The case for early identification and intervention of chronic kidney disease: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference

Michael G. Shlipak[1,2], Sri Lekha Tummalapalli[1,2], L. Ebony Boulware[3], Morgan E. Grams[4,5], Joachim H. Ix[6,7,8], Vivekanand Jha[9,10,11], Andre-Pascal Kengne[12,13], Magdalena Madero[14], Borislava Mihaylova[15,16], Navdeep Tangri[17], Michael Cheung[18], Michel Jadoul[19], Wolfgang C. Winkelmayer[20] and Sophia Zoungas[21,22]; for Conference Participants[23]

[1]*Kidney Health Research Collaborative, Department of Medicine, University of California San Francisco, San Francisco, California, USA;* [2]*General Internal Medicine Division, Medical Service, San Francisco Veterans Affairs Health Care System, San Francisco, California, USA;* [3]*Department of General Internal Medicine, Duke University School of Medicine, Durham, North Carolina, USA;* [4]*Division of Nephrology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA;* [5]*Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA;* [6]*Division of Nephrology-Hypertension, Department of Medicine, University of California San Diego, San Diego, California, USA;* [7]*Nephrology Section, Veterans Affairs San Diego Healthcare System, La Jolla, California, USA;* [8]*Division of Preventive Medicine, Department of Family Medicine and Public Health, University of California San Diego, San Diego, California, USA;* [9]*George Institute for Global Health, University of New South Wales, New Delhi, India;* [10]*University of Oxford, Oxford, UK;* [11]*Department of Nephrology, Manipal Academy of Higher Education, Manipal, India;* [12]*Non-Communicable Diseases Research Unit, South African Medical Research Council, Cape Town, South Africa;* [13]*Department of Medicine, University of Cape Town, Cape Town, South Africa;* [14]*Department of Medicine, Division of Nephrology, Instituto Nacional de Cardiología Ignacio Chávez, Mexico City, Mexico;* [15]*Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK;* [16]*Institute of Population Health Sciences, Blizard Institute, Queen Mary University of London, London, UK;* [17]*Department of Community Health Services, Max Rady College of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada;* [18]*Kidney Disease: Improving Global Outcomes, Brussels, Belgium;* [19]*Cliniques Universitaires Saint Luc, Université Catholique de Louvain, Brussels, Belgium;* [20]*Selzman Institute for Kidney Health, Section of Nephrology, Department of Medicine, Baylor College of Medicine, Houston, Texas, USA;* [21]*Diabetes and Vascular Medicine Unit, Monash Health, Clayton, Victoria, Australia; and* [22]*School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia*

Chronic kidney disease (CKD) causes substantial global morbidity and increases cardiovascular and all-cause mortality. Unlike other chronic diseases with established strategies for screening, there has been no consensus on whether health systems and governments should prioritize early identification and intervention for CKD. Guidelines on evaluating and managing early CKD are available but have not been universally adopted in the absence of incentives or quality measures for prioritizing CKD care. The burden of CKD falls disproportionately upon persons with lower socioeconomic status, who have a higher prevalence of CKD, limited access to treatment, and poorer outcomes. Therefore, identifying and treating CKD at the earliest stages is an equity imperative. In 2019, Kidney Disease: Improving Global Outcomes (KDIGO) held a controversies conference entitled "Early Identification and Intervention in CKD." Participants identified strategies for screening, risk stratification, and treatment for early CKD and the key health system and economic factors for implementing these processes. A consensus emerged that CKD screening coupled with risk stratification and treatment should be implemented immediately for high-risk persons and that this should ideally occur in primary or community care settings with tailoring to the local context.

Kidney International (2021) **99,** 34–47; https://doi.org/10.1016/j.kint.2020.10.012

KEYWORDS: albuminuria; chronic kidney disease; cost-effectiveness; creatinine; cystatin; detection; kidney failure; glomerular filtration rate; health systems; interventions; proteinuria; risk models; risk stratification; screening

Correspondence: *Michael G. Shlipak, University of California, 4150 Clement Street, Building 2, Room 455C, San Francisco, California, USA 94121. E-mail: michael.shlipak@ucsf.edu; or Sophia Zoungas, School of Public Health and Preventive Medicine, Monash University, Level 3, 553 St. Kilda Road, Melbourne 3004 Australia. E-mail: sophia.zoungas@monash.edu*

[23]See Appendix for list of Conference Participants.

Received 15 July 2020; revised 5 October 2020; accepted 13 October 2020; published online 27 October 2020

Nearly 700 million persons worldwide have chronic kidney disease (CKD), and the burden falls disproportionately upon socially disadvantaged and other vulnerable groups.[1] In many regions, persons with lower socioeconomic status have a higher prevalence of CKD, limited access to treatment, and poorer outcomes.[2–5] Early

Table 1 | Key conclusions from the Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference on Early Identification and Intervention

Populations for CKD screening, risk stratification, and treatment
Conclusion 1. Persons with hypertension, diabetes, or cardiovascular disease should be screened for CKD.
Conclusion 2. CKD screening and treatment programs should also be implemented in other high-risk individuals and populations based on comorbidities, environmental exposures, or genetic risk factors.
Conclusion 3. The initiation, frequency, and cessation of CKD screening should be individualized based on kidney and cardiovascular risk profiles and individual preferences.
Measurements for early CKD
Conclusion 4. CKD screening and risk stratification must consist of a dual assessment of estimated glomerular filtration rate (eGFR) and albuminuria (UACR).
Conclusion 5. Accurate GFR estimation includes both creatinine and cystatin C measurements for initial diagnosis and staging.
Conclusion 6. The combination of creatinine, cystatin C, and UACR for CKD screening is affordable in high-income settings.
Interventions for CKD
Conclusion 7. A key rationale for CKD screening is the availability of many effective interventions to delay CKD progression and reduce cardiovascular risk.
Conclusion 8. Accurate diagnosis and staging of CKD are necessary to utilize treatments effectively.
Conclusion 9. Patient engagement is a critical component of efforts to screen for and treat CKD.
Health system and economic factors
Conclusion 10. CKD screening and treatment efforts require multi-stakeholder implementation strategies to overcome barriers to high-quality CKD care.
Conclusion 11. Financial and nonfinancial incentives need to be aligned toward CKD screening, risk stratification, and treatment.
Conclusion 12. CKD screening in high-risk groups is likely to be cost-effective.
Conclusion 13. CKD screening approaches may differ in LMIC countries.

CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; GFR, glomerular filtration rate; LMICs, low- and middle-income countries; UACR, urine albumin-to-creatinine ratio.

identification of CKD by screening for kidney disease, followed by risk stratification and treatment, offers the potential to substantially reduce the morbidity and mortality from CKD and its related complications, such as cardiovascular disease.[6,7] However, at present, there is no accepted systematic strategy for CKD screening and treatment.

Despite effective methods to diagnose and treat CKD at its earliest stages, there is a lack of consensus on whether health systems and governments should implement CKD screening programs. Professional societies have been discordant on whether or not to screen for CKD.[8–10] To address this ongoing controversy, in October 2019, the Kidney Disease Improving Global Outcomes (KDIGO) held a Controversies Conference entitled "Early Identification and Intervention in CKD." Meeting participants represented a global multidisciplinary panel of clinicians and scientists. The rationale for CKD screening strategies was evaluated within the context of the World Health Organization (WHO) principles of screening for disease.[11,12] Four major topics were then addressed: (i) the selection of populations for CKD screening, (ii) the relative diagnostic and predictive characteristics of tests for kidney disease, (iii) the evidence base for treatments that reduce the risk of CKD progression and cardiovascular disease, and (iv) implementation strategies for CKD screening, risk stratification, and treatment programs and the key factors determining resource allocation and cost-effectiveness. The conference agenda, discussion questions, and plenary session presentations are available on the KDIGO website: https://kdigo.org/conferences/early-identification/.

After a comprehensive review of the conference topics by 4 breakout groups and by all attendees in collective discussion, consensus was reached to endorse a broad and proactive plan for CKD screening, risk stratification, and treatment with the goal of reducing the global burden of kidney disease (Table 1; Figure 1). The conference participants agreed that CKD met the WHO principles of screening for disease, as early CKD is asymptomatic, there are accurate and low-cost diagnostic tests, and effective treatments can be initiated in the early stages (Table 2[1,10,13–33]). Furthermore, the development of a CKD screening program was considered to be an equity imperative, particularly because socially disadvantaged and other vulnerable populations experience a disproportionate burden of CKD and are the least likely to receive effective treatments to reduce the risk of complications and improve outcomes.[34–36]

Two additional themes to emerge from the conference serve as important underlying principles for CKD screening strategies. First, as highlighted by patient representatives and advocates, there was a strong belief that patients overwhelmingly prefer earlier CKD screening and diagnosis and that patient education has the potential to improve self-management and disease prognosis. Second, a significant economic rationale was evident for early CKD screening, risk stratification, and treatment, particularly given the costs of kidney failure to individuals, health care systems, and society. For example, in the United States alone, Medicare spending for those with CKD or kidney failure is estimated to be more than 114 billion dollars annually.[37] The economic burden is

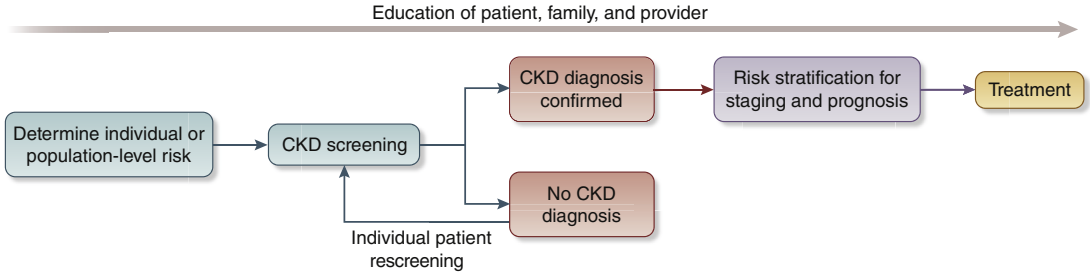

Figure 1 | Conceptual framework of a chronic kidney disease (CKD) screening, risk stratification, and treatment program.

particularly acute in low and middle income countries (LMICs); an estimated 188 million people experience catastrophic health expenditure annually as a result of kidney diseases across LMICs, the greatest of any disease group.[38] Specific components of an optimal CKD screening, risk stratification, and treatment strategy are described in the sections that follow.

Populations for CKD screening, risk stratification, and treatment

Conference participants reviewed different potential objectives for CKD screening and treatment initiatives: the identification of all persons with CKD, the identification of individuals within high-risk populations to maximize testing yield, or the identification of individuals with CKD who are most likely to progress to kidney failure or experience cardiovascular or other CKD complications. Additional topics of discussion included the definitions of high-risk populations and the optimal frequency of rescreening. The conference participants concluded that individual and population-level risk of having CKD and experiencing its complications should inform whether persons should be screened for CKD. Decisions concerning the age to initiate testing, the frequency of repeat testing, and the time to forgo or end testing should all be individualized based upon risk factors, preferences, and life expectancy.

Conclusion 1. Persons with hypertension, diabetes, or cardiovascular disease should be screened for CKD. The conference participants concluded that efforts for early CKD detection should first be implemented in individuals with established CKD risk factors, given the higher expected prevalence of CKD among those individuals. Conference participants recognized that identifying and treating all CKD cases would be the most complete approach to improving kidney health and reducing the burden of kidney disease. However, population-wide CKD screening programs were noted to have potential drawbacks, including higher costs and greater barriers to implementation than targeted high-risk CKD screening. A strategy of CKD screening in populations with common and important CKD risk factors will prioritize the identification of cases at high risk for CKD progression and cardiovascular events and with established treatment strategies. This approach would detect individuals with CKD at a lower cost per case identified but could miss CKD cases that are attributable to less common or unrecognized risk factors.

After reviewing data on worldwide CKD prevalence in high-risk populations, the conference participants concluded that CKD screening should be implemented for groups with these well-accepted CKD risk factors: hypertension, diabetes, and/or cardiovascular disease.[39] For example, in the United States, over 20% of individuals with hypertension have increased albuminuria (urine albumin-to-creatinine ratio [UACR] $\geq$ 30 mg/g), yet only 7% of those with hypertension are tested for albuminuria, representing potentially millions of patients with undiagnosed CKD.[40] Similarly, participants with self-reported cardiovascular disease in the National Health and Nutrition Examination Survey in the United States had a high CKD prevalence of over 40%.[37] The conference participants recognized that a CKD screening strategy based upon the presence of specific risk factors could miss a large population of patients who have not yet been diagnosed with those CKD risk factors or who are not followed in primary care.

Conclusion 2. CKD screening and treatment programs should also be implemented in other high-risk individuals and populations based upon comorbidities, environmental exposures, or genetic factors. Several other characteristics that may elevate an individual's likelihood of having CKD, including older age, race/ethnicity, other systemic diseases that impact kidneys (e.g. systemic lupus erythematosus, HIV infection), family history of kidney disease, genetic risk factors, poor access to health care or low socioeconomic status, high-risk occupations and environmental exposures, prior acute kidney injury,[33] preeclampsia, exposure to nephrotoxins, and obesity, were identified.[25] CKD screening for persons with these risk factors should be guided by an individualized clinical assessment and joint decision making, rather than with a uniform approach.

Conclusion 3. The initiation, frequency, and cessation of CKD screening should be individualized based on kidney and cardiovascular risk profiles and individual preferences. The time in the life course when screening for CKD should begin should be based upon the estimated likelihood of having CKD for that individual, rather than chronologic age alone. Thus, conference participants favored initiating CKD screening based on comorbidities and individualized risk assessment rather than at a specific chronologic age.

The frequency of testing is a critical aspect of CKD screening programs and has substantial impact on costs. The conference

Table 2 | CKD screening meets the WHO's principles of screening for disease

WHO criteria	CKD screening
1. The condition sought should be an important public health problem.	• CKD is highly prevalent, costly, and its worldwide disease burden is increasing.[1]
2. There should be an accepted treatment for patients with recognized disease.	• Treatments for recognized CKD can be initiated during early stages, are accepted, and are highly effective. Use of ACEi/ARBs in CKD patients substantially reduces risk of kidney failure. A meta-analysis of randomized trials showed that ACEi/ARB therapy lowered the odds of kidney failure by 30%–39% and of CVD events by 18%–24%.[13] • Statin use has been shown to significantly decrease risk of cardiovascular events and mortality in patients with CKD.[14] • In randomized trials, intensified blood pressure control in CKD patients reduced rates of fatal and nonfatal CVD events and all-cause mortality.[15] • In type 1 diabetes, long-term glycemic control reduces incidence of CKD.[16] In type 2 diabetes, glycemic control may be particularly beneficial in earlier CKD stages.[17] • SGLT2 inhibitors in patients with CKD with proteinuria and diabetes reduce the risk of kidney failure by 30%–40%.[18,19]
3. Facilities for diagnosis and treatment should be available.	• CKD screening and treatment in earlier stages could occur in primary-care practices or community-based settings.
4. There should be a recognizable latent or early symptomatic stage.	• CKD is asymptomatic until late stages. The asymptomatic stage contributes to low awareness of CKD in patients with the diagnosis.[20] Therefore, a screening program could shift recognition of CKD into much earlier stages relative to current practice.
5. There should be a suitable test or examination.	• There are low-cost and accurate tests for CKD. Serum creatinine and cystatin C are accurate tests to estimate GFR. Quantitative UACR is a sensitive measurement of kidney damage, whereas urine dipstick proteinuria is lower-cost but has lower sensitivity.[21]
6. The test should be acceptable to the population.	• Testing for CKD is accepted by the population. CKD is tested through standard venipuncture and non-invasive urine testing, and individuals with CKD express a preference for early communication about a CKD diagnosis.[22]
7. The natural history of the condition, including development from latent to declared disease, should be adequately understood.	• The natural history of CKD, including consequences of inadequate treatment, are well-understood. A large body of evidence has shown that CKD increases the risk of cardiovascular events and mortality. If not adequately treated, CKD may progress to kidney failure, and this risk can be predicted using risk equations.[23,24]
8. There should be an agreed policy on whom to treat as patients.	• There are clear guidelines for CKD treatment upon CKD detection. KDIGO guidelines recommend ACEi/ARB therapy in diabetic and nondiabetic adults with CKD and urine albumin excretion >300 mg/24 hours (or equivalent), and suggest ACEi/ARB use in diabetic adults with CKD and urine albumin excretion 30–300 mg/24 hours (or equivalent).[25] • Statin use is recommended for all patients with CKD above the age of 50 or with intermediate or high atherosclerotic cardiovascular disease risk, and CKD is a risk enhancer for patients with borderline atherosclerotic cardiovascular disease.[26,27] • Current guidelines from the American College of Cardiology/American Heart Association (ACC/AHA) recommend that adults with hypertension and CKD should be placed in a high-risk category and be treated to a BP goal of less than 130/80 mm Hg, without delaying treatment for a trial of nonpharmacologic interventions.[28] Forthcoming KDIGO BP guideline update will advocate targeting systolic BP < 120 mm Hg. • SGLT2 inhibitors are recommended in persons with diabetes and CKD.[10]
9. The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.	• CKD screening in high-risk groups is likely to be cost-effective. Several cost-effectiveness analyses of urine dipsticks and other methods of CKD screening have found that screening persons with hypertension or diabetes was cost-effective in simulation models.[29,30] Given the low cost of testing, strategies can be tailored to the resources of the specific health care system, as with other screening targets.[31]
10. Case-finding should be a continuing process and not a "once and for all" project.	• If a screening program is implemented, repeated screening is necessary to detect incident CKD in individuals with an initial negative screen. One-time screening does not capture high lifetime risk of CKD.[32] • An efficient detection strategy could tailor the timing of the next testing to the probability of new CKD based upon risk factors and current test results.[33]

ACEi, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; BP, blood pressure; CKD, chronic kidney disease; CVD, cardiovascular disease; GFR, glomerular filtration rate; KDIGO, Kidney Disease: Improving Global Outcomes; SGLT2, sodium-glucose cotransporter-2; UACR, UACR, urine albumin-to-creatinine ratio; WHO, World Health Organization.

participants agreed that the frequency of repeat testing should not be uniform for all persons, but rather should be guided by each individual's risk of developing CKD, based in part upon the results of previous testing. Risk equations that estimate 5-year CKD probabilities could be used to guide the timing of subsequent testing.[33] For example, individuals with a high probability of incident CKD at 5 years could be rescreened at 1- or 2-year intervals, whereas those at low risk of incident CKD could be rescreened after several years.

The diagnosis of CKD is currently a controversial topic among older adults, who experience the greatest burden of CKD and are at the highest risk for certain complications, such as cardiovascular disease and heart failure.[41–43] Some nephrologists are concerned that CKD is overdiagnosed among older adults and have called for an age-adapted definition.[44] However, underdiagnosis of CKD in older adults also carries consequences, as older adults have the highest prevalence of CKD and CKD impacts their physical and cognitive function, medication safety, and cardiovascular prognosis. Conference participants acknowledged that there are potential harms associated with CKD overdiagnosis in older adults; in response, members emphasized that in older adults CKD should be appropriately diagnosed and risk stratified using all available measurements. For example, among those with an estimated glomerular filtration rate based on serum creatinine ($eGFR_{Cr}$) 45–59 ml/min per 1.73 m^2 and UACR $<$ 30 mg/g, the diagnosis of CKD should require confirmation by cystatin C testing, as recommended by the KDIGO 2012 guidelines (recommendation 1.4.3.5).[25] Second, in the setting of limited life expectancy, CKD treatment strategies should carefully weigh the risks and benefits of treatment. Older adults should not be excluded from CKD screening programs, and whether or not to screen for CKD in an older individual should be determined with a holistic approach to patient treatment goals as with other population-based screening programs, like cancer detection.

Measurements for early CKD

The ideal measurements for CKD testing would accurately screen, confirm, and stage CKD; risk stratify for important outcomes; and guide treatments relevant to prevention of kidney and cardiovascular complications. The conference participants reviewed the breadth of evidence related to testing methods for CKD and concluded that the ideal initial screening approach must consist of both eGFR (by creatinine, cystatin C, or both) and albuminuria measurement. Similarly, conference attendees agreed that at the stage of confirmation, the ideal CKD diagnosis would consist of the "triple marker" panel of serum creatinine, serum cystatin C, and UACR.

Conclusion 4. CKD screening and risk stratification must consist of a dual assessment of estimated glomerular filtration rate (eGFR) and albuminuria (UACR). Assessing both kidney function by eGFR and kidney damage by measuring albuminuria are critical both to detect and to risk stratify CKD.[45] Lower eGFR and higher albuminuria are both strongly associated with risks of cardiovascular events, kidney failure, and mortality, so their measurement is crucial for effective risk stratification of persons with CKD. The presence and severity of albuminuria also guides the use and dosage of treatments such as angiotensin-converting enzyme inhibitors and angiotensin receptor blockers (ACEi/ARBs). Despite guideline recommendations, many clinicians currently fail to assess albuminuria in patients with reduced eGFR[46–48] or diabetes.[49]

The preferred method for initial screening for CKD, in addition to eGFR measurement, is the UACR. UACR is preferred over urine protein-to-creatinine ratio because UACR specifically detects albumin, which has stronger associations with clinical outcomes and is required to appropriately stage CKD according to the KDIGO 2012 CKD guidelines.[25] Uurine protein-to-creatinine ratio also measures non-albumin proteins and may be less sensitive for albumin at the lower ranges of detection. Urine dipstick assays are inexpensive but less sensitive for detecting clinically important albuminuria, especially in the 30–300 mg/g range.[21]

Conclusion 5. Accurate GFR estimation includes both creatinine and cystatin C measurements for initial diagnosis and staging. There was strong agreement by the conference attendees that GFR should be estimated using creatinine and cystatin C in parallel, which would more accurately diagnose and stage CKD, with less misclassification, than a strategy based upon creatinine alone.[50,51] There is a low number needed to test with cystatin C to avoid CKD misclassification. For example, in a meta-analysis of 90,750 participants across 16 cohort studies, 23% of persons with eGFR 60–74 ml/min per 1.73 m^2 based on creatinine had eGFR $<$60 ml/min per 1.73 m^2 based on cystatin C.[50] Conversely, among persons with eGFR 45–59 ml/min per 1.73 m^2 by creatinine, 42% had eGFR $\geq$60 ml/min per 1.73 m^2 by cystatin C. Similarly, in the UK Biobank study of 440,526 participants, 14.5% of persons with an eGFR 60–75 ml/min per 1.73 m^2 by creatinine had eGFR $<$60 ml/min per 1.73 m^2 by cystatin C, and a total of 53.7% of persons with eGFR 45–59 ml/min per 1.73 m^2 by creatinine were reclassified to eGFR $\geq$60 ml/min per 1.73 m^2 by cystatin C.

