PREDICTING KIDNEY TUMOR SUBTYPE FROM CT IMAGES USING RADIOMICS AND CLINICAL FEATURES

by

Duygu ŞİRİN

B.S., in Industrial Engineering, Galatasaray University, 2019

Submitted to the Institute of Biomedical Engineering
in partial fulfillment of the requirements
for the degree of
Master of Science
in
Biomedical Engineering

Boğaziçi University
2022

ACKNOWLEDGMENTS

I would like to thank my advisor, my family and friends for all their help and support throughout the preparation of this thesis work.

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, Duygu Şirin, hereby certify that I am aware of the Academic Ethics and Integrity Policy issued by the Council of Higher Education (YÖK) and I fully acknowledge all the consequences due to its violation by plagiarism or any other way.

Name :

Signature:

Date:

ABSTRACT

PREDICTING KIDNEY TUMOR SUBTYPE FROM CT IMAGES USING RADIOMICS AND CLINICAL FEATURES

This study aims to evaluate the performance of machine learning methods in predicting the subtype (clear-cell vs. non-clear-cell) of kidney tumors using clinical patient and radiomics data from CT images. CT images of 192 malignant kidney tumor cases (142 clear-cell, 50 other) from TCIA's KiTS-19 Challenge were used in the study. There were several different tumor subtypes in the other group, most of them being chromophobe or papillary RCC. Patient clinical data were combined with the radiomic features extracted from CT images. Features were extracted from 3D images and all of the slices were included in the feature extraction process. Initial dataset consisted of 1157 features of which 1130 were radiomics and 27 were clinical. Features were selected using Kruskal Wallis - ANOVA test followed by Lasso Regression. After feature selection, 8 radiomic features remained. None of the clinical features were considered important for our model as a result. Training set classes were balanced using SMOTE. Training data with the selected features were used to train the Coarse Gaussian SVM and Subspace Discriminant classifiers. Coarse Gaussian SVM was faster compared to Subspace Discriminant with a training time of 0.47 sec and 11000 obs/sec prediction speed. Training duration of Subspace Discriminant was 4.1 sec with 960 obs/sec prediction speed. For Coarse Gaussian SVM was found as 0.86 while for Subspace Discriminant AUC was 0.85. Both models produced promising results on classifying malignant tumors as ccRCC or non-ccRCC.

Keywords: Kidney Tumor, Clear-Cell, Machine Learning, CT Imaging.

ÖZET

RADYOMİK VE KLİNİK ÖZELLİKLER KULLANILARAK BT GÖRÜNTÜLERİNDEN BÖBREK TÜMÖRÜ TİPİNİN BELİRLENMESİ

Bu çalışma, radyomik ve klinik veriler kullanılarak BT görüntülerinden böbrek tümörü tipinin (berrak hücreli veya berrak hücreli olmayan) tespit edilmesinde makine öğrenimi algoritmalarının performanslarını karşılaştırmayı hedeflemektedir. Çalışmada TCIA'nın KiTS-19 Yarışması'nda bulunan 192 malign böbrek tümörü BT görüntüsü (142 berrak hücreli, 50 diğer tip) kullanılmıştır. Özellikler 3 boyutlu görüntü kesitlerinin tümü kullanılarak çıkarılmıştır. Başlangıçta çıkarılan özellik sayısı, 1130 radyomik ve 27 klinik olmak üzere toplam 1157'dir. Kruskal Wallis - ANOVA testi ardından Lasso Regresyon metodu uygulanarak modellerde kullanılacak 8 özellik seçilmiştir. Bu aşamada tüm klinik özellikler elenmiştir. Eğitim kümesi sınıfları SMOTE metodu kullanılarak dengelenmiştir. Seçilen özellikler için eğitim kümesi verileri kullanılarak Coarse Gaussian SVM ve Subspace Discriminant algoritmaları eğitilmiştir. Coarse Gaussian SVM için eğitim süresi 0.47 sn ve tahmin hızı yaklaşık 11000 göz/sn iken; Subspace Discriminant için bu değerler sırasıyla 4.1 sn ve yaklaşık 960 göz/sn olarak gözlemlenmiştir. Coarse Gaussian SVM için ROC puanı 0.86 bulunurken, Subspace Discriminant için 0.85 olarak hesaplanmıştır. İki model de çalışmanın amacı olan böbrek tümörü tipi sınıflandırması için umut verici sonuçlar sağlamıştır.

Anahtar Sözcükler: Böbrek Tümörü, Berrak Hücre, Makine Öğrenimi, BT Görüntüleme.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AICP	Artificial Intelligence Care Provider
ANOVA	Analysis Of Variance
AUC	Area Under The Curve
C4KC	Climb 4 Kidney Cancer
CCRCC	Clear Cell Renal Cell Carcinoma
CHRCC	Chromophobe Renal Cell Carcinoma
CT	Computed Tomography
GLCM	Gray Level Co-Occurrence Matrix
GLDM	Gray Level Dependence Matrix
GLRLM	Gray Level Run-Length Matrix
GLSZM	Gray Level Size Zone Matrix
IDN	Inverse Difference Normalized
KiTS	Kidney And Kidney Tumor Segmentation Challenge
KW	Kruskal Wallis
LASSO	Least Absolute Shrinkage And Selection Operator
LoG	Laplacian Of Gaussian
MRI	Magnetic Resonance Imaging
NGTDM	Neighboring Gray Tone Difference Matrix
NLP	Natural Language Processing
PET	Positron Emission Tomography
PRCC	Papillary Renal Cell Carcinoma
RCC	Renal Cell Carcinoma
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
SMOTE	Synthetic Minority Oversampling Technique
SPECT	Single Photon Emission Computerized Tomography
SVM	Support Vector Machine

TCIA	The Cancer Imaging Archive
US	Ultrasound
VOI	Volume Of Interest
VR	Virtual Reality

1. INTRODUCTION

More than 400 000 patients are diagnosed with kidney cancer each year, more than 90% of them being renal cell carcinoma (RCC). RCC is known to be the most common type (approximately 75%) of kidney cancer as well having the highest mortality rate among genitourinary cancers. It is also one of the 10 most common cancers. RCC has more than 10 histological and molecular subtypes. The progression of its molecular and histopathological description has changed the way it is classified. Clear cell RCC (CCRCC), papillary RCC (PRCC) and chromophobe RCC (CHRCC) have the highest incidences among all subtypes [1].

In cases where RCC is localized; active surveillance, nephrectomy or ablation are the treatment options. If the cancer metastasizes (nearly 30% of localized RCC cases), systemic therapy alternatives are needed.

Over the past 30 years, mortality rates of RCC have been dropping globally. On the contrary, incidence rates continue to rise by dint of changing healthcare and peoples' way of living. For instance; abdominal imaging in hospitals has become more common, which increases the chance of uncovering small renal masses. In addition, the rising number of obesity cases all over the world may have an impact on the incidence of RCC.

