Comparative Performance of Urine and Blood-based Attributes in Chronic Kidney Disease Prediction using Artificial Neural Network

M. H. A. Ghafar , A. H. A. Razak, M. S. A. Megat Ali, S. A. M. Al Junid, A. Ahmad, M. F. A. Latip, M. N. Taib, M. Fathimah

Abstract—**Chronic Kidney Disease is one of the leading causes of death worldwide. An intelligent diagnostic system capable of early detection is therefore, becoming increasingly important. These would allow effective intervention to be delivered to patients, thus prolonging kidney function and reducing risk of mortality. The system should be non-invasive, convenient, accurate, and reliable in detecting the required attributes. This study compares between urine- and blood-based attributes in acute renal failure prediction using artificial neural network. A total of 400 sample data is obtained from UCI Machine Learning Repository. Multiple imputation is then implemented to generate synthetic data. These overcome the issue of missing datapoints and unbalanced sample distribution. Two artificial neural network models are trained. One using the urine-based attributes and the other, using blood-based attributes. Both models attained excellent classification accuracies of 96.0% and 98.0%, respectively. However, the ANN model developed based on urine-based attributes are recommended intelligent diagnostic systems due to lower computational requirements and the sample acquisition protocol is much convenient for patients and medical practitioners.**

Index Terms—**Acute renal failure, blood, urine, multiple imputation, artificial neural network**

I. INTRODUCTION

IDENTIFYING and profiling important attributes of chronic diseases have led to the development intelligent diagnostic systems that can perform early detection. These allow medical practitioners to anticipate and deliver effective intervention to patients, thus prolonging the function of critical organs and reducing risk of death [1]. Chronic Kidney Disease (CKD) is

This manuscript is submitted on 7th August 2020 and accepted on 24th November 2020. M. H. A. Ghafar, A. H. A. Razak, and S. A. M. Al Junid is with Electronics Architecture and Application Research Group (EArA), Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia. A. H. A. Razak is the corresponding author (e-mail: hadi@ieee.org).M. S. A. Megat Ali is with Microwave Research Institute, Universiti Teknologi MARA, Shah Alam, Malaysia.

A. Ahmad, M. F. A. Latip, and M. N. Taib is with Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia.

M. Fathimah is with the Department of Pathology, Faculty of Medicine, Universiti Teknologi MARA, Sungai Buloh, Malaysia.

the fastest growing cause of death with at least 2.4 million cases reported every year [2].

CKD is defined as kidney damage, or glomerular filtration rate of less than 60ml/min/1.73m² for duration of more than three months [3]. Excessive consumption of sugar can lead the causes of kidney damage. High volume of glucose will increase the function of kidney to filtering wastes and extra fluid from the blood [4]. Thus, the initial sign of CKD is existence protein in urine called albumin. A healthy kidney does not let albumin pass from blood into the urine [5]. CKD is divided into five stages with the worse being the end-stage renal disease. This requires the patient to undergo kidney replacement [6].

Based on the reports from American Society of Nephrology, International Society of Nephrology, European Renal Association-European Dialysis and Transplant Association in 2017, the total number of individuals with CKD, acute kidney injury, and those on renal replacement therapy exceeds 850 million. The figure doubles the number of diabetes cases. The Associations also predicted that CKD will be the top 5 causes of death worldwide [7]. The Institute of Health Metric and Evaluation reported 14% of women and 12% of men are affected by the diseases [8]. Based on the 24th Report of Malaysian Dialysis and Transplant Registry in 2016, the number of CKD cases steadily increased for the past 10 years [9].

Previous work related to CKD focuses on detection of the disease and its stages from various diagnostic methods. Among of the diagnosis method is feature extraction from the ultrasound kidney images [10]. In other method, CKD also can be predicted by the parameter of two main clinical tests, which are the urine and blood tests. Urine test or Urinalysis includes three parts of exam which are visual, microscopic and dipstick test. Several parameters can be extracted from urine test such that albuminuria, specific gravity, protein and glucose. Albumin is a protein contained in the blood, but if it is found too much in urine, the kidney would not function normally and

will progress to damage. Amount of albumin in the urine is dependent on the condition of the kidney individually. In [11] states that "the less albumin in your urine, the better". As a result, albuminuria or proteinuria urine test is more significant to determine whether patient have kidney diseases or not. Another type of urine test that significant to diagnose chronic kidney disease is specific gravity. Specific gravity is a ratio of urine to water density. The range of this measurement is between 1.005 and 1.030 [12].

Meanwhile, there are several clinical blood tests for kidney diseases such as random blood glucose, blood urea nitrogen, serum creatinine, serum sodium and potassium lab test. Both blood test significant to determine kidney level because the type of data is based on continuous value. For example, blood urea nitrogen (BUN) is used to measure the amount of nitrogen in blood [13]. The level of BUN is rising when kidney is not able to remove urea from the blood normally.

These tests can be complemented by machine learning techniques that enable generalization of attributes between normal subjects and those with CKD [14]. In fact, there are already attempts to model clinical traits with different stages of the disease [15]. The established machine learning methods include convolutional neural network [16], support vector machines [17], decision tree classifier [18] and artificial neural network (ANN) [19]. The goals are to improve diagnostic systems and reduce cost of medical treatment through early interventions [20]. Despite the continual effort by various researchers, the systems have yet to attain its optimum potential. Majority of previous works are built from open-source repositories that contains missing data. The unrecorded test attributes have been not only among the types of attributes, but also across individual [21]. Studies have shown that these deficiencies have an adverse effect on prediction performance and needs to be solved prior to development of the computational model [22].

Among the widely used method for dealing with missing datapoints is multiple imputation. When compared to single imputation method, multiple imputation enables analysis of dataset and utilize it to obtain consistent synthetic data points. The method introduces statistical inference and illustrate references to the respective cases [23]. These can be achieved through three to five repetitions by adding average values to the missing data [24]. Furthermore, it can reduce errors caused by unbalanced distribution among the control groups, thus improving the model performance [25]. Multiple imputation is, therefore, an important method to improve validity of the study and reduce waste of resources caused by missing data [26].

ANN is one of the established machine learning techniques. The approach mimics the biological functioning of neurons in the brain and learn from given examples. Although considered as one of the pioneering artificial intelligence paradigms, it remains relevant due to its ability to generalize solution to a problem [27]. The method has been widely implemented in modelling physiological phenomena such as cognitive behaviour [28], dengue prediction [29], and heart diseases [30]. Thus far, ANN has also been used in studies involving CKD. However, the findings are debatable as it lacks technical clarity on the pre-processing and modelling aspects [22].

