

An Adaptive Neuro-Fuzzy Expert System for Clinical Staging of Patients with Chronic Kidney Disease

Bernard Andoh (✉ bernard.andoh@uenr.edu)

University of Energy and Natural Resources

Adebayo Felix Adekoya

University of Energy and Natural Resources

Benjamin Asubam Weyori

University of Energy and Natural Resources

Elliot Koranteng Tannor

Komfo Anokye Teaching Hospital

Research Article

Keywords: kidney disease, modifiable lifestyle, MDRD method, health records

Posted Date: December 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1144843/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

An Adaptive Neuro-Fuzzy Expert System for Clinical Staging of Patients with Chronic Kidney Disease.

Bernard Andoh^{1*}, Adebayo Felix Adekoya¹, Benjamin Asubam Weyori¹ and Elliot Koranteng Tannor²

¹University of Energy and Natural Resources.
Department of Computer Science and Informatics.

²Komfo Anokye Teaching Hospital
Renal Clinic- Kumasi.

Abstract

Background: Chronic kidney disease (CKD) is a health complication faced by almost every nation, as Ghana is no exception. The Ghana Dialysis Service Foundation (DSF) indicated that an average of 12,000 kidney-failure cases is diagnosed among Ghanaians every year. An adequate diagnosis of suspected CKD will play a critical role in saving many lives. This study widens the predictive factors grouped as; personal lifestyle, laboratory findings, and medical history of a patient. The glomerulus filtration rate addresses the controversy of diagnosing CKD without considering aetiology.

Methods: The study adopted a structured, guided questionnaire to obtain the dataset from health records of 180 patients diagnosed with CKD. The Renal clinic of Komfo Anokye Teaching hospital gave access to the patients' health records following the clearance from the ethical committee in Kwame Nkrumah University of Science and Technology. The health records were categorized into Personal Lifestyle, Laboratory findings and Medical history. Eleven factors that influence the incidence of chronic kidney disease were identified and analyzed in this work. The developed model is based on a hybrid of Artificial Neural Network and Fuzzy logic (ANFIS) techniques.

Results: An experimental result of the proposed predictive model recorded an average Root Mean Square Error of 0.85014 for training and 1.3983 for testing. The result showed a close relationship between the actual and predicted outcomes for the training and testing. The Modification of Diet for Renal Disease (MDRD) model is applied to the dataset to estimate the Glomerulus Filtration Rate to determine the stage of the Chronic Kidney Disease. The developed model (NANFIS) gave a better performance than the Modification of Diet for Renal Disease MODEL results.

Conclusions: Considering the aetiology that widens the predictive factors provides an improved perspective in predicting the stage of the disease. The Modification of Diet for Renal Disease model considers Age, race, sex and creatinine to determine the location of Chronic Kidney Disease. The increase in data would help to improve the performance of the developed model better.

Keywords: kidney disease, modifiable lifestyle, MDRD method, health records.

*Correspondence: bernard.andoh@uenr.edu.gh

¹University of Energy and Natural Resources.
Department of Computer Science and Informatics.

Background

Chronic kidney disease (CKD) is a health complication faced by almost every nation, as Ghana is no exception [1]. There were approximately 35 million deaths in 2005, as stated by the World Health Organization (WHO) [2]. Over two million people worldwide need dialysis or a kidney transplant to stay alive, with these patients forming just about 10% of people who need these treatments to live [3]. The individual's lifestyle remains a crucial determinant for a person diagnosed with chronic kidney disease [4]. Chronic kidney disease with the stages of 1-3 (milder CKD) affects approximately 7% of the world's population, with the highest cases recorded in developing countries. Predicting the clinical stage of the chronic kidney disease as well as the progression of the disease with some level of accuracy requires an elaborate consideration of factors that may influence the incidence of the disease. The stage of CKD determines the severity of the disease. The rate of progression and the definition of the stage is quite critical as it gives several indications that support the required intervention and treatments.

The lifestyle decisions and practices of an individual directly influence health and wellbeing. Practices such as consuming high amounts of salt, alcohol and self-medication do not promote healthy kidney life. The human body is most active during the day, the same as its ability to carry out its metabolic activities. The individual's lifestyle burdens the kidney if the practices that promote quality health are not encouraged. It can be seen as the keeping of unhealthy way of life. Hence, the absence of physical activity or exercise, eating late into the night, having bedtime snacks cause Chronic Kidney Disease, which otherwise will be different when the person maintains a healthy lifestyle

[5](Michishita et al., 2016). Besides lifestyle decisions and practices, many medical conditions affect the health of the kidney, i.e. diabetes, high blood pressure (hypertension), Glomerulonephritis, Polycystic Kidney Disease and kidney stones is a direct product of one's lifestyle practices.

Institutional Background

The data source is the renal clinic department of the Komfo Anokye Teaching Hospital in the Ashanti Region. The hospital's geographic location makes it accessible to all the towns or areas that share boundaries with the region. Referrals are received from all upper areas, including Northern, Upper East and West, Bono and Ahafo, formally Brong Ahafo and some part of the Volta region. [6].

Methods

Data

A retrospective review of patients' health records spanning over ten years (2000-2020) is carried out to obtain the dataset needed to build the predictive model. The risk factors of chronic kidney disease as identified from literature are used to collect the data. These attributes were divided into three main categories; Personal Lifestyle, Laboratory findings, and Medical history.