Three most important RCC risk factors were determined as tobacco use, hypertension and being obese/overweight. The following are also correlated with RCC: hemodialysis, diabetes mellitus, chronic kidney disease and kidney transplantation. In addition to these, genetic factors partake in the risk of RCC development. For example; having a family member which has or had RCC makes an individual two times more likely to acquire it as well [1].

Nowell claimed that the majority of neoplasms originate from a single cell. The

genetic diversity of this origin cell allows the tumor to progress. Therefore, due to differences of tumor evolution from patient to patient, as well as within the same case, specific treatments should be implemented [2]. In one of the studies which was conducted in order to understand tumor genetic diversity in RCC; it was discovered that VHL mutation and 3p loss of heterozygosity were present across all regions - in each tumor- for all four patients. On the other hand, mutations such as BRM1, KDM5C, PTEN and SETD2 resided in the primary tumor as well as some of the places to which it had spread.

In a study which included 19 different cancer types, CCRCC was found to have the highest T cell infiltration median [3]. According to this paper, use of immune checkpoint inhibitors might be effective for the treatment of CCRCC. This is due to the strong antigen recognition and presentation, besides the robust immune filtration which is already present.

Currently, a great effort is being put into the studies concerning how different kidney tumor morphology might affect the treatment process. Surgery, chemotherapy and targeted drugs are used for treatment and a variety of new, more effective drugs are continued to be developed. There has been a big improvement in the median survival of the disease in the past few years, thanks to the targeted drug development [4].

In the past, patients with RCC would seek medical help because of hematuria, a conspicuous abdominal mass and flank pain. Today, most patients are diagnosed thanks to the imaging results carried out in pursuance of other medical findings. These results may raise suspicion even though radiographic presentations of different RCCs are changeful. Generally, CCRCC presents features such as high contrast enhancement agent uptake, heterogeneity and outward growth [1].

During the process of diagnosis of renal cancer; laboratory tests and biopsy are used in addition to radiology. Presently, biopsy is obligatory in order to deduce key information as whether the cancer is invasive, its grade, its stage and spread to lymph nodes. In addition to these, biopsy must be performed to identify specific

proteins, genes and other factors which are unique to the tumor. These factors play an important role in prognosis prediction and in the construction of a treatment plan. They can also provide a clearer view on how to design more effective drugs targeting specific intracellular pathways [5].

Biopsy is a highly invasive diagnosis tool which carries a small risk of infection and bleeding. Moreover, for cancer cases, reaching a result might take several days. For these reasons, there is a need for obtaining the necessary information for diagnosis by using better tools.

RCC is staged considering several factors: presence of metastasis, spread to lymph nodes, tumor volume outside the kidney, size of the tumor. Computed tomography is an indispensable tool in order to reach a healthy staging decision. Moreover, a variety of predictors for prognosis evaluation can be analyzed via laboratory testing. In general, the Fuhrman grade is preferred for RCC, which has a significant meaning for prognosis.

In the present, RCCs are managed by a range of treatment options which are suitable for varying cases. When the tumor is operable, nephrectomy is performed. In cases where surgery cannot be considered practicable, RCC is commonly handled by systemic treatment. Further, ablative therapies or active surveillance are also possible choices especially for certain patient types. As regards the non-CCRCCs, disease management cannot be based on sufficient data, as very few patients were included in drug trials in the past. In addition, there are many non-CCRCC subtypes with different characteristics, which implies that each one might respond in a different manner to the same treatment. Largely, the need for efforts to research and develop effective targeted drugs for non-CCRCCs still remains.

Some of the treatments which might be important options in the future are; targeted radiotherapy, personalized vaccines and selective cytoreductive nephrectomy. Moreover, combinations of various therapies might take place; such as adding immunotherapy or targeted therapy to the existing processes. At present, it is known

that treating RCC with a collection of distinct targeted drugs is more effective than using only one drug. However, this comes with the disadvantage of higher toxicity development in the patient's body. If drug resistance is to be avoided, optimal targeted drug combinations must be established. Immune checkpoint inhibitors may be the primary and the most suitable answer for reaching this goal [1].

Imaging is essential in the process of diagnosis and treatment of RCCs. A variety of medical imaging modalities play different roles when it comes to detection of renal masses, as well as deciding on the course of action. First of all, it is important to discriminate between benign and malignant tumors. Oncocytoma and angiomyolipoma are benign renal tumors, with oncocytoma showing a high level of similarity to RCC on medical images. Secondly, discrimination among RCC histologic subtypes becomes a critical target. Different imaging techniques have variable performances for reaching these two goals.

Ultrasound is one the modalities used for renal imaging. It does not require nephrotoxic contrast agents and there is no radiation involved in the process. Moreover, it is an economical option and easily reachable. When the patient's tumor is large enough, its detection possibility with the US is high. On the other hand, small tumors are comfortably missed on US images. Differentiation between malignant and benign cases is undependable; as it is when discriminating between subtypes. Some innovative US techniques are considered relatively better than gray-scale US for these objectives.

Computed Tomography (CT) is widely used for clinical diagnosis, localization of pathology, observation of anatomical structure, surveillance of therapy evolution and planning the optimal treatment in cancer. CT is generally the first choice for imaging the evolution of renal tumor because it is more available than Magnetic Resonance Imaging (MRI) and is more useful than Ultrasound (US) Imaging [6]. Moreover, CT imaging provides shorter acquisition time than MRI. The most prominent disadvantages are exposure to radiation and nephrotoxic contrast agents.

Scanning protocol has a determinative impact for detecting tumors, as well as

discriminating between malignant, benign and RCCs of different subtypes. Larger tumors are more easily detected. There are studies with high success which use CT imaging for both goals. It can be profitably used to determine whether a tumor is RCC or angiomyolipoma. However, it cannot be used by itself for detecting a papillary RCC. Furthermore, it is not plausible to discriminate between oncocytoma and chromophobe RCC due to their resembling appearances. Also, there are a few novel techniques which ameliorate the efficacy of discrimination. Conclusively, when RCC is suspected, biopsy or surgery should be performed for additional diagnostic information.

In general, MRI imaging is preferred when a renal tumor cannot be characterized with CT or US. In addition, it can be needed for patients who have other renal disorders. Despite its acquisition taking long, it is superior to CT in terms of not having to expose the patient to radiation. MRI capability is not as common at clinics as CT imaging, which can be considered as another convenience of CT. Regarding the identification of tumors MRI and CT perform similarly for the ones which require surgery.

There is no solid evidence showing apparent diffusion coefficient values can be used for distinguishing between ccRCC and other RCC subtypes. On the other hand, MRI might be an effective tool to differentiate ccRCC and papillary. As for chromophobe RCC, it is hard to recognize because of its image characteristics being similar to other RCC subtypes. Some of the angiomyolipoma can be told from RCCs, even though biopsy is encouraged for final diagnosis. Oncocytoma is very hardly discriminated from RCC due to their resembling features, with chromophobe RCC being the most alike among all RCCs. Some of these challenges may be overcome to improve renal tumor distinction using new MRI imaging techniques as several studies have shown.

