Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 4, 7264 – 7273

An Ensemble based technique (EARLNP) for recommendations of Lupus activity in Humans from kidney datasets

[1]Dr. A. MALARVIZHI

[1]Assistant Professor, P.G. and Research Department of Computer Science, H.H. The Rajah's College (A), Pudukkottai, malarvizhiselvam@gmail.com

Abstract

Global healthcare systems are perfect examples of digital technology evolutions. These systems analyze large amounts of patient data for deriving insights and assisting clinicians in prediction of diseases. Automated health recommenders are becoming popular in healthcare where intelligent systems have significant importance in their capability to aid decision making processes about illnesses. The recommender system using patient's lifestyles or physical health records forecast health issues including the presence of LNs (Lupus Nephritis), a severe form of SLE (Systemic Lupus Erythematosus) which caused by immune complex deposits in human kidneys. In their acute phases, they cause substantial injuries and nephron losses and when not treated adequately, kidneys turn into chronic or irreversibly damaged.Though therapies for handling LNs have improved, it is imperative for systems which predict consensual outcomes. Hence, the main objective of this paper is to propose a non-invasive health recommender technique called EARLNAP, an ensemble technique for implementations in kidneys malfunctioning recommender systems. The results of this system are favorable in terms of its performances and performance metrics.

Keywords: LNs (Lupus Nephritis), SLE (Systemic Lupus Erythematosus), EARLNAP.

INTRODUCTION

LNs, a serious form of SLEs, affect 35 to 60% of the people globally irrespective of ethnicities, sex, and ages [1]. Importantly, LNs flare primarily due to nephron losses contributing to worsening of renal functions. It is very essential to determine therapy/non-therapy based treatmenttrials which show success. The basic aim of managing patients with LN can be categorized into shortterm therapies for avoiding flares and longterm therapies for safe guarding kidneys. SLE categorizations [2] incorporate cellular microscopic inspections of urine sediments and are useful in diagnosing affected kidneys. LNs under a microscope are depicted in Figure 1.

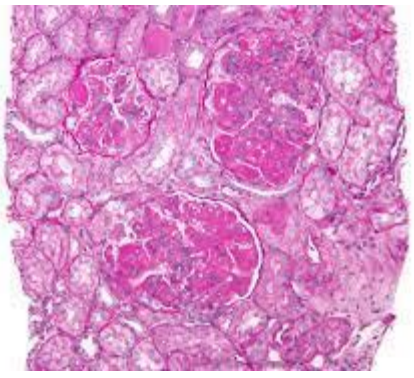

Fig. 1 – *Microscopic image of LN*

Urinalysis of abnormalities in urea can assist in diagnose of sick kidneys. These can be observed from urine sediment findings/examinations which significantly vary, though clinical urinalysis are being questioned. Large quantities of urine dilute pellet's cellular

constituents impacting cellular count assay accuracies. Urine's higher concentrations in mornings are void when compared with randomized collections through the day making urine's collection time a limitation. As a result, discrepancies of subsequent assessments are needed for quantifying urine's analytical components. Hence, urinalysis was recently removed for categorizing SLEs [3] due to differences in analysis by laboratories and absence of homogenized clinical data for studies. The main element of finding LNs is the presence of protein in urea which fails in most occasions of LN analyses. Moreover, current treatments of LNs have been ineffective in producing remissions or avoiding future flares. The patients also do not respond to therapies adequately as only 1/3rd of LN patients achieve full remissions after six months of treatments.In spite of successful therapies, about 1/5th patients have renal failures in the first decade of their illness [4]. Figure 2 shows LNs or Kidney inflation's influence on the human body.

