

**Application of Deep Learning to Classify Disease Progression in Patients with
Autosomal Dominant Polycystic Kidney Disease**

Hyun Bae Jang

Division of Experimental Medicine

Faculty of Medicine

McGill University

Montréal, QC, Canada

*A thesis submitted to McGill University in partial fulfillment of the requirements
of the degree of Master of Science (M.Sc.) in Experimental Medicine*

©Copyright Hyun Bae Jang, June 2023

TABLE OF CONTENTS

Table of Contents	ii
Abstract.....	vi
Abstrait	viii
Acknowledgement.....	x
Contribution of Authors.....	xii
List of Figures and Tables.....	xiv
Abbreviations and Acronyms	xvi
1. Introduction.....	1
1.1 Study Rationale.....	1
1.2 Objective	1
2. Comprehensive Review of Relevant Literature	1
2.1. Autosomal Dominant Polycystic Kidney Disease.....	1
2.2. Normal Kidney	2
2.3. Clinical Manifestations.....	2
2.4. Diagnosis	3
2.4.1. Imaging Tools.....	3
2.4.2. Diagnostic Guidelines	4
2.5. Prognosis.....	5
2.5.1. Factors Associated with Disease Progression.....	5

2.5.2. Prognostic Biomarkers	5
2.5.3. Prognostic Imaging Tools.....	6
2.6. Classification of Risk	7
2.7. Limitations of the MIC.....	9
2.8. Available Treatment.....	12
2.8.1. Non-Pharmacological Interventions.....	12
2.8.2. Pharmacological Interventions.....	14
2.9. Risk Associated with the Current Prognostic Tool and Clinical Management	17
2.10. Current Literature on eTKV Calculation Using the Ellipsoid Equation	17
2.11. Alternative Conventional Models to Predict Disease Progression of ADPKD.....	19
2.12. Proposed Solution	22
2.13. Machine Learning.....	23
2.13.1. Introduction.....	23
2.13.2. Supervised Learning.....	23
2.13.3. Unsupervised Learning	25
2.13.4. Semi-supervised Learning.....	26
2.13.5. Reinforcement Learning	26
2.14. Stages of Machine Learning Model Development	26
2.15. Performance Evaluation Metrics in Machine Learning.....	29
2.16. Current Literature on Machine Learning and Predicting Outcome in ADPKD..	31
3. Objective and Hypothesis.....	34

3.1.	Objective	34
3.2.	Study Aims and Hypotheses.....	34
4.	Methods.....	35
4.1.	Study Cohort	35
4.2.	Variables of Interest for Machine Learning.....	35
4.3.	Clinical Data Collection and Imaging Information Extraction.....	36
4.4.	Clinical MRI Image Extraction	37
4.5.	Manual Segmentation.....	38
4.6.	Computer Features	39
4.7.	Data Pre-Processing.....	39
4.8.	Machine Learning Model Development.....	41
4.8.1.	Clinical Model	41
4.8.2.	Deep Model Architecture	41
4.9.	Statistical Analysis	43
5.	Results	44
5.1.	Study Participant Characteristics	45
5.2.	Manual Kidney and Cyst Segmentations of MRI Images.....	58
5.3.	Aim 1: Manual Segmentation vs. Ellipsoid Equation.....	60
5.4.	Aim 2: Machine Learning Model Performance	64
5.4.1.	Random Forest Approach Using Only Clinical Variables.....	64

5.4.2. Support Vector Machine Approach Using Only Clinical Variables	64
5.4.3. Deep Learning Approach Using Clinical Variables and Imaging Features	64
6. Discussion	67
6.1. Baseline Characteristics	68
6.2. Manual Segmentation vs. Ellipsoid Equation	69
6.3. Machine Learning Models	71
6.3.1. Clinical Model	71
6.3.2. Deep Model	73
6.4. Strengths	74
6.5. Limitations	74
7. Future Steps	75
8. Conclusions	76
9. References	77

ABSTRACT

Autosomal dominant polycystic kidney disease (ADPKD) is the most common monogenic cause of kidney failure. It is associated with progressive kidney cyst growth eventually leading to kidney failure in many patients. The standard clinical risk stratification tool uses estimated total kidney volume (eTKV) adjusted for height and age; however, it does not consider the complex interplay between demographic, imaging, or clinical features that may be pertinent. We explored whether a machine learning strategy that integrates clinical and imaging features may achieve a more accurate and reliable prediction of renal disease progression in ADPKD. Our study cohort consisted of 145 patients from the MUHC PKD clinic with a confirmed diagnosis of ADPKD and at least one magnetic resonance imaging (MRI) scan where manual segmentation was performed for total kidney volume (TKV) and total cyst volume. Our first aim was to compare eTKV with the manual segmented (i.e., ground truth) TKV. Pearson's correlation coefficient between the TKVs was 0.96 ($p < 0.05$). However, the Bland-Altman analysis displayed wide limits of agreement [-40.43%, 25.15%] and 25 patients (17.2%) with greater than or equal to a 20% difference. In total, 24 patients (16.6%) were misclassified by one risk class. Of those, the clinically significant misclassifications that affect disease-modifying therapy were 8 (5.5%). For our second aim, we developed three different machine and deep learning models using clinical features and MRI slices as inputs for the outcome of rapid estimated glomerular filtration rate (eGFR) decline, defined as 4 mL/min/1.73m² per year or greater. The weighted-average F1 scores, a machine learning metric that measures the model's accuracy by combining precision and recall, from the random forest and support vector machine classifiers were 0.68 and 0.81, respectively. A deep-learning model provided the most accurate result with a weighted-average F1 score of 0.87 and true positive and true negative values of 0.93 and 0.80, respectively. Our

study determined that risk stratification using eTKV is useful but can misclassify risk in many patients. A deep learning approach that integrates clinical information with MRI can successfully classify patients most at risk for renal progression. Our model needs to be further validated in external cohorts but offers the potential to better identify patients most at risk of renal progression with ADPKD.

ABSTRAIT

La polykystose rénale autosomique dominante (PKRAD) est la cause monogénique la plus fréquente d'insuffisance rénale. Elle est associée à une croissance progressive des kystes rénaux conduisant éventuellement à une insuffisance rénale chez de nombreux patients. L'outil standard de stratification du risque clinique utilise le volume rénal total estimé (VTKe) ajusté en fonction de la taille et de l'âge ; cependant, il ne tient pas compte de l'interaction complexe entre les caractéristiques démographiques, d'imagerie ou cliniques qui peuvent être pertinentes. Nous avons exploré si une stratégie d'apprentissage automatique qui intègre des fonctionnalités cliniques et d'imagerie peut permettre une prédiction plus précise et fiable de la progression de la maladie rénale dans la PKRAD. Notre cohorte d'étude était composée de 145 patients de la clinique PKD du Centre Universitaire de Santé McGill (CUSM) avec un diagnostic confirmé de PKRAD et au moins d'une imagerie par résonance magnétique (IRM) où une segmentation manuelle a été effectuée pour le volume rénal total (VTK) et le volume total du kyste. Notre premier objectif était de comparer VTKe avec le VTK manuel segmenté (c'est-à-dire la vérité terrain). Le coefficient de corrélation de Pearson entre les VTK était de 0.96 ($p < 0.05$). Cependant, l'analyse de Bland-Altman a montré de larges limites d'accord [-40.43%, 25.15%] et 25 patients (17.2%) avec une différence supérieure ou égale à 20%. Au total, 24 patients (16.6%) ont été mal classés dans la classe d'un risque. Parmi eux, le taux des erreurs de classification cliniquement significatives affectant la thérapie modificatrice de la maladie était de 8 (5.5%). Pour notre deuxième objectif, nous avons développé trois modèles différents d'apprentissage automatique et profond en utilisant des caractéristiques cliniques et des tranches d'IRM comme entrées pour le résultat du déclin rapide du débit de filtration glomérulaire estimé (eGFR), défini comme 4 ml/min/1.73 m² par an ou plus. Les scores F1 moyens pondérés, une métrique

d'apprentissage automatique qui mesure la précision du modèle en combinant précision et rappel, à partir des classificateurs de la forêt aléatoire et du vecteur de support étaient de 0.68 et 0.81, respectivement. Un modèle d'apprentissage en profondeur a fourni le résultat le plus précis avec un score F1 moyen pondéré de 0.87 et des valeurs vraies positives et vraies négatives de 0.93 et 0.80, respectivement. Notre étude a déterminé que la stratification du risque à l'aide de le VTKe est utile, mais peut entraîner une mauvaise classification du risque chez de nombreux patients. Une approche d'apprentissage en profondeur qui intègre les informations cliniques à l'IRM peut classer avec succès les patients les plus à risque de progression rénale. Notre modèle doit être davantage validé dans des cohortes externes, mais offre le potentiel de mieux identifier les patients les plus à risque de progression rénale avec ADPKD.

ACKNOWLEDGEMENT

I want to thank Dr. Ahsan Alam for taking me on as his mentee and providing me with this research opportunity in the field of nephrology. Starting from PHGY 508 course during my undergraduate years, Dr. Alam has continuously guided me to flourish my interest and knowledge in clinical research and Autosomal Dominant Polycystic Kidney Disease (ADPKD). Through the clinical shadowing opportunities he offered, I got to interact with patients and better understand the importance of my current research. Throughout our many discussions, Dr. Alam always attentively listened and carefully considered my ideas; he made me feel like a friend and colleague. Most importantly, he taught me the importance of conducting research with the highest ethical standards. As an aspiring clinician and researcher, Dr. Alam inspires me to approach every research study, patient, colleague, mentee, family, and friend with the same care, passion, empathy, composure, and professionalism that he displays daily.

Thank you, Farhad Maleki, Shiva Heydari, and Keyhan Najafian, for developing the machine learning models, reviewing my thesis sections related to machine learning, and clarifying any misconceptions. Through your help, I acquired so much knowledge about machine learning in clinical studies. You equipped me with the lens that now allows me to carefully see through the supposedly high accuracy and F1 scores and better distinguish a “good” model from a “bad” model.

I would also like to extend my gratitude to Dr. Caroline Reinhold, Evan McNabb, and the AIPHL members —Rita Zakarian and Nikesh Muthukrishnan — for helping me with image extractions, adjudication of my manual segmentations, and any further discussions we had regarding patient MRI images. I would also like to give special thanks to Dr. Tomoko Takano, a true veteran in the Division of Experimental Medicine, for reminding me of the mandatory steps

that need to be taken and answering any questions that Dr. Alam and I had. Without you, I cannot imagine the number of “unsatisfactory” I would have received from the division. I also want to thank the committee members and my academic advisor, Dr. Thomas Mavrakanas, Dr. Ives Levesque, and Dr. Evelyn Vinet. You challenged me with questions and critiques and provided incredible insight that rerouted my research to where it is now.

Finally, I would like to thank my family (Soo, Ji, Eddie, and Jeany), friends, and God for always supporting me throughout challenging times, whether financially or physically or emotionally. I would not know where I would be without you all.

CONTRIBUTION OF AUTHORS

Shiva Heydari: Developed the clinical random forest and support vector machine models.

Obtained the evaluation metric scores for the models. Reviewed and edited Chapters 1.12., 1.13., 1.14., 3.6., 3.7., 3.8.1., 4.4.1., 4.4.2., and 5.3.1.

Keyhan Najafian: Developed the deep learning model. Obtained the evaluation metric scores for the model. Designed Figures 3, 10, 11, and 12. Wrote Chapter 3.8.2.

Rita Zakarian and Nikesh Muthukrishnan: Extracted MRI images from the MUHC Oacis portal and anonymized the files in preparation for manual segmentation.

Evan McNabb: Helped troubleshoot a minor issue regarding the 3D Slicer program and the MRI scan.

Farhad Maleki: Consulted on which machine learning approach would be most appropriate for each sub-aim. Recommended Shiva Heydari and Keyhan Najafian for collaboration.

Caroline Reinhold: Provided mentorship as the co-supervisor via meetings and discussions. Trained the main author on how to maneuver through the MRI scans and discern kidneys and cysts from other organs and physiological structures. Adjudicated the final manual segmentations.

Ahsan Alam: Provided mentorship as the primary supervisor via meetings and discussions. Co-designed the research project with the author. Adjudicated the final manual segmentations. Reviewed and edited the thesis.

Hyun Bae Jang: Co-designed the research project with Dr. Ahsan Alam. Completed the literature review. Filtered eligible patients from the MUHC PKD clinic based on the selection criteria. Extracted clinical data of all eligible patients from Oacis. Completed the manual segmentation task of baseline and follow-up T-2 weighted MRIs. Discussed and troubleshoot with

Shiva Heydari and Keyhan Najafian throughout the development and optimization steps of the clinical machine learning and deep models. Wrote all chapters of the thesis and designed all figures and tables, excluding Chapter 3.8.2. and Figures 3, 10, 11, and 12. Reviewed and edited Chapter 3.8.2.

LIST OF FIGURES AND TABLES

Figure 1. The measurements of the maximal kidney a) coronal length, b) sagittal length, and c) width and depth using axial planes in an MRI scan.....	7
Figure 2. Unique Clinical Presentations of ADPKD in Patients from the MUHC PKD Clinic...	12
Figure 3. The confusion matrix depicting the performance of the predictive model by Raj et al. ¹⁰⁹	33
Figure 4. The architecture of the deep learning model for processing 2D images (2DExNet), their corresponding clinical features (FuseNet), and the classifier to categorize patients as either rapid or non-rapid progressors.	43
Figure 5. Enrollment, Exclusion, and Inclusion Based on Screening Criteria	46
Figure 6. Mayo Imaging Class, TKV, and Annual Rate of eGFR Decline, and MRI Segmentations. The green label outlines the kidney parenchyma, while the red label outlines only the cysts.....	59
Figure 7. Linear Regression Analysis Comparing the Height-Adjusted Segmented TKV from Manual Segmentation with Height-Adjusted eTKV from the Ellipsoid Equation.....	61
Figure 8. Giavarina Plot Comparing the Height-Adjusted Segmented TKV with Height-Adjusted eTKV.....	62
Figure 9. Confusion Matrix Highlighting the Misclassifications Resulting from the Use of the Ellipsoid Equation.....	63
Figure 10. Inconsistency in the Ellipsoid Equation is Observed in Two Patients with Similar Age, Height-Adjusted TKV, and Identical Mayo Imaging Class.....	63
Figure 11. Confusion matrix comparing the predicted classification of rapid and non-rapid progressors from ground-truth in the validation set (N=61).....	65

Figure 12. AUROC graph comparing the TP rate with the FP rate from the deep learning model in the test set..... 66

Figure 13. Confusion matrix comparing the predicted classification of rapid and non-rapid progressors from ground-truth in the test set (N=72). 67

Table 1. Advantages and Disadvantages of Each Imaging Modality in ADPKD 4

Table 2. The Annual Rate of TKV Growth Stratified by Mayo Imaging Class²⁴ 9

Table 3. The Annual Rate of eGFR Decline Stratified by Mayo Imaging Class and Sex²⁴ 9

Table 4. Baseline Characteristics of Patients from the MUHC PKD Clinic, Study Aim 1, and Study Aim 2 (Left to Right). Plus-minus values are means±SD. Square bracket values are median [1st quartile-3rd quartile]. 53

Table 5. Baseline Characteristics of Rapid Progressors and Non-Rapid Progressors from Aim 2. Plus-minus values are means±SD. Square bracket values are median [1st quartile-3rd quartile].57

Table 6. Annual Rates of eGFR Decline and Proportion of Rapid Progressors Within Each Mayo Imaging Class Using eTKV and manually segmented TKV. Square bracket values are median [1st quartile-3rd quartile]. 57

Table 7. Long-term trajectory statistical model for each decade of age and MIC using the 14-year data from the CRISP dataset.¹¹ 69

ABBREVIATIONS AND ACRONYMS

ACEi: Angiotensin-Converting Enzyme Inhibitor

ADPKD: Autosomal Dominant Polycystic Kidney Disease

ARB: Angiotensin Receptor Blocker

AUROC: Area under the receiver operating characteristic curve

BMI: Body Mass Index

CKD: Chronic Kidney Disease

CNN: Convolutional Neural Network

CRISP: Consortium for Radiologic Imaging of Polycystic Kidney Disease

CT: Computed Tomography/Tomographic

DBP: Diastolic Blood Pressure

DICOM: Digital Imaging and Communications in Medicine

eGFR: Estimated Glomerular Filtration Rate

ESKD: End-Stage Kidney Disease

eTKV: Estimated TKV

FN: False Negative

FP: False Positive

HALT-PKD: Halt Progression of Polycystic Kidney Disease

HASTE: Half-Fourier Acquisition Single-Shot Turbo Spin Echo

hteTKV: Height-Adjusted Estimated TKV

htSegTKV: Height-Adjusted Manual Segmented TKV

htTKV: Height-Adjusted TKV

KDIGO: Kidney Disease Improving Global Outcomes

LH: Lachine Hospital

MGH: Montreal General Hospital

MIC: Mayo Imaging Classification

MLP: Multi-Layer Perceptron

MRI: Magnetic Resonance Imaging

MUHC: McGill University Health Centre

Nrrd: Nearly Raw Raster Data

PACS: Picture Archiving and Communications Systems

RAAS: Renin-Angiotensin-Aldosterone System

RAASi: Renin-Angiotensin-Aldosterone System inhibitor

REPRISE: Replicating Evidence of Preserved Renal Function: An Investigation of Tolvaptan Safety and Efficacy in ADPKD

RF: Random Forest

RNN: Recurrent Neural Network

RVH: Royal Victoria Hospital

SBP: Systolic Blood Pressure

SegTKV: Manual Segmented TKV

SSFSE: Single-Shot Fast Spin-Echo

SVM: Support Vector Machine

TEMPO 3:4: Tolvaptan Efficacy and Safety in Management of Autosomal Dominant Polycystic Kidney Disease and Its Outcomes 3:4

TKV: Total Kidney Volume

TN: True Negative

TP: True Positive

US: Ultrasonography

1. INTRODUCTION

1.1 Study Rationale

Tolvaptan is a disease-modifying drug that has been proven to delay the onset of kidney failure in patients with autosomal dominant polycystic kidney disease (ADPKD).¹ Due to the cost and side effects of the drug, only those who meet the criteria of being at high risk for disease progression are eligible for treatment. Therefore, correctly identifying patients most likely to progress is imperative to plan treatment and inform patients. The Mayo Imaging Classification (MIC) tool, the current prognostication tool, applies total kidney volume (TKV), height, and age to predict the disease progression in patients diagnosed with ADPKD.² It plays an important role in helping clinicians determine whether the patient is eligible for tolvaptan therapy.³ However, we believe that including other clinical and imaging factors that are known or suspected to impact the disease progression of ADPKD can improve prediction and better identify which patients are most likely to progress.

1.2 Objective

The main objective of this study is to develop machine-learning models derived from different learning algorithms and incorporate various clinical and imaging features that the current predictive statistical model does not factor in. We aim to accurately classify whether a patient diagnosed with ADPKD will experience rapid disease progression by implementing machine learning.

2. COMPREHENSIVE REVIEW OF RELEVANT LITERATURE

2.1. Autosomal Dominant Polycystic Kidney Disease

ADPKD is the most common inherited kidney disease, which affects an estimated one in 1000 live births or approximately 12.5 million people worldwide.⁴ It is also the fourth leading cause of

end-stage kidney disease (ESKD). Approximately 50% of patients diagnosed with ADPKD reach ESKD by the age of 60 and require either dialysis treatment or kidney transplantation.^{4, 5 6, 7}

2.2. Normal Kidney

Each kidney contains approximately a million nephrons, the kidney's functional units, that work together to form a filtration system.⁸ In general, the blood drains through the capillaries between the afferent and efferent arterioles and into the glomeruli of the nephron.⁶ As the blood travels through the proximal tubule, ascending and descending loop of Henle, distal convoluted tubule, and the collecting duct, various ions, minerals, water, and waste products are reabsorbed and secreted, effectively achieving the balance mentioned previously. Subsequently, the collecting ducts from many nephrons converge into the renal papilla, followed by the renal pelvis, and finally drain into the ureter. The estimated glomerular filtration rate (eGFR) depends on the number of functioning nephrons.⁸ Therefore, kidney damage can directly lead to a decrease in eGFR, indicating worsening kidney function.

2.3. Clinical Manifestations

Kidney-related complications of ADPKD typically begin with the bilateral formation of kidney cysts and exponential growth in TKV and cystic burden.⁹ Unlike the normal kidneys, polycystic kidneys can weigh up to 30 pounds or 13 kg.⁷ Urologic complications involve acute and chronic pain, gross hematuria, cyst-related infection, recurrent urinary tract infection, and nephrolithiasis.^{9, 10} Over time, patients experience a decline in the eGFR.

Hypertension, defined by blood pressure above 140/90 mmHg, is a common extrarenal complication that presents in approximately 50% of patients with ADPKD between the ages of 20-34 years.⁹ Furthermore, more than 80% of adults with ADPKD present with liver cysts.¹¹

Pancreatic cysts, cardiovascular disease, abdominal hernia, and intracranial aneurysm are other extrarenal manifestations of ADPKD.¹⁰

2.4. Diagnosis

2.4.1. Imaging Tools

Ultrasonography (US) is the preferred diagnostic tool for ADPKD for its wide availability, low cost, and non-invasiveness.^{12, 13} However, it lacks precision and reproducibility, especially in patients with smaller cysts.^{3, 13, 14} US can only detect cysts of 10 mm or greater in diameter and is prone to inter-observer variability. In contrast, computed tomography (CT) and magnetic resonance imaging (MRI) offer higher resolution and can detect smaller cysts of 2-3 mm in diameter with higher sensitivity and specificity.¹⁵⁻¹⁷ Of the two, MRI is a safer diagnostic approach because CT exposes patients to radiation. Additionally, MRI can help diagnose ADPKD with 100% specificity and sensitivity in at-risk individuals between the ages of 16 and 40 with more than ten kidney cysts and a family history.¹⁶ Therefore, MRI is considered the gold-standard modality in research studies.¹⁸ More specifically, T2-weighted MRI is recommended in patients with ADPKD for two reasons: 1) it does not require gadolinium, a potential risk factor for nephrogenic systemic fibrosis when using older gadolinium formulations, and 2) it provides better visualization and delineation of kidney tissues and cysts.¹⁹⁻²¹ However, MRI machines are not widely available in some centres, and imaging acquisition comes with a higher cost and longer acquisition time.^{22, 23}

	Advantage	Disadvantage
Ultrasonography	<ol style="list-style-type: none"> 1. Widely available 2. Low cost 3. Non-invasive 	<ol style="list-style-type: none"> 1. Older ultrasound machines cannot detect cysts with a diameter smaller than 10 mm 2. Prone to inter-observer variability 3. Lacks precision and accuracy
MRI	<ol style="list-style-type: none"> 1. High sensitivity 2. High resolution of kidney tissues and cysts 3. Detects smaller cysts of 2-3 mm in diameter 4. Absence of radiation and nephrotoxic agents 5. Precise, accurate, and reliable measurement of TKV 	<ol style="list-style-type: none"> 1. Not widely available 2. High cost 3. Long acquisition time
CT	<ol style="list-style-type: none"> 1. High sensitivity 2. Detects smaller cysts of 2-3 mm in diameter 3. Precise, accurate, and reliable measurement of TKV 	<ol style="list-style-type: none"> 1. Exposure to radiation

Table 1. Advantages and Disadvantages of Each Imaging Modality in ADPKD

2.4.2. Diagnostic Guidelines

In at-risk individuals with positive family history aged 15 to 39, the detection of at least three kidney cysts in both kidneys by US confirms the diagnosis of ADPKD.¹³ In those aged 40 to 59, two cysts in each kidney confirm the diagnosis. Finally, four or more cysts in each kidney confirm the diagnosis in those older than 60. For the remaining 10-15% of patients with negative or unknown family history, genetic tests, specialized consultations, and imaging are recommended to confirm the diagnosis of ADPKD while ruling out other cystic kidney diseases.^{24, 25}

Patients can be diagnosed as either typical (Class 1) or atypical (Class 2) ADPKD based on the patterns of their cysts.²⁶ Typical ADPKD cases are defined by the bilateral, symmetric distribution of kidney cysts and represent 90-95% of patients diagnosed with ADPKD.²⁷ In contrast, atypical cases are characterized by irregular distribution of cysts: unilateral, segmental, asymmetric, lopsided, bilateral with unilateral atrophy, and bilateral with bilateral atrophy. They represent the remaining 5-10% of patients. Class 2 patients are generally categorized as slower progressors in eGFR decline than Class 1 patients.²⁶ Imaging-based risk stratification tools have not been developed for patients with atypical or Class 2 ADPKD. Currently, criteria to formally differentiate the two classes have not been established. The absence of a standard classification method promotes subjective bias and can negatively impact clinical decision-making.