The inclusion of cystatin C in CKD diagnosis and treatment initiatives is consistent with the KDIGO 2012 Guidelines[25] and is a critical component of accurate risk stratification, as cystatin C markedly strengthens the association between eGFR and cardiovascular events, kidney failure, and death.[50,51] Accurate eGFR staging has implications for treatment, as certain medications have indications that are determined in part by the GFR stage. Cystatin C has the additional advantage of offering GFR estimates that do not require the incorporation of a race coefficient, as is required for creatinine.[52] Although cystatin C is an integral test for the initial diagnosis and staging of CKD, conference participants agreed that it may not be required for routine monitoring after the diagnosis of CKD is made.

Conclusion 6. The combination of creatinine, cystatin C, and UACR for CKD screening is affordable in high-income settings. The cost of the tests used to screen for CKD has a major influence on the affordability and the cost-

effectiveness of a CKD screening and treatment program. Because these are all automated tests, the cost of each measure will decrease proportionately to the volume of tests that are conducted.[53] The actual costs to perform diagnostic tests often differ substantially from the reimbursement rate negotiated or set by payors for the tests, and costs are typically much lower than what a health system or laboratory might charge a private customer in the absence of health insurance or a national health plan. To assess international CKD testing availability and costs, KDIGO collaborated with the International Society of Nephrology (ISN) to conduct a global survey. Of 24 respondents (73% response rate), there was high availability of CKD laboratory tests globally and a low cost of serum creatinine and UACR testing. Reimbursement rates for the "triple marker" panel of serum creatinine, serum cystatin C, and UACR are likely affordable in high-income settings.

Interventions for CKD

A fundamental justification for the early detection of CKD is the availability of evidence-based interventions to slow the progression of CKD and reduce its complications. Accurate diagnosis and staging of CKD impact the choice of treatments. Conference participants reviewed the evidence for effective interventions in CKD.

Conclusion 7. A key rationale for CKD screening is the availability of many effective interventions to delay CKD progression and reduce cardiovascular risk. Because the utility of CKD screening and risk stratification depend upon the potential for effective treatments, the conference summarized available evidence for interventions that delay CKD progression and reduce cardiovascular risk (Figure 2[54,55]). A cornerstone of managing CKD is lifestyle modification, and although the benefits are largely inferred from observational studies,[56–60] implementing lifestyle recommendations poses no or low risk for patients. For all persons with CKD, recommendations include smoking cessation, regular exercise, and a healthy diet (fruit, vegetables, legumes, and whole grains) (Figure 3).

Intensive blood pressure lowering has been shown to reduce cardiovascular events and all-cause mortality in nondiabetic CKD, and is likely to be cost-effective.[15,61] Use of statins and ezetimibe lower the risk of major adverse cardiovascular events,[14,62,63] and low-cost statins are cost-effective in CKD.[64] Evidence is emerging on the role of intensive glucose control in reducing the risk of kidney events; however the renal benefits of intensive glucose control should be weighed with risks when using agents that cause hypoglycemia.[65] More recently, SGLT2 inhibitors show a strong benefit in slowing the progression of diabetic kidney disease, with the greatest absolute benefits in those with higher albuminuria; evidence for SGLT2 inhibitors in nondiabetic CKD is emerging.[18,19,66–68] Another class of glucose-lowering agents, glucagon-like peptide-1 (GLP-1) receptor agonists, significantly reduce cardiovascular outcomes among persons with CKD[69] and may slow diabetic kidney disease progression.[70,71] Treatment of metabolic acidosis in CKD G3 with fruits and vegetables or oral bicarbonate may preserve GFR and additional studies are ongoing.[72] In summary, contrary to the previous nihilism about the utility of CKD screening programs, there are numerous effective treatments now available for management of CKD.

Conclusion 8. Accurate diagnosis and staging of CKD are necessary to utilize treatments effectively. Accurate diagnosis and staging of CKD are critical in choosing effective treatments, dosing medications appropriately, and minimizing nephrotoxic burden. Early identification of CKD is important in the consideration of blood pressure management and targets. Identification of albuminuria impacts the choice of antihypertensive agent, as ACEi/ARBs are first-line agents in those with albuminuria, even in the absence of hypertension.

Accurate GFR estimates are crucial for safe and effective medication dosing in persons with CKD. For example, gabapentinoid use is prevalent among patients with CKD,[73] and drug accumulation increases risks of side-effects such as sedation. Similarly, the association of baclofen with risk of encephalopathy increases with higher CKD stage.[74] Diagnosis and staging of CKD also influences the need for nephrotoxin avoidance, including minimization of nonsteroidal anti-inflammatory drug exposure and i.v. contrast.

Conclusion 9. Patient engagement is a critical component of efforts to screen for and treat CKD. Patient and family education and engagement are critical to the success of early CKD care. When implemented appropriately, education can have a number of potential benefits, including improved patient activation, improved access to health care, improved access and adherence to medications, timely nephrology referral, dietician referral, and diabetes education. There is evidence suggesting that well designed, interactive, frequent, and multifaceted educational interventions that include both individual and group participation can improve knowledge and self-management for secondary prevention of CKD.[75,76] Although there is limited reimbursement for patient education on CKD in its earliest stages,[77] education programs are accessible across multiple languages and cultures and vital to providing patients information about kidney health. Barriers to educating patients about early CKD include the use of literature that is not at an appropriate level or in a familiar language,[78] poor access to multidisciplinary teams,[79] or lack of health system preparedness. Nephrology societies in individual countries can work to ensure that information is culturally appropriate.[80,81] Public education, shared decision making, culturally appropriate messaging, use of mobile technology, support groups, community outreach, community resources (food banks, housing), indigenous navigators/health coaches, and social marketing all have potential benefits for generating and maintaining engagement.

Several online tools and mobile applications are now available to engage people in the detection and management of early CKD.[82] The National Kidney Disease Education Program (NKDEP) sponsors an initiative to promote kidney

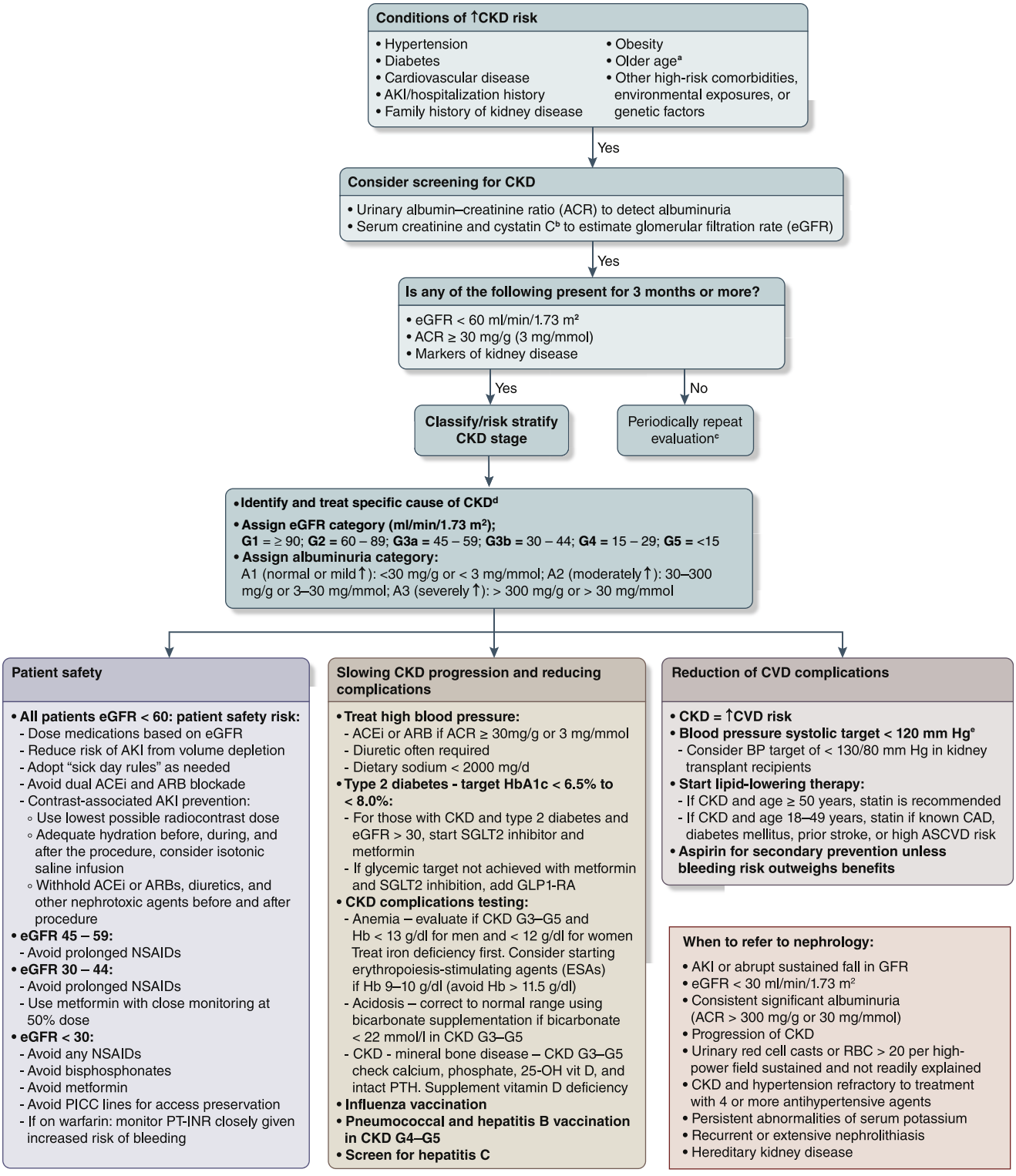

Figure 2 | Algorithm to screen, risk stratify, and treat chronic kidney disease (CKD). Based on Kidney Disease: Improving Global Outcomes (KDIGO) guidelines and adapted from Vassalotti *et al.*[54,55] [a]Care for older adults or those with limited life expectancy, including necessity of screening for CKD and treatment recommendations, should be individualized based on clinical status and patient preference. [b]If estimated glomerular filtration rate (eGFR)cr 45–59 ml/min per 1.73 m^2 or in individuals with low muscle mass, chronic illness, malnutrition, or other circumstances, order cystatin C to estimate glomeular filtration rate (GFR). [c]There are no current evidence-based recommendations regarding frequency of screening. Consider using risk equations (e.g., from CKD Prognosis Consortium) to estimate interval risk of developing CKD. [d]Cause of CKD is classified based on clinical evaluation. [e]Based on level 2 recommendation, which means a majority of people would adopt systolic blood pressure target $<$ 120 mm Hg while many others may opt for a less-intensive approach. As there is likely to be marked variability in how individual patients weigh and value the potential benefits and harms of intensive blood pressure (BP) control (continued)

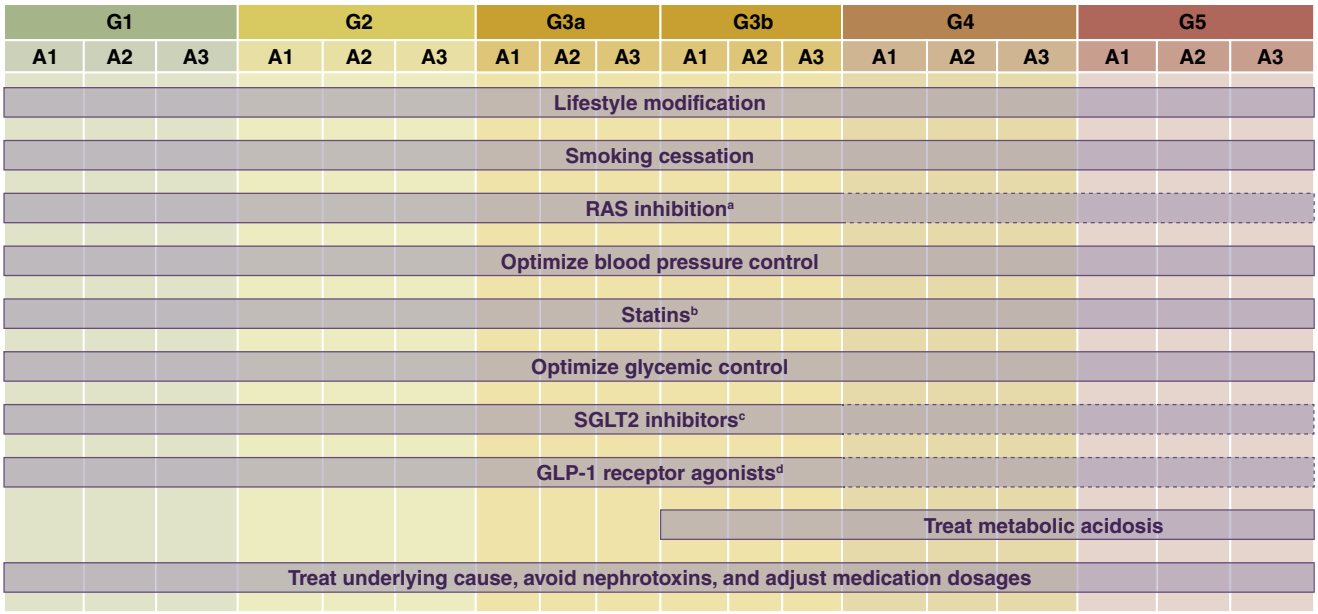

Figure 3 | Interventions to slow chronic kidney disease (CKD) progression and/or reduce cardiovascular risk. [a]Unclear if and when to discontinue renin-angiotensin system (RAS) inhibition in advanced CKD. [b]Statins should not be initiated for those beginning dialysis therapy. However, patients already receiving statins at the time of dialysis initiation can continue their statin treatment. [c]Applies to CKD patients with type 2 diabetes only. SGLT2 inhibitor is recommended as first-line treatment with metformin and may also have benefits in persons with CKD and no diabetes.[67,68] Sodium-glucose cotransporter-2 (SGLT2) inhibitors should be initiated if estimated glomerular filtration rate (eGFR) 30 ml/min per 1.73 m^2 and can be continued through G4–G5 until dialysis initiation, at which point SGLT2 inhibitor should be discontinued. There is no evidence for initiation of SGLT2 inhibitors if eGFR $<$ 30 ml/min per 1.73 m^2. [d]Applies to CKD patients with type 2 diabetes only. Glucagon-like peptide-1 (GLP-1) receptor agonist can be considered when SGLT2 inhibitor and/or metformin is not tolerated or glycemic target is not reached. Dulaglutide can be used if eGFR $>$ 15 ml/min per 1.73 m^2; exenatide can be used if creatinine clearance $>$ 30 ml/min; there are limited data for use of liraglutide, lixisenatide, or semaglutide in severe CKD. Consult dosing recommendations for use of these agents in CKD G4 and G5.

disease education via digital media and curates educational topics targeted to patients rather than providers.[83] Patient portals allowing direct access to providers and medical information can facilitate patient self-management and communication with providers.[84] As mobile technologies emerge to support health care delivery, a large portion of the CKD population may have difficulty accessing them because individuals with CKD are frequently older, may have lower socioeconomic status and health literacy,[82,84–86] and may have privacy concerns about disease labeling and health information.[85–89] Moving forward, the co-design of interventions with patients embedded in local communities will be important for ensuring utility and acceptability.

Health system and economic factors

The KDIGO 2012 guideline for the evaluation and management of CKD[25] was a major step toward defining high-quality CKD care. However, uptake in clinical practice has been suboptimal, with clinicians typically failing both to assess albuminuria for CKD diagnosis and staging[90] and to confirm the diagnosis of CKD in persons with $eGFR_{Cr}$ of 45–59 ml/min per 1.73 m^2 with measurements of cystatin C, where available.[91–94] Recognizing these large implementation gaps, the KDIGO conference participants devoted substantial effort towards considering effective implementation strategies and summarizing the barriers and facilitators for effective CKD care in the context of health system and economic factors.

Conclusion 10. CKD screening and treatment efforts require multi-stakeholder implementation strategies to overcome barriers to high-quality CKD care. The conference participants identified key strategies for CKD screening program development and success. Success of any CKD screening program will depend upon the engagement of health professionals in the local context, particularly primary care clinicians and allied health professionals, including front-line health workers. Patient-related barriers in CKD care include low patient knowledge of CKD and its associated risks and social risk factors, such as limited financial resources and low health

Figure 2 | (continued) and since this may vary with age, culture, number of drugs (both BP-lowering and other) and other factors, shared decision-making between individual patients and clinicians should be emphasized. ACEi, angiotensin converting enzyme inhibitor; ACR, albumin-creatinine ratio; AKI, acute kidney injury; ARB, angiotensin receptor blocker; ASCVD, atherosclerotic cardiovascular disease; CAD, coronary artery disease; CVD, cardiovascular disease; GLP-1 RA, glucagon-like peptide-1 receptor agonist; Hb, hemoglobin; NSAID, nonsteroidal anti-inflammatory drugs; PICC, peripherally inserted central catheter line; PT-INR, prothrombin time-international normalized ratio; PTH, parathyroid hormone; RBC, red blood cell; SGLT2, sodium-glucose cotransporter-2.

Table 3 | Health system–level approaches for improving early CKD identification and management

Framework for developing initiatives
Conceptual
• Understand patient flow through the health system and identify possible tactics for engagement
• Develop risk-based approaches to identifying and treating CKD
• Integrate novel program processes with existing health services and processes
○ Augment/strengthen in view of early CKD identification and intervention
• Actively engage with:
○ Patients
○ Health care clinicians
○ Health system administrators and policy makers
• Develop monitoring and improvement strategies
Practical
• Determine population for screening based on local risk factors
• Identify existing screening programs for other diseases, such as cardiovascular
• Assess whether there is necessary political commitment
• Specify available resources (workforce/material/funding)
• Develop a technical package, strategies using the best available evidence, for CKD screening and management
• Provide specific, actionable recommendations with level of evidence
• Develop targeted versions of the guideline summary aimed at patients and primary-care providers[103,104]
• Develop visually appealing infographics and apps to aid in knowledge translation
• Integrate guideline recommendations in laboratory information systems and electronic health records with clinical decision support
• Engage medical educators to teach guideline-based CKD care in medical schools
• Engage all stakeholders—professional societies (such as ISN), patients, payers, health systems, and disease-specific foundations—in dissemination strategies
• Establish governance for monitoring, evaluation, and improvement
Framework for continued advocacy and expansion of efforts
• Identify the full health and economic burdens of kidney diseases
• Establish kidney disease registries and use collected data to drive surveillance, feedback, and integration
• Collaborate with other guideline bodies and professional societies to maximize consistency in recommendations (e.g., primary care, cardiology, endocrinology, geriatrics)
• Develop evidence-based quality measures for CKD care
• Document real-world health and economic consequences of successful interventions and models of care
• Identify methods for sustainable financing of optimal services
• Generate and promote evidence linking health promotion to improved health and economic outcomes regarding kidney diseases
• Continue advocacy from researchers, clinicians, and policy makers for healthy environments and lives
• Focus investments and reforms to develop effective primary-care systems, including pharmaceuticals and behavioral interventions
• Invest in research to identify novel risk factors for kidney diseases
• Implement cost-effective strategies to target care to individuals at increased risk of kidney disease

CKD, chronic kidney disease; ISN, International Society of Nephrology.

literacy. Health system–related barriers include perceived lack of urgency for detecting early CKD among primary care clinicians, lack of knowledge of CKD guidelines, lack of incentives for CKD interventions, lack of CKD-specific clinical quality measures, and suboptimal communication between specialties.[79,95,96]

Several efforts can improve CKD management in the primary care setting. Identifying high-risk individuals for CKD screening will require that clinicians are educated about CKD risk factors. Effective CKD risk stratification will require education about CKD staging, particularly the importance of albuminuria.[48,49,96,97] To bridge the education gap, existing guidelines could be simplified with quick reference guides for primary care clinicians. Approaches for CKD screening, risk stratification, and treatment should be integrated with existing health services and processes.[98] For example, automated laboratory reporting[55,99] and the use of risk equations and clinical decision support tools could be embedded in existing electronic health records to guide the selection of individuals for testing and the frequency of repeat testing. Subsequently, the extensively validated Kidney Failure Risk Equation[24,100] could risk-stratify patients for appropriate referral to specialty care and frequency of follow-up, thereby improving efficiency.[101] Selection of high-risk individuals and populations should be tailored to local populations, who may face distinct biological, environmental, and social risk factors.