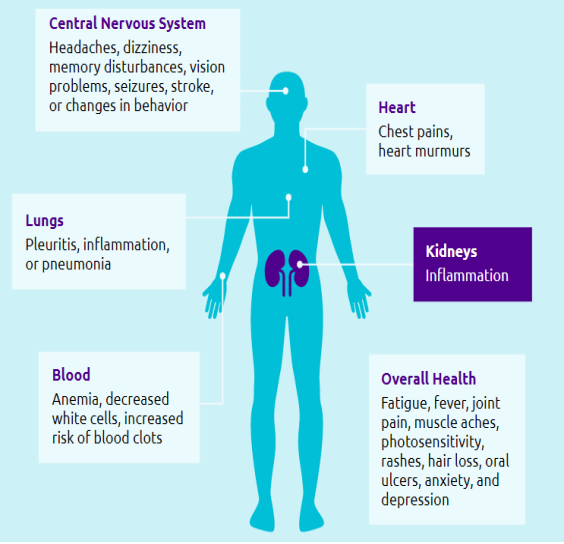

Fig. 2 – *Humans Body parts affected by Kidney Inflammations*

Anticipations and long term renal prognosis is critical for detecting LNs in its early stages. This has resulted in recent initiatives focusing on understanding molecular processes that happen in individuals with renal problems and non-renal SLEs. The study in [5] used high throughputs of kidney biopsies from cohorts of patients with LNs to evaluate intra renal molecular patterns associated with disease severities. The study found that patients with histological signs of tubular injuries had worst functioning of the kidneys. Another study on of intra-renal transcript expressions [6] used multiple kidney biopy outcomes to link differentially expressed genes in individuals matched against favourable and poor clinical responses. Many studies have aimed to uncover and link early clinical characteristics of poor renal functions from laboratory tests, and genetic pathways for improving patient surveillances and therapies. The implementation of recommender systems with clear instructions and uniformity of patient data can help in overcoming hurdles faced by heath care services where major entities namely patients and diagnostics play critical roles in recommender systems. Patients can select treatment preferences and ratings in patient based recommender systems [7]. The earliest step in recommender system predictions is collection of data where the study in [8] used SLE dataset for its recommendations. There has been an increase in the application of analytics using DMTs (Data Mining Techniques) in healthcare due to the rapid development of these technologies. Data analytics in many industries which have voluminous data is becoming increasingly popular where its connectivity has resulted in making recommender systems extremely popular. This work examines the working principles of building and creating a data analytics using DMTs in the area of healthcare. The proposed HRS examines health concerns and builds an intelligent prediction using association gained by the HRSs (Health Recommender Systems) to warn LN patients. The main objective of this paper is to propose a non-invasive health recommender technique called EARLNAP (Ensemble Approach for Recommendation of Lupus Nephritus Patients, an ensemble technique for implementations HRSs (Health Recommender Systems) for identifying affected kidneys and hence predict LN patients.This introductory section is followed by a description of HRSs which includes studies related to HRSs. Section three details on the methodology of this study followed by implementation results and a discussion. The

paper is concluded in section five with opportunities for future work.

HRSs:Currently, internet is the main sources of information in all domains including buying and selling. When people are doubtful or confused about a subject, internet is the primary source they refer where recommender systems provide a platform to learn about a topic of interest like features of an item, patient/treatment preferences. Recommender systems filter vast information found on the internet to suit the searcher's queries and provide them with vital information [9]. Data analytics is not an unfamiliar concept, but has been evolving characterized by continuous variations. Various approaches have been used in IRs (Information Retrievals) as the internet has data that is unstructured or unprocessed and need processing before their applications. The finest example of how this type of analytics can be used by healthcare was shown in [10]. Data sent by healthcare institutions, hospitals, and clinics need to show consistency and trustworthiness [11]. Based on the patient's test results or other information, HRSs can predict whether or not a person is affected by a disease. data. HRSs can be implemented based on patient profilestowards specific elements. Millions of patients suffer from diseases with multiple test results. Even clinicians refer to previous histories of patients for further treatments and in such a scenario; it becomes difficult for pathologists to conclude quickly on the line of treatment. Using HRSs can increase their decision convergences. HRSs not only can assist in decisions for avoid hazards or failures, but also assist in monitoring patients and provide needed therapies based on monitored vital signs and thus provide appropriateness of HRSs functions [12]. HRSs can be seen of as having three separate phases: data gathering, learning, and predictions/recommendations [13]. During the Information Collection Phase, crucial information on patients is gathered, including patient's personal characteristics, habits, and resources. Recommender engines cannot function without well defined patient data as in the Learning Phase it exploits patient's attributes acquired from the previous phase. Preferable outcomes for patients are indicated in the Prediction/Recommender Phase. HRSs can forecast based on models or observed patient actions.