Nuclear imaging might also be performed in the course of diagnosis and treatment surveillance of renal tumors. PET/CT is not mainly suggested for the first diagnosis of renal tumors considering its poor performance of identification. On the other hand, SPECT/CT and PET/CT can be used when differentiating RCCs and oncocytomas, as well as detecting ccRCC by the aid of correct tracer choices.

Overall, medical imaging has encouraged advancements which may bring the world closer to distinguish between malignant and benign cases, and between different RCC subtypes. However, currently, depending solely on imaging for diagnosis in cases where RCC is suspected would not be a possible medical decision-making process.

Data collection and analysis gains more importance day by day. It is accepted that data can be a powerful tool to show the causal relationships between incidents, help detect errors, provide guidance for decision-making, predict the outcome of actions, and partition instances into desired classes. As a result, many disciplines began to increase their efforts to collect and manipulate data for these purposes. Medicine is one of the fields that is being influenced by this trend.

Obtaining standardized and clean digital data is highly crucial for the sake of sustainability and reliability. Therefore, data collectors should be attentive and well-informed. Dedicated control mechanisms and data health checks might be implemented in order to avoid errors as much as possible. In medicine, clinical processes are often frail to these sorts of errors due to its nature. However, gathering and storage of digital data, and implementing these to usual practice routines, presents several advantages. A variety of medical data are collected such as test findings, clinical observations, diagnoses, current health issues, medications as well as data obtained via different medical devices which usually have integrated software systems. Real-time data acquisition is especially valuable in medicine and it should be supported by the authorities to be made apprehensible. Digital health data are crucial when it comes to personalized medicine, as well as detecting diseases at early stages and personalized preventative medicine. Furthermore, recent medical findings and research results should be reachable freely by the actors in medicine and science fields, so that relevant data can be used to enable further advancements and additional data.

As the volume of data increases, the need for smart and automated tools becomes more crucial in order to exploit it in the most efficient way possible. Today, many artificial intelligence applications are used for this purpose in a variety of work areas. By most, the birth of robots is recognized as the root of artificial intelligence. It was

first outlined as “the science and engineering of making intelligent machines”. It is established by several disciplines coming together such as psychology, philosophy and computer science.

With the progress of artificial intelligence, computers have become able to solve complicated mathematical problems which have multiple branches. Moreover, it started to be used in medical diagnostic processes as well as data mining. The advancements resulted in the emergence of instruments which have increasing computation capacities. Currently, artificial intelligence moves rapidly by enforcing new concepts and discovering innovative solutions to complex problems using engineering methods [7].

Cybernetics is a research area which focuses on conceptual organization roles of complicated systems. It aims to understand the functioning principles of a system, rather than the components which make it up. Cybernetics tries to answer the question of how systems consume information, construct models and manage actions in order to reach a goal. The aforementioned question might be about systems from many different disciplines such as: biology, physics, psychology, technology, sociology or a mix of these. Cybernetics is not solely concerned with engineered systems, but also with organic, evolving systems which are self-controlled.

Second-order cybernetics mainly focuses on the impact and performance of human observation when building system models. It was realized that the systems which were built, would only acknowledge facts that are significant for its objective. This situation stems from their dependency on their designers. These kinds of models are inactive and open to be steered by the designers. Second-order cybernetics leans on the idea of a system being alive, having interactions with other systems and the observers. Artificial intelligence is among the fields on which modern cybernetics have an important influence [8].

In medicine, artificial intelligence is used in two major forms: virtual and physical. The physical form consists of medical equipment, physical entities and care-giver

robots which become more advanced day by day. Robots can take part in surgeries either as the main surgeons or as assistants. Artificial intelligence may also be used in the process of guided drug delivery to specific organs or tissues. It is a necessity to interrogate the impact of these applications on the physical and psychological state of patients, adverse events, quantifiable results and health signs [7].

The virtual part primarily corresponds to machine learning and deep learning applications, which involve algorithms ameliorating their learning via knowledge and training. Machine learning can be described as the study domain which enables computers to learn without programming them specifically. In other words, it aims to program computers so that they can optimize models by the aid of data samples and past evidence. To sum up, the concept of machine learning is based on tutoring a computer to wisely carry out tasks by experience it gains from an external environment.

A machine learning application must be able to learn from the information it gathers from its surroundings to be successful. In this particular framework, learning corresponds to projecting dependencies from given information.

Machine learning cannot be defined without mentioning data mining. Data mining employs machine learning methods in order to investigate extensive databases and discover hard-to-find information within those databases. On the other hand, machine learning applications use data mining during data preprocessing as a preparatory step for learning specific tasks. Nonetheless, one should keep in mind that machine learning offers solutions to complicated artificial intelligence problems via adaptation to constantly evolving cases, as well as resolving complications concerning databases.

Machine learning methods might be classified according to how they label data. Supervised algorithms aim to label an output, using the labeled data which are present. On the contrary, unsupervised methods take solely the input instances of a sample to learn from them. The third class would be semi-supervised methods which is a mixture of the aforementioned two classes. These kinds of methods use data which have both labeled and unlabeled data; and the labeled part is exploited to deduce information

about the unlabeled part.

Machine learning methods can also be classified based on their probabilistic aspects. There are discriminant methods which aim to find the probability of an output dependent on the inputs. Support vector machines might be an example belonging to this class. In the other class, there are generative methods which are entirely probabilistic. These methods may use Bayesian networks, or a technique that does not include graph modeling such as naive Bayes.

With regards to the methodology that should be used when constructing a successful machine learning application, primarily, the essence of the problem must be clearly defined. Input samples and the outputs are considered in this step. Moreover, the quality of data which are fed to the model, is crucial. The impact of bad quality input data utilization cannot be compensated by a good model. Even though the model is robust to noise, it is principally based on estimates from the input data. Lastly, a successful model should be optimally simple, whereas more complex models tend to be too much dependent on the input samples, which results in overfitting. If the model's simplicity is preserved, the data which are not present may also be taken into account [9].

Ensemble learning is a methodology of machine learning which builds a collection of classifiers and decides on the class of a new data instance by voting these classifiers' prediction results. That is, predictions of multiple classifiers are blended into one final decision using weighted/unweighted voting. Studies have been and are being made about the development of successful classifier ensembles. Generally, ensembles have a tendency to achieve higher accuracies than the classifiers of which they consist.