Despite the growing body of literature related to CKD prediction using artificial intelligent technique, the following gaps have been identified. 1) The literature has presented a myriad of testing methods to obtain the desired attributes. However, none has emphasized on the importance of selecting the most convenient, non-invasive, and cost-effective attributes; yet maintaining the prediction accuracy and reliability. 2) In studies related to CKD, most of the work on handling missing data and unbalanced dataset involves the use of primitive techniques such as attribute removal, and correlation process to identify the significant parameters. However, methods such as multiple imputation is equally important for accurately generating reliable data points. 3) Although there are already studies on CKD involving ANN, the findings were not well presented as it lacks the technical descriptions on both data pre-processing and development of model, rendering the results debatable.

To solve these issues, the following objectives are proposed. First, the study aims to characterize the attributes based on urine and blood tests. Imaging techniques such as computed tomography has been ruled out as it relatively more expensive than clinical laboratory tests. Second, the study aims to implement multiple imputation method to solve missing data points within the dataset and balance out data distribution between those with CKD and healthy controls. Third, two prediction models are developed. One model uses the urine-based attributes, while the other model uses the blood-based attributes. A comparative analysis is then performed to reach a conclusive recommendation.

II. METHODS

The study comprises of data acquisition, characterization of urine- and blood-based attributes, implementation of multiple imputation method for generating synthetic data, development of CKD prediction for urine- and blood-based attributes, as well as comparative analysis. These are summarized in the flowchart shown in Figure 1.
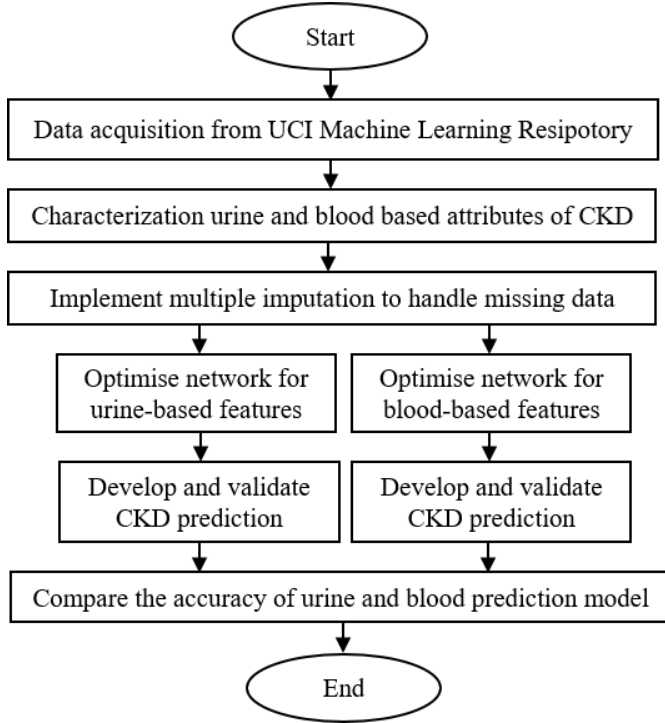
```
            ┌───────────┐
            │   Start   │
            └─────┬─────┘
                  ↓
┌─────────────────────────────────────────────────┐
│ Data acquisition from UCI Machine Learning       │
│ Resipotory                                       │
└─────────────────────┬───────────────────────────┘
                      ↓
┌─────────────────────────────────────────────────┐
│ Characterization urine and blood based attributes│
│ of CKD                                           │
└─────────────────────┬───────────────────────────┘
                      ↓
┌─────────────────────────────────────────────────┐
│ Implement multiple imputation to handle missing  │
│ data                                             │
└──────────┬──────────────────────┬───────────────┘
           ↓                      ↓
┌────────────────────┐  ┌────────────────────┐
│ Optimise network   │  │ Optimise network   │
│ for urine-based    │  │ for blood-based    │
│ features           │  │ features           │
└─────────┬──────────┘  └─────────┬──────────┘
          ↓                       ↓
┌────────────────────┐  ┌────────────────────┐
│ Develop and        │  │ Develop and        │
│ validate CKD       │  │ validate CKD       │
│ prediction         │  │ prediction         │
└─────────┬──────────┘  └─────────┬──────────┘
          ↓                       ↓
┌─────────────────────────────────────────────────┐
│ Compare the accuracy of urine and blood          │
│ prediction model                                 │
└─────────────────────┬───────────────────────────┘
                      ↓
                ┌───────────┐
                │    End    │
                └───────────┘
```

Fig.1. Overview of research methods.

A. Data Acquisition

A total of 400 data samples were acquired from the open-source UCI Machine Learning Repository [31]. The dataset is further segregated into two control groups: 250 samples in the CKD group and the remaining 150 samples as healthy controls. The database provides thirteen critical attributes. Table I shows the five urine-based and nine blood-based attributes.

TABLE I
ATTRIBUTES

Urine-based	Blood-based
Specific gravity	Blood glucose random
Albumin	Blood urea
Sugar	Serum creatinine
Pus Cell	Sodium
Pus Cell Clumps	Potassium
	Hemoglobin
	Packed cell volume
	White blood cell count
	Red blood cell count

B. Multiple Imputation

The missing data is generated by considering the average value from the overall data. Subsequently, five repetitions are made through the multiple imputation process: enhancing the total number of samples to 1000 for each control group. The mean values are only implemented for numerical attributes. Meanwhile, nominal attributes remain unchanged. The concept of imputation is based on the resulting model, a new regression model used to impute the missing value for each variable [32].

This can be expressed by (1). Y_j with missing value, a model is fitted with the non-missing observations. The fitted model has the regression parameters estimates $(\beta_o, \beta_1, \dots, \beta_{(j-1)})$ and associated covariance matrix $\sigma_j Y_j$, where Y_j is the usual matrix from the intercept and variable $Y_1, Y_2, \dots, Y_{(j-1)}$.

$$Y_j = \beta_o + \beta_1 Y_1 + \beta_2 Y_2 + \dots + \beta_{(j-1)} Y_{(j-1)} \tag{1}$$

For each imputation, new parameters $(\beta_{*o}, \beta_{*1}, \dots, \beta_{*(j-1)})$ and σ_{*j} are drawn from the posterior predictive distribution of the missing data. The missing values of the original data are then replaced by expression of (2)

$$\beta_{*o} + \beta_{*1} y_1 + \beta_{*2} y_2 + \dots + \beta_{*(j-1)} y_{(j-1)} + z_i \sigma_{*j} \tag{2}$$

Where $y_1, y_2, \dots, y_{(j-1)}$ are the covariate values of the first (j-1) variables and z_i is a simulated normal deviate.

C. Artificial Neural Network

The modelling tasks in this study is performed using MATLAB. Generally, the ANN comprises of an input layer, several hidden layers, and an output layer [33]. However, a study had shown that a network with single hidden layer is sufficient to approximate and arbitrary function of to an acceptable level of accuracy [34]. The number of input nodes will depend on the number of attributes to the network. Meanwhile, single output node is used for the study. The optimum number of hidden nodes is determined based on a separate set of optimization experiment which considers unique traits observed from network training [35].