Several models of care could be used to optimize care for newly diagnosed patients with CKD. These include co-management of care between primary care and nephrology, nephrology specialist consultation for high-risk CKD patients, and team-based multidisciplinary care.[102] Interventions for CKD early identification and intervention should be integrated into existing workflows using quality-improvement principles or as effectiveness-implementation hybrid studies. Overarching principles for development include efficiency, equity, ethics, education, sustainability, scalability, cultural appropriateness, access, and quality (Table 3[103,104]). Interventions based upon detection and management of other chronic diseases could be applied or extended to include CKD screening, risk stratification, and treatment.[105]

Table 4 | Defining success for CKD screening, risk stratification, and treatment programs

Outcomes	Indicators
Process (pertaining to health systems, providers, or patients)	
	• Screening with the correct tests (eGFR and UACR) • Timely and appropriate follow-up testing • Dietary, exercise, and smoking cessation counseling • Clinician CKD awareness as measured by documentation • Patient adherence to treatment plan • Appropriate nephrology/kidney transplant referrals • Availability of essential medicines and testing
Patient-centered	
	• Patient awareness of and attitudes toward CKD diagnosis • Patient experience and satisfaction • CKD-specific knowledge • Trust in physician • Quality of life • Shared decision-making for modality choice, including kidney replacement therapy and conservative management
Intermediate clinical	
	• Blood pressure control • Glycemic control • Statin use • ACEi/ARB use • SGLT2 inhibitor use • Vaccinations • Management of CKD-specific complications • Drug dosing/adverse drug events
Clinical	
	• CKD progression—eGFR slope, 40% decline in eGFR, doubling of serum creatinine, kidney failure • Hospitalization or emergency department visits • Cardiovascular events • Acute kidney injury events • Emergency dialysis starts • Pre-emptive transplant rates • All-cause mortality

ACEi, angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; SGLT2, sodium-glucose cotransporter-2; UACR, urine albumin-to-creatinine ratio.

After implementation, CKD screening programs should be rigorously evaluated. There are a large variety of potential outcomes that can be studied for determining program success (Table 4). Categories of relevant endpoints include process measures (persons diagnosed with CKD, new treatments initiated), patient-centered measures (knowledge, awareness), intermediate clinical measures (blood pressure control, glycemic control), and clinical outcomes (cardiovascular events, kidney failure, and death). In addition, the resource impact and costs to the health system should be evaluated. Process, patient-centered, and intermediate clinical outcomes may be more feasible short- and medium-term targets than clinical endpoints to quantify the efficacy of CKD screening programs as they occur on shorter time horizons. In order to demonstrate an effect on clinical outcomes, studies would need to have large sample sizes and long durations of follow-up.

Conclusion 11. Financial and nonfinancial incentives need to be aligned toward CKD screening, risk stratification, and treatment. For health systems to be incentivized toward CKD early identification, payors would need to adopt a lifetime risk perspective, as the benefits of CKD screening and early intervention accrue slowly. Currently in many countries, physicians receive greater reimbursement for time spent treating patients with kidney failure than for an equal amount of time treating CKD at earlier stages.[106] To achieve success in CKD screening programs, resources need to be allocated toward prevention of kidney-related and cardiovascular adverse outcomes. CKD screening and treatment programs may benefit from utilizing financial incentives and/or penalties through pay-for-performance programs or alternative payment models. Tactics for reducing program costs include high-throughput testing, leveraging generic treatments including fixed-dose combinations, price negotiations, and strategic purchasing.

Conclusion 12. CKD screening in high-risk groups is likely to be cost-effective. The conference included experts in health economics who reviewed existing simulation models that evaluated the cost-effectiveness of CKD screening strategies and their impact on downstream health outcomes. Existing CKD detection and treatment simulation models indicate that screening high-risk individuals for CKD, including those with hypertension or diabetes, is likely to be cost-effective in high-income countries based upon thresholds of approximately $50,000-$150,000 USD per quality-adjusted life year.[30,107] However, it was emphasized that cost-effectiveness is not a binary assessment but rather represents a likelihood of cost-effectiveness at each $/quality-adjusted life year threshold, sensitive to the assumptions of the model. Important drivers of cost-effectiveness for CKD screening are the frequency of

testing, expected prevalence and incidence (testing yield) of CKD, and costs of confirmatory tests and treatment, among others.[30,108,109]

Several limitations of existing cost-effectiveness studies were identified and their correction suggested as future research directions.[109] First was the predominant focus on proteinuria, primarily dipstick proteinuria, which is insensitive for low levels of albuminuria. Second, there was very limited inclusion of cardiovascular outcomes, which are key complications of CKD that drive hospitalizations and mortality; most CKD patients are at much higher risk for cardiovascular events than progression to kidney failure. Third, few cost-effectiveness studies included LMICs. Fourth, models did not incorporate patient perspectives and patient-reported outcomes. Last, many models had outdated or incomplete evidence and potentially inaccurate assumptions.

The conference participants concluded that the appropriate time frame for simulating the effect of a CKD screening program is at minimum a decade and preferably a lifetime horizon. In addition, the health system perspective may undervalue the cost-effectiveness of CKD screening programs due to benefits in reducing disability and unemployment, so the societal perspective should also be considered. Projecting and monitoring cost-effectiveness may be done using high-quality natural history observational studies, clinical trial data on intervention effects, and local health system data on outcomes and costs.

Conclusion 13. CKD screening approaches may differ in LMIC countries. The conference members concluded that CKD screening approaches depend on the screening setting and available resources.[110] Managing the cost of early CKD screening programs is important for their successful implementation, especially in LMICs.[111,112] The conference participants recognized that the cost of measuring creatinine, cystatin C, and UACR in LMICs may be prohibitive.[113] In such settings, creatinine testing and dipstick screening for albuminuria were deemed to be an acceptable first step if followed by appropriate confirmatory testing. The conference participants called for universal availability of eGFR and albuminuria testing in primary care globally. The commitment and resources to treat persons who have newly detected CKD are critical components for any CKD screening program, as screening must be linked directly with risk-stratification and treatment capabilities.

In some countries, widespread testing for low eGFR by creatinine or cystatin C is already in place. In such settings, addition of a screening program designed to identify individuals with albuminuria but normal eGFR could identify additional high-risk individuals among whom treatments such as ACEi/ARBs may be beneficial. In resource-poor settings with elevated risk of proteinuric kidney disease, such as populations with prevalent *APOL1* high-risk variants, widespread screening with urine dipstick may be beneficial, ideally accompanied by eGFR measurement. Individuals identified by such screening approaches could then be evaluated with assessment of both eGFR (by creatinine, cystatin C, or both) and UACR to confirm the diagnosis of CKD and accurately risk stratify.

In communities with limited availability of primary care clinicians, community-based screening using nonphysician health care workers may be an option for CKD screening, risk stratification, and treatment. In some settings, treatments for late stages of CKD, such as dialysis or transplantation, may not be widely available, underscoring the importance of early CKD identification and intervention.[7]

Conclusions

The KDIGO Controversies Conference participants were unanimous that the bulk of evidence supports systematic approaches to screen for, risk stratify, and treat persons with CKD. Because interventions to slow CKD progression and reduce cardiovascular risk are evidence based and have been shown to improve outcomes, conference attendees agreed that the focus should be on strategies to maximize deployment of CKD screening, risk stratification, and treatment efforts. Ideally, these implementation efforts will be launched across multiple countries and myriad health systems. A worldwide effort will generate lessons learned and opportunities to share and disseminate strategies that are successful. Pragmatic trials should be designed to test CKD early-identification and intervention programs across various high-risk populations using different combinations of measures. Implementation efforts should engage policy makers, local clinicians, the community at large, and broader stakeholders in an iterative process. Ultimately, we call for large-scale randomized controlled trials to evaluate the effects of CKD screening, risk-stratification, and treatment programs compared with usual care on clinical endpoints. Implementing and evaluating systematic CKD screening efforts across large health care systems will build definitive evidence regarding their effectiveness to reduce morbidity and mortality from kidney disease for the global population.

APPENDIX

Other Conference Participants

Georgi Abraham, India; Zanfina Ademi, Australia; Radica Z. Alicic, USA; Ian H. de Boer, USA; Raj Deo, USA; Xiaoqiang Ding, China; Natalie Ebert, Germany; Kevin J. Fowler, USA; Linda F. Fried, USA; Ron T. Gansevoort, The Netherlands; Guillermo Garcia-Garcia, Mexico; Brenda R. Hemmelgarn, Canada; Jessica Lee Harding, USA; Joanna Q. Hudson, USA; Kunitoshi Iseki, Japan; Vasantha Jotwani, USA; Leah S. Karliner, USA; Andrew S. Levey, USA; Adrian Liew, Singapore; Peter J. Lin, Canada; Andrea O.Y. Luk, Hong Kong; Verónica Martínez, Mexico; Andrew E. Moran, USA; Mai Nguyen, USA; Gregorio T. Obrador, Mexico; Donal O'Donoghue, UK; Meda E. Pavkov, USA: Jessie Pavlinac, USA; Neil R. Powe, USA; Jesse C. Seegmiller, USA; Jenny I. Shen, USA; Rukshana Shroff, UK; Laura Solá, Uruguay; Maarten W. Taal, UK; James Tattersall, UK; Joseph A. Vassalotti, USA; Matthew R. Weir, USA; and Ella Zomer, Australia

DISCLOSURE

MGS declared having consultancy fees from Intercept Pharmaceuticals and University of Washington; stock equity from Cricket Health and TAI Diagnostics; research support from Booz Allen Hamilton; and future research support from Bayer U.S. SLT declared having consultancy fees from Bayer AG and research support from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). MEG declared having research support from the National

Kidney Foundation/National Institutes of Health (NIH). JHI declared having consultancy fees from AstraZeneca; future consultancy fees from Ardelyx; research support from NIDDK; and serving as Data and Safety Monitoring Board (DSMB) member for Sanifit. VJ declared having consultancy fees from AstraZeneca and Baxter Healthcare; and research support from Baxter Healthcare and GSK. A-PK declared having research support from European and Developing Countries Clinical Trials Partnership and future research support from the NIH. MM declared having consultancy fees from AstraZeneca and Bayer; future consultancy fees from AstraZeneca and Bayer; and research support from Fundación Gonzalo Río Arronte. NT declared having consultancy fees from AstraZeneca, Boehringer Ingelheim-Lilly, Janssen, and Otsuka; stock equity from Mesentech, Pulsedata, Rénibus, and Tricida; and research support from AstraZeneca, Janssen, Otsuka, and Tricida. MJ declared having consultancy fees from Amgen, AstraZeneca, Boehringer Ingelheim, Mundipharma, Merck Sharp & Dohme (MSD), and Vifor Fresenius Medical Care Renal Pharma; future consultancy fees from Astellas; speakers bureaus from Amgen, AstraZeneca, Menarini, Mundipharma, and MSD; research support from Amgen, MSD, and Otsuka; and future research support from AstraZeneca. WCW declared having consultancy fees from Akebia, AstraZeneca, Bayer, Janssen, Merck, Relypsa, and Vifor Fresenius Medical Care Renal Pharma; and research support from the NIH. All the other authors declared no competing interests.

ACKNOWLEDGMENTS

This conference was sponsored by KDIGO and supported in part by unrestricted educational grants from Akebia Therapeutics, AstraZeneca, Bayer HealthCare, Boehringer Ingelheim, Cricket Health, CVS Kidney Care, Fresenius Medical Care, Gentian, Janssen, Kyowa Kirin, Mundipharma, Reata, Relypsa, and Tricida. We thank Jennifer King, PhD, for assistance with manuscript preparation and Dr. Jennifer Lees for providing data from the UK Biobank Study. KDIGO also thanks the International Society of Nephrology (Dr. Aminu Bello and Sandrine Damster) and the International Society of Nephrology regional experts in the field for their participation in the CKD testing costs survey.

REFERENCES

1. Bikbov B, Purcell CA, Levey AS, et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2020;395:709–733.
2. Palmer Alves T, Lewis J. Racial differences in chronic kidney disease (CKD) and end-stage renal disease (ESRD) in the United States: a social and economic dilemma. *Clin Nephrol*. 2010;74(suppl 1):S72–S77.
3. Caskey FJ. Prevalence and incidence of renal disease in disadvantaged communities in Europe. *Clin Nephrol*. 2016;86:34–36.
4. Garcia M, Hernandez B, Ellington TG, et al. A lack of decline in major nontraumatic amputations in Texas: contemporary trends, risk factor associations, and impact of revascularization. *Diabetes Care*. 2019;42:1061–1066.
5. Crews DC, Bello AK, Saadi G, World Kidney Day Steering Committee. Burden, access, and disparities in kidney disease. *Kidney Int Rep*. 2019;4:372–379.
6. Curtis S, Komenda P. Screening for chronic kidney disease: moving toward more sustainable health care. *Curr Opin Nephrol Hypertens*. 2020;29:333–338.
7. Li PK, Garcia-Garcia G, Lui SF, et al., World Kidney Day Steering Committee. Kidney health for everyone everywhere-from prevention to detection and equitable access to care. *Kidney Int*. 2020;97:226–232.
8. Moyer VA. Screening for chronic kidney disease: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012;157:567–570.
9. Qaseem A, Hopkins RH Jr, Sweet DE, et al. Screening, monitoring, and treatment of stage 1 to 3 chronic kidney disease: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2013;159:835–847.
10. American Diabetes Association. 11. Microvascular complications and foot care: standards of medical care in diabetes—2020. *Diabetes Care*. 2020;43(suppl 1):S135–S151.
11. Wilson JMG, Jungner G, World Health Organization. Principles and practice of screening for disease. 1968. Available at: https://apps.who.int/iris/handle/10665/37650. Accessed October 22, 2019.
12. Andermann A, Blancquaert I, Beauchamp S, Dery V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ*. 2008;86:317–319.
13. Xie X, Liu Y, Perkovic V, et al. Renin-angiotensin system inhibitors and kidney and cardiovascular outcomes in patients with CKD: a bayesian network meta-analysis of randomized clinical trials. *Am J Kidney Dis*. 2016;67:728–741.
14. Herrington WG, Emberson J, Mihaylova B, et al. Impact of renal function on the effects of LDL cholesterol lowering with statin-based regimens: a meta-analysis of individual participant data from 28 randomised trials. *Lancet Diabetes Endocrinol*. 2016;4:829–839.
15. Wright JT Jr, Williamson JD, Whelton PK, et al. A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med*. 2015;373:2103–2116.
16. Ruospo M, Saglimbene VM, Palmer SC, et al. Glucose targets for preventing diabetic kidney disease and its progression. *Cochrane Database Syst Rev*. 2017;6:CD010137.
17. Coca SG, Ismail-Beigi F, Haq N, et al. Role of intensive glucose control in development of renal end points in type 2 diabetes mellitus: systematic review and meta-analysis intensive glucose control in type 2 diabetes. *Arch Intern Med*. 2012;172:761–769.
18. Perkovic V, Jardine MJ, Neal B, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *N Engl J Med*. 2019;380:2295–2306.
19. Neal B, Perkovic V, Mahaffey KW, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med*. 2017;377:644–657.
20. Tuot DS, Plantinga LC, Hsu CY, et al. Chronic kidney disease awareness among individuals with clinical markers of kidney dysfunction. *Clin J Am Soc Nephrol*. 2011;6:1838–1844.
21. White SL, Yu R, Craig JC, et al. Diagnostic accuracy of urine dipsticks for detection of albuminuria in the general community. *Am J Kidney Dis*. 2011;58:19–28.
22. Wright Nunes J, Roney M, Kerr E, et al. A diagnosis of chronic kidney disease: despite fears patients want to know early. *Clin Nephrol*. 2016;86:78–86.
23. Go AS, Chertow GM, Fan D, et al. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med*. 2004;351:1296–1305.
24. Tangri N, Grams ME, Levey AS, et al. Multinational assessment of accuracy of equations for predicting risk of kidney failure: a meta-analysis. *JAMA*. 2016;315:164–174.
25. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int Suppl*. 2013;3:1–150.
26. Wanner C, Tonelli M. KDIGO clinical practice guideline for lipid management in CKD: summary of recommendation statements and clinical approach to the patient. *Kidney Int*. 2014;85:1303–1309.
27. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol*. 2019;73:3168–3209.
28. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension*. 2018;71:1269–1324.
29. Boulware LE, Jaar BG, Tarver-Carr ME, et al. Screening for proteinuria in US adults: a cost-effectiveness analysis. *JAMA*. 2003;290:3101–3114.
30. Komenda P, Ferguson TW, Macdonald K, et al. Cost-effectiveness of primary screening for CKD: a systematic review. *Am J Kidney Dis*. 2014;63:789–797.
31. Wolf AMD, Fontham ETH, Church TR, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA Cancer J Clin*. 2018;68:250–281.
32. Grams ME, Chow EK, Segev DL, Coresh J. Lifetime incidence of CKD stages 3–5 in the United States. *Am J Kidney Dis*. 2013;62:245–252.
33. Nelson RG, Grams ME, Ballew SH, et al. Development of risk prediction equations for incident chronic kidney disease. *JAMA*. 2019;322:2104–2114.
34. Crews DC, Liu Y, Boulware LE. Disparities in the burden, outcomes, and care of chronic kidney disease. *Curr Opin Nephrol Hypertens*. 2014;23:298–305.

35. Morton RL, Schlackow I, Mihaylova B, et al. The impact of social disadvantage in moderate-to-severe chronic kidney disease: an equity-focused systematic review. *Nephrol Dial Transplant*. 2016;31:46–56.
36. Nicholas SB, Kalantar-Zadeh K, Norris KC. Socioeconomic disparities in chronic kidney disease. *Adv Chronic Kidney Dis*. 2015;22:6–15.
37. United States Renal Data System. 2018 USRDS annual data report: epidemiology of kidney disease in the United States. Bethesda, MD: National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health; 2018. Available at: https://www.usrds.org/annual-data-report/previous-adrs/.
38. Essue BM, Laba T-L, Knaul F, et al. Economic burden of chronic ill health and injuries for households in low-and middle-income countries. In: Jamison DT, Gelband H, Horton S, et al., eds. *Disease Control Priorities: Improving Health and Reducing Poverty*. 3rd ed. Washington, DC: International Bank for Reconstruction and Development/World Bank; 2017. chapter 6.
39. Zelnick LR, Weiss NS, Kestenbaum BR, et al. Diabetes and CKD in the United States population, 2009–2014. *Clin J Am Soc Nephrol*. 2017;12:1984–1990.
40. United States Renal Data System. *2019 USRDS annual data report: epidemiology of kidney disease in the United States*. Bethesda, MD: National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health; 2019. Available at: https://www.usrds.org/annual-data-report/.
41. Coresh J, Selvin E, Stevens LA, et al. Prevalence of chronic kidney disease in the United States. *JAMA*. 2007;298:2038–2047.
42. Hallan SI, Matsushita K, Sang Y, et al. Age and association of kidney measures with mortality and end-stage renal disease. *JAMA*. 2012;308:2349–2360.
43. Matsushita K, Ballew SH, Coresh J, et al. Measures of chronic kidney disease and risk of incident peripheral artery disease: a collaborative meta-analysis of individual participant data. *Lancet Diabetes Endocrinol*. 2017;5:718–728.
44. Delanaye P, Jager KJ, Bokenkamp A, et al. CKD: a call for an age-adapted definition. *J Am Soc Nephrol*. 2019;30:1785–1805.
45. Seidu S, Barrat J, Khunti K. Clinical update: the important role of dual kidney function testing (ACR and eGFR) in primary care: identification of risk and management in type 2 diabetes. *Prim Care Diabetes*. 2020;14:370–375.
46. Litvin CB, Hyer JM, Ornstein SM. Use of clinical decision support to improve primary care identification and management of chronic kidney disease (CKD). *J Am Board Fam Med*. 2016;29:604–612.
47. Perkins RM, Chang AR, Wood KE, et al. Incident chronic kidney disease: trends in management and outcomes. *Clin Kidney J*. 2016;9:432–437.
48. Abdel-Kader K, Greer RC, Boulware LE, Unruh ML. Primary care physicians' familiarity, beliefs, and perceived barriers to practice guidelines in non-diabetic CKD: a survey study. *BMC Nephrol*. 2014;15:64.
49. Stevens PE, O'Donoghue DJ, de Lusignan S, et al. Chronic kidney disease management in the United Kingdom: NEOERICA project results. *Kidney Int*. 2007;72:92–99.
50. Shlipak MG, Matsushita K, Arnlov J, et al. Cystatin C versus creatinine in determining risk based on kidney function. *N Engl J Med*. 2013;369:932–943.
51. Lees JS, Welsh CE, Celis-Morales CA, et al. Glomerular filtration rate by differing measures, albuminuria and prediction of cardiovascular disease, mortality and end-stage kidney disease. *Nat Med*. 2019;25:1753–1760.
52. Ebert N, Shlipak MG. Cystatin C is ready for clinical use. *Curr Opin Nephrol Hypertens*. 2020;29:591–598.
53. Ismail OZ, Bhayana V, Kadour M, et al. Improving the translation of novel biomarkers to clinical practice: the story of cystatin C implementation in Canada: a professional practice column. *Clin Biochem*. 2017;50:380–384.
54. Vassalotti JA, Centor R, Turner BJ, et al. Practical approach to detection and management of chronic kidney disease for the primary care clinician. *Am J Med*. 2016;129:153–162.
55. Vassalotti JA, Thavarajah S. Combined albuminuria and estimated GFR laboratory reporting impacts primary care management of CKD. *Kidney Med*. 2020;2:235–238.
56. Xia J, Wang L, Ma Z, et al. Cigarette smoking and chronic kidney disease in the general population: a systematic review and meta-analysis of prospective cohort studies. *Nephrol Dial Transplant*. 2017;32:475–487.
57. Thompson S, Wiebe N, Padwal RS, et al. The effect of exercise on blood pressure in chronic kidney disease: a systematic review and meta-analysis of randomized controlled trials. *PLoS One*. 2019;14:e0211032.
58. vanden Wyngaert K, van Craenenbroeck AH, van Biesen W, et al. The effects of aerobic exercise on eGFR, blood pressure and VO_2 peak in patients with chronic kidney disease stages 3–4: a systematic review and meta-analysis. *PLoS One*. 2018;13:e0203662.
59. Bach KE, Kelly JT, Palmer SC, et al. Healthy dietary patterns and incidence of CKD: a meta-analysis of cohort studies. *Clin J Am Soc Nephrol*. 2019;14:1441–1449.
60. Kelly JT, Palmer SC, Wai SN, et al. Healthy dietary patterns and risk of mortality and ESRD in CKD: a meta-analysis of cohort studies. *Clin J Am Soc Nephrol*. 2017;12:272–279.
61. Bress AP, Bellows BK, King JB, et al. Cost-effectiveness of intensive versus standard blood-pressure control. *N Engl J Med*. 2017;377:745–755.
62. Messow CM, Isles C. Meta-analysis of statins in chronic kidney disease: who benefits? *QJM*. 2017;110:493–500.
63. Sanguankeo A, Upala S, Cheungpasitporn W, Ungprasert P, Knight EL. Effects of statins on renal outcome in chronic kidney disease patients: a systematic review and meta-analysis. *PLoS One*. 2015;10:e0132970.
64. Schlackow I, Kent S, Herrington W, et al. Cost-effectiveness of lipid lowering with statins and ezetimibe in chronic kidney disease. *Kidney Int*. 2019;96:170–179.
65. Zoungas S, Arima H, Gerstein HC, et al. Effects of intensive glucose control on microvascular outcomes in patients with type 2 diabetes: a meta-analysis of individual participant data from randomised controlled trials. *Lancet Diabetes Endocrinol*. 2017;5:431–437.
66. Neuen BL, Ohkuma T, Neal B, et al. Effect of canagliflozin on renal and cardiovascular outcomes across different levels of albuminuria: data from the CANVAS program. *J Am Soc Nephrol*. 2019;30:2229–2242.
67. Heerspink HJL, Stefánsson BV, Correa-Rotter R. DAPA-CKD Trial Committees and Investigators. Dapagliflozin in patients with chronic kidney disease. *N Engl J Med*. 2020;383:1436–1446.
68. Wheeler DC, Stefansson BV, Jongs N, et al. Effects of dapagliflozin on major adverse kidney and cardiovascular events in patients with diabetic and non-diabetic chronic kidney disease: a prespecified analysis from the DAPA-CKD trial. *Lancet Diabetes Endocrinol*. 2021;9:22–31.
69. Marso SP, Daniels GH, Brown-Frandsen K, et al. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med*. 2016;375:311–322.
70. Mann JF, Ørsted DD, Brown-Frandsen K, et al. Liraglutide and renal outcomes in type 2 diabetes. *N Engl J Med*. 2017;377:839–848.
71. Tuttle KR, Lakshmanan MC, Rayner B, et al. Dulaglutide versus insulin glargine in patients with type 2 diabetes and moderate-to-severe chronic kidney disease (AWARD-7): a multicentre, open-label, randomised trial. *Lancet Diabetes Endocrinol*. 2018;6:605–617.
72. Raphael KL, Isakova T, Ix JH, et al. A randomized trial comparing the safety, adherence, and pharmacodynamics profiles of two doses of sodium bicarbonate in CKD: the BASE pilot trial. *J Am Soc Nephrol*. 2020;31:161–174.
73. Novick TK, Surapaneni A, Shin JI, et al. Prevalence of opioid, gabapentinoid, and NSAID use in patients with CKD. *Clin J Am Soc Nephrol*. 2018;13:1886–1888.
74. Muanda FT, Weir MA, Bathini L, et al. Association of baclofen with encephalopathy in patients with chronic kidney disease. *JAMA*. 2019;322:1987–1995.
75. Lopez-Vargas PA, Tong A, Howell M, Craig JC. Educational interventions for patients with CKD: a systematic review. *Am J Kidney Dis*. 2016;68:353–370.
76. Narva AS, Norton JM, Boulware LE. Educating patients about CKD: the path to self-management and patient-centered care. *Clin J Am Soc Nephrol*. 2016;11:694–703.
77. Kramer H, Jimenez EY, Brommage D, et al. Medical nutrition therapy for patients with non-dialysis-dependent chronic kidney disease: barriers and solutions. *J Acad Nutr Diet*. 2018;118:1958–1965.
78. Teasdale EJ, Leydon G, Fraser S, et al. Patients' experiences after CKD diagnosis: a meta-ethnographic study and systematic review. *Am J Kidney Dis*. 2017;70:656–665.
79. Sperati CJ, Soman S, Agrawal V, et al. Primary care physicians' perceptions of barriers and facilitators to management of chronic kidney disease: a mixed methods study. *PLoS One*. 2019;14:e0221325.
80. Narva AS, Briggs M. The National Kidney Disease Education Program: improving understanding, detection, and management of CKD. *Am J Kidney Dis*. 2009;53(3 suppl 3):S115–S120.

