Advanced Machine Learning Models for Accurate Kidney Cancer Classification Using CT Images

Article *in* Mesopotamian Journal of Big Data · January 2025

DOI: 10.58496/MJBD/2025/001

2 authors:

Dhuha Abdalredha
Nahrain University
3 PUBLICATIONS **12** CITATIONS
SEE PROFILE

Mazin Abed Mohammed
University of Anbar
350 PUBLICATIONS **13,303** CITATIONS
SEE PROFILE

Mesopotamian journal of Big Data
Vol. **(2025)**, 2025, **pp**. 1–25
DOI: https://doi.org/10.58496/MJBD/2025/001 ISSN: 2958-6453
https://mesopotamian.press/journals/index.php/BigData

Research Article

Advanced Machine Learning Models for Accurate Kidney Cancer Classification Using CT Images

Dhuha Abdalredha Kadhim [1], Mazin Abed Mohammed [2, *],

[1] *Informatics Institute for Postgraduate Studies, Iraqi Commission for Computer & Informatics, Baghdad, Iraq.*

[2] *Department of Artificial Intelligence, College of Computer Science and Information Technology, University of Anbar, Anbar, Iraq.*

ARTICLE INFO

Article History

Received 20 Oct 2024
Accepted 11 Dec 2024
Published 10 Jan 2025

Keywords

Kidney Cancer
Support Vector Machine
Machine Learning
CT Images
Classification

ABSTRACT

Kidney cancer, particularly renal cell carcinoma (RCC), poses significant challenges in early and accurate diagnosis due to the complexity of tumor characteristics in computerized tomography (CT) images. Traditional diagnostic approaches often struggle with variability in data and lack the precision required for effective clinical decision-making. This study aims to develop and evaluate machine learning (ML) models for the accurate classification of kidney cancer using CT images, focusing on improving diagnostic precision and addressing potential challenges of overfitting and dataset heterogeneity. Two ML models, Support Vector Machines (SVM) and Multi-Layer Perceptrons (MLP), were employed for classification. Key attribute extraction techniques, including grayscale-level co-occurrence matrix (GLCM) and Gabor filters, were utilized to capture texture and structural features of CT images. Data normalization and preprocessing ensured consistency and enhanced model reliability. The SVM model achieved an accuracy of 93%, while the MLP model demonstrated superior performance with a 99.64% accuracy rate. These results highlight the MLP model's ability to capture complex patterns in the data. However, the exceptional accuracy of the MLP model raises concerns about potential overfitting, warranting further evaluation on more diverse datasets. This study underscores the potential of ML techniques, particularly MLP, in enhancing the accuracy of kidney cancer diagnosis. Integrating such advanced ML models into clinical workflows could significantly improve patient outcomes.

1. INTRODUCTION

Kidney cancer is a pressing problem on the global level as it is among ten most frequent cancers across the globe. It affects a growing number of people, especially over the last few decades, thanks to improvements in imaging and diagnostic tools [1]. K Styx renal cell carcinoma is the most frequent form of kidney cancer in adults and is estimated to affect around 90% of all patients. There are also other less common types of the RCC that make hard to distinguish: transitional cell carcinoma and renal sarcoma [2]. Kidney cancer staging is essential in the pre-operative management of the procedure because it determines the choice of operation and survival of the patient. However, the conventional methods of diagnosis involving imaging and histopathological comparisons have their own problems, which include subjectivity in examinations, differences in readings by different observers and most importantly, there exists no clear means of differentiating between one pathology and another, especially where they are almost similar [3][24].

The development of ML has introduced a significant change in medical diagnostic planning as it provides capable methods for handling large datasets [4, 5]. A particular strength of the ML approaches is based on using computational algorithms to derive subtle features in large data sets, serving as strong indicators to improve diagnostic accuracy, as well as reduce the erroneous input from humans. In the context of Kidney cancer classification where medical image data are common, other advanced Machine learning models such as Support Vector Machines (SVM) and Multi Layer Perceptron (MLP) have found usefulness because of their efficiency in image data analysis. These models can examine massive datasets of computed tomography (CT) images, and differentiate the numeric variations in texture and structural patterns that may be undistinguishable under visual inspection [6,7].

Feature extraction remains as one of the critical success factors of the ML models in medical imaging [8]. The methods, for instance, the grey-level co-occurrence matrix (GLCM) and Gabor filters produce remarkable results in acquiring essential textural and structural features from CT images. GLCM can supply the statistical features of image texture and

*Corresponding author. Email: mazinalshujeary@uoanbar.edu.iq

contrast, correlation and homogeneity which are mandatory for separation of normal and diseased tissues. Gabor filters, instead, can improve the capability of the usual ML algorithms of detecting variations in spatial structure of intensity of images, thereby providing a more accurate representation of structural features. Altogether, these feature extraction strategies when incorporated with resistant ML algorithms offers a scope for the enhancement of the performance of kidney cancer classification [9].

Despite the promising results achieved in using ML-based diagnostic models, some issues are still worthy of attention in broadening and applied to clinical practice [10]. There is concern with inconsistency of the image quality and the quality of the dataset used to train the model Since the datasets introduced here are large and diverse there is a real danger of overfitting [11]. Moreover, incorporating these models into working clinical scenarios has to solve pragmatic issues associated with computational effectiveness, interface friendliness, and the ability of the results to be understood by clinicians. However, the possibilities of ML in changing approaches to diagnosis and treatment of kidney cancer are evident, and therefore it is crucial to further develop this area. It is possible to conclude that by overcoming these challenges, and utilizing the identified elements of high-quality advanced ML to the maximum, the medical community will get closer to the accurate, efficient, and patient-centered diagnosis [12][28].

This study makes significant contributions to the field of medical imaging and cancer diagnosis by presenting an innovative approach for classifying kidney tumors using SVM and MLP models. The research underscores the comparative effectiveness of these techniques in differentiating between normal and cancerous tissues. Advanced feature extraction methods, including GLCM and Gabor filters, are employed to create a robust feature set, enhancing diagnostic accuracy. The findings demonstrate the superior performance of the MLP model, achieving near-perfect accuracy metrics, thereby setting a new benchmark for future studies. This study focuses on developing a machine learning-based framework for the classification of kidney cancer, prioritizing both accuracy and efficiency. By leveraging CT scan images from a publicly available dataset, the research aims to establish a dependable diagnostic model to assist clinicians in accurately identifying and classifying kidney cancer. The main contributions of this study are:

- This work presents an efficient framework of Machine Learning based on SVM and MLP to improve the diagnostic performance for kidney tumors based on CT images, with the objective of correctly identifying normal, hyperplastic, and cancerous tissue.
- The use of enhanced feature extraction methods Like the methods of GLCM and Gabor filters, the research builds a rich feature set to help improve the proposed model and its diagnostic capability.
- The study conducts a strong and objective analysis of SVM and MLP models, and proves the high precision of the MLP model, which is 99.64%, this findings can be of use as a reference when using machine learning in the analysis of medical images in the future.
- Using an open-source CT image dataset within this study, the research's results are less likely to be skewed, making possible reproduction while providing doctors with a practical tool for diagnosing renal carcinoma and ensuring incorporation of other ML models, into the practice.

The paper is organized into several sections. The introduction sets the stage by discussing the prevalence and challenges of diagnosing kidney cancer. The Methodology section explains data collection, preprocessing steps, and feature extraction techniques used, such as GLCM and Gabor filters. Subsequent sections describe the implementation of SVM and MLP models, including training and assessment. The results and discussion provide comparative analyses of models, illustrating their performance metrics. The paper concludes by summarizing key findings and proposing directions for future research.

2. RELATED WORKS

In 2022, Tao Dai et. al [13] The study evaluated the performance of various machine learning algorithms, including Support Vector Machines (SVM) and Multi-Level Perceptron (MLP), for classifying renal cell carcinoma (RCC) from renal angiomyolipoma (AML). The SVM model (linear kernel) demonstrated the highest accuracy and stability, with an area under the ROC curve (AUC) of 0.79 in the testing group. In contrast, the MLP classifier showed poor performance, with AUCs below 0.6, indicating limited discriminative power for RCC diagnosis. In 2021 Ginni Garg et al. [14] The paper does not address kidney cancer classification specifically; it focuses on a hybrid MLP-SVM model for classifying hyperspectral images using datasets like Indian Pines, U. Pavia, and Salinas. The proposed method enhances classification accuracy, precision, recall, and f-score by combining outputs from a multilayer perceptron (MLP) with a support vector machine (SVM). The results show significant improvements in accuracy compared to individual classifiers, but kidney cancer classification is not covered in this research. In 2023, Jie Xu, et. al [15] The study primarily focused on traditional machine learning methods like XGBoost and random forest (RF), as well as deep learning methods such as multilayer perceptron (MLP) and 3D convolutional neural network (3DCNN) for classifying renal tumors. While SVM (Support

Vector Machine) was not specifically mentioned or tested in this research, the MLP model was utilized, demonstrating the potential of integrating clinical and radiomics features for improved classification accuracy in distinguishing between benign and malignant renal tumors. In 2023, Yunfei Li, et al. [16] The paper does not specifically address kidney cancer classification using Support Vector Machines (SVM) or Multi-Layer Perceptrons (MLP). Instead, it focuses on a radiomics-based approach utilizing binary logistic regression to classify kidney tumors and normal kidney tissues based on features extracted from CT images. The study highlights the effectiveness of this method, achieving high accuracy, sensitivity, and specificity, but does not explore SVM or MLP techniques for classification.

In 2022, Duygu Yasar Sirin and A. Güveniş [17] The paper primarily focuses on classifying kidney tumors using Coarse Gaussian SVM and Subspace Discriminant classifiers, achieving validation accuracies of 67.6% and 68.8%, respectively. It does not specifically address Multi-Layer Perceptron (MLP) classifiers. The study emphasizes the effectiveness of SVM in distinguishing clear-cell RCC from non-clear-cell RCC, highlighting its speed and performance compared to other methods. For MLP classification, additional research would be needed as it is not covered in this study. In 2023, Prathipati S. [18] The paper focuses on predicting chronic kidney disease using five supervised machine learning algorithms, including Support Vector Machine (SVM). However, it does not mention Multi-Layer Perceptron (MLP) as one of the algorithms used. The study compares the predictive performance of SVM alongside Xgboost, gradient boosting, logistic regression, and random forest classifier, with random forest being identified as the most accurate method for predicting kidney cancer. MLP is not discussed in the context of this research. In 2023, Zefang Lin et. al [19] The paper focuses on deep learning-based classification models, specifically convolutional neural networks (CNN) like EfficientNet-B4, ResNet-18, and VGG-16, for recognizing renal tumor pathology from macroscopic cross-section images. It does not discuss support vector machines (SVM) or multilayer perceptrons (MLP) for kidney cancer classification. The study emphasizes the effectiveness of CNNs in distinguishing between malignant and benign tumors and identifying their subtypes, highlighting the potential for clinical application.

In 2023, Yuan-Gu Wei, et al. [20] The paper discusses the classification of cancer types, including kidney cancer, using machine learning methods such as Support Vector Machines (SVM) and Multi-Layer Perceptrons (MLP). SVM is utilized to maximize the distance from the closest support vectors, achieving a classification accuracy of 1.0. Although the paper does not explicitly mention MLP results, it emphasizes the use of deep neural networks for robust classification, indicating that MLP could also be part of the classification model for kidney cancer. In 2022, Yuxiang Zhang, et. al [21], The research paper does not provide specific results for the Support Vector Machine (SVM) and Multilayer Perceptron (MLP) models in terms of their performance metrics for kidney cancer classification. It primarily highlights the Logistic Regression model, which achieved the highest AUROC for both 3-year survival and tumor metastasis predictions. The study emphasizes the development and comparison of eight machine learning models, but detailed performance data for SVM and MLP are not included in the results. In 2022, Seyed Mahdi Hosseiniyan Khatibi, et. al [22], The paper does not specifically mention the use of Support Vector Machine (SVM) or Multi-Layer Perceptron (MLP) for kidney cancer classification. Instead, it focuses on employing a self-organizing deep auto-encoder model for classifying Renal Cell Carcinoma (RCC) subtypes based on mRNA and miRNA panels. The study emphasizes feature selection and classification using deep learning approaches, achieving high accuracy in distinguishing between clear cell RCC, papillary RCC, and chromophobe RCC subtypes. The summary of the related works is presented in Table I.