Vector of input variables x_i is transformed into a vector of hidden variables, h_j, via activation function, Γ_1. This can be expressed by (3). w_{ij} is the connecting weights from ith input node to jth hidden node, and θ_j represents the biases. Γ_1 employs the hyperbolic tangent function.

$$h_j = \Gamma_1 \left(\sum_{i=1}^{M} w_{ij} x_i + \theta_j \right) \tag{3}$$

The vector of hidden variables, h_j, is then transformed into the output, y_k, through activation, Γ_2. This is expressed by (4). w_{jk} is the connecting weights between jth hidden node to kth output node, and θ_k represents the bias. Γ_2 employs the pure linear function.

$$y_k = \Gamma_2\left(\sum_{j=1}^{N} w_{jk}h_j + \theta_k\right) \quad (4)$$
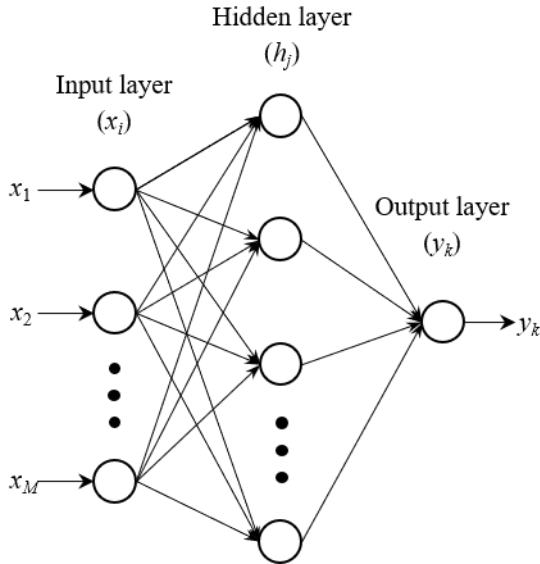
The general structure of ANN is illustrated by Fig. 2.

Fig.2. General structure of ANN

As shown by (5), the error, e is obtained as the difference between the output, y_k, and the desired output, y_t.

$$e = y_k - y_d \quad (5)$$

The error is then used in the back-propagation weight-update procedure through suitable learning algorithms. In this study, Levenberg-Marquardt algorithm is adopted. Network training will continue in an iterative manner until the error is being minimized [36]. Data is initially randomized. 70% is used for training, 15% is used for validation and the remaining 15% is used for testing. The validation set is used to avoid the network from over-fitting. Throughout training, data from the validation set is periodically used to gauge the network's generalization ability. If the validation error increases, network training is stopped and the set of network weights prior to that is used in the final structure. The performance of the model is then assessed using the testing sets [37].

Two models will be developed using ANN: one using the urine-based attributes and the other using blood-based attributes. The first model adopts a five-input structure, and the second model will take on a nine-input structure. The number of hidden nodes, however, relies on an experiment performed using the rule of thumb. The lower limit is set as 2/3 the size of input and output layers. Meanwhile, the upper limit should be less than twice the size of the input layer [38]. The experiment starts by training the ANN using the lower limit setting for 40 iterations and averages its performance. As shown in Fig.3, the procedure is then repeated until it reaches the maximum limit.
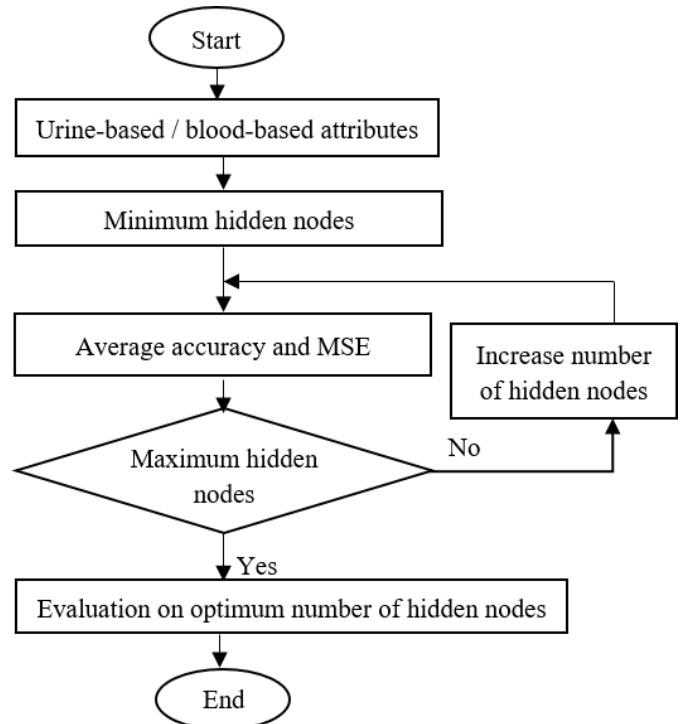
Fig.3. Flowchart to determine optimum number of hidden nodes.

The evaluation will consider the best average training accuracy and mean squared error (MSE). Accuracy (*Acc*) is expressed by (6), where *TP* is the true positives, *TN* is the true negatives, *FP* is the false positives, and *FN* is the false negatives.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (6)$$

Meanwhile, MSE is presented by (7), where $e(t)$ is the error at iteration instant, t, and N is the total number of iterations.

$$MSE = \frac{1}{N}\sum_{t=1}^{N} e(t) \quad (7)$$

In this study, samples with CKD is given the output index of 1, while healthy samples are given the output index of 2. Once the optimum number of hidden nodes is finalized, the network is trained to obtain the best results. These are assessed not only in terms of accuracy, but also sensitivity (*Se*) and positive predicitivity (*Pp*). Both parameters are each expressed by (8) and (9).

