Preemptive Diagnosis of Chronic Kidney Disease Using Machine Learning Techniques

Reem A. Alassaf [1], Khawla A. Alsulaim [2] ,Noura Y. Alroomi [3], Nouf S. Alsharif [4] , Mishael F. Aljubeir [5], Sunday O. Olatunji [6], Alaa Y. Alahmadi [7] , Mohammed Imran[8], Rahma A. Alzahrani[9], Nora S. Alturayeif [10]
College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, Kingdom of Saudi Arabia
[1] Reemalassaf17@gmail.com ,[2] khawla.95@windowslive.com ,[3] roominy@gmail.com,[4] n13s@hotmail.com , [5]mishael.aljubair@hotmail.com ,, [6]osunday@iau.edu.sa ,[7] ayalahmadi@iau.edu.sa , [8] mbahmed@iau.edu.sa , [9]razahrani@iau.edu.sa [10]nsalturayeif@iau.edu.sa,

***Abstract*— Chronic Kidney Disease (CKD) is a major public health concern with rising prevalence. In Saudi Arabia, approximately 2 Billion Riyals are solely allocated for renal replacement therapy which is required for patients with advanced stages of CKD. Therefore, this study aims to decrease the number of patients and the expenses needed for treatment by preemptively diagnosing chronic kidney disease accurately using data mining and machine learning techniques. Data have been collected from a region that has never been explored before in literature. This study uses Saudi data retrieved from King Fahd University Hospital(KFUH) in Khobar to carry out the experiment. Experimental Results show that ANN, SVM, Naïve Bayes achieved a testing accuracy of 98.0% while k-NN has achieved an accuracy of 93.9%.**

***Keywords*— Artificial Neural Network, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes, Chronic Kidney Disease**

I. Introduction

Chronic Kidney Disease (CKD) is a major public health concern with rising prevalence. Kidney disease is when the kidneys are damaged and could not filter the blood properly. This damage could cause the wastes to build up in the body [1]. There are five stages of CKD, the most serious one is stage 5 because, at this stage, the kidneys are unable to do most of their functions. It is difficult to pinpoint the CKD stage of each patient especially at the early-stages [2]. It also causes a high possibility of death within a short period of time, a patient must be hospitalized and appropriately cured. The most common causes of kidney disease are diabetes and high blood pressure [1].

In Saudi Arabia, the number of people who need kidney replacement due to CKD is growing day by day. CKD is ranked the fourth leading cause of death in Saudi Arabia with the mortality rate of 5.44% [3]. Moreover, the escalating prevalence of CKD poses challenges on professionals and the health care system [2]. Approximately two billion Saudi Riyals are allocated for renal replacement therapy which is required for patients with advanced stages of CKD [4]. Hence, the need for a preemptive diagnostic predictive solution that will facilitate early and easier accurate diagnosis of CKD cannot be overemphasized.

Machine learning is a field of computer science that gives the ability of machines to learn without being explicitly programmed. By using computational methods, machine learning has shown success in providing solutions for early-stage diagnosis in a variety of medical domains. These methods are used to find hidden patterns from data and mine these data for decision- makers [5].

In this study, the work in the subsequent sections is organized in the following manner. The second section covers a review of related work in literature while the third section includes the description of the four proposed machine learning techniques: ANN, SVM, NB, and k-NN. After that, the fourth section presents the empirical study that includes the dataset description and experimental setup. Then, the fifth section demonstrates and discusses the results of this study. And finally, the last section includes the conclusion and recommendation for future work based on this study.

II. Related Work

The following are some of the earlier works in the field of using machine-learning algorithms to diagnose CKD. They used the same dataset from UCI Machine Learning Repository with different machine learning algorithms. The dataset has been collected from Apollo hospital (Tamilandu) [6]. It has 25 attributes and 400 total instances, out of which 250 instances are classified as CKD and 150 instances as non-CKD [6].