Proposed EARLNAP Methodology:

This work proposes new HRSs to enhance the health care implementations of predicting / recommending LN affected patients. EARLNAP is based on predictive analytics for providing patients with suitable medications/precautions. The therapies for LNs have improved and evolved over the years, although consensus outcome metrics are still needed. While it is well recognized that early diagnosis and treatments have better renal outcomes, there are very few early indicators for renal function degradations. The proposed methodology consists ofdifferent phases pre-processing, extraction of features and Predictions and recommendations. Figure 3 depicts a DMT based Recommender System.

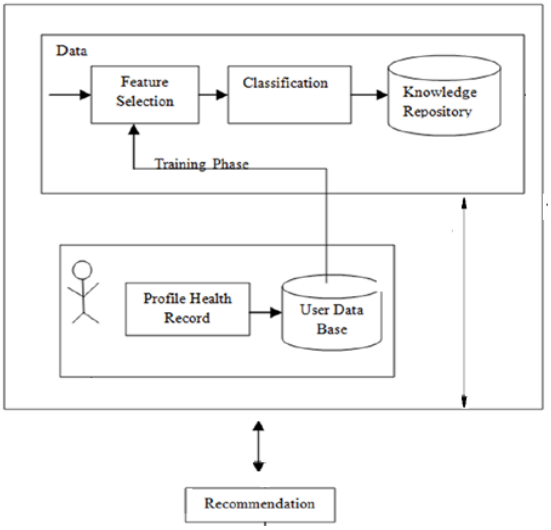

Fig. 3 - *Recommendations based on DMTs.*

The proposed recommender's collection phase starts from healthcare kidney diseases dataset (SLE). In the second phase, the learning phase, EARLNAP trains on acquired data encompassing data cleaning, feature extractions, and hyper tuning before training. The outputs of the training phase are then used for recommendations which can generalized by clinicians specifically to diseases like LN or

SLE or CKDs (Chronic Kidney Diseases). Thus the proposed EARLNAP framework uses Pre-processing, Feature Extractions/Dimensionality reductions, Parameter Tuning, Training/Predictions and Recommendations in its framework. These recommendations can help in early predictions for valuable clinical guidelines and high-quality healthcare treatments.

EARLNAP Pre-processing: Clinical datasets in real world scenarios are prone to discrepancies, resulting in low-quality outcomes. As a result, the initial stage in DMTs is to investigate datasets, learn its properties, and prepare it for modeling by cleaning or nullifying discrepancies where Missing data is a fairly common problem of datasets with patient records. Their attributesmay also contain missing values [14] where single imputations like mean andmedianare not enough for medical datasets [15]. Hence, EARLNAP uses multiple imputations for replacingmissing values in the values of features.

EARLNAP Feature Extractions:A Feature's importance is computed where higher the value, more important the feature. This work uses correlation between features to select and extract them. Feature subsets that are substantially connected with the class but not with other aspects of the class are regarded as excellent outputs. The following is an operational description of the aforementioned hypothesis where feature evaluations can be mathematically formulated as Equation (1):

$$r_{fc} = \frac{k\overline{r_{fc}}}{\sqrt{(k + k(k-1)\overline{r_{ff}})}} \qquad \text{……..(1)}$$

Where rfc is the correlation between the summed features and the class variable; k is the number of features rfc and above line is the average of the correlation between the features and class variable; and rff and above line is the average inter-correlations between features [16]. Figure 4 depicts the correlations between features/predictors.

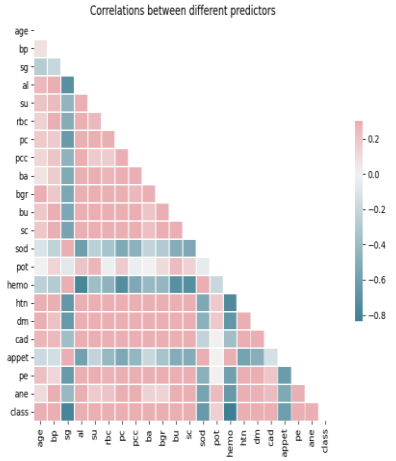

Fig. 4 – *Correlation between Predictors*

EARLNAP Parameter Tuning:The selected features are then tuned for optimality using grid search as it is common that a major portion of the time is spent on collecting, cleaning, and organizing data. Until model's performances are deemed to be satisfactory, they are trained, tested, validated, and then re-trained. MLTs (Machine Learning Techniques) rely heavily on data for their decisions. Hyperparameters are used to improve their performances as finding optimum hyperparameters aids in the development of highperforming models where searches include Random, Grid and Manual models Searches. In additions, Bayesian Optimizations are also used to select suitable hyperparameters for a model.Figure 5 depicts Grid Search across Two Parameters.