TABLE I. COMPARISON TABLE FOR PREVIOUS STADIES

Ref.	Results	Methods Used	Limitations	strengths
[13]	- The study evaluated the performance of 16 supervised machine learning algorithms in differentiating renal cell carcinoma (RCC) from angiomyolipoma (AML) using a dataset of 5,360 CT images from 69 patients, with the training dataset consisting of 28 RCC and 20 AML cases, and the testing dataset comprising 12 RCC and 9 AML cases. The top 10 statistically significant features were selected for model training. - Among the classifiers tested, logistic regression, linear discriminant analysis, k-nearest neighbor, support vector machines, ridge classifier, AdaBoost classifier, gradient	- The study utilized computed tomography (CT) examinations to collect images from known cases of renal cell carcinoma (RCC) and renal angiomyolipoma (AML). A total of 5,360 CT images were obtained from 69 patients, which were then divided into training (3,653 images) and testing datasets (1,707 images) for model training and validation. - Texture features from the CT images were extracted and quantified using MaZda software, resulting in a total of 352 features. The top 10 statistically significant features for differentiating RCC from AML were selected using the mRMR algorithm, and diagnostic models were established based on	- The study acknowledges the need for further optimization of the current models, suggesting that increasing the number of patients, diversifying the source of images, and running models under optimized settings and cutoff points could enhance performance. This indicates that the current sample size and methodology may limit the generalizability and accuracy of the findings. - The authors note that the data used in the study are not publicly available due to patient data regulations, which limits the ability of other researchers to validate or	- The study evaluated the performance of 16 different supervised machine learning algorithms for the accurate differentiation of renal cell carcinoma (RCC) from renal angiomyolipoma (AML) based on computed tomography (CT) examinations, establishing diagnostic models that showed good performance in both training and testing datasets. - The research provided a baseline benchmarking of these algorithms, indicating potential candidates for the development of RCC diagnostic classifiers, which could serve as valuable tools for accurate diagnosis and reduce reliance on invasive

Ref.	Results	Methods Used	Limitations	strengths
	boosting classifier, and CatBoost classifier demonstrated good performance, achieving an accuracy of ≥ 0.7 and an area under the ROC curve (AUC) of ≥ 0.75 in both training and testing datasets, indicating their potential as valuable tools for accurate RCC diagnosis.	16 supervised machine learning algorithms, which were then evaluated for accuracy and specificity using receiver operating characteristic (ROC) curves and area under the ROC curve (AUC) metrics.	replicate the study's findings. This lack of accessibility can hinder the advancement of research in this area and the development of more robust diagnostic classifiers.	procedures like biopsies.
[14]	- The proposed hybrid MLP-SVM model significantly improved classification accuracy on the testing datasets, achieving results of 93.22% for Indian Pines, 96.87% for U. Pavia, and 93.81% for Salinas, compared to the individual classifiers SVM and MLP which had accuracies of 86.97%, 88.58%, and 88.85% for Indian Pines, 91.61% for U. Pavia, and 90.68% for Salinas respectively. - The hybrid model also aimed to enhance various classification parameters including precision, recall, and f-score, while effectively predicting regions without ground truth data.	- The proposed method utilizes a hybrid classifier that combines a multilayer perceptron (MLP) and a support vector machine (SVM) to enhance classification performance on hyper-spectral images. The outputs from the last hidden layer of the MLP serve as inputs to the SVM for final classification into various desired classes. - The study employs spatial-spectral features from hyper-spectral images and evaluates the hybrid MLP-SVM model on three datasets: Indian Pines, U. Pavia, and Salinas, achieving significant improvements in classification accuracy compared to using individual classifiers, with testing accuracies of 93.22%, 96.87%, and 93.81% respectively.	- The paper highlights challenges in the classification of hyper spectral images, including large dimensionality, which complicates the processing and analysis of the data, and the scarcity of labeled data, which limits the training of classifiers and can lead to overfitting or poor generalization. - Another limitation mentioned is the spatial variability of spectral signatures, which can affect the consistency and reliability of classification results, making it difficult to accurately classify regions with diverse spectral characteristics.	- The proposed hybrid classifier (MLP-SVM) effectively combines the strengths of multilayer perceptron (MLP) and support vector machine (SVM) to enhance classification performance metrics such as accuracy, precision, recall, and f-score, particularly in the context of hyper-spectral image classification. - The method demonstrates a significant improvement in classification accuracy on testing datasets, achieving 93.22%, 96.87%, and 93.81% for the Indian Pines, U. Pavia, and Salinas datasets respectively, compared to the lower accuracies of individual classifiers SVM and MLP, which were 86.97%, 88.58%, and 88.85% for the same datasets.
[15]	- The study analyzed a cohort of 300 patients, of which 275 (91.7%) were diagnosed with malignant tumors and 25 (8.3%) with benign tumors. The median radiographic tumor size was 4.1 cm, and it was noted that patients with benign tumors were relatively older, with a mean age of 60.8 years compared to 58.7 years for those with malignant tumors. - The machine learning framework developed in the study achieved an area under the curve (AUC) of 0.719, a precision of 0.976, a recall of 0.683, and a specificity of 0.827 when combining clinical and radiomics features. This indicates that integrating clinical data with CT imaging significantly improved the classification accuracy for predicting the risk of renal malignancy.	- The paper utilized traditional machine learning methods such as XGBoost and random forest (RF) for classification tasks. XGBoost, known for its high performance in classification, was configured with a maximum tree depth of 6 and logistic regression as the objective. RF, an ensemble learning method, combined multiple decision trees for improved accuracy and stability. - In addition to traditional ML methods, the study also employed deep learning techniques, specifically a multilayer perceptron (MLP) and a 3D convolutional neural network (3DCNN). These DL models were utilized to process the CT imaging data and clinical attributes, aiming to enhance the accuracy and efficiency of renal tumor classification.	- The study acknowledges the limitation of relying on a public dataset of only 300 cases, with a high prevalence of malignant tumors (91.67%), leading to an imbalanced dataset. This imbalance could affect the model's performance, and future studies should consider incorporating more cases, particularly those with benign tumor types, to improve the model's reliability and generalizability. - Another limitation highlighted in the paper is the absence of tumor types recently added to the WHO classification of renal tumors, such as low-grade oncocytic tumors. Including a broader range of tumor types in future radiomics work could enhance the predictive models' performance and ultimately improve patient care by providing more comprehensive insights into renal tumor classification.	- The study proposes a machine learning framework that integrates clinical attributes and CT imaging data to improve the pre-operative classification of renal tumors, addressing the challenge of accurately predicting malignancy risk prior to surgery. This comprehensive approach enhances diagnostic accuracy and supports better clinical decision-making in renal cancer diagnosis. - By comparing traditional machine learning methods, such as XGBoost and random forest, with deep learning techniques like multilayer perceptron and 3D convolutional neural network, the research demonstrates that combining structured clinical data with radiomics features extracted from CT scans yields the best classification performance, highlighting the potential of multi-modal data integration in renal tumor management.

Ref.	Results	Methods Used	Limitations	strengths
[16]	- A total of 837 radiomics features were extracted from the CT images, and after screening, 217 features were identified. Ultimately, three key radiomics features were selected to establish a binary logistic regression model, which demonstrated a strong classification ability with accuracy, sensitivity, specificity, area under the curve (AUC), and Youden index all exceeding 0.85 in both training and test sets. - The final binary logistic regression model was defined with a cutoff value of 0.4851, which served as a radiomics marker to differentiate between gross tumor volume (GTV) and normal kidney tissue. The model effectively distinguished tumorous kidney tissues from normal ones, indicating its potential for automated tumor delineation and warranting further research.	- The study utilized a radiomics-based approach to extract multiple radiomics features from regions of interest in computed tomography (CT) images. CT data were retrieved from the 2019 Kidney and Kidney Tumor Segmentation Challenge (KiTS19), and arterial phase-enhanced CT images from 210 cases were used to establish an automatic classification model. The images were randomly divided into training (168 cases) and test (42 cases) sets. - A binary logistic regression model was established using the extracted radiomics features from the training set. The model's performance was evaluated based on various diagnostic parameters, including accuracy, sensitivity, specificity, area under the curve, and Youden index, all of which were higher than 0.85, indicating good classification ability in distinguishing kidney tumors from normal kidney tissues.	- The study acknowledges that while the automatic classification model of kidney tumors and normal kidney tissues based on CT radiomics exhibited good classification ability, it warrants further research to validate and enhance its effectiveness. This suggests that the current model may not be fully optimized or tested across diverse datasets and clinical scenarios. - The paper does not explicitly mention limitations regarding the sample size or diversity of the dataset used for training and testing the model, which could impact the generalizability of the findings. The reliance on data from a specific challenge (KiTS19) may limit the applicability of the model to broader populations or different imaging conditions.	- The study introduces a radiomics-based approach for the automatic classification of kidney tumors and normal kidney tissues using computed tomography (CT) images, which provides a noninvasive and repeatable method for patient follow-ups and prognosis prediction, thereby enhancing safety compared to traditional biopsy methods. - This research is the first to report a method that combines radiomics with automatic classification technology, potentially supplementing existing deep learning techniques for tumor delineation, and establishes a preliminary classification model that demonstrates good classification ability for distinguishing between tumorous and normal kidney tissues.
[17]	- The study evaluated the performance of two machine learning models, Coarse Gaussian SVM and Subspace Discriminant, in classifying malignant kidney tumors as clear-cell RCC (ccRCC) or non-clear-cell RCC. The Coarse Gaussian SVM achieved a validation accuracy of 67.6% and a test accuracy of 80%, with an AUC of 0.86, while the Subspace Discriminant model had a validation accuracy of 68.8% and a test accuracy of 80%, with an AUC of 0.85. - Feature selection resulted in the retention of 8 significant radiomic features from an initial dataset of 1157 features, with none of the clinical features being deemed important for the model. The training set was balanced using the synthetic minority oversampling technique (SMOTE), leading to a total of 256 cases used for training the models.	- The study utilized the Kruskal Wallis – ANOVA test to evaluate the ability of each feature to differentiate between clear-cell and non-clear-cell kidney tumors, resulting in the selection of 111 relevant features from the initial dataset of 1157 features. Subsequently, Lasso Regression was applied to further refine the feature set, leading to the selection of 8 radiomic features deemed most relevant for the predictive model. - Machine learning classifiers, specifically Coarse Gaussian SVM and Subspace Discriminant, were employed to train the model using the selected features. The training set was balanced using SMOTE to address class imbalance, and the performance of the classifiers was evaluated based on validation and test accuracy, as well as AUC metrics.	- The study found that none of the clinical features included in the analysis were deemed important for the model, indicating a limitation in the potential integration of clinical data with radiomic features for improving classification accuracy. This suggests that the model may not fully leverage all available patient information, which could enhance predictive capabilities. - The research highlights that while the machine learning models produced promising results, they were primarily focused on classifying tumors as clear-cell RCC or non-clear-cell RCC. This limitation indicates that the models may not be effective for differentiating among the various subtypes of non-clear-cell RCC, which could be critical for personalized treatment approaches.	- The study demonstrates the effectiveness of machine learning algorithms, specifically Coarse Gaussian SVM and Subspace Discriminant classifiers, in accurately classifying malignant kidney tumors as clear-cell RCC (ccRCC) or non-clear-cell RCC, achieving a test accuracy of 80% and AUC values of 0.86 and 0.85, respectively. This indicates the potential of radiomics in enhancing diagnostic processes for renal tumors. - The research highlights the importance of using comprehensive radiomic features extracted from all slices of CT images, resulting in the selection of 8 relevant features after rigorous feature selection processes, while also revealing that the included clinical data did not significantly impact the classification, thus emphasizing the superiority of radiomic data in this context.