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, the time taken to test the model, and the mean absolute error [7].

In another related work, 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 (RBFClassifier), 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,

978-1-5386-6673-9/18/$31.00 ©2018 IEEE

while SMO and RBF achieved a lower classification accuracy [8].

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 [9].

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 hemoglobin. By using this approach they achieved a detection accuracy of 0.993 using F-measure [10].

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 [11] 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% [12].

III. Description of Proposed Techniques

A. Artificial Neural Network

Artificial neural networks (ANNs) are a branch of machine learning that are statistical-based learning algorithms which were designed to simulate the properties of the biological neural networks [13]. One of the most widely used neural network is Multilayer Perceptron (MLP). MLP consists mainly of three types of layers made up of artificial neurons and connected by weighted links as shown in figure 1 [14]. Depending on the weights and a specific value called the activation value, some neurons will be activated to some value and others will not. The activation pattern of a layer affects that of the next layer [15].

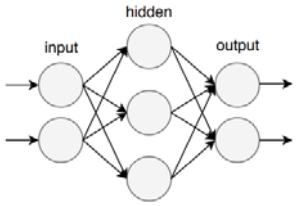

Fig. 1. Basic Structure of MLP

B. Support Vector Machine

The original Support Vector Machine (SVM) algorithm was first introduced by Russian mathematicians Vladimir Vapnik and Alexey Chervonenkis based on decades of research in computational learning theory [16]. SVM is a supervised machine learning algorithm used in both classification and regression problems [17]. Focusing on classification problems, SVM became a popular choice among many researchers, since it often outperforms other classification algorithms [17]. Moreover, SVM performs well even when the number of samples is low [17]. SVM has been used in many fields such as optical character recognition [18], email spam detection [19] , and medical diagnosis [20]. For further details readers should refer to [16], [21], [22].

C. Naïve Bayes

Naïve Bayes is a classification algorithm that is based on the Bayesian probability theorem. The classifier operates under the fundamental Naive Bayes assumptions which are independent and equal of feature contribution to the outcome, which means that feature presence or absence is unrelated to the presence or absence of any other feature. Naïve Bayes classifies an instance by calculating the probability of its belonging to each target class, where the instance will be considered as belonging to the target class with the highest probability using the following rule (1):

$$P(A|B) = \frac{P(B|A).\ P(A)}{P(B)} \quad (1)$$

In the rule above, A is the target class and B is the features vector describing an instance, P(A|B) is the probability of the instance B belonging to the target class, A. P(A) is the prior probability of the target class in the training set. P(B) is the probability of the features vector given the target class in the training set [23].

D. K-Nearest Neighbors

K-NN is a simple classifier which increasingly became a popular choice in practice. The idea of K-NN was first introduced by Fix and Hodges in 1951 as a non-parametric method for pattern classification [24]. It is considered a lazy classifier as it does not require building a training model, instead, for each query, the k nearest neighbors are located from the training database regardless of their class label, then the majority vote predicts the class label of the instance [25].

IV. Empirical Study

A. Data Description

After going through literature, it was found that Saudi data has never been explored to diagnose chronic kidney disease. Upon this finding, the authors decided to investigate this region by requesting data from KFUH. Patient's laboratory

visits were considered as records in the dataset instead of a single record to represent each patient due to the limited digitalized patients' information in the hospital's database. The obtained dataset contained 244 records with a binary target class variable. 118 records were classified as CKD Patients and the reaming 126 records were classified as non-CKD Patients. Initially, the dataset contained 491 features. After going through several data cleaning processes to downsize the dataset dimensionality such as removing irrelevant features, for example: surgeon name, resident on call, etc., and removing features with missing values more than 50%. The remaining features were 57 features. All the remaining features along with their correlation to the target variable are shown in table I. The dataset contains 48 numeric features and 9 nominal features.