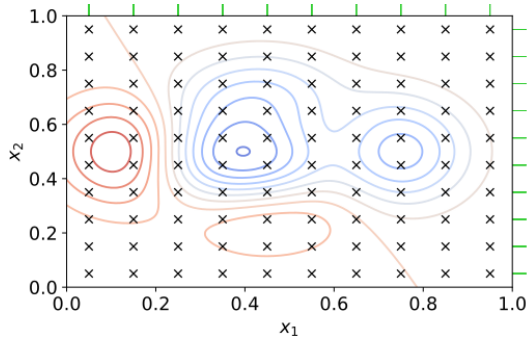

Fig. 5 - *Grid Search*

EARLNAP Training/Predictions:Ensemble learning improve MLT's outcomes by joining several model outcomes. They produce better predictive performanceswhen compared to single models. They work on the principle of

using classifiers and then allowing them to vote. They are used to overcome Statistical and representational issues while analyzing data. The goal of ensembles isn't to get very accurate base models, but rather to get base models that do mistakes. If ensembles are employed for classifications, high accuracies can be achieved even if base classifier's accuracy is low and they misclassify different training samples. Ensembles can be constructed using different techniques including majority Votes, bagging and RFs (Random Forests), RandomInjections, FeatureSelections and Error-Correcting Output Coding. This work uses RFs for classifications. The steps for RFs implementations are listed below:

- From the original data set, many subsets are produced by replacing observations.
- A subset of features is chosen at random, and the feature with the best split is utilised to iteratively divide the node.
- The tree has reached its full size.
- Repeat steps 1-3 until a forecast based on the aggregate of predictions from n trees is obtained.

RFs areextensions of bagging. Each classifier in the ensemble uses DTs (Decision Trees) generated using random selection of attributes. To create the final forecast, the mode of the classes for classification or the mean prediction for regression, the predictions from all trees are combined. Ensembles are RFs that employ a set of outcomes to arrive at a final judgement. They construct many individual DTs while training and for each tree a node's importance is computed using Gini Importance (binary tree with 2 nodes) based on Equation (1):

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$$

……(1)

Where, nij is node j's importance, w jis the weighted number of samples reaching node j, Cj is the impurity value of node j, leftjis child node from left split on node j and rightjis child node from right split on node j. The importance of features on DTs are computed using Equation (2)

$$fi_i = \frac{\sum_{j:node\ j\ splits\ on\ feature\ i} ni_j}{\sum_{k \in all\ nodes} ni_k}$$

……..(2)

Where, fii is the importance of feature i and nij is the importance of node j. These values are normalized in the interval [0,1] by dividing by the sum of all feature importance values as depicted in Equation (3).

$$normfi_i = \frac{fi_i}{\sum_{j \in all\ features} fi_j}$$

…….(3)

The final feature importance, at the Random Forest level, is it's average over all the trees. The sum of the feature's importance value on each trees is calculated and divided by the total number of trees as depicted in Equation (4)

$$RFfi_i = \frac{\sum_{j \in all\ trees} normfi_{ij}}{T}$$

……(4)

Where, RFfii stands for the importance of feature i calculated from the model's all trees, normfiijis the normalized feature importance for i in tree j and T is the total number of trees. For each DT, EARLNAP computes feature's importance by summing the gain, scaled by the number of samples passing through the node as shown in Equation (5)

$$fi_i = \sum_{j:nodes\ j\ splits\ on\ feature\ i} s_j C_j$$

……..(5)

Where, fii is the importance of feature i, sj is the number of samples reaching node j and Cj is the impurity value of node j.