The primary requirements of an ensemble to perform better than all of its classifiers is that each classifier is sufficiently accurate and distinct. Sufficient accuracy can be described as predicting with less error than random prediction. Classifiers being distinct signifies their difference of prediction errors for new instances. For example; if there were five classifiers which all made the same incorrect prediction for a new

data point, the ensemble result would also be incorrect. Contrarily, if the classifiers are unique, the majority of them could classify the new instance correctly while the others are incorrect. This would result in the final decision to be correct, which means the ensemble is successful thanks to this diversity.

A machine learning algorithm's goal is to find the optimum hypothesis in a pool of hypotheses. It becomes a challenge when this pool of hypotheses is wider than the size of the training sample. If the sample is too small, the algorithm faces the problem of detecting too many identical hypotheses which have the exact same accuracy. One way of increasing the chance of selecting the correct classifier is to create a set of accurate learners. This is one of the reasons why an ensemble might work wonderfully.

Another argument is in the computational aspect. A lot of learners apply local searches, which raises the risk of getting stuck on the minimum or maximum of the objective function for a particular region of the input space. The risk is lowered if the local search gets started off from various distinct points by establishing an ensemble. This method may provide a superior estimation of the correct function than the classifiers within the ensemble.

The last rationale is from a representational stand point. It is likely to acquire a broader representation of functions when the hypotheses in the pool are given weights and then summed up. The preceding points are the essential problems that cause the current individual learners to fall flat. On that account, ensemble learning offers the opportunity to remove or at least weaken these deficiencies that typical learning algorithms have.

There are different ways an ensemble can be built. One method influences the training data and develops several hypotheses. Multiple subsets of training data are used to run the classifier numerous times. This method might be the most preferable in case of unsteady algorithms. If the output classifier varies immensely in return of a little modification to the training sample, the algorithm is considered to be unstable.

Another approach to building an ensemble is introducing randomness into the algorithm. In the process of neural networks training, errors from output nodes are back propagated to the input nodes. During the backpropagation, the network's beginning weights are determined at random. This technique enables the classifier to be at variance with each different beginning weighting [10].

Random subspace is among the approaches to ensemble methods. It works by randomizing the algorithm through the medium of constructing feature subsets and using these subsets to train the classifier. Afterwards, the results of these classifiers are merged into one final output through majority voting.

Random subspace may be used with different algorithms such as discriminant analysis. With the aim of reducing error rates, random subset of features can be fed into different learners. The disadvantage of this technique is that, in certain cases, the capability of differentiation is weak because of random picking. When this situation occurs, the final output is also inadequate. However, the inconvenience can be defeated by majority voting. Often, an individual learner of the ensemble consumes a limited portion of the existing features. Furthermore, each of them is capable of classifying new data points. The majority voting process enables the classifiers to distinctly classify new data points. The definitive result is determined as the prediction which gets the majority of votes among all classification decisions [11].

In medicine, there are numerous successful artificial intelligence applications and one of them is an unsupervised methodology using unsupervised protein-protein interaction graphs. This application encouraged innovative findings of therapeutic targets. The proposed methodology integrated an adaptive evolutionary algorithm with advanced clustering techniques [12]. These kinds of methods are more vigorous and less likely to overfit, which may happen in case the training sample is not sufficiently large compared to the number of parameters.

Another AI application in medicine is its use for providing care for patients. This can be in the form of avatars which are usually VR simulations, robots or systems

with physical sensors and sound. Natural language processing (NLP), VR simulation and artificial intelligence built on human expert knowledge are rapidly progressing research areas. Advances in these fields also give rise to the development of artificial intelligence care providers. AICPs have been established and evaluated for clinical exercise, ability attainment and presenting information on mental health and assistance to patients in need. Additional purposes of artificial intelligence care providers are tracking, evaluation and consulting about self-care [13].

An emerging field, Radiomics, is concerned with reaching useful diagnostic, prognostic and predictive information. It has the ability to provide many advantages in cancer imaging. Radiomics focuses on obtaining quantitative information from clinical images. This approach helps characterize the image phenotypes of the tumor in a more detailed way.

In radiomics, the goal is to exploit the data present in the clinical environment and draw powerful insights from it based on state-of-the-art, occasionally non-intuitive mathematical interpretation. Medical images incorporate information which is useful for distinguishing diseases. Human eye is able to perceive only a part of this useful information. Therefore, there is a need for tools which are capable of capturing and presenting the knowledge that may have a shaping impact on diagnosis and treatment of diseases. Radiomics offers this opportunity in oncology for this reason.

Radiomics obtains quantitative texture information using artificial intelligence to extract intensity spatial distribution and pixel correlations in an image. Different experts might get varying insights from a medical image, hence, the decisions made according to the insights may largely depend on the expert's judgment. Inconsistencies that may arise because of this situation can be averted by quantifying the visually detectable elements such as texture, form and intensity changes in medical images. In this fashion, radiomics enhances the data which are needed during the diagnosis routine.

Medical images obtained via various imaging modalities such as CT, PET and

MRI can be put through radiomics methodology. Radiomics presents the opportunity to make analyses based on the combined information gathered from different imaging methods. Therefore, it becomes more advantageous than analyzing data from each imaging modality separately. On the other hand, the circumstances chosen for the analysis may differ among researchers, which have a high impact on the outcomes. For this reason; standardization, steadiness and universality between all studies must be established.

A radiomics analysis starts with determining the region or volume of interest. This is an important part of the process as the features that will be used are extracted within ROI or VOI. Images may be segmented by an expert which is called manual segmentation. It can also be performed automatically; either semi or fully. Level set based active contouring, localized region based active contouring and seeded region growing are some of the semi-automatic image segmentation methods. Generally, deep learning is used for segmenting images completely automatically [14].

The main objective of texture segmentation is to differentiate the components in the image so as to focus on the ones that are more important for the purposes of the analysis. Moreover, differentiation of foreground and background contents can be possible with segmentation which is required for some analyzes. Segmentation simplifies the analysis of an image by converting it into an image with the desired number of regions with respect to texture characteristics [15].

In Medical Imaging, segmentation serves the goal of recognizing anatomical structures. It presents solutions to problems such as detecting tumors, tissue analysis and bone fractures. One of the factors complicating segmentation in medical images is noise, which is related to the sensors and electronic system used. Noise makes it harder to decide in which class a pixel/voxel belongs. Another factor is the partial volume effect which is present both in CT and MR images. There are also the problems arising from the presence of different artifacts such as motion and ring artifacts [16]. In kidney tumor segmentation specifically, the non-uniform motion, various shapes and similar appearance produce extra difficulties [17]. Search for finding new segmentation

methods which can overcome the complications caused by these factors is a matter of focus in medical image analysis.

After segmentation, images need to be processed prior to feature extraction. Image pre-processing renders the images more homogenized in relation to pixel distribution, gray level intensity and histogram bins. This step has great importance as it affects the potency of the extracted features. For instance, texture features can be made stable in terms of rotation by interpolation. Normalization is another pre-processing technique. It enables the elimination of pixels with gray levels higher or lower than the predefined limits. Finally, discretization is performed by creating bins of gray level intensities.