81. Kidney Health Initiative. Technology roadmap for innovative approaches to renal replacement therapy. October 2018. Available at: https://www.asn-online.org/g/blast/files/KHI_RRT_Roadmap1.0_FINAL_102318_web.pdf. Accessed December 10, 2020.
82. Diamantidis CJ, Becker S. Health information technology (IT) to improve the care of patients with chronic kidney disease (CKD). *BMC Nephrol*. 2014;15:7.
83. National Kidney Disease Education Program. Available at: http://www.nkdep.nih.gov. Accessed September 16, 2020.
84. Tuot DS, Boulware LE. Telehealth applications to enhance CKD knowledge and awareness among patients and providers. *Adv Chronic Kidney Dis*. 2017;24:39–45.
85. Singh K, Diamantidis CJ, Ramani S, et al. Patients' and nephrologists' evaluation of patient-facing smartphone apps for CKD. *Clin J Am Soc Nephrol*. 2019;14:523–529.
86. Siddique AB, Krebs M, Alvarez S, et al. Mobile apps for the care management of chronic kidney and end-stage renal diseases: Systematic search in app stores and evaluation. *JMIR Mhealth Uhealth*. 2019;7:e12604.
87. Lewis RA, Lunney M, Chong C, Tonelli M. Identifying mobile applications aimed at self-management in people with chronic kidney disease. *Can J Kidney Health Dis*. 2019;6, 2054358119834283.
88. Lambert K, Mullan J, Mansfield K, Owen P. Should we recommend renal diet-related apps to our patients? An evaluation of the quality and health literacy demand of renal diet-related mobile applications. *J Ren Nutr*. 2017;27:430–438.
89. Rogers D. Patient perspective of smartphone-based apps for CKD self-care. *Clin J Am Soc Nephrol*. 2019;14:483–484.
90. Tuttle KR, Alicic RZ, Duru OK, et al. Clinical characteristics of and risk factors for chronic kidney disease among adults and children: an analysis of the CURE-CKD registry. *JAMA Netw Open*. 2019;2:e1918169.
91. Fraser SD, Parkes J, Culliford D, et al. Timeliness in chronic kidney disease and albuminuria identification: a retrospective cohort study. *BMC Fam Pract*. 2015;16:18.
92. Ravera M, Noberasco G, Weiss U, et al. CKD awareness and blood pressure control in the primary care hypertensive population. *Am J Kidney Dis*. 2011;57:71–77.
93. Razavian M, Heeley EL, Perkovic V, et al. Cardiovascular risk management in chronic kidney disease in general practice (the AusHEART study). *Nephrol Dial Transplant*. 2012;27:1396–1402.
94. Nash DM, Brimble S, Markle-Reid M, et al. Quality of care for patients with chronic kidney disease in the primary care setting: a retrospective cohort study from Ontario, Canada. *Can J Kidney Health Dis*. 2017;4, 2054358117703059.
95. Boulware LE, Carson KA, Troll MU, Powe NR, Cooper LA. Perceived susceptibility to chronic kidney disease among high-risk patients seen in primary care practices. *J Gen Intern Med*. 2009;24:1123–1129.
96. Boulware LE, Troll MU, Jaar BG, Myers DI, Powe NR. Identification and referral of patients with progressive CKD: a national study. *Am J Kidney Dis*. 2006;48:192–204.
97. Greer RC, Liu Y, Cavanaugh K, et al. Primary care physicians' perceived barriers to nephrology referral and co-management of patients with CKD: a qualitative study. *J Gen Intern Med*. 2019;34:1228–1235.
98. Tsang JY, Blakeman T, Hegarty J, et al. Understanding the implementation of interventions to improve the management of chronic kidney disease in primary care: a rapid realist review. *Implement Sci*. 2016;11:47.
99. Hallan SI, Rifkin DE, Potok OA, et al. Physicians' recognition and management of kidney disease: a randomized vignette study evaluating the impact of the KDIGO 2012 CKD classification system. *Kidney Med*. 2020;2:258–266.
100. Tangri N, Stevens LA, Griffith J, et al. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA*. 2011;305:1553–1559.
101. Smekal MD, Tam-Tham H, Finlay J, et al. Patient and provider experience and perspectives of a risk-based approach to multidisciplinary chronic kidney disease care: a mixed methods study. *BMC Nephrol*. 2019;20:110.
102. Reilly R, Evans K, Gomersall J, et al. Effectiveness, cost effectiveness, acceptability and implementation barriers/enablers of chronic kidney disease management programs for indigenous people in Australia, New Zealand and Canada: a systematic review of mixed evidence. *BMC Health Serv Res*. 2016;16:119.
103. Flodgren G, Hall AM, Goulding L, et al. Tools developed and disseminated by guideline producers to promote the uptake of their guidelines. *Cochrane Database Syst Rev*. 2016;8:CD010669.
104. Gagliardi AR, Alhabib S, members of Guidelines International Network Implementation Working G. Trends in guideline implementation: a scoping systematic review. *Implement Sci*. 2015;10:54.
105. Mytton OT, Jackson C, Steinacher A, et al. The current and potential health benefits of the National Health Service Health Check cardiovascular disease prevention programme in England: a microsimulation study. *PLoS Med*. 2018;15:e1002517.
106. Centers for Medicare and Medicaid Services, Health and Human Services. Medicare PROGRAM; revisions to payment policies under the physician fee schedule and other revisions to Part B for CY 2018; Medicare Shared Savings Program requirements; and Medicare Diabetes Prevention Program. Final rule. *Fed Regist*. 2017;82(219):52976–53371.
107. Farmer AJ, Stevens R, Hirst J, et al. Optimal strategies for identifying kidney disease in diabetes: properties of screening tests, progression of renal dysfunction and impact of treatment—systematic review and modelling of progression and cost-effectiveness. *Health Technol Assess*. 2014;18:1–128.
108. Yarnoff BO, Hoerger TJ, Simpson SK, et al. The cost-effectiveness of using chronic kidney disease risk scores to screen for early-stage chronic kidney disease. *BMC Nephrol*. 2017;18:85.
109. Sutton AJ, Breheny K, Deeks J, et al. Methods used in economic evaluations of chronic kidney disease testing—a systematic review. *PLoS One*. 2015;10:e0140063.
110. Lunney M, Alrukhaimi M, Ashuntantang GE, et al. Guidelines, policies, and barriers to kidney care: findings from a global survey. *Kidney Int Suppl*. 2018;8:30–40.
111. Jha V, Arici M, Collins AJ, et al. Understanding kidney care needs and implementation strategies in low- and middle-income countries: conclusions from a "Kidney Disease: Improving Global Outcomes" (KDIGO) Controversies Conference. *Kidney Int*. 2016;90:1164–1174.
112. George C, Mogueo A, Okpechi I, et al. Chronic kidney disease in low-income to middle-income countries: the case for increased screening. *BMJ Glob Health*. 2017;2:e000256.
113. Bello AK, Levin A, Tonelli M, et al. Assessment of global kidney health care status. *JAMA*. 2017;317:1864–1881.

International Journal of Multimedia and Ubiquitous Engineering
Vol.12, No.8 (2017), pp.23-36
http://dx.doi.org/10.14257/ijmue.2017.12.8.03

Risk Level Prediction of Chronic Kidney Disease Using Neuro-Fuzzy and Hierarchical Clustering Algorithm (s)

Kerina Blessmore Chimwayi, Noorie Haris, Ronnie D. Caytiles*
and N. Ch. S. N Iyengar**

School of Computer Science and Engineering, VIT University, Vellore, T.N., India,
** Multimedia Engineering department, Hannam University, Daejeon, Korea*
***Professor, Department of Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, India*
kchimwayi@gmail.com, noorienittoor@gmail.com, rdcaytiles@gmail.com, srimannarayanach@sreenidhi.edu.in

Abstract

Chronic Kidney Disease (CKD) is usually characterized by a gradual loss of the functioning which the kidney does over time due to various factors. Early prediction and treatment save the kidney and halts the progress of CKD. CKD disease is being viewed as global public health issue for the past decade. The greatest threat for this deadly disease is developing countries where getting therapy is very expensive. The importance of predicting individuals who are at risk of CKD as well as applying clustering techniques cannot be underestimated since these can modify the progression of the disease. Identifying the silent killer disease early offers best opportunities for implementing possible strategies for lessening the probability of kidney loss. Neuro-fuzzy algorithm is applied to determine the risk of CKD in patients. Predictions done using neuro-fuzzy gave an accuracy of 97 percent. Using selected features, prediction for CKD disease is done so as to identify the risk. The results of the prediction are clustered to identify the percentage of patients with a high risk of having kidney disease who have a higher probability of being diabetic. Using hierarchical clustering three clusters formed show that there is a strong relationship between chronic kidney and diabetes.

Keywords*: Nuero-fuzzy, CKD, Diabetes, Clustering, ANFIS, Random Forest*

1. Introduction

In this century, healthcare big data analytics is inevitable due to the huge amounts of data that has variety, velocity, veracity and volume as their core characteristic. This healthcare data is being collected each second from various healthcare organizations all over the world. Utilizing big data for healthcare analytics offers great benefits to all stakeholders in healthcare which include healthcare providers, payers, patients and management. With terabytes of health data information being generated, especially through various electronic health records of huge hospitals, great insight about high risk disease predictions and much better ways of diagnosing them are acquired. One problem in the healthcare domain is the CKD that the world faces now. Huge amount of data is being generated related to CKD, from which various insights can be drawn for an effective solution through decision making among the health care sectors.

CKD is damage to kidneys due to factors related to lifestyle changes. Nowadays due to exposure to environmental changes, there are changes even observed in health which many do not recognize due to busy lifestyles. CKD can be caused due to lack of water consumption, smoking, improper diet, loss of sleep and many other factors. Researches also state that being diabetic is a highest risk which by time leads to kidney failure. CKD is unique in its nature among most diseases since it is mostly discovered when it is in the

ISSN: 1975-0080 IJMUE

final stages of progression whereby it will be much risky as well as expensive to treat due to being in the final stage called kidney failure.

In the domain of healthcare, this paper aims in building a model for risk level prediction in CKD considering all of the symptoms and causes contributing to it. The symptoms are the attributes that will define different stages of kidney diseases. Based on the different stages, one can classify a set of patient records to identify to which class of kidney disease a patient may belong to. On classifying patients, it results in easy recognition of the dominant attributes of CKD. Certain solutions can be provided with respect to the dominant attributes to avoid progression of CKD.

To construct a model on risk prediction of kidney disease, various machine learning techniques can be applied and then their performance can be compared with respect to accuracy, specificity and sensitivity of the models. Before application of any machine learning technique, there is a need of doing feature selection to understand the dominant attributes. Feature selection method called random forest is used to achieve selection of dominant attributes. This paper is mainly concerned with the use of machine learning techniques namely neuro-fuzzy systems and clustering which is termed unsupervised learning. These techniques will predict the class of kidney disease to which a patient can belong to, based on the symptoms. Finally, a risk analysis can be made based on all kinds of machine learning techniques applied.

2. Literature Survey

Dhifaf Azeez *et al.*, [1] proposed a model which does categorize patients who are in the emergency department by developing an intelligent triage system in an emergency department. Triage system is where there are a number of patients waiting for the treatment because of disease outbreaks. Triage system is integrated to the neural network system and also to the neuro-fuzzy system that can classify by capturing the signs of the patients and their appearance. This will reduce the time for examining the emergency of patients staying in the queue. Proposed work is used to examine the feasibility of ANFIS which is not explored previously in categorizing the patients.

Medical data was extracted from the Objective Primary Triage Scale (OPTS) data sheet from the Emergency Department that consists of categorical, text and continuous data. It was found from the accuracy results that ANN worked better in comparison to ANFIS for triage prediction. It has been concluded that ANN fits the output better when compared with ANFIS for new unseen data.

M. Zhamim Hossain *et al.*, [2] proposed a big data healthcare framework through the use of voice pathology assessment (VPA) as a case study. In this system, two features which include MPEG-7 low-level audio as well as the interlaced derivative pattern have been used for processing voice and speech signals. Support vector machines and a Gaussian model are used as the classifiers. This worked also includes VPA system which shows its efficiency as far as accuracy and time requirements are considered.

With the new era of the internet of things (IOT), electronic health records, genomic data, behavioral data and public health data are generating huge amounts of data which are rich with hidden patterns in which big data health analytics can be exploited to bring evidence and insights which will reduce the average cost of healthcare as well as yield improved outcomes through smarter decisions.

Dr. S. Vijayarani and Mr.S.Dhayanand [3] predicted the type of chronic kidney disease into four different types of diseases using MATLAB to implement ANN and SVM. The major aim was to compare the performance of these techniques in predicting CKD in terms of accuracy and performance. ANN networks proved to be superior to SVM machines in terms of accuracy.

Manish Kumar [4] in his work predicted the risk of chronic kidney disease by comparing numerous algorithms which were implemented using WEKA tool. The

researcher concentrated on applying various algorithms using WEKA tool then comparing these with the numerous methodologies applied. The researcher further used tenfold cross validation to classify each classifier. The only methodology that was detailed in explanation is that of Random forest for prediction yet the conclusion and analysis went on to discuss about other techniques that were not on the waiting list.

K. R. Anantha Padmanaban and G. Parthiban [5] in their work had an aim of predicting the early detection of chronic kidney disease also known as chronic renal disease for diabetic patients with the help of machine learning methods. In their research they suggested a decision tree to arrive at concrete results with desirable accuracy by measuring its performance to its specification and sensitiveness. The tool that was used with the diabetes dataset is WEKA. Naïve Bayes was compared with decision trees in predicting the risk of CKD in diabetic patients.

Adler Perotte *et al.* [6] implemented risk prediction on CKD progression with by using unrelated electronic health records accompanied with time series analysis. Here the researchers focused on using heterogeneous data sources for calculating the risk of progression from stage 3 to 5. The paper is more statistical at the same time no design was implemented for the paper.

Pushpa M. Patil [7] wrote a review paper on prediction of chronic kidney disease. From their review majority of researchers implemented the work using WEKA tool. Whilst the majority of researchers in trying to find the best solution, implemented the project using WEKA, some used MATLAB tool. Studies prove and show that for the prediction and diagnosis of this deadly CKD, these two tools were used in comparison to R which is good for data analysis and easy integration with big data technologies such as Mongo and Hadoop. In the review methodologies highlighted include Multilayer Perceptron, Logistic Regression, and Support Vector Machines (SVM) among other techniques.

Neha Sharma, Er and Rohit Kumar Verma [8] aimed at improving the prediction of kidney disease in old ages by using MATLAB to implement neuro-fuzzy algorithm which they said is better than existing probabilistic methods. In this paper they state that since this area is dealing with patient life, the use of if then rules reduces accuracy as compared to the use of mathematical models which they proposed. The researchers used the ANFIS (Adaptive Neuro-Fuzzy Inference Systems) model for implementation in MATLAB.

Abushariah, M.A.M *et al.* [9] Worked on research whereby they emphasized on designing an Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and ANFIS Approaches. The project was implemented using MATLAB with a dataset from UCI.

Abeer Y. Al-Hyari *et al.* [10] worked on a system which diagnosed renal failure, the main aim being to make it easy to identify CDK. In their research they compared three techniques which include artificial neural networks (ANN), Naïve Bayes and Decision trees using WEKA tool.

Naganna Chetty *et al.*, [11] in their research, the Role of Attributes Selection in Classification of Chronic Kidney Disease Patients, classification models with different classification algorithms which included wrapper subset and best first search method were explored in order to predict and classify CKD and non CKD patients. From their results it was evident that they obtained better accuracy on a dataset where they applied feature selection on the dataset as compared to when they did not apply any attribute selection mechanism.

A.Q. Ansari and Neeraj Kumar Gupta [12] proposed a computationally intelligent system which integrated neural networks and fuzzy logic for predicting the risk of Coronary Heart Disease (CHD). So as to automate CKD diagnosis, a simulation was done. Results of the research suggested that hybrid systems are suitable for identifying high or low cardiac risk.

From the analysis and review of work being done with respect to healthcare domain, many machine learning techniques are being implemented to predict possibility of diseases. This shortcoming will be tackled by a different approaches whereby the researchers propose to implement ANFIS for prediction of CKD and risk determination then clustering the output of the prediction to identify meaningful clusters of diabetic patients from the CDK outcome. Identification of relationships and clusters within high risk CKD patients to other chronic diseases such as sugar is therefore important. This is due to the fact that CKD is highly associated with a higher risk of complications which are related to those diseases.

3. Algorithms

In clustering a group of objects that belong to the same class is termed as a cluster. Similar objects are grouped in one cluster and in another cluster contains the dissimilar objects. The process of clustering is to create classes of similar objects by abstracting objects into a group. Analysis of cluster involves partitioning dataset into groups based on similarity among them followed by assigning labels to the groups formed. Clustering is adaptable to changes and helps derive useful features that differentiates groups.

