576. [CrossRef]
72. Pedersen, A.B.; Mikkelsen, E.M.; Cronin-Fenton, D.; Kristensen, N.R.; Pham, T.M.; Pedersen, L.; Petersen, I. Missing data and multiple imputation in clinical epidemiological research. *Clin. Epidemiol.* **2017**, *9*, 157. [CrossRef]
73. Sanchez-Pinto, L.N.; Venable, L.R.; Fahrenbach, J.; Churpek, M.M. Comparison of variable selection methods for clinical predictive modeling. *Int. J. Med. Inform.* **2018**, *116*, 10–17. [CrossRef]
74. Senan, E.M.; Al-Adhaileh, M.H.; Alsaade, F.W.; Aldhyani, T.H.; Alqarni, A.A.; Alsharif, N.; Uddin, M.I.; Alahmadi, A.H.; Jadhav, M.E.; Alzahrani, M.Y. Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques. *J. Healthc. Eng.* **2021**, *2021*, 1004767. [CrossRef] [PubMed]
75. Plackett, R.L. Karl Pearson and the chi-squared test. In *International Statistical Review/Revue Internationale de Statistique*; International Statistical Institute: The Hague, The Netherlands, 1983; pp. 59–72.

76. Devi, L.; Subathra, P.; Kumar, P. Tweet sentiment classification using an ensemble of machine learning supervised classifiers employing statistical feature selection methods. In *Proceedings of the Fifth International Conference on Fuzzy and Neuro Computing (FANCCO-2015)*; Springer: Hyderabad, India, 2015.
77. Nissim, N.; Moskovitch, R.; Rokach, L.; Elovici, Y. Detecting unknown computer worm activity via support vector machines and active learning. *Pattern Anal. Appl.* **2012**, *15*, 459–475. [CrossRef]
78. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
79. Shen, F.; Zhao, X.; Kou, G.; Alsaadi, F.E. A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique. *Appl. Soft Comput.* **2021**, *98*, 106852. [CrossRef]
80. He, H.; Bai, Y.; Garcia, E.A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–6 June 2008.
81. Lu, C.; Lin, S.; Liu, X.; Shi, H. Telecom fraud identification based on ADASYN and random forest. In Proceedings of the 2020 5th International Conference on Computer and Communication Systems (ICCCS), Shanghai, China, 22–24 February 2020.
82. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]
83. Pearson, K., III. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [CrossRef]
84. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417. [CrossRef]
85. Vidal, R.; Ma, Y.; Sastry, S.S. Principal component analysis. In *Generalized Principal Component Analysis*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 25–62.
86. Zahid, L.; Maqsood, M.; Durrani, M.Y.; Bakhtyar, M.; Baber, J.; Jamal, H.; Mehmood, I.; Song, O.-Y. A spectrogram-based deep feature assisted computer-aided diagnostic system for Parkinson's disease. *IEEE Access* **2020**, *8*, 35482–35495. [CrossRef]
87. Lee, S.; Jin, H.; Vecchietti, L.F.; Hong, J.; Park, K.-B.; Har, D. Power Management of Nanogrid Cluster with P2P Electricity Trading Based on Future Trends of Load Demand and PV Power Production. *arXiv* **2020**, arXiv:2009.00863.
88. Heikal, M.; Torki, M.; El-Makky, N. Sentiment analysis of Arabic tweets using deep learning. *Procedia Comput. Sci.* **2018**, *142*, 114–122. [CrossRef]
89. Gunawan, W.; Suhartono, D.; Purnomo, F.; Ongko, A. Named-entity recognition for indonesian language using bidirectional lstm-cnns. *Procedia Comput. Sci.* **2018**, *135*, 425–432. [CrossRef]
90. Arabzad, S.M.; Tayebi Araghi, M.E.; Sadi-Nezhad, S.; Ghofrani, N. Football match results prediction using artificial neural networks; the case of Iran Pro League. *J. Appl. Res. Ind. Eng.* **2014**, *1*, 159–179.
91. Helwan, A.; Tantua, D.P. IKRAI: Intelligent knee rheumatoid arthritis identification. *Int. J. Intell. Syst. Appl.* **2016**, *8*, 18. [CrossRef]
92. Suparwito, H.; Thomas, D.T.; Wong, K.W.; Xie, H.; Rai, S. The use of animal sensor data for predicting sheep metabolisable energy intake using machine learning. *Inf. Process. Agric.* **2021**, *8*, 494–504. [CrossRef]
93. Mirjalili, S.; Seyed, M.M.; Andrew, L. Greywolfoptimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]
94. Rao, P.S.; Parida, P.; Sahu, G.; Dash, S. A multi-view human gait recognition using hybrid whale and gray wolf optimization algorithm with a random forest classifier. *Image Vis. Comput.* **2023**, *136*, 104721. [CrossRef]
95. Abualigah, L.; Shehab, M.; Alshinwan, M.; Alabool, H. Salp swarm algorithm: A comprehensive survey. *Neural Comput. Appl.* **2020**, *32*, 11195–11215. [CrossRef]
96. Mirjalili, S. Sca: A sine cosine algorithm for solving optimization problems. *Knowl. Based Syst.* **2016**, *96*, 120–133. [CrossRef]
97. Yang, X.S.; Deb, S. Cuckoo search via Lévy flights. In Proceedings of the 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), Coimbatore, India, 9–11 December 2009; pp. 210–214.
98. Tong, F.; Shahid, M.; Jin, P.; Jung, S.; Kim, W.H.; Kim, J. Classification of the urinary metabolome using machine learning and potential applications to diagnosing interstitial Cystitis. *Bladder* **2020**, *7*, e43. [CrossRef]
99. Yu, X.; Wang, R.; Han, C.; Wang, Z.; Jin, X. A panel of urinary long non-coding RNAs differentiate Bladder Cancer from uroCystitis. *J. Cancer* **2020**, *11*, 781. [CrossRef]
100. Lalkhen, A.G.; McCluskey, A. Clinical tests: Sensitivity and specificity. *Contin. Educ. Anaesth. Crit. Care Pain* **2008**, *8*, 221–223. [CrossRef]
101. Ribeiro, M.T.; Singh, S.; Guestrin, C. Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
102. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS2017), Long Beach, CA, USA, 4–9 December 2017.
103. Musić, A.; Telalović, J.H.; Đulović, D. The Influence of Stringency Measures and Socio-Economic Data on COVID-19 Outcomes. In *Communications in Computer and Information Science, Proceedings of the Mediterranean Forum—Data Science Conference, First International Conference, MeFDATA 2020, Sarajevo, Bosnia and Herzegovina, 24 October 2020*; Hasic Telalovic, J., Kantardzic, M., Eds.; Springer: Cham, Switzerland, 2021; Volume 1343. [CrossRef]

104. Yuan, H.; Liu, M.; Kang, L.; Miao, C.; Wu, Y. An empirical study of the effect of background data size on the stability of SHapley Additive exPlanations (SHAP) for deep learning models. *arXiv* **2022**, arXiv:2204.11351.
105. Karthik, S.; Singh Bhadoria, R.; Lee, J.G.; Kumar Sivaraman, A.; Samanta, S.; Balasundaram, A.; Chaurasia, B.K.; Ashokkumar, S. Prognostic Kalman Filter Based Bayesian Learning Model for Data Accuracy Prediction. *Comput. Mater. Contin.* **2022**, *72*, 244–259. [CrossRef]