Image pre-processing is followed by radiomic feature extraction which indicates the computation of feature values. There are a variety of features such as texture, shape, histogram and radial. Usually, the number of features extracted need to be reduced in order to avoid overfitting. Feature selection signifies the choosing of features which will be meaningful for the model. Correlation between the features should also be considered and significantly correlated features must be excluded from the model.

In conclusion, a radiomics application should fulfill an existing clinical necessity. Moreover, it should be based on sufficient data to justify the inferences of the study. If possible, including data other than the ones obtained from medical images might be considered. Finally, for an optimal radiomics study, images should be acquired in a standard way or if not, the methodology should compensate for the changeable procedures which may have been used [14].

One of the the main objectives of the presented work is to perform a comparison among existing machine learning methods for the classification of tumor histologic subtypes of renal cell carcinoma (RCC) patients. It also focuses on the question of which radiomic features and patient clinical data provide meaningful information about the histologic subtypes.

2. LITERATURE REVIEW

In a 2019 study [18], Convolutional Neural Networks were used to classify the tumor histologic subtypes of RCC cases. The data set included clear-cell, chromophobe and papillary subtypes. The model was fed with three-phase CT images and one slice from each phase was used. AUC values for differentiating clear-cell from non-clear-cell, papillary from non-papillary and chromophobe from non-chromophobe were; 0.93, 0.91 and 0.88 respectively.

Another paper by Kocak et al. focused on classifying the tumors as ccRCC or Non-ccRCC, as well as differentiation of ccRCC, pcRCC and chcRCC from each other. Artificial Neural Networks classifier predicted the subtypes as ccRCC or non-ccRCC with an AUC of 0.92, while the AUC of Support Vector Machine classifier was 0.79. Both of them performed worse in the three-class models [19].

Zhang et al. evaluated several models incorporating SVMs for classifying tumors as ccRCC or non-ccRCC, and chromophobe or papillary RCC. Slices with the largest cross-sectional area of the lesion from 3-phase CT images were used. Top 3 features were selected by Mann-Whitney U-tests, ROC curves and Pearson's correlation coefficient methods. An SVM with a nonlinear radial basis function kernel was implemented. Best results were achieved using the corticomedullary phase images. AUC for ccRCC vs. non-ccRCC classification was 0.94 [20].

Hoang et al., conducted a study which used random forest models for three classifications: oncocytoma vs. RCC, oncocytoma vs. ccRCC and papillary vs. ccRCC. Three consecutive slices containing the largest cross-sectional area from each of the four phases of MR images of 142 lesions from 41 patients were included. Pairwise Wilcoxon rank test, modified false discovery rate adjustment, Lasso regression were used for feature selection. ccRCC cases were distinguished from oncocytomas with an average accuracy of 77.9% [21].

3. MATERIALS AND METHODS

3.1 Data Sets

CT images and patient clinical data from the Climb 4 Kidney Cancer-Kidney Tumor Segmentation Challenge (C4KC-KiTS) database [22],[23] were acquired. 210 patients were included in the C4KC-KiTS database. In this study, 192 of the cases which had malignant tumors were used.

3.2 Image Pre-processing

Resampling, intensity normalization and gray level discretization were applied before starting the feature extraction process. Images had different slice thicknesses (0.5 mm to 5 mm) and different pixel sizes (0.438 mm to 1.04 mm). After reconstruction and resampling, 1 mm x 1 mm x 1 mm spatial resolution was achieved. Python software was used to perform resampling, and the new values of the resampled images were obtained by Cubic B-Spline interpolation method [24]. Z-score normalization was used for the normalization of image intensity values. For gray level discretization, bin width was adjusted to be 0.01 on 3D Slicer software [25]. Gray level discretization lessens the heterogeneity influences on the images, resulting from acquisition and reconstruction protocols [26].

3.3 Feature Extraction

Radiologic images carry relevant and significant clinical information [27]. Feature extraction is an important step for finding the link between disease and image attributes, on the grounds that its enablement to obtain solid, quantitative representations.

Features were extracted using PyRadiomics extension on 3D Slicer. Three types of images were subject to feature extraction: original, Laplacian of Gaussian (LoG) and wavelet-transformed. Laplacian of Gaussian filter values were 2 mm, 4 mm, and 6 mm in order to explain patterns with various sizes.

After all radiomic features were extracted, certain patient clinical data were added to get a combined data set. Clinical data included information such as age, sex, body mass index; as well as presence of several diseases, alcohol and tobacco use. Afterwards, the combined data were split into training and test sets as 85% and 15% respectively. As a result, 162 training cases included 128 clear-cell RCC and 34 other histologic subtypes (chromophobe, papillary, clear-cell papillary, multilocular cystic, urothelial, wilms). Further, 15 of the test cases were clear-cell and the remaining 15 were other (chromophobe, papillary, clear-cell papillary).

3.4 Feature Selection

Feature selection process was executed on Matlab (R2021b) software. Kruskal-Wallis (KW) test was conducted as the first step of feature selection. KW compares the medians of the groups of data to determine if the samples come from the same population. In this case, each feature was tested for its ability to differentiate between the data classified as clear-cell and other with $p = 0.05$. Only 111 features among the 1157 were decided to be relevant. Moreover, none of the patient clinical features were selected.

Secondly, least absolute shrinkage and selection Operator (LASSO) was used at the next phase for selecting features. Lasso is an improved version of ordinary least squares estimates in regression analysis combining subset selection and ridge regression [28]. It causes some regression coefficients to shrink and set some of them to zero. At the end, coefficients belonging to the less important features become zero. Lambda with the minimum standard error was chosen to obtain the optimal set of coefficients. Subsequently, 8 features were selected as the most relevant for our model.

3.5 Model Training and Evaluation

Coarse Gaussian Support Vector Machine and Subspace Discriminant classifiers were trained with the selected features in the Classification Learner App of Matlab. SVM classifier aims to find the optimal hyperplane in the N-dimensional space that distinctly classifies the data points. The optimal hyperplane can be described as the most distant of all possible ones to both classes. The data points closest to this hyperplane are defined as support vectors. In some cases simple hyperplanes do not show sufficient separation performance. Hence, kernels are reproduced which use several functions. In this study, the SVM classifier with a Gaussian (radial basis function) kernel was used. Box constraint level was 1 and kernel scale was chosen as 11 for the classifier. 5-fold cross validation was used in the training process.

Discriminant classifiers assume that different classes have different Gaussian distributions. Their objective is to classify the data points while minimizing the classification cost. Ensemble learning combines several classifiers to improve the prediction performance. Each learner, discriminant classifier, is trained using a random subset of features among the selected ones. At the end, the best model is introduced. The model included 30 learners with 4 subspaces and the training was performed using 5-fold cross validation. Previously reserved test data set was used to test the model performances. Accuracy and area under the curve (AUC) for both models were calculated for evaluation.

