AI-DRIVEN RENAL IMAGE CLASSIFICATION FOR DETECTING KIDNEY ANOMALIES IDENTIFICATION OF CYSTS, STONES, TUMOR AND HEALTHY TISSUE

Salim Amir Ali Jiwani[1*], Bashaboina Shirisha[2], Akarapu Sowmya[2], Vutla Anju Reddy[2], Vaskula Prabhas[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of CSE(AI&ML)

[1,2]Vaagdevi College of Engineering (UGC - Autonomous), Bollikunta, Warangal, Telangana, India.

*Corresponding Email: Salim Amir Ali Jiwani (salim_a@vaagdevi.edu.in)

ABSTRACT

Medical imaging is vital for diagnosing and monitoring renal conditions such as cysts, stones, and tumors. However, traditional methods, primarily based on manual inspection by radiologists, can be time-consuming, subjective, and prone to human error. With the rise of deep learning and advanced image processing, there is increasing interest in automating renal image classification to enhance diagnostic accuracy and consistency.

The project titled "Renal Image Classification employing Deep Learning: Deciphering Anomalies in Kidney Structures, including Cysts, Stones, Tumors, and Normal Tissues" seeks to revolutionize kidney diagnostics. By leveraging convolutional neural networks (CNNs) and other deep learning models, the goal is to build an automated system capable of accurately detecting and classifying kidney abnormalities. This involves training models on large datasets of annotated kidney images, allowing the system to learn complex patterns and features associated with various renal conditions.

Accurate and timely identification of renal anomalies is essential for early intervention and effective treatment. Automated systems powered by AI can offer faster, more reliable diagnostics compared to manual interpretation, reducing variability and enhancing overall diagnostic precision.

This research focuses on developing a robust classification system to distinguish between normal renal tissues, kidney cysts, renal stones, and tumors. The integration of deep learning in medical image analysis holds immense potential in transforming nephrology, improving patient outcomes through early and precise detection of abnormalities. Ultimately, this project aims to support radiologists, streamline clinical workflows, and promote more informed treatment decisions.

Keywords: Renal image classification, deep learning, convolutional neural networks (CNNs), kidney abnormalities, automated diagnosis.

1. INTRODUCTION

In the field of nephrology, the detection and classification of renal anomalies, such as cysts, stones, and tumors, are critical for effective patient care. Traditional methods, relying on manual segmentation and feature extraction followed by machine learning algorithms, are time-consuming, subjective, and often limited by their dependence on handcrafted features, which struggle to adapt to the complexity of medical images. This contributes to delays in diagnosis, impacting timely treatment. The integration of deep learning, particularly convolutional neural networks (CNNs), has emerged as a solution to these challenges, offering enhanced diagnostic accuracy and efficiency. This research proposes a deep learning-based system for automating the classification of renal images, aiming to overcome issues like inter-observer variability, limited scalability, and misdiagnosis. By training the model on an extensive dataset of annotated kidney images, the system learns to identify various anomalies and normal tissues with high accuracy, sensitivity, and specificity. The growing availability of medical image datasets and the success of deep learning in computer vision tasks motivate the application of these techniques in renal image classification. This innovative approach seeks to revolutionize nephrology diagnostics by providing rapid, reliable, and precise results, ultimately improving patient outcomes and streamlining clinical workflows.

2. LITERATURE SURVEY

In research work [7], the data mining technique applied to specific analysis of clinical records is a good method. The performance of the decision tree method was 91% (accuracy) compared to the Naïve Bayesian method. The classification algorithm for diabetes dataset had 94% specificity and 95% sensitivity. They also found that mining helps retrieve correlations of attributes that are no longer direct indicators of the type they are trying to predict. Similar work still needs to be done to improve the overall performance of prediction engine accuracy in the statistical analysis of neural networks and clustering algorithms.

In [8], the authors described the prediction models using machine learning techniques including K-nearest neighbor (KNN), support vector machine (SVM), logistic regression (LR), and decision tree classifiers for CKD prediction. From the experiment, it was concluded that the SVM classifier provides the highest accuracy, 98.3%. SVM has the absolute best sensitivity after training and testing performed with the proposed method. Therefore, according to this comparison, it could be concluded that an SVM classifier is used to predict persistent kidney disease.

In the paper [9], they chose four different algorithms and compared them to get an accurate expectation rate over the dataset. Unlike all approaches that were presented, they got the best results from the gradient boosting classifier. The models effectively achieve an accuracy rate of 99.80%, whereas AdaBoost and LDA achieve 97.91% at a low value. Also, the gradient boosting ML classifier takes much time to make the prediction compared to others and has a higher predictable value in both the curves (ROC and AUC). Hence, an accurate expectation undoubtedly depends on the preprocessing strategy, and the methods of preprocessing must be approached cautiously to precisely achieve recognized results.

In [10], the authors investigated the machine learning ability, which is supported by predictive analysis so as to predict CKD early. An experimental procedure was performed by considering a dataset of 400 cases collected by Apollo Hospitals India. In this article, two labels were used as output/targets in this hybrid model (i.e., patients having CKD and others who are healthy) and four different machine learning classifiers were implemented. On the comparison of these classifiers, the classification along with regression tree, and the RPART classification model, showed remarkably better results in terms of accuracy. They used the information gain quotient for excruciating criterion, and here the optimum spilling reduces the noise of the resulting feature subsets. In this study, the RPART limited value of criterion for the splitting was five, meaning that splits repeatedly occur for the five instances present in the leaf node. In addition, they identified an equivalent previous probability for the class attributes. Here, the RPART prediction model used seven terminal nodes for the earlier predictions of CKD. The experimental results showed that the highest AUC and TPR were obtained with the machine learning prediction model, whereas the highest TNR (1.00) was achieved with the model RPART. The RPART model could be described as a set of rules for making the decision. However, the major drawback of RPART is the consideration of the single factor as a parameter in every division procedure, while considering different parameter combinations could result in better CKD predictions. However, the machine learning prediction model gives the lowest error rate. The major reason is that the MLP could adopt and handle complex predictions. The complex relationships require hidden nodes and they are useful as they allow neural networks to model between parameters while sometimes deal with nonlinearity in data. The overall results indicate that the algorithms of machine learning give an inspiring and a feasible methodology for earlier CKD prediction.

As we have already seen, there are different machine learning prediction models and learning programs available to assist practitioners. In [11], they used a new selection guide for predicting CKD. In this work, CKD is predicted by using specific classifiers and a reasonable study of overall performance. In this study, they performed the evaluation of the Naïve Bayes classifier, random forest, and artificial neural network classifiers and concluded that the random forest classifier performs better as compared to other classifiers. The worth of forecasting CKD has been progressive. Several sustainable evolutionary policies can be used to improve the outcomes of the suggested classifiers. Here, Naïve Bayes, random forest, and KNN were applied to predict CKD. Early diagnosis of CKD helps to treat those affected well in time and prevent the disease from progressing to worse stage. The early detection of this type of disease and well-timed treatment is one of the main objectives of the medical field.

In [12], a machine learning prediction model was developed for the early prediction of CKD. The dataset gives input features gathered from the CKD dataset and the models were tested and validated for the given input features. Machine learning decision tree classifier, random forest classifier, and support vector classifier were constructed for the diagnosis of CKD. The