$$Se = \frac{TP}{TP + FN} \times 100\% \tag{8}$$

$$Pp = \frac{TP}{TP + FP} \times 100\% \tag{9}$$

III. RESULTS AND DISCUSSION

A. Data Acquisition

The five urine-based attributes are specific gravity, albumin, sugar, pus cell and pus cell clumps. These are comprised of both numerical and nominal data. The distribution of attributes is shown in Fig. 4 to Fig. 8
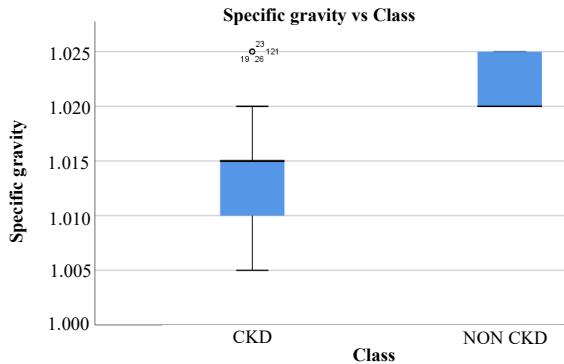
Fig.4. Specific gravity (N=400).
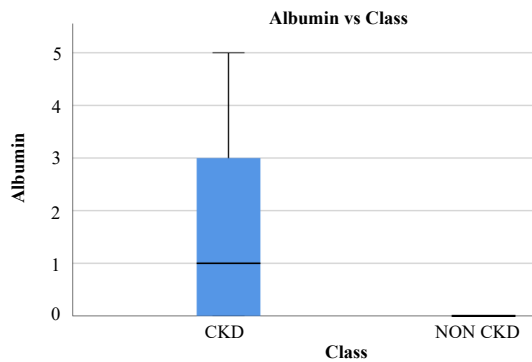
Fig.5. Albumin (N=400).
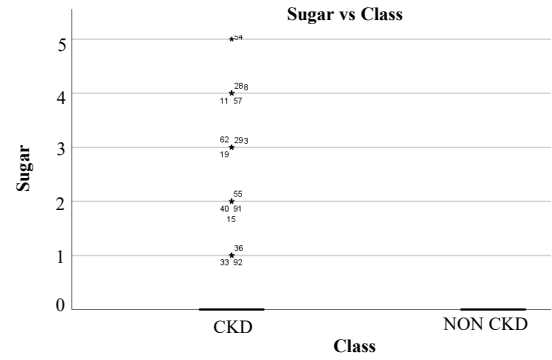
Fig.6. Sugar (N=400).
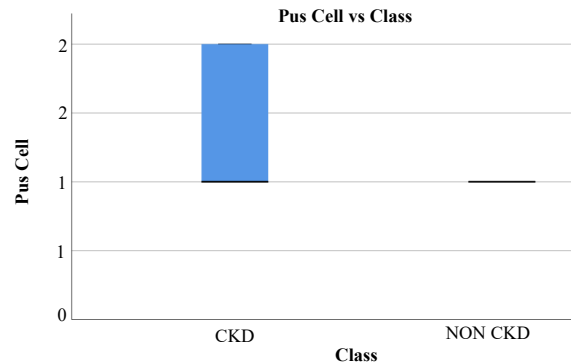
Fig.7. Pus cell (N=400).
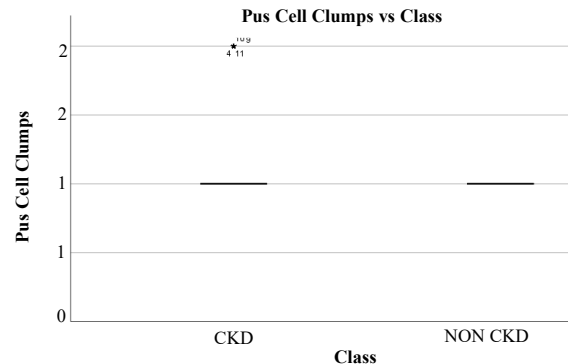
Fig.8. Pus cell clumps (N=400).

Meanwhile, the blood-based attributes consist of blood glucose, blood urea, serum creatinine, sodium, potassium, haemoglobin, packed cell volume, white blood cell count and red blood cell count. Compared to the urine, blood-based attributes are all composed of numerical data. The distribution of attributes is shown in Fig. 9 to Fig. 17.
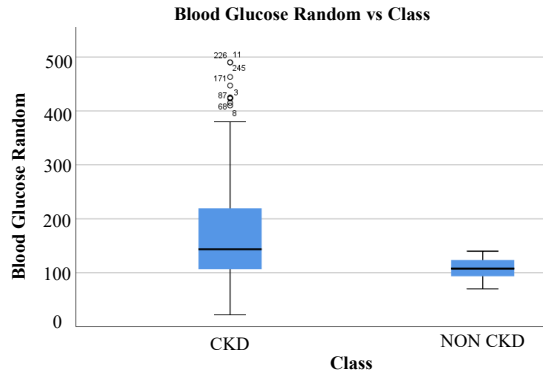
Fig.9. Blood glucose random (N=400).
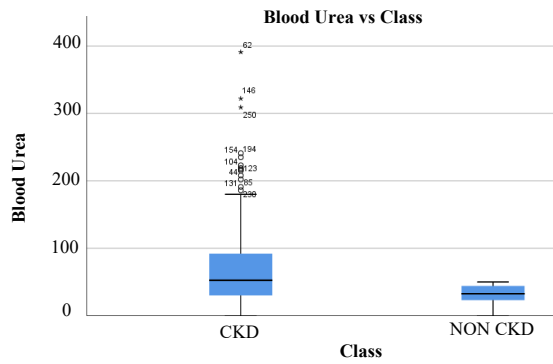
Fig.10. Blood urea (N=400).

Fig.11. Serum creatinine (N=400).

Fig.12. Sodium (N=400).

Fig.13. Potassium (N=400).
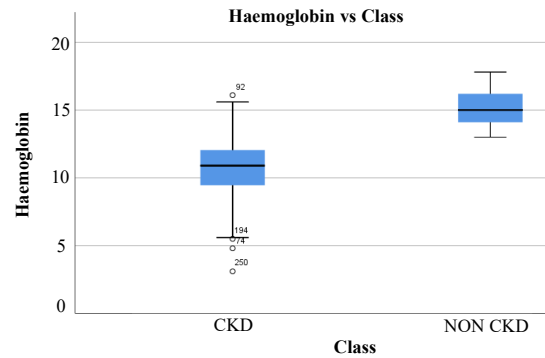
Fig.14. Haemoglobin (N=400).
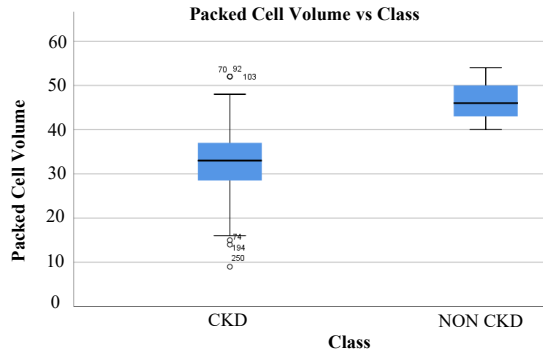
Fig.15. Packed cell volume (N=400).
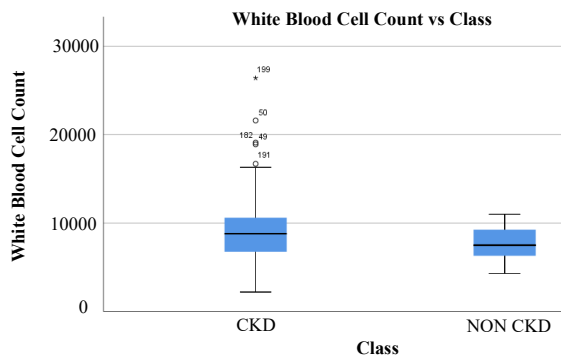
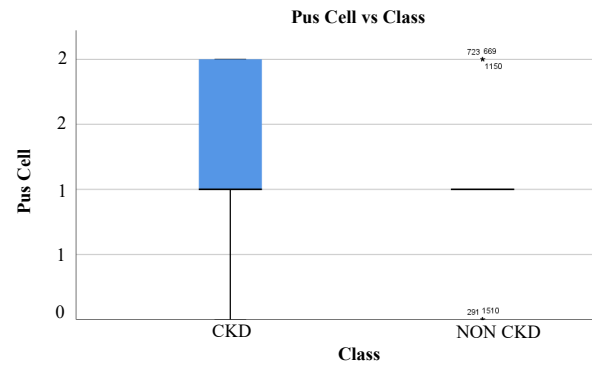