A fusion of neural networks with fuzzy logic is generally defined as a system trained using a particular learning algorithm which is derived from neural network foundations. Learning in this case is done on local information whereas modifications are done locally. This approach combine artificial neural networks (ANN) with Fuzzy Rule Based Systems (FRBS). FRBS is laid upon the structure of ANN and the learning algorithm is used to adapt FRBS parameters which can be membership functions. Adaptive Neural Fuzzy Inference System (ANFIS) method for implementing neuro-fuzzy on a medical charges dataset is considered.

3.1. Neuro-fuzzy

Neuro-fuzzy is a popular algorithm in the field of artificial intelligence that can be used in predictions of certain data in every domain that we work in. ANFIS works on the basis of training the neural network which is tuned by the membership function parameters on integrating with the fuzzy system. Neuro-adaptive learning techniques have been used to tune membership function parameters. In general it can be said that ANFIS is similar to that of fuzzy inference system in terms of functionality. Sugeno and Tsukamoto models are represented by ANFIS. Hybrid learning algorithm is adopted by ANFIS.

The assumption is that the FIS has got two inputs which are a and b and an output c. On considering a model called sugeno fuzzy model, it will generate the following two rules:

First Rule:

If a is R_1 and b is S_1 then $g1=x_1a+y_1 b+z_1$.

Second Rule:

If a is R_2 and b is Q_2 then $g1=x_2a+y_2 b+z_2$.

Each of the working of the layers in ANFIS architecture can be explained as follows:

First Layer :

- $Z_{L,j}$ which is output from the jth node of a layer named L.
- Here all the nodes are called adaptive and they have a node function associated to them

$Z_{L,j} = \mu P_i$ (a) for j = 1, 2, or

$Z_{L,j} = \mu Q_{i-2}$ (a) for j = 3, 4

- a (or b) are the input nodes j and Pj (or Qj−2) are linguistic labels associated with this node.
- $Z_{L,j}$ is the membership function of fuzzy set (P1, P2, Q1, Q2).

$$\text{Membership function } \mu(a) = \frac{1}{1+\left|\frac{a-r_j}{p_j}\right|^{2q_j}} \quad (1)$$

P_j,q_j and r_i are the premise parameters in this case.

Second Layer:

- Here we have all nodes that are fixed and have a label which is named p.
- Output receives a product of the input for all the coming si. Product of all the incoming signals are forwarded to the output.

$Z_{2,j} = W_j = \mu P_j$ (a). μQ_j (b) for j = 1, 2

- All nodes resemble and carry the firing strength of each rule.
- T-norm operator can be used like AND operator.

Third Layer:

- In this layer we have fixed nodes that are labelled as Norm.
- These node calculate the ratio of the j^{th} rule's firing strength in line with the sum of all rules' firing strengths.

$$Z_{3,j} = \overline{W_j} = \frac{w_j}{w_1+w_2}, \; j = 1, 2 \quad (2)$$

- Normalized firing strengths represents the output.

Fourth Layer:

- All nodes belonging to this layer are adaptive nodes associated with a node function:

$Z_{4,j} = W_j G_i = W_j(x_j a + y_j b + z_j)$

- Layer 3 produces the normalized firing strength, W_j .
- The parameter set of node is $\{x_j , y_j , z_j\}$ and these termed as consequents.

Fifth Layer:

- Here we have a node which is single and also fixed with a S. All the incoming signals are given to output and they are given considering that they are summations of all signals.

$$\text{Output of layer 5} = Z_{5,j} = \sum_j \overline{W_j} g_j = \frac{\sum_j w_j g_j}{\sum_j w_j} \quad (3)$$

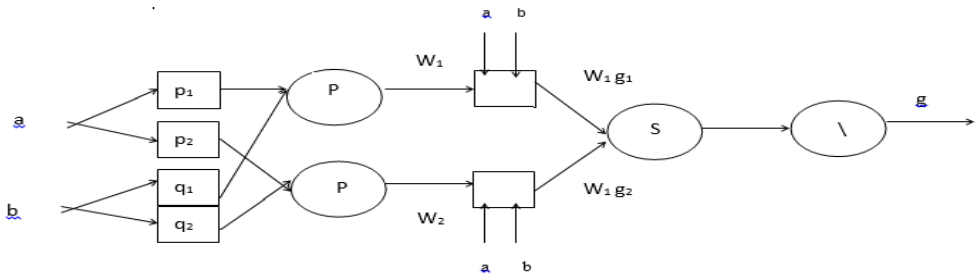

Figure 1. Neuro-Fuzzy Architecture

3.2. Hierarchical Clustering

In this type of analysis, each case of the CKD dataset starts as its own cluster. The clusters are combined on at a time till a single cluster is formed.

In this case Y = {y1, y2, y3, ..., y n} are CKD data points.

1) The first step is a disjoint clustering having level as zero and having a sequence n = 0 number.
2) Second step involves finding the minimum distance between pair for clusters within the current cluster, for example pair (e), (f), according to the distance d [(e),

(f)] = min d [(i), (j)] and in the current cluster the minimum distance is of all is considered.

3) In this stage sequence number is incremented as: n = n +1. Upon finishing this step clusters (e) and (f) are merged into a solitary cluster forming the next clustering n. The level of clustering is established to L (n) = d [(e), (f)].
4) For CKD dataset the matrix for distance is updated, and it is represented by D, whereby it is done deleting columns and the rows belonging to clusters (e) and (f) and at the same time a row is added as well as the column which corresponds to the new cluster. Distance between the new cluster, denoted (e, f) and old cluster (l) is therefore given by the following equation: d [(l), (e, f)] = min (d [(l), (e)], d [(l), (f)]).
5) Repeat the steps 2 to 4 till all clusters have been merged to form one cluster of patients else stop.

3.3. Random Forest

Random forest (RF) is a well-known ensemble learning method which is being applied in different fields which include high-dimensional classification and pattern recognition. When using RF many single individual decision trees will be created utilizing classification as well as regression algorithms. CART as an example of rule based systems splits a node by recursive partitioning process with regard to the no and yes answer which is from predictors. The class purity is maximized at each step of rule generation within the two resulting subsets. Based on the independent rules each subset is then further split. To measure the data partition's impurity CART uses Gini index. Maximizing the difference of heterogeneity is the main aim of CART, but it usually stumbles when real world datasets are presented dealing whereby high error of prediction comes as a result of over fitting. Bagging is a technique used in RF which provide a solution by deriving classifiers for data with high dimensions in a faster way. Through voting within the ensemble, accuracy of decision for classification has to be obtained from the individual classifiers.

4. Problem Definition

It is noted that CKD has been a global health concern therefore there is greater need to understand the relationship between CKD and other diseases through prediction of risk and clustering the results since it will play a greater role in developing public health policies that improve outcomes. Fig 1 below showing the relationship of CKD to chronic diseases such as diabetes. Among patients with chronic diseases there is an increased risk of complication related to those diseases.

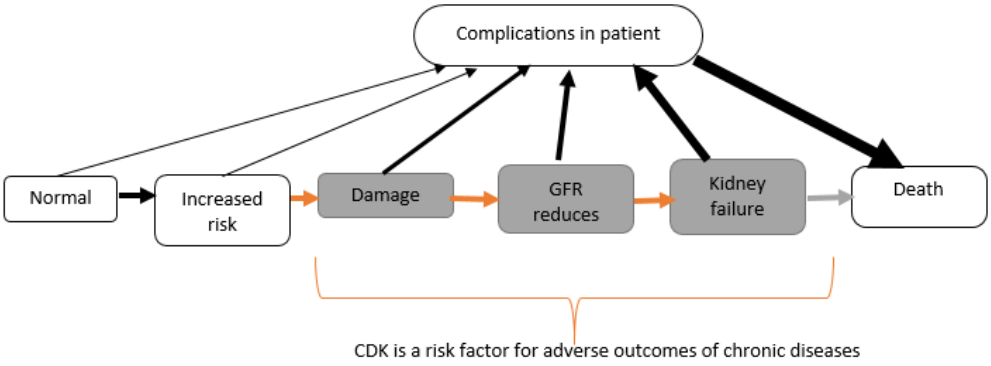

Figure 2. Relationship of CKD with Other Diseases

5. Design Methodology

The figure below resemble the design of the ANFIS system for the prediction, risk identification and clustering for the chronic kidney disease dataset.

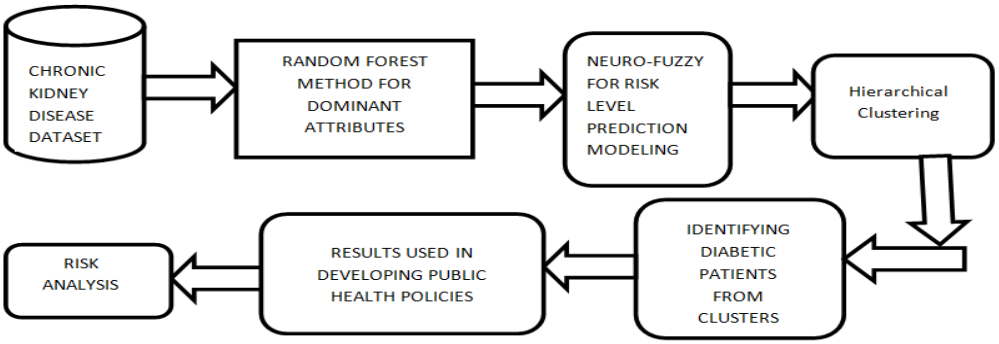

Figure 3. Methodology Design

6. Implementation Procedure

Using big data healthcare analytics, we can utilize huge amounts of patients' data so as to contribute in providing the right intervention to the right patients' at the most appropriate time. Proposed work consists of extracting essential attributes using random forest feature selection method. Following a thorough feature selection the data will be used for risk level prediction of kidney diseases using ANFIS. The applied techniques will predict the classification of patients based on different stages of kidney diseases with respect to the symptoms. The results from implementing ANFIS will be clustered using hierarchical clustering to identify valuable clusters.

The dataset for CKD is obtained from UCI Repository having 400 instances and 25 attributes on which the above mentioned techniques are to be applied. R tool will be used for the implementation of risk level prediction of CKD.

6.1. Dataset Collection

Upon definition of the problem, the appropriate dataset was chosen. CKD dataset has got a lot of useful variables which are key and necessary for the identification of the disease in patients. The dataset was initially in Attribute-Relation File Format (ARFF) which is the format for WEKA and was hence transformed into a suitable comma separated value format for use in R. Each and every feature in the Table 1 has a contribution to the class label being CKD or NOTCKD.

Table 1. Dataset Attributes

Attribute	Description
Age in years	Number of years
Blood pressure of patient	Given in mm/Hg
Specificity Gravity	Ranges from 1005 to 10025 (the higher the risk)
Albumin	Range is 0 to 5 (the higher the better)
Sugar level	5 levels indicating severity

Red Blood Cells	Is abnormal or normal
Pus Cell	Is normal or not normal (high number lead to urinary tract)
Pus Cell clumps	Can be present or not present
Bacteria	Can be present or not present
Blood Glucose	It is in mgs/dl
Blood urea	It is in mgs/dl
Serum creatinine	High level is not good
Sodium	It is measured in mEq/L
Potassium	It is measured in mEq/L
Haemoglobin	Less than 15 is kidney failure
Packed Cell Volume	This is numerical
White Blood Cell Count	This is numerical cell count
Red Blood Cell Count	Should not be higher or less than normal
Hypertension	It is categorical (yes or no)
Diabetes (mellitus)	It is categorical (yes or no)
Artery disease (Coronary)	It is categorical (yes or no)
Appetite	Is it poor or good (yes or no)
Pedal Edema	It is categorical (yes or no)
Anemia	It is categorical (yes or no)
Class	Given as ckd or notckd

6.2. Data Pre-Processing and Feature Selection

To obtain accurate results which are not misleading, sufficient data pre-processing should be done. Many healthcare datasets usually have missing values, noisy and inconsistent data. All these contribute to the quality of data, low quality data results in extremely poor machine learning results. The data was cleaned and missing values were replaced with column means.

The data under consideration initially had 25 attributes (14 nominal and 11 numeric) describing early stages of CKD in Indians. Of the 400 instances of the dataset, 150 are not having CKD and 250 are having CKD.

Random Forest (RF) feature selection is implemented to give a reduced feature set for model development. As shown in Figure 4 below, RF represents the importance of each feature by using mean decrease accuracy and gini. Accuracy in this case is measured in terms of the impact that each feature in the dataset is having directly on accuracy of the model and hence unimportant variable have less effect on ANFIS model. Using Mean

decrease gini on the other hand gives variables represented in way such that the most important attribute is at the top at the same time giving this as a dot on an x-axis and it uses gini impurity on each node during the growing of the trees. The first ten features are used for ANFIS model development that plots the membership functions for each attribute as in Figure 5.

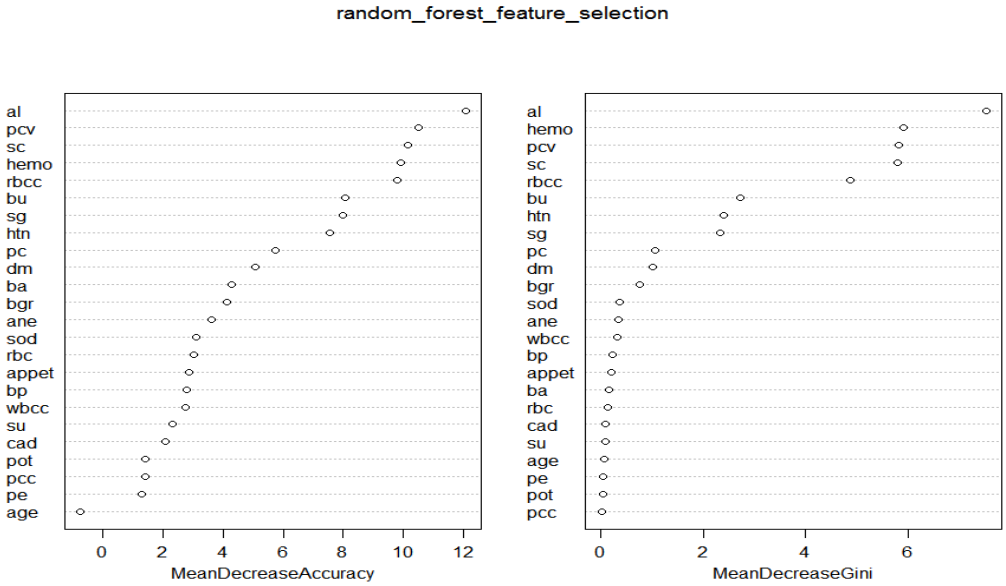

Figure 4. Random Forest Feature Selection

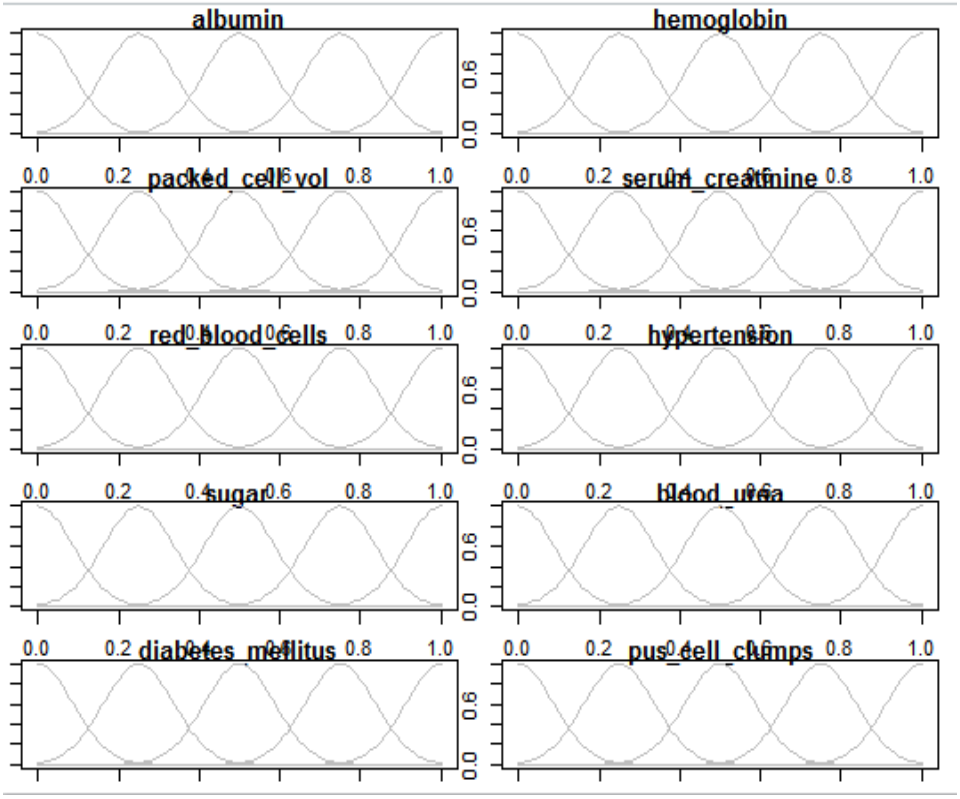

Figure 5. Plot of the Gaussian Membership Functions

6.3. Hierarchical Clustering from ANFIS Results

Authors determine the number of clusters using nbclust. Using the elbow method, it is seen that the optimal number of clusters for the ANFIS predicted results is 3 as shown in Figure 7. This shows therefore that to find the maximum similarity the k of hierarchical clustering is set to 3. This will enable identification for diabetic and other chronic diseases clusters.

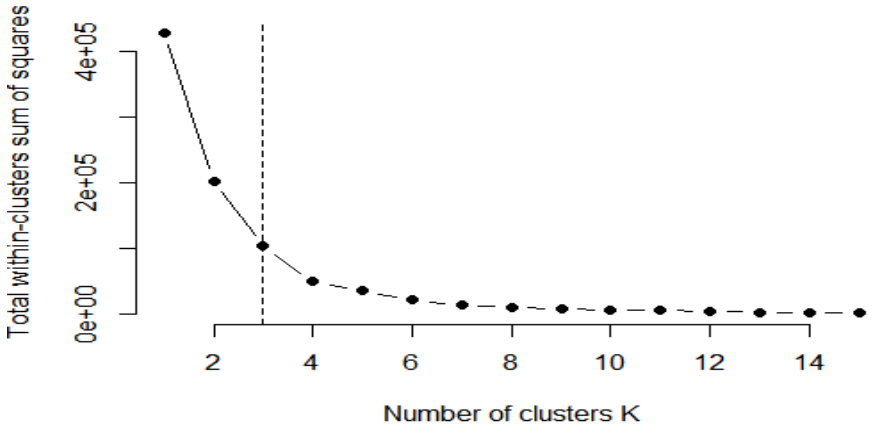

Figure 6. Determining Optimal Number of Clusters

The dendrogram below show the clusters formed by using hierarchical clustering.

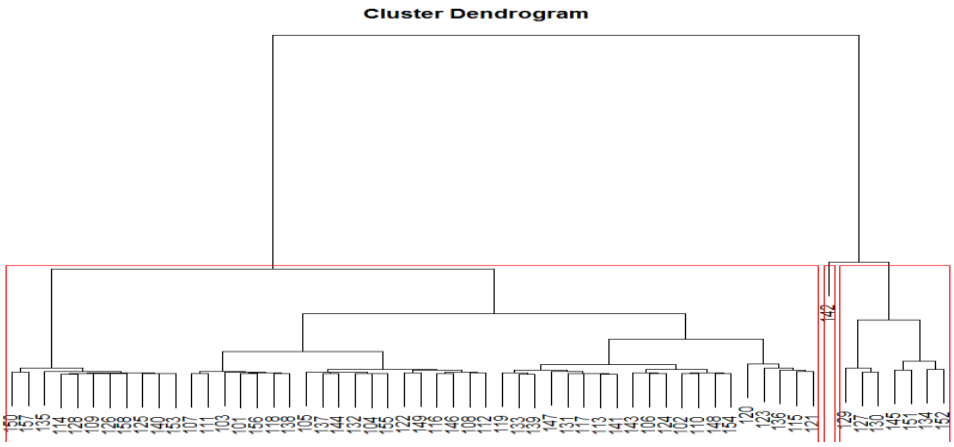

Figure 7. Dendogram of Predicted Results

The figure below show the cluster details.

```
> clusterCut <- cutree(clusters, 3)
> rect.hclust(clusters,k=3)
> table(clusterCut, clust$pred)

clusterCut  1  2
         1  0 50
         2  0  7
         3  1  0
```

Figure 8. Clusters Formed as Per Class

```
> aggregate(clust,by=list(cluster=clusterCut),median)
  cluster albumin hemoglobin packed_cell_vol serum_creatinine red_blood_cells hypertension
1       1       0      14.65             46              0.9             5.1            0
2       2       3      10.30             33              5.3             3.6            1
3       3       4       3.10              9             13.3             2.1            1
  sugar blood_urea diabetes_mellitus pus_cell_clumps pred
1  1.02       37.5                 0               1    2
2  1.01      125.0                 1               0    2
3  1.01      309.0                 1               0    1
> |
```

Figure 9. Analysis of Clusters Formed

7. Results and discussions

Using neuro-fuzzy system, predictions were done with an overall accuracy of 97 percent. The results of the prediction showing the risk of any patient having CKD given the ten features selected from feature selection are clustered using hierarchical clustering. The 3 clusters formed identify a group of patients who have higher risk of other related chronic diseases. Cluster 1 shows that those in this cluster are similar in nature in that they have high hemoglobin, high sugar level, low blood urea, high packed cell, high red blood cell count, with pus cell clumps and without having diabetes mellitus form cluster 1 and these have CKD.

The majority in cluster 2 have sugar which is a little bit lesser than those in cluster 1. However because the sugar level is 1.01 they form cluster 2 which is also classified as having CKD. Those with lower serum creatinine also formed a cluster. It has been shown from the results of this study that a patient with hypertension, having diabetes mellitus and high sugar level there is a greater risk of kidney failure.

