LITERATURE SURVEY

Related Work In recent years, few studies have been done on the classification or diagnosis of chronic kidney disease. In 2013, T. Di Noia et al. [1], presented a software tool that used the artificial neural network ANN to classify patient status, which is likely to lead to end-stage renal disease (ESRD). The classifiers were trained using the data collected at the University of Bari over a 38-year period, and the evaluation was done based on precision, recall, and F-measure. The presented software tool has been made available as both an Android mobile application and online web application.

Later in 2016, K. A. Padmanaban and G. Parthiban [2] aimed in their work to detect chronic kidney disease for diabetic patients using machine learning methods. In their research, they used 600 clinical records collected from a leading Chennai-based diabetes research center. The authors have tested the dataset using the decision tree and Naïve Bayes methods for classification using the WEKA tool. They concluded that the decision tree algorithm outweighs the Naïve Bayes with an accuracy of 91%.

A. Salekin and J. Stankovic [3] evaluated three classifiers: random forest, K-nearest neighbors, and neural network to detect the CKD. They used a dataset with 400 patients form UCI with 24 attributes. By using the wrapper method, a feature reduction analysis has been performed to find the attributes that detect this disease with high accuracy. By considering: albumin, specific gravity, diabetes mellitus, hemoglobin, and hypertension as features, they can predict the CKD with .98 F1 and 0.11 RMSE.

In the study carried out by W. Gunarathne, K. Perera, and K. Kahandawaarachchi [4], Microsoft Azore has been used to predict the patient status of CKD. By considering 14 attributes out of 25, they compared four different algorithms, which were Multiclass Decision Forest, Multiclass Decision Jungle, Multiclass Decision Regression, and Multiclass Neural Network. After comparison, they found that Multiclass Decision Forest performed the best with 99.1% accuracy.

In the study carried out by W. Gunarathne, K. Perera, and K. Kahandawaarachchi [5], Microsoft Azore has been used to predict the patient status of CKD. By considering 14 attributes out of 25, they compared four different algorithms, which were Multiclass Decision Forest, Multiclass Decision Jungle, Multiclass Decision Regression, and Multiclass Neural Network. After comparison, they found that Multiclass Decision Forest performed the best with 99.1% accuracy.

P. Yildirim [6] studied the effect of sampling algorithms in predicting chronic kidney disease. The experiment was done by comparing the effect of the three sampling algorithms: Resample, SMOTE, and Spread Sup Sample on the prediction by multilayer perceptron classification algorithm. The study showed that sampling algorithms could improve the classification algorithm performance, and the resample method has a higher accuracy among the sampling algorithms. On the other hand, Spread Sub Sample was better in terms of execution time.

A. J. Aljaaf et al. [7] examined in their study the ability of four machine learning (ML) models for early prediction of CKD, which were: support vector machine (SVM), classification and regression tree (CART), logistic regression (LR), and multilayer perceptron neural network (MLP). By using the CKD dataset from UCI and seven features out of 24, they compared the performance of these ML models. The results showed that the MLP model had the highest AUC and sensitivity. It was also noticeable that logistic regression almost had the same performance as MLP but with the advantage of the simplicity of the LR algorithm. Therefore, in our study, we can use the LR algorithm as a start or a benchmark and then use more complex algorithms.

Lastly in 2019, J. Xiao et al. [8] in their study established and compared nine ML models, including LR, Elastic Net, ridge regression lasso regression SVM, RF, XGBoost, knearest neighbor and neural network to predict the progression of CKD. They used available clinical features from 551 CKD follow-up patients. They conclude that linear models have the overall predictive power with an average AUC above 0.87 and precision above 0.8 and 0.8, respectively.

N.A. Almansour et al. intend to aid in the prevention of CKD by using machine learning techniques to detect CKD at an early stage, focusing on applying multiple machine learning classification algorithms to a dataset of 400 patients and 24 CKD-related variables. As classification algorithms, they employ Artificial Neural Network and support vector machine, and any missing values in the dataset are replaced by the relevant characteristics. The best collected parameters and characteristics were used to construct the final model of

the two suggested strategies. Experiment results showed that ANN performed better than SVM, with accuracy of 99.75 percent and 97.75 percent, respectively.