TABLE I. THE CORRELATION BETWEEN EACH ATTRIBUTE AND THE TARGET VARIABLE

Attribute	Correlation with Class
Creat (mg/dL)	0.75748
BUN (mg/dL)	0.67843
RBC (mg/dL)	0.38104
Hct (%)	0.36307
Cl (mEq/L)	0.32744
Albumin (g/dL)	0.312377
Urobilinogen	0.307190
Lymph (K/uL)	0.282077
CO2 (mEq/L)	0.269131
T Bili (mg/dL)	0.258459
Na (mEq/L)	0.253291
WBC (k/ul)	0.244342
Plt (k/ul)	0.241742
Hct	0.240818
Hgb (g/dL)	0.231426
Mono (K/uL)	0.227483
T Protein (g/dL)	0.223650
Bilirubin	0.210866
Phosphorus (mg/dL)	0.190466
D Bili (mg/dL)	0.185668
PT (sec)	0.172599
Baso (K/uL)	0.169541
Protein	0.164905
INR	0.163532
Bacteria	0.159646
Hgb	0.158284
Random Glucose(mg/dL)	0.147174
Ca (mg/dL)	0.143652
Neu (K/uL)	0.143451
Nitrite	0.132075
Leukocytes	0.130377
RDW (%)	0.121898
Urine Color	0.115892
Lymph	0.112596
Specific Gravity	0.101573
MPV	0.085770
MCHC	0.084675
Glucose (%)	0.082059
Seg	0.077571
PH	0.074320
MPV (fL)	0.067394
GGTP (U/L)	0.046533
RDW	0.043670
Anion Gap	0.043646
Mg (mg/dL)	0.036164
MCH	0.033628
Eos (K/uL)	0.022601
MCV (fL)	0.018273
MCV	0.014062
MCHC (g/dL)	0.013750
Aptt (sec)	0.012598
Alk Phos (U/L)	0.011477
K (mEq/L)	0.011061
Clarity	0.010502
MCH (pg)	0.001734
Mono	0.000081
Normal Control aPTT (sec)	0.000000

B. Experimental Setup

The experiment was carried out using Weka [26] and Python machine learning library: scikit-learn [27]. Weka was used for data preprocessing and features ranking by calculating the correlation between every feature and the class label. While Python was used for parameter tuning, feature selection, and for the training and testing process of the four classifiers: ANN, SVM, K-NN, and Naïve Bayes. Both 10-fold cross validation and direct partitioning schemes were used to partition the dataset. The 10-fold cross validation was utilized in both parameter tuning, and features selection. While direct partitioning was used in the classification process with 80:20 training-testing split. Each one of the four classification algorithms was trained with the optimized values of the classifiers' parameters that resulted from the optimization strategy.

C. Performance Measures

The study uses four performance measures to evaluate the classification, namely: testing accuracy, precision, recall, and

f-measure. The testing accuracy is the main performance measure which calculates the ratio of correctly classified instances over all instances of the testing set (2). The precision calculates the ratio of True Positive (TP) over all predicted positive outcomes (3). The recall calculates the ratio of TP over all actual positive outcomes (4). F-measure assesses the performance of each class (5). All the performance measures described are derived from the four possible outcomes:

True Positive (TP): The number of instances correctly classified with CKD.
False Positive (FP): The number of instances incorrectly classified with CKD.
Ture Negative (TN): the number of instances correctly classified as non-CKD.
False Negative (FN): The number of instances incorrectly classified as non-CKD.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$f-measure = \frac{2\times precision \times recall}{precision \times recall} \quad (5)$$

V. Results and Discussion

A. Optimization Strategy

To achieve models that can optimally solve a classification problem, the parameters of the learning algorithms have to be tuned. For that purpose, exhaustive grid search with cross validation was used. Grid search is used to set the search space with the specified parameters and their range of possible values. Then it generates all the possible combinations of the parameters' values to search the parameters simultaneously [28]. For each combination, 10-fold cross validation was used for validation and the average of the models' accuracy is calculated.