Results and Discussion:

This section displays stage wise experimental results of the proposed scheme EARLNAP was implemented and executed using Python 3.7.5 on an AMD athelon processor with 4 GB memory. The experiments were coded for CKD Data Set found in the UCI repository where its attributes were used to predict the presence of LN. The dataset has twenty four health related attributes of four hundred with complete information on 158 patients while the

remaining needed pre-processing. Figure 6 depicts a snapshot of the Dataset.

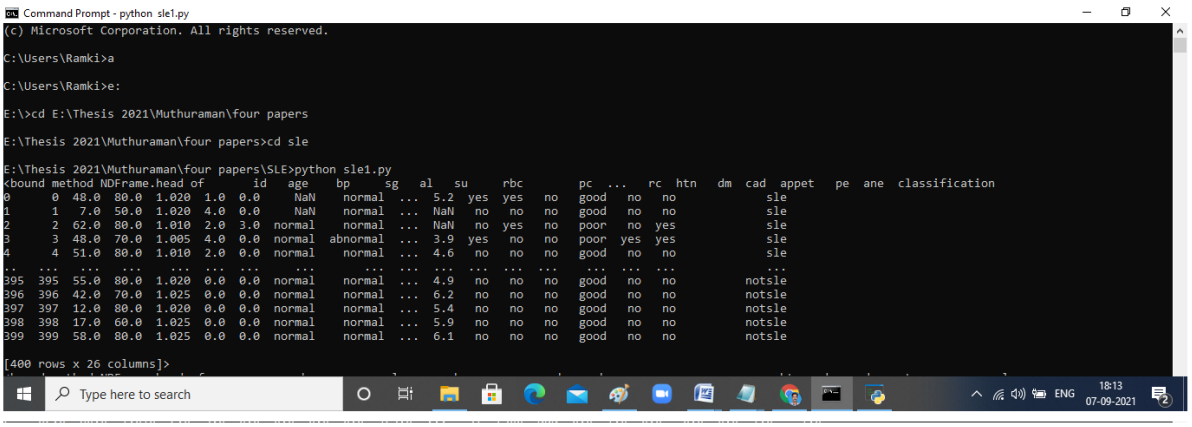

Fig. 6 – *CKD Dataset*

EARLNAP Pre-processing:Te dataset had missing values in many attributes with at least one missing value in each attribute. Table 1 lists missing values of attributes. This work also used multiple imputations for replacing missing values in the dataset. The replaced values were between 2 to 9.

Table 1 – *Missing Values List in the Dataset*

Attribute Name	Missing	
	Number	*Percent*
Red blood cells	159	38.00%
Red blood cell count	137	32.80%
White blood cell count	111	26.50%
Potassium	94	22.50%
Sodium	92	22.00%
Packed cell volume	74	17.80%
Pus cell	68	16.30%
Hemoglobin	54	13.00%
Sugar	51	12.30%
Specific gravity	49	11.80%
Albumin	48	11.50%
Blood glucose	46	11.00%
Blood urea	19	4.80%
Serum creatinine	18	4.50%
Blood pressure	12	3.00%
Age	9	2.30%
Bacteria	4	1.00%
Pus cell clumps	4	1.00%
Coronary artery disease	2	0.50%
Diabetes mellitus	2	0.50%
Hypertension	2	0.50%
Anemia	1	0.30%
Pedal Edema	1	0.30%
Appetite	1	0.30%

Data was cleaned further by converting text to numeric values. Red blood and pus cells were coded as one for 'abnormal' and 0 for 'normal'. The values of pus cell clumps and bacteria were set to 1 for presence and 0 for absent. Values which were above normal for features were set to 1 while good or normal feature values were set to 0. Features with improper values were labeledas non-numeric. Figure 7 depicts EARLNAP Pre-processing Output.