7.1. Confusion matrix

The confusion matrix for the neuro-fuzzy algorithm on application of algorithm on the dataset gave the following results of classifying the risk as indicated in Table .2.

Table 2. Confusion Matrix

	Predicted	
Actual	1	2
1	14	0
2	0	44

Table 3. Accuracy Measures for ANFIS

Accuracy	100%
Sensitivity	100%
Specificity	97%

The error rate for prediction is found to be 0. Accuracy of the neuro-fuzzy model can be summarized as in table.3.

8. Conclusions and Future Work

The results of this research can be added to the domain of healthcare and can be used for providing suggestions in the domain by making it easy for healthcare professionals in diagnosis and treatment of patients as well as for identifying relationships within diseases that patients have. Future work should mainly focus on implementing more big data

oriented tools and techniques which makes the process much faster and effective. The greatest challenge in healthcare domain is the data, provided enough and appropriate data is available, many applications can be implemented which take the healthcare industry to an advanced level.

Acknowledgements

The author Dr. N. Ch. S. N. Iyengar is thankful to Dr. P. Narasimha Reddy, Executive Director, SNIST and Dr. G. Viswanathan, Chancellor, VITU for their support and encouragement.

References

[1] D. Azeez, M. A. Ali, K .B. Gan and I. Saiboon, "Comparison of adaptive neuro-fuzzy inference system and artificial neutral networks model to categorize patients in the emergency department", Springer Plus 2013.
[2] M. Zhamim Hossain and G. Muhammad, "Big Data Voice Pathology Assessment Framework", IEEE Access, **(2016)**.
[3] Dr. S. Vijayarani1 and S.Dhayanand, "Kidney disease prediction using SVM and ANN algorithms", International Journal of Computing and Business Research (IJCBR) ISSN (Online), vol. 6, no. 2, **(2015)**, pp. 2229-6166.
[4] M. Kumar, "Prediction of Chronic Kidney Disease Using Random Forest Machine Learning Algorithm", "International Journal of Computer Science and Mobile Computing", vol. 5, no. 2, **(2016)**, pp. 24-33.
[5] K. R. Anantha Padmanaban and G. Parthiban, "Applying Machine Learning Techniques for Predicting the Risk of Chronic Kidney Disease", Indian Journal of Science and Technology, DOI: 10.17485/ijst/2016/v9i29/93880, vol. 9, no. 29.
[6] A. Perotte1, R. Ranganath, J. S. Hirsch1, D. Blei and N. Elhadad1, "Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis", Oxford University press.
[7] N. Sharma and Er. Rohit Kumar Verma, "Prediction of Kidney Disease by using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 9, September 2016, ISSN: 2277 128X.
[8] M.A.M. Abushariah, A.A.M. Alqudah, O.Y. Adwan, and Yousef, "Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches", Journal of Software Engineering and Applications, pp. 1055-1064.
[9] A.Y. Al-Hyari, M. Al-Taee and M. A. Al-Taee, "Clinical Decision Support System for Diagnosis and Management of Chronic Renal Failure", IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), **(2013)**.
[10] N. Chetty, K.S. Vaisla and S.D. Sudarsan, "Role of Attributes Selection in Classification of Chronic Kidney Disease Patients", IEEE, **(2015)**.
[11] A. Q. Ansari and N. K. Gupta, "Automated Diagnosis of Coronary Heart Disease Using Neuro-Fuzzy Integrated System".
[12] Welling Healthcare Private Limited available online at http://homeopathyclinic.co.in/undiagnosed-chronic-kidney-disease-ckd-india/.
[13] J. Sun and C. K. Reddy, "Bigdata Analytics for Healthcare", Tutorial presentation at the SIAM International Conference on Data Mining, Austin, TX, **(2013)**.
[14] J. Jang, C. Sun and E. Mizutani, "Neuro-fuzzy and Soft Computing", Prentice Hall, **(1999)**.
[15] P. M. Patil, "Review on Prediction of Chronic Kidney disease Using Data Mining Techniques", International Journal of Computer Science and Mobile Computing, vol. 5, no. 5, **(2016)**, pp. 135.

Authors

Kerina Blessmore Chimwayi, she did her B. Tech in Electronic Commerce from Harare Institute of Technology, Harare, Zimbabwe and is currently pursuing M. Tech in Computer Science & Engineering (with Specialization in Big Data Analytics) at VIT University, Vellore. Her research interests include Machine Learning, Deep Learning and Spatial Data Mining in Healthcare.

Noorie Haris, she did her B.tech in Computer Science and Engineering from MES College of Engineering, Kuttipuram, Kerala and is currently pursuing M.tech in Computer Science and Engineering with specialization in Big Data Analytics at VIT University, Vellore. Her research interests include advanced analytics using big data technologies.

Ronnie D. Caytiles, he had his Bachelor of Science in Computer Engineering- Western Institute of Technology, Iloilo City, Philippines, and Master of Science in Computer Science– Central Philippine University, Iloilo City, Philippines. He finished his Ph.D. in Multimedia Engineering, Hannam University, Daejeon, Korea. Currently, he serves as an Assistant Professor at Multimedia Engineering department, Hannam University, Daejeon, Korea. His research interests include Mobile Computing, Multimedia Communication, Information Technology Security, Ubiquitous Computing, Control and Automation.

N. Ch. S. N. Iyengar, (b 1961), currently Professor, Information Technology, Sreenidhi Institute of Science (SNIST) and Technology, Yamnapet, Ghatkesr, Hyderabad-501301,Telengana, India He is a doctorate in both Applied Mathematics and Computer Science and Engineering. His research interests include Agent-Based Distributed Computing, Intelligent Computing, Network Security, Cloud Computing, Big Data Analytics and Fluid Mechanics. He had 32+ years of experience in teaching to B.Tech and M.Tech students. He guided 12 Ph.Ds, 5 M. Phils and 75 + M.Tech Projects apart from authoring several textbooks. He had 200+ research publications in reputed peer reviewed international journals along with students. He organized many conferences/workshops and continuing education programmes He served as Keynote speaker/ / Invited speaker //PCM/reviewer for many International conferences. He serves as an Editor in chief/Guest editor /Editorial board member for many international journals. He is the professional member of many bodies.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Machine learning Approach of Chronic Kidney Disease Prediction using Clustering Technique

S.Gopika [1], Dr.M.Vanitha [2]

Research Scholar, PG and Research Department of Computer Science, J.J College of Arts and Science (Autonomous), Pudukottai, TamilNadu, India[1]

Assistant Professor, Department of Computer Applications, Alagappa University, Karaikudi, TamilNadu, India[2]

ABSTRACT: Chronic Kidney Disease (CKD) is a gradual decrease in renal function over a period of several months or years. Diabetes and high blood pressure are the most common causes of chronic kidney disease. The main objective of this work is to determine the kidney function failure by applying the clustering algorithm on the test result obtained from the patient medical report. The aim of this work is to reduce the diagnosis time and to improve the diagnosis accuracy using clustering algorithms. The proposed work deals with clustering of different stages in chronic kidney disease according to its severity. The experiment is performed on different algorithms like k-means, k-medoids and Fuzzy C Means. The experimental results show that the Fuzzy c means algorithm gives better result than the other clustering algorithms and produces 87% accuracy.

KEYWORDS: Chronic Kidney Disease (CKD), Clustering, Data mining, Machine Learning (ML), Fuzzy C Means

I. INTRODUCTION

Data Mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

This has led to the emergence of Knowledge Discovery in Databases (KDD) which is responsible for transforming low-level data into high-level knowledge for decision making. Knowledge discovery in databases consists of the list of iterative sequence steps of processes and data mining is one of the KDD processes. Data mining is the application of algorithms for extracting patterns from large volume of data. There is a wealth of data available within the healthcare systems. The healthcare environment is information rich yet knowledge poor. Hence, for healthcare research, data driven statistical research has become a complement. As with the use of computers powered with automated tools the large volumes of healthcare data are being collected and made available to the medical research groups. As a result, Knowledge Discovery in Databases (KDD), which includes data mining techniques, has become a popular research tool for healthcare researchers to identify and exploit patterns and relationships among large number of variables, and also made them able to predict the outcome of a disease using the historical cases stored within datasets.

Clustering is an important area of application for a variety of fields including data mining, knowledge discovery, statistical data analysis, data compression and vector quantization. Clustering has been formulated in various ways in machine learning, pattern recognition, optimization and statistics literature. Clustering is the most common form of

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

unsupervised learning. According to the rule of the unsupervised learning, clustering does not require supervision. No supervision means that there is no human expert who has assigned documents to classes. In clustering, it is the distribution and makeup of the data that will determine cluster membership. The notion of what constitutes a good cluster depends on the application and there are many methods for finding clusters subject to various criteria. These include approaches based on splitting and merging such as ISODATA, randomized approaches such as CLARA, CLARANS, and methods based on neural nets, and methods designed to scale to large databases, including DBSCAN, BIRCH and ScaleKM. Among clustering formulations that are based on minimizing a formal objective function, perhaps the most widely used and studied is partition based algorithms like K-Means, and Fuzzy C-Means clustering. In a partitioned algorithm, given a set of n data points in real d-dimensional space, and an integer k, the problem is to determine a set of k points in Rd, called centers, so as to minimize the mean squared distance from each data point to its nearest center. This measure is often called the squared-error distortion and this type of clustering falls into the general category of variance based clustering.

Chronic kidney diseases have become a major public health problem. Chronic diseases are a leading cause of morbidity and mortality in India. Chronic kidney diseases account for 60% of all deaths worldwide. Eighty percentage of chronic disease deaths worldwide occur in low - and middle-income countries [2]. The National Kidney foundation determines the different stages of chronic kidney disease based on the presence of kidney damage and glomerular filtration rate (GFR), which is measure a level of kidney function. There are five stages of chronic kidney disease.

The remaining paper is organized as follows: Section II discusses literature survey of the research. In section III describes the methodologies used for clustering chronic kidney disease are discussed. Section IV deals with the experiments and its results for parameter measures used in clustering, and the result obtained in each clustering algorithms. Section V describes the conclusion of the proposed work along with the feature enhancement.

II. LITERATURE REVIEW

Mohammed Abdul Khaleel et al. This paper presents a performance analysis on various clustering algorithm namely K-means, expectation maximization, and density based clustering in order to identify the best clustering algorithm for microarray data. Sum of squared error, log likelihood measures are used to evaluate the performance of these clustering methods. This paper conducted an empirical study on various clustering algorithms in order to observe their performance on gene expression data in terms of sum of squared error and log likelihood. In this empirical study, the performance of the clustering algorithms namely density based clustering, expectation maximization clustering and K-means clustering are evaluated on various gene expression data. From this evaluation, it is observed that the performance of expectation maximization clustering algorithm is comparatively better than the density based clustering algorithm in terms of log likelihood [1].

Veerappan et al. This paper analyze the three major clustering algorithms: K-Means, Farthest First and Hierarchical clustering algorithm and compare the performance of these three major clustering algorithms on the aspect of correctly class wise cluster building ability of algorithm. The result analysis shows that K-means algorithm performs well without inserting the principle component analysis filter as compared to the Hierarchical clustering algorithm and Farthest first clustering since it have less instances of incorrectly clustered objects on the basis of class clustering. Hierarchical clustering as compared to Farthest fast clustering gives better performance. Also this algorithm performs better after merging principle component analysis filter with it. Farthest first clustering though gives a fast analysis when taken an account of time domain, but makes comparatively high error rate. [2]

Lambodar Jena et al. This paper mainly presents an overview of types of clustering techniques and some of the applications of data mining where clustering techniques can be applied. The main goal of clustering is to produce a good and high quality clusters that depends mainly on the similarity measure which has the ability to discover some or all hidden patterns and also make the analysis of data easy. The quality of clusters produced by clustering method is

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

measured by its ability to discover some or all of the hidden patterns. It has been observed that, the most common type of clustering technique that has been used by different applications of data mining is the k-means clustering technique [3].

Abeer et al. In this paper we take diabetes and heart datasets relate with their matching fields then apply the classification algorithm in kidney dataset in software tool finding weather people affected by diabetes are getting chance to get kidney disease or not, output are evaluated as Tested Negative (No Diabetes), Tested Normal(Not affected), Tested High(affected). A new approach for efficiently predicting the diabetes, kidney disease from some medical records of patients. Dataset has designed with matching attributes applied in classification algorithms like J48, Random Tree, Random Forest, REP, Naïve Bayesian algorithm [4].

Jerlin Rubini et al. In this study, we have made a comprehensive comparative analysis of 14 different classification algorithms and their performance has been evaluated by using 3 different cancer data sets. The results indicate that none of the classifiers outperformed all others in terms of the accuracy when applied on all the 3 data sets. Most of the algorithms performed better as the size of the data set is increased. We recommend the users not to stick to a particular classification method and should evaluate different classification algorithms and select the better algorithm. This study focuses on finding the right algorithm for classification of data that works better on diverse data sets. However, it is observed that the accuracies of the tools vary depending on the data set used. It should also be noted that classifiers of a particular group also did not perform with similar accuracies [5].

III. METHODOLOGY

A. Cluster Analysis

The objective of cluster analysis is the classification of objects according to similarities among them, and organizing of data into groups. Clustering techniques are among the unsupervised methods, they do not use prior class identifiers. The main potential of clustering is to detect the underlying structure in data, not only for classification and pattern recognition, but for model reduction and optimization. Various definitions of a cluster can be formulated, depending on the objective of clustering. Generally, one may accept the view that a cluster is a group of objects that are more similar to one another than to members of other clusters. The term "similarity" should be understood as mathematical similarity, measured in some well-defined sense. In metric spaces, similarity is often defined by means of a distance norm.

The proposed system compare K-Means, K-Means++, Fuzzy C-Means clustering algorithms for the chronic kidney disease from UCI machine repository , in these algorithms group the dataset with its matrix values to calculate the PC (Partition Coefficient),CE (Classification Entropy), SC (Partition Index), S (Separation Index), XB (Xie and Beni's Index), DI (Dunn's Index), ADI(Alternative Dunn Index). To compare those algorithms with various validity indices.

B. K-means Algorithm

The k-means algorithm takes the input parameter, k, and partitions a set of n objects into k clusters so that the resulting intracluster similarity is high but the intercluster similarity is low. Cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster's centroid or center of gravity.

Algorithm steps

Given the data set X, choose the number of clusters $1 < c < N$.

Initialize with random cluster centers chosen from the data set. Repeat for $l = 1; 2;$

Step 1 Compute the distances

$$D_{ik}^2 = \left(x_k - v_i\right)^T \left(x_k - v_i\right), \qquad 1 \le i \le c, \qquad 1 \le k \le N.$$

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Step 2 Select the points for a cluster with the minimal distances, they belong to that cluster.

Step 3 Calculate cluster centers

$$v_i^{(l)} = \frac{\sum_{j=1}^{N_i} x_i}{N_i}$$

Until

$$\prod_{k=1}^{n} max \left| v^{(l)} - v^{(l-1)} \right| \neq 0$$

Ending Calculate the partition matrix

C. K-Medoids Algorithm

The *k*Medoidsalgorithm is clustering algorithm related to the *k*-means algorithm and the medoid shift algorithm. Both the *k*-means and *k*-medoids algorithms are partition (breaking the dataset up into groups) and both attempt to minimize the distance between points labeled to be in a cluster and a point designated as the center of that cluster. In contrast to the *k*-means algorithm. A medoids can be defined as the object of a cluster whose average dissimilarity to all the objects in the cluster is minimal. I.e. it is a most centrally located point in the cluster.

Algorithm Steps:

1: Arbitrarily choose k data items as the initial medoids.

2: Assign each remaining data item to a cluster with the nearest medoid.

3. Randomly select a non-medoid data item and compute the total cost of swapping old medoid data item with the currently selected non-medoid data item.

4. If the total cost of swapping is less than zero, then perform the swap operation to generate the new set of k-medoids.

5. Repeat steps 2, 3 and 4 till the medoids stabilize their locations.

D. Fuzzy C-Means (FCM)

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. Straightly speaking, this algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one.

Algorithm steps

1. Randomly select cluster centre
2. Initialize U=[u_{ij}] matrix, $U^{(0)}$

$$u_{ij} = \frac{1}{\sum_{k-1}^{c} \left(\frac{|| xi - cj ||}{|| xi - ck ||} \right)^{\frac{2}{2\pi - 1}}}$$

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

3. Calculate the u_{ij} using:

4. At k-step: calculate the centre vectors $C^{(k)}=[c_j]$ with $U^{(k)}$

$$c_j = \frac{\sum_{i-1}^{N} u_{ij}^{m} \; xi}{\sum_{i-1}^{N} u_y^{m}}$$

5. Update $U^{(k)}$,$U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k-1}^{c} \left(\frac{|| xi - cj ||}{|| xi - ck} \right)^{2m-1}}$$

6. If $|| U^{(k+1)} - U^{(k)} ||< \varepsilon$ or the minimum J is achieved, then STOP; otherwise return to step 2.

7. In the end, you will get a result like:

E. Validity Indices

Different scalar validity measures have been proposed in the literature, none of them is perfect by oneself, and therefore we used several indexes in our Toolbox, which are described below:

1. Partition Coefficient (PC): measures the amount of "overlapping" between clusters.

$$PC(c) = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N} (\mu_{ij})2$$

Where μ_{ij} is the membership of data point *j* in cluster *i*. The disadvantage of PC is lack of direct connection to some property of the data themselves. The optimal number of cluster is at the maximum value.

2. Classification Entropy (CE): it measures the fuzzyness of the cluster partition only, which is similar to the Partition Coefficient.

$$CE(c) = -\frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{ij} \log(\mu_{ij})$$

3. Partition Index (SC): is the ratio of the sum of compactness and separation of the clusters. It is a sum of individual cluster validity measures normalized through division by the fuzzy cardinality of each cluster.

$$SC(c) = \sum_{i=1}^{c} \frac{\sum_{j=1}^{N} (\mu_{ij})^{m} \; || x_{j-} v_i ||^2}{N_i \; \sum_{k=1}^{c} || v_{k-} v_i ||^2}$$

SC is useful when comparing different partitions having equal number of clusters. A lower value of *SC* indicates a better partition.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

4. Separation Index (S): on the contrary of partition index (SC), the separation index uses a minimum-distance separation for partition validity.

$$S(c) = \frac{\sum_{i=1}^{c} \sum_{j=1}^{N} \left(\mu_{ij}\right)^{2} \| x_{j-} v_{i} \|^{2}}{N \min_{i,k} \| v_{k-} v_{i} \|^{2}}$$

5. Xie and Beni's Index (XB): it aims to quantify the ratio of the total variation within clusters and the separation of clusters.

$$XB(c) = \frac{\sum_{i=1}^{c} \sum_{j=1}^{N} \left(\mu_{ij}\right)^{m} \| x_{j-} v_{i} \|^{2}}{N \min_{i,j} \| x_{j-} v_{i} \|^{2}}$$

The optimal number of clusters should minimize the value of the index.

6. Dunn's Index (DI): this index is originally proposed to use at the identification of "compact and well separated clusters". So the result of the clustering has to be recalculated as it was a hard partition algorithm.

$$DI(c) = min_{i\epsilon c}\{min_{j\epsilon c, i\neq j}\{\frac{min_{x\epsilon C_i, y\epsilon C_j} d(x,y)}{max_{k\epsilon c}\{max_{x,y\epsilon C^d(x,y)}\}}\}\}$$

The main drawback of Dunn's index is computational since calculating becomes computationally very expansive as *c* and *N* increase.

7. Alternative Dunn Index (ADI): the aim of modifying the original Dunn's index was that the calculation becomes simpler, when the dissimilarity function between two clusters ($min_{x\epsilon Ci, y\epsilon Cj} d(x, y)$) is rated in value from beneath by the triangle-non equality:

$$d(x, y) \geq |d(y, v_j) - d(x, v_j)|$$

Where *vj* is the cluster center of the *j*-th cluster.

$$ADI(c) = min_{i\epsilon c}\{min_{j\epsilon c, i\neq j}\{\frac{min_{x_i\epsilon C_i, x_j\epsilon C_j | d(y,v_j) - d(x_i,v_j)|}}{max_{k\epsilon c}\{max_{x,y\epsilon C^d(x,y)}\}}\}\}$$

IV. EXPERIMENTAL RESUTS

The dataset for diagnosis of chronic kidney disease is obtained from medical reports of the patients collected from UCI machine learning repository. There are 400 instances with 25 different attributes related to kidney disease like PID (patients ID), Age, Gender, Weight, Serum - albumin, Serum - sodium, Blood urea nitrogen, Serum creatinine, Serum uric acid, Sodium urine, Urine urea nitrogen, Urine creatinine, Urine uric acid ,EGFR and Kidney failure. The main contributing attribute to identify the chronic kidney disease is EGFR. Based on the value of the EGFR, the instances are classified as Low, Mild, Moderate, Normal and Severe.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

Table 1: Description of Dataset

Dataset	Number of Objects	Number of Attributes	Number of Clusters
Chronic Kidney Disease	400	25	2

A. Validity Measure

Table 2: Clustering Validity Measure for Chronic Kidney Disease Dataset and Algorithm

Dataset	Algorithm	PC	CE	SC	S	XB	DI	ADI
Chronic Kidney Disease	**K-Means**	1	NaN	0.7265	0.0014	3.1784	0.6481	0.5636
	K-Means++	1	NaN	0.7237	0.0014	Inf	0.6481	0.5501
	FCM	0.8088	0.3311	0.9548	0.0018	2.6115	0.64808	0.5471

Table 3: Performance Measure for Chronic Kidney Disease Dataset and Algorithm

Performance Measure	K-Means	K-Medoids	FCM
Sensitivity	0.89	0.73	0.92
Specificity	0.83	0.80	0.87
Accuracy	86%	76.5%	89%

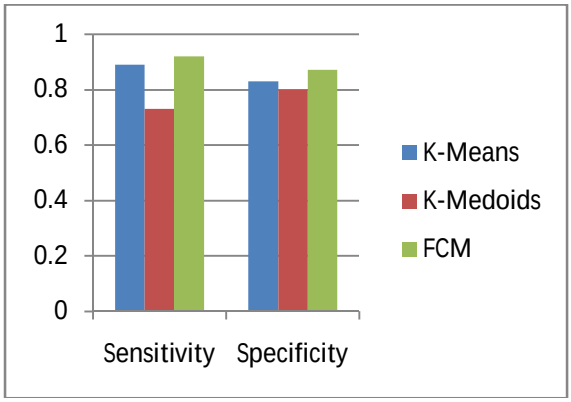

Fig 1.Performance Measure for sensitivity and Specificity

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

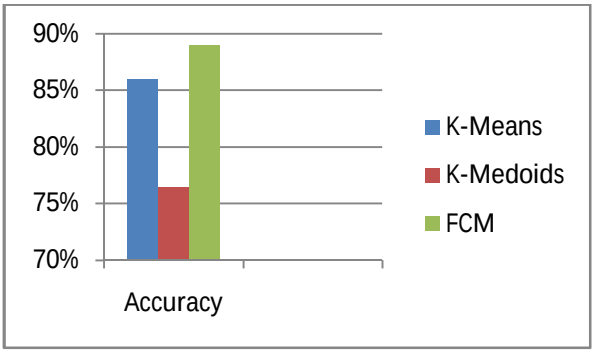

Fig 2.Performance Measure for Accuracy

Prediction of chronic kidney disease is one of the essential topics in medical diagnosis. The proposed work is to cluster the different stages of chronic kidney disease according to its severity. The clustering algorithms that have been considered for predicting chronic kidney disease are k-means, k-medoids and FCM. The models are evaluated with four different measures like Accuracy, Sensitivity and Specificity. From the experimental result, the FCM Function is the better accuracy for predicting chronic kidney disease and it attains the accuracy of 89%.