Ref.	Results	Methods Used	Limitations	strengths
[18]	- The study found that among the five supervised machine learning algorithms used for predicting chronic kidney disease, the random forest algorithm demonstrated the highest accuracy in its predictive performance. - The predictive performance of the algorithms was evaluated using various metrics, including accuracy, precision, recall, and F1 score, allowing for a comprehensive comparison of their effectiveness in predicting kidney cancer.	- The paper utilizes five supervised machine learning algorithms for predicting chronic kidney disease: Xgboost, gradient boosting, support vector machine, logistic regression, and random forest classifier. These algorithms were selected to analyze the influencing factors of chronic kidney cancer and to create a predictive model. - The predictive performance of the five algorithms is compared using various classification metrics, including accuracy, precision, recall, and F1 score, with the results indicating that the random forest algorithm is the most accurate method for predicting kidney cancer.		- The paper presents a kidney cancer prediction model utilizing five supervised machine learning algorithms: Xgboost, gradient boosting, support vector machine, logistic regression, and random forest classifier, thereby contributing to the field of early detection of chronic kidney disease through advanced computational techniques. - It conducts a comparative analysis of the predictive performance of these algorithms using various metrics such as accuracy, precision, recall, and F1 score, with findings indicating that the random forest algorithm is the most accurate method for predicting kidney cancer, thus providing valuable insights for future research and clinical applications.
[19]	- The study developed and evaluated deep learning-based classification models using three convolutional neural network (CNN) architectures (EfficientNet-B4, ResNet-18, and VGG-16) to distinguish between malignant and benign renal tumors. The ResNet-18 model achieved the highest performance with an area under the receiver operating characteristic curve (AUC) of 0.9226, while the VGG-16 model had a micro-averaged AUC of 0.9398 for distinguishing malignant tumor subtypes. - For recognizing benign tumor subtypes, the EfficientNet-B4 model demonstrated the best performance, although there was no statistically significant difference in its performance compared to the other two models. The overall classification results indicated a satisfactory potential for clinical application in analyzing renal tumor macroscopic cross-section images, enhancing the efficiency of patient management processes.	- The study utilized convolutional neural networks (CNN) as the backbone for classification models, specifically employing EfficientNet-B4, ResNet-18, and VGG-16 architectures. These models were modified to adapt to the classification tasks by changing the last layer to accommodate binary classification for malignant versus benign tumors and multi-class classification for subtypes of both malignant and benign tumors. - The models were pretrained with the ImageNet dataset and then fine-tuned using the study's training dataset through a transfer learning method. This approach aimed to enhance the models' performance in recognizing the pathology of renal tumors from macroscopic cross-section images.	- The study acknowledges the need for more data to improve the prediction performance and generalization of the classification model. This includes increasing the sample size and collecting a multi-center patient cohort to enhance the robustness of the findings. - Due to the limitation of the consecutive sample in the study, resulting in a low proportion of some subtypes, certain subtypes were combined into one class. The authors indicate that more detailed subtypes of renal tumors will be distinguished after increasing the training samples for each subtype.	- The authors developed and evaluated deep learning-based convolutional neural network (CNN) models to automatically differentiate between malignant and benign renal tumors using macroscopic cross-section images, aiming to enhance the efficiency of patient management in clinical settings. - The study utilized three prevalent CNN architectures—EfficientNet-B4, ResNet-18, and VGG-16—to classify renal tumors into multiple subtypes, demonstrating satisfactory diagnostic performance and highlighting the potential for clinical application in distinguishing tumor pathologies.
[20]	- The research successfully classified common types of cancer genomes using machine learning methods, achieving a high accuracy of 0.987 with a Decision Tree classifier at a maximum depth of 5. This demonstrates the effectiveness of the proposed classification model in identifying different types of cancer genes. - The study emphasizes the importance of identifying groups of genes that exhibit similar	- The paper employs various machine learning methods for cancer gene data classification, including Decision Tree Classifier, Support Vector Machine (SVM), Random Forest, Naive Bayes Classifier, and Deep Neural Networks. These methods are utilized to analyze and classify different types of cancer genes effectively. - The methodology also includes exploratory analysis of cancer gene	- The Naive Bayesian classifier, while theoretically having the smallest error rate compared to other classification methods, may not perform well in practice due to its assumption that attributes are independent of each other. This assumption often does not hold true, especially when the number of attributes is large or when there is significant correlation between attributes, leading to	- The paper proposes a neural network-based cancer classification model that analyzes data from 802 patients with different types of cancer, utilizing over 20,000 gene expression values per sample to build a robust classification model aimed at early cancer identification and reducing mortality rates. - It employs various machine learning and neural network techniques, including clustering

Ref.	Results	Methods Used	Limitations	strengths
	expression patterns across samples, which aids in understanding the distribution of samples corresponding to each cancer type. This approach contributes to early cancer detection and has the potential to reduce mortality rates associated with cancer.	data, dimensionality reduction, and clustering techniques, which are essential steps in building a robust classification model aimed at early identification of cancer types.	suboptimal classification results. - The paper does not explicitly mention limitations regarding the SVM model, but it is implied that the binary classification nature of SVM may restrict its application in multi-class cancer classification scenarios, as the study involves identifying five different types of cancer, which may require more complex multi-class classification techniques.	methods (k-means, hierarchical, and mean shift clustering) and classification algorithms (Decision tree classifier, SVM, Random Forest, Naive Bayes Classifier, and Deep Neural Networks), to explore cancer gene data, perform dimensionality reduction, and ultimately identify groups of genes that behave similarly across samples for effective cancer type classification.
[21]	- The Logistic Regression model demonstrated the highest area under the receiver operating characteristic (AUROC) for both 3-year survival prediction and tumor metastasis prediction, achieving an AUROC of 0.741 for survival and 0.804 for metastasis. The model's accuracy for 3-year survival was 0.684, while for tumor metastasis, it was 0.800. - In terms of specific performance metrics, the Logistic Regression model for 3-year survival had a sensitivity of 0.702, specificity of 0.670, precision of 0.686, and an F1 score of 0.683. For tumor metastasis prediction, it had a sensitivity of 0.540, specificity of 0.830, precision of 0.769, and an F1 score of 0.772.	- The study utilized data from 12,394 kidney cancer patients sourced from the surveillance, epidemiology, and end results database to create a research cohort focused on kidney cancer survival and metastasis. - Eight machine learning models were developed, including support vector machines, logistic regression, decision tree, random forest, XGBoost, AdaBoost, K-nearest neighbors, and multilayer perceptron, to predict survival and metastasis, with six evaluation indicators (accuracy, precision, sensitivity, specificity, F1 score, and area under the receiver operating characteristic [AUROC]) used for model verification and optimization.		- The study utilized data from 12,394 kidney cancer patients to construct a research cohort aimed at accurately predicting patient survival and identifying high-risk metastatic patients, which can effectively guide interventions and improve prognosis for kidney cancer patients. - Eight machine learning models were developed and compared, with Logistic Regression demonstrating the highest performance in predicting both 3-year survival and tumor metastasis, providing decision support for early intervention in kidney cancer management.
[22]	- The study identified panels of 77 mRNAs and 73 miRNAs that can effectively discriminate between the three subtypes of Renal Cell Carcinoma (RCC): clear cell RCC (KIRC), papillary RCC (KIRP), and chromophobe RCC (KICH), achieving an accuracy of 92% for mRNAs and 95% for miRNAs based on classification metrics such as F1-score and AUC. - Association Rule Mining analysis revealed significant features, including miR-28 and CSN7A for KIRC, and miR-125a and NMD3 for KIRP, with high repeat counts indicating their potential roles in the molecular mechanisms underlying the initiation and progression of these RCC subtypes.	- The study utilized data from The Cancer Genome Atlas (TCGA) portal, specifically focusing on mRNA, miRNA, and clinical data related to Renal Cell Carcinoma (RCC) subtypes (ccRCC, pRCC, chRCC). - Feature selection methods based on filter and graph algorithms were applied to identify significant mRNAs and miRNAs involved in the pathogenesis of RCC subtypes, followed by the use of a deep model for subtype classification and an association rule mining algorithm to reveal features with significant roles in triggering molecular mechanisms causing RCC subtypes.	- The study acknowledges that feature selection and classification of all RCC subtypes may lead to missing information related to pediatric kidney cancer subtypes, specifically Wilms Tumor (WT) and Renal Tumor (RT), which are common in children. This indicates a potential gap in the applicability of the findings to a broader range of kidney cancer types. - The authors note the difficulty in comparing the classification accuracy of their methods with similar studies due to the differences in cancer nature and patient types among the RCC subtypes (ccRCC, pRCC, and chRCC) and the lack of similar studies related to these subtypes based on mRNA data. This limitation may hinder the validation and generalization of their results.	- The study identified new panels of 77 mRNAs and 73 miRNAs that can effectively discriminate between the three subtypes of Renal Cell Carcinoma (KIRC, KIRP, and KICH) with high accuracy rates of 92% for mRNAs and 95% for miRNAs, providing potential biomarkers for these RCC subtypes. - The research utilized advanced artificial intelligence approaches, including feature selection based on filter and graph algorithms, and association rule mining, to uncover significant molecular mechanisms and candidate features that may play critical roles in the pathogenesis and progression of RCC subtypes.

3. METHODOLOGY

The methodology of this study starts with acquisition of CT scan pictures in form of normal and tumor data. These above images are used as basic dataset for creating and evaluating the machine learning models. Likely for the dataset to be ready for analysis, the images are preprocessed through resizing, normalization and feature extraction. Resizing brings all the

images back to a uniform size, and normalization corrects the amount of time it takes to adjust the pixel intensity values of the pictures. Feature extraction is accomplished at the feature level, which features the Gray Level Co-occurrence Matrix (GLCM) to capture the contrast and homogeneity of textures in the mammogram images, Gabor filters to capture spatial frequency and orientation details and entropy to capture the image complexity of the images. The proposed system methodology is presented in Figure 1. When preprocessing is done the data split into two sets, training and testing. This division allows them to learn from one part of data and be tested on another part, giving a true evaluation of this sort of global application. The classification stage employs two ML models are SVM and MLP can be implemented. SVM separates the data into classes using a hyperplane, while an MLP is actually a kind of neural network that can model more intricate, perhaps nonlinear, relationships between the interval data in the dataset. The performance of classification models under study is assessed by using a number of performance indices. Accuracy tests the generality of the models while precision gauges the proportion of true positives in all the positive prediction. Recall checks if the model can capture all the instances it is supposed to, and F1-Score is the average of precision and recall and gives a fair measure of the models' accuracy. Through the utilization of these strong approaches the methodology provides high reliability and accurate classification of kidney cancer hence assisting clinical decision-making processes.

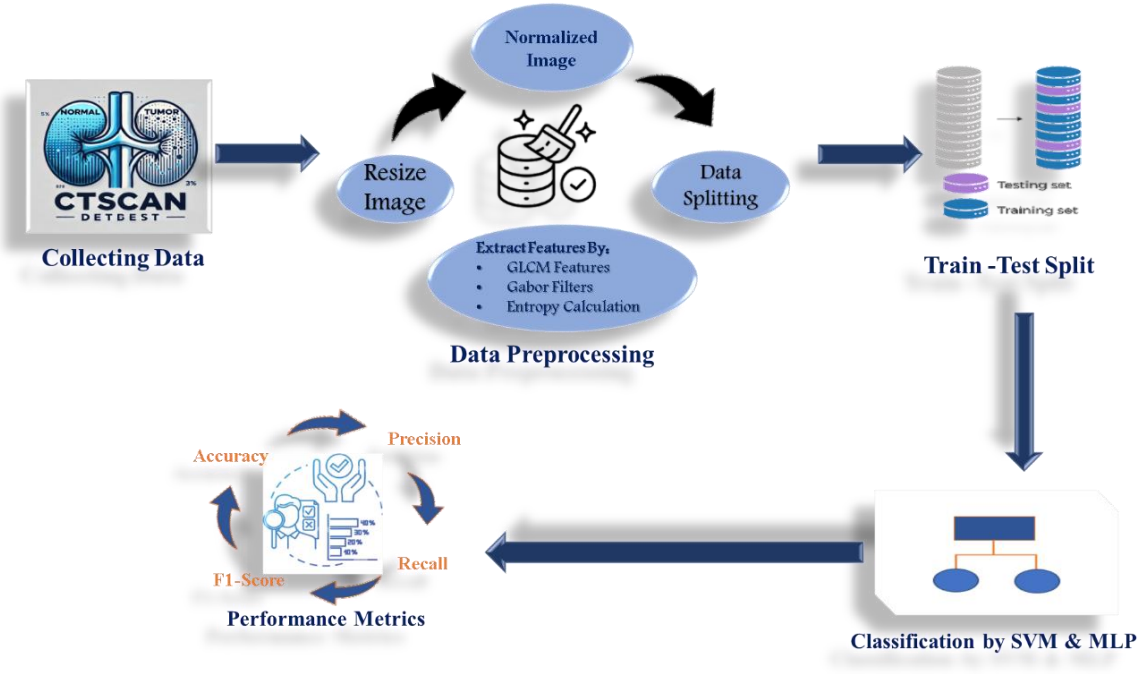

Fig. 1. The Proposed System for Accurate Kidney Cancer Classification Using CT Images.