The laboratory and medical history data are made up of the following attributes; Age, Gender, Family History (per CKD), Alcohol History, and rate of Salt Intake for the Lifestyle, Serum creatinine, Specific Gravity, Albumin and Urea/BUN for Laboratory result and year of diagnosis for Hypertension and Diabetes. The attributes listed above are the medical history data components required to predict chronic kidney disease. Table 1 gives detail on the features or risk factors for the development of the predictive model.

Ageing or **Age** as a risk factor influences the level of the glomerulus filtration rate. The increase in Age causes a reduction in the value of the estimated glomerulus filtration rate (eGFR), which increases the risk of developing CKD. The genetic predisposition that results from specific genetic variations, often inherited from a parent (**Family History**), also increases an individual's susceptibility to developing chronic kidney disease. An example is the rare types of kidney disease inherited from one's parent, i.e., Polycystic kidney disease (PKD) and Fabry disease [7].

Creatinine is a waste product formed by the normal breakdown of muscle cells in the body [8]. Laboratory test of **Serum Creatinine** serves to prompt most people who may not be aware of the danger their kidney poses to them.

Symptoms identified with too little kidney function include loss of appetite, vomiting, itching, weakness, and flu-like symptoms. Swelling in the legs and shortness of breath may occur if water builds up in the body [9].

Salt Intake or the amount of salt consumption also plays a vital role in the proper functioning of various organs in the body. It is a significant electrolyte source for controlling fluids entering and leaving the

body tissues and cells. The human body excretes unwanted fluid by filtering the blood through the kidneys. Extra fluid not needed by the body is sucked out and put into the bladder to be removed as urine. The process continues by the kidneys using osmosis to draw the extra water out of the blood. This process uses a delicate balance of sodium and potassium to pull the water across a wall of cells from the bloodstream into a collecting channel that leads to the bladder [10]. Properly managing the salt intake of patients plays a fundamental role in maximizing the beneficial effect of Angiotensin-converting enzyme (ACE) inhibitors on CKD progression as well [11].

The **Specific Gravity** measurement allows physicians to know the concentration of all particles present in the patient's urine. A high level of Specific Gravity indicates that there is a contamination of glucose. The measurement can also indicate diabetes insipidus, glomerulonephritis, etc.

Albumin, a significant protein present in the blood, help determine how the kidney is functioning. Kidney disease is also associated with reduced urea (**Urea/BUN**) excretion, leading to a consequent rise in blood concentration.

Table 1. Features and respective value ranges and units

Age	18 -79	Years
Gender	0 or 1	Female or Male
Family History	0 - 1	Yes/No
Hypertension	0 - 20	Years
Alcohol History	0 - 2	N / S/ F
Salt Intake	1 - 3	Sm / Avg/ H
Serum Creatinine	57 - 3300	61.9–114.9 mol/L in men and 53.0-97.2 mol/L in women
Specific Gravity	1.005 - 1.030	-
Albumin	10 - 55	g/L
Urea/Blood Urea Nitrogen (BUN)	2.1 - 7.5	mg/dL
Diabetes Milletus	0 - 1	Yes/No
Output Feature		
Chronic Kidney Disease	1 - 5	Stage
N - never, S - sometimes, F - frequent, Sm - small, Avg - Average, H -High		

Chronic Kidney Disease (CKD)

The Glomerular Filtration Rate (GFR) is the most common criteria to measure how healthy one's kidney is performing, estimated using the Modification of Diet in Renal Disease (MDRD) model. MDRD is the most preferred method for estimating the GFR of CKD patients with GFR less than 60ml/min/1.73m². [12] (Matsushita et al. from Chiu et al., 2013). The severity of CKD of a patient is dependent on the stage of the disease (Table 2). A person with a persistent malfunctioning kidney for three months or more qualifies to be labelled as a CKD patient. The evidence relates to some diseases such as cancer, cardiovascular disease, diabetes, etc. These diseases share common modifiable risk factors: alcohol consumption,

Body Mass Index (BMI), cigarette smoking, unhealthy diet, and physically inactive [13]. Modifiable lifestyle risk factors are the lifestyles of an individual exhibited in their daily activities. The modifiable lifestyle risk factors include Body Mass Index (BMI), level of consumption of alcohol and cigarette, daily fruit and vegetable consumption (diet), physical activity [13].

The MDRD method estimates the GFR using creatinine measurement, sex, race and Age, which can be seen to limit the number of factors needed to predict the stage of CKD efficiently, a challenge identified as a controversy, that is, diagnosing CKD without any consideration of aetiology [14]. This work improves the number of features to determine the stage of Chronic Kidney

Disease, which is a limitation of the MDRD model.

Table 2. Classification of CKD defined by

Stage	State of kidney function	Severity by Glomerulus Filtration Rate (GFR)
1	Normal or Increasing levels	GFR \geq 90
2	Mild damage	GFR of 60–89
3	Moderate function	GFR of 30–59
4	Severe damage	GFR of 15–29
5	Renal failure	GFR of 15 (or dialysis)

KDOQI.

Features/Variable Statistics

The statistical analysis of the collected data reveals that the average Age of patients considered in this work is forty-nine years (49yrs), with an average serum creatinine measure at 529 mol/L, which is also very high, having a more significant influence on the incidence of chronic kidney disease. The mean values of diabetes also show that the patients diagnosed with a diabetic condition is most likely to have a high value for the serum creatinine. The standard deviation for the various attributes, especially for the attribute age, shows that, even though the minimum Age is eighteen years (18yrs), most of the patients with chronic kidney disease can be thought to be within Age thirty-four sixty-four (64yrs). Table 6 gives the exact values for the statistical results of the collected data.