V. CONCLUSION

Cluster analysis is one of the major tasks in various research areas. The clustering aims at identifying and extract significant groups in underlying data. This is based on a certain clustering criterion the data are grouped so that data points in a cluster are more similar to each other than points in different clusters. Since clustering is applied in many fields, a number of clustering techniques and algorithms have been proposed and are available in literature. In the proposed system to analysis the major clustering algorithms such as K-Means, K-Medoids and Fuzzy C-Means with Euclidean distance measure by using Chronic Kidney Disease UCI dataset. It illustrates the efficiency of clustering algorithm with its validity measures. It shows the Fuzzy C-Means clustering algorithm had better than other clustering algorithms. The experimental result shows the performance of the Fuzzy C-Means algorithm was improved significantly.

REFERENCES

[1]Mohammed Abdul Khaleel and Sateesh Kumar Pradham, "A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases", International Journal of Advanced Research in Computer Science and Software Engineering, 2013, Vol.3, No. 8, pp. 149-153.
[2]Veerappan, Ilangovan and Abraham Georgi," Chronic Kidney Disease: Current Status, Challenges and Management in India", Indian Journal of Public Health Research and Development, 2014, Vol.6, No.1, pp. 1694-1702.
[3]Lambodar Jena and Narendra Ku. Kamila, "Distributed Data Mining Classification Algorithms for Prediction of Chronic Kidney Disease", International Journal of Emerging Research in Management and Technology, 2015, Vol. 4, No. 11, pp. 110-118.
[4]Abeer, Ahmad and Majid," Diagnosis and Classification of Chronic Renal Failure Utilizing Intelligent Data Mining Classifiers", International Journal of Information Technology and Web Engineering, 2014, Vol.9, No.4, pp. 1-12.
[5]Jerlin Rubini and Eswaran," Generating Comparative Analysis of Early Stage Prediction of Chronic Kidney Disease", International Journal of Modern Engineering Research, 2015, Vol. 5, No. 7, pp. 49-55.
[6]Ruey Kei Chiu and Renee Yu - Jing," Constructing Models for Chronic Kidney Disease Detection and Risk Estimation", Proceedings of 22nd IEEE International Symposium on Intelligent Control, Singapore, 2011, pp. 166-171.
[7]Jayalakshmi and Santhakumaran, "Improved Gradient Descent Back Propagation Neural Networks for Diagnoses of Type II Diabetes Mellitus", Global Journal of Computer Science and Technology, 2010, Vol. 9, No. 5, pp. 94-97.
[8]Koushal Kumar and Abhishek, "Artificial Neural Networks for Diagnosis of Kidney Stones Disease", International Journal Information Technology and Computer Science, 2012, Vol. 7, No. 3, pp. 20-25.
[9]Rajalakshmi, Neelamegam and Bharathi, "Diagnosis and Classification of Level of Kidney function Using Associative Neural Network and Polynomial Neural Network", Research Journal of Pharmaceutical, Biological and Chemical Sciences, 2013, Vol. 4,No. 4, pp. 724-738.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 7, July 2017

[10]Vijayarani and Dhayanand, "Kidney disease prediction using Support Vector Machine and Artificial Neural Network algorithms", International Journal of Computing and Business Research (IJCBR), 2015, Vol. 1, No. 3, pp.1765-1771.
[11]Shweta Kharya, "Using data mining techniques for diagnosis and prognosis of cancer disease", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), 2012, Vol. 2, No. 2, pp. 55-66.
[12]Abhinandan Dubey, "A Classification of CKD Cases Using Multi Variant K-Means Clustering", International Journal of Scientific and Research Publications, 2015, Vol. 5, No. 8, pp. 1-5.
[13]Yashpal Singh and Alok Singh Chauhan," Neural Networks in Data Mining", Journal of Theoretical and Applied Information Technology, 2005, Vol.5, No.6, pp. 37-42.
[14]Baylor, Konukseven and Koku," Control of a Differentially Driven Mobile Robot Using Radial Basis Function Based Neural Networks", World Scientific and Engineering Academy and Society Transactions on Systems and Control, 2008, Vol. 3, No. 12, pp. 1002-1013.
[15]Shubham Bind, Arvind Kumar Tiwari and Anil Kumar Sahani, " A Survey of Machine Learning Based Approaches for Parkinson Disease Prediction", International Journal of Computer Science and Information Technologies, 2015,Vol. 6, No. 2, pp.1648-1655.

Informatics in Medicine Unlocked 15 (2019) 100178

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Prediction of kidney disease stages using data mining algorithms

El-Houssainy A. Rady[a], Ayman S. Anwar[b,*]

[a] Institute of Statistical Studies & Research, Cairo University, Giza, Egypt
[b] Magrabi Hospitals & Centers, Cairo, Egypt

ARTICLE INFO

Keywords:
Prediction of kidney disease stages
Data mining techniques
Probabilistic neural networks
Multilayer perceptron
Support vector machine
Radial basis function

ABSTRACT

Early detection and characterization are considered to be critical factors in the management and control of chronic kidney disease. Herein, use of efficient data mining techniques is shown to reveal and extract hidden information from clinical and laboratory patient data, which can be helpful to assist physicians in maximizing accuracy for identification of disease severity stage. The results of applying Probabilistic Neural Networks (PNN), Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Radial Basis Function (RBF) algorithms have been compared, and our findings show that the PNN algorithm provides better classification and prediction performance for determining severity stage in chronic kidney disease.

1. Introduction

A global health problem which is steadily growing is Chronic kidney disease (CKD). It is a chronic condition associated with increased morbidity and mortality, a high risk of many other diseases including cardiovascular disease, and high health care costs. Over two million people worldwide receive dialysis or kidney transplant treatment to stay alive, yet this number may represent only 10% of people who need treatment to live [9]. The majority of the 2 million people who receive treatment for kidney failure are in only five relatively wealthy countries, which represent 12% of the global population. By comparison, only 20% of the world's population is treated in about 100 developing countries, and they represent almost half the global population. Annually, more than one million people in 112 lower-income countries die from untreated kidney failure, due to the huge financial burden of dialysis or kidney transplantation treatment [9].

Thus, there is significant importance in the early detection, controlling, and managing of the disease. It is necessary to predict the progression of CKD with reasonable accuracy because of its dynamic and covert nature in the early stages, and patient heterogeneity. CKD is often described by severity stages. Clinical decisions are influenced by the stage, whether a patient is progressing, and the rate of progression. Also, defining the disease stage is quite crucial as it gives several indications that support the determination of required intervention and treatments.

Therefore, data mining can play a major role in extracting hidden data from the large patient medical and clinical dataset that physicians frequently collect from patients to obtain insights about the diagnostic information, and to implement precise treatment plans. Data mining can be defined as the process of extracting hidden data from a large dataset. Data mining techniques are applied and used widely in various contexts and fields. With data mining techniques we could predict, classify, filter and cluster data. The goal or prediction attribute refers to the algorithm processing of a training set containing a set of attributes and outcomes.

Machine learning algorithms have been used to predict and classify in the healthcare field. Yu et al. [17] have used the Support Vector Machine Algorithm to classify and predict diabetes and pre-diabetes patients, and the results show that SVM is useful to classify patients with common diseases. Similarly, Magnin et al. [19] have classified Alzheimer's disease by using a Support Vector Machine (SVM) to analyze whole-brain anatomical magnetic resonance imaging (MRI) for a set of patients, and the results shows that SVM is a promising approach for Alzheimer's disease early detection. Dessai et al. [18] have done heart disease prediction using the Probabilistic Neural Network Algorithm, Decision tree Algorithm, and Naïve Bayes Algorithm, and PRNN provides the best results compared with other algorithms for heart disease prediction. Cao et al. [20] have done prediction of HBV-induced liver cirrhosis using the Multilayered Perceptron (MLP) Algorithm and the results shows that the MLP classifier gives satisfactory prediction outputs for liver disease, mostly in HBV-related liver cirrhosis patients.

2. Materials & methods

Data Mining was utilized in our study because it is a process of identifying novel, potentially useful, valid and ultimately understandable patterns in data [4]. Supervised and unsupervised learning

* Corresponding author.
E-mail address: ayman.anwarr@gmail.com (A.S. Anwar).

https://doi.org/10.1016/j.imu.2019.100178
Received 9 December 2018; Received in revised form 11 March 2019; Accepted 1 April 2019
Available online 07 April 2019

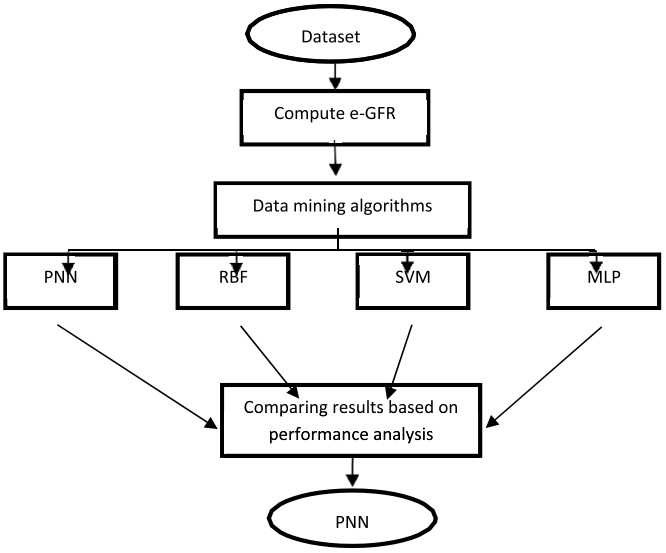

Fig. 1. Methodology workflow.

techniques are used for data mining classification. A "supervised" learning technique requires the building of a model based on previous performance analysis and is used in both medical and clinical research for classification, statistical regression and association rules [5]. On the other hand, the "unsupervised" learning technique is not guided by prior analysis and does not create a pre-analysis hypothesis. A model can be constructed based upon the results and is useful for clustering [6].

Three different types of the most commonly used artificial neural network algorithms and support vector machine algorithms have been used for this study, to determine which algorithm will give the best classification results, so as to identify the stage of chronic kidney disease, based on patient clinical and laboratory data. (see Fig. 1)

Machine learning techniques employ two phases to build the predictive/classification model as follows:

- A training phase that learns algorithmically how to build the model by using training datasets with expected outputs.
- A validation phase that estimates how well the model has been trained by using validation datasets without the expected outputs.

2.1. Probabilistic Neural Networks

Probabilistic Neural Networks (PNN) are a kind of Radial Basis Function neural network with a one pass learning algorithm and highly parallel structure. PNN was introduced by Donald F. Specht in 1990 as a memory-based network that provides estimates of categorical variables. The algorithm provides a smooth approximation of a target function, even with sparse data in a multidimensional space [16]. The advantages of PNN are fast learning and easy tuning. The PNN is composed of four layers: input, pattern (RBF kernel function), summation, and output, as shown in Fig. 2. Each neuron of the pattern layer uses a radial basis function as an activation function. This function is commonly taken to be Gaussian.

2.2. Multilayer Perceptron algorithm

The Multilayer Perceptron (MLP) is one of most important class of neural networks, consisting of an input layer, one or more hidden layers, and the output layer, as shown in Fig. 3. MLPs have been applied successfully to solve difficult and diverse problems, by training them in a supervised manner using a well-known algorithm i.e., the error back-propagation algorithm [3]. This algorithm is based on the error

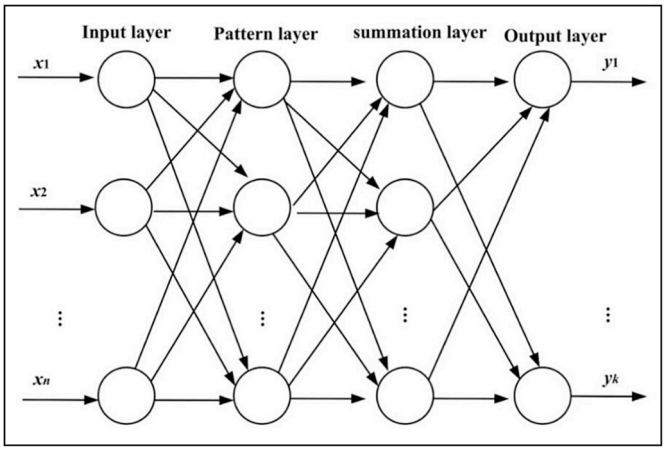

Fig. 2. Probabilistic neural networks (PNN) Layers.

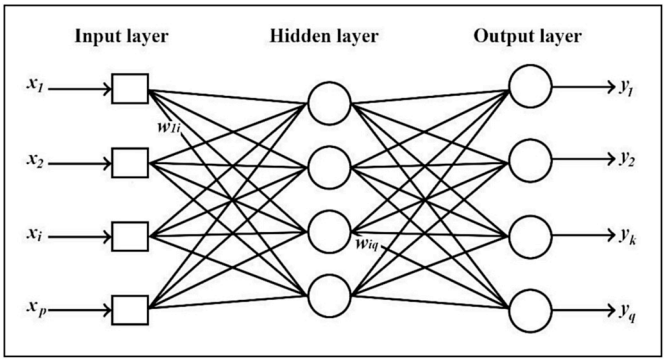

Fig. 3. Multilayer perceptron (MLP) layers.

correction learning rule. As such, it may be viewed as a generalization of an adaptive filtering algorithm.

2.3. Support vector machine algorithm

The SVM is a method for the classification of both linear and nonlinear data [7]. The SVM algorithm works as follows. It uses a nonlinear mapping to renovate the unique training data into a higher dimension. Surrounded by this new dimension, it examines the linear optimal separating hyperplane as shown in Fig. 4, i.e., a "decision boundary" sorting out the tuples of one class from another. With a suitable

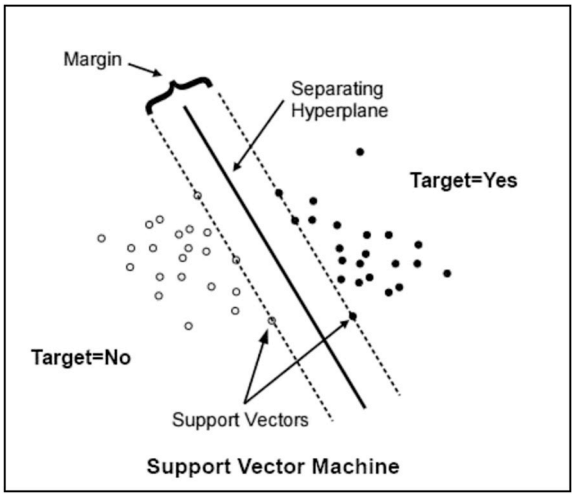

Fig. 4. Support Vector Machine (SVM) optimal hyperplane.

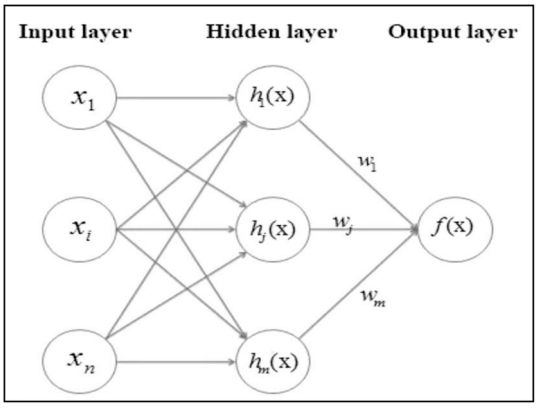

Fig. 5. Radial basis function (RBF) layers.

nonlinear mapping to a necessarily high dimension, data from two classes can always be separated by a hyperplane. The SVM finds the hyperplane using support vectors and margins [13]. Although the training time of even the fastest SVMs can be exceedingly slow, they are accurate, and exemplary in their ability to model complex nonlinear decision boundaries. They are much less prone to overfitting as compared with other methods. SVM initiates also provide a compact description of the learned model. SVMs can be used for prediction, along with classification. They have been applied to several areas, including handwritten digit recognition, object recognition, and speaker identification, as well as benchmark time-series prediction tests.

2.4. Radial basis function algorithm

The Radial Basis Function (RBF) is a neural network algorithm which requires less computing time for network training [10]. It consists of three layers: input layer, hidden layer, and output layer, as shown in Fig. 5 The nodes within each layer are fully connected to the previous layer [15]. The input variables in the input layer pass directly to the hidden layer without weights. The transfer functions of the hidden nodes are RBFs. The parameters associated with the RBFs are optimized during the network training. These parameter values are not necessarily the same throughout the network, nor are they directly related to or constrained by the actual training vectors. When the training vectors are assumed to be accurate, it is desirable to perform a smooth interpolation between them, then linear combinations of RBFs can be found which give no error at the training vectors. The method of fitting RBFs to data, for function approximation, is closely related to distance weighted regression.

3. Chronic kidney disease

CKD progression can be considered as a function of various parameters including underlying renal diseases, blood pressure, hypertension, proteinuria, and age. Early diagnosis of the CKD requires great attention among physicians, especially in determining the appropriate time to apply medical treatments and to control identified risk factors that reflect on the disease progression to End Stage Renal Disease (ESRD), such hypertension, proteinuria, and hyperphosphatemia.

3.1. Stages of chronic kidney disease

The stages of Chronic Kidney Disease (CKD) are mainly based on measured or estimated Glomerular Filtration Rate (eGFR). There are five stages, but kidney function is normal in Stage 1, and minimally reduced in Stage 2.

The KDOQI (Kidney Disease Outcomes Quality Initiative) stages of kidney disease are (see Table 1):

Definition of chronic: Labelling someone as having CKD requires two samples at least 90 days apart. Historical values can be used. The estimated Glomerular Filtration Rate (eGFR) depends on creatinine measurement, sex, race and age. One of the most accurate methods to calculate the eGFR is the Modification of Diet in Renal Disease (MDRD) [12].

$$eGFR = 186 \times (Creatinine / 88.4)^{-1.154} \times (Age)^{-0.203} \times (0.742 \text{ if female}) \times (1.210 \text{ if black})$$

4. Results

The following analysis was performed using the DTREG Predictive Modeling System. The experimental comparison of the utilized algorithms was done based on the performance measures of classification accuracy and execution time. Model testing and validation was performed by a V-fold cross validation technique. Missing predictor variable values were replaced by medians during the analysis.

The dataset used in the analysis consisted of 361 CKD Indian patients and contained 25 variables (11 numerical, 14 categorical). Before starting the analysis, eGFR were calculated to identify the severity stage of the kidney disease for each patient by applying the eGFR formula described in section 3 on the used dataset. Dataset source is available on UCI machine learning repository.

4.1. Variables description

Follow in Table 2 is a description of variables used in the analysis, which contains the variable name, class, type, number of Missing rows, and categories, according to DTREG output.

4.2. Sensitivity and specificity

Sensitivity, Specificity and Accuracy percentage were employed to evaluate the performance of the utilized classification algorithms.

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN}$$

Table 1
CKD Stages according to GFR measurement value.

Stage	GFR	Description	Treatment stage
1	90+	Normal kidney function but urine findings or structural abnormalities or genetic trait point to kidney disease	Observation, control of blood pressure.
2	60–89	Mildly reduced kidney function, and other findings (as for stage 1) point to kidney disease	Observation, control of blood pressure and risk factors.
3A	45–59	Moderately reduced kidney function	Observation, control of blood pressure and risk factors.
3B	30–44		
4	15–29	Severely reduced kidney function	Planning for end stage renal failure.
5	< 15 or on dialysis	Very severe, or end stage kidney failure (sometimes call established renal failure)	Treatment choices.

Table 2
Variables description used in the analysis.