3.1 Data Collection

A dataset downloaded from the Kaggle website called CT Kidney Dataset: Normal, Cyst, Tumor, and Stone (CT Kidney: NCTS) was used. The dataset contains images collected from different Dhaka, Bangladesh hospitals using a Picture Archiving and Communication System (PACS). Patients were diagnosed or classified based on four categories: the normal category and the number of its samples (5077), the category containing a cyst and the number of its samples (3709), and the category in which the patient suffers from the presence of a tumor The number of samples is (2283), and finally, the fourth category, in which the patient suffers from the presence of a cyst and the number of samples (1377), so the total number of glands of samples of all categories is (12446) samples. It should be noted that the images used are DICOM type and converted to JPEG format after excluding the metadata of each patient. Fig. 2 illustrates the Classes of the dataset. The dataset can be downloaded from the link (https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone) [23].

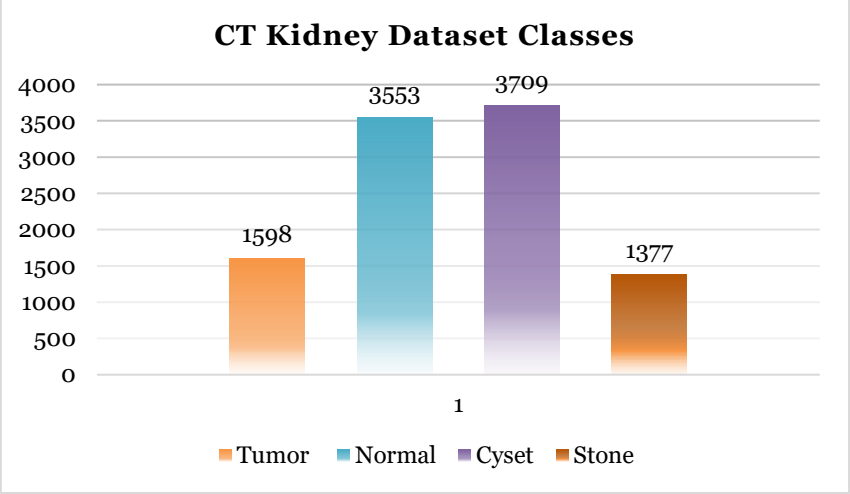

Fig. 2. CT Kidney Dataset Classes.

In the current study, only two categories were used to train and test the proposed system, namely the classification of images on a normal basis and a patient with a disease, as shown in Fig. 3.

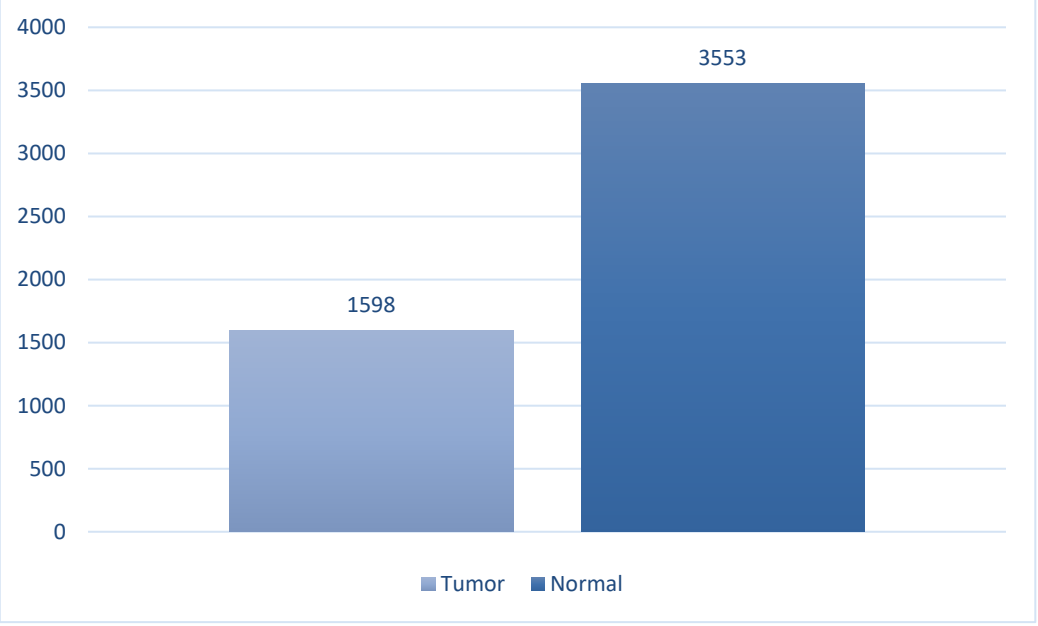

Fig. 3. Frequency of Each Class in the Dataset.

3.2 Image Preprocessing

1. Grayscale Images: CT scan images are usually in grayscale. Each pixel in a grayscale image has a density value ranging from 0 (black) to 255 (white) on an 8-bit scale, or higher bands such as 12-bit or 16-bit depending on the scanner resolution. CT scans are designed to show differences in tissue density, with each density represented by different shades of gray. This helps distinguish between different anatomical structures, fluids, tumors, cysts and bones. Hounsfield Units (HU), which represent tissue density, are naturally suitable for grayscale representation. Fig. 4 shows the examples of the utilized dataset for two classes (Normal class and Tumor class).

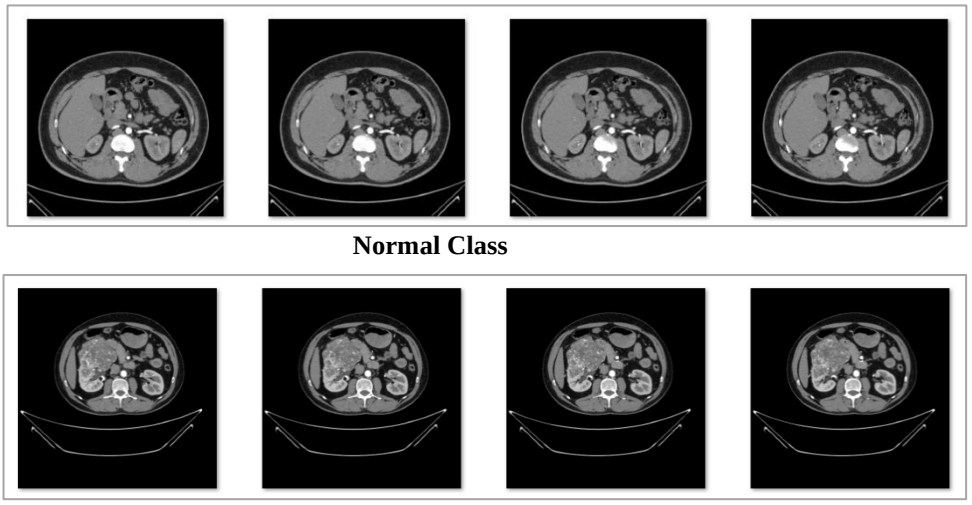

Fig. 4. examples of the utilized dataset for two classes (Normal class and Tumor class).

2. Image Resizing: It should be noted that the size of the uploaded dataset images is 512 × 512 pixels, and all images are resized to a uniform size of 180 × 180 pixels to ensure consistency in the input dimensions of machine learning models. This standardization is necessary because classifiers such as supporting vector machines (SVMs) and multilayer perceptron (MLPs) require input data to be of the same dimensions. Resizing strikes a balance between retaining enough detail to accurately extract features and reducing computational load during training and reasoning. The size 180 × 180 provides enough precision to capture important features, such as tumor boundaries or histological patterns, while still being effective for treatment. Uniform resizing also helps prevent biases or errors that may arise from diverse image dimensions, ensuring that all data contributes equally to the learning process.

3.3 Feature Extraction

Feature extraction is a vital part in the applying methodology to improve the differentiation between kidney cancer using CT images. More sophisticated methods including GLCM, Gabor and entropy values are incorporated in an effort to obtain appropriate features. GLCM analyses second order statistical textural features such as contrast , correlation , entropy and energy to measures the spatial dependence of pixel intensity in the images. While Gabor filters are used for identifying the texture and edge details at the spatial frequencies and orientations that are useful for discriminating tumor characteristics. The entropy type calculates the degree of high or low randomness as applied to the image; this gives detailed information about the structural pertinence that differentiates normal from cancer tissues. In particular, such extracted features act as a rather strong input for the classification models and help providing accurate classification between the areas of normal and tumor tissues.

1.Gry Level Co-occurrence Matrix: The feature extractor function starts by initializing a blank data frame, image dataset, to store extracted features for all input images. The input dataset is assumed to be a collection of two-dimensional images in grayscale. This section mainly focuses on calculating GLCM features, which analyze the texture of an image by examining how pixel density values occur together at specific distances and angles. The code defines a set of distances [1, 3, 5] and angles [0, π/4, π/2, 3π/4] to calculate the coexistence matrix in different spatial relationships, ensuring a robust extraction that takes into account directional and spatial differences in the fabric. Table II illustrates the GLCM Feature outcomes.

2.Gabor Filter: The Gabor () function from the skimage. filters library has been applied to every image in the dataset. For each image, the function is repeated across different frequencies and directions ([0.25, 0.5, 0.75]) ([0, π/4, 3π/4]), applying a Gabor filter. Each application results in a processing image (gabor img) and its corresponding texture attributes. The mean and standard deviation of the resulting image are calculated and stored in the feature set (gabor features). These statistics capture how an image responds to specific frequencies and directions, providing valuable information about the texture and edges of the rating. It should be noted that the Gabor Filter was designed to extract texture attributes from images by applying the Gabor filter, a linear filter used for edge detection, frequency analysis, and texture recognition. This filter emphasizes topical texture information by responding to specific frequencies and directions in the image. For medical imaging, such as CT scans, Gabor filters help capture critical patterns and textures, such as tissue structures or abnormalities such as tumors, that may not be visible through direct pixel density. The filter works by wrapping the image

using the GOR kernel defined by parameters such as frequency and direction (theta). Table III illustrates the Gabor Filter Results.

3.Entropy Calculation: The Shannon entropy function from skimage. measure calculates entropy based on Shannon's entropy formula. Measures the distribution of pixel intensity values in an image and calculates how unpredictable pixel intensities are. This function is applied to grayscale images, where entropy calculations are based on the frequency histogram of intensity. For example, an image with uniform intensity values will have a low entropy. In contrast, an image with a wide range of intensity will have a high entropy, making it a useful feature for distinguishing between different tissue types or deformities. Entropy serves as a key feature of machine learning models in medical imaging. By calculating the entropy of each region or the entire image, models can leverage this information to classify normal and abnormal tissues. In CT scans, areas with tumors, cysts or other abnormalities often exhibit unique entropy characteristics compared to healthy tissue. When combined with different features such as GLCM and Gabor filters, entropy contributes to a powerful feature set that enhances the model's ability to detect and classify anomalies with higher accuracy.