Machine Learning Techniques

The adaptive Neuro-Fuzzy Inference System is the machine learning technique used in this study. It is a hybrid of an Artificial Neural Network and a Fuzzy Inference System. The strength of fuzzy logic to reason under uncertainty and that of the artificial neural network to learn from a given dataset is combined to help predict the risk of chronic kidney disease of an individual. The Artificial Neural Network and the Fuzzy Inference System are supervised and unsupervised learning techniques, respectively. The fuzzy Inference System uses the subtractive clustering algorithm to generate the rules of the Fuzzy Inference System [15].

The machine-learning algorithm uses two phases; the learning phase and the testing /validation phase. The training phase ensures that the predictive or classification model is built by using the training dataset. The testing /validation phase, on the testing dataset, estimates the performance of the generated/built models.

Artificial Neural Network.

The work of Warren McCulloch and Walter Pitts in 1948 laid the foundation of Artificial Neural Network (ANN), which aimed to describe how the primary processing of the elements of the brain function. The ANN is made up input layer, hidden layer(s) and the output layer. Each layer consists of many nodes with its associated parameters. The nodes are connected through directional links identified with neurons that make up the brain. ANN can implicitly detect complex nonlinear relationships between dependent and independent variables, i.e. detecting all possible interactions between predictor variables and the availability of multiple training algorithms. The adaptive ability of a node in the neural network is due to the

output of each node. The result of each node is dependent on its associated parameters from the data set. The neural network computes parameters for each node, and these parameters are updated based on the learning rule to minimize the error measure. The learning rule for a given ANN is usually Gradient Descent and the chain rule. Still, due to the limitation of the Gradient Descent, which is slowness and tendency to be trapped in local minima, Hybrid Learning Rule is preferred.

2.2. Fuzzy Inference System.

Fuzzy inference is formulating the mapping from a given input to an output using fuzzy logic. Fuzzy logic is the technique used to derive logic described as fuzziness describing a phenomenon or quantity ratio. Unlike the usual uncertainty in the computation of quantitative analysis, the fuzzy inference system employs the fuzzy if-then rules to model the qualitative aspect of human knowledge and reasoning processes. This provides the mathematical tools to build a fuzzy model of a system. A fuzzy if-then rule consists of ANTECEDENT (premises) and CONSEQUENT. The premise component of the rule is made up of a Linguistic variable

representing a feature and linguistic value/label.

The fuzzy Inference system has five functional blocks, as illustrated in Figure 1. These include fuzzy rules, fuzzy inference engine, fuzzification database, and defuzzification. Fuzzification is the interface of the inference system that transforms the crisp inputs into the degree of association or match with the linguistic values leveraging on the definitions of Membership Functions (MFs) of the fuzzy sets needed for the fuzzification the crisp input values contained in the Database. Fuzzy Inference Engine is the decision-making unit of the system that performs inference operations on the rules captured in the fuzzy rule base. The fuzzy if-then rules or fuzzy conditional statements are expressions of the form IF A THEN B, where A and B are labels of fuzzy sets characterized by appropriate membership functions. The Defuzzification component transforms the unclear results into crisp output. The defuzzification of the outcomes of the fuzzy system into crisp output is by using a defuzzification method; Max-membership, Centroid Method, Mean-Max membership and Weighted-Average Method

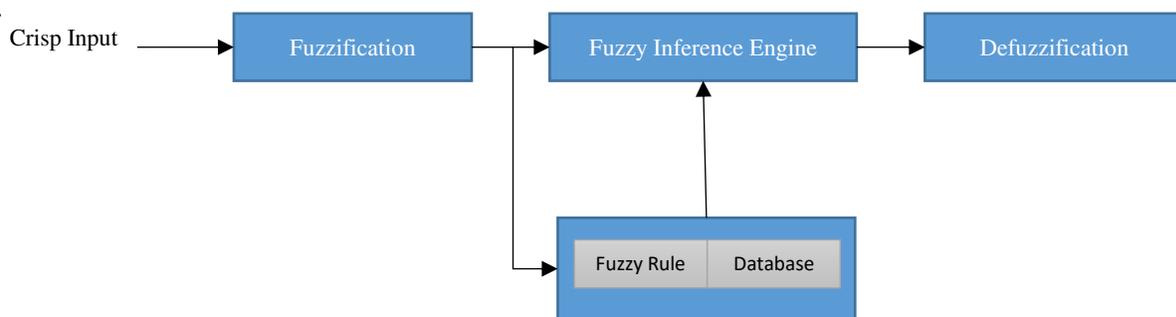


Figure 1. Fuzzy Inference System.

Different inference systems follow the type of defuzzification method used; Type 1 uses Weighted Average, Type 2 uses the centroid of area, mean of maxima, bisector of area, and maximum criteria. Type 3 of the fuzzy

The inference system is the Takagi-Sugeno-Kang inference system whose output is by a linear combination of the result of each rule plus a constant term. The final output of the Takagi-Sugeno-Kang is the weighted average of the output of each rule.

The Adaptive Neuro-Fuzzy Inference System is based on the Takagi-Sugeno Inference system. Takagi-Sugeno-Kang (TSK) inference system, besides Mamdani fuzzy system, is one of the widely used fuzzy models. The approach of the TSK-style fuzzy system produces the crisper output directly as the rule consequents use polynomials.

