LITERATURE SURVEY

Jena and Kamila proposed a method for predicting chronic kidney disease using SVM, Naïve Bayes, Multilayer Perceptron, J48 (a type of decision tree), Conjunctive Rule and Decision Table. From the experimental result, the Multilayer Perception algorithm gives a better classification accuracy of 99.7%. The performance of these algorithms was measured by classification accuracy, the time taken to build the model, test the model, and the mean absolute error .

Manish Kumar has made some research on other authors' studies and reported that SVM performed best compared to other classifiers. The authors used six machine learning algorithms namely: Random Forest (RF), Naïve Bayes, Sequential Minimum Optimization (SMO), Radial Basis Function (RBF Classifier), Multilayer Perceptron Classifier (MLPC) and Simple Logistic (SLG). The author compared the performance of the six classifiers with SVM. The results showed that RF achieved a performance of 100% classification accuracy.

In another study conducted by Celik et al, the dataset was partitioned into training and testing data twice, in two different sizes. Test-1 used approximately 66% of the data to train the model, and the remaining 34% was used for testing. In Test-2 on the other hand, only 10% of the data was used for training and 90% of data was used for testing. Then, SMO and J48 classification algorithms were applied. Both Test- 1 and Test-2 showed that using J48, a decision tree algorithm, resulted in a higher accuracy rate of 100%, better than the results obtained from SMO. Furthermore, the accuracy obtained from Test-1 was better compared to Test-2 concluding that training data with more instances results in higher classification accuracy.

Using the same dataset, Salekin and Stankovic have developed an automated machine learning solution to detect CKD and explore 24 parameters related to kidney disease. The dataset used for evaluation suffers from noisy and missing data. They evaluate solutions with three different classifiers: K-NN, RF, and neural nets. To reduce over-fitting as well as to identify the most important predictive attributes for chronic kidney disease, they have performed feature reduction using two methods: the wrapper method and LASSO regularization. Also, through cost analysis considering all 24 attributes they identify a cost-effective highly accurate detection classifier using only 5 attributes: specific gravity, albumin, diabetes mellitus, hypertension, and haemoglobin. By using this approach they achieved a detection accuracy of 0.993 using F-measure.

Ramya and Radha used a different dataset to diagnose CKD. The dataset is obtained from the medical reports of patients from different laboratories in Coimbatore. It consists of 1000 instances and 15 attributes related to kidney disease. The instances are classified as low, mild, moderate, normal, and sever based on the value of the attribute EGFR (estimated glomerular filtration rate). Also, this work differs from the previous ones in the tool used to determine the accuracy of the machine learning algorithms. Most of the previous works used Weka, while R tool was used in this study for comparing different algorithms. Four machine learning algorithms were used in this work: Back Propagation Neural Network, Radial Basis Function Neural Network, and RF. From the experimental result, the Radial Basis Function has better performance with accuracy of 85.3% .