2.5. Prognosis

2.5.1. Factors Associated with Disease Progression

The patient's age of diagnosis, male sex, early-onset hypertension, urologic events (e.g., urinary tract infections, nephrolithiasis, and gross hematuria), left ventricular hypertrophy, hepatic cysts in women, and enlarged kidney volume are independent factors associated with worse kidney function in ADPKD.²⁸ The type of ADPKD gene mutation (i.e., *PKD1* truncating mutation vs. *PKD1* non-truncating vs. *PKD2*) can also be a prognostic factor. However, genetic testing is not routinely performed or indicated and can be costly.²⁹

2.5.2. Prognostic Biomarkers

As the size and the number of cysts progressively grow over time, the cysts compress the kidney parenchyma.³⁰ Consequently, the proportion of healthy kidney tissue decreases, and kidney function diminishes. However, throughout the early stages of ADPKD, the remaining nephrons hyperfiltrate to compensate for the lost kidney function and maintain a clinically-standard level

of eGFR.^{31, 32} The eGFR only rapidly declines once the patient enters the late stage of ADPKD, eventually leading to ESKD and the need for dialysis treatment or kidney transplantation. Thus, even though eGFR is typically the absolute indicator of one's kidney function, it can provide an insensitive representation of the patient's kidney disease, especially during the early stages of ADPKD.

The most important prognostic biomarker of ADPKD in clinical practice is the TKV.^{33, 34} TKV is a more reliable biomarker of ADPKD progression than eGFR because it continuously increases across all stages of ADPKD.³¹ The Consortium for Radiologic Imaging of Polycystic Kidney Disease (CRISP) study discovered a positive association between the baseline height-adjusted TKV (htTKV) at a given age and the rate of kidney and cyst enlargements.³¹ Moreover, the rates of TKV growth and eGFR decline were inversely correlated. The outcomes from the 14.5-year follow-up study of CRISP were consistent with previously-stated findings, providing strong evidence that htTKV is an independent prognostic biomarker in ADPKD.³⁵ Additionally, they concluded that baseline TKV and the rate of TKV growth were strongly associated with the development of advanced stages of CKD.

2.5.3. Prognostic Imaging Tools

MRI or CT scans can be used to measure the TKV and prognosticate ADPKD disease progression. However, most large research studies have resorted to MRI modality, likely due to its absence in radiation exposure.³⁶ A repeat measurement every 1-2 years is sufficient in observing patient's growth in TKV.

Stereology or manual segmentation of the images is the gold-standard approach in research studies to determine ground truth manual segmented TKV (SegTKV). However, this approach is rarely used in clinical settings as it is labour- and time-intensive to complete the task (i.e., up to

50 minutes per analysis).^{15, 26, 37} For convenience, clinicians use the ellipsoid equation to calculate the estimated TKV (eTKV): Left or Right Kidney Volume = $(\pi/6) \times (L_{\text{coronal}} + L_{\text{sagittal}})/2 \times W \times D$ (L = maximum longitudinal length; W = maximum width perpendicular to L; D = maximum depth).^{26, 38} The lengths can be measured by a radiologist from the three orthogonal dimensions (sagittal, coronal, and axial planes) from CT or MRI scans (Figure 1).

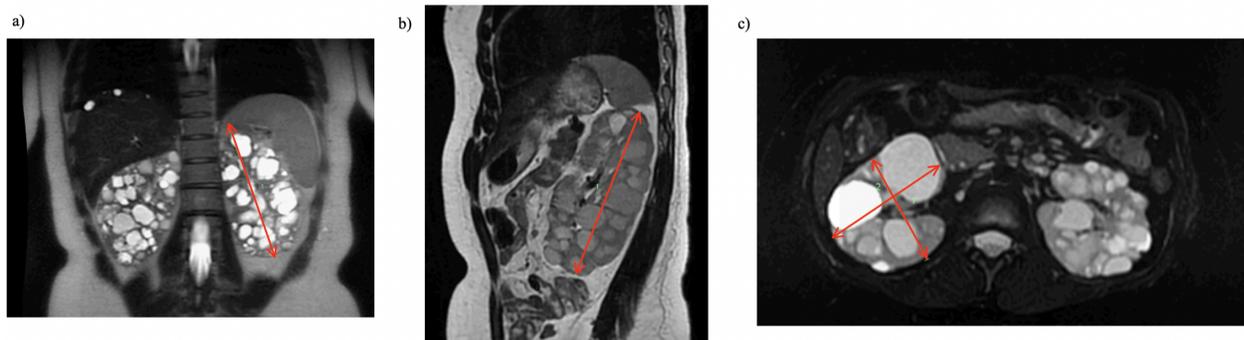


Figure 1. The orthogonal measurements of the maximal kidney a) coronal length, b) sagittal length, and c) width and depth using axial planes in an MRI scan.

2.6. Classification of Risk

The MIC tool, created by the *Mayo Foundation for Medical Education and Research* (<https://www.mayo.edu/research/documents/pkd-center-adpkd-classification/doc-20094754>), is currently the best predictive model that assesses disease progression in patients with typical (Class 1) ADPKD.³⁹ The MIC tool stratifies patients into five subclasses (Class 1A-1E) based on their TKV (mL), height (m), and age (years). Classes 1A and 1B are considered slow progressors of ADPKD, and Classes 1C, 1D, and 1E are considered rapid progressors. Each class exhibits a unique linear path of annual rates of TKV growth (Table 2) and eGFR decline (Table 3). Although the MIC is not designed to predict the disease progression of atypical (Class 2) patients, they are generally at lower risk than typical (Class 1) patients.

	Predicted TKV slope (% per year) for both sex
Class 1A	<1.5
Class 1B	1.5-3
Class 1C	3-4.5
Class 1D	4.5-6
Class 1E	>6

Table 2. The Annual Rate of TKV Growth Stratified by Mayo Imaging Class²⁶

	Predicted eGFR slope for men (mL/min/1.73m² per year)	Predicted eGFR slope for women (mL/min/1.73m² per year)
Class 1A	-0.23	0.03
Class 1B	-1.33	-1.13
Class 1C	-2.63	-2.43
Class 1D	-3.48	-3.29
Class 1E	-4.78	-4.58

Table 3. The Annual Rate of eGFR Decline Stratified by Mayo Imaging Class and Sex²⁶

2.7. Limitations of the MIC

The MIC tool utilizes the eTKV, calculated from the ellipsoid equation, to classify patients with ADPKD. However, the ellipsoid equation presents poor reproducibility and repeatability compared to the gold-standard manual segmentation approach because it assumes an elliptical growth of kidneys.⁴⁰ However, in clinical settings, cystic kidney growth is highly irregular and unique to each patient.⁴¹ Furthermore, the ellipsoid equation only considers three orthogonal lengths from one to two image slices.^{42, 43} If the ellipsoid equation is used to calculate the TKV

of patients with exophytic cysts, the deformation in their kidneys' shape can misrepresent the true axial measurements.⁴⁴ The main clinical concern that could arise from inaccurate TKV measurements is the risk of misclassification and its impact on the decision to offer disease-modifying therapy to a patient.

The MIC tool also faces the issue of generalizability in specific patient populations. Despite its wide usage in clinical settings, it has not been formally validated in prospective clinical trials. Moreover, black patients and atypical (Class 2) patients were underrepresented in the study population used to design the predictive statistical model. Therefore, the MIC tool should be used with caution for patients who share these characteristics.²⁶

Finally, the MIC tool does not include other important demographic, imaging, and clinical features that have previously been described as independent risk factors of disease progression in ADPKD.²⁸ Cyst distribution, size, and number are also examples of imaging features that have been hypothesized to influence a more rapid kidney function decline.⁴⁵ Theoretically, cysts in the collecting duct, where many nephrons converge, could impair kidney function more than cysts in a single nephron.⁴⁶ Likewise, cysts in the renal papilla, where collecting ducts drain, could further worsen kidney function. Understanding the renal outcomes in those with few large cysts or innumerable small cysts should provide important insights.⁴⁵ However, these associations have not been formally studied to our knowledge.

The limitations are also apparent in patients from the McGill University Health Centre (MUHC) PKD Clinic. Figure 2 displays an illustrative example describing three Class 1C female patients in the same decade for age and were followed for at least two years. According to the MIC tool, they are all projected to experience an annual increase in TKV by 3-4.5% per year and an annual

decrease in eGFR by $-2.43 \text{ mL/min/1.73m}^2$ per year. However, unlike the MIC prediction, their true clinical course deviates significantly from one another.

Patient A presents with the highest TKV among the three, described by a dominant cyst in the left kidney and numerous smaller bilateral cysts. However, she has a well-preserved eGFR of $94 \text{ mL/min/1.73m}^2$ with no observed annual kidney function decline. The Class 2-like presentation of cysts and the relatively large proportion of healthy tissue on the right kidney may explain this phenomenon. However, there is no established study that confirms this hypothesis.

Patients B and C share similar TKV and bilateral, symmetric distribution of innumerable small- and medium-sized cysts. Interestingly, Patient B has a much lower baseline eGFR than Patient C ($59 \text{ mL/min/1.73m}^2$ vs. $80 \text{ mL/min/1.73m}^2$), yet Patient C demonstrated a worse kidney function outcome than Patient B ($-9 \text{ mL/min/1.73m}^2$ per year vs. $-7 \text{ mL/min/1.73m}^2$ per year). The older age and presence of liver cysts in Patient B may identify someone at risk of worsening renal function. The differences in clinical factors — not accounted for in the MIC model algorithm — such as race, obesity, diet, or past medical history (e.g., hematuria, nephrolithiasis, hypertension) and their potentially complex associations may also help explain the different annual eGFR decline observed in each patient.²⁸ However, given that Patient C is younger than Patient B with a faster annual decline in eGFR, the kidney functions of the two patients may converge over time.

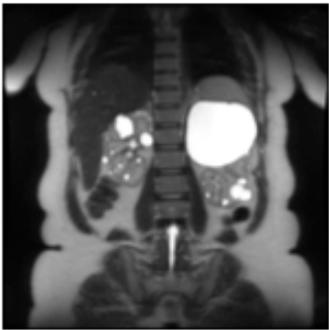
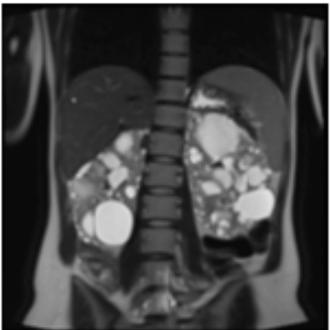
		
Patient A	Patient B	Patient C
49-year-old TKV 1522 mL Mayo Imaging Class 1C eGFR 94 ml/min/1.73 m ² GFR decline 0 ml/min/year	47-year-old TKV 1196 mL Mayo Imaging Class 1C eGFR 59 ml/min/1.73 m ² GFR decline 7 ml/min/year	41-year-old TKV 1176 mL Mayo Imaging Class 1C eGFR 80 ml/min/1.73 m ² GFR decline 9 ml/min/year

Figure 2. Unique Clinical Presentations of ADPKD in Patients from the MUHC PKD Clinic

Many of these associations are speculative, but this clinical example highlights the limitations of the MIC tool and the need for a better predictive modelling tool. TKV, age, and height do not provide enough detail on patients' progression of ADPKD. A prognostic tool that can consider various factors — in addition to the input variables of MIC — may improve the prediction of disease progression in ADPKD, pushing for more individualized patient care.

2.8. Available Treatment

2.8.1. Non-Pharmacological Interventions

Hypertension is an independent risk factor for progression to ESKD.^{28, 47, 48} Therefore, blood pressure control is important for managing ADPKD. The current guidelines suggest restricting salt intake to a maximum of 5 g per day. The goal is to decrease the plasma osmolality level, inhibit the renin-angiotensin-aldosterone system (RAAS), and reduce vasopressin, the main driver of cyst proliferation.^{49, 50} A study illustrated a worsening eGFR of -0.11 mL/min/1.73m²

per year per 1 g of salt in both early- and late-stages of ADPKD.⁵⁰ A post hoc analysis of the Halt Progression of Polycystic Kidney Disease (HALT-PKD) study also discovered an increase in TKV in early-stage ADPKD and a decrease in eGFR in late-stage ADPKD with higher salt intake.⁵¹ A systematic review, which included eight randomized clinical trials that compared two or more levels of salt intake, observed a significant decrease in systolic blood pressure, proteinuria, and adverse kidney outcomes when salt intake was restricted.⁵²

Other recommended lifestyle modifications involve lowering body mass index (BMI), transitioning to a healthy lifestyle (e.g., cardiovascular exercise) and diet, quitting smoking, and avoiding non-steroidal anti-inflammatory drugs.^{25, 53-55} However, there is limited evidence and their effects on the disease progression of ADPKD should be further examined.

Increased water intake can reduce the recurrence of kidney stones. An animal study demonstrated a lower level of vasopressin and a delay in cyst proliferation with higher water intake.^{56, 57} However, the results from a human trial were not consistent with previous findings.⁵⁸ This could be attributed to the small sample size, challenges in achieving a much higher water intake, and a short study period. A long-term trial with a larger, more generalizable patient cohort and rigorous methodology is required to confirm any benefits of prescribed water therapy.

High protein intake can induce hyperfiltration in the kidneys, increase the level of vasopressin, and worsen disease progression.⁵⁹ A low protein diet of less than 1.3 g per kg of body mass per day is advised by the Kidney Disease Improving Global Outcomes (KDIGO), but these are not specific to those with ADPKD.⁶⁰ Once the patient reaches an eGFR of 30 mL/min/1.73m², further restriction in protein consumption (less than 0.8 g per kg of body mass) is advised. A randomized controlled trial demonstrated a faster decline in the eGFR in the first four months,

followed by a slower decline in those with moderate renal insufficiency when they maintained a low protein diet (0.58g per kg per day) and low blood pressure (mean arterial pressure of 125 mmHg or less).⁶¹ However, compared to a low-protein diet, a very low-protein diet (0.28g per kg per day) had no significant effect on the renal progression of patients with severe renal insufficiency.

Although there is increasing evidence that cyst growth in ADPKD may be associated with aerobic glycolysis, commonly seen in cancer cells, its effect on the disease progression is not well-established.⁶²

2.8.2. Pharmacological Interventions

Hypertension can also be managed pharmacologically with renin-angiotensin-aldosterone-system inhibitors (RAASi). Hypertensive patients with ADPKD who are younger than 50 years with a preserved eGFR of greater than 60 mL/min/1.73m² and without significant cardiovascular comorbidities can be prescribed either angiotensin-converting enzyme inhibitor (ACEi) or angiotensin receptor blocker (ARB).¹⁰ The HALT-PKD Study A, a double-blind, placebo-controlled trial, demonstrated that strict blood pressure control (between 95/60 and 110/75 mmHg) using ACEi or ARB reduced the annual rate of TKV increase by 1% but did not affect the annual rate of eGFR decrease.⁶³ Moreover, strict blood pressure control lowered renal vascular resistance and reduced proteinuria. Patients did not experience significant improvement when prescribed a combination of ACEi and ARB. Thus, dual angiotensin blockade is not recommended.

Tolvaptan, a selective arginine vasopressin type 2 receptor antagonist, is currently the only approved drug in North America that can slow the disease progression in ADPKD.⁶⁴ Unlike

other pharmacological approaches, it directly suppresses the release of vasopressin into the distal nephrons, thereby reducing cyst proliferation in the kidneys and additional fluid secretion into existing cysts.⁶⁵ Patients who receive tolvaptan therapy achieve a slower annual rate of growth in TKV and annual rates of eGFR and kidney function decline.^{66, 67} In the Tolvaptan Efficacy and Safety in Management of Autosomal Dominant Polycystic Kidney Disease and Its Outcomes 3:4 (TEMPO 3:4) randomized, double-blinded placebo-controlled trial, patients with TKV greater or equal to 750 mL and creatinine clearance of 60 mL/min or more were enrolled for three years to test the efficacy of tolvaptan. Patients who received tolvaptan experienced an average annual rate of TKV increase of 2.7% compared to 5.5% in the placebo group.⁶⁶ Moreover, the average annual rate of eGFR decline was -2.72 mL/min/1.73m² per year vs. -3.70 mL/min/1.73m² per year in the intervention and placebo groups, respectively. Additionally, patients who were prescribed tolvaptan reported fewer incidents of kidney pain that generally requires medical attention and analgesic agents.

The efficacy of tolvaptan in rapid progressors defined by the MIC tool became more evident in the post hoc analysis of the TEMPO 3:4 trial.² When they excluded slow progressors, the annual rates of TKV growth (5.8% per year vs. 2.9% per year) and eGFR decline (-3.93 mL/min/1.73m² per year vs -2.78 mL/min/1.73m² per year) were significantly lower in the tolvaptan group compared to the placebo group. Furthermore, the Replicating Evidence of Preserved Renal Function: An Investigation of Tolvaptan Safety and Efficacy in ADPKD (REPRISE) trial, a multicentre, placebo-controlled, double-blinded study, tested the efficacy and safety of tolvaptan in more advanced patients with eGFR between 25 and 65 mL/min/1.73m².⁶⁷ After a year, the average annual rate of eGFR decline was lower in the tolvaptan group (-2.34 mL/min/1.73m²) than in the placebo group (-3.61 mL/min/1.73m²). It

is estimated that the onset of kidney failure can be delayed by an average of one year for every four years of tolvaptan treatment.^{64, 66} In other words, physicians can potentially delay the median age of ESKD onset by 6.5 years and increase the life expectancy of patients with advanced ADPKD by 2.6 years. Currently, tolvaptan therapy is considered the first line of drug treatment only for patients under 65 years old who are classified as rapid progressors (Classes 1C, 1D, or 1E) or show clear evidence of rapid eGFR decline.^{32, 49, 68}

While tolvaptan is efficacious in patients at risk of disease progression, it is also associated with significant aquaresis-related side effects such as thirst, polyuria, polydipsia, and pollakiuria.⁶⁹ Moreover, patients must receive monthly liver enzyme level assessments to mitigate the risk of reversible drug-induced liver toxicity. The cost-effectiveness of tolvaptan, which costs up to \$5760 (in 2010 US dollars) per patient, is also considered low compared to other therapeutic interventions.⁷⁰ Ineligible patients are prescribed with RAASi with aggressive blood pressure management if they have hypertension and are recommended general non-pharmacological interventions. Therefore, if the rapid progressors can be better distinguished, tolvaptan therapy could be applied more appropriately, improving cost-effectiveness and risk-benefit.

mTOR inhibitors (e.g., sirolimus and everolimus) and somatostatin analogues (e.g., octreotide) have been shown to slow the annual rate of TKV increase but not improve the rate of eGFR decline in randomized clinical trials.⁷¹⁻⁷⁴ However, KDIGO advises against the prescription of these drugs for treating ADPKD until longer clinical trials with larger sample sizes support previous findings.²⁵

2.9. Risk Associated with the Current Prognostic Tool and Clinical Management

Due to the differences in clinical management strategies between slow- and fast-progressors, accurate classification is imperative in patients with ADPKD.⁴⁰ Although misclassifications can occur between all MIC classes from the clinical application of the ellipsoid equation, the misclassification between Class 1B and 1C poses the greatest concern as this threshold often triggers the decision to treat a patient. Misclassifying patients as Class 1B when their true MIC is Class 1C can delay the benefits they could receive from disease-modifying tolvaptan therapy. In contrast, misclassifying patients as Class 1C when their true MIC is Class 1B exposes them to a potential for unnecessary adverse events and adds emotional and financial burden to patients.⁶⁹

2.10. Current Literature on eTKV Calculation Using the Ellipsoid Equation

A retrospective study by Shi et al. compared the TKV measurements and prognostic performances between the ellipsoid equation and manual segmentation approaches.⁴⁰ 308 patients from the Toronto PKD clinic with confirmed diagnoses of typical ADPKD entered the study. Apart from the five patients with contrast-enhanced CT images, all patients possessed T2-weighted MRI. Manual segmentation and axial measurement tasks were performed by a highly experienced radiologist in separate sittings while blinded from patient information.

The Bland-Altman analysis illustrated a mean percentage difference in TKV of only -0.6% between the two methods. However, the limits of agreement ranged between -20% and +20%, and 17 cases (5.5%) had a TKV difference of greater than 20%. Furthermore, 42 patients (13.6%) were misclassified when the ellipsoid equation was used. More importantly, seven and four patients were misclassified as Classes 1B and 1C, respectively. These findings suggest that the ellipsoid equation is inconsistent in accounting for certain irregularities in the cystic kidneys and directly contributes to the misclassification of patients.⁴⁰

Demoulin et al. also compared the TKV measurement approaches in 140 patients with confirmed diagnoses of typical ADPKD from a clinic located in Brussels.⁴³ Those who had undergone kidney failure or unilateral nephrectomy were excluded from the study. A junior and senior radiologist measured the axial lengths and segmented the kidneys of 53 and 140 patients, respectively, to assess reproducibility (inter-reader agreement). Subsequently, they each remeasured 10 and 22 patients' kidneys to assess repeatability (intrareader agreement). However, only the measurements performed by the senior radiologist were used to assess misclassification.