For ANN, a single hidden layer with 100 neurons, identity activation function, and Limited-memory BFGS (L-BFGS) optimization algorithm gave the best score among the available activation functions and optimizers. Identity activation function is a linear function denoted as $f(x)=x$ and is also known as step function [29]. The output of the identity function is either 1 or 0 which determines if a neuron is activated or not. Therefore, it is suitable for binary classification [30]. L-BFGS optimization algorithm is a Newton's method that updates the weights based on the loss function which can result in a more efficient update [31]. Also, L-BFGS converges faster and performs better for small datasets [32].

The same procedure was applied to SVM and the searched parameters are the kernels each with its parameters, the penalty parameter C, and epsilon. Similarly, the parameters of K-NN were tuned using the same procedure. These parameters include the value of K and the weight which indicates whether the neighbor points are weighted equally or depending on their distance from the query point [33]. Table II shows the optimal parameters chosen for each classifier.

TABLE II. Optimal Parameters For Each Classifier

Classifier	Parameter	Value
ANN	Number of Layers	1
	Number of Neurons	100
	Activation Function	Linear
	Optimization Algorithm	L-BFGS
SVM	Kernel Type	Linear
	C	0.001
	Epsilon	0.001
K-NN	K	5
	Weight	Distance

B. Feature Selection

The feature selection method applied in this study uses correlation coefficient and recursive feature elimination. Correlation coefficient is used to rank the attributes from the highest to lowest correlation with the target variable as shown in table I. Then, the recursive feature elimination procedure produces subsets by recursively eliminating the bottom half of the ranked attributes until a single feature remained. Each subset runs all the four classifiers with 10-fold cross validation to find the subset achieving the highest performance.

Table III shows that the highest average accuracy was achieved using the top 2 features: Creatinine, and Blood Urea Nitrogen (BUN).

TABLE III. Result of different feature subsets

Number of Attributes	Accuracy of Classifier (%)				Average Accuracy
	ANN	*SVM*	*NB*	*k-NN*	
57 (All)	0.926	0.938	0.737	0.754	0.83875
29	0.934	0.931	0.762	0.837	0.866
15	0.656	0.939	0.774	0.881	0.8125
8	0.946	0.942	0.623	0.902	0.85325
4	0.943	0.946	0.959	0.910	0.9395
2	0.951	0.943	0.956	0.939	0.94725
1	0.943	0.931	0.935	0.939	0.937

C. Classification

The performance of ANN, SVM, K-NN, and NB on the optimal parameters and features with direct partitioning is shown in table IV. It is important to note that the parameters for each classifier were investigated again to ensure there are still the optimal for our classification, in the case of SVM we find that changing C=100 gave a better result than our initial assumption.

TABLE IV. Classification performance on optimal feaures

Classifier	Accuracy	Precision	Recall	f-measure
ANN	0.980	0.964	1	0.98167
SVM	0.980	0.964	1	0.98167
NB	0.980	0.964	1	0.98167
k-NN	0.939	0.929	0.963	0.9457

According to table IV, ANN, SVM, and Naïve Bayes show the highest performance in terms of all measures (accuracy, precision, recall, and f-measure). The confusion matrices below exhibit a further look into the actual and predicted classes by ANN, SVM, Naïve Bayes, and k-NN in tables V, VI, VII, VIII respectively.

The most significant metric to examine in the confusion matrices below is the false negative (FN) rate. Since the failure to detect a CKD patient may cause them the agony of further complications [34]. The lowest FN rate was found using ANN, SVM and Naïve Bayes followed by K-NN.

TABLE V. ANN CONFUSION MATRIX

		Predicted	
		CKD	**NonCKD**
Actual	**CKD**	27 (TP)	0 (FN)
	NonCKD	1 (FP)	21 (TN)

TABLE VI. SVM CONFUSION MATRIX

		Predicted	
		CKD	**NonCKD**
Actual	**CKD**	27 (TP)	0 (FN)
	NonCKD	1 (FP)	21 (TN)

a.