A TSK-style fuzzy rule base comprising of n rules with m antecedents each takes the form:

$$R_1: \text{IF } x_1 \text{ is } A_{11} \text{ and } \dots \text{ and } x_m \text{ is } A_{m1} \text{ THEN } y = f_1(x_1, \dots, x_m) = \beta_{01} + \beta_{11}x_1 + \beta_{21}x_2 + \dots + \beta_{m1}x_m \quad (1)$$

$$R_2: \text{IF } x_1 \text{ is } A_{12} \text{ and } \dots \text{ and } x_m \text{ is } A_{m2} \text{ THEN } y = f_2(x_1, \dots, x_m) = \beta_{02} + \beta_{12}x_1 + \beta_{22}x_2 + \dots + \beta_{m2}x_m \quad (2)$$

·
·
·

$$R_m: \text{IF } x_1 \text{ is } A_{1n} \text{ and } \dots \text{ and } x_m \text{ is } A_{mn} \text{ THEN } y = f_n(x_1, \dots, x_m) = \beta_{0n} + \beta_{1n}x_1 + \beta_{2n}x_2 + \dots + \beta_{mn}x_m \quad (3)$$

Where R_i and β_{0i}, β_{si} are the number of rules and the constant parameters of the linear function of the rule consequents, respectively, $i \in \{1, 2, \dots, n\}$ and $s \in \{1, 2, \dots, m\}$.

Adaptive Neuro-Fuzzy Inference System.

The ANFIS technique as applied in this work is identified by the style of the rule and the consequent, which is based on the TSK-fuzzy style (type 3 fuzzy inference system). Figure 2 shows the ANFIS generalized architecture.

Let assume a fuzzy inference system with;

Two inputs x and y and one output, z .

Rule base containing two rules.

Rule 1: If x is A_1 and y is B_1 Then $f_1 = p_1x + q_1y + r_1$.

Rule 2: If x is A_2 and y is B_2 Then $f_2 = p_2x + q_2y + r_2$

LAYER 1

Every node i in this layer is a square node with a node function

$$O_i^1 = \mu_{A_i}(x) \quad (4)$$

Where x is the input to node i and A_i is the linguistic label/value. In other words, is the membership function of O_i^1 and it specifies the degree to which the given x satisfies the quantifier A_i . The bell-shaped function is chosen for μ_{A_i} .

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{x - c_i}{a_i} \right)^2 \right]^{b_i}} \quad (5)$$

Or

$$\mu_{A_i}(x) = \exp \left\{ - \left(\frac{x - c_i}{a_i} \right)^2 \right\} \quad (6)$$

Where $\{a_i, b_i, c_i\}$ is the parameter set. Changes in the parameters also influence the shape of the bell-shape function on the linguistic label A_i . Other membership functions such trapezoidal, triangular-shaped. Parameters in this layer are referred to as premise parameters.

LAYER 2

Every node in this layer is a circle node labelled Π , which multiplies the incoming signals and sends the product out as an output

$$w_i = \mu_{A_i}(x) \times \mu_{B_i}(y) \quad \text{for } i = 1, 2 \quad (7)$$

The output of every node represents the firing strength of a rule. T-norm operator that performs generalize AND can be used as node function in this layer.

LAYER 3

Every node in this layer is a circle node labelled N . The i th node calculates the ratio of the ... rule's firing strength to the sum of all firing strengths.

$$\bar{w}_i = \frac{w_i}{w_1 + w_2} \quad \text{for } i = 1, 2 \quad (8)$$

LAYER 4

Every node I in this layer is a square node with a node function

$$O_i^4 = \bar{w}_i f_i = w_i (p_i x + q_i y + r_i) \quad (9)$$

Where \bar{w}_i is the output of layer three and $\{p_i, q_i, r_i\}$ is the parameter set. The parameter in this layer is the consequent parameter.

LAYER 5

The single node in this layer is a circle node labelled \sum that computes the overall output as the summation of all incoming signals.

$$O_i^5 = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (10)$$

The learning algorithm implemented by the ANFIS technique is the Hybrid Learning Algorithm. In the forward pass of the Hybrid learning algorithm, functional signals go forward till Layer 4, and the consequent parameters are identified by Least Square Estimate (LSE). In the backward pass, the error rate propagates backwards, and the gradient descent updates the premise. There are four methods of updating the parameters in the ANFIS technique. These include using only the Gradient Descent, Using the Gradient Descent and One Pass of LSE, using both Gradient Descent and the LSE. The fourth method is by using the Sequential LSE only.