4. RESULTS

4.1 Feature Extraction and Selection

In addition to the 27 clinical features (see Figure 4.1), 1130 radiomic features were extracted from the CT images, adding up to 1157 features in total. Among the radiomic features; 744 were from wavelet-transformed images with 8 distinct filters and 6 classes of features (first-order, gray level dependence matrix (gldm), gray level co-occurrence matrix (glcm), gray level run-length matrix (glrlm), gray level size zone matrix (glszm) and neighboring gray tone difference matrix (ngtdm). Laplacian of Gaussian (LoG) filtered images produced 279 features, while 107 features were extracted from original images.

Feature Name	Feature Name
gender	malignant_lymphoma
body_mass_index	localized_solid_tumor
myocardial_infarction	metastatic_solid_tumor
congestive_heart_failure	moderate_to_severe_liver_disease
peripheral_vascular_disease	smoking_history_never_smoked
cerebrovascular_disease	smoking_history_previous_smoker
copd	smoking_history_current_smoker
connective_tissue_disease	chewing_tobacco_use_never_or_not_in_last_3mo
peptic_ulcer_disease	chewing_tobacco_use_quit_in_last_3mo
uncomplicated_diabetes_mellitus	alcohol_use_two_or_less_daily
diabetes_mellitus_with_end_organ_damage	alcohol_use_never_or_not_in_last_3mo
chronic_kidney_disease	alcohol_use_more_than_two_daily
hemiplegia_from_stroke	radiographic_size
leukemia	

Figure 4.1 List of clinical features.

As a result of the Kruskal-Wallis test, 111 features were eliminated as they were not significant for the problem in question. The process was followed by Lasso regression to detect the most useful features, which left 8 of them (see Figure 4.2) to be used in the classification models. These included: First Order Interquartile Range, GLCM Inverse Difference Normalized and GLRLM Run Variance. Prior to model training, new instances belonging to the minority class were created synthetic minority

Image Type	Feature Name
Original	First Order - Interquartile Range
Original	GLCM - IDN
Log filtered (sigma: 2 mm)	3D GLRLM – Long Run Emphasis
Log filtered (sigma: 2 mm)	3D GLRLM – Run Variance
Log filtered (sigma: 2 mm)	3D GLDM - Dependence Variance
Log filtered (sigma: 4 mm)	3D First Order - Kurtosis
Log filtered (sigma: 6 mm)	3D First Order - 90^{th} Percentile
Log filtered (sigma: 6 mm)	3D First Order - Maximum

Figure 4.2 Selected features for the models.

oversampling technique (SMOTE) to balance the training set. Ultimately, both classes consisted of 128 cases and the models were trained with a total number of 256 cases.

4.2 Performance Evaluation

Coarse Gaussian SVM was faster compared to Subspace Discriminant with a training time of 0.47 sec and 11000 obs/sec (observations per second) prediction speed. Training duration of Subspace Discriminant was 4.1 sec with 960 obs/sec prediction speed. For Coarse Gaussian SVM; validation accuracy was 67.6% while the accuracy of test was 80%, with and AUC of 0.86. Similarly, Subspace Discriminant had 68.8% validation accuracy and 80% test accuracy; AUC was 0.85. Figure 4.1 shows the confusion matrices of the two models. The receiving operator characteristic (ROC) curves can be seen on Figure 4.2.

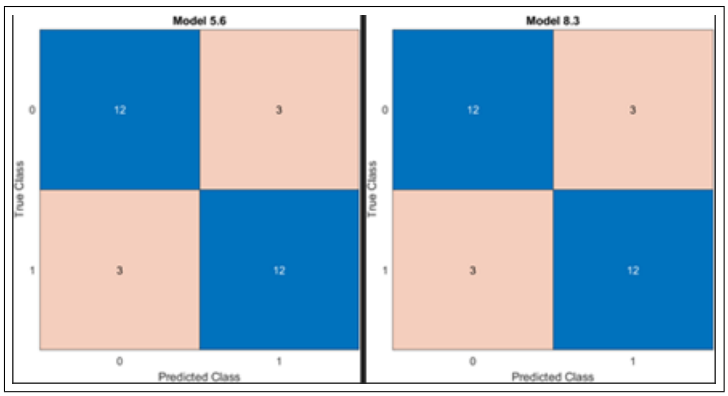

Figure 4.3 Confusion matrices for Coarse Gaussian SVM and Subspace Discriminant on test data set.

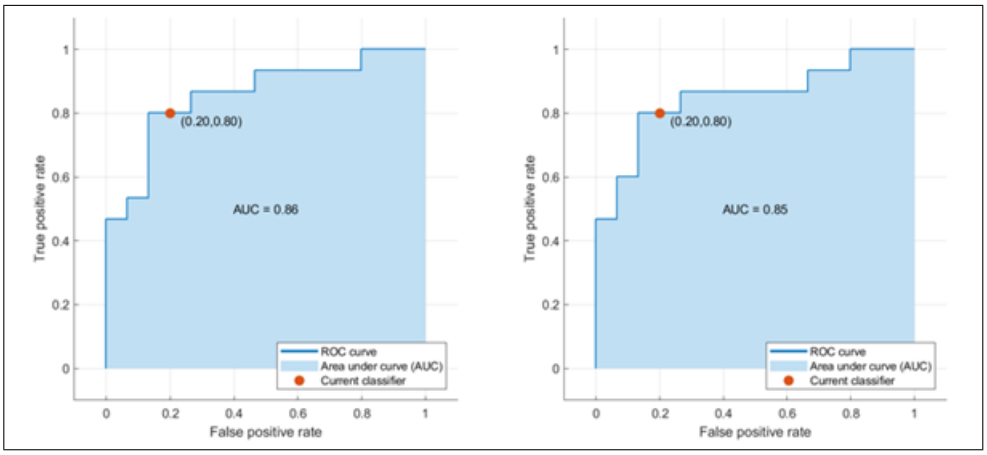

Figure 4.4 ROC curves of classification models on the test dataset.

5. DISCUSSION

For metastatic cases, management of surgical process and cases for which surgery is not an option; differentiating RCC histologic subtypes can be of great importance. Subtype information may also be considerably helpful for new targeted immunotherapy opportunities which are specific to histologic subtypes. There are other emerging therapies which address the therapeutic demands of particular subtypes. Due to the several disadvantages of biopsy such as mistakes of sampling, complication risk, necessity to obtain adequate samples and its potential impact on the treatment process; its comprehensive use is questionable.

This study investigates the usefulness of machine learning algorithms for malignant kidney tumor histologic subtype classification. In consideration of the performance evaluation, both models demonstrated promising results when classifying the tumors as clear-cell RCC or non-clear-cell RCC. Nonetheless, Coarse Gaussian SVM might be slightly more preferable because of its training and prediction speed.