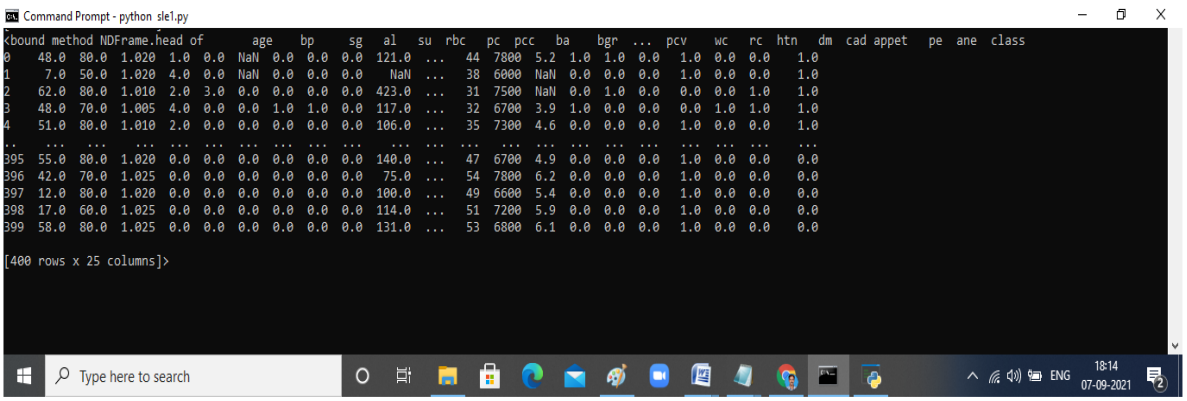

Fig. 7 - *EARLNAP Pre-processing Output*

EARLNAP Feature Extractions: In EARLNAP, if the feature's correlation with the class is larger than the highest correlation between the feature and any of the previously picked features, the feature is considered a subset. Continuous features were changed to categorical features using Equation (1) to accommodate all kinds of features using a discretization method. This proposed scheme used Exhaustive Search for its feature selections. Figure 8 depicts computed importance of features in EARLNAP.

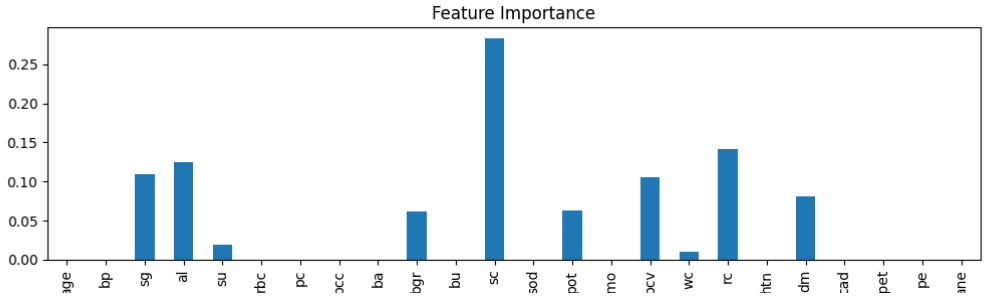

Fig. 8 - *computed importance of features in EARLNAP.*

EARLNAP Parameter Tuning:EARLNAP uses Grid Search which locates best hyperparameters for improving performances. Grid Searches compute performances for each combination of hyperparameters and their values, and then selects optimum value for hyperparameters. Based on the hyperparameters count processing is executed. EARLNAP used a max depth of two for tree nodes and eight estimators to decrease the number of trees. The final optimal parameters chosen for predictions were The parameters chosen were specific gravity, albumin, sugar, blood glucose random, serum creatinine, potassium, packed cell volume, white blood cell count, red blood cell count. Figure 9 depicts the output of tuned parameters that can optimize predictions.

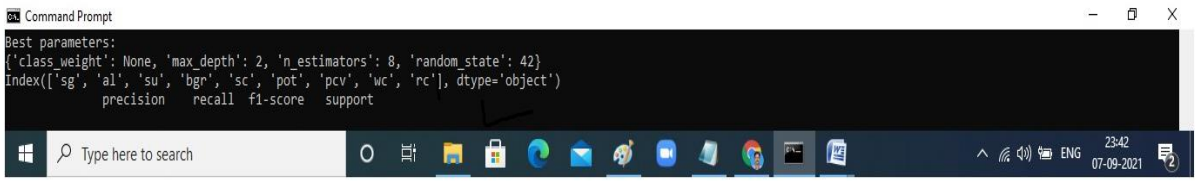

Fig. 9 – *EARLNAP Hyperparameter Output*

Evaluation of EARLNAP: The success of the proposed EARLNAP was evaluated based on criteria borrowed from IRs [18] which are Precision (Measure of relevant retrieved instances), Recall (Correctly recommended items in the collection of recommended items),

F-Measure (Measure of a test's accuracy and the weighted harmonic mean of precision and recall values) and ROC-Curve (Curve for comparing diagnostic tests and a plot of TPRs (True positive rates) against FPRs (False Positive Rates) . The proposed scheme achieved an accuracy of 0.888430. Figure 10 depicts the ROC of the proposed EARLNAP while Table 2 shows the evaluation metric values before and after parameter optimizations.