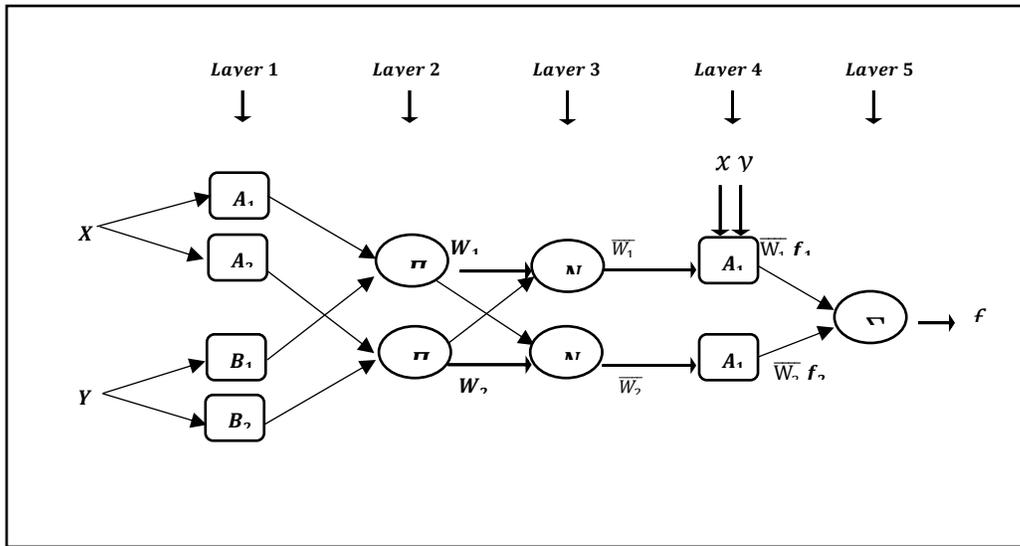


Figure 2. Generalized ANFIS Architecture (Source: J.S.R Jang)

Rule Generation by Takagi-Seguno-Kang (TSK) Fuzzy System.

The classification of a given data point by the ANFIS technique is by the design of the fuzzy rules. The TSK fuzzy inference system

generates the rules for a fuzzy system automatically by using the available dataset. TSK attempts to obtain fuzzy rules by categorizing and organizing the given dataset into clusters using a data clustering algorithm

[16]. The algorithm proposed in [17], outline the step by step sequence followed by the TSK for the generation of the rules of the fuzzy system.

2.3.1.1 Algorithm for Rule generation

Given $D = \{\text{input-output data pairs from unknown function of system } F\}$, $X = \{x_1, \dots, x_n\} \subset R^p$ and $Y = \{y_1, y_2, \dots, y_n\} \subset R$.

p is the dimension of the input vector.

1. Data pre-processing.
2. Generate initial rule-based which is just one rule by the Least Square Method. The fuzzy antecedents A_{ij} s in the initial rule are determined by

$$m_{A_{ij}} = \frac{\sum_{k=1}^n x_{kj}}{n} \quad (11)$$

$$\sigma_{A_{ij}} = \sqrt{\frac{\sum_{k=1}^n (x_{kj} - m_{A_{ij}})^2}{n-1}} \quad (12)$$

Where $j = 1, \dots, p$. The Least Square technique is used as the Consequent parameter.

3. Construct a new rule and add it to the fuzzy rule base. The vector with the worse error is identified (z_i') and taken as the candidate center for a new rule.

$$z_i' = \{z_i | MSE_{D_i} = \max_{k=1, \dots, n} (MSE_{D_k})\} \quad (13)$$

$$(MSE_{D_k}) = (y_k - \hat{y}_k)^2 \quad (14)$$

Fuzzy antecedent $A_{newrule,j}$ in a rule is characterized by

$$(m_{newrule,j}, \sigma_{newrule,j})_j = 1, \dots, p \quad (15)$$

Center of the new rule is identified by:

$$m_{newrule,j} = x_{i,j} \quad (16)$$

$$\sigma_{newrule,j} = \sqrt{\frac{\sum_{k=1}^n u_{k,i}^2 \|x_{k,j} - m_{newrule,j}\|^2}{\sum_{k=1}^n u_{k,i}^2}} \quad (17)$$

The consequent parameters are estimated using Locally Weighted Regression (LWR), a way to construct regression surface through a multivariate smoothing procedure. The resulting parameters of the new rule are computed by minimizing the objective function J , defined as:

$$J = \sum_{k=1}^n w_k (y_k - \hat{y}_k)^2 \quad (18)$$

w_k is the weighting factor which is assumed and is the same value as the degree of membership. The weighting factor ensures that the consequence parameters are influenced only by the data point within the fuzzy set that defines the region of validity of the new rule. The fuzzy rule acts like an independent model that is only related to a subset of training data.

4. Tune the system parameters which lead to the generation of the rule base, $R = \{R_1, R_2, \dots, R_M\}$. $M = \text{number of rules}$. The tuning of the parameters on the antecedent and consequent of the whole rule base, R , involves the Hybrid learning algorithm, which ensures the Neural Network's learning technique improves approximation precision.

The Hybrid learning algorithm combines a recursive singular value decomposition-based least-squares estimator and the gradient descent to tune or refine the parameters of the rule base.

5. The last step is to measure the performance of the generated model. This is by the use of Mean Square Error evaluation criteria, given as:

$$MSE = \frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2 \quad (19)$$

If the $MSE \geq \varepsilon$ and $M < M_{max}$ then the algorithm returns to step 3; otherwise, the

given termination conditions are satisfied, and the fuzzy model is constructed.

Results

The result section elaborates the data, evaluation matrix, and predictive model performance generated in this research work. The dataset consisted of 180 records of patients that generated 294 data points.

Evaluation Matrix

The key performance indicator (KPI) for the system's predictive model is the Root Mean Square Error (RMSE) and the Confusion Matrix.