The Bland-Altman plot illustrated a mean percentage difference of 9.5% between the two measurements, with the limits of agreement between -10.4% and 29.4% for the junior radiologist. For the senior radiologist, the mean percentage difference was 6.9%, and the limit of agreement was between -12.0% and 25.7%. Although the average time to complete a TKV measurement using the ellipsoid equation was approximately 35 minutes faster, the repeatability and reproducibility were significantly greater for the manual segmentation approach. The repeatability coefficients for manual segmentation vs. the ellipsoid equation were 2.4% vs. 14% for the junior radiologist and 4.6% vs. 17% for the senior radiologist. The reproducibility coefficients between the two readers for manual segmentation and ellipsoid equation measurements were 6.7% and 15%, respectively. Further analyses illustrated that manual segmentation by a junior and a senior radiologist could detect changes in TKV of greater than 6.6% and 3.5% in a patient with 95% confidence. Meanwhile, the orthogonal and eTKV measurements by a junior and senior radiologist can detect changes in TKV of greater than 22% and 19% in a patient with 95% confidence.⁴³

There was generally a good agreement between the MIC using eTKV and SegTKV (intraclass coefficient = 0.924). However, the ellipsoid equation misclassified 21 out of the 140 patients (15%) by one risk class. Of those, eight were misclassified as Class 1B instead of Class 1C.⁴³

Even though the ellipsoid equation performs well in most cases, large discrepancies in certain cases and misclassifications should be minimized by exploring alternative approaches to measure TKV. Semi-automated or fully automated segmentations using machine learning could improve accuracy and precision.^{41, 75, 76} Future studies should also compare the two measurement approaches within their PKD clinic cohorts to externally validate the findings. Moreover, analyzing images with discrepancies of greater than 20% can be a future step to identifying any patterns the ellipsoid equation fails to capture.

2.11. Alternative Conventional Models to Predict Disease Progression of ADPKD

In 2017, Kline et al. applied image texture features from 122 T2-weighted MRIs from the CRISP study to predict whether patients with ADPKD will progress to chronic kidney disease (CKD) Stage 3A, 3B, and 30% decline in eGFR after eight years.⁴⁵ The study chose entropy (degree of disorder within the kidney), correlation (grayscale value dependence of kidney voxels), and energy (a measure of tissue uniformity) from the nine image texture features to examine their predictive power, along with the traditional clinical variables (htTKV, age, and baseline eGFR); entropy signals randomness in cyst distribution; correlation illustrates the regions with similar appearances within the kidney; energy indicates the differences in cyst size and number.

For CKD Stage 3A classification, entropy, correlation, and energy displayed the area under the receiver operating characteristic curve (AUROC) of 0.93, 0.72, and 0.80, respectively. For CKD Stage 3B classification, the AUROC were 0.86, 0.79, and 0.80. For classifying the 30% decrease

in eGFR, the AUROCs were 0.82, 0.69, and 0.75. The linear regression model of entropy, correlation, and energy with eGFR percentage change displayed slopes of -0.52, -0.43, and -0.52.⁴⁵

When the three image texture features were added, the AUROC of the traditional model (htTKV, age, and baseline eGFR) for CKD Stage 3A classification improved from 0.86 to 0.94. For CKD Stage 3B classification, it improved from 0.90 to 0.96. For classifying the 30% decrease in eGFR, it improved from 0.75 to 0.85. Finally, multiple linear regression of the traditional biomarkers with eGFR percentage change obtained a Pearson's correlation coefficient of -0.51. It improved to -0.70 when the image textures were added.⁴⁵

There are some limitations to this study. Since the image acquisition protocol is not standardized across clinics, and the model relies on image texture features relevant to image signal intensities, its performance may depend heavily on the quality of the image resolution. Moreover, the CRISP study cohort only includes patients under the age of 46 and is enriched with those considered to be at higher risk of disease progression. Therefore, this predictive model should be externally validated to test its generalizability.⁷⁷

In another study, McEwan et al. used patients from the placebo arm of the TEMPO 3:4 trial (N=484) to develop a predictive multivariable regression model.⁷⁸ The model used age, gender, and baseline TKV to predict the disease progression until the patient reached 80 years old or ESKD, defined by eGFR less than 15 mL/min/1.73m². The two equations in the model were fitted to annual changes in TKV +500 and eGFR +60 to avoid calculating the log of negative numbers. Moreover, the model's uncertainty in prediction was calculated and incorporated into

the algorithm to mimic inter-patient variability since patients who share the same baseline characteristics may vary in their TKV and eGFR in real clinical settings.

The model aggregated estimates from a hypothetical cohort of up to 10,000 patients.

Subsequently, the model's accuracy was externally validated on patient cohorts from the CRISP study, HALT-PKD A and B, and THIN studies, which included patients at various stages and severity of ADPKD. Finally, the model was tested on a hypothetical cohort that matched the baseline characteristics of the overall study population from the TEMPO 3:4 trial (N=1445).

The statistical model predictions corresponded well with ground truth, although the 95% prediction intervals widened over time. The predictions were also consistent with the HALT-PKD cohort. However, since the HALT-PKD B study consisted of rapid progressors, the eGFR decline prediction deviated in the final two years of the trial. In the same HALT-PKD cohort, the predictions were better when only those with baseline TKV of 1000-1500 mL were included. Similarly, the predictions matched well with the THIN database when compared against the sub-populations with baseline TKV of 1500 mL.⁷⁸

The study illustrated that age had a modest impact on the onset age of ESKD when eGFR and TKV were the same.⁷⁸ Sex positively correlated with TKV growth but was not statistically significant. Finally, as expected from previous studies, baseline TKV and current TKV were strong predictive risk factors for annual rates of TKV growth and eGFR decline, respectively. Even though the statistical model performed well, the study has some limitations. The TEMPO 3:4 trial only included patients who were lower or equal to 50 years old with TKV greater or equal to 750 mL. Furthermore, a limited number of patients with baseline TKV of 750-850 mL were enrolled in the study. Therefore, if patients from the external cohort had TKV less than 850 mL, the equation assumed the value of 850 mL. Similarly, the annual changes in TKV and eGFR

for patients older than 50 years old were assumed as the annual change at 50 years old. In addition, the time-varying covariate model may need to be explored to account for the effects of RAASi and tolvaptan on patients' disease progression in HALT-PKD and TEMPO 3:4 studies, respectively.

Nevertheless, the results from both studies suggest that the current biomarkers — htTKV, age, and baseline eGFR — are not enough to stratify all ADPKD patients into five subclasses and predict their disease progression. Both studies acknowledge that additional features, such as image features (e.g., cystic location and distribution) and clinical factors (e.g., hypertension history, polycystic liver disease, and family history) should be explored to assess their relevance in prediction.

2.12. Proposed Solution

So far, many individual clinical factors (e.g., early-onset hypertension and urologic events, sex, and age of diagnosis) and some image features (e.g., baseline TKV and image texture features) have been identified as factors affecting the renal progression of ADPKD.²⁸ It would be interesting to examine how other imaging features (e.g., cyst distribution, size, load, composition, and location), which have been postulated but not confirmed, affect one's disease progression.^{31, 45, 46} The interaction between demographics, imaging, and clinical variables can also be explored.

Developing a machine learning model that can predict disease progression across the full spectrum of ADPKD by identifying important features and complex associations within the clinical variables and medical imaging could move practice towards a precision medicine approach. Furthermore, a more accurate classification of these patients may enrich clinical trials when studying new investigational treatments.

2.13. Machine Learning

2.13.1. Introduction

Humans are adept at deciphering information and creating associations from small datasets. However, the task becomes more challenging as the dataset becomes larger. Machine learning is a powerful tool that can process large, heterogenous datasets, discover, and learn complex relationships between various features, and ultimately outperform humans in a short period of time.

Traditional machine learning models were initially designed to make decisions based on specific rule-based systems coded by the engineer.⁷⁹ Now, they possess algorithms and architectures that allow them to generalize information and identify patterns throughout the training and validation steps.⁸⁰ Predictive models can then apply the acquired knowledge and patterns on unseen data to accurately predict the outcomes of interest or label. The prediction tasks of machine learning involve classification, regression, and novelty detection.^{80, 81} In contrast, descriptive models simply gain knowledge from past data and understand complex relationships.⁸⁰ There are four categories of machine learning approaches: supervised learning, unsupervised learning, semi-supervised, and reinforcement learning.⁸²

2.13.2. Supervised Learning

In supervised learning, the model uses a dataset with labelled input and its corresponding outcome of interest.⁸⁰ Moreover, it is directly informed of the appointed task. The model is trained using either a classification or regression algorithm, depending on whether the outcome variable is categorical or continuous.⁸³ In the testing stage, the model's predictive performance is evaluated on unseen data from the test set.⁸⁴ Linear and logistic regression, support vector

machine (SVM), K-nearest neighbours, naïve Bayes, decision tree, random forest (RF), and neural networks are common supervised learning approaches.

SVM is a type of supervised learning algorithm used for classification.⁸² In SVM, a hyperplane, which is a solid line acting as the decision boundary, separates the classes in high-dimensional data, a dataset containing many features. The model's performance is greatest when the margin between the hyperplane and the closest data points from each class is maximized. The hyperplane can be adjusted to maximize the margin and improve generalizability. However, the hyperplanes may not separate the two classes completely if the data contains noise and outliers. Soft-margin SVM approach can help mitigate this problem at the cost of a decrease in sensitivity and an increase in bias. SVMs can also linearly separate non-linear data by transforming the data into a higher dimensional space using kernel functions. The advantages of SVM are their ability to handle high-dimensional data and perform well with small datasets. However, it can be sensitive to the choice of kernel function and is computationally expensive for larger datasets.

The decision tree approach classifies data points by splitting them into nodes, branches, and leaves based on their attributes.⁸⁵ In other words, each decision tree path that each data takes can be understood as an if-then conditional statement. Decision trees are easily comprehensible and interpretable, require few data pre-processing and hyperparameter tuning compared to other approaches, can handle missing values, and are robust to outliers.^{82, 85} However, they do not perform well with imbalanced datasets, are sensitive to noise, and are susceptible to overfitting, especially with smaller datasets.⁸²

The RF approach independently grows many decision trees and uses an ensemble learning technique called bagging or bootstrap aggregation to make predictions.⁸² Bagging creates

random subsets of data and features with replacements from the training dataset.^{82, 86} The model is trained on these subsets and builds its decision trees. For regression tasks, it averages the results obtained from the trees.⁸² For classification tasks, it selects the majority vote. For these reasons, RF models can outperform a single decision tree, handle high-dimensional data, and mitigate the risk of overfitting. Furthermore, they can report and rank the importance of each input feature in carrying out the prediction task.⁸⁷

Neural networks and deep learning are supervised learning approaches inspired by the architecture of the human brain.⁸² Neural networks have input, hidden, and output layers that are analogous to dendrites, soma, a nucleus, and an axon in the neuron. The data in the input layer is processed throughout the hidden layer(s), and the results for the given input data are generated in the output layer. Widely used deep learning approaches are multi-layer perceptron (MLP), recurrent neural network (RNN), and convolutional neural network (CNN). For instance, MLP starts from the input layer and calculates the linear function according to the network parameters (e.g., weight and bias values). A non-linear function is subsequently applied to all neurons within the hidden layer, followed by the output layer. The parameters are constantly adjusted based on the loss function measured in each backpropagation and random iteration process. The training is complete once it satisfies the pre-determined criteria set by the developers and the model is ready for testing.

2.13.3. Unsupervised Learning

Unlike supervised learning models, unsupervised learning models work with unlabelled data that consists only of input variables. While supervised learning models receive target outputs and feedback (e.g., rewards and punishment) from the environment, unsupervised learning models are left to identify associations, learn features, label the inputs, and produce the output

independently.^{82, 88} Unsupervised learning approaches are mainly applied for association mining, clustering, and feature reduction tasks.^{89, 90} While the model's autonomy helps identify hidden patterns within large, high-dimensional datasets, it may be less accurate than a supervised model. Moreover, it does not promise that the associations will be useful or related to the study. The main pitfall is the possibility of the Blackbox Effect. If the model grows too complex, it becomes impossible to identify the labels and weight added to each variable and understand the logic behind the results even if it performs well.⁹¹ K-means clustering and principal component analysis are examples of unsupervised learning approaches.

2.13.4. Semi-supervised Learning

Semi-supervised learning involves a mix of labelled and unlabelled input features, but most remain unlabelled.⁸⁸ This approach is preferred when the data is large and unlabelled, but the labelling process requires expert knowledge or can be labour-intensive.⁹⁰

2.13.5. Reinforcement Learning

Contrary to other machine learning models, a training dataset does not need to be provided externally for reinforcement learning models.⁸² Instead, it can create its own dataset during training, receive feedback from its environment, and optimize its performance via a trial-and-error method. Although it reduces the time and labour involved in labelling data, this machine-learning approach can only be applied in settings where numerous trial-and-error are acceptable without any major consequences.

2.14. Stages of Machine Learning Model Development

Initial data analyses and appropriate preprocessing steps are crucial for improving the model's performance.⁸² Since most machine learning techniques cannot handle missing values, two

strategies can be applied: 1) data removal and 2) data imputation (e.g., mean, median, or mode).

Typically, the latter method is preferred since the former method is only tolerable in large sample sizes.

Cross-validation is a popular strategy that can prevent overfitting and improve the robustness of the model, especially with smaller datasets.⁹² More specifically, k-fold cross-validation is one commonly used in machine learning. The data sample with size n is randomized and divided into k subsets of equal size, where k can be any arbitrary number ranging from 2 to $n-1$.⁹² The first subset is set aside for testing, and the remaining $k-1$ subsets are used for training.⁹³ The model's performance on the test set is recorded. The steps are repeated k times until all subsets have been utilized once as a test set. All the results are averaged and recorded as a single performance metric of the model.

High-dimensional data can bring challenges to building a high-performing model.⁸² If too many features are included in the model, it could increase the risk of overfitting, complexity, and interpretability of the model. Excluding features irrelevant to predicting the outcome or redundant will simplify the model and improve performance. Another technique to better handle high-dimensional data is through dimensionality reduction. Feature extraction and feature selection techniques can help lower the dimensions of the dataset while preserving the fundamental properties that contribute to the model's accuracy.

Class imbalance is another problem that can lead to poor generalizability, especially in the minority class.⁹⁴ It is especially common in medical data and there is increasing evidence that current clinical models are disproportionately affecting specific populations in the minority groups.^{95,96} A balanced training set is crucial for improving the performance and generalizability

of machine learning models.⁹⁷ However, if an imbalanced dataset is inevitable due to various reasons, oversampling and undersampling techniques can be sought out to mediate this issue. Random oversampling will increase the sample size of the minority class by duplicating observations from the minority variable. However, it comes with the cost of increased training time, memory complexity, and the risk of overfitting. Meanwhile, undersampling decreases the number of samples from the majority class to level the sample sizes of both classes. However, undersampling can result in the loss of pertinent information and is not recommended for use in datasets with small sample sizes.⁸²

Normalization is an essential preprocessing step for datasets that consist of features with diverse scales, units of measurement, and range of values.⁹⁸ Feature-based machine-learning approaches tend to place greater weight on features with greater numerical values throughout the training process and bias the final prediction. Hsu et al. highlighted the vast improvement in the SVM models' performances when scaling was performed on real-world data on astroparticles, bioinformatics, and vehicles.⁹⁹ The accuracies of the model improved from 75.2% to 96.9%, 36% to 85.2%, and 4.88% to 87.8%, respectively. Scaling all features within a common range ensures equal numerical importance and reduction of bias in the model.¹⁰⁰ It is important to understand that equal numerical importance does not equate to equal contribution to the model's decision-making process. Log scaling, min-max scaling and Z-score scaling are examples of common normalization techniques.¹⁰¹ The min-max method scales the values to a fixed range between 0 and 1 and reduces the standard deviation. Meanwhile, Z-score scaling transforms the features to the standard normal distribution, where the mean is 0, and the standard deviation is 1. Once the data has been analyzed, cleaned, and preprocessed, it is randomly split into training, validation, and test sets. For smaller datasets, the conventional split ratio is 70% training, 15%

validation, and 15% test. The test set must be locked away until the training and validation stages are completed to mitigate bias.⁸² The model is first exposed to the training set. The model learns from this data and adjusts its internal parameters to minimize prediction error. After training, the quality of the model's performance is evaluated on the validation set, and its performance is measured. If the model performed well on the training but poorly on the new, independent validation subset, it would be a sign of overfitting. In these cases, developers can make necessary adjustments. The validation step also provides the opportunity for hyperparameter tuning, which are steps to adjust parameters pre-determined by the developers and hence independent of the dataset and the model.⁸² Once the model prototypes are fully trained and validated, the best-performing model from the validation step is selected based on the various performance metrics (e.g., accuracy, precision, recall, and F1 score).⁹⁰ The selected model is tested on the testing set, which was previously locked away to avoid data leakage. The evaluation metrics obtained from this step will provide a rough estimate of the model's performance and generalizability in the real world.

2.15. Performance Evaluation Metrics in Machine Learning

Various metric tools can be used to assess the machine learning model's performance. Accuracy ($\frac{TP + TN}{TP + FP + TN + FN}$) is one of the most commonly used metrics for evaluating categorical predictions.¹⁰² However, accuracy should be used with caution for imbalanced datasets as it may not be representative of the model's true performance.¹⁰³ For instance, if a model predicts all cases as "no disease" for a disease with very low prevalence, an accuracy of 99% overestimates its performance on the majority class while overlooking its true performance on the minority class. Another pitfall of accuracy is that it combines the model's performance in predicting TP and TN rates rather than displaying its individual performance for each class.¹⁰⁴ If one of the

rates is significantly lower than the other, only reporting the accuracy could obscure the model's true performance. Therefore, sensitivity ($\frac{TP}{TP + FN}$) and specificity ($\frac{TN}{TN + FP}$) should be reported along with accuracy.

Precision ($\frac{TP}{TP + FP}$), recall ($\frac{TP}{TP + FN}$), and F1 score ($2 * \frac{Precision * Recall}{Precision + Recall} = \frac{TP}{TP + 0.5(FP + FN)}$), which is the harmonic mean of precision and recall, are widely used in machine learning studies. The F1 score is highest when recall and precision are equal to one another. However, presenting the F1 score alone may fail to interpret the true performance of the model because it focuses only on the true positives and obscures the discrepancy between recall and precision values.¹⁰⁵ As a resolution, precision, recall, and confusion matrix, which illustrates the proportion of TP, TN, FP, and FN, can be presented with the F1 score. In cases of class-imbalanced data, a weighted-average F1 score is a more reliable metric. It accounts for the ratio of the number of samples per class among the total number of samples, which is termed support.¹⁰⁶

$$Weighted\ F1\ Score = \sum_{i=1}^N Support_i \times F1\ Score_i$$

$$Support_i = \frac{Number\ of\ Samples\ in\ Class\ i}{Total\ Number\ of\ Samples}$$

The AUROC, another commonly used metric, illustrates the trade-off between the TP rate and FP rate for each threshold value.¹⁰⁷ Subsequently, the AUROC value will help visualize the sensitivity and specificity across all thresholds. Since the AUROC depends on TP and FP rates, it is insensitive to changes in class distribution. Therefore, while the accuracy, precision, and F1 score are susceptible to change based on the class distributions of the training and testing set, the

AUROC may remain more robust. However, similar to F1 scores, AUROC also weighs equal importance on sensitivity and specificity for all problems, which may not always be the case.¹⁰⁸

2.16. Current Literature on Machine Learning and Predicting Outcome in ADPKD

A comprehensive literature search of PubMed and Google Scholar was conducted. The keywords “autosomal dominant polycystic kidney disease”, “machine learning”, “automated”, “disease progression”, “prognosis”, and “prediction” were used in both “AND” and “OR” combinations. There was only one pre-print study on a machine learning predictive model in ADPKD at the time of writing this thesis. Similar to the image texture feature study by Kline et al., Raj et al. developed a model using the T2-weighted MRIs from the CRISP study (N=135). They established the same prediction criteria: whether the patient will reach CKD Stage 3A, 3B, and a 30% decline in eGFR after eight years.¹⁰⁹ In addition, the model predicted the CKD stage of each patient and the percentage decline in the eGFR after eight years. The model consisted of two networks: 1) Attention U-Net that automatically segments the kidneys from the MRI and calculates the htTKV (mL/m) by multiplying the number of segmented kidney voxels by the voxel volume (mm³) and dividing by the product of 1000 and the height of the patient (m) and 2) A two-part network of CNN and MLP that uses image segmentation, predicted htTKV, age, and baseline eGFR to predict the outcomes of patients.

The training, validation, and testing data for the Attention U-Net were split into 70, 10, and 20 subjects per fold in a 5-fold cross-validation. Early stopping was implemented to prevent overfitting. The network with the highest dice score, a metric that evaluates the similarity between automated and ground-truth manual segmentations, was selected. The training, validation, and testing data for the two-part network were split into 95, 13, and 27 subjects per fold in a 5-fold stratified cross-validation. Weighted F1 scores and AUROC evaluated the

classification performance while Pearson's correlation coefficient and Bland-Altman plots evaluated the regression performance.

The comparison between ground truth and predicted htTKV from the Attention U-Net displayed a Pearson's correlation coefficient of 0.98 with a mean percentage difference of $13.47 \pm 13.70\%$. In the two-part network, the AUROC for predicting whether a patient will reach CKD Stage 3A, 3B, and 30% eGFR decline were 0.965, 0.960, and 0.952, which performed close to or better than the performance of the image texture feature statistical model developed by Kline et al.⁴⁵ Furthermore, the precision and recall ranged between 85-90% for each prediction criterion. The comparison between ground truth and predicted eGFR percent changes displayed a Pearson's correlation coefficient of 0.81. The Bland-Altman analysis provided a mean percentage difference of $1.12 \pm 15.58\%$, with most observations placed within the 95% confidence interval. Moreover, the model reached a weighted-average F1 score of 0.851 and AUROC of 0.972 in predicting the CKD stage for each patient after eight years. Finally, Raj et al. compared their model's prediction of eGFR with the MIC tool and obtained Pearson's correlation coefficients of 0.86 and 0.64, respectively. The MIC tool underestimated the eGFR (bias of $-1.76 \text{ mL/min/1.73m}^2$), while the machine learning model overestimated (bias of $1.18 \text{ mL/min/1.73m}^2$). The Bland-Altman analyses displayed a smaller 95% confidence interval for the machine learning model compared to the MIC tool (-26.87 - $+29.22 \text{ mL/min/1.73m}^2$ vs. -44.88 - $+41.37 \text{ mL/min/1.73m}^2$). To summarize, the evaluation metrics indicate that the model is highly accurate and precise in both the automated segmentation and the prediction of the patient's kidney function outcome after eight years.