TABLE VII. NAÏVE BAYES CONFUSION MATRIX

		Predicted	
		CKD	**NonCKD**
Actual	**CKD**	27 (TP)	0 (FN)
	NonCKD	1 (FP)	21 (TN)

b.

TABLE VIII. K-NN CONFUSION MATRIX

		Predicted	
		CKD	**NonCKD**
Actual	**CKD**	26 (TP)	1 (FN)
	NonCKD	2 (FP)	20 (TN)

VI. CONCLUSION AND RECOMMENDATION

In this study, Saudi medical records were investigated for the first time in the process of diagnosing CKD using machine learning techniques. Authors used correlation coefficient and recursive feature elimination for feature selection. Then, four classification algorithms were explored, namely: ANN, SVM, Naïve Bayes, and k-NN. The performance of each of these classifiers was examined by the classification accuracy, precision, recall, and f-measure achieved by the classifier. ANN, SVM, and NB all achieved an accuracy of 98% while k-NN achieved an accuracy of 93.9%. Further research can be done to exceed the classification accuracy currently achieved, by using different classifiers or feature selection methods.

REFERENCES

[1] I. Veerappan and G. Abraham, "Chronic kidney disease: current status , challenges and management in India.," *Nephrology*, no. Table 1, pp. 593–7, 2008.

[2] H. Almutary, A. Bonner, and C. Douglas, "Chronic kidney disease in Saudi Arabia: a nursing perspective," *Middle East J Nurs.*, vol. 7, no. 6, pp. 17–26, 2013.

[3] IHME, "Saudi Arabia | IHME." [Online]. Available: http://www.healthdata.org/saudi-arabia. [Accessed: 31-Oct-2017].

[4] A. a Al-Sayyari and F. a Shaheen, "End stage chronic kidney disease in Saudi Arabia. A rapidly changing scene.," *Saudi Med. J.*, vol. 32, pp. 339–346, 2011.

[5] G. D. Magoulas and A. Prentza, "Machine Learning in Medical Applications," Springer, Berlin, Heidelberg, 2001, pp. 300–307.

[6] L.Jerlin Rubini, "UCI Machine Learning Repository: Chronic_Kidney_Disease Data Set," 2015. .

[7] L. Jena and N. Ku Kamila, "Distributed Data Mining Classification Algorithms for Prediction of Chronic-Kidney-Disease," *Int. J. Emerg. Res. Manag. &Technology*, vol. 9359, no. 4, pp. 2278–9359, 2015.

[8] M. Kumar, "Prediction of Chronic Kidney Disease Using Random Forest Machine Learning Algorithm," *Int. J. Comput. Sci. Mob. Comput.*, vol. 5, no. 2, pp. 24–33, 2016.

[9] E. Celik, M. Atalay, and A. Kondiloglu, "The Diagnosis and Estimate of Chronic Kidney Disease Using the Machine Learning Methods," no. 4, pp. 27–31, 2016.

[10] A. Salekin and J. Stankovic, "Detection of Chronic Kidney Disease and Selecting Important Predictive Attributes," *Proc. - 2016 IEEE Int. Conf. Healthc. Informatics, ICHI 2016*, pp. 262–270, 2016.

[11] "R: What is R?" [Online]. Available: https://www.r-project.org/about.html. [Accessed: 20-Dec-2017].

[12] S.Ramya and N.Radha, "Diagnosis of Chronic Kidney Disease Using Machine Learning Algorithms," *Res. J. Pharm. Biol. Chem. Sci.*, vol. 7, no. 3, pp. 365–373, 2016.

[13] A. S. Nur, N. H. Mohd Radzi, and A. O. Ibrahim, "Artificial Neural Network Weight Optimization: A Review," *TELKOMNIKA Indones. J. Electr. Eng.*, vol. 12, no. 9, pp. 6897–6902, 2014.