Ser	Variable	Class	Type	Missing rows	Categories
1	Age	Predictor	Numerical	0	66
2	Blood_Pressure	Predictor	Numerical	5	10
3	Specific_Gravity	Predictor	Categorical	7	5
4	Albumin	Predictor	Categorical	3	6
5	Sugar	Predictor	Categorical	4	6
6	Red_Blood_Cells	Predictor	Categorical	0	2
7	Pus_Cell	Predictor	Categorical	2	2
8	Pus_Cell_Clumps	Predictor	Categorical	4	2
9	Bacteria	Predictor	Categorical	4	2
10	Blood_Glucose_Random	Predictor	Numerical	5	143
11	Blood_Urea	Predictor	Numerical	3	116
12	Serum_Creatinine	Predictor	Numerical	0	83
13	Sodium	Predictor	Numerical	0	34
14	Potassium	Predictor	Numerical	7	39
15	Hemoglobin	Predictor	Numerical	8	111
16	Packed_Cell_Volume	Predictor	Numerical	10	41
17	White_Blood_Cell_Count	Predictor	Numerical	7	86
18	Red_Blood_Cell_Count	Predictor	Numerical	7	45
19	Hypertension_	Predictor	Categorical	2	2
20	Diabetes_Mellitus_	Predictor	Categorical	2	2
21	Coronary_Artery_Disease_	Predictor	Categorical	2	2
22	Sex_	Predictor	Categorical	1	2
23	Pedal_Edema_	Predictor	Categorical	1	2
24	Anemia_	Predictor	Categorical	1	2
25	CKD_STAGE_	Target	Categorical	0	5

Table 3
Summary of Algorithms classification outputs for classifying the CKD patients with stage 1 disease severity.

Sensitivity & Specificity – CKD Stage 1	PNN		SVM		RBF		MLP	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
Total records	361	361	361	361	361	361	361	361
Accuracy	100.00%	99.72%	99.45%	91.14%	98.89%	95.84%	77.56%	77.29%
True positive (TP)	79 (21.88%)	79 (21.88%)	78 (21.61%)	57 (15.79%)	76 (21.05%)	72 (19.94%)	0 (0.00%)	11 (3.05%)
True negative (TN)	282 (78.12%)	281 (77.84%)	281 (77.84%)	272 (75.35%)	281 (77.84%)	274 (75.90%)	280 (77.56%)	268 (74.24%)
False positive (FP)	0 (0.00%)	1 (0.28%)	1 (0.28%)	10 (2.77%)	1 (0.28%)	8 (2.22%)	2 (0.55%)	14 (3.88%)
False negative (FN)	0 (0.00%)	0 (0.00%)	1 (0.28%)	22 (6.09%)	3 (0.83%)	7 (1.94%)	79 (21.88%)	68 (18.84%)
Sensitivity	100.00%	100.00%	98.73%	72.15%	96.20%	91.14%	0.00%	13.92%
Specificity	100.00%	99.65%	99.65%	96.45%	99.65%	97.16%	99.29%	95.04%
Geometric mean of sensitivity and specificity	100.00%	99.82%	99.19%	83.42%	97.91%	94.10%	0.00%	36.38%
Positive Predictive Value (PPV)	100.00%	98.75%	98.73%	85.07%	98.70%	90.00%	0.00%	44.00%
Negative Predictive Value (NPV)	100.00%	100.00%	99.65%	92.52%	98.94%	97.51%	77.99%	79.76%
Geometric mean of PPV and NPV	100.00%	99.37%	99.19%	88.72%	98.82%	93.68%	0.00%	59.24%
Precision	100.00%	98.75%	98.73%	85.07%	98.70%	90.00%	0.00%	44.00%
Recall	100.00%	100.00%	98.73%	72.15%	96.20%	91.14%	0.00%	13.92%
F-Measure	1	0.9937	0.9873	0.7808	0.9744	0.9057	0	0.2115

Table 4
Summary of Algorithms classification outputs for classifying the CKD patients with stage 2 disease severity.

Sensitivity & Specificity – CKD Stage 2	PNN		SVM		RBF		MLP	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
Total records	361	361	361	361	361	361	361	361
Accuracy	100.00%	98.89%	99.45%	85.60%	93.63%	90.58%	72.85%	71.47%
True positive (TP)	81 (22.44%)	78 (21.61%)	80 (22.16%)	53 (14.68%)	80 (22.16%)	75 (20.78%)	75 (20.78%)	58 (16.07%)
True negative (TN)	280 (77.56%)	279 (77.29%)	279 (77.29%)	256 (70.91%)	258 (71.47%)	252 (69.81%)	188 (52.08%)	200 (55.40%)
False positive (FP)	0 (0.00%)	1 (0.28%)	1 (0.28%)	24 (6.65%)	22 (6.09%)	28 (7.76%)	92 (25.48%)	80 (22.16%)
False negative (FN)	0 (0.00%)	3 (0.83%)	1 (0.28%)	28 (7.76%)	1 (0.28%)	6 (1.66%)	6 (1.66%)	23 (6.37%)
Sensitivity	100.00%	96.30%	98.77%	65.43%	98.77%	92.59%	92.59%	71.60%
Specificity	100.00%	99.64%	99.64%	91.43%	92.14%	90.00%	67.14%	71.43%
Geometric mean of sensitivity and specificity	100.00%	97.96%	99.20%	77.35%	95.40%	91.29%	78.85%	71.52%
Positive Predictive Value (PPV)	100.00%	98.73%	98.77%	68.83%	78.43%	72.82%	44.91%	42.03%
Negative Predictive Value (NPV)	100.00%	98.94%	99.64%	90.14%	99.61%	97.67%	96.91%	89.69%
Geometric mean of PPV and NPV	100.00%	98.84%	99.20%	78.77%	88.39%	84.33%	65.97%	61.40%
Precision	100.00%	98.73%	98.77%	68.83%	78.43%	72.82%	44.91%	42.03%
Recall	100.00%	96.30%	98.77%	65.43%	98.77%	92.59%	92.59%	71.60%
F-Measure	1	0.975	0.9877	0.6709	0.8743	0.8152	0.6048	0.5297

Table 5
Summary of Algorithms classification outputs for classifying the CKD patients with stage 3 disease severity.

Sensitivity & Specificity - CKD Stage 3	PNN		SVM		RBF		MLP	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
Total records	361	361	361	361	361	361	361	361
Accuracy	100.00%	96.95%	100.00%	73.68%	94.18%	92.52%	82.55%	77.29%
True positive (TP)	82 (22.71%)	81 (22.44%)	82 (22.71%)	48 (13.30%)	62 (17.17%)	61 (16.90%)	63 (17.45%)	56 (15.51%)
True negative (TN)	279 (77.29%)	269 (74.52%)	279 (77.29%)	218 (60.39%)	278 (77.01%)	273 (75.62%)	235 (65.10%)	223 (61.77%)
False positive (FP)	0 (0.00%)	10 (2.77%)	0 (0.00%)	61 (16.90%)	1 (0.28%)	6 (1.66%)	44 (12.19%)	56 (15.51%)
False negative (FN)	0 (0.00%)	1 (0.28%)	0 (0.00%)	34 (9.42%)	20 (5.54%)	21 (5.82%)	19 (5.26%)	26 (7.20%)
Sensitivity	100.00%	98.78%	100.00%	58.54%	75.61%	74.39%	76.83%	68.29%
Specificity	100.00%	96.42%	100.00%	78.14%	99.64%	97.85%	84.23%	79.93%
Geometric mean of sensitivity and specificity	100.00%	97.59%	100.00%	67.63%	86.80%	85.32%	80.44%	73.88%
Positive Predictive Value (PPV)	100.00%	89.01%	100.00%	44.04%	98.41%	91.04%	58.88%	50.00%
Negative Predictive Value (NPV)	100.00%	99.63%	100.00%	86.51%	93.29%	92.86%	92.52%	89.56%
Geometric mean of PPV and NPV	100.00%	94.17%	100.00%	61.72%	95.82%	91.95%	73.81%	66.92%
Precision	100.00%	89.01%	100.00%	44.04%	98.41%	91.04%	58.88%	50.00%
Recall	100.00%	98.78%	100.00%	58.54%	75.61%	74.39%	76.83%	68.29%
F-Measure	1	0.9364	1	0.5026	0.8552	0.8188	0.6667	0.5773

Table 6
Summary of Algorithms classification outputs for classifying the CKD patients with stage 4 disease severity.

Sensitivity & Specificity – CKD Stage 4	PNN		SVM		RBF		MLP	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
Total records	361	361	361	361	361	361	361	361
Accuracy	100.00%	99.72%	100.00%	80.33%	99.45%	96.95%	86.43%	85.32%
True positive (TP)	57 (15.79%)	56 (15.51%)	57 (15.79%)	15 (4.16%)	56 (15.51%)	49 (13.57%)	12 (3.32%)	9 (2.49%)
True negative (TN)	304 (84.21%)	304 (84.21%)	304 (84.21%)	275 (76.18%)	303 (83.93%)	301 (83.38%)	300 (83.10%)	299 (82.83%)
False positive (FP)	0 (0.00%)	0 (0.00%)	0 (0.00%)	29 (8.03%)	1 (0.28%)	3 (0.83%)	4 (1.11%)	5 (1.39%)
False negative (FN)	0 (0.00%)	1 (0.28%)	0 (0.00%)	42 (11.63%)	1 (0.28%)	8 (2.22%)	45 (12.47%)	48 (13.30%)
Sensitivity	100.00%	98.25%	100.00%	26.32%	98.25%	85.96%	21.05%	15.79%
Specificity	100.00%	100.00%	100.00%	90.46%	99.67%	99.01%	98.68%	98.36%
Geometric mean of sensitivity and specificity	100.00%	99.12%	100.00%	48.79%	98.96%	92.26%	45.58%	39.41%
Positive Predictive Value (PPV)	100.00%	100.00%	100.00%	34.09%	98.25%	94.23%	75.00%	64.29%
Negative Predictive Value (NPV)	100.00%	99.67%	100.00%	86.75%	99.67%	97.41%	86.96%	86.17%
Geometric mean of PPV and NPV	100.00%	99.84%	100.00%	54.38%	98.96%	95.81%	80.76%	74.43%
Precision	100.00%	100.00%	100.00%	34.09%	98.25%	94.23%	75.00%	64.29%
Recall	100.00%	98.25%	100.00%	26.32%	98.25%	85.96%	21.05%	15.79%
F-Measure	1	0.9912	1	0.297	0.9825	0.8991	0.3288	0.2535

Table 7
Summary of Algorithms classification outputs for classifying the CKD patients with stage 5 disease severity.

Sensitivity & Specificity - CKD Stage 5	PNN		SVM		RBF		MLP	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation
Total records	361	361	361	361	361	361	361	361
Accuracy	100.00%	98.06%	100.00%	90.58%	100.00%	98.06%	95.84%	91.69%
True positive (TP)	62 (17.17%)	55 (15.24%)	62 (17.17%)	46 (12.74%)	62 (17.17%)	57 (15.79%)	58 (16.07%)	52 (14.40%)
True negative (TN)	299 (82.83%)	299 (82.83%)	299 (82.83%)	281 (77.84%)	299 (82.83%)	297 (82.27%)	288 (79.78%)	279 (77.29%)
False positive (FP)	0 (0.00%)	0 (0.00%)	0 (0.00%)	18 (4.99%)	0 (0.00%)	2 (0.55%)	11 (3.05%)	20 (5.54%)
False negative (FN)	0 (0.00%)	7 (1.94%)	0 (0.00%)	16 (4.43%)	0 (0.00%)	5 (1.39%)	4 (1.11%)	10 (2.77%)
Sensitivity	100.00%	88.71%	100.00%	74.19%	100.00%	91.94%	93.55%	83.87%
Specificity	100.00%	100.00%	100.00%	93.98%	100.00%	99.33%	96.32%	93.31%
Geometric mean of sensitivity and specificity	100.00%	94.19%	100.00%	83.50%	100.00%	95.56%	94.92%	88.47%
Positive Predictive Value (PPV)	100.00%	100.00%	100.00%	71.88%	100.00%	96.61%	84.06%	72.22%
Negative Predictive Value (NPV)	100.00%	97.71%	100.00%	94.61%	100.00%	98.34%	98.63%	96.54%
Geometric mean of PPV and NPV	100.00%	98.85%	100.00%	82.46%	100.00%	97.47%	91.05%	83.50%
Precision	100.00%	100.00%	100.00%	71.88%	100.00%	96.61%	84.06%	72.22%
Recall	100.00%	88.71%	100.00%	74.19%	100.00%	91.94%	93.55%	83.87%
F-Measure	1	0.9402	1	0.7302	1	0.9421	0.8855	0.7761

Where:

- TP is Number of true positive classification cases
- FN is Number of false negative classification cases
- TN is Number of true negative classification cases
- FP is Number of false positive classification cases

Algorithm classification results are exhibited in Table (3) for patients with CKD stage 1 disease severity, and shows that PNN Algorithm gives the highest classification accuracy of 99.7%, Precision 98.7% and F-Measure 99.37% as compared with all other algorithm results.

Algorithm classification results are displayed in Table (4) for patients with CKD stage 2 disease severity, and shows that PNN algorithm

Table 8
Overall classification accuracy percentage and analysis execution time for all used algorithms.

Algorithm	Overall Accuracy	Total Execution Time
PNN	96.7%	0:00:12
SVM	60.7%	0:00:40
RBF	87%	2:29.6
MLP	51.5%	00:03.5

Fig. 6. CKD Stage 1 Classification accuracy % for all used algorithms.

Fig. 7. CKD Stage 2 Classification accuracy % for all used algorithms.

Fig. 8. CKD Stage 3 Classification accuracy % for all used algorithms.

provides a highest classification accuracy 98.9%, Precision 98.7% and F– Measure 97.5% as compared with all other algorithm results.

Algorithm classification results in Table 5 for the patients with CKD stage 3 disease severity shows that the PNN algorithm gives highest classification accuracy with percentage 96.9%, Precision 89% and F-Measure 93.6% as compared with all other algorithm results.

Algorithm classification results in Table 6 for patients with CKD stage 4 disease severity shows that the PNN algorithm gives the highest classification accuracy with a percentage 99.7%, Precision 100% and F-Measure 99.1% as compared with all other algorithm results.

Fig. 9. CKD Stage 4 Classification accuracy % for all used algorithms.

Fig. 10. CKD Stage 5 Classification accuracy % for all used algorithms.

Fig. 11. Overall Classification accuracy % for all used algorithms.

Algorithm classification results in Table 7 for patients with CKD stage 5 disease severity shows that the PNN algorithm provides the highest classification accuracy with percentage 98%, Precision 100% and F-Measure 94% as compared with all other algorithm results.

5. Discussion

Our study suggests that the severity stages of chronic kidney disease

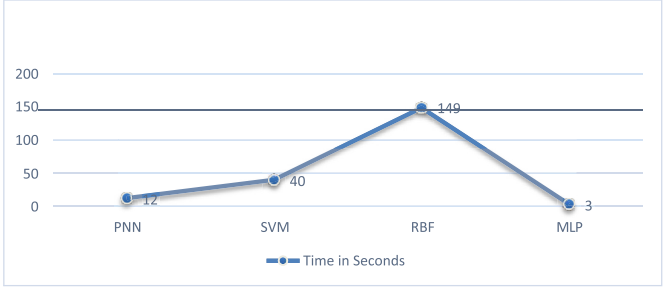

Fig. 12. Total analysis execution time in seconds for all used algorithms.

Table 9
Overall importance of predictive variables in building the classification model.

Predictive Variables	Importance %
Serum Creatinine	100.000
Blood Urea	38.455
Albumin	22.034
Age	20.394
Hemoglobin	10.359
Hypertension	9.256

can be accurately classified and predicted by using data mining techniques. The above-mentioned results from Tables 3–8 suggest that the Probabilistic Neural Networks and Radial Basis Function techniques are providing the most accurate classification, precision, and highest F-Measure, comparable with the Support Vector Machine and Multilayer Perceptron techniques. On the other hand, the Radial Basis Function technique requires more processing time than the Probabilistic Neural Network technique.

The Probabilistic Neural Network technique gives best classification results as compared with all other used techniques in classifying CKD stages (see Figs. 6-12), as follows:

- Accuracy percentage 99.7%, Precision 98.7% and F-Measure 99.37% in classifying Stage 1 CKD patients.
- Accuracy percentage 98.9%, Precision 98.7% and F-Measure 97.5% in classifying Stage 2 CKD patients.
- Accuracy percentage 96.9%, Precision 89% and F-Measure 93.6% in classifying Stage 3 CKD patients.
- Accuracy percentage 99.7%, Precision 100% and F-Measure 99.1% in classifying Stage 4 CKD patients.
- Accuracy percentage 98%, Precision 100% and F-Measure 94% in classifying Stage 5 CKD patients.

The results of Table 9 shows that the following predictor variables are the most important variables during construction of the classification model: Serum Creatinine (100%), Blood Urea (38.5%), Albumin (22%), Age (20%), Hemoglobin (10%) and Hypertension (9%).

The Probabilistic Neural Networks technique can be readily implemented for classifying the severity stages of chronic kidney disease patients.

6. Conclusion

Finally, as observed from Table 8, the Probabilistic Neural Networks algorithm gives the highest overall classification accuracy percentage of 96.7%, compared to other algorithms in classifying the stages of CKD patients. On the other hand, the Multilayer Perceptron requires a minimum execution time (3 s) whereas the Probabilistic Neural Network requires 12 s to finalize the analysis.

These algorithms have been compared with classification accuracy based on correctly classified stages of CKD patients, time taken to construct the model, and time taken to test the model. The Probabilistic Neural Networks algorithm yields a better classification accuracy and prediction performance to predict the stages of chronic kidney disease patients.

Significance Statement: The current study applied four data mining algorithms on a clinical/laboratory dataset consisting of 361 chronic kidney disease patients. The results of the addressed algorithms have been compared to define the most accurate algorithm results in classifying the severity stage of CKD. This study recommends that the Probabilistic Neural Networks algorithm is the best algorithm that can be used by physicians in order to eliminate diagnostic and treatment errors.

Conflicts of interest

No competing interest exists.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2019.100178.

References

[3] Ramchoun H, Amine M, Idrissi J, Ghanou Y, Ettaouil M. Multilayer Perceptron: Architecture optimization and training. IJIMAI 2016;4(1):26–30.
[4] Dhamodharan S. Liver disease prediction using bayesian classification. 4th national conference on advance computing. Application Technologies; 2014, May.
[5] Joshi J, Doshi R, Patel J. Diagnosis and prognosis breast cancer using classification rules. Int J Eng Res Gen Sci 2014;2(6):315–23.
[6] Solanki AV. Data mining techniques using WEKA classification for sickle cell disease. Int J Comput Sci Inf Technol 2014;5(4):5857–60.
[7] Aljahdali S, Hussain SN. Comparative prediction performance with support vector machine and random forest classification techniques. Int J Comput Appl 2013;69(11).
[9] Couser WG, Remuzzi G, Mendis S, Tonelli M. The contribution of chronic kidney disease to the global burden of major noncommunicable diseases. Kidney Int 2011;80(12):1258–70.
[10] Schaback R. A practical guide to radial basis functions. Electronic Resource 2007;11.
[12] Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, et al. National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Ann Intern Med 2003;139(2):[137]–47].
[13] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge university press; 2000.
[15] Fukumizu K. Active learning in multilayer perceptrons. Advances in neural information processing systems. 1996. p. 295–301.
[16] Specht DF. Probabilistic neural networks. Neural Network 1990;3(1):109–18.
[17] Yu W, Liu T, Valdez R, Gwinn M, Khoury MJ. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre- diabetes. BMC Med Inf Decis Mak 2010;10(1):16.
[18] Dessai ISF. Intelligent heart disease prediction system using probabilistic neural network. Int J Adv Comp Theory Eng (IJACTE) 2013;2(3):2319–526.
[19] Magnin B, Mesrob L, Kinkingnéhun S, Pélégrini-Issac M, Colliot O, Sarazin M, et al. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 2009;51(2):73–83.
[20] Cao Y, Hu ZD, Liu XF, Deng AM, Hu CJ. An MLP classifier for prediction of HBV-induced liver cirrhosis using routinely available clinical parameters. Dis Markers 2013;35(6):653–60.

Review Nephrology (Carlton). 2009 Jun;14(4):367-73. doi: 10.1111/j.1440-1797.2009.01113.x.

Review article: Early detection of chronic kidney disease in Australia: which way to go?

Timothy Mathew [1], Olivia Corso

Affiliations

PMID: 19563377 DOI: 10.1111/j.1440-1797.2009.01113.x

Abstract

Early detection of chronic kidney disease (CKD) followed by appropriate clinical management appears the only means by which the increasing burden on the health-care system and affected individuals will be reduced. The asymptomatic nature of CKD means that early detection can only occur through testing of individuals. The World Health Organization principles of screening for chronic disease can now be largely fulfilled for CKD. The risk groups to be targeted, the expected yield and the tests to be performed are reviewed. For a screening programme to be sustainable it must carry a greater benefit than risk of harm for the participant and be shown to be cost-effective from the community point of view. Whole population screening for CKD is impractical and is not cost-effective. Screening of those at increased risk of CKD could occur either through special events run in the community, workplace or in selected locations such as pharmacies or through opportunistic screening of high-risk people in general practice. Community screening programmes targeted at known diabetics, hypertensives and those over 55 years have been described to detect 93% of all CKD in the community. The yield of CKD stages 3-5 from community screening has been found to vary from 10% to 20%. The limitations of screening programmes including the cost and recruitment bias are discussed. The most sustainable and likely the most cost-efficient model appears to be opportunistic general practice screening. The changing structure of general practice in Australia lends itself well to the requirements for early detection of CKD.

Related information

MedGen

LinkOut - more resources

Full Text Sources

Ovid Technologies, Inc.
Wiley

Medical

Genetic Alliance
MedlinePlus Health Information