While the results look promising, the confusion matrix illustrated in the study suggests that the results may be unreliable (Figure 3). It includes the entire study population (N=135) rather than

stratifying them by training, validation, and testing subsets. Without knowledge of the exact number of TP and TN from the testing set, it becomes difficult to assess the model’s true performance on real-world data. For the same reason, the study does not mention whether the other numerical evaluation metrics are simply the mean from all subsets combined or solely from the testing set.

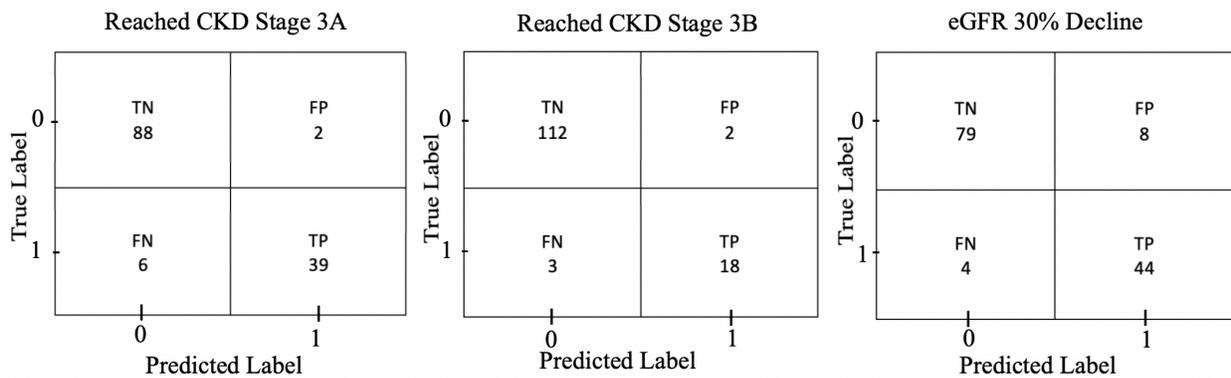


Figure 3. The confusion matrix depicting the performance of the predictive model by Raj et al.¹⁰⁹

There are several other limitations. The study is a pre-print and has yet to be peer-reviewed.

Therefore, the results should be accepted with caution. The study also only reports AUROC in predicting whether a patient will reach CKD Stage 3A, 3B, and 30% eGFR decline after eight

years. Providing other metrics, such as the F1 score, accuracy, precision, and recall, would

provide a more robust assessment of the model’s performance.¹¹⁰ Furthermore, the model should

be externally validated using cohorts from other PKD clinics. The sample size of less than 150 is

considered small in machine learning, and similar to the image texture study by Kline et al., there

is a selection bias for the CRISP study cohort. As a future step, it would be interesting to test the

model’s performance after adding other relevant clinical factors and cyst segmentations of the

MRI.

There are other predictive machine learning models in the field of nephrology unrelated to ADPKD.¹¹¹⁻¹¹³ However, ADPKD brings unique challenges to developing reliable machine learning models because of the irregularity in shape and growth of the cystic kidneys, composition of cysts, and extrarenal complications.⁷⁶ Therefore, more studies are required to determine whether machine learning can outperform existing statistical models in predicting the disease progression of ADPKD .

3. OBJECTIVE AND HYPOTHESIS

3.1. Objective

This is an exploratory study that aims to develop various machine learning models that may aid clinical decision-making in ADPKD and perform better than the current MIC tool in the future.

3.2. Study Aims and Hypotheses

Aim 1: Compare the TKV measured from the ground-truth manual segmentation to the eTKV obtained from the ellipsoid equation.

Hypothesis 1: Discordances between estimated and ground truth TKV result in misclassification in MIC, justifying the need for a different approach to estimate the TKV.

Aim 2a: Develop a machine learning model that utilizes clinical features to classify the rate of eGFR decline across the full spectrum of patients with ADPKD.

Hypothesis 2a: A machine learning approach that includes clinical features will classify the rate of eGFR decline in patients with ADPKD.

Aim 2b: Develop a deep learning model that utilizes MRI and clinical features to classify the rate of eGFR decline across a spectrum of patients with ADPKD.

Hypothesis 2b: Adding the MRI, along with clinical features, will perform better than the machine learning models that only use clinical features.

4. METHODS

The MUHC Research Ethics Board approved this retrospective cohort study without the need for individual patient consent (Project Number: 2022-8080).

4.1. Study Cohort

The study participants were derived from the MUHC PKD clinic. The MUHC PKD clinic, which opened in 2015, accepts referrals from primary care physicians and specialists. Patients are seen by one of two nephrologists who have expertise in PKD. All patients followed in the clinic were screened for eligibility. Patients over the age of 15 with a confirmed diagnosis of ADPKD and readily extractable coronal-cut abdominal MRI in Digital Imaging and Communications in Medicine (DICOM) format between December 1, 2015 and November 1, 2022 were included in the study. Patients who had transitioned to dialysis or transplantation clinics in the past due to ESKD were also included if they had a baseline MRI prior to starting kidney replacement therapies. Patients with suspected but unconfirmed diagnoses of ADPKD or who had other forms of imaging (e.g., US or CT) were excluded. Patients also required a minimum follow-up time of 1 year or at least three creatinine values.

4.2. Variables of Interest for Machine Learning

Continuous input variables for the model included age at the event of baseline MRI, BMI, the period under tolvaptan therapy, htSegTKV, eGFR (\pm 6 months from baseline MRI), systolic blood pressure (SBP), diastolic blood pressure (DBP), and manual segmented cyst volume.

Categorical input variables included sex, family history of ADPKD, tolvaptan treatment at any point in time, and diabetes and smoking statuses.

The binary output was whether the patient is a rapid or non-rapid progressor, indicated by their annual rate of eGFR decline. Patients with an annual rate below $-4 \text{ mL/min/1.73m}^2$ per year were considered rapid progressors, and vice versa.

While the definition of rapid progression using MIC class is consistent across the Asian, European, and North American consensus (Class 1C, 1D, or 1E), its definition using the rate of eGFR decline is not.³ For instance, while the Canadian expert consensus defines rapid progressors as patients with $-2.5 \text{ mL/min/1.73m}^2$ per year or greater, the PBS Australian consensus uses the threshold of $-5 \text{ mL/min/1.73m}^2$ per year or greater. Therefore, using the approximate average between the variable range of clinical threshold values ($-4 \text{ mL/min/1.73m}^2$ per year) was most appropriate.

4.3. Clinical Data Collection and Imaging Information Extraction

Patients' clinical and imaging information was extracted, between their time of baseline MRI to November 1, 2022 from Oacis, the MUHC's electronic medical record system. Only the clinical information from lab tests available to the nephrologists from the MUHC was recorded. For patients who had already transitioned to ESKD, data was recorded until the initiation of dialysis or transplantation or death or loss to follow-up.

The shortest time interval between two consecutive lab reports was three months to avoid capturing cases of acute kidney injury. If more than one lab test was done within that interval, the serum creatinine and urine albumin-to-creatinine ratio values closest to the previous test values were recorded.

The period under tolvaptan therapy in months was calculated by subtracting the treatment discontinuation date from the initiation date. However, if the patient was still under tolvaptan

treatment by the study end date, their most recent clinical lab date replaced the discontinuation date in approximating the number of months on tolvaptan therapy.

We applied the three orthogonal axes of the kidneys, previously recorded by the designated radiologist at the MUHC, to the ellipsoid equation to calculate the eTKV. If the values were unavailable or were discrepant between the progress note and the imaging report, a senior radiologist (Dr. Reinhold) repeated the measurements while remaining blinded. The eTKV and SegTKV were each used to obtain two separate MIC tool classes.

We used the CKD-Epi equation, which considers sex, age, race, and serum creatinine, to calculate patients' eGFR.¹¹⁴ This study was initiated prior to the race variable being dropped from the refit CKD-Epi study equation. We used Python (version 3.8.8.) to calculate the annual rate of eGFR decline. The np.polyfit() code fits the dates of clinical lab findings and eGFR within the polynomial function. We then multiplied 365 days per year to the slope, which was the daily rate of eGFR decline, to calculate the annual eGFR decline. Since eGFR can vary based on patients' nutrition, diet, and hydration on the lab test day, the annual rate of eGFR decline was calculated only for patients with three or more serum creatinine values. If the patient had less than three serum creatinine measurements, they were only included in Aim 1 of the study.

4.4. Clinical MRI Image Extraction

All MRIs meeting the study's inclusion criteria between December 1, 2015 and November 1, 2022 were extracted as DICOM files for manual segmentation. The extracted DICOM files were stripped of identifiers, anonymized, and coded (e.g., PKD1, PKD2...). If a patient had follow-up images, they were coded as PKD1a, PKD1b, and onwards. The patients could only be identified

using the Picture Archiving and Communications Systems (PACS) number or accession number of the images. The data scientists were not provided with these identifiers.

The extracted MRI scans were performed at three hospitals affiliated with the MUHC: Royal Victoria Hospital (RVH), Montreal General Hospital (MGH), and Lachine Hospital (LH). At the RVH, 1.5-T GE Signa Artist and 3.0-T Siemens Skyra scanners were used. The acquisition matrices were 288x224 and 256x256, respectively. The MRI images were reconstructed to 512x512 for the 1.5-T GE Signa Artist scanner. The GE and Siemens scanners reported bandwidths of 325Hz per pixel and 700Hz per pixel, respectively. At the MGH, 1.5-T GE Signa Artist was used. The acquisition matrices were 320x192 but were reconstructed to 512x512. The bandwidth was reported to be 325Hz per pixel. At the LH, 1.5-T Siemens Aera was used. The acquisition matrices were 256x256. The bandwidth was reported to be 700Hz per pixel.

The MRI was performed with the following parameters for all scanners: T2-weighted Half-Fourier Acquisition Single-Shot Turbo Spin Echo Imaging (HASTE)/ Single-Shot Fast Spin-Echo (SSFSE), 90° flip angle, 90-100ms echo time, and 2000ms repetition time.

4.5. Manual Segmentation

The coronal T2-weighted HASTE/SSFSE images were manually segmented by the M.Sc. candidate after appropriate training from a radiologist (Dr. Reinhold) using a commercially-available 3D Slicer image computing software (versions 4.11.20210226 and 5.1.0) and Wacom Intuos Wireless Graphic Tablet (10.4" x 7.8"), a free-hand drawing tool. Two separate segmentation files were created for each MRI case. The first segmentation outlined the kidney parenchyma and included exophytic cysts. The kidney pelvis and hilar vascular structures outside of the parenchyma were excluded. The second segmentation outlined only kidney cysts. Hemorrhagic, calcified, fluid-filled, and exophytic cysts were included in the cyst segmentation.

The segmentations were revised and corrected a few months after completion to avoid human error and improve the quality of the segmentation task. Ambiguous cases were adjudicated by the project supervisors, an experienced radiologist (Dr. Reinhold) and a nephrologist (Dr. Alam) with experience in ADPKD imaging. The final segmentations were saved in nearly raw raster data (.Nrrd) format.

The SegTKV and cyst volume were computed using the *Segmentation Statistics* tool pre-installed in the *3D Slicer* program. The tool multiplied the voxel volume — obtained by the image spacing — by the number of segmented kidney voxels to calculate the SegTKV in cm³. The height-adjusted manual segmented TKV (htSegTKV) was obtained by dividing the SegTKV in mL by the patient’s height in metres.

4.6. Computer Features

Apple M2 with 8GB of RAM and 256GB disk was used to calculate the annual eGFR decline on R Studio Version 2022.12.0+353. Intel Core™ i7-7700 with 8GB of RAM and 238GB disk was used to train, validate, and test machine learning models on the clinical features. For developing the deep learning model, the NVIDIA Tesla V100S PCIe GPU, featuring 32GB of VRAM, and an Intel® Xeon® Gold 5220 CPU at 2.20GHz, was used. We used the PyTorch 1.10.1+CU102 framework in Python 3.8. to implement the pipeline and develop the model. The Albumentation package was used for image augmentation.¹¹⁵

4.7. Data Pre-Processing

The output, the annual rate of eGFR decline, was simplified into a binary class: rapid and non-rapid progressors. The absence of documentation in patients’ progress notes for family history, hypertension, nephrolithiasis, and smoking status was assumed to be negative. Missing height

was imputed as the average of all female or male heights, depending on the patient's sex.

Missing SBP and DBP were imputed using the mode of all patients to avoid bias from outliers.

Missing BMI was imputed as the average value since it followed a normal distribution. Patients on tolvaptan for three months or less were assumed to be non-takers. For patients without an eGFR within the six months of baseline imaging, any eGFR closest to the baseline MRI date and similar to the remaining eGFR values was used.

For the clinical model, Z-score normalization (*sklearn.preprocessing.StandardScaler* on Python) was used on time on tolvaptan treatment, age, htSegTKV, eGFR, SBP, DBP, cyst volume, and BMI. For the deep model, we normalized all clinical features individually to map all the features to the same range of 0 and 1.

We alleviated the complexity of processing volumetric 3D MRI images by sampling three key 2D image slices to represent the patient's MRI. For each patient, we extracted the 2D MRI slice with the largest manual segmentation area, n , and its two neighbouring slices, $n-1$ and $n+1$. The

equation $I = \frac{I - \min(I)}{\max(I) - \min(I)}$, where I is the image slice, was used to normalize and map I 's pixel

intensities to a range of 0 and 1; $\min(I)$ and $\max(I)$ calculated the minimum and maximum pixel intensity from the image slice. Moreover, the corresponding 2D mask and patient clinical features were saved along with the extracted 2D slices. It is worth noting that the clinical features are consistent across the three 2D slices as they pertain to the same patient. By employing this strategy, the sample size tripled the original sample size.

Data augmentation techniques, such as padding, resizing, and normalization transformations were applied to our training, validation, and test sets while an extensive list of pixel-level (color) and spatial-level transformations were only applied to the images in the training set. These implementations further increased the variation of our small dataset and assisted the model with

more training samples. This would decrease the chance of overfitting and increase the deep model's generalizability and robustness.

To avoid any information leak in our data processing pipeline for the deep model, we utilized a group-based stratified random splitting approach to divide the dataset into train:validation:test subsets.

4.8. Machine Learning Model Development

4.8.1. Clinical Model

The clinical model applied clinical and imaging variables as the input to categorize patients as rapid or non-rapid progressors. Two different models were implemented using the RF and SVM Linear Kernel approaches. The model with the higher evaluation metric was chosen as the final model and compared against the deep model.

5-fold cross-validation was performed, and the train:validation:test split was 80:10:10 per fold in both model techniques. There was no hyperparameter tuning during the validation stage. For each technique, 10 models of the same hyperparameters were trained independently with random subsets of the split data. After each model had the opportunity to tune the parameters in the validation stage, they were all tested on the same 10% of the unseen data. The performance metrics among the 10 models were averaged for reporting purposes, but the best-performing model was selected as the final model.

4.8.2. Deep Model Architecture

Our deep model classifier consisted of three components: 1) EfficientNet-b2, 2) FuseNet, and 3) Classifier (Figure 4). We used the pre-trained EfficientNet-b2 model (listed as 2DExNet in Figure 4) as our feature extractor to extract 1000 features from the MRI slices.¹¹⁶ This entirely pre-trained model with its fully connected layer was used without making any modifications to

its architecture. Additionally, the FuseNet model performed the feature fusion to generate 15 new features from the input clinical features. The FuseNet architecture was composed of three fully connected layers with neuron sizes of 15, 32, and 15, respectively. The feature maps from the two aforementioned models were concatenated and fed as input to the final Classifier, which consisted of three fully connected layers with neuron sizes 1015, 256, and 2, respectively. This approach effectively combines information from both the MRI and clinical features and accurately classifies the patient as either a rapid or non-rapid progressor.

The train:validation:test subset ratio was 63%:17%:20%. We trained the model for 50 epochs using the Adam optimizer with a learning rate of 1E-4 and a batch size of 32. The ExponentialLR scheduler with a gamma size of 9E-1 was employed to adjust the learning rate. Cross Entropy Loss was used to calculate the loss function throughout training. The model's performance was assessed on the validation subset during each training epoch by calculating the accuracy, precision, recall, and F1 scores. During validation, early stopping was employed to avoid overfitting and the model that produced the highest F1 score was chosen as the best-performing model for testing. During the testing stage, the final prediction label for each patient was determined by taking the majority vote between the predicted labels for the three 2D MRI slices (n-1, n, n+1) extracted from each patient, and considered as the final patient class label.

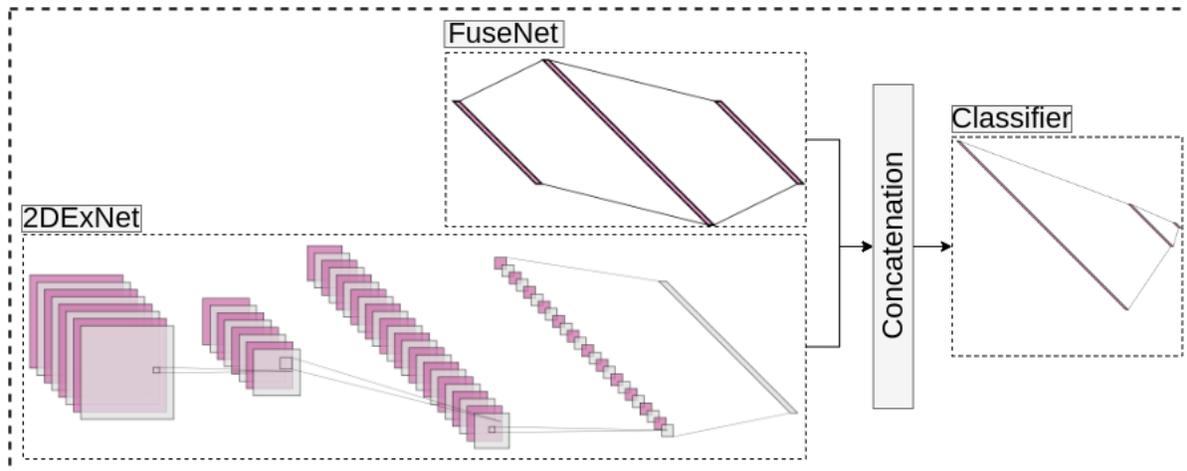


Figure 4. The architecture of the deep learning model for processing 2D images (2DExNet), their corresponding clinical features (FuseNet), and the classifier to categorize patients as either rapid or non-rapid progressors.

4.9. Statistical Analysis

The statistical calculations were made using Excel, Version 16.74 (Microsoft Corporation 2023) and R Studio Version 2022.12.0+353. A p-value less than or equal to 0.05 was considered statistically significant. The descriptive statistics were expressed as mean \pm standard deviation (SD) or median and interquartile range for continuous variables and frequency in percentage for categorical variables. The independent samples t-test and chi-square test were performed to compare the baseline characteristics between rapid and non-rapid progressors in the Aim 2 study cohort. The independent samples t-test and chi-square test were also performed to compare the annual rate of eGFR decline and the proportion of rapid progressors, respectively, for each MIC class.

A correlation plot and Pearson's correlation coefficient were used to illustrate and assess the concordance between htSegTKV and height-adjusted estimated TKV (hteTKV). Moreover, a confusion matrix was used to compare the misclassifications in MIC between the ground-truth

manual segmentation and the ellipsoid equation. Finally, a Giavarina plot was created to analyze the bias between the mean percentage differences between htSegTKV and hteTKV and evaluate the agreement between the two values.

For the RF approach, the accuracy, precision, recall, F1 score, AUROC, and feature importance from the test set were reported to evaluate the models' abilities to classify patients correctly. For the SVM approach, the accuracy, precision, recall, and F1 score from the test set were reported. Note that we used the weighted average of these four metrics for all models. For the deep learning model, the accuracy, precision, recall, F1 score, and AUROC from the aggregated labels of the test set were reported. Furthermore, the confusion matrices from both the validation and test subsets were presented.

$$Precision = \left(\frac{TP}{TP + FP} \right)$$

$$Recall = \left(\frac{TP}{TP + FN} \right)$$

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} = \frac{TP}{TP + 0.5(FP + FN)}$$

$$Weighted\ F1\ Score = \sum_{i=1}^{n=2} Support_i \times F1\ Score_i$$

$$Support_i = \frac{Number\ of\ Samples\ in\ Class\ i}{Total\ Number\ of\ Samples}$$

5. RESULTS

5.1. Study Participant Characteristics

231 patients were registered in the MUHC PKD clinic (Figure 5). Of those, 38 patients with currently undiagnosed/unconfirmed diagnoses of ADPKD, 37 patients with no MRI or restricted access, and 11 patients with MRI performed at a non-MUHC institution were excluded from the study. Of the eligible 145 patients, 25 patients had less than three serum creatinine values and were excluded from the second aim of the study. Therefore, 145 patients and 120 patients were included in the study's first (Manual Segmentation vs. Ellipsoid Equation) and second aims (Machine Learning Model), respectively. A total of 229 MRI images (145 baseline and 84 follow-up images) were extracted. However, two follow-up MRI were excluded during manual segmentation due to low resolution, leaving a total of 145 baseline and 82 follow-up images for analysis in the study.

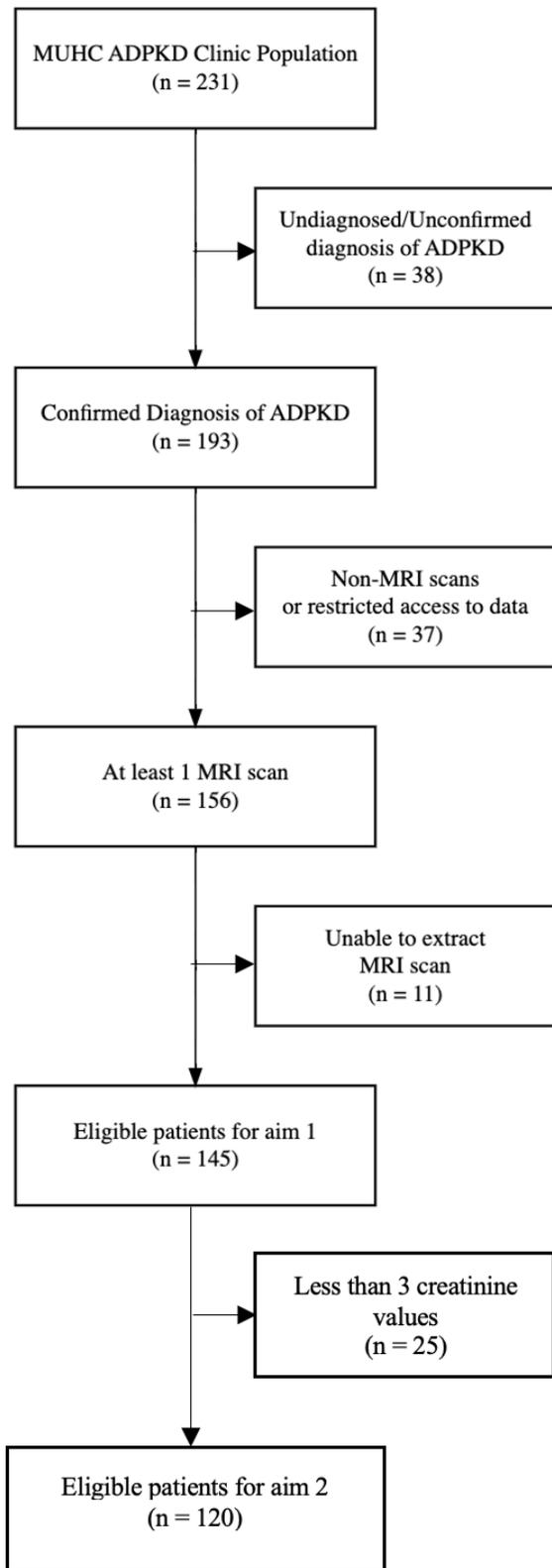


Figure 5. Enrollment, Exclusion, and Inclusion Based on Screening Criteria

Table 4 illustrates the baseline characteristics of the MUHC PKD clinic (N=231), Aim 1 patient cohort (N=145), and Aim 2 patient cohort (N=120). In the Aim 1 patient cohort, the htSegTKV and hteTKV were 865.0 mL [490.3-1306.6] and 773.5 mL [435.6-1244.4]. Two patients had missing height values; consequently, their heights were imputed to calculate htSegTKV and hteTKV. For the Aim 1 study, the MIC using SegTKV was as follows: Class 1A (N=11), Class 1B (N=27), Class 1C (N=56), Class 1D (N=32), Class 1E (N=15), Class 2 (N=2). the MIC using eTKV was as follows: Class 1A (N=14), Class 1B (N=32), Class 1C (N=54), Class 1D (N=27), Class 1E (N=14), Class 2 (N=2).