Root Mean Square Error (RMSE) =

$$\sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (20)$$

$$\text{Precision, } P_i = \frac{TTP_{all}}{TTP_{all} + TFP_i} \quad (19)$$

$$\text{Recall, } R_i = \frac{TTP_{all}}{TTP_{all} + TFN_i} \quad (20)$$

$$\text{Specificity, } S_i = \frac{TTN_{all}}{TTP_{all} + TFP_i} \quad (21)$$

Table 3. Confusion Matrix Result for NANFIS

	Stage	Predicted Stages				
		One	Two	Three	Four	Five
Actual Stages	One	1	1	0	0	0
	Two	2	2	2	1	0
	Three	4	3	8	3	1
	Four	1	11	3	5	5
	Five	0	10	0	17	18

$$\text{Overall Accuracy} = \frac{TTP_{all}}{\text{Total number of Testing Entries}} \quad (24)$$

Where

Total False Negative (*TFN*) is total false-negative classification cases

Total False Positive (*TFP*) is total false positive classification cases

Total True Negative (*TFN*) is total false-negative classification cases

Total True Positive (*TTP*) is total true positive classification cases

Table 4. Result for the Confusion Matrix

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
$P_i = \frac{TTP_{all}}{TTP_{all} + TFP_i}$	0.56	0.50	0.74	0.77	0.90
$R_i = \frac{TTP_{all}}{TTP_{all} + TFP_i}$	0.13	0.07	0.62	0.77	0.95
$S_i = \frac{TTN_{all}}{TTP_{all} + TFP_i}$	0.50	0.29	0.85	0.78	0.84
Overall accuracy = $\frac{TTP_{all}}{TTP_{all}}$					0.52
95 % Confidence Interval (CI)					(0.41, 0.62)

The model's performance based on the confusion matrix (Table 3 and 4) indicates that Stage 5 had the highest correctly predicted case with a True Positive value of eighteen (18). Stage 4 had the highest False Negative with 17 cases all from the Stage 5

class. This could be attributed to the fact that most of the cases from the data set were at Stage 5 of chronic kidney disease, alluding to the notion of detecting the disease at an advanced stage.

The measure of Recall (also known as Sensitivity) of the predictive indicates how the model correctly classifies the patient when the condition for the various stages is satisfied. The predictive model recorded higher Recall values for classifying Stage 5 – Stage 3. This implies that the NANFIS model can detect the deterioration in kidney performance as early Stage 3, which is an excellent stage to treat the disease to avoid dialysis and kidney transplant.

The model's performance in not classifying one stage as another is good, especially for the three steps discussed above(Stage 3 – Stage 5).

The performance of the NANFIS model is compared to the acceptable Modification of Diet in Renal Disease (MDRD) model for the prediction of the stage of chronic kidney disease using the estimated Glomerulus Filtration Rate (eGFR). The accuracy of the MDRD based on the data from the Komfo Anokye Teaching Hospitals is 0.427 (42.7%) as indicated in Table 5. compared to the 5.196 (51.69%) of the NANFIS model. The MDRD predicted thirty-eight (38) correctly, whereas the NANFIS model correctly predicted forty-six (46). Based on the available data, it can be concluded that the NANFIS model is a better predictor than the MDRD equation, which can be attributed to the NANFIS model considering the patient's lifestyle and medical history.

The NANFIS model identified the most influencing features and the relationship

between those features and the stage of chronic kidney disease. As indicated in Fig. 3 below, the features with the most influence include Age, Serum Creatinine, Gender, and the amount of Salt-Intake.

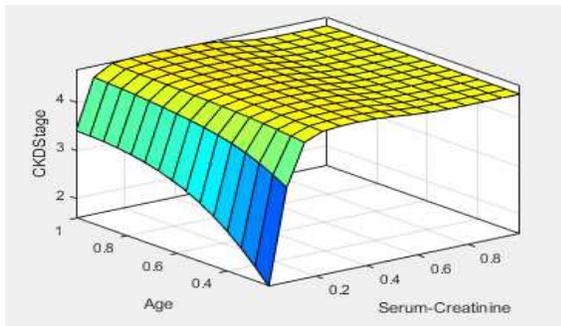
The stage of chronic kidney disease that the Age of a person most influences in conjunction with Serum Creatinine is across all the steps, as shown in Figure 3(a). The graph concerning the Age of the patient indicates that the younger the patient, the lower the setting of chronic kidney disease. The older patients with the disease usually have stage three and above cases. The Gender of the patient, as shown in Figure 3 (b), indicates that most men are diagnosed with chronic kidney disease at a later stage compared with their female counterparts. This calls for the attention of most men to have periodic medical screening to know how their kidney is performing and to get help as early as possible before their kidney become worse.

The Salt-Intake also influence the incidence of chronic kidney disease across all the stages. Consuming too much salt poses a significant challenge to the filtering ability of the kidney. On the other hand, hypertension accelerates the kidney's rate of damage, considering Figure 3(c). The CKD stage associated with hypertension begins from stage 3 and goes above stage 4. Thus, the most likely stage of prediction based on these variables in conjunction with Serum Creatinine is across all stages of chronic kidney disease.

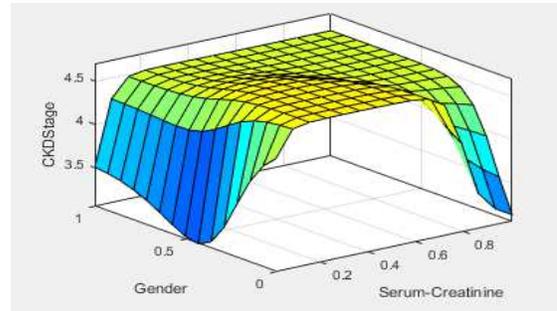
Table 5. Comparison of NANFIS and MDRD Glomerulus Filtration Rate Equation.