For the method that Han et al. introduced, medical experts needed to determine the positions of the tumors [18]. On the other hand, the classification of different tumor subtypes was successful. Therefore, the method can provide assistance for radiologists on the next steps of the disease. It may also aid radiologists with little experience to diagnose cancer.

Their method efficiently classified papillary RCC and clear cell RCC; yet, the efficiency was low for classifying chromophobe RCC. This may happen because of having an uneven number of instances in classes. In this case, each class had approximately the same number of instances. The reason might also be that chromophobe RCC would need other features in order to be differentiated from other subtypes, which were not included in this study. Conducting research with additional data may increase the performance.

The approach was based on linearly combining three channel images so as to make them more suitable for machine learning models. The study demonstrated that three channel images might be altered by means of achieving improved results, even though it is not possible to fix their weights entirely. It was also deduced that the optimal weight might be various for different tasks. The linear combination weights were determined in the course of learning and the best weights were automatically chosen based on the training sample. Expert opinion can be gathered in order to find out if the linearly combined images make it simpler to distinguish the suspected region.

In their paper, researchers showed that approaching the discrimination goal as a three-class problem decreases the model performance when compared to the two-class classification model. This can be a sign of high similarity between some tumor subtypes. In addition, they discovered that a bigger volume of data is needed in order to optimize the three-class model than the two-class model. Also, the weight quantity is greater in the three-class model than it is in the two-class model, though, the reason for this could not be explained with the present study.

In another study, Kocak B.et al. prepared two-class models to classify three RCC subtypes providing a comparison of their performances [19]. They expressed that the models which aimed to differentiate clear cell RCC from non-clear cell RCC outperformed the models with other classification goals. The performance gap was especially big for the three-class models as they performed unsatisfactorily when compared against other models.

Kocak B.et al. also deduced that utilization of corticomedullary phase CT images enabled the models to show superior performance as opposed to the utilization of unenhanced CT images. This situation might suggest occurence of generalizability issues in case of erroneous features selection when unenhanced CT images are used in these models. In order to compensate for this issue, they claimed that methods such as bagging or adaptive boosting may be useful.

Validating the model performances externally is beneficial in terms of gener-

alizability, yet, it does not solve all the related problems. For instance, overfitting becomes more probable because of the fact that a small sized training sample was used. Moreover, segmentation of two image slices may not be clinically applicable although the suggested models require them. It should also be taken into account that external validation images were pre-processed as the internal images with the exception of reconstruction.

Although renal cell carcinoma has many other uncommon subtypes, these were not included in the study because of insufficient cases. This reasoning goes for our study as well. In the paper of Kocak B.et al., it was also pointed out that working with cases which are divided into two separate groups in terms of small and large sized tumors might answer other specific questions.

Hoang et al. conducted a research which argued for the usefulness of analyzing contrast enhanced magnetic resonance image textures in order to distinguish renal tumor histologic subtypes [21]. They chose to construct a random forest model for this purpose. Also, they took advantage of the tumor's complete cross sectional area to extract features.

One of the main findings was that clear cell RCC exhibited weaker arterial enhancement compared to oncocytoma. There are studies that have reached the opposite of this result, as well as the ones that have reached the same conclusion. This might mean that it is difficult to differentiate ccRCCs and oncocytomas by their imaging features. On the other hand, their study had promising outcomes as their models were significantly more successful to differentiate between oncocytomas and two subtypes of renal cell carcinomas than a research with expert opinions.

Another discovery they made was the existence of significant image texture features for discriminating clear cell RCC from papillary RCC. This discovery is consistent with the opinion that papillary RCC has higher homogeneity than clear cell RCC, especially for small kidney tumors. The research was limited with the inclusion of only small renal masses because they detected a particular pressing lack in this area.

The same study had a few restraints such as the exclusion of chromophobe RCC and angiomyolipoma due to insufficient data. Secondly, the number of patients being too low against the number of lesions might have resulted in misleading accuracy for detecting oncocytomas. For this reason, the model should be evaluated with a wider patient sample to apprehend the variation among patients and achieve better precision.

Zhang et al. found that standard deviation, entropy and mean of positive pixels provide useful information about non clear cell RCCs on images [20]. In addition, skewness and mean of positive pixels can be included in models to distinguish between chromophobe RCC and papillary RCC. They also discovered that clear cell RCC presented notably higher mean, standard deviation, mean of positive pixels, entropy and lower kurtosis than non clear cell RCC.

According to the results of the paper; raise in the amount of highlighted features during filtration and the difference of mean brightness in comparison with background might cause heterogeneity between tumor histologic subtypes. These may also give rise to the observation of greater standard deviation, mean, entropy, mean of positive pixels and smaller kurtosis values for clear cell RCC.

In the study, authors claimed that combining different texture features might be useful for the distinction of clear cell and non-clear cell RCC even though each feature does not have the power to make the distinction by itself alone. Moreover, contrast-enhanced CT appeared to provide texture features with stronger differentiation ability when compared to unenhanced CT images. The accuracy of diagnosis was also greater for the models which incorporated enhanced images.

One of the disadvantages of the study was the number of cases being low, which may decrease the accuracy of cross validation. Moreover, due to the fact that some clear cell RCC cases were left out, the scope of this study might not reflect some of the clear cell RCC texture traits. Authors also pointed out that the diagnostic accuracy might have been overestimated because they used data which was collected in the past.

Our methodology produced similar results as other studies focusing on the similar questions. Therefore, we can deduce that machine learning in radiomics is a viable method for determining the histologic tumor subtype of renal tumors. However, our study differs from others with the data source which was used, as well as other dimensions such as having a high number of cases. Dissimilar to the studies of Kocak B. et al., Zhang G. et al., Hoang et al. and Han et al., this study used all slices of the CT images as an input to the models. Furthermore, we tested if the inclusion of patient clinical data would be useful. Our study found that the specific clinical data included did not have an impact on the classification.

In the future, improved models might be constructed by the addition of blood and urine biomarkers as clinical features. Our methodology can be seen as the next course of action after determining that the tumor is malignant. Therefore, performing a classification for differentiating benign and malignant cases prior to the application of our method can be considered. In addition, automated segmentation is crucial in order for this methodology to be applicable in a real clinical environment. Increasing the size of the data set to achieve better representation of other histologic subtypes can also be considered in order to answer different classification problems.

6. CONCLUSION

We proposed two different models based on machine learning algorithms to label the malignant tumor cases as ccRCC or non-ccRCC using relevant radiomic features extracted from renal CT images. Both models produced similar results which can be considered as encouraging. These types of classifiers were considered for the first time. This work supports the objective of having a fast and non-invasive technique in the diagnosis process of RCC patients; specifically for deciding the tumor histologic subtype.