The mean age of patients from Aim 2 patients was 46.3 ± 14.2 years. 73.3% had a family history of ADPKD, and 79.2% were diagnosed with hypertension. The median htSegTKV and hteTKV were 939.5 mL [542.0-1372.5] and 852.5 mL [481.7-1273.4], respectively. One patient had a missing height value; once again, the height was imputed to calculate htSegTKV and hteTKV.

The manual segmented cyst volume was 960.6 mL [390.7-1606.1]. For the Aim 2 study, the MIC using htSegTKV was as follows: Class 1A (N=10), Class 1B (N=20), Class 1C (N=48), Class 1D (N=27), Class 1E (N=12), Class 2 (N=2). The MIC using hteTKV was as follows: Class 1A (N=12), Class 1B (N=23), Class 1C (N=47), Class 1D (N=24), Class 1E (N=11), Class 2 (N=2).

The baseline eGFR was 67.3 mL/min/1.73m² [45.1-98.5], and the annual rate of eGFR decline was -3.44 mL/min/1.73m² per year [-5.42--1.65]. Of the 120 patients, 50 patients were classified as rapid progressors and 70 patients were classified as non-rapid progressors.

Table 5 describes the baseline characteristics of the Aim 2 patient cohort stratified by disease progression. In rapid progressors, the mean age was 48.4 ± 13.4 years. 72.0% had a family history of ADPKD, and 96.0% were diagnosed with hypertension. The median htSegTKV and hteTKV were 1185.5 mL [939.5-1637.9] and 1136.0 mL [834.7-1418.5]. The manual segmented

cyst volume was 1305.7 mL [960.1-2095.0]. The MIC using htSegTKV was as follows: Class 1A (N=0), Class 1B (N=5), Class 1C (N=20), Class 1D (N=16), Class 1E (N=7), Class 2 (N=1). The MIC using hteTKV was as follows: Class 1A (N=0), Class 1B (N=6), Class 1C (N=22), Class 1D (N=14), Class 1E (N=6), Class 2 (N=1). The baseline eGFR was 57.8 mL/min/1.73m² [43.8-75.1], and the annual rate of eGFR decline was -5.79 mL/min/1.73m² per year [-4.60--7.36]. In non-rapid progressors, the mean age was 44.8 ± 14.6 years. 74.3% had a family history of ADPKD, and 67.1% were diagnosed with hypertension. The median htSegTKV and hteTKV were 653.8 mL [384.3-1092.9] and 601.4 mL [351.8-1033.5]. The manual segmented cyst volume was 475.7 mL [186.0-1418.4]. The MIC using htSegTKV was as follows: Class 1A (N=10), Class 1B (N=15), Class 1C (N=28), Class 1D (N=11), Class 1E (N=5), Class 2 (N=1). The MIC using hteTKV was as follows: Class 1A (N=12), Class 1B (N=17), Class 1C (N=25), Class 1D (N=10), Class 1E (N=5), Class 2 (N=1). The baseline eGFR was 82.7 mL/min/1.73m² [47.0-106.8], and the annual rate of eGFR decline was -3.45 mL/min/1.73m² per year [-3.00--0.88]. The differences in hypertension, polycystic liver disease, RAASi, tolvaptan at any given time, time on tolvaptan, SegTKV, eTKV, cyst volume, htSegTKV, hteTKV, htSegTKV MIC Class 1A and Class 1D, hteTKV MIC Class 1A, eGFR, and the annual rate of eGFR decline between rapid and non-rapid progressors were statistically significant.

Table 6 further describes the Aim 2 patients' observed annual rate of eGFR decline and the proportion of rapid progressors in each MIC class. A negative correlation was found between each MIC and the annual rate of eGFR decline. When stratified using hteTKV, the annual rate of eGFR decline and proportion of rapid progressors in each MIC class was as follows: Class 1A (-0.98 mL/min/1.73m² per year, 0.0%), Class 1B (-2.94 mL/min/1.73m² per year, 26.0%), Class 1C (-3.87 mL/min/1.73m² per year, 47.9%), Class 1D (-4.46 mL/min/1.73m² per year, 58.3%),

Class 1E (-6.47 mL/min/1.73m² per year, 54.5%), and Class 2 (-2.42 mL/min/1.73m² per year, 50.0%). When stratified using htSegTKV, the annual rate of eGFR decline and proportion of rapid progressors in each MIC class was as follows: Class 1A (-1.43 mL/min/1.73m² per year, 0.0%), Class 1B (-2.46 mL/min/1.73m² per year, 25.0%), Class 1C (-3.40 mL/min/1.73m² per year, 42.9%), Class 1D (-4.44 mL/min/1.73m² per year, 59.3%), Class 1E (-5.33 mL/min/1.73m² per year, 58.3%), and Class 2 (-2.42 mL/min/1.73m² per year, 50.0%). There was no statistically significant difference between patients in each class for both the annual rate of eGFR decline and the proportion of rapid progressors.

	PKD Clinic (N=231)	Study-Eligible Patients (N=145)	Model Development (N=120)
<u>DEMOGRAPHICS</u>			
Age at baseline MRI (years)		44.8 ± 14.7	46.3 ± 14.2
Missing, n (%)		0 (0.0)	0 (0.0)
Male, n (%)	122 (52.8)	67 (46.2)	54 (45.0)
Missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
Non-Black, n (%)	221 (95.7)	138 (95.2)	114 (95.0)
Missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
Family History, n (%)	154 (67.0)	108 (74.5)	88 (73.3)
Missing, n (%)	1 (0.4)	0 (0.0)	0 (0.0)
<u>COMORBIDITIES AND CLINICAL VARIABLES</u>			
Hypertension, n (%)	155 (67.4)	103 (71.0)	95 (79.2)
Missing, n (%)	1 (0.4)	0 (0.0)	0 (0.0)
Hypertension before the age of 35, (%)	74 (32.2)	46 (31.7)	42 (35.0)
Missing, n (%)	1 (0.4)	0 (0.0)	0 (0.0)
Urologic Event before the age of 35, n (%)	56 (24.5)	35 (24.1)	30 (25.0)
Missing, n (%)	2 (0.8)	0 (0.0)	0 (0.0)
Nephrolithiasis, n (%)	28 (12.2)	16 (11.0)	13 (10.8)
Missing, n (%)	2 (0.8)	0 (0.0)	0 (0.0)
Polycystic Liver Disease, n (%)	153 (66.8)	109 (75.2)	93 (77.5)
Missing, n (%)	2 (0.8)	0 (0.0)	0 (0.0)
Intracranial Aneurysm, n (%)	15 (6.6)	9 (6.2)	8 (6.7)

Missing, n (%)	2 (0.8)	0 (0.0)	0 (0.0)
Renin-Angiotensin-Aldosterone-System Inhibitor, n (%)	131 (57.2)	90 (62.1)	83 (69.2)
Missing, n (%)	2 (0.8)	0 (0.0)	0 (0.0)
Systolic Blood Pressure (mmHg)		121 [113-130]	122 [115-132]
Missing, n (%)		9 (6.2)	7 (5.8)
Diastolic Blood Pressure (mmHg)		78 [72-83]	80 [72-84]
Missing, n (%)		9 (6.2)	7 (5.8)
Tolvaptan at any given time, n (%)	74 (32.2)	65 (44.8)	63 (52.5)
Missing, n (%)	1 (0.4)	0 (0.0)	0 (0.0)
Time on Tolvaptan (months)		42.3 ± 24.5	43.7 ± 24.0
Missing, n (%)		0 (0.0)	0 (0.0)
Body Mass Index, (kg/m²)		26.6 [23.6-30.3]	26.6 [24.0-31.3]
Missing, n (%)		27 (18.6)	25 (20.8)
Diagnosis of Diabetes, n (%)	17 (7.4)	9 (6.21)	9 (7.5)
Missing, n (%)	2 (0.8)	0 (0.0)	0 (0.0)
Smoking Status, n (%)			
Non-Smoker	180 (78.6)	116 (80.0)	97 (80.8)
Smoker	35 (15.3)	21 (14.5)	17 (14.2)
Ex-Smoker	14 (6.1)	7 (4.8)	6 (5.0)
Missing, n (%)	2 (0.8)	0 (0.0)	0 (0.0)
Height (m)		1.70 [1.64-1.78]	1.70 [1.64-1.78]
Missing, n (%)		2 (1.4)	1 (0.8)
<u>MRI MEASUREMENTS</u>			

Manual-Segmented TKV (mL)		1435.8 [819.4-2284.7]	1576.1 [921.4-2381.0]
Missing, n (%)		0 (0.0)	0 (0.0)
Estimated TKV		1272.0 [746.0-2145.1]	1397.0 [821.0-2178.6]
Missing, n (%)		0 (0.0)	0 (0.0)
Cyst Volume (mL)		882.9 [329.2-1503.0]	960.6 [390.7-1606.1]
Missing, n (%)		0 (0.0)	0 (0.0)
Height-adjusted Manual-Segmented TKV (mL)		865.0 [490.3-1306.6]	939.5 [542.0-1372.5]
Missing, n (%)		2 (1.4)	1 (0.8)
Height-adjusted Estimated TKV (mL)		773.5 [435.6-1244.4]	852.5 [481.7-1273.4]
Missing, n (%)		2 (1.4)	1 (0.8)
Mayo Imaging Class Using Manual-Segmented TKV, n (%)			
Class 1A, n (%)		11 (7.6)	10 (8.3)
Class 1B, n (%)		27 (18.6)	20 (16.7)
Class 1C, n (%)		56 (38.6)	48 (40)
Class 1D, n (%)		32 (22.1)	27 (22.5)
Class 1E, n (%)		15 (10.3)	12 (10.0)
Class 2, n (%)		2 (1.4)	2 (1.7)
Missing, n (%)		2 (1.4)	1 (0.8)
Mayo Imaging Class Using Estimated TKV, n (%)			
Class 1A, n (%)		14 (9.7)	12 (10.0)
Class 1B, n (%)		32 (22.1)	23 (19.2)
Class 1C, n (%)		54 (37.2)	47 (39.2)

Class 1D, n (%)		27 (18.6)	24 (20.0)
Class 1E, n (%)		14 (9.7)	11 (9.2)
Class 2, n (%)		2 (1.4)	2 (1.7)
Missing, n (%)		2 (1.4)	1 (0.8)
<u>RENAL OUTCOME</u>			
eGFR (ml/min/1.73m²)		76.5 [45.7-106.9]	67.3 [45.1-98.5]
Missing, n (%)		10 (6.9)	0 (0.0)
Annual Rate of eGFR Decline (ml/min/1.73m² per year)		-3.45 [-5.42--1.65]	-3.44 [-5.42--1.65]
Missing, n (%)		25 (17.2)	0 (0.0)

Table 4. Baseline Characteristics of Patients from the MUHC PKD Clinic, Study Aim 1, and Study Aim 2 (Left to Right). Plus-minus values are means±SD. Square bracket values are median [1st quartile-3rd quartile].

	Rapid Progressors (N=50)	Non-Rapid Progressors (N=70)	p-value
<u>DEMOGRAPHICS</u>			
Age at baseline MRI (years)	48.4±13.4	44.8±14.6	0.17
Missing, n (%)	0 (0.0)	0 (0.0)	
Male, n (%)	25 (50.0)	41 (58.6)	0.35
Missing, n (%)	0 (0.0)	0 (0.0)	
Non-Black, n (%)	47 (94.0)	67 (95.7)	0.68
Missing, n (%)	0 (0.0)	0 (0.0)	
Family History, n (%)	36 (72.0)	52 (74.3)	0.78
Missing, n (%)	0 (0.0)	0 (0.0)	
<u>COMORBIDITIES AND CLINICAL VARIABLES</u>			
Hypertension, n (%)	48 (96.0)	47 (67.1)	<0.001
Missing, n (%)	0 (0.0)	0 (0.0)	
Hypertension before the age of 35, (%)	22 (44.0)	20 (28.6)	0.08
Missing, n (%)	0 (0.0)	0 (0.0)	
Urologic Event before the age of 35, n (%)	16 (32.0)	14 (20.0)	0.14
Missing, n (%)	0 (0.0)	0 (0.0)	
Nephrolithiasis, n (%)	6 (12.0)	7 (10.0)	0.73
Missing, n (%)	0 (0.0)	0 (0.0)	
Polycystic Liver Disease, n (%)	46 (92.0)	47 (67.1)	<0.05
Missing, n (%)	0 (0.0)	0 (0.0)	
Intracranial Aneurysm, n (%)	6 (12.0)	2 (2.9)	0.05
Missing, n (%)	0 (0.0)	0 (0.0)	

Renin-Angiotensin-Aldosterone-System Inhibitor, n (%)	43 (86.0)	40 (57.1)	<0.001
Missing, n (%)	0 (0.0)	0 (0.0)	
Systolic Blood Pressure (mmHg)	123 [115-131]	122 [115-132]	0.58
Missing, n (%)	2 (4.0)	5 (7.1)	
Diastolic Blood Pressure (mmHg)	80 [73-84]	78 [72-84]	0.68
Missing, n (%)	2 (4.0)	5 (7.1)	
Tolvaptan at any given time, n (%)	35 (70.0)	28 (40.0)	<0.05
Missing, n (%)	0 (0.0)	0 (0.0)	
Time on Tolvaptan (months)	33.0±29.5	15.7 ± 24.9	<0.001
Missing, n (%)	0 (0.0)	0 (0.0)	
Body Mass Index (kg/m²)	26.0 [23.6-29.8]	26.9 [24.3-31.2]	0.59
Missing, n (%)	9 (18.0)	16 (22.9)	
Diagnosis of Diabetes, n (%)	4 (8.0)	5 (7.1)	0.85
Missing, n (%)	0 (0.0)	0 (0.0)	
Smoking Status, n (%)			
Non-Smoker	40 (80.0)	57 (81.4)	0.85
Smoker	10 (20.0)	7 (10.0)	0.12
Ex-Smoker	0 (0.0)	6 (8.6)	0.03
Missing, n (%)	0 (0.0)	0 (0.0)	
Height (m)	1.68 [1.64-1.78]	1.70 [1.64-1.77]	0.74
Missing, n (%)	1 (2.0)	0 (0.0)	
MRI MEASUREMENTS			
Manual-Segmented TKV (mL)	2008.5 [1569.4-2757.9]	1117.3 [642.7-1909.6]	<0.001
Missing, n (%)	0 (0.0)	0 (0.0)	

Estimated TKV (mL)	1946.6 [1350.3-2582.0]	1012.4 [598.4-1832.7]	<0.05
Missing, n (%)	0 (0.0)	0 (0.0)	
Cyst Volume (mL)	1305.7 [960.1-2095.0]	475.7 [186.0-1418.4]	<0.05
Missing, n (%)	0 (0.0)	0 (0.0)	
Height-adjusted Manual-Segmented TKV (mL)	1185.5 [939.5-1637.9]	653.8 [384.3-1092.9]	<0.001
Missing, n (%)	1 (2.0)	0 (0.0)	
Height-adjusted Estimated TKV (mL)	1136.0 [834.7-1418.5]	601.4 [351.8-1033.5]	<0.05
Missing, n (%)	1 (2.0)	0 (0.0)	
Mayo Imaging Class Using Manual-Segmented TKV, n (%)			
Class 1A, n (%)	0 (0.0)	10 (14.3)	<0.05
Class 1B, n (%)	5 (10.0)	15 (21.4)	0.10
Class 1C, n (%)	20 (40.0)	28 (40.0)	1.00
Class 1D, n (%)	16 (32.0)	11 (15.7)	<0.05
Class 1E, n (%)	7 (14.0)	5 (7.1)	0.22
Class 2, n (%)	1 (2.0)	1 (1.4)	0.80
Missing, n (%)	1 (2.0)	0 (0.0)	
Mayo Imaging Class Using Estimated TKV, n (%)			
Class 1A, n (%)	0 (0.0)	12 (17.1)	<0.05
Class 1B, n (%)	6 (12.0)	17 (24.3)	0.10
Class 1C, n (%)	22 (44.0)	25 (35.7)	0.36
Class 1D, n (%)	14 (28.0)	10 (14.3)	0.07
Class 1E, n (%)	6 (12.0)	5 (7.1)	0.36
Class 2, n (%)	1 (2.0)	1 (1.4)	0.80

Missing, n (%)	1 (2.0)	0 (0.0)	
<u>RENAL OUTCOME</u>			
eGFR (ml/min/1.73m²)	57.8 [43.8-75.1]	82.7 [47.0-106.8]	<0.05
Missing, n (%)	0 (0.0)	0 (0.0)	
Annual Rate of eGFR Decline (ml/min/1.73m² per year)	-5.79 [-4.60--7.36]	-3.45 [-3.00--0.88]	<0.001
Missing, n (%)	0 (0.0)	0 (0.0)	

Table 5. Baseline Characteristics of Rapid Progressors and Non-Rapid Progressors from Aim 2.

Plus-minus values are means±SD. Square bracket values are median [1st quartile-3rd quartile].

	Annual eGFR Decline (Height-Adjusted Estimated TKV) (mL/min/1.73m² per year) (N = 120)	Annual eGFR Decline (Height-Adjusted Manual-Segmented TKV) (mL/min/1.73m² per year) (N = 120)	p-value	Rapid Progressors (Height-Adjusted Estimated TKV), n (%)	Rapid Progressors (Height-Adjusted Manual-Segmented TKV), n (%)	p-value
Class 1A	-0.98 [-3.35--0.45]	-1.43 [-3.35--0.45]	0.97	0, (0.0)	0, (0.0)	
Class 1B	-2.94 [-4.40--0.98]	-2.46 [-4.40--0.84]	0.94	6, (26.0)	5 (25.0)	0.86
Class 1C	-3.87 [-5.20--1.65]	-3.40 [-5.20--1.65]	0.79	23 (47.9)	21 (42.9)	0.44
Class 1D	-4.46 [-7.60--2.70]	-4.44 [-6.60--2.70]	0.96	14 (58.3)	16 (59.3)	0.88
Class 1E	-6.47 [-8.55--2.68]	-5.33 [-8.55--3.24]	0.70	6 (54.5)	7 (58.3)	0.55
Class 2	-2.42 [-7.41-2.57]	-2.42 [-7.41-2.57]	1.00	1 (50.0)	1 (50.0)	1.00

Table 6. Annual Rates of eGFR Decline and Proportion of Rapid Progressors Within Each Mayo

Imaging Class Using eTKV and manually segmented TKV. Square bracket values are median [1st quartile-3rd quartile].

5.2. Manual Kidney and Cyst Segmentations of MRI Images

Figure 6 displays an example of the manual segmentation tasks. It highlights the irregularity and diversity of kidney size, cystic burden, and disease progression among patients with ADPKD.

The green label outlines the kidney parenchyma, while the red label outlines only the cysts. The Class 1A (TKV=370 mL) patient presents with a symmetrical bilateral distribution of innumerable small cysts with a few medium-sized cysts. Meanwhile, the Class 1D patient (TKV=5503 mL) exhibits a symmetrical bilateral distribution of larger kidney cysts that replace most of the kidney parenchyma. In contrast, the Class 2 patient (2247.4 mL) presents with much larger cysts, but similar in number, in the left kidney relative to the right kidney. Furthermore, the annual eGFR decline is unique to each patient (-2.19 mL/min/1.73m² per year vs. -6.60 mL/min/1.73m² per year vs. 2.57 mL/min/1.73m² per year).