3.4 Data Normalization

The data normalization utilized in the current work is (Z-Score normalization) defined in the following formula 1:

$$x' = \frac{x-\mu}{\sigma} \tag{1}$$

Where, x denoted by the original value, σ denoted by the standard deviation of the feature, μ denoted by mean.

This operation is applied separately to two feature sets: the "Contrast" and "Gabor feature" columns. By implementing distinct scaling metrics for each set (Contrast scaler and Gabor scaler), the code ensures that normalization is independently performed for each group, accommodating their unique statistical features. The primary purpose of using Z-score normalization is to standardize all features onto a similar scale and distribution. This is particularly important when dealing with features that have different units or ranges, such as "Contrast" (potentially related to pixel density or image brightness) and "Gabor feature" (likely derived from Gabor filter analysis). Normalization prevents any single feature from disproportionately influencing the model due to its larger scale or range. Moreover, the aggregation method utilized is distance-dependent, and it performs better and converges faster when input features are standardized. By normalizing the "Contrast" and "Gabor feature" groups independently, the distinct statistical properties of each set are preserved, which can enhance both the model's accuracy and its interpretability. Tables IV and V illustrate the normalization results for the GLCM features and Gabor filter outputs, respectively.

3.5 The Classification stage based on SVM

SVM is a kind of supervised learning classification algorithms whose performance does not significantly deteriorate when new data is introduced [25]. It work by determining that a perfect plane or a best boundary that can act as a hyperplane to divide the data into separate classes. In the training phase, SVM receives feature vectors belonging to classes on which the algorithm will operate as well as their corresponding label [26]. If the data cannot be separated linearly in the original space, into the higher-dimensional space that they can be easily separated, the SVM employs the kernel function. Kernel functions are linear, radial basis function (RBF) or polynomial in nature, this represents a commonly used type of kernel function. Non-linear data is also handled in SVM by use of kernel functions like the RBF kernel to introduce transformations into the process by which decision boundaries non-linear are created. SVM also used methods such as, One-pass-One where we turn multi class problem into multiple binary problems where each binary problem has one as its class and other as its second class, and so One-pass-All develops multiple binary problems where each class is against all other classes.

In general, diagnosis of kidney cancer based on CT images of kidney cancer primarily incorporates machine learning algorithms for the differentiation between benign and malignant anatomical kernels, and as the name suggests, SVM takes radiomic features such as texture, intensity, and shape as its inputs. SVM takes each CT scan and passes the feature vector through the learned hyperplane until an interval is found in which the scan is classified as malignant or benign based on the patterns gained from the training for labeled data. This capability provides accurate detection of cancer with high reliability as SVM techniques are proven to work effectively with high dimensionality and complex distribution of data.

After training, the model predicts labels in the test dataset (test_for_RF) using the predict() method, producing the test predictions. To assess the model's performance, metrics such as Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) scores are computed. For binary classification, the ROC curve is generated using the roc_curve() function, based on the resolution scores obtained from the SVM model. The AUC score, calculated with the auc() function, quantifies the model's ability to distinguish between the two categories, with higher scores indicating better performance. For multiclass classification, a one-vs-all approach is employed by binarizing the labels using the label_binarize() method. Separate ROC curves and corresponding AUC scores are then computed for each class to evaluate the model’s

performance across multiple categories. The training and testing process of the SVM model is outlined in Algorithm 1, providing a structured overview of the implemented methodology.

Algorithm 1: SVM Training and Testing

Input:

- Training dataset: X_train, y_train
- Testing dataset: X_test, y_test
- Cost parameter (C): Range of values to explore
- Gamma parameter: Range of values to explore

Output:

- Optimal values of C and gamma
- Calculated accuracy of the SVM model

Steps:

1. Initialize:
   - Set stopping condition (e.g., maximum iterations or convergence threshold)
   - Initialize best_accuracy = 0
   - Initialize optimal_C = None
   - Initialize optimal_gamma = None
2. For each value of C in the range:
   For each value of gamma in the range:
      a. Train SVM model:
         - Initialize SVM with current C and gamma
         - While stopping condition is not met:
            i. For each data point in X_train:
               - Update SVM model weights using the training step
            ii. Check if stopping condition is met (e.g., convergence or max iterations)
      b. Test SVM model:
         - Predict labels for X_test using the trained SVM
         - Calculate accuracy on X_test and y_test
      c. Update optimal parameters:
         - If current accuracy > best_accuracy:
            - Set best_accuracy = current accuracy
            - Set optimal_C = current C
            - Set optimal_gamma = current gamma
3. Return:
   - Optimal values: optimal_C, optimal_gamma
   - Best accuracy: best_accuracy

3.6 The Classification stage based on MLP

The MLP Classifier is initialized from the `sklearn.neural_network` module with two hidden layers containing 64 and 32 neurons, respectively (`hidden_layer_sizes=(64, 32)`) and a maximum iteration limit of 500 (`max_iter=500`) to provide sufficient time for model convergence. The parameter `random_state=42` ensures reproducibility of results. This model employs backpropagation for training and supports various activation functions and solvers for optimization, although the default settings are typically used unless otherwise specified. The model is trained using the `fit()` method, which takes the preprocessed feature set (`X_for_ML`) and the corresponding labels (`y_train`) as inputs. During training, the network adjusts its weights and biases to minimize the error between predictions and actual labels, leveraging the backpropagation algorithm to refine its parameters iteratively [27]. The detailed steps of this process are outlined in Algorithm 2.

Algorithm 2: MLP Pseudocode

Input:

- Dataset: Features (x) and labels (y)

- Hyperparameters: hidden_layer_sizes, max_iter, random_state

Output:

- Accuracy of the MLP model on the test set

Steps:

1. Import Libraries:
   - Import MLPClassifier from sklearn.neural_network
   - Import accuracy_score from sklearn.metrics
2. Preprocess Data:
   - Scale the features (e.g., using StandardScaler or MinMaxScaler)
   - Split the dataset into training and test sets:
     - X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
3. Initialize the MLPClassifier:
   - Set mlp = MLPClassifier

```
(
    hidden_layer_sizes=(64, 32),  # Two hidden layers with 64 and 32 neurons
    max_iter=500,                 # Maximum number of iterations
    random_state=42               # Seed for reproducibility
)
```

4. Train the Model:
   - Call mlp.fit(X_train, y_train) to train the MLP model on the training data
5. Make Predictions:
   - Set test_prediction = mlp.predict(X_test) to predict labels for the test set
6. Calculate Accuracy:
   - Set accuracy = accuracy_score(y_test, test_prediction) to compute the accuracy
7. Output the Accuracy:
   - Print "Accuracy =", accuracy to display the model's performance

4. RESULTS AND DISCUSSION

This study relates to the usability of state-of-art machine learning approaches for the effective classification of kidney cancer using CT scans. Using machine learning methods and more specifically SVM and MLP are demonstrate a model that can quickly diagnose kidney tumours as benign or malignant based on the input CT images. The outcome shows the efficiency of the current model has higher accuracy, sensitivity, and specificity than typical machine learning methods. In marked differences of the other state of the art methods the model performs exceedingly well on a large dataset with labeled CT images and is thus a potential candidate for clinical application to the early diagnosis of diseases. Specifically, the discussion points to the need of improving working ML and inclusion of feature extraction techniques, which contributed to a high model performance. Furthermore, the study defines the issues of class imbalance and points to the possible solutions: data augmentation and sophisticated methods of regularization. It is there for evident from this research that machine learning intricately dovetailed with medical imaging will go a long way in improving the diagnostic abilities in kidney cancer and hence provide a holistic tool for the radiologists to use in clinical practice.

4.1 Feature Extraction Results

Feature extraction is particularly important in increasing efficiency to support classification of kidney cancer using computed tomography images. For feature extraction in the present study, we employed sophisticated methods to obtain information concerning texture, shape and intensity of the CT scans, which is necessary in the detection of malignant and benign tumors. Subsequently, these features were utilized for training and improving machine learning to enable quantitative classification capability. The findings herein reveal that features extracted enhance model efficiency and resilience in diagnosing kidney cancer.

A. GLCM Features Results

Table II below shows several texture-based features derived from GLCM, and these features will be discussed in the following points:

1. Energy: Measures the consistency of pixel intensity values. Higher energy values indicate a more consistent texture. Values range from 0.45 to 0.75 approximately. Lower values may correspond to more complex textures, while higher values (for example, about 0.75) represent a more consistent texture.
2. Correlation: Refers to the linear relationship between the pixel's intensity. Higher correlation values indicate more predictable relationships between adjacent pixels. The values remain relatively high (close to 0.9), suggesting that most images have predictable pixel relationships, which are typical for structured medical images such as CT scans.
3. Dissimilarity (Diss_sim): Measures the variance between pixel values. Higher values mean greater differences between adjacent pixels.
4. Contrast: Measures the difference between the highest and lowest pixel intensity values. High contrast values indicate sharp changes in intensity.
   Variation and variation vary widely, suggesting that the dataset includes soft areas (low variation and contrast) and areas with sharp transitions or irregularities.
5. Homogeneity: Reflects the proximity of the distribution of elements in GLCM to the diameter of the GLCM. Higher values indicate smoother textures. Higher values (close to 1.0) at the bottom of the table indicate a softer texture of images.

From the above, we find diversity in the values of this feature that the dataset contains a mix of textures and patterns, which are likely to represent different categories (for example, normal versus abnormal tissues). These GLCM features are critical to differentiating between these categories, as they capture texture-based properties unique to each image. For example, tumors may show higher variation and variation, while normal tissues may show higher homogeneity and energy. These features will be critical to training the machine learning model to effectively classify images, as they provide rich information about the basic textures of the data.

TABLE II. GLCM FEATURE RESULTS.

Sample ID	Energy_d1_a0	Corr_d1_a0	Diss_sim_d1_a0	Homogen_d1_a0	Contrast_d1_a0
0	0.508142	0.911112	10.362632	0.550152	616.795779
1	0.459036	0.92581	13.409435	0.520337	883.82005
2	0.459229	0.925412	13.473557	0.518786	895.262259
3	0.458996	0.926313	13.385971	0.52094	890.908132

4	0.459761	0.926598	13.408132	0.52306	894.279268
...	...	...	...	...	...
5146	0.750402	0.875601	6.466356	0.771028	391.117815
5147	0.752513	0.874009	6.440968	0.773339	394.455431
5148	0.755032	0.872509	6.374395	0.776326	394.17874
5149	0.757693	0.872173	6.315084	0.778185	387.355556
5150	0.759271	0.874544	6.248914	0.779595	377.565922

B. Gabor Filter Results

Table III below includes multiple Gabor features (e.g., Gabor_feature_14, Gabor_feature_15, etc.), which are derived from applying Gabor filters at specific parameters such as frequency and direction. These features represent texture information such as edge strength, frequency patterns, and spatial structure in images. Each feature column has numerical values that correspond to the mean or standard deviation of the image filtered by Gabor, capturing the prevailing patterns and differences in texture. This makes Gabor filters particularly effective for analyzing medical images such as CT scans, where subtle differences in texture can indicate abnormalities such as tumors or lesions. Gabor's feature values vary widely across images, reflecting the diversity of texture and styles within the dataset. For example:

- Higher feature values (for example, around 88 in Gabor_feature_15) may correspond to strong edges or high-frequency texture in the image.
- Lower feature values (e.g., about 10 per Gabor_feature_18) may represent smoother or homogeneous areas with fewer differences in density. This difference in feature values indicates that the dataset includes a mix of images with distinct textures, which can represent different categories (for example, normal, tumor, cyst). In addition, trends in values (for example, a gradual decline toward the bottom rows) may reflect patterns in the dataset, such as moving from more complex textures to simpler textures.

These Gabor features are critical to training machine learning models for image classification. The combination of average standard deviation values at different frequencies and directions provides a rich set of features that capture global and local strength information. For example, images with tumors may display higher feature values in certain directions due to irregular textures, while normal tissue may show lower and more consistent values. By integrating these features into a workbook such as SVM or MLP, the model can leverage texture-based distinctions to achieve an accurate classification. This highlights the usefulness of Gabor filters as a powerful feature extraction tool in medical imaging.