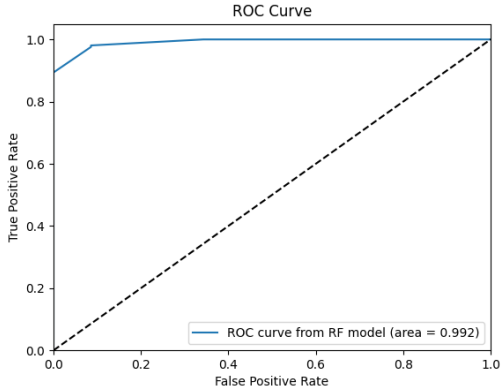

Fig. 10 – *EARLNAP ROC*

Table 2 –*EARLNAP evaluation scores before and after parameter optimizations*

Evaluation Metric Values Before Optimizations				
	precision	recall	f1-score	Support
0.0 – 1.0	1.00 – 1.00	1.00 – 1.00	1.00-1.00	39-14
Avg / Total	1.00	1.00	1.00	53
Evaluation Metric Values Before Optimizations				
	precision	recall	f1-score	Support
0.0 – 1.0	0.56 – 1.00	1.00 – 0.87	0.72-0.93	35-207
Avg / Total	0.94	0.89	0.90	242

EARLNAP Recommendations: The guidelines are based on 2012 EULAR/ERA-EDTA recommended update for LN's care [17]. The nature of LNs necessitates multi-disciplinary approaches by rheumatologists and nephrologists for care of patients in shared decisionmaking. Kidney involvements in LNs was established by a nephropathologist's histologic examination, and that disease treatment and periodic monitoring of these patients be done in specialist clinics..

- Investigation of Suspected LNs:In individuals with SLEs who have glomerular hematuria and/or cellular casts, proteinuria >0.5 g per 24 hours or a urine protein-to-creatine ratio (UPCR) of less than 500 mg/g, or unexplained declines in glomerular filtration rate, the task force recommends kidney biopsy (GFR). Antiphospholipid antibodies (aPL) should be tested in all patients with SLE, especially those with probable kidney involvement, because the results might be diagnostic and predictive.

- Recommendations for Treatment of LNs in Adultss: Long-term kidney function stabilities, better proteinuria levels by 3 months, and a 50% reduction in proteinuria (partial clinical responses) by 6 months should be the therapy goals. Proteinuria of 0.5 to 0.7 g per 24 hours should be achieved by 12 months (complete clinical response). The task force advised that individuals with nephrotic-range proteinuria at baseline may need a further 6 to 12 months to achieve full clinical improvement. Because these patients' proteinuria recovers more slowly, switching medicines is not essential as long as proteinuria improves.

- Recommendations for Monitoring/Prognosis of LNs: Patients with lupus nephritis should be evaluated in specialty centres on a regular basis; each visit should include a urinalysis to assess proteinuria (quantified by spot UPCR or 24-hour urine collection), the presence of glomerular hematuria, and the presence of cellular casts, which are indicators of an impending kidney flare. At these visits, serum C3/C4 and anti-dsDNA titers should be checked. A second kidney biopsy may be undertaken in individuals who have not responded to immunosuppressive medication or to distinguish between continued histologic activity and irreparable damage. Rebiopsy of the protocol may be necessary to evaluate histologic class transition, activity changes, or the requirement for ongoing therapy.