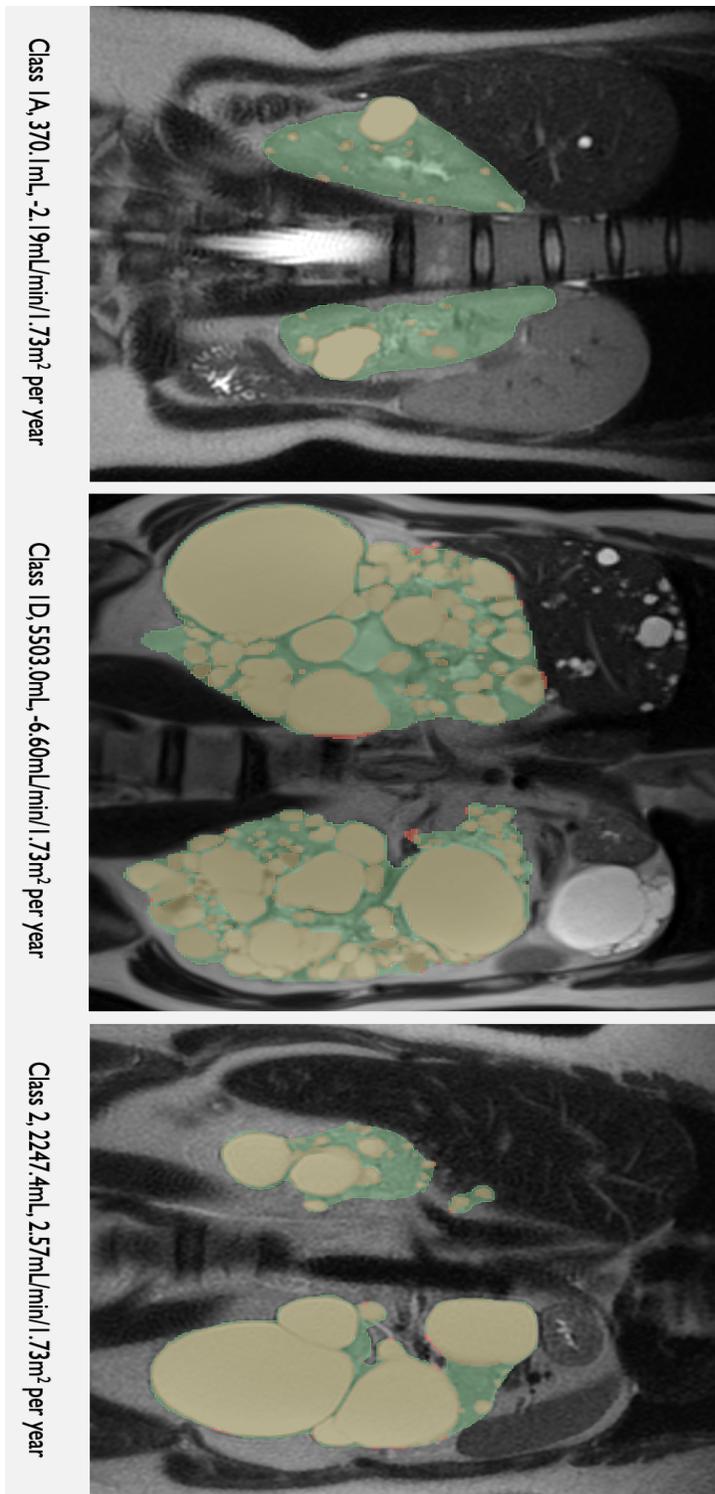


Figure 6. Mayo Imaging Class, TKV, and Annual Rate of eGFR Decline, and MRI Segmentations. The green label outlines the kidney parenchyma, while the red label outlines only the cysts.

5.3. Aim 1: Manual Segmentation vs. Ellipsoid Equation

Pearson's correlation coefficient of 0.96 ($p < 0.05$) indicated a high concordance between htSegTKV and hteTKV with few outliers (Figure 7). The Giavarina Plot reported a mean percentage difference of -7.64% between hteTKV and htSegTKV (Figure 8). However, the limits of agreement were between -40.43% and +25.15%, with six cases outside of the 95% confidence interval. Moreover, 25 patients (17.2%) had greater or equal to 20% absolute difference between hteTKV and htSegTKV. Both the Giavarina Plot and confusion matrix illustrated that the ellipsoid equation underestimates the kidney volume at lower volumes and could overestimate at higher volumes compared to manual segmentation.

The confusion matrix displayed 24 misclassifications (16.6%) within one risk class between the ellipsoid equation and manual segmentation approaches (Figure 9). Of those, most misclassifications lie between Class 1C and Class 1D. The ellipsoid equation method misclassified two Class 1C patients as Class 1D and eight Class 1D patients as Class 1C. Moreover, the ellipsoid equation misclassified eight Class 1C patients (33.3% of all misclassifications) as Class 1B. The htSegTKV classified 27 (18.6%) and 56 (38.6%) patients in Classes 1B and 1C, respectively. Meanwhile, hteTKV classified 32 (22.1%) and 54 (37.2%) patients in Classes 1B and 1C.

Of the 25 patients (17%) who exhibited $\geq 20\%$ difference between the two measures, 23 patients possessed irregular kidney shapes due to large exophytic cysts. Figure 10 provides an example of two patients within the study cohort of similar age and htSegTKV. It is important to note that both patients possess dominant exophytic cysts. The percentage difference between the two kidney measurements is 57.1% for the leftmost patient. According to the MIC tool and the eTKV, this patient is classified as Class 1B, which predicts an annual rate of eGFR decline of -

1.33 mL/min/1.73m² per year.²⁶ In contrast, when htSegTKV is used, the patient is classified as Class 1C, which predicts an annual eGFR decline of -2.63 mL/min/1.73m² per year. The patient's clinical data from the progress note indicates their true eGFR decline as -2.94 mL/min/1.73m² per year. Although the rightmost patient also possesses numerous large exophytic cysts, the percentage difference for the rightmost patient in the kidney measurements is only 1.5%. Moreover, the patient is classified correctly, and the annual eGFR decline of prediction and ground truth are similar in value.

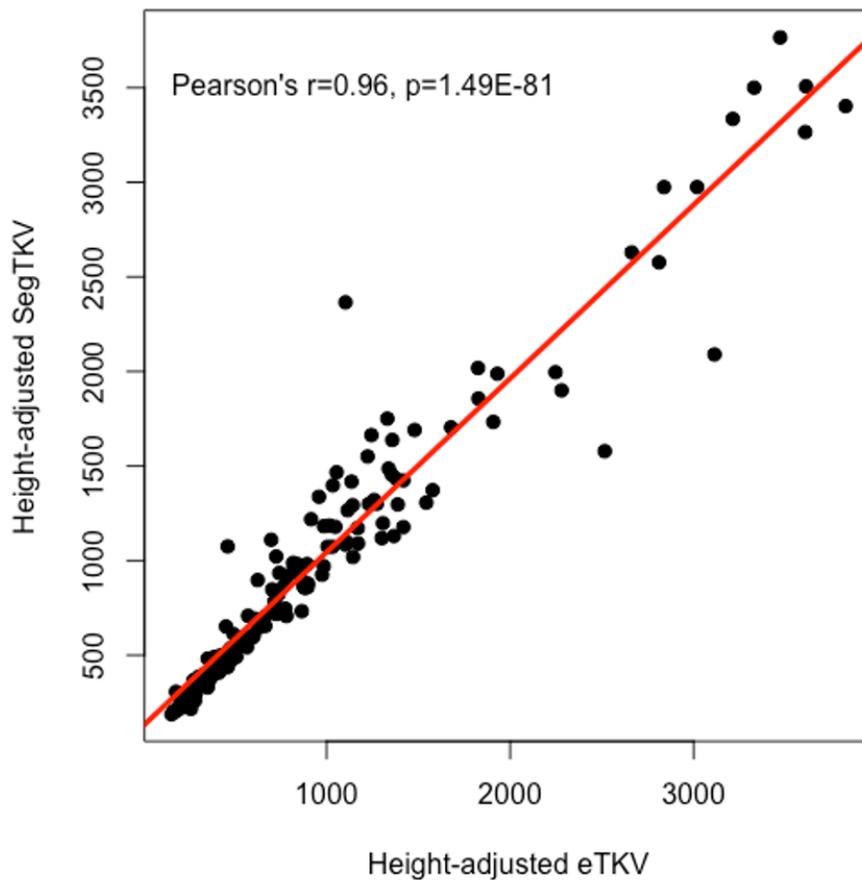


Figure 7. Linear Regression Analysis Comparing the Height-Adjusted Segmented TKV from Manual Segmentation with Height-Adjusted eTKV from the Ellipsoid Equation

Giavarina Plot for hteTKV and htSegTKV

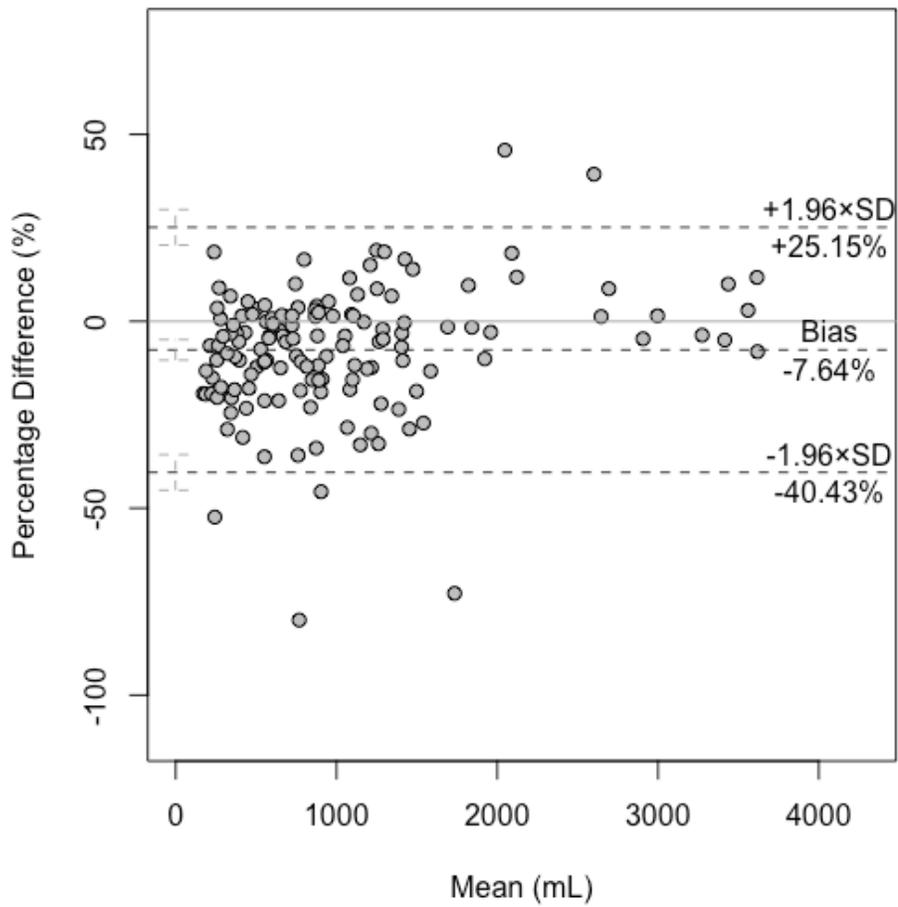


Figure 8. Giavarina Plot Comparing the Height-Adjusted Segmented TKV with Height-Adjusted eTKV

SegTKV	eTKV MIC							
	MIC	1A	1B	1C	1D	1E	2	Missing
1A		11	0	0	0	0	0	0
1B		3	24	0	0	0	0	0
1C		0	8	46	2	0	0	0
1D		0	0	8	23	1	0	0
1E		0	0	0	2	13	0	0
2		0	0	0	0	0	2	0
Missing		0	0	0	0	0	0	2

Figure 9. Confusion Matrix Highlighting the Misclassifications Resulting from the Use of the Ellipsoid Equation

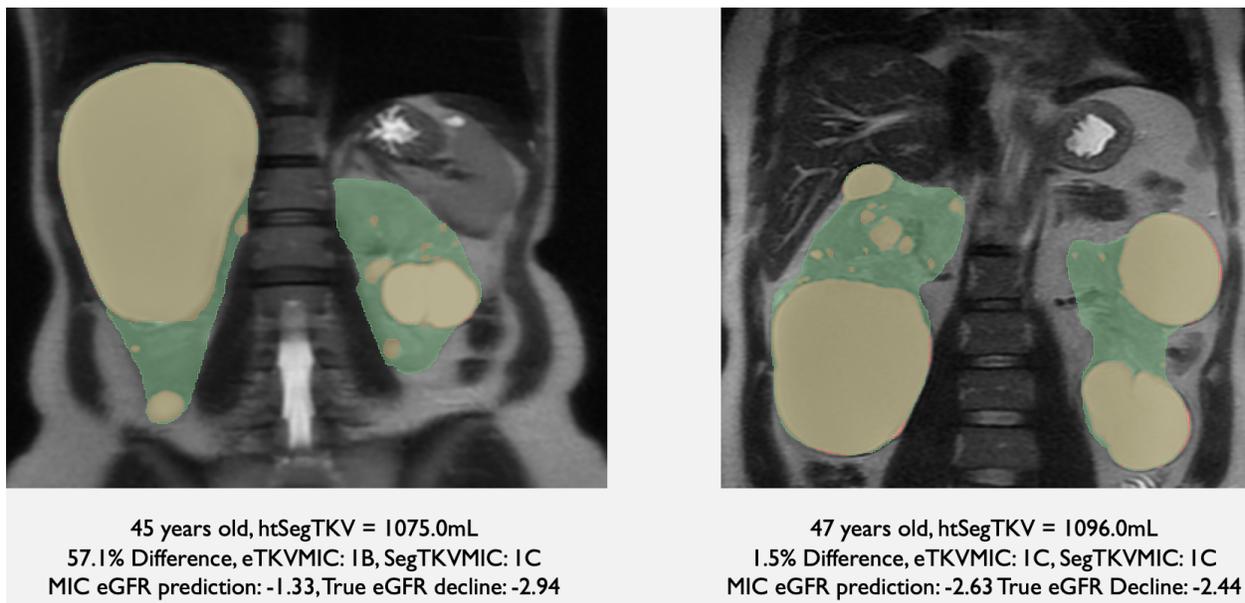


Figure 10. Inconsistency in the Ellipsoid Equation is Observed in Two Patients with Similar Age, Height-Adjusted TKV, and Identical Mayo Imaging Class

5.4. Aim 2: Machine Learning Model Performance

5.4.1. Random Forest Approach Using Only Clinical Variables

Before normalizing the data, the RF model predicted the rate of disease progression with accuracy and AUROC of 0.64 and 0.72 on the testing set. The precision, recall, and F1 score were 0.75, 0.62, and 0.67. However, after normalization, the accuracy and AUROC improved to 0.65 and 0.73. The precision, recall, and F1 score improved to 0.76, 0.63, and 0.68. The top six features were hypertension history, whether the patient was on tolvaptan, the time of tolvaptan treatment, htSegTKV, cyst volume, and baseline eGFR.

5.4.2. Support Vector Machine Approach Using Only Clinical Variables

Before normalization, the SVM model displayed an accuracy, precision, recall, and F1 score of 0.54, 0.69, 0.54, and 0.52, respectively, on the testing set. After normalization, all values improved to 0.81.

5.4.3. Deep Learning Approach Using Clinical Variables and Imaging Features

The best-performing model obtained accuracy, precision, recall, and F1 score of 0.84, 0.85, 0.84, and 0.85, respectively, on the validation set. The TP, TN, FP, and FN values were 0.85, 0.92, 0.08, and 0.15, respectively (Figure 11).

The same model obtained the accuracy, precision, recall, F1 score, and AUROC (Figure 12) of 0.88, 0.88, 0.88, 0.87, and 0.86 on the testing set. The TP, TN, FP, and FN values were 0.93, 0.80, 0.20, and 0.07, respectively (Figure 13).

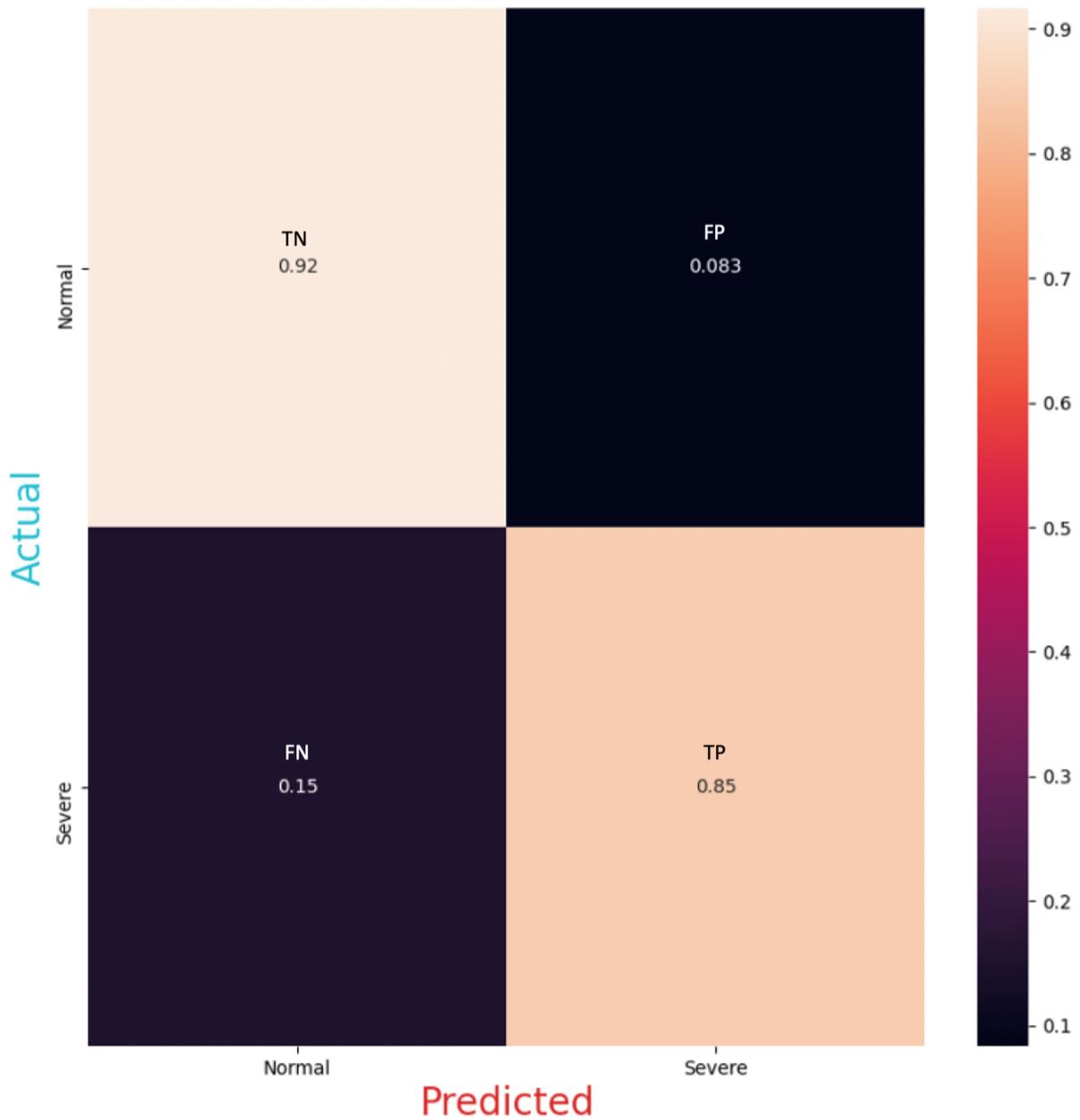


Figure 11. Confusion matrix comparing the predicted classification of rapid and non-rapid progressors from ground-truth in the validation set (N=61).

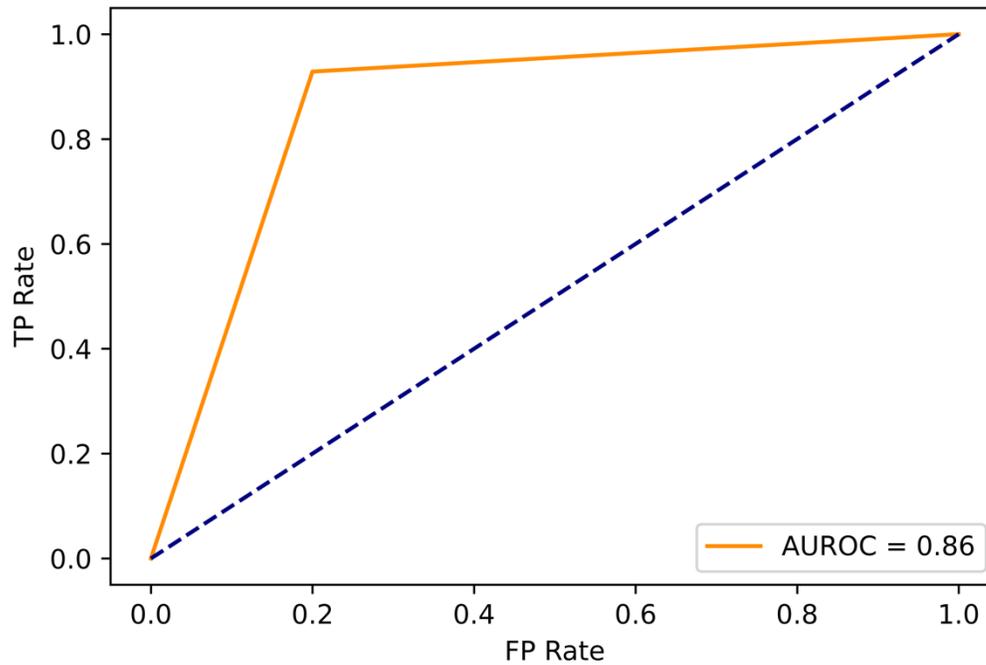


Figure 12. AUROC graph comparing the TP rate with the FP rate from the deep learning model in the test set.

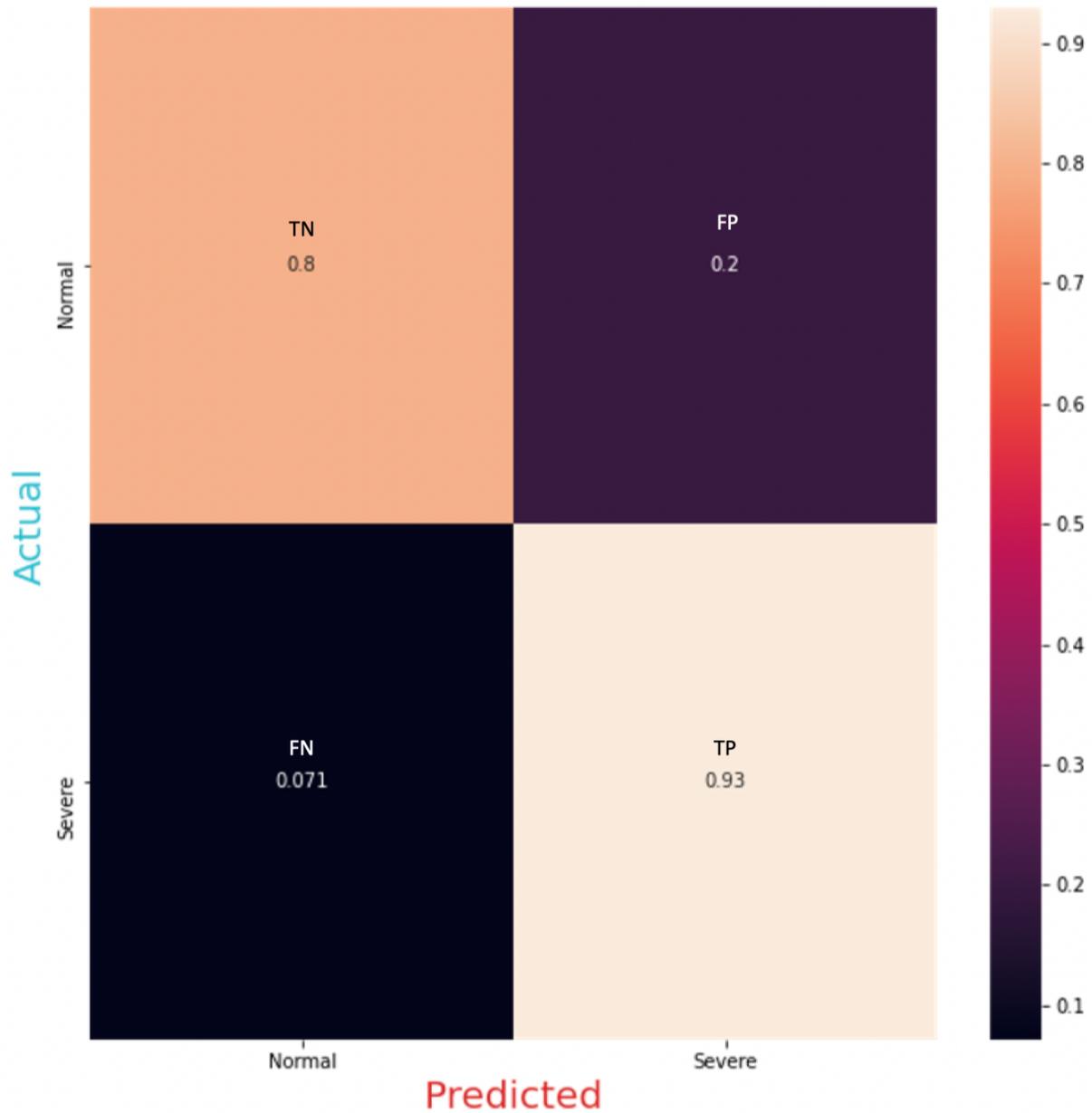


Figure 13. Confusion matrix comparing the predicted classification of rapid and non-rapid progressors from ground-truth in the test set (N=72).