[14] S. O. Olatunji, M. S. Al-Ahmadi, M. Elshafei, and Y. A. Fallatah, "Forecasting the Saudi Arabia stock prices based on artificial neural networks model," *Int. J. Intell. Inf. Syst.*, vol. 2, no. 5, pp. 77–86, 2013.

[15] D. Kriesel, "A Brief Introduction to Neural Networks," *Retrieved August*, p. 244, 2005.

[16] D. Ciobanu, "Mathematical and Quantative Methods Using SVM for Classification 1 Introduction," vol. 8,

no. 13, pp. 207–222.

[17] J. M. Moguerza and A. Muñoz, “Support Vector Machines with Applications,” *Stat. Sci.*, vol. 21, no. 3, pp. 322–336, 2006.

[19] S. O. Olatunji, “Extreme Learning machines and Support Vector Machines models for email spam detection,” *Can. Conf. Electr. Comput. Eng.*, pp. 1–6, 2017.

[20] A. Kampouraki, D. Vassis, P. Belsis, and C. Skourlas, “e-Doctor: A Web based Support Vector Machine for Automatic Medical Diagnosis,” *Procedia - Soc. Behav. Sci.*, vol. 73, pp. 467–474, 2013.

[21] R. Amami, D. Ben Ayed, and N. Ellouze, “An Empirical Comparison of SVM and Some Supervised Learning Algorithms for Vowel recognition,” *Int. J. Intell. Inf. Process.*, vol. 3, no. 1, pp. 63–70, 2012.

[22] Z. Gavrilov, “SVM tutorial,” *Slides*, p. 33, 2003.

[23] E. Krishnaveni, K;Radhamani, “Diagnosis and evaluation of ADHD using Naïve Bayes and J48 classifiers,” in *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on*, 2016, pp. 1809–1814.

[24] T. I. Cannings, “New Approaches to Modern Statistical Classification Problems,” no. August, 2015.

[25] H. Sethi, A. Goraya, and V. Sharma, “Artificial Intelligence based Ensemble Model for Diagnosis of Diabetes,” *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 4, pp. 1540–1549, 2017.

[26] E. Frank, M. A. Hall, and I. H. Witten, “The WEKA Workbench,” *Morgan Kaufmann, Fourth Ed.*, pp. 553–571, 2016.

[27] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt, and G. Varoquaux, “API design for machine learning software: experiences from the scikit-learn project,” pp. 1–15, 2013.

[28] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, “Algorithms for Hyper-Parameter Optimization,” in *Advances in Neural Information Processing Systems (NIPS)*, 2011, pp. 2546–2554.

[29] K. . Vijayarekha, “Activation Functions,” *NPTEL – Electron. Commun. Eng. – Pattern Recognit.*, pp. 1–6.

[30] Lars Hulstaert, “Understanding Activation Functions in Neural Networks,” pp. 1–10, 2017.

[31] A. Karpathy, “CS231n Convolutional Neural Networks for visual recognition.,” 2017. .

[32] P. Moritz, R. Nishihara, and M. I. Jordan, “A Linearly-Convergent Stochastic L-BFGS Algorithm,” vol. 41, 2015.

[33] K. Hechenbichler and K. Schliep, “Weighted k-Nearest-Neighbor Techniques and Ordinal

[18] M. Rajnoha, R. Burget, and M. K. Dutta, “Offline handwritten text recognition using support vector machines,” *2017 4th Int. Conf. Signal Process. Integr. Networks, SPIN 2017*, pp. 132–136, 2017.

Classification,” *Mol. Ecol.*, vol. 399, p. 17, 2004.

[34] M. M. Roche and P. P. Wang, “Factors associated with a diabetes diagnosis and late diabetes diagnosis for males and females,” *J. Clin. Transl. Endocrinol.*, vol. 1, no. 3, pp. 77–84, 2014.