TABLE III. GABOR FILTER RESULTS.

Sample ID	Gabor_feature_14	Gabor_feature_15	Gabor_feature_16	Gabor_feature_17	Gabor_feature_18	Gabor_feature_19	Gabor_feature_20	Gabor_feature_21	Gabor_feature_22	Gabor_feature_23
0	39.064938	90.788302	22.864012	33.734283	38.504938	88.569841	27.140617	45.11682	38.582469	88.648973
1	37.055494	88.947037	36.021111	44.4669	33.859815	83.687059	33.761049	38.859805	33.876605	83.703597
2	37.249167	89.122567	36.074784	44.55301	34.43679	84.275538	33.863519	39.069378	34.490432	84.337515
3	37.164537	89.046657	36.055278	44.452159	34.024475	83.852051	33.914012	39.115947	33.994506	83.816234
4	37.543056	89.427936	36.166728	44.574986	35.058611	84.959723	34.05784	39.333409	35.010154	84.909078
...	...	...	...	...	...	...	...	...	...	...
5146	27.922593	78.744977	10.288889	30.539892	28.579568	78.480574	11.964784	36.417606	28.697407	78.629583

5147	27.929599	78.74566	10.282068	30.752284	28.521049	78.412529	11.855772	36.273622	28.449691	78.320676
5148	27.164506	77.804573	10.246296	30.910016	28.044352	77.861362	11.701389	36.013099	27.949846	77.740423
5149	26.712531	77.25253	10.170957	30.972664	27.479259	77.183852	11.452716	35.487628	27.479506	77.182242
5150	26.453765	76.922124	10.120247	30.92372	28.109321	78.009732	11.502531	35.773455	28.102438	78.00247

C. Normalization Results

Normalization contributes significantly into the pre-processing step of the feature extraction process for highly complex models applied in accurate classification of Kidney cancer from CT origin. It makes sure that features whose units are different, ranges of values, or distributions are differ, for example, texture features which were extracted from GLCM and the spatial features which were extracted using Gabor filters are normalized to a similar scale. It is even more relevant to models that employ distance-based methods, because standardized data result in higher convergence rates and less fluctuations during training sessions. Nonetheless, one normalizes these features individually; thereby statistical differences between various feature sets are resolved, making the classification model more exploitable or interpretable. The proposed normalization has shown how this set normalized the features of this dataset as presented in Tables IV and V enabling the model to improve both in accuracy and robustness in detecting the kidney cancer type.

1. Normalization For GLCM

Table IV shows the normalization results for GLCM, including attributes such as energy, correlation, variation, homogeneity, and variance. These normal values indicate that standard score normalization is likely to be applied, with attributes centered around 0 with a standard deviation of 1. Energy results (Energy_d1_a0) the values in this trait are within a narrow range around the average, usually between 0.4 and 0.7, as shown in the normal results. This suggests that the original distribution of attributes was consistent across samples. After normalization, the scope was scaled to make the attribute comparable to others. Natural correlation (Corr_d1_a0) values are also relatively high, with values such as 0.9111 and 0.9258 appearing frequently. This suggests that the original feature values were close to their average, with less variation. Normalization ensures that these values are appropriately measured without being dominated by outliers. Variation values (Diss_sim_d1_a0) show a wider range of standard values, such as 10.36 and 6.44. These large values likely indicate that the original data for this feature was of higher variance. Normalization helps mitigate the effect of this variability during comparisons or model training. While homogeneity (Homogen_d1_a0) has standard values ranging from about 0.5 to 0.77. These results highlight the moderate variability in the original data, and normalization ensures that these Values are equal when used in conjunction with other GLCM features. Finally, the variance results (Contrast_d1_a0) show marked variation, with values ranging from -1.15 to 0.67. This indicates that the original feature had negative and positive values, with a significant deviation from the mean. Normalization ensures that the feature corresponds to the standardized scale of other attributes.

TABLE IV. NORMALIZATION RESULTS FOR GLCM FEATURES

Sample ID.	Energy_d1_a0	Corr_d1_a0	Diss_sim_d1_a0	Homogen_d1_a0	Contrast_d1_a0
0	0.508142	0.911112	10.362632	0.550152	-0.337902
1	0.459036	0.92581	13.409435	0.520337	0.63209
2	0.459229	0.925412	13.473557	0.518786	0.673654
3	0.458996	0.926313	13.385971	0.52094	0.657838
4	0.459761	0.926598	13.408132	0.52306	0.670084
...	...	...	...	...	...
5146	0.750402	0.875601	6.466356	0.771028	-1.157698
5147	0.752513	0.874009	6.440968	0.773339	-1.145574
5148	0.755032	0.872509	6.374395	0.776326	-1.146579
5149	0.757693	0.872173	6.315084	0.778185	-1.171365
5150	0.759271	0.874544	6.248914	0.779595	-1.206927

2. Normalization For Gabor Filter

Table V provides uniform values for different Gabor filter properties, such as Gabor_feature_14 to Gabor_feature_23. These properties have been normalized, likely using standard degree normalization, to extend each property to mean 0 and standard deviation 1. The results reflect how the outputs of the original Gabor candidate, which may have varied significantly in size and scope, were standardized for consistency. The results of normalization indicate that most of the

characteristics now fall within a range close to -2 to +2. For example, Gabor_feature_14 and Gabor_feature_15 contain uniform values such as 1.14 and 0.81, while Gabor_feature_16 and Gabor_feature_17 display greater variance, with values ranging from -1.5 to 1.5. This shows that normalization effectively concentrates properties and adjusts differences in scale. Some features, such as Gabor_feature_20, show wider uniform ranges, including negative values such as -1.977 and positive values such as 1.360. This suggests that the original Gabor output for this feature had greater contrast compared to features such as Gabor_feature_18, which have a narrower spread around 0. Normalization ensures that these differences are mitigated, allowing all features to contribute equally to subsequent tasks. When uniform values are observed across samples (for example, sample IDs from 0 to 5150), there is a smooth transition without sudden jumps. This suggests that the original data was persistent and likely distributed without anomalous values, making normalization effective in maintaining underlying patterns.

TABLE V. NORMALIZATION RESULTS FOR GABOR FILTER

Sample ID	Gabor_feature_14	Gabor_feature_15	Gabor_feature_16	Gabor_feature_17	Gabor_feature_18	Gabor_feature_19	Gabor_feature_20	Gabor_feature_21	Gabor_feature_22	Gabor_feature_23
0	1.142173	1.11449	-0.041585	-0.386741	1.176105	1.137465	0.346343	0.544287	1.186448	1.145866
1	0.805845	0.84186	1.495048	0.925269	0.450054	0.484743	1.360035	-0.555435	0.452193	0.486411
2	0.838261	0.86785	1.501316	0.935796	0.540238	0.56341	1.375725	-0.518601	0.547968	0.570942
3	0.824096	0.85661	1.499038	0.923467	0.475791	0.506799	1.383456	-0.510416	0.470589	0.50143
4	0.88745	0.913065	1.512054	0.938482	0.637431	0.654871	1.405478	-0.472195	0.62906	0.647159
...	...	...	...	...	...	...	...	...	...	...
5146	-0.72276	-0.668724	-1.510247	-0.77724	-0.375269	-0.211252	-1.977315	-0.984671	-0.355916	-0.190199
5147	-0.721588	-0.668623	-1.511044	-0.751276	-0.384416	-0.220348	-1.994007	-1.009978	-0.394567	-0.231391
5148	-0.849644	-0.807967	-1.515222	-0.731994	-0.458926	-0.294027	-2.017645	-1.055767	-0.472558	-0.308766
5149	-0.925293	-0.889706	-1.524021	-0.724336	-0.547252	-0.384595	-2.055721	-1.148123	-0.545945	-0.383199
5150	-0.968603	-0.938628	-1.529943	-0.730319	-0.448771	-0.274193	-2.048093	-1.097886	-0.448749	-0.273823

4.2 The SVM classification Results

The results of the confusion matrix of the SVM model in the following Fig. 5 show strong performance with high accuracy and balanced error rates across the binary classification task. A true negative (TN) count of 1416 and a true positive (TP) count of 629 indicates that the model correctly classifies the majority of negative and positive samples respectively. With only 56 false negatives (FN) and 108 false positives (FP), the model shows reliable detection for both categories. This is also supported by derived metrics, such as an accuracy of about 92.7%, demonstrating that the model is highly effective in classifying cases correctly overall. Accuracy and recall values provide more insights into model performance. The accuracy of 85.3% reflects that the majority of the expected positives are indeed true, although some false positives persist. A 91.8% recall indicates the model's strong ability to identify true positive states, reducing false negatives. The F-1 score of 88.4% balances accuracy and recall, making it a powerful measure of performance in scenarios where false positive and negative results cause concern. Overall, the results suggest that the SVM model is well calibrated for this classification task, with slightly better performance in identifying negative samples than positive samples.

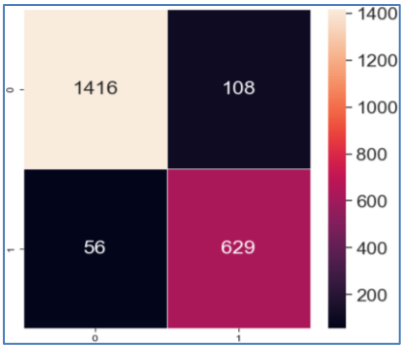

Fig. 5. Confusion Matrix Results for SVM Model.

The ROC (Future Operating Characteristics) curve of the SVM model in Fig. 6 shows the balance between true positivity (TPR) and false positivity (FPR) across different classification thresholds. The curve shows strong performance, approaching closely from the upper-left corner of the chart, demonstrating a pattern that achieves a high real positive rate (TPR) while maintaining a low false positive rate (FPR). The area under the curve (AUC) is 0.97, reflecting an excellent level of separability. An AUC value close to 1 indicates that the model can effectively distinguish between positive and negative categories. The sharp rise of the ROC curve at the beginning suggests that the pattern achieves high recall with minimal false positives at lower thresholds. As the threshold is loosened, the curve flattens, showing decreasing returns in increasing the true positivity rate while the false positivity rate grows. This behaviour corresponds to a well-calibrated model that balances sensitivity and privacy. The high AUC value confirms the model's reliability in handling binary classification tasks, making it suitable for scenarios where class differentiation is critical.

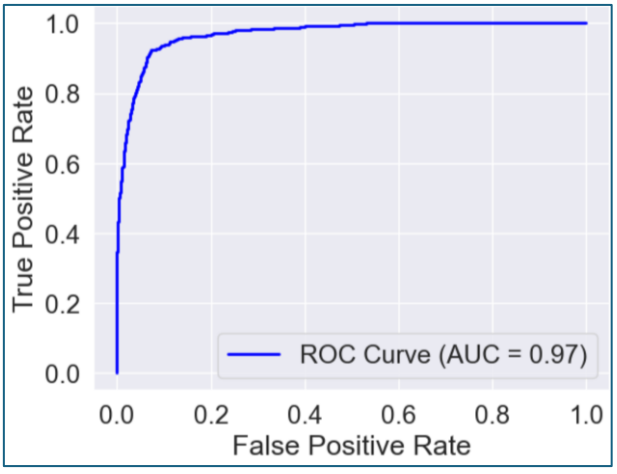

Fig. 6. ROC for SVM Model.

Fig. 7 below, the learning curve of the SVM model illustrates the relationship between training accuracy and verification accuracy with increasing the proportion of training set used. Initially, when only a small portion of training data is used, the accuracy of the training is very high, approaching 94%, suggesting that the model fits overly with the limited data. However, verification accuracy starts low, around 89%, due to insufficient data for generalization. As the proportion of training data increases, the accuracy of verification steadily improves, suggesting better generalization as the model learns from more examples. In later stages, both training accuracy and verification converge with the use of the full dataset. The final verification accuracy stabilizes slightly below the training accuracy, indicating that the model is well-optimized and avoids too much over-proportionality. The gap between the two curves is minimal, indicating a good balance between bias and contrast. However, slight fluctuations in training accuracy in the mid-range can be attributed to the sensitivity of the model to specific subsets of training data. Overall, the learning curve shows that the SVM model benefits from increased training data and achieves strong performance with proper generalization.