Expected Output	Predictions		
	NANFIS Model		MDRD Model
	Initial	Optimized	
Correctly Predicted	26	46	38
Wrongly Predicted	63	43	51
Test Data size	89	89	89
Accuracy (%)	30	52	43

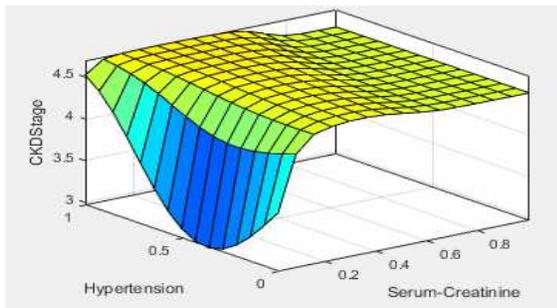
The Salt-Intake also influence the incidence of chronic kidney disease across all the stages. Consuming too much salt poses a significant challenge to the filtering ability of the kidney. On the other hand, hypertension accelerates the rate at which the kidney gets damage considering Figure 3(c). The CKD stage associated with hypertension begins from stage 3 and goes above stage 4. Thus, the most likely stage of prediction based on these variables in conjunction with Serum Creatinine is across all stages of chronic kidney disease.



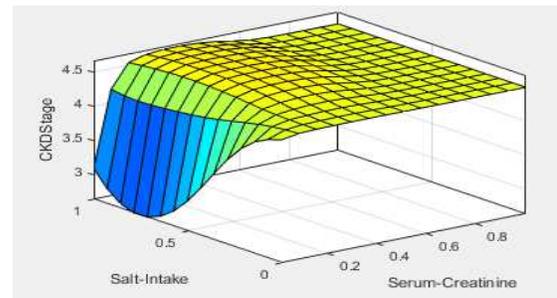
a



b



c



d

Figure 3. Surface view of the relationships between input and output variables (Serum Creatinine and others).

Table 6. Summary of Data statistics of the model variable considered in the ANFIS Modeling

<i>Data Statistics</i>	<i>Input Variable</i>										
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
	Age	Gender	Family History	Hypertension	Alcohol Intake	Salt Intake	Serum Creatinine	Specific Gravity	Albumin	Urea/BUN	Diabetes
Mean	49.2	0.5	0.2	3.9	0.6	1.9	529.0	1.0	16.2	22.3	3.0
Standard Error	0.9	0.0	0.0	0.3	0.1	0.0	40.3	0.0	0.7	4.8	0.3
Median	51.0	0.0	0.0	2.0	0.0	2.0	274.0	1.0	10.0	11.3	0.5
Mode	18.0	0.0	0.0	0.5	0.0	2.0	53.0	1.0	10.0	2.1	0.5
Standard Deviation	15.5	0.5	0.4	4.9	0.9	0.5	689.6	0.5	11.8	81.7	5.6
Sample Variance	241.4	0.3	0.1	23.6	0.7	0.3	475577.5	0.3	139.1	6682.9	31.6
Kurtosis	-0.7	-2.0	1.7	2.0	-1.2	5.3	41.7	292.9	0.9	205.5	6.4
Skewness	-0.2	0.1	1.9	1.6	0.8	-1.9	4.9	17.1	1.6	13.7	2.6
Range	61.0	1.0	1.0	20.0	2.0	3.0	7710.8	9.1	48.1	1291.3	26.0
Minimum	18.0	0.0	0.0	0.0	0.0	0.0	11.2	1.0	1.9	1.8	0.0
Maximum	79.0	1.0	1.0	20.0	2.0	3.0	7722.0	10.2	50.0	1293.0	26.0
Sum	14406.0	138.0	45.0	1153.5	188.0	545.0	154984.9	304.3	4735.3	6524.1	887.0
Count	293	293	293	293	293	293	293	293	293	293	293

Conclusion.

Chronic Kidney Disease has several factors. Limiting the number of factors to four which is the case of the Modification of Diet in Renal Disease (MDRD) model, is not enough to comprehensively determine the stage of the disease for a patient.

Considering the aetiology (which includes the lifestyle, laboratory test result, and the patient's medical history with respect to hypertension and diabetes) in this work provides an improved perspective in predicting the stage of the disease.

The accuracy of the prediction of the developed model (NANFIS) compared to the MDRD is better even though the dataset was not big enough.

The model performs excellently with high accuracy in correctly classifying the patient when the condition for each stage is satisfied especially for Stages 3, 4 and 5.

This model can be improved with increased data, which can be considered an improvement to the NANFIS model.

Abbreviations

NANFIS: Nephrology Adaptive Neuro-Fuzzy Inference System; MDRD: Modification of Diet for Renal Disease; RMSE: Root Mean Square Error; ANFIS: Adaptive Neuro-Fuzzy Inference System.

Acknowledgement:

We thank the staff of Komfo Anokye Renal Clinic for their support and cooperation during the data collection.

Declaration

Ethics approval and Consent for publication

Committee on Human Research, Publication and Ethics- School of Medical Science/Komfo Anokye Teaching Hospital of Kwame Nkrumah University of Science and Technolog.

The Committee considered the ethical merit of the application for study and approved the protocol.

Data collected did not interact directly with the patients.