Conclusion And Future Work

The capacity to detect LNs utilising HRSs with the fewest amount of tests or characteristics was investigated in this study. The relationship between variables was investigated in order to minimise the number of characteristics and eliminate redundancy. Specific gravity, albumin, sugar, blood glucose random, serum creatinine, potassium, packed cell volume, white blood cell count, and red blood cell count were determined to have the most influence on predicting LN patients using a filter feature selection technique. Using 10-fold cross-validation, the classifiers were trained, tested, and validated. Highperformance was achieved by the proposed EARLNAP (88.8%) after optimization of parameters. Therefore, it can be conclude that LN can be detected with these features. The study has introduced medical recommendations based on predictionsd which can always be added to the implemented systems to warn clinicians and patients alike when affected with LNs. This implementation can also try deep convolution neural networks in the diagnosis of LN in the future.

Reference

[1] Anders H--J, Saxena R, Zhao M--H, et al. Lupus nephritis. Nat Rev Dis Primers 2020;6:7.

[2] Petri M, Orbai A--M, Alarcón GS, et al. Derivation and validation of the systemic lupus international collaborating clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64:2677–86.

[3] Aringer M, Costenbader K, Daikh D, et al. European League against Rheumatism/American College of rheumatology classification criteria for systemic lupus erythematosus. Ann Rheum Dis 2019;2019:1151–9.

[4] Zhang L, Lee G, Liu X, et al. Long--Term outcomes of end--stage kidney disease for patients with lupus nephritis. Kidney Int 2016;89:1337–45.

[5] Pamfil C, Makowska Z, De Groof A, et al. Intrarenal activation of adaptive immune effectors is associated with tubular damage and impaired renal function in lupus nephritis. Ann Rheum Dis 2018;77:1782–9

[6] Parikh SV, Malvar A, Song H, et al. Molecular imaging of the kidney in lupus nephritis to characterize response to treatment. Transl Res 2017;182:1–13.

[7] Martínez-Pérez, B.; De La Torre-Díez, I.; López-Coronado, M. Privacy and security in mobile health APPs: A review and recommendations. J. Med. Syst. 2015, 39, 181.

[8] Mu, R.; Zeng, X.; Han, L. A Survey of Recommender Systems Based on Deep Learning. IEEE Access 2018, 6, 69009–69022.

[9] Isinkaye, F.O.; Folajimi, Y.O.; Ojokoh, B.A. Recommender systems: Principles, methods and evaluation. Egypt. Inform. J. 2015, 16, 261–273, ISSN 1110-8665

[10] Li, S.; Kawale, J.; Fu, Y. Deep collaborative filtering via marginalized denoising auto-encoder. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 18–23 October 2015; pp. 811–820.

[11] Priyadarshini, R.; Barik, R.K.; Panigrahi, C.; Dubey, H.; Mishra, B.K. An investigation into the e cacy of deep learning tools for big data analysis in health care. Int. J. Grid High Perform. Comput. 2018, 10, 1–13.

[12] Wang, Y.; Hajli, N. Exploring the path to big data analytics success in healthcare. J. Bus. Res. 2017, 70, 287–299.

[13] Gope, J.; Jain, S.K. A survey on solving cold start problem in recommender systems. In Proceedings of the 2017 International Conference on Computing, Communication, and Automation (ICCCA), Greater Noida, India, 5–6 May 2017; pp. 133–138.

[14] A. J. M. Kaky, "Intelligent Systems Approach for Classification and Management of Patients with Headache," Liverpool John Moores University, 2017.

[15] W. Gunarathne, K. Perera and K. Kahandawaarachchi, "Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD)," in Bioinformatics and Bioengineering (BIBE), 2017 IEEE 17th International Conference On, 2017,

[16] Polat K, Sahan S, Kodaz H, Gunes S. LNCS 3611. Berlin; Heidelberg: Springer; 2005. A new classification method for breast cancer diagnosis: Feature selection artificial immune recognition system (FS-AIRS). Advances in Neural Computation; pp. 830–8.

[17] . Fanouriakis A, Kostopoulou M, Cheema K, et al. 2019 update of the joint European League Against Rheumatism and European Renal Association–European Dialysis and Transplant Association (EULAR/ERA–EDTA) recommendations for the management of lupus nephritis [published online March 27, 2020]. Ann Rheum Dis. doi:10.1136/annrheumdis-2020-216924.

[18] Canny, J. Collaborative filtering with privacy via factor analysis. In Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland, 11–15 August 2002; pp. 238–245.