6. DISCUSSION

Our study first compared the accuracy of the ellipsoid equation against the gold-standard manual segmentation to justify the need for a novel approach to estimate TKV. We then developed three

classification machine learning models. The first two models incorporated clinical variables with the traditional predictive biomarkers (htSegTKV, age, and height). The third model used both the baseline MRI and clinical variables. We examined their performances in classifying whether the patient will progress below or above the annual rate of eGFR decline of $-4 \text{ mL}/\text{min}/1.73\text{m}^2$ per year.

6.1. Baseline Characteristics

The study-eligible patients from our MUHC PKD clinic had mean baseline age of 44.8 ± 14.7 years and a median BMI of $26.6 \text{ kg}/\text{m}^2$ [23.6-30.3] (Table 4). Furthermore, the median SegTKV, total cyst volume, and eGFR were 1435.8 mL [819.4-2284.7], 882.9 mL [329.2-1503.0], and $76.4 \text{ mL}/\text{min}/1.73\text{m}^2$, respectively. In contrast, the CRISP study cohort had a mean baseline age of 33.8 ± 8.9 years and a mean BMI of $25.9 \pm 5.2 \text{ kg}/\text{m}^2$.^{31, 77} Their mean SegTKV, total cyst volume, and eGFR were $1076 \pm 670 \text{ mL}$, $534 \pm 529 \text{ mL}$, and $98.2 \pm 24.9 \text{ mL}$. The baseline characteristics of our study cohort were older than the CRISP cohort with larger kidney volumes and lower eGFR.

Recently, Yu et al. developed a long-term trajectory statistical model for each decade of age and MIC using the 14-year data from the CRISP dataset (Table 7).¹¹⁷ The model was further validated using the HALT-PKD dataset (N=558). Our cohort's annual eGFR decline closely matched the 40-50 years decade group in the long-term model.

Annual Rate of eGFR Decline (mL/min/1.73m² per year)

	20-30 yrs	30-40 yrs	40-50 yrs	50-60 yrs
Class 1A	0.77	-0.63	-2.03	-3.42
Class 1B	0.21	-1.19	-2.58	-3.98
Class 1C	-0.52	-1.92	-3.32	-4.71
Class 1D	-1.97	-3.37	-4.77	-6.16
Class 1E	-3.25	-4.65	-6.05	
Class 2	NA	NA	NA	NA

Table 7. Long-term trajectory statistical model for each decade of age and MIC using the 14-year data from the CRISP dataset.³

6.2. Manual Segmentation vs. Ellipsoid Equation

Our results were consistent with the findings from the comparative study conducted by Shi et al. and Demoulin et al.^{40, 43} Although the eTKV and SegTKV correlated well, the ellipsoid equation tended to underestimate the true TKV and large limits of agreement were observed. In both studies, 3.6% and 5.7% of the patients, respectively, were misclassified as either Class 1B or Class 1C. Similarly, 8 of 145 patients (5.5%) in our study cohort were misclassified as Class 1B. This is concerning because Class 1C is the threshold that determines a patient’s eligibility for tolvaptan treatment. If patients are misclassified as Class 1B, it could delay the initiation of tolvaptan therapy, and patients may experience more rapid growth in the cystic kidneys, rapid decline in eGFR, and earlier onset of kidney failure.^{3, 40, 66, 67} Conversely, if patients are misclassified as Class 1C and prescribed tolvaptan, they may experience a lower quality of life

from the unnecessary side effects and cost of tolvaptan. Therefore, minimizing the classification error will improve patient care and management of ADPKD.

Although the ellipsoid approach was efficient and generally reliable for calculating the TKV, our analysis of the MRI scans of the 25 patients with $\geq 20\%$ difference between the two measurement approaches highlighted the lack of precision and accuracy in a minority of patients with exophytic cysts. Figure 10 illustrates two patients with similar ages, htSegTKV, large exophytic cysts, and irregular kidney shapes. However, the percentage difference between the two measurements was 1.5% for one patient and more than 50% for the other. Moreover, despite the similarity in ground-truth MIC (Class 1C), age, and htSegTKV, they experienced different rates of annual eGFR decline. Patients from the MUHC PKD clinic with large discrepancies between the two measurement approaches generally possessed large exophytic cysts. Our finding is consistent with the study findings by O'Neill et al.¹¹⁸ Since the ellipsoid equation applies the orthogonal measurements from the mid-slice of the MRI scan, it fails to capture the shape irregularities secondary to cysts seen in other MRI slices. Although the ellipsoid equation is time-efficient and accurate, these limitations result in avoidable misclassifications in MIC and bring concerns to the accuracy of current clinical decision-making.

Resorting to manual segmentation for patients who exhibit large exophytic cysts or irregular kidney shapes may be the short-term solution to minimizing the number of misclassifications. However, further studies should examine the impact of exophytic cysts on one's kidney disease progression relative to cysts located within the kidney tissue. Understanding whether they should be included in the TKV calculation will aid in determining if an alternative approach should be explored for the small cohort exhibiting irregular cyst growth. These findings should be accounted for while exploring new prognostic tools. Applying machine learning is a practical

approach that could be explored to identify the associations between these factors less visible to the human eyes or conventional statistical models.

6.3. Machine Learning Models

6.3.1. Clinical Model

We first included only the clinical features to assess the predictive performance of the machine learning models. Since our dataset was high-dimensional relative to its sample size, it was most appropriate to explore the RF and SVM approaches. Before normalization, the RF model performed better than the SVM model. However, after normalization, the SVM model's performance surpassed that of the RF model. As well, the precision and accuracy became similar in value after normalization for the SVM model, indicating a more reliable predictive performance.

Before normalization, the continuous variables, such as htSegTKV, cyst volume, SBP, DBP, time on tolvaptan, and BMI, were expressed in a wide range of units and values. As expected, normalizing the input features to a common range and reducing numerical bias improved the SVM model's classification performance.¹¹⁹ However, the same was not observed because RF classifiers are tree-based algorithms that produce the output based on majority votes.⁸² In other words, the differences in values and scales across features do not significantly affect the RF model's performance regardless of feature normalization.

Although the RF model performed worse than the SVM, the feature importance obtained from the RF model still aids us in deducing the possible logic behind the model. Since the MUHC PKD clinic had an average baseline eGFR of $-67.3 \text{ mL/min/1.73m}^2$ [45.1-98.5], we could deduce that most patients exhibited more renal progression than typical early-stage patients. Although the RF approach does not provide information on the exact weight each feature has in

the classification task, the statistically significant difference in the baseline eGFR between rapid and non-rapid progressors may indicate that it helped the model to better classify rapid progressors.⁴² In contrast, the model may have performed worse in classifying those in the earlier stages with more steady eGFR, which could explain the lower evaluation metric scores.

Hypertension is a well-established independent risk factor for TKV increase in ADPKD and progression to ESKD, even in non-ADPKD patients.^{28, 120} If the patient is diagnosed with hypertension before the age of 35, the risk of early onset of ESKD increases.²⁸ There was also a statistically significant difference in the hypertension history between the two groups.

Therefore, we could speculate that understanding the patient's hypertension history may be helping the model classify patients. History of tolvaptan at any given time and time on tolvaptan followed as the third and fourth significant features, respectively. Since tolvaptan is recommended to high-risk patients determined by the MIC criteria, the model may have identified fast-progressing patients as those who have taken tolvaptan longer.¹ However, it is also important to note that not all patients at Class 1C or higher take tolvaptan. In our study cohort, 25 out of 82 patients at Class 1C or higher are not taking tolvaptan for various reasons. Therefore, it is important to assess how the model will classify patients from an external cohort where criteria for tolvaptan prescribing may be different. Interestingly, baseline htSegTKV, the important prognostic biomarker for ADPKD, was the fifth significant feature. Cyst volume was the final feature that contributed to the model's prediction. Many studies speculate the importance of cysts in determining the disease progression, but only a few studies have examined its impact.^{31, 121} This may be because cyst segmentations are time-consuming and laborious.⁷⁶ Although Lee et al. observed a correlation of $r=-0.60$ between cyst volume and creatine clearance, the sample size was only 56 patients.¹²¹ Furthermore, King et al. discovered a relationship between cyst volumes

and the slope of eGFR decline, but only nine patients were studied.¹²² The CRISP study discovered the direct correlation between the change in TKV and cyst volume and the relationship between TKV and eGFR decline. Although this implies a potential relationship between cyst volume and eGFR decline, it has not been explored directly. Therefore, the effect of cyst volume on disease progression should be explored further in other ADPKD cohorts.

6.3.2. Deep Model

Our deep learning model provided more promising results than the RF and SVM models which only included clinical features. The features extracted from the MRI may have provided the model with additional information and patterns concerning the irregular shapes of the cystic kidneys and the distributions of cysts. Furthermore, despite our small sample size in the training set, there is no evidence of overfitting due to the high accuracy, precision, recall, and F1 scores between 0.85 and 0.90 obtained in both validation and testing sets. According to the confusion matrix from the testing phase, the model learned to distinguish rapid progressors more accurately than non-rapid progressors despite having more non-rapid progressors (N=70) than rapid progressors (N=50) in the total sample size. Similar to the previous two models, we speculate that clinical features that are unique to rapid progressors, such as the history of tolvaptan at any given time and time on tolvaptan, may have positively contributed to the model accuracy in identifying rapid progressors more easily.

We cannot compare our model's performance with the prognostic model developed by Raj et al. because it combines the number of TP, TN, FP, and FN from the training, validation, and testing subsets altogether.¹⁰⁹ Moreover, our deep learning model cannot be compared with the MIC tool as it is a regression model.

Although our study contained a small number of patients, a common limitation in clinical deep-learning models, our model performed well without overfitting. This may be attributed to a few strategies employed during pre-processing and model training stages: 1) Simplifying the 3D images into the largest segmentation 2D slice and its two neighbouring slices for each patient resolved the issue of small sample size while preserving highly informative features relevant to the image textures. 2) The pixel-level and spatial-level transformations applied to the images diversified the images observed by the model during training and validation, which helped with generalizability. 3) The strategy of aggregating the model prediction for three slices chosen from the same patient's MRI scan improved the robustness of our model.

6.4. Strengths

Our study explores a novel field of incorporating machine learning in classifying disease progression in well-characterized, carefully followed patients with ADPKD. It is the first to develop a deep model that incorporates the MRI and many clinical features, not limited to the traditional variables used in the MIC tool, to predict the disease progression of ADPKD. We clearly illustrate that including MRI enhances the model's predictive performance in machine learning and the power of deep learning approaches.

6.5. Limitations

Our study contains several limitations. The main issue that emerges with all deep learning models is the Blackbox Effect.⁹¹ Although our model performed well, we cannot fully decipher which imaging features were extracted from the EfficientNet-b2 model and what associations were identified between and within the clinical and imaging features in finalizing prediction. To mitigate the issue of small sample size, we augmented our dataset and discretized the continuous output feature, the annual eGFR decline, into a binary class. Although this step

simplified and optimized the learning process of the model, important information can be lost, explanatory power may diminish, and non-linear relationships may be concealed.¹²³ For instance, the distances between individuals with similar renal function decline and those with distinct renal function decline will be viewed as the same by the model.^{124, 125} Finally, the cut-off points used in discretization are arbitrary.¹²³ The discrepancies that arise across various studies can challenge the ability to compare the performances of the machine learning models, as seen previously.

Although the output feature was well-balanced, some input features (e.g., history of hypertension, and family history) were imbalanced. However, this is the inevitable nature of medical data and will always bring bias in developing a machine-learning model.¹²⁶

As with any cohort study, there are also clinical limitations. The MUHC PKD clinic is a referral clinic and may only generalize to select ADPKD populations. For instance, our study cohort mainly consists of patients who are non-Black. Therefore, our model may be less generalizable toward Black patients. Moreover, since higher-risk patients are usually referred to a specialty clinic for therapy, this cohort will likely still identify those who are more eligible for tolvaptan therapy. Given that the follow-up time of patients was less than 10 years, future studies could train the model using well-characterized patients who have been followed at the clinic for at least 10 years. The availability and access to MRI tests may also create bias when the model is generalized. MRI is expensive, and some regions or communities may have limited access. A more generalizable model can be developed for other imaging modalities, such as CT and US.

7. FUTURE STEPS

Despite using an available and well-characterized cohort from the MUHC PKD clinic, a separate external validation cohort should be identified to further assess the performance of our deep learning model.

Given that TKV increases more consistently throughout the disease than eGFR, it would be interesting to compare the model's performance with our deep learning model using the annual rate of eGFR decline. Our study was not able to develop a separate deep-learning model that predicts the annual increase in htSegTKV because our PKD clinic lacked patients with at least two MRI images. Again, this may be feasible if other cohorts are identified, or longer follow-up is available from the MUHC PKD clinic cohort.

8. CONCLUSIONS

In conclusion, our deep model successfully classified the severity of disease progression in patients with ADPKD by integrating clinical information and MRI. We incorporated MRI image features that a conventional statistical model cannot account for by implementing a deep learning approach. Furthermore, we identified the inconsistency in the eTKV calculation from the ellipsoid equation approach and its impact on MIC misclassification. Our study justifies the need for a better prognostic tool and provides evidence that deep learning has the potential to predict disease progression better than the conventional MIC risk classification tool, which can help to individualize and improve patient care and the clinical management of ADPKD in the future.

9. REFERENCES

1. Torres VE, Higashihara E, Devuyst O, Chapman AB, Gansevoort RT, Grantham JJ, Perrone RD, Ouyang J, Blais JD, Czerwiec FS: Effect of tolvaptan in autosomal dominant polycystic kidney disease by CKD stage: results from the TEMPO 3: 4 trial. *Clinical Journal of the American Society of Nephrology*, 11: 803-811, 2016
2. Irazabal MV, Blais JD, Perrone RD, Gansevoort RT, Chapman AB, Devuyst O, Higashihara E, Harris PC, Zhou W, Ouyang J: Prognostic enrichment design in clinical trials for autosomal dominant polycystic kidney disease: The TEMPO 3: 4 clinical trial. *Kidney International Reports*, 1: 213-220, 2016
3. Chebib FT, Torres VE: Assessing Risk of Rapid Progression in Autosomal Dominant Polycystic Kidney Disease and Special Considerations for Disease-Modifying Therapy. *American Journal of Kidney Diseases*, 2021
4. Chebib FT, Torres VE: Autosomal Dominant Polycystic Kidney Disease: Core Curriculum 2016. *Am J Kidney Dis*, 67: 792-810, 2016 [10.1053/j.ajkd.2015.07.037](https://doi.org/10.1053/j.ajkd.2015.07.037)
5. Churchill DN, Bear JC, Morgan J, Payne RH, McManamon PJ, Gault MH: Prognosis of adult onset polycystic kidney disease re-evaluated. *Kidney international*, 26: 190-193, 1984
6. Finco DR: Chapter 17 - Kidney Function. In: *Clinical Biochemistry of Domestic Animals (Fifth Edition)*. edited by Kaneko JJ, Harvey JW, Bruss ML, San Diego, Academic Press, 1997, pp 441-484
7. Clearinghouse NKaUDI: Polycystic Kidney Disease. 2007
8. Preuss HG: Basics of Renal Anatomy and Physiology. *Clinics in Laboratory Medicine*, 13: 1-11, 1993 [https://doi.org/10.1016/S0272-2712\(18\)30456-6](https://doi.org/10.1016/S0272-2712(18)30456-6)

9. Torres VE, Harris PC, Pirson Y: Autosomal dominant polycystic kidney disease. *Lancet*, 369: 1287-1301, 2007 10.1016/s0140-6736(07)60601-1
10. Cornec-Le Gall E, Alam A, Perrone RD: Autosomal dominant polycystic kidney disease. *The Lancet*, 393: 919-935, 2019
11. Bae KT, Tao C, Zhu F, Bost JE, Chapman AB, Grantham JJ, Torres VE, Guay-Woodford LM, Meyers CM, Bennett WM: MRI-based kidney volume measurements in ADPKD: reliability and effect of gadolinium enhancement. *Clinical Journal of the American Society of Nephrology*, 4: 719-725, 2009
12. Ravine D, Sheffield L, Danks D, Gibson R, Walker R, Kincaid-Smith P: Evaluation of ultrasonographic diagnostic criteria for autosomal dominant polycystic kidney disease 1. *The Lancet*, 343: 824-827, 1994
13. Pei Y, Obaji J, Dupuis A, Paterson AD, Magistroni R, Dicks E, Parfrey P, Cramer B, Coto E, Torra R, San Millan JL, Gibson R, Breuning M, Peters D, Ravine D: Unified criteria for ultrasonographic diagnosis of ADPKD. *J Am Soc Nephrol*, 20: 205-212, 2009
10.1681/asn.2008050507
14. Emamian SA, Nielsen MB, Pedersen JF: Intraobserver and interobserver variations in sonographic measurements of kidney size in adult volunteers. A comparison of linear measurements and volumetric estimates. *Acta Radiol*, 36: 399-401, 1995
15. Sharma K, Rupprecht C, Caroli A, Aparicio MC, Remuzzi A, Baust M, Navab N: Automatic Segmentation of Kidneys using Deep Learning for Total Kidney Volume Quantification in Autosomal Dominant Polycystic Kidney Disease. *Sci Rep*, 7: 2049, 2017
10.1038/s41598-017-01779-0

16. Pei Y, Hwang YH, Conklin J, Sundsbak JL, Heyer CM, Chan W, Wang K, He N, Rattansingh A, Atri M, Harris PC, Haider MA: Imaging-based diagnosis of autosomal dominant polycystic kidney disease. *J Am Soc Nephrol*, 26: 746-753, 2015
10.1681/asn.2014030297
17. McIsaac HK, Thordarson DS, Shafran R, Rachman S, Poole G: Claustrophobia and the magnetic resonance imaging procedure. *Journal of Behavioral Medicine*, 21: 255-268, 1998
18. Akbari P, Nasri F, Deng S, Khowaja S, Lee S, Warnica W, Lu H, Rattansingh A, Atri M, Khalili K, Pei Y: Total Kidney Volume Measurements in ADPKD by 3D and Ellipsoid Ultrasound in Comparison to Magnetic Resonance Imaging. *Clin J Am Soc Nephrol*, 2022 10.2215/cjn.14931121
19. Kaewlai R, Abujudeh H: Nephrogenic systemic fibrosis. *American Journal of Roentgenology*, 199: W17-W23, 2012
20. Kline TL, Korfiatis P, Edwards ME, Warner JD, Irazabal MV, King BF, Torres VE, Erickson BJ: Automatic total kidney volume measurement on follow-up magnetic resonance images to facilitate monitoring of autosomal dominant polycystic kidney disease progression. *Nephrology Dialysis Transplantation*, 31: 241-248, 2016
21. Bae KT, Grantham JJ: Imaging for the prognosis of autosomal dominant polycystic kidney disease. *Nature Reviews Nephrology*, 6: 96-106, 2010 10.1038/nrneph.2009.214
22. de Amorim Paiva CC, de Mello Junior CF, Guimarães Filho HA, de Brito Gomes CA, Junior LRS, Junior GMB, Paiva CSM: Reproducibility of Renal Volume Measurement in Adults Using 3-Dimensional Sonography. *Journal of Ultrasound in Medicine*, 33: 431-435, 2014

23. Kim HC, Yang DM, Lee SH, Cho YD: Usefulness of renal volume measurements obtained by a 3-dimensional sonographic transducer with matrix electronic arrays. *Journal of Ultrasound in Medicine*, 27: 1673-1681, 2008
24. Iliuta I-A, Kalatharan V, Wang K, Cornec-Le Gall E, Conklin J, Pourafkari M, Ting R, Chen C, Borgo AC, He N: Polycystic kidney disease without an apparent family history. *Journal of the American Society of Nephrology*, 28: 2768-2776, 2017
25. Chapman AB, Devuyst O, Eckardt K-U, Gansevoort RT, Harris T, Horie S, Kasiske BL, Odland D, Pei Y, Perrone RD: Autosomal-dominant polycystic kidney disease (ADPKD): executive summary from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. *Kidney international*, 88: 17-27, 2015
26. Irazabal MV, Rangel LJ, Bergstralh EJ, Osborn SL, Harmon AJ, Sundsbak JL, Bae KT, Chapman AB, Grantham JJ, Mrug M: Imaging classification of autosomal dominant polycystic kidney disease: a simple model for selecting patients for clinical trials. *Journal of the American Society of Nephrology*, 26: 160-172, 2015
27. Bae KT, Shi T, Tao C, Yu ASL, Torres VE, Perrone RD, Chapman AB, Brosnahan G, Steinman TI, Braun WE, Srivastava A, Irazabal MV, Abebe KZ, Harris PC, Landsittel DP, Consortium HP: Expanded Imaging Classification of Autosomal Dominant Polycystic Kidney Disease. *Journal of the American Society of Nephrology : JASN*, 31: 1640-1651, 2020 10.1681/ASN.2019101121
28. Gabow PA, Johnson AM, Kaehny WD, Kimberling WJ, Lezotte DC, Duley IT, Jones RH: Factors affecting the progression of renal disease in autosomal-dominant polycystic kidney disease. *Kidney international*, 41: 1311-1319, 1992