Fig.16. White blood cell count (N=400).
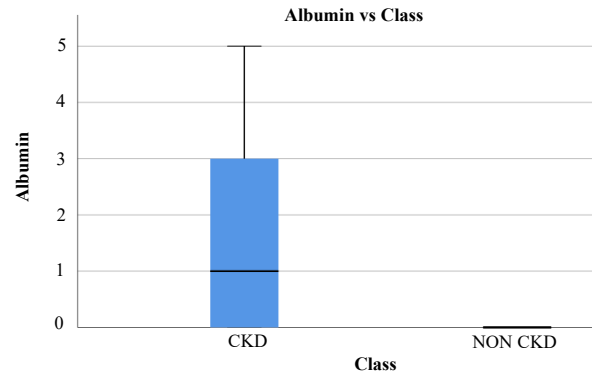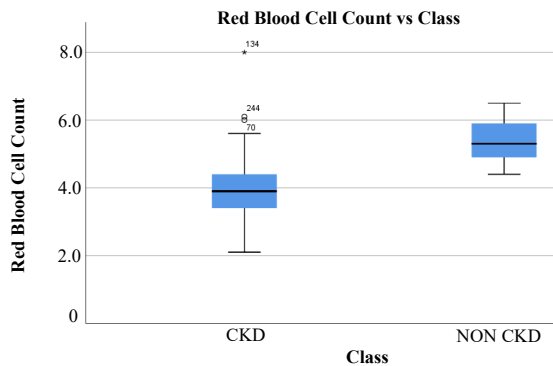
Fig.17. Red blood cell count (N=400).

B.   Multiple Imputation

There are missing data points from the original dataset. Hence, multiple imputation generates the synthetic version of them, and the overall dataset is enhanced to N=2000. The new distribution of attributes for urine-based attributes are shown in Fig. 18 to Fig. 22. The pattern of albumin, sugar, pus cell and pus cell clumps remain unchanged. Although specific gravity is takes on a nominal value, the pattern slightly differs from the original dataset.
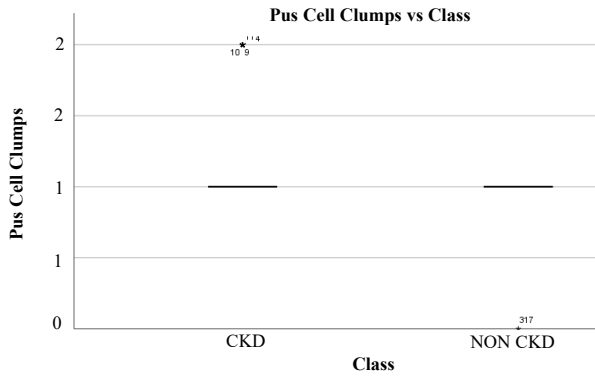

Fig.18. Specific gravity (N=2000).

Fig.19. Albumin (N=2000).
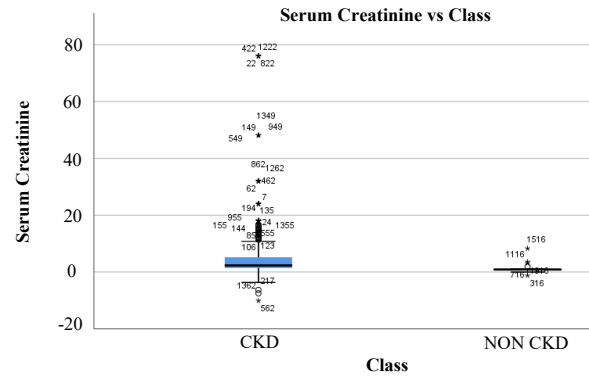

Fig.20. Sugar (N=2000).

Fig.21. Pus cell (N=2000).

Fig.22. Pus cell clumps (N=2000).

Meanwhile, the new distribution of blood-based attributes is shown in Fig. 23 to Fig. 31. The pattern for these attributes slightly differs from the original dataset but contains more outlier due to the increasing volume of samples.
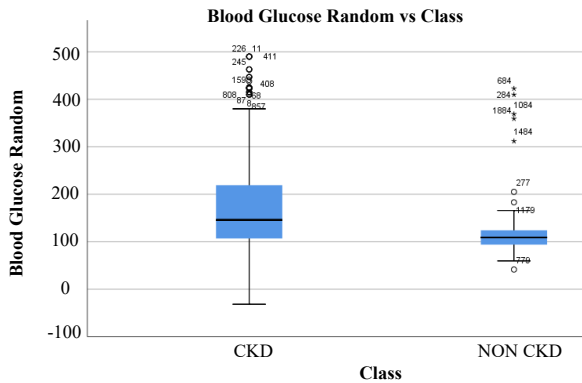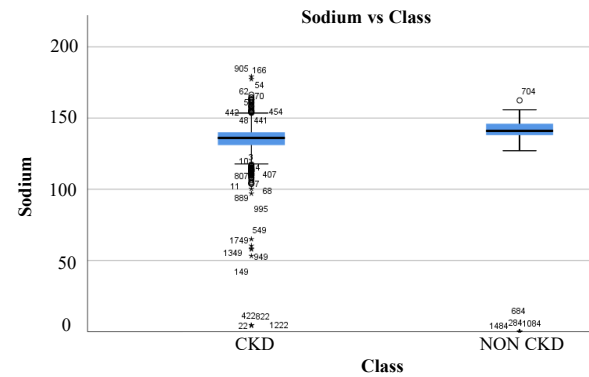

Fig.23. Blood glucose random (N=2000).
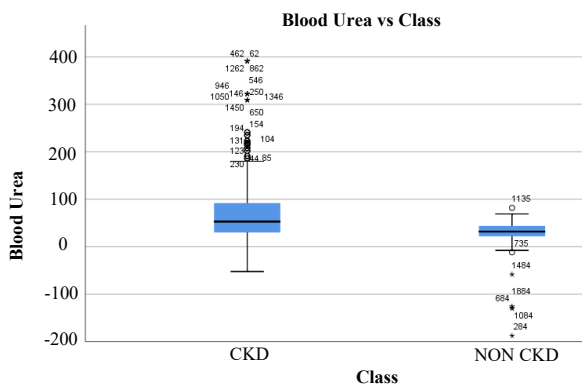

Fig.24. Blood urea (N=2000).