7. List of publications produced from the thesis

1. Evaluation Of Machine Learning Algorithms For Predicting Kidney Tumor Subtype From CT Images Using Radiomics and Clinical Features, D. Şirin, A Güveniş, " *The International Izmir Democracy University Engineering Symposium (IES'21) Proceedings Abstract Book*, pp. 112, Dec. 13–18, 2021.

2. Predicting Kidney Tumor Subtype From CT Images Using Radiomics and Clinical Features, D. Şirin, A Güveniş, *Natural and Applied Sciences Journal*, Vol. 5, pp: 29-37, 2022.

REFERENCES

1. Hsieh, J. J., M. P. Purdue, S. Signoretti, and C. S. et al., "Renal cell carcinoma," *Nature Reviews Disease Primers*, Vol. 3, Mar 2017.

2. Nowell, P. C., "The clonal evolution of tumor cell populations: Acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression," *Science*, Vol. 194, no. 4260, pp. 23–28, 1976.

3. Şenbabaoğlu, Y., R. S. Gejman, A. G. Winer, and M. L. et al., "Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger rna signatures," *Genome Biology*, Vol. 17, no. 1, pp. 1–25, 2016.

4. Le, V. H., and J. J. Hsieh, "Genomics and genetics of clear cell renal cell carcinoma: A mini-review," *Journal of Translational Genetics and Genomics*, Vol. 2, Nov 2018.

5. H B Ökmen, A. Güveniş, and H. Uysal, "Predicting the polybromo-1 (pbrm1) mutation of a clear cell renal cell carcinoma using computed tomography images and knn classification with random subspace," *Vibroengineering Procedia*, Vol. 26, pp. 30–34, 2019.

6. van Oostenbrugge, T. J., J. J. Fütterer, and P. F. A. Mulders, "Diagnostic imaging for solid renal tumors: a pictorial review," *Kidney Cancer*, Vol. 2, no. 2, pp. 79–93, 2018.

7. Hamet, P., and J. Tremblay, "Artificial intelligence in medicine," *Metabolism*, Vol. 69, pp. S36–S40, Apr 2017. Insights Into the Future of Medicine: Technologies, Concepts, and Integration.

8. Heylighen, F., and C. Joslyn, "Cybernetics and second-order cybernetics," in *Encyclopedia of Physical Science and Technology (Third Edition)* (Meyers, R. A., ed.), pp. 155–169, New York: Academic Press, third edition ed., 2003.

9. el Naqa, I., and M. J. Murphy, *What Is Machine Learning?*, pp. 3–11. Cham: Springer International Publishing, 2015.

10. Dietterich, T. G., "Ensemble methods in machine learning," in *Multiple Classifier Systems*, (Berlin–Germany), pp. 1–15, Springer Berlin Heidelberg, 2000.

11. Ashour, A. S., Y. Guo, A. R. Hawas, and G. Xu, "Ensemble of subspace discriminant classifiers for schistosomal liver fibrosis staging in mice microscopic images," *Health Information Science and Systems*, Vol. 6, Nov 2018.

12. Theofilatos, K., N. Pavlopoulou, C. Papasavvas, and S. L. et al., "Predicting protein complexes from weighted protein–protein interaction graphs with a novel unsupervised methodology: evolutionary enhanced markov clustering," *Artificial Intelligence In Medicine*, Vol. 63, no. 3, pp. 181–189, 2015.

13. Luxton, D. D., "Recommendations for the ethical use and design of artificial intelligent care providers," *Artificial Intelligence in Medicine*, Vol. 62, pp. 1–10, Sep 2014.

14. Van-Timmeren, J. E., D. Cester, S. Tanadini-Lang, and H. A. et al., "Radiomics in medical imaging - âhow-toâ guide and critical reflection," *Insights Into Imaging*, Vol. 11, no. 1, pp. 1–16, 2020.

15. Bhosle, V. V., and V. P. Pawar, "Texture segmentation: different methods," *International Journal of Soft Computing and Engineering*, Vol. 3, pp. 69–74, Nov 2013.

16. Sharma, N., and L. M. Aggarwal, "Automated medical image segmentation techniques," *Journal of Medical Physics*, Vol. 26, 2010.

17. Sabarinathan, D., M. Parisa-Beham, and S. M. Mansoor-Roomi, "Hyper vision net: kidney tumor segmentation using coordinate convolutional layer and attention unit," in *National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics*, pp. 609–618, Springer, 2019.

18. Han, S., S. I. Hwang, and H. J. Lee, "The classification of renal cancer in 3-phase ct images using a deep learning method," *Journal of Digital Imaging*, Vol. 32, pp. 638â–643, May 2019.

19. Kocak, B., A. H. Yardimci, C. Turan-Bektas, and M. H. T. et al., "Textural differences between renal cell carcinoma subtypes: Machine learning-based quantitative computed tomography texture analysis with independent external validation," *European Journal of Radiology*, Vol. 107, pp. 149–157, 2018.

20. Zhang, G. M. Y., B. Shi, H. D. Xue, and B. G. et al., "Can quantitative ct texture analysis be used to differentiate subtypes of renal cell carcinoma?," *Clinical Radiology*, Vol. 74, no. 4, pp. 287–294, 2019.

21. Hoang, U. N., S. M. Mirmomen, O. Meirelles, and J. Y. et al., "Assessment of multiphasic contrast-enhanced mr textures in differentiating small renal mass subtypes," *Abdominal Radiology*, Vol. 43, pp. 3400–3409, June 2018.

22. Clark, K., B. Vendt, and K. S. et al., "The cancer imaging archive (tcia): Maintaining and operating a public information repository," *Journal of Digital Imaging*, Vol. 26, pp. 1045–57, Dec 2013.

23. Heller, N., N. Sathianathen, A. Kalapara, and E. W. et al., "Data from c4kc-kits [data set]," *The Cancer Imaging Archive*, Vol. 10, 2019.

24. Wang, Z., K. Wang, and S. An, "Cubic b-spline interpolation and realization," in *International Conference on Information Computing and Applications*, pp. 82–89, Springer, 2011.

25. Fedorov, A., R. Beichel, J. Kalpathy-Cramer, and J. F. et al., "3d slicer as an image computing platform for the quantitative imaging network," *Magnetic Resonance Imaging*, Vol. 30, pp. 1323–1341, Nov 2012. Quantitative Imaging in Cancer.

26. Larue, R. T. H. M., J. E. van Timmeren, E. E. C. de Jong, and G. F. et al., "Influence of gray level discretization on radiomic feature stability for different ct scanners, tube currents and slice thicknesses: a comprehensive phantom study," *Acta Oncologica*, Vol. 56, pp. 1544–1553, Sep 2017.

27. Tomaszewski, M. R., and R. J. Gillies, "The biological meaning of radiomic features," *Radiology*, Vol. 298, no. 3, p. 505, 2021.

28. Tibshirani, R., "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 58, no. 1, pp. 267–288, 1996.