29. Cornec-Le Gall E, Audrézet M-P, Rousseau A, Hourmant M, Renaudineau E, Charasse C, Morin M-P, Moal M-C, Dantal J, Wehbe B: The PROPKD score: a new algorithm to predict renal survival in autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology*, 27: 942-951, 2016
30. Bae KT, Zhou W, Shen C, Landsittel DP, Wu Z, Tao C, Chapman AB, Torres VE, Alan S, Mrug M: Growth pattern of kidney cyst number and volume in autosomal dominant polycystic kidney disease. *Clinical Journal of the American Society of Nephrology*, 14: 823-833, 2019
31. Grantham JJ, Torres VE, Chapman AB, Guay-Woodford LM, Bae KT, King BF, Jr., Wetzel LH, Baumgarten DA, Kenney PJ, Harris PC, Klahr S, Bennett WM, Hirschman GN, Meyers CM, Zhang X, Zhu F, Miller JP: Volume progression in polycystic kidney disease. *N Engl J Med*, 354: 2122-2130, 2006 10.1056/NEJMoa054341
32. Chebib FT, Perrone RD, Chapman AB, Dahl NK, Harris PC, Mrug M, Mustafa RA, Rastogi A, Watnick T, Yu ASL, Torres VE: A Practical Guide for Treatment of Rapidly Progressive ADPKD with Tolvaptan. *Journal of the American Society of Nephrology*, 29: 2458-2470, 2018 10.1681/asn.2018060590
33. Irazabal MV, Rangel LJ, Bergstralh EJ, Osborn SL, Harmon AJ, Sundsbak JL, Bae KT, Chapman AB, Grantham JJ, Mrug M, Hogan MC, El-Zoghby ZM, Harris PC, Erickson BJ, King BF, Torres VE, Investigators tC: Imaging Classification of Autosomal Dominant Polycystic Kidney Disease: A Simple Model for Selecting Patients for Clinical Trials. *Journal of the American Society of Nephrology*, 26: 160-172, 2015 10.1681/asn.2013101138

34. Grantham JJ, Chapman AB, Torres VE: Volume progression in autosomal dominant polycystic kidney disease: the major factor determining clinical outcomes. *Clin J Am Soc Nephrol*, 1: 148-157, 2006 10.2215/cjn.00330705
35. Yu ASL, Shen C, Landsittel DP, Harris PC, Torres VE, Mrug M, Bae KT, Grantham JJ, Rahbari-Oskoui FF, Flessner MF, Bennett WM, Chapman AB: Baseline total kidney volume and the rate of kidney growth are associated with chronic kidney disease progression in Autosomal Dominant Polycystic Kidney Disease. *Kidney International*, 93: 691-699, 2018 <https://doi.org/10.1016/j.kint.2017.09.027>
36. Tangri N, Hougen I, Alam A, Perrone R, McFarlane P, Pei Y: Total Kidney Volume as a Biomarker of Disease Progression in Autosomal Dominant Polycystic Kidney Disease. *Can J Kidney Health Dis*, 4: 2054358117693355, 2017 10.1177/2054358117693355
37. Magistroni R, Corsi C, Marti T, Torra R: A review of the imaging techniques for measuring kidney and cyst volume in establishing autosomal dominant polycystic kidney disease progression. *American journal of nephrology*, 48: 67-78, 2018
38. Higashihara E, Nutahara K, Okegawa T, Tanbo M, Hara H, Miyazaki I, Kobayasi K, Nitatori T: Kidney volume estimations with ellipsoid equations by magnetic resonance imaging in autosomal dominant polycystic kidney disease. *Nephron*, 129: 253-262, 2015
39. Available at: <https://www.mayo.edu/research/documents/pkd-center-adpkd-classification/doc-20094754>.
40. Shi B, Akbari P, Pourafkari M, Iliuta I-A, Guiard E, Quist CF, Song X, Hillier D, Khalili K, Pei Y: Prognostic Performance of Kidney Volume Measurement for Polycystic Kidney Disease: A Comparative Study of Ellipsoid vs. Manual Segmentation. *Scientific Reports*, 9: 10996, 2019 10.1038/s41598-019-47206-4

41. Kline TL, Korfiatis P, Edwards ME, Blais JD, Czerwiec FS, Harris PC, King BF, Torres VE, Erickson BJ: Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys. *Journal of digital imaging*, 30: 442-448, 2017
42. Grantham JJ, Torres VE: The importance of total kidney volume in evaluating progression of polycystic kidney disease. *Nature Reviews Nephrology*, 12: 667-677, 2016
10.1038/nrneph.2016.135
43. Demoulin N, Nicola V, Michoux N, Gillion V, Ho TA, Clerckx C, Pirson Y, Annet L: Limited Performance of Estimated Total Kidney Volume for Follow-up of ADPKD. *Kidney International Reports*, 6: 2821-2829, 2021
<https://doi.org/10.1016/j.ekir.2021.08.013>
44. Bae KT, Tao C, Wang J, Kaya D, Wu Z, Bae JT, Chapman AB, Torres VE, Grantham JJ, Mrug M: Novel approach to estimate kidney and cyst volumes using mid-slice magnetic resonance images in polycystic kidney disease. *American journal of nephrology*, 38: 333-341, 2013
45. Kline TL, Korfiatis P, Edwards ME, Bae KT, Yu A, Chapman AB, Mrug M, Grantham JJ, Landsittel D, Bennett WM: Image texture features predict renal function decline in patients with autosomal dominant polycystic kidney disease. *Kidney international*, 92: 1206-1216, 2017
46. Grantham JJ, Mulamalla S, Swenson-Fields KI: Why kidneys fail in autosomal dominant polycystic kidney disease. *Nat Rev Nephrol*, 7: 556-566, 2011 10.1038/nrneph.2011.109
47. Chapman AB: Hypertension in autosomal dominant polycystic kidney disease. *Kidney Int*, 61: S71-S73, 1997

48. Schrier RW: Renal volume, renin-angiotensin-aldosterone system, hypertension, and left ventricular hypertrophy in patients with autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology*, 20: 1888-1893, 2009
49. Soroka S, Alam A, Bevilacqua M, Girard L-P, Komenda P, Loertscher R, McFarlane P, Pandeya S, Tam P, Bichet DG: Updated Canadian expert consensus on assessing risk of disease progression and pharmacological management of autosomal dominant polycystic kidney disease. *Canadian journal of kidney health and disease*, 5: 2054358118801589, 2018
50. Kramers BJ, Koorevaar IW, Drenth JP, de Fijter JW, Neto AG, Peters DJ, Vart P, Wetzels JF, Zietse R, Gansevoort RT: Salt, but not protein intake, is associated with accelerated disease progression in autosomal dominant polycystic kidney disease. *Kidney International*, 98: 989-998, 2020
51. Torres VE, Abebe KZ, Schrier RW, Perrone RD, Chapman AB, Yu AS, Braun WE, Steinman TI, Brosnahan G, Hogan MC, Rahbari FF, Grantham JJ, Bae KT, Moore CG, Flessner MF: Dietary salt restriction is beneficial to the management of autosomal dominant polycystic kidney disease. *Kidney Int*, 91: 493-500, 2017
10.1016/j.kint.2016.10.018
52. McMahon EJ, Campbell KL, Bauer JD, Mudge DW, Kelly JT: Altered dietary salt intake for people with chronic kidney disease. *Cochrane Database of Systematic Reviews*, 2021
53. Orth SR, Stöckmann A, Conradt C, Ritz E, Ferro icwM, Kreusser W, Piccoli G, Rambausek M, Roccatello D, Schäfer K: Smoking as a risk factor for end-stage renal failure in men with primary renal disease. *Kidney international*, 54: 926-931, 1998

54. Ozkok A, Akpınar TS, Tufan F, Kanitez NA, Uysal M, Guzel M, Caliskan Y, Alisir S, Yazici H, Ecder T: Clinical characteristics and predictors of progression of chronic kidney disease in autosomal dominant polycystic kidney disease: a single center experience. *Clinical and experimental nephrology*, 17: 345-351, 2013
55. Bajwa ZH, Gupta S, Warfield CA, Steinman TI: Pain management in polycystic kidney disease. *Kidney international*, 60: 1631-1644, 2001
56. Torres VE, Bankir L, Grantham JJ: A case for water in the treatment of polycystic kidney disease. *Clinical Journal of the American Society of Nephrology*, 4: 1140-1150, 2009
57. Nagao S, Nishii K, Katsuyama M, Kurahashi H, Marunouchi T, Takahashi H, Wallace DP: Increased water intake decreases progression of polycystic kidney disease in the PCK rat. *Journal of the American Society of Nephrology*, 17: 2220-2227, 2006
58. Higashihara E, Nutahara K, Tanbo M, Hara H, Miyazaki I, Kobayashi K, Nitatori T: Does increased water intake prevent disease progression in autosomal dominant polycystic kidney disease? *Nephrol Dial Transplant*, 29: 1710-1719, 2014 10.1093/ndt/gfu093
59. van Gastel MD, Meijer E, Scheven LE, Struck J, Bakker SJ, Gansevoort RT: Modifiable factors associated with copeptin concentration: a general population cohort. *American Journal of Kidney Diseases*, 65: 719-727, 2015
60. Kalantar-Zadeh K, Fouque D: Nutritional management of chronic kidney disease. *New England Journal of Medicine*, 377: 1765-1776, 2017
61. Klahr S, Levey AS, Beck GJ, Caggiula AW, Hunsicker L, Kusek JW, Striker G: The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease. *New England Journal of Medicine*, 330: 877-884, 1994

62. Meijer E, Gansevoort RT: Emerging non-pharmacological interventions in ADPKD: an update on dietary advices for clinical practice. *Current Opinion in Nephrology and Hypertension*, 30: 482-492, 2021
63. Schrier RW, Abebe KZ, Perrone RD, Torres VE, Braun WE, Steinman TI, Winklhofer FT, Brosnahan G, Czarnecki PG, Hogan MC: Blood pressure in early autosomal dominant polycystic kidney disease. *New England Journal of Medicine*, 371: 2255-2266, 2014
64. Chebib FT, Torres VE: Assessing Risk of Rapid Progression in Autosomal Dominant Polycystic Kidney Disease and Special Considerations for Disease-Modifying Therapy. *Am J Kidney Dis*, 78: 282-292, 2021 10.1053/j.ajkd.2020.12.020
65. Chow CL, Ong AC: Autosomal dominant polycystic kidney disease. *Clinical medicine*, 9: 278, 2009
66. Torres VE, Chapman AB, Devuyst O, Gansevoort RT, Grantham JJ, Higashihara E, Perrone RD, Krasa HB, Ouyang J, Czerwiec FS: Tolvaptan in patients with autosomal dominant polycystic kidney disease. *New England Journal of Medicine*, 367: 2407-2418, 2012
67. Torres VE, Chapman AB, Devuyst O, Gansevoort RT, Perrone RD, Koch G, Ouyang J, McQuade RD, Blais JD, Czerwiec FS: Tolvaptan in later-stage autosomal dominant polycystic kidney disease. *New England Journal of Medicine*, 377: 1930-1942, 2017
68. Gordon CE, Perrone RD: Tolvaptan or transplant: why wait? *Kidney International*, 98: 286-289, 2020
69. Berl T, Quittnat-Pelletier F, Verbalis JG, Schrier RW, Bichet DG, Ouyang J, Czerwiec FS, Investigators S: Oral tolvaptan is safe and effective in chronic hyponatremia. *Journal of the American Society of Nephrology*, 21: 705-712, 2010

70. Erickson KF, Chertow GM, Goldhaber-Fiebert JD: Cost-effectiveness of tolvaptan in autosomal dominant polycystic kidney disease. *Annals of internal medicine*, 159: 382-389, 2013
71. Liu Y-M, Shao Y, He Q: Sirolimus for treatment of autosomal-dominant polycystic kidney disease: a meta-analysis of randomized controlled trials. *Transplantation proceedings*. Elsevier, 2014 pp 66-74
72. Walz G, Budde K, Mannaa M, Nürnberger J, Wanner C, Sommerer C, Kunzendorf U, Banas B, Hörl WH, Obermüller N: Everolimus in patients with autosomal dominant polycystic kidney disease. *New England Journal of Medicine*, 363: 830-840, 2010
73. Caroli A, Perico N, Perna A, Antiga L, Brambilla P, Pisani A, Visciano B, Imbriaco M, Messa P, Cerutti R: Effect of longacting somatostatin analogue on kidney and cyst growth in autosomal dominant polycystic kidney disease (ALADIN): a randomised, placebo-controlled, multicentre trial. *The Lancet*, 382: 1485-1495, 2013
74. Hogan MC, Masyuk TV, Page LJ, Kubly VJ, Bergstralh EJ, Li X, Kim B, King BF, Glockner J, Holmes DR: Randomized clinical trial of long-acting somatostatin for autosomal dominant polycystic kidney and liver disease. *Journal of the American Society of Nephrology*, 21: 1052-1061, 2010
75. Bevilacqua V, Brunetti A, Cascarano GD, Palmieri F, Guerriero A, Moschetta M: A deep learning approach for the automatic detection and segmentation in autosomal dominant polycystic kidney disease based on magnetic resonance images. *International Conference on Intelligent Computing*. Springer, 2018 pp 643-649
76. Kline TL, Edwards ME, Fetzer J, Gregory AV, Anaam D, Metzger AJ, Erickson BJ: Automatic semantic segmentation of kidney cysts in MR images of patients affected by

- autosomal-dominant polycystic kidney disease. *Abdom Radiol (NY)*, 46: 1053-1061, 2021
10.1007/s00261-020-02748-4
77. Chapman AB, Guay-Woodford LM, Grantham JJ, Torres VE, Bae KT, Baumgarten DA, Kenney PJ, King Jr BF, Glockner JF, Wetzel LH: Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): The Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) cohort. *Kidney international*, 64: 1035-1045, 2003
78. McEwan P, Bennett Wilton H, Ong ACM, Ørskov B, Sandford R, Scolari F, Cabrera M-CV, Walz G, O'Reilly K, Robinson P: A model to predict disease progression in patients with autosomal dominant polycystic kidney disease (ADPKD): the ADPKD Outcomes Model. *BMC Nephrology*, 19: 37, 2018 10.1186/s12882-017-0804-2
79. Rajkomar A, Dean J, Kohane I: Machine Learning in Medicine. *New England Journal of Medicine*, 380: 1347-1358, 2019 10.1056/NEJMra1814259
80. Alpaydin E: *Introduction to machine learning*, MIT press, 2020
81. Smola A, Vishwanathan S: *Introduction to machine learning*. Cambridge University, UK, 32: 2008, 2008
82. Maleki F, Ovens K, Najafian K, Forghani B, Reinhold C, Forghani R: Overview of Machine Learning Part 1: Fundamentals and Classic Approaches. *Neuroimaging Clin N Am*, 30: e17-e32, 2020 10.1016/j.nic.2020.08.007
83. Lison P: An introduction to machine learning. *Language Technology Group (LTG)*, 1: 1-35, 2015
84. Learned-Miller EG: *Introduction to supervised learning. I: Department of Computer Science, University of Massachusetts: 3*, 2014

85. Song YY, Lu Y: Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*, 27: 130-135, 2015 10.11919/j.issn.1002-0829.215044
86. Strobl C, Malley J, Tutz G: An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, 14: 323-348, 2009 10.1037/a0016973
87. Qi Y: Random forest for bioinformatics. In: *Ensemble machine learning: Methods and applications*. Springer, 2012, pp 307-323
88. Ghahramani Z: Unsupervised learning. *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany, August 4-16, 2003, Revised Lectures: 72-112*, 2004
89. Mahesh B: Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)[Internet]*, 9: 381-386, 2020
90. Alzubi J, Nayyar A, Kumar A: Machine learning from theory to algorithms: an overview. *Journal of physics: conference series*. IOP Publishing, 2018 pp 012012
91. Azodi CB, Tang J, Shiu S-H: Opening the Black Box: Interpretable Machine Learning for Geneticists. *Trends in Genetics*, 36: 442-455, 2020
<https://doi.org/10.1016/j.tig.2020.03.005>
92. Marcot BG, Hanea AM: What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Computational Statistics*, 36: 2009-2031, 2021
10.1007/s00180-020-00999-9
93. Yadav S, Shukla S: Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. *2016 IEEE 6th International conference on advanced computing (IACC)*. IEEE, 2016 pp 78-83

94. Lanera C, Berchiolla P, Sharma A, Minto C, Gregori D, Baldi I: Screening PubMed abstracts: is class imbalance always a challenge to machine learning? *Systematic reviews*, 8: 1-9, 2019
95. Casal MA, Ivy SP, Beumer JH, Nolin TD: Effect of removing race from glomerular filtration rate-estimating equations on anticancer drug dosing and eligibility: a retrospective analysis of National Cancer Institute phase 1 clinical trial participants. *The Lancet Oncology*, 22: 1333-1340, 2021
96. Courtland R: The bias detectives. *Nature*, 558: 357-360, 2018
97. Rahman MM, Davis DN: Addressing the class imbalance problem in medical datasets. *International Journal of Machine Learning and Computing*, 3: 224, 2013
98. Ali PJM, Faraj RH, Koya E, Ali PJM, Faraj RH: Data normalization and standardization: a technical report. *Mach Learn Tech Rep*, 1: 1-6, 2014
99. Hsu C-W, Chang C-C, Lin C-J: A practical guide to support vector classification. Taipei, Taiwan, 2003
100. Singh D, Singh B: Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97: 105524, 2020
<https://doi.org/10.1016/j.asoc.2019.105524>
101. Jo J-M: Effectiveness of normalization pre-processing of big data to the machine learning performance. *The Journal of the Korea institute of electronic communication sciences*, 14: 547-552, 2019
102. Dinga R, Penninx BW, Veltman DJ, Schmaal L, Marquand AF: Beyond accuracy: measures for assessing machine learning models, pitfalls and guidelines. *BioRxiv*: 743138, 2019

103. Gupta A, Tatbul N, Marcus R, Zhou S, Lee I, Gottschlich J: Class-Weighted evaluation metrics for imbalanced data classification. *arXiv preprint arXiv:201005995*, 2020
104. Novaković JD, Veljović A, Ilić SS, Papić Ž, Tomović M: Evaluation of classification models in machine learning. *Theory and Applications of Mathematics & Computer Science*, 7: 39, 2017
105. Hand D, Christen P: A note on using the F-measure for evaluating record linkage algorithms. *Statistics and Computing*, 28: 539-547, 2018 [10.1007/s11222-017-9746-6](https://doi.org/10.1007/s11222-017-9746-6)
106. Rafael-Palou X, Turino C, Steblin A, Sánchez-de-la-Torre M, Barbé F, Vargiu E: Comparative analysis of predictive methods for early assessment of compliance with continuous positive airway pressure therapy. *BMC medical informatics and decision making*, 18: 1-14, 2018
107. Fawcett T: An introduction to ROC analysis. *Pattern Recognition Letters*, 27: 861-874, 2006 <https://doi.org/10.1016/j.patrec.2005.10.010>
108. Halligan S, Altman DG, Mallett S: Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: a discussion and proposal for an alternative approach. *Eur Radiol*, 25: 932-939, 2015 [10.1007/s00330-014-3487-0](https://doi.org/10.1007/s00330-014-3487-0)
109. Raj A, Tollens F, Caroli A, Nörenberg D, Zöllner FG: Automated Prognosis of Renal Function Decline in ADPKD Patients using Deep Learning. *medRxiv*: 2023.2001.2013.23284471, 2023
110. Fernández A, García S, Galar M, Prati RC, Krawczyk B, Herrera F: *Learning from imbalanced data sets*, Springer, 2018
111. Debal DA, Sitote TM: Chronic kidney disease prediction using machine learning techniques. *Journal of Big Data*, 9: 1-19, 2022

112. Ilyas H, Ali S, Ponum M, Hasan O, Mahmood MT, Iftikhar M, Malik MH: Chronic kidney disease diagnosis using decision tree algorithms. *BMC nephrology*, 22: 1-11, 2021
113. Rady E-HA, Anwar AS: Prediction of kidney disease stages using data mining algorithms. *Informatics in Medicine Unlocked*, 15: 100178, 2019
114. Levey AS, Stevens LA: Estimating GFR using the CKD epidemiology collaboration (CKD-EPI) creatinine equation: more accurate GFR estimates, lower CKD prevalence estimates, and better risk predictions. *American Journal of Kidney Diseases*, 55: 622-627, 2010
115. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA: Alumentations: fast and flexible image augmentations. *Information*, 11: 125, 2020
116. Tan M, Le Q: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Kamalika C, Ruslan S (Eds.) *Proceedings of the 36th International Conference on Machine Learning*. Proceedings of Machine Learning Research, PMLR, 2019 pp 6105--6114
117. Yu ASL, Shen C, Landsittel DP, Grantham JJ, Cook LT, Torres VE, Chapman AB, Bae KT, Mrug M, Harris PC, Rahbari-Oskoui FF, Shi T, Bennett WM: Long-term trajectory of kidney function in autosomal-dominant polycystic kidney disease. *Kidney International*, 95: 1253-1261, 2019 <https://doi.org/10.1016/j.kint.2018.12.023>
118. O'Neill WC, Robbin ML, Bae KT, Grantham JJ, Chapman AB, Guay-Woodford LM, Torres VE, King BF, Wetzel LH, Thompson PA, Miller JP: Sonographic assessment of the severity and progression of autosomal dominant polycystic kidney disease: the Consortium of Renal Imaging Studies in Polycystic Kidney Disease (CRISP). *Am J Kidney Dis*, 46: 1058-1064, 2005 [10.1053/j.ajkd.2005.08.026](https://doi.org/10.1053/j.ajkd.2005.08.026)

119. Ben-Hur A, Ong CS, Sonnenburg S, Schölkopf B, Rätsch G: Support vector machines and kernels for computational biology. *PLoS computational biology*, 4: e1000173, 2008
120. Iglesias CG, Torres VE, Offord KP, Holley KE, Beard CM, Kurland LT: Epidemiology of adult polycystic kidney disease, Olmsted County, Minnesota: 1935–1980. *American Journal of Kidney Diseases*, 2: 630-639, 1983
121. Lee YR, Lee K-B: Reliability of magnetic resonance imaging for measuring the volumetric indices in autosomal-dominant polycystic kidney disease: correlation with hypertension and renal function. *Nephron Clinical Practice*, 103: c173-c180, 2006
122. King BF, Reed JE, Bergstralh EJ, Sheedy PF, Torres VE: Quantification and longitudinal trends of kidney, renal cyst, and renal parenchyma volumes in autosomal dominant polycystic kidney disease. *Journal of the American Society of Nephrology*, 11: 1505-1511, 2000
123. Royston P, Altman DG, Sauerbrei W: Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in medicine*, 25: 127-141, 2006
124. Altman DG, Royston P: The cost of dichotomising continuous variables. *Bmj*, 332: 1080, 2006
125. Sauerbrei W, Royston P: Continuous variables: to categorize or to model. *The 8th International Conference on Teaching Statistics–Data and Context in statistics education: Towards an evidence based society International Statistical Institute, Voorburg*. 2010
126. Li D-C, Liu C-W, Hu SC: A learning method for the class imbalance problem with medical data sets. *Computers in Biology and Medicine*, 40: 509-518, 2010
<https://doi.org/10.1016/j.combiomed.2010.03.005>