Fig.25. Serum creatinine (N=2000).
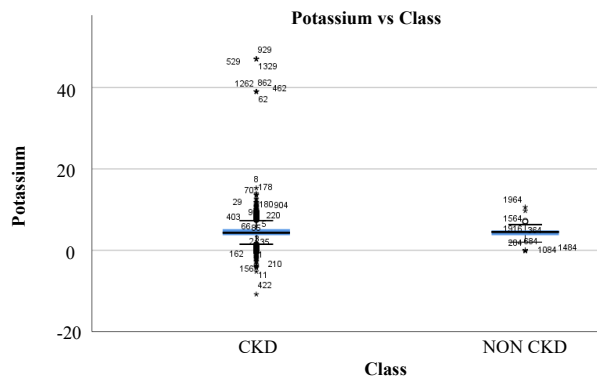

Fig.26. Sodium (N=2000).
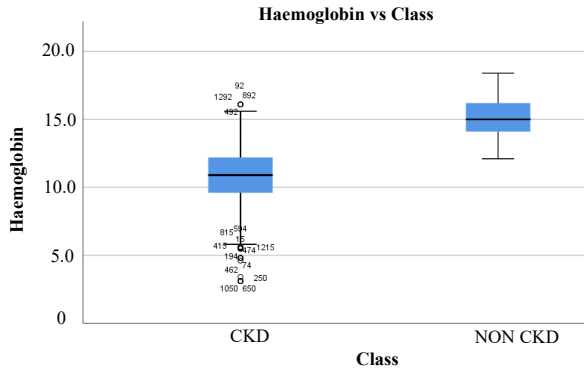

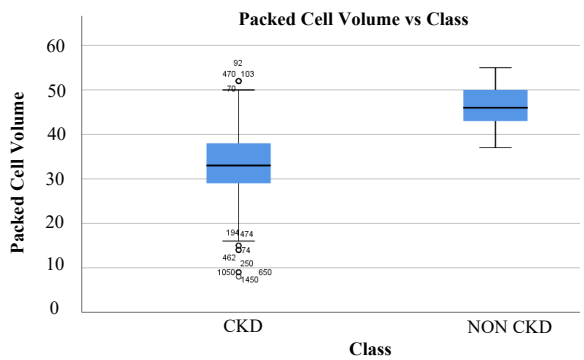Fig.27. Potassium (N=2000).

Fig.28. Haemoglobin (N=2000).
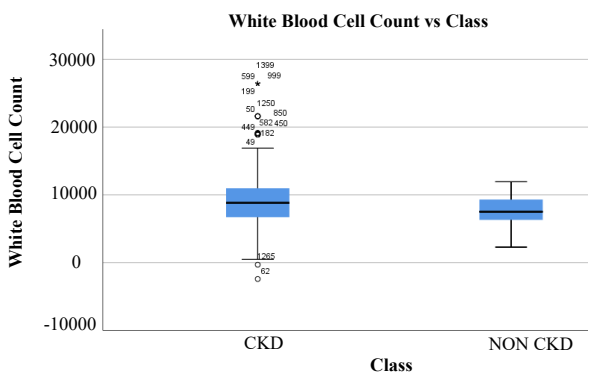

Fig.29. Packed cell volume (N=2000).
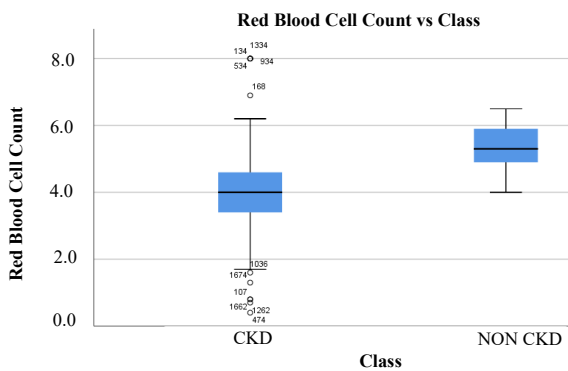

Fig.30. White blood cell count (N=2000).

Fig.31. Red blood cell count (N=2000).

C. Artificial Neural Network

Initially, the experiment is conducted to determine the optimum number of hidden nodes for both urine-based attributes and blood-based attributes. The lower- and upper limits for urine-based attributes was set at four and nine hidden nodes. The average training accuracy and MSE is shown in Fig. 32 and Fig. 33. Based on the results obtained, the optimum number of hidden nodes for the ANN model is eight.
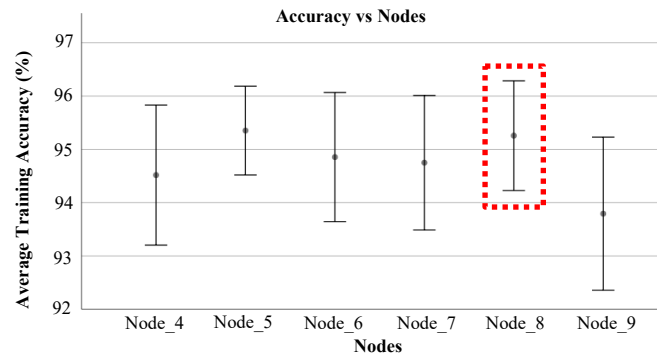

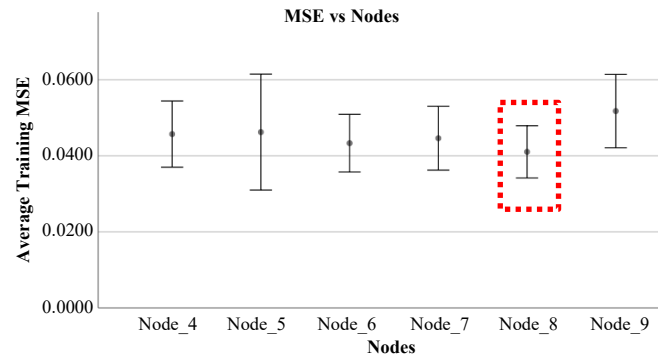Fig.32. Average training accuracy for urine-based attributes.

Fig.33. Average training MSE for urine-based attributes.

The network is then trained, validated, and tested with the optimum structure, yielding the results shown in Table II.