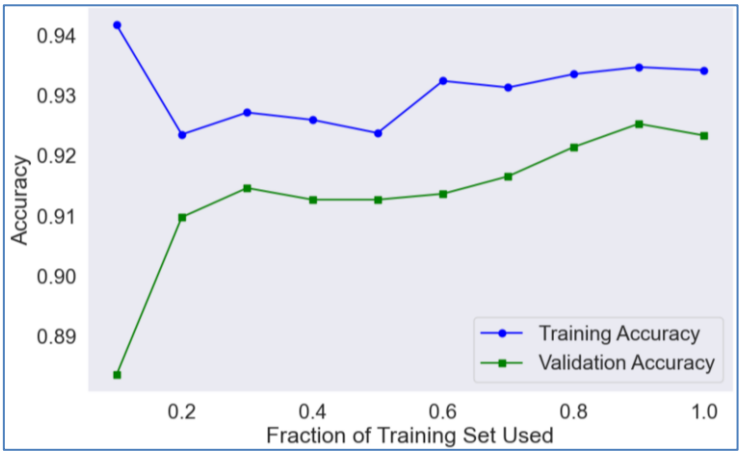

Fig. 7. Learning Curve for SVM Model.

The learning curve of the SVM model in fig.8 with cross-validation provides insights into model performance as the training group size increases. The blue curve represents the degree of training, which starts with near-perfect accuracy when the training set is small in size. This behavior indicates that the model over relates to smaller datasets and captures specific patterns in finite data. However, as the volume of training grows, the accuracy of the training gradually decreases, converging towards a more realistic value that reflects the complexity of the overall problem. The green curve, which represents the degree of cross-verification, starts low, suggesting that the model is struggling to generalize with smaller datasets. However, the degree of cross-verification steadily increases as more data becomes available, demonstrating improved generalization and reduced over relevance. Narrowing the gap between training scores and cross-verification as the dataset size approaches 4000 indicates better model stability and a balance between bias and variability. Shaded areas around the curves illustrate the variation in the results, where the cross-check curve is initially more variable, which stabilizes with more data. These results underscore the importance of a larger training suite to achieve consistent and reliable performance in the SVM model.

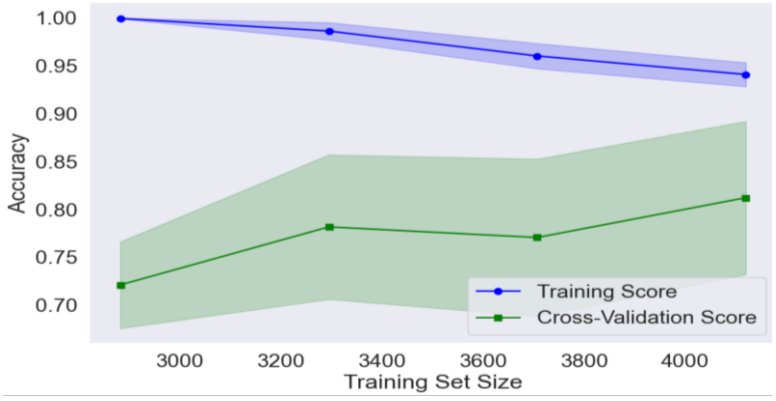

Fig. 8. Learning Curve for SVM Model with Cross-Validation.

4.3 The MLP classification Results

The confusion matrix for the multilayer sensory perception model (MLP) in Fig. 9 below shows exceptional classification performance in a binary task, as shown by the distribution of values. The model achieved 1524 true negative (TN) and 677 true positive results (TP), meaning that it correctly classified a large majority of samples in both negative (category 0) and positive (category 1) categories. There must be no false positive (FP) results, which indicates the ideal accuracy of the negative category. This suggests that all samples predicted as negative were indeed negative, a desirable trait in applications where the erroneous classification of negative cases as positive leads to serious consequences. The presence of 8 false-negative (FN) results, although small in number, indicates that some positive samples were misclassified as negative. This slight deficiency affects the recall of the positive category, suggesting that despite the accuracy of the model, it sometimes fails to detect some positive cases. Despite this, recall remains high, and the low number of false negative results reflects a strong generalization. Overall, the results suggest that the MLP model is well-tuned, effectively avoiding over relevance or inappropriateness, as evidenced by the absence of false positives and a low error rate in detecting positives. The balance between real pros and real negatives underscores the model's ability to generalize across both categories, making it highly reliable for tasks that require accurate and accurate classification.

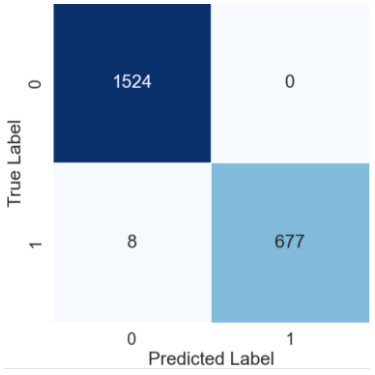

Fig. 9. Confusion Matrix Results

The ROC curve (future operating characteristics) shown in Fig. 10 shows the performance of the classification model, where the orange line represents the ROC curve and the area under the curve (AUC) is 1.00. This result indicates optimal rating performance, as the model achieves a true positive rate (TPR) 1 while maintaining a false positive rate (FPR) 0. The curve that hugs the upper-left corner of the graph indicates that the pattern distinguishes between positive and negative categories without error. Such a finding is rare and suggests that the model is either exceptionally calibrated or may have been overloaded on the dataset. The diagonal dotted line represents the basic performance of a random classifier, which achieves an area under the curve 0.5. The large gap between the model curve and this baseline highlights the model's superior discriminatory capability. However, while AUC 1.00 is perfect in theory, it raises concerns about potential overprocessing, especially if the dataset is small or lacks sufficient variability. This finding is particularly relevant for real-world applications, where perfect class segregation is difficult due to noise and overlapping features. Thus, while the figure reflects excellent performance, further investigation of the dataset and cross-verification results will be useful to ensure model strength in broader or invisible datasets.

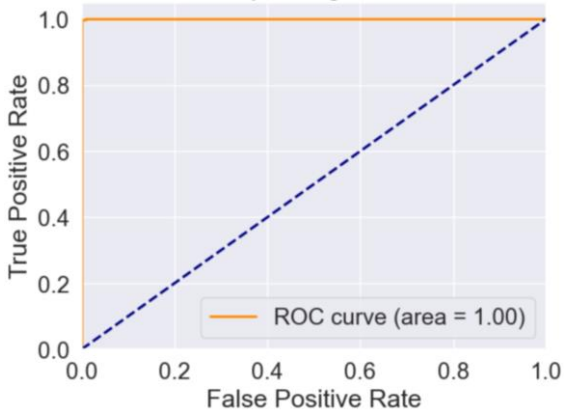

Fig. 10. Receive Operating Characteristic

The learning curve in Fig. 11 shows the relationship between the training score and the degree of mutual verification as a function of the training group size. The training score, represented by the red line, remains roughly constant at a very high value, close to 1.0, across all training group sizes. This suggests that the model fits the training data perfectly, which is often a sign of low bias. However, the lack of a decrease in training accuracy as the training group grows may indicate potential over-processing, as the model is more likely to memorize training data rather than generalize it to invisible examples. By contrast, the cross-verification score, represented by the green line, starts remarkably low, around 0.70, when the training group size is small. As the training group size increases, the degree of cross-verification is steadily improving, suggesting that the model begins to generalize better with more data. However, there is still a significant gap between the training score and the degree of mutual verification, even in larger dataset sizes, suggesting that the model is struggling to fully generalize. This discrepancy highlights the problem of variance where the model works very well on training data but is less effective on validation data. To address this issue, techniques such as organization, increasing the diversity of the training dataset, or reducing the complexity of the model may be necessary to enhance generalization.

Fig. 11. Learning Curve

The ROC curve in Fig. 12 shows the classification performance of the MLP (Multilayer Calculator) model, achieving an area under the curve (AUC) of 1.00. This suggests that the MLP model separates the two categories altogether, achieving a true positive rate (TPR) of 1.0 without experiencing any false positive rate (FPR). The curve that hugs the upper-left corner of the chart reflects the ideal rating performance, with the model accurately identifying all positive states while avoiding false positive predictions altogether. This result is highly desirable in applications where sensitivity and privacy are paramount. However, while an ideal AUC score is theoretical idealism, it is important to critically evaluate this outcome. Achieving AUC 1.00 may indicate the possibility of over-processing, especially if the model is trained on a dataset that does not represent real-world variation or contains limited noise. Over-processing means that the model works exceptionally on the training or test suite but may have difficulties dealing with invisible or more complex data. To ensure robustness, additional evaluation through independent datasets, cross-verification, or external calibration is essential. While the ROC curve shows exceptional performance, the broader context of the data set and the possibility of real-world generalization must be considered to validate this result.

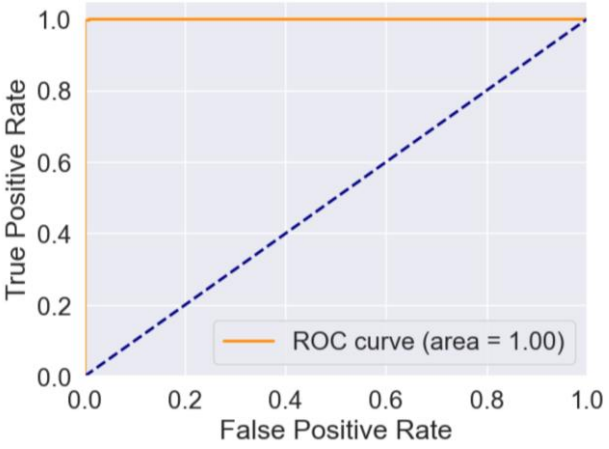

Fig. 12. Receiver Operating Characteristic for MLP

The learning curve of the MLP model in Fig. 13 shows the model's training accuracy (red line) and cross-checking accuracy (green line) as a function of the training group size. The training score remains consistently high, close to 1.0, regardless of the size of the training group. This suggests that the MLP model is able to fit the training data perfectly, reflecting a model with low bias. However, this high training accuracy indicates the possibility of over-processing, as the model seems to save training data rather than generalizing well to invisible data. On the other hand, the cross-check score starts remarkably low, around 0.70, when the training group is small, and steadily improves as the training group size increases. This trend suggests that as the model is exposed to more training data, it generalizes better, resulting in higher validation accuracy. However, there is still a noticeable gap between training scores and cross-verification, even with larger data sets, suggesting that the model did not fully overcome over-processing. To mitigate this problem, strategies such as organizing, exclusion, or reducing model complexity can help improve generalization. Overall, while MLP shows strong performance on the training suite, the increased degree of mutual verification highlights the importance of larger datasets for better real-world applicability.

Fig. 13. Learning Curve & Cross-Validation for MLP

The learning curve in Fig. 14 shows the relationship between training accuracy (red line) and test accuracy (green line) with increasing the training group portion of the MLP model. Training accuracy remains constant at approximately 1.0 across all parts of the training suite. This suggests that the MLP model fits the training data perfectly, suggesting that the model has low bias. However, consistently high training accuracy may indicate potential overprocessing, as the model saves training data rather than generalizing to invisible data. On the other hand, the accuracy of the test starts much lower, around 0.70, when only a small part of the training kit is used. As the training data portion increases, the accuracy of the test gradually improves, with a significant increase observed as the training group size approaches its maximum. Ultimately, the accuracy of the test converges with the accuracy of the training, suggesting that the additional data helps the model to generalize better. However, the initial gap between training and test accuracy in most of the curve highlights over-processing problems, with the model working perfectly on training data but less effectively on invisible data. This gap narrows as training data increases, suggesting that collecting more data or applying regulation techniques can help improve test accuracy and further reduce overprocessing. The convergence of the two lines near the training set break limit shows that the model is approaching a better generalization with sufficient data.