Availability of data and materials

The study data is obtained from the Record Department- Renal Clinic of Komfo Anokye Teaching Hospital. The Pre-processed dataset is provided as a supplementary material.

Competing interests

The authors declared no potential conflicts of interest in the authorship and publication of this manuscript.

Authors' Contributions

Prof. Adekoya and Dr Weyori contributed to the conception of the topic of study. They also contributed in terms of tools for data collection.

Dr. Tannor was very instrumental in the data collection and data interpretation. He works at Komfo Anoky Teaching Hospital. Bernard Andoh was involved in the conception and design of the research, collection of data, and analysis.

Reference

- [1] Lokpo, S. Y., Osei-Yeboah, J., Owiredo, W. K. B. A., Ussher, F. A., Orish, V. N., Gadzeto, F., ... Asumbasiya Aduko, R. (2018). Renal dysfunction among ghanaians living with clinically diagnosed hypertension in the asutifi-south district: A cross-sectional descriptive study at the St. Elizabeth hospital, Hwidiem. *International Journal of Hypertension*, 2018(November 2018). <https://doi.org/10.1155/2018/8428063>
- [2] Levey, A. S., Atkins, R., Coresh, J., Cohen, E. P., Collins, A. J., Eckardt, K. U., ... Eknoyan, G. (2007). Chronic kidney disease as a global public health problem: Approaches and initiatives - A position statement from Kidney Disease Improving Global Outcomes. *Kidney International*, 72(3), 247–259. <https://doi.org/10.1038/sj.ki.5002343>
- [3] Couser, W. G., Remuzzi, G., Mendis, S., & Tonelli, M. (2011). The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. *Kidney International*, 1–13. <https://doi.org/10.1038/ki.2011.368>
- [4] Dunkler, D., Kohl, M., Heinze, G., Teo, K. K., Rosengren, A., Pogue, J., ... Mann, J. F. E. (2015). Modifiable lifestyle and social factors affect chronic kidney disease in high-risk individuals with type 2 diabetes mellitus. *Kidney International*, 87(4), 784–791. <https://doi.org/10.1038/ki.2014.370>
- [5] Michishita, R., Matsuda, T., Kawakami, S., Kiyonaga, A., Tanaka, H., Morito, N., & Higaki, Y. (2016). The association between unhealthy lifestyle behaviors and the prevalence of chronic kidney disease (CKD) in middle-aged and older men. *Journal of Epidemiology*, 26(7), 378–385. <https://doi.org/10.2188/jea.JE20150202>
- [6] Ministry of Health. (2019, August 07). Komfo Anokye Teaching Hospital. Retrieved July 06, 2020, from <https://www.moh.gov.gh/komfo-anokye-teaching-hospital/>
- [7] National Kidney Foundation. (2018, June 26). Fabry Disease. Retrieved July 09, 2020, from <https://www.kidney.org/atoz/content/fabry>
- [8] National Kidney Foundation. (2020, June 25). How Your Kidneys Work. Retrieved July 03, 2020, from <https://www.kidney.org/kidneydisease/howkidneyswork>
- [9] Davita Kidney Care. (2020). What is creatinine. Retrieved July 09, 2020, from <https://www.davita.com/education/kidney-disease/symptoms/what-is-creatinine>
- [10] Blood Pressure Association. (2008). Salt's effects. Retrieved July 09, 2020, from <http://www.bloodpressureuk.org/microsites/salt/Home/Whysaltisbad/Saltseffects>
- [11] Mallamaci, F. (2014). Salt Restriction in Chronic Kidney Disease : A Simple Need or a Must ? 124–128. <https://doi.org/10.1159/000355786>
- [12] Chiu, R. K., Chen, R. Y., Wang, S.-A., Chang, Y.-C., & Chen, L.-C. (2013). Intelligent Systems Developed for the Early Detection of Chronic Kidney Disease. *Advances in Artificial Neural Systems*, 2013, 1–7. <https://doi.org/10.1155/2013/539570>
- [13] Ng, R., Sutradhar, R., Yao, Z., Wodchis, W. P., & Rosella, L. C. (2019). Smoking, drinking, diet and physical activity—modifiable lifestyle risk factors and their associations with Age to first chronic disease. *International Journal of Epidemiology*, 1–18. <https://doi.org/10.1093/ije/dyz078>
- [14] Polkinghorne, K. R. (2011). Controversies in Chronic Kidney Disease Staging. 32(May), 55–59.
- [15] Kour, H., Manhas, J., & Sharma, V. (2019). Brief Paper: Evaluation of Subtractive Clustering based Adaptive Neuro-Fuzzy Inference System with Fuzzy C-Means based ANFIS System in Diagnosis of Alzheimer. *Journal of Multimedia Information System*, 6(2), 87–90. <https://doi.org/10.33851/jmis.2019.6.2.87>
- [16] Tawafan, A. H., Profesor, E., Bin, M., Madya, P., Bin, Z., Tawafan, A. H., Sulaiman, M., Ibrahim, Z. Bin, Engineering, E., & Teknikal, U. (N.D.). Detection Of High Impedance Fault On Power Distribution System Using Probabilistic Neural Network Detection Of High Impedance Fault On Power Distribution System Using Probabilistic Neural Network.
- [17] Rezaee, Babak & Zarandi, Mohammad. (2010). Data-driven fuzzy modeling for Takagi-Sugeno-Kang fuzzy system. *Inf. Sci.*. 180. 241-255. 10.1016/j.ins.2009.08.021.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryData.docx](#)