TABLE II
CLASSIFICATION PERFORMANCE FOR URINE-BASED ATTRIBUTES

Parameters		CKD	Healthy	*Acc*
Training	*Se*	94.6%	97.7%	95.8%
	Pp	98.6%	91.6%	
Validation	*Se*	95.9%	98.1%	96.7%
	Pp	98.9%	92.7%	
Testing	*Se*	95.6%	96.6%	96.0%
	Pp	97.7%	93.5%	

TABLE III
CLASSIFICATION PERFORMANCE FOR BLOOD-BASED ATTRIBUTES

Parameters		CKD	Healthy	*Acc*
Training	*Se*	97.7%	98.9%	98.1%
	Pp	99.3%	96.3%	
Validation	*Se*	95.9%	99.0%	97.0%
	Pp	99.5%	92.8%	
Testing	*Se*	97.2%	99.2%	96.0%
	Pp	99.4%	95.9%	

Meanwhile, the lower- and upper limits for urine-based attributes was set at seven and seventeen hidden nodes. The average training accuracy and MSE is shown in Fig. 34 and Fig. 35. Based on the results obtained, the optimum number of hidden nodes for the ANN model is seventeen
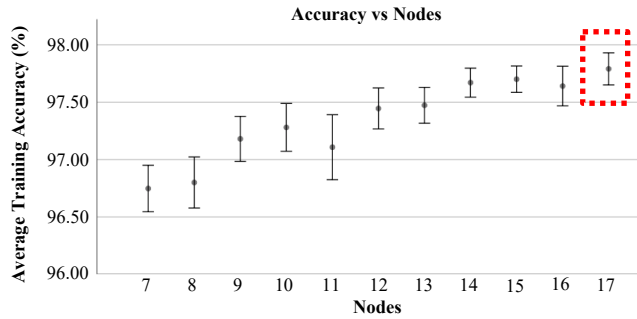

Fig.34. Average training accuracy for blood-based attributes.
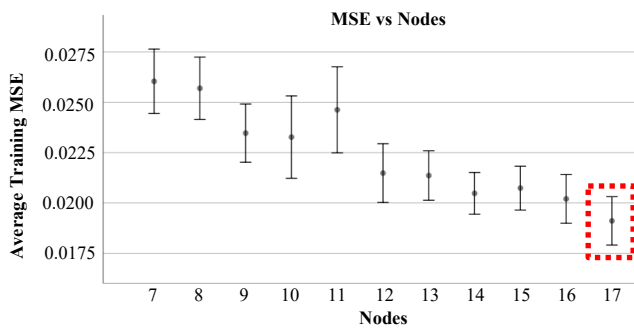

Fig.35. Average training MSE for blood-based attributes.

The network is then trained, validated, and tested with the optimum structure, yielding the results shown in Table III.

D. Comparative Analysis

Both the urine-based and blood-based attributes has each attained excellent overall accuracy of 96.0% and 98.0%. The ANN prediction model developed using blood-based attributes has shown superior performance, although by a small margin of 2.0%. However, this study would recommend the model developed based on urine-based attributes based on the following reasons: 1) The model only requires five input and eight hidden nodes. These reduces the computational complexity of the model and is relatively more suitable to be implemented in intelligent diagnostic systems. 2) The use of urine-based attributes presents a more convenient and non-invasive method for both patients and medical practitioners, therefore rendering data collection and analysis more effective than acquisition of blood-based attributes.

IV. CONCLUSION

The study initially sets out with the following objectives. First, the study aims to characterize the attributes based on urine and blood tests. Second, the study aims to implement multiple imputation method to solve missing data points within the dataset and balance out data distribution between those with CKD and healthy controls. Third, the study aims to develop two ANN prediction models based on urine-based and blood-based attributes. A comparative analysis is then performed to reach a conclusive recommendation.

Generally, of those objectives have been achieved. The study was able to characterize the distribution of attributes between samples with CKD and healthy controls. The missing datapoints are dealt with using multiple imputation technique. The method was also able to generate synthetic samples, enhancing the dataset to N=2000. Prior to developing the prediction models, the ANN structure is optimized for urine-based and blood-based attributes. Two ANN prediction model

is then trained, validated, and tested; yielding excellent performance. Despite attaining inferior accuracy, the model developed based on urine-based attributes is recommended for implementation in intelligent diagnostic systems due to the efficient computational requirements. Furthermore, the sample acquisition protocol is more convenient for both patients and medical practitioners.

V. ACKNOWLEDGEMENT

VI. REFERENCES

[1] S. D. Arasu and R.Thirumalaiselvi, *et al.*, "Review of Chronic Kidney Disease based on Data Mining Techniques," *International Journal of Applied Engineering Research*, vol. 12, pp. 13498–13505, 2017.

[2] International Society of Nephrology (2019,Jan.), "Kidney Health for Everyone Everywhere," *ISN – Global Operations Center*. [Online]. Available: https://www.worldkidneyday.org/2019-campaign/2019-wkd-theme/.

[3] Vassalotti JA, Centor R, Turner BJ, Greer RC, Choi M, Sequist TD, and the U.S. Kidney Disease Outcomes Quality Initiative, *et al.*, "A Practical Approach to Detection and Management of Chronic Kidney Disease for the Primary Care Clinician," *Am. J. Med.*, vol. 1, pp. 1–23, 2015.

[4] C. Y. Ng, Z. S. Lee, and K. S. Goh, "Cross-sectional study on knowledge of chronic kidney disease among medical outpatient clinic patients," *Medical Journal of Malaysia*, vol. 71, no. 3, pp. 99–104, 2016.

[5] Urology Kidney Foundation (2020,Jul.), "What is Kidney (Renal) Failure?," *Urology Kidney Foundation*. [Online]. Available:https://www.urologyhealth.org/urologic-conditions/kidney-(renal)-failure.

[6] D. K. Ng, Y. Xu, J. Hogan,J. M. Saland, L. A. Greenbaum, S. L. Furth, B. A. Warady, C. S. Wong, *et al.*, "Timing of patient-reported renal replacement therapy planning discussions by disease severity among children and young adults with chronic kidney disease," *Pediatric Nephrology*, 2020.

[7] C. Kovesdy, R. Langham, M. Rosenberg, and V. Jha, "A single number for advocacy and communication-worldwide more than 850 million individuals have kidney diseases," *Kidney Int.*, vol. 96, no. 5, pp. 1048–1050, 2019, doi: 10.1016/j.kint.2019.07.012.

[8] Bernama, "Kidney failure strikes women more than men, study finds," *The Star*, 2019. https://www.thestar.com.my/news/nation/2019/06/03/kidney-failure-strikes-women-more-than-men/ (accessed Jul. 11, 2019).

[9] H. S. Wong and B. L. Goh, "24th Report of the Malaysian Dialysis & Transplant 2016," 2018.

[10] C. J. Chen, T. W. Pai, H. H. Hsu, C. H. Lee, K. S. Chen, and Y. C. Chen, "Prediction of chronic kidney disease stages by renal ultrasound imaging," *Enterprise Information Systems*, vol. 14, no. 2, pp. 178–195, 2020.