Fig. 14. Learning Curve for MLP.

Finally, in the present a table VI showing the performance metrics to show the difference between the SVM and MLP models in terms of accuracy, Precision, recall and F1 coefficient.

TABLE VI. COMPARISON BETWEEN SVM AND MLP MODELS RESULTS.

Metrics / Model	Accuracy	Precision	Recall	F1 Score
SVM	93%	90.5%	92.5%	91.5%
MLP	99.64%	99.5%	99.5%	99.5%

Table VI compares the performance metrics of two classification models, Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP), in terms of accuracy, precision, recall, and F1 score. These metrics provide a comprehensive evaluation of the models' effectiveness in correctly classifying data and handling both positive and negative cases. The SVM model achieves 93% accuracy, 90.5% precision, 92.5% recall, and an F1 score of 91.5%, indicating a well-balanced performance with a strong trade-off between precision and recall. However, its overall performance is lower than the MLP model, particularly in accuracy and recall, suggesting that SVM may misclassify some positive cases or produce false positives. In contrast, the MLP model demonstrates significantly superior performance across all metrics, achieving 99.64% accuracy, 99.5% precision, 99.5% recall, and an F1 score of 99.5%. These results reflect near-perfect classification accuracy, with minimal false positives and false negatives. The close alignment of precision, recall, and F1 score further highlights the MLP model's balanced sensitivity and specificity. The superior performance of the MLP model can be attributed to its neural network architecture, which effectively captures complex and nonlinear patterns in the data, making it particularly advantageous for intricate classification tasks.

However, the MLP model's near-ideal metrics raise concerns about potential overfitting, especially if the training and testing datasets lack diversity or fail to represent real-world scenarios adequately. On the other hand, while the SVM model shows lower performance, it may generalize better to unseen data and exhibit greater robustness against overfitting, depending on the application. Ultimately, the choice between these models should consider the task's specific

requirements, the dataset's complexity, and the importance of balancing performance with the risk of overfitting. The limitations of this study include:

- The SVMs and the MLPs may not perform well when it comes to generalization of the unseen data in particular when the feature space of the dataset is limited in sample size, or restricted to demography and imaging characteristics of patients. This can lead to additional imprecision since the employed model is not as accurate when applied to other CT imagery samples.
- Though both the SVM and MLP models are equally effective in such classification tasks done on large CT image datasets with many numbers of features, both models are computationally very intensive and time consuming. The major drawback is the necessity of hyperparameter tuning and large memory storage that is a problem when using high resolution medical image data.
- The assessments reveal that both SVM and MLP are vulnerable to noise and outliers in the data. However, when using CT images, noises within the images or some small shifts in the image position can significantly degrade the feature extraction step and even decrease the classification accuracy of the models.

5. CONCLUSION

This study also determines the capacity in which machine learning models, namely SVM and MLP could revolutionise the accurate classification of kidney cancer using relevant CT images. The study complemented the extraction of feature variables by using higher order texture and structural features such as GLCM and Gabor filters. Here the MLP Model outperformed the general model with 99.64% thus showing a high level of efficiency as compared to the SVM model at 93%. These results confirm the effectiveness of the MLP to capture the complex relationships between the input and output, as well as the enhanced efficiency of the MLP in solving classification problems in the medical image field. Nonetheless, the high performance of the MLP increases questions of overfitting, suggesting more external verifications on various and larger databases. This work also presents one way of how the clinical process can be extended using machine learning in an optimization and accuracy manner to seize a better diagnosis for patients. In further research, collecting a larger sample size of data, tweaking the hyperparameters of the model greater with more rigorous testing of feature selection methods such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) should be the future direction of the research. Moreover, using superior deep learning networks for feature extraction, may prove to have even better performance. In conclusion, this dissertation aids in the creation of new machine learning techniques that may greatly enhance the prognosis for the early detection and prognosis of kidney cancer and individualized course of therapy.

Conflicts Of Interest

The author's disclosure statement confirms the absence of any conflicts of interest.

Funding

The authors had no institutional or sponsor backing.

Acknowledgment

The authors expresses appreciation to the institution for their continuous support and access to relevant research materials.

References

[1] S. Mahmud, T. O. Abbas, A. Mushtak, J. Prithula, and M. E. H. Chowdhury, “Kidney Cancer Diagnosis and Surgery Selection by Machine Learning from CT Scans Combined with Clinical Metadata,” *Cancers (Basel)*, vol. 15, no. 12, 2023, doi: 10.3390/cancers15123189.

[2] Kadhim, D.A. and Mohammed, M.A., 2024. A Comprehensive Review of Artificial Intelligence Approaches in Kidney Cancer Medical Images Diagnosis, Datasets, Challenges and Issues and Future Directions. *International Journal of Mathematics, Statistics, and Computer Science*, *2*, pp.199-243.

[3] M. Ferro *et al.*, “Artificial intelligence and radiomics in evaluation of kidney lesions: a comprehensive literature review,” 2023. doi: 10.1177/17562872231164803.

[4] F. TÜRK, “Machine Learning of Kidney Tumors and Diagnosis and Classification by Deep Learning Methods,” *Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi*, 2019, doi: 10.29137/umagd.640667.

[5] Mohammed, M.A., Lakhan, A., Abdulkareem, K.H., Deveci, M., Dutta, A.K., Memon, S., Marhoon, H.A. and Martinek, R., 2024. Federated-Reinforcement Learning-Assisted IoT Consumers System for Kidney Disease Images. *IEEE Transactions on Consumer Electronics*.

[6] R. M. Haralick, I. Dinstein, and K. Shanmugam, "Textural Features for Image Classification," *IEEE Trans Syst Man Cybern*, vol. SMC-3, no. 6, 1973, doi: 10.1109/TSMC.1973.4309314.
[7] N. N. Al-Mayahi and F. G. Mohammed, "Esophageal Cancer Segmentation Based on FCM Algorithm," in *AIP Conference Proceedings*, American Institute of Physics Inc., Feb. 2024. doi: 10.1063/5.0185313.
[8] Ali, A.M. and Mohammed, M.A., 2024. A comprehensive review of artificial intelligence approaches in omics data processing: evaluating progress and challenges. *International Journal of Mathematics, Statistics, and Computer Science*, *2*, pp.114-167.
[9] M. Saini and R. Chhikara, "DWT Feature based Blind Image Steganalysis using Neural Network Classifier," *International Journal of Engineering Research & Technology (IJERT)*, vol. 4, no. 4, pp. 776–782, 2015, [Online]. Available: www.ijert.org
[10] A. S. Mahmoud, O. Lamouchi, and S. Belghith, "Advancements in Machine Learning and Deep Learning for Early Diagnosis of Chronic Kidney Diseases: A Comprehensive Review," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 149–156, Sep. 2024, doi: 10.58496/BJML/2024/015.
[11] W. Abbaoui, S. Retal, B. El Bhiri, N. Kharmoum, and S. Ziti, "Towards revolutionizing precision healthcare: A systematic literature review of artificial intelligence methods in precision medicine," 2024. doi: 10.1016/j.imu.2024.101475.
[12] C. He, X. Wu, J. Zhou, Y. Chen, and J. Ye, "Raman optical identification of renal cell carcinoma via machine learning," *Spectrochim Acta A Mol Biomol Spectrosc*, vol. 252, 2021, doi: 10.1016/j.saa.2021.119520.
[13] T. Dai *et al.*, "Benchmarking Machine Learning Algorithms for Diagnosis of Renal Cell Carcinoma," *Iranian Journal of Radiology*, vol. 19, no. 3, 2022, doi: 10.5812/iranjradiol-119266.
[14] G. Garg, D. Kumar, Y. Sonker, and R. Garg, "A Hybrid MLP-SVM Model for Classification using Spatial-Spectral Features on Hyper-Spectral Images," *arXiv:2101.00214 Search...*, pp. 1–9, 2022.
[15] J. Xu, X. He, W. Shao, J. Bian, and R. Terry, "Classification of Benign and Malignant Renal Tumors Based on CT Scans and Clinical Data Using Machine Learning Methods," *Informatics*, vol. 10, no. 3, 2023, doi: 10.3390/informatics10030055.
[16] Y. Li, X. Gao, X. Tang, S. Lin, and H. Pang, "Research on automatic classification technology of kidney tumor and normal kidney tissue based on computed tomography radiomics," *Front Oncol*, vol. 13, 2023, doi: 10.3389/fonc.2023.1013085.
[17] D. ŞİRİN and A. GÜVENİŞ, "PREDICTING KIDNEY TUMOR SUBTYPE FROM CT IMAGES USING RADIOMICS AND CLINICAL FEATURES," *Natural and Applied Sciences Journal*, vol. 5, no. 1, 2022, doi: 10.38061/idunas.1084748.
[18] S. Prathipati, "A COMPARATIVE STUDY ON CHRONIC KIDNEY CANCER PREDICTION USING SUPERVISED MACHINE LEARNING ALGORITHMS," *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, vol. 07, no. 04, Apr. 2023, doi: 10.55041/IJSREM18629.
[19] Z. Lin *et al.*, "Recognizing pathology of renal tumor from macroscopic cross-section image by deep learning," *Biomed Eng Online*, vol. 22, no. 1, 2023, doi: 10.1186/s12938-023-01064-4.
[20] Y. Wei, M. Gao, J. Xiao, C. Liu, Y. Tian, and Y. He, "Research and Implementation of Cancer Gene Data Classification Based on Deep Learning," *Journal of Software Engineering and Applications*, vol. 16, no. 06, 2023, doi: 10.4236/jsea.2023.166009.
[21] Y. Zhang *et al.*, "Machine learning-based prognostic and metastasis models of kidney cancer," *Cancer Innovation*, vol. 1, no. 2, 2022, doi: 10.1002/cai2.22.
[22] S. M. Hosseiniyan Khatibi, M. Ardalan, M. Teshnehlab, S. Z. Vahed, and S. Pirmoradi, "Panels of mRNAs and miRNAs for decoding molecular mechanisms of Renal Cell Carcinoma (RCC) subtypes utilizing Artificial Intelligence approaches," *Sci Rep*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-20783-7.
[23] Islam MN, Hasan M, Hossain M, Alam M, Rabiul G, Uddin MZ, Soylu A. Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography. Scientific Reports. 2022 Jul 6;12(1):1-4.
[24] M. A. . Ahmed, "Design and Synthesis of Biodegradable Polymeric Nanoparticles for Targeted Delivery of Chemotherapeutics in Triple-Negative Breast Cancer", SHIFAA, vol. 2023, pp. 32–40, Apr. 2023, doi: 10.70470/SHIFAA/2023/004.
[25] Abbas, T., Ali, S.F., Mohammed, M.A., Khan, A.Z., Awan, M.J., Majumdar, A. and Thinnukool, O., 2022. Deep learning approach based on residual neural network and SVM classifier for driver's distraction detection. Applied Sciences, 12(13), p.6626.
[26] Ali, A.M. and Mohammed, M.A., 2024. Optimized Cancer Subtype Classification and Clustering Using Cat Swarm Optimization and Support Vector Machine Approach for Multi-Omics Data. Journal of Soft Computing and Data Mining, 5(2), pp.223-244.

[27] Saeed, M.H., Kosar, N., Hassan, S.U., Nadeem, S., Mohammed, M.A., Abd Ghani, M.K. and Abdulkareem, K.H., 2023. Determination of bandgap of period 3, 4, and 5 transition metal dopants on zinc oxide using an artificial neural network based approach. Chemometrics and Intelligent Laboratory Systems, 242, p.104983.
[28] T. Al-Quraishi, C. Keong NG, O. A. Mahdi, A. Gyasi, and N. Al-Quraishi, “Advanced Ensemble Classifier Techniques for Predicting Tumor Viability in Osteosarcoma Histological Slide Images”, Applied Data Science and Analysis, vol. 2024, pp. 52–68, May 2024.