[11] National Institute of Diabetes and Digestive and Kidney Diseases, "Albuminuria: Albumin in the Urine," 2020.

[12] Rugheed Ghadban, "Specific Gravity," *MedScape*, 2014.

[13] Osvaldo Padilla, "Normal Laboratory Values : Blood , Plasma , and Serum," *MSD Manual Professional Version*, 2020.

[14] A. J. Aljaaf, D. Al-jumeily, H. M. Haglan, M. Alloghani, T. Baker, and A. J. Hussain, *et al.*, "Early Prediction of Chronic Kidney Disease Using Machine Learning Supported by Predictive Analytics," *IEEE Congress on Evolutionary Computation*, pp. 1–9, 2018.

[15] N. Myakina, A. Vigel, I. Lots, *et al.*, "Time in Range Analysis and Glucose Variability in Type 1 Diabetic Patients with Different Stages of Chronic Kidney Disease Natalia," *International Multiconference Bioinformatics of Genome Regulation and Structure*, vol. 11, pp. 106–109, 2018.

[16] J. Park, S. Bae, S. Seo, S. Park, J. Bang, J. H. Han, W. Lee, J. S. Lee, *et al.*, "Measurement of Glomerular Filtration Rate using Quantitative SPECT/CT and Deep-learning-based Kidney Segmentation," *Scientific Reports*, vol. 9, no. 1, pp. 1–8, 2019.

[17] H. Polat, H. D. Mehr, and A. Cetin, *et al.*, "Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods," *Journal of Medical System*, vol. 4, no. 41, pp. 1–11, 2017.

[18] S. Sharma, V. Sharma, A. Sharma, *et al.*, "Performance Based Evaluation of Various Machine Learning Classification Techniques for Chronic Kidney Disease Diagnosis," *IEEE 17th International Conference on Bioinformatics and Bioengineering*, vol. 17, no. 17, pp. 291–296, 2017.

[19] S. D. Vijayarani, "Kidney Disease Prediction using SVM and ANN Algorithms," *International Journal of Computing and Business Research*, vol. 6, no. 2, pp. 12, 2015.

[20] T. M. Rahman, S. Siddiqua, S. E. Rabby, N. Hasan, and M. H. Imam, *et al.*, "Early detection of kidney disease using ECG signals through machine learning based modelling," *International Conference on Robotics,Electrical and Signal Processing Techniques*, vol. 1, pp. 319–323, 2019.

[21] M. S. Başarslan and F. Kayaalp, *et al.*, "Performance analysis of fuzzy rough set-based and correlation-based attribute selection methods on detection of Chronic Kidney Disease with various classifiers," *Scientific Meeting on Electrical-Electronics and Biomedical Engineering and Computer Science*, pp.1-5, 2019.

[22] T. Mahboob, A. Ijaz, A. Shahzad, and M. Kalsoom, "Handling Missing Values in Chronic Kidney Disease Datasets Using KNN , K-Means and K-Medoids Algorithms," in *2018 12th International Conference on Open Source Systems and Technologies (ICOSST)*, 2018, pp. 76–81.

[23] D. B. Rubin, "An Overview of Multiple Imputation," *Multiple Imputation*, vol. 1, no. 1, pp. 79–84, 1987.

[24] James Lani, "Multiple Imputation for Missing Data," *Statistics Solutions*, 2019. .

[25] P. Yildirim, "Chronic Kidney Disease Prediction on Imbalanced Data by Multilayer Perceptron," in *IEEE 41st Annual Computer Software and Applications Conference*, 2017, pp. 193–198, doi: 10.1109/COMPSAC.2017.84.

[26] J. A. C. Sterne *et al.*, "Research & reporting: Multiple imputation for missing data in epidemiological and clinical research : potential and pitfalls," vol. 339, no. July, pp. 157–160, 2009, doi: 10.1136/bmj.b2393.

[27] G. S and D. V. M., "Survey on Prediction of Kidney Disease by using Data Mining Techniques," *IJARCCE*, vol. 118, no. 8, pp. 198–201, 2017, doi: 10.17148/ijarcce.2017.6138.

[28] S. Chen and D. Tan, "A SA-ANN-Based Modeling Method for Human Cognition Mechanism and the PSACO Cognition Algorithm," *Complexity*, vol. 2018, p. 21, 2018, doi: 10.1155/2018/6264124.

[29] F. Ibrahim, M. N. Taib, W. A. B. W. Abas, C. C. Guan, and S. Sulaiman, "A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN)," *Comput. Methods Programs Biomed.*, vol. 79, no. 3, pp. 273–281, 2005, doi: 10.1016/j.cmpb.2005.04.002.

[30] Miray Akgül; Özlen Erkal Sönmez; Tuncay Özcan, "Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making," *Intell. Fuzzy Tech. Big Data Anal. Decis. Mak.*, vol. 1029, pp. 313–321, 2020, doi: 10.1007/978-3-030-23756-1.

[31] C. Dua, Dheeru and Graff, "UCI Machine Learning Repository," *UCI Machine Learning Repository*, 2017. .

[32] Y. Yuan, "Multiple imputation for missing data: Concepts and new

development," *SAS Inst. Inc, Rockville, MD*, pp. 1–13, 2010, [Online]. Available: http://www.math.montana.edu/~jimrc/classes/stat506/notes/multipl eimputation-SAS.pdf.

[33] M. Negnevitsky, *Artificial Intelligence*, Second Edi. Harlow: Pearson Education, 2005.

[34] O. Contribution, "On the Approximate Realization of Continuous Mappings by Neural Networks," *Neural Networks*, vol. 2, pp. 183–192, 1989, doi: 10.1016/0893-6080(89)90003-8.

[35] F. S. Panchal and M. Panchal, "Review on Methods of Selecting Number of Hidden Nodes in Artificial Neural Network," *Int. J. Comput. Sci. Mob. Comput.*, vol. 3, no. 11, pp. 455–464, 2014, [Online]. Available: www.ijcsmc.com.

[36] W. H. Qazi, S. B. S. Osman, and M. B. Memon, "Improvement of Correlation using Artificial Neural Networks Technique for the Prediction of Resistivity against Soil Strength Properties," *Indian J. Sci. Technol.*, vol. 9, no. October, p. 10, 2016, doi: 10.17485/ijst/2016/v9i37/94351.

[37] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2019, doi: 10.1016/j.ipm.2009.03.002.

[38] S. Xu and L. Chen, "A Novel Approach for Determining the Optimal Number of Hidden Layer Neurons for FNN ' s and Its Application in Data Mining," in *5th International Conference on Information Technology and Applications (ICITA 2008)*, 2008, no. July 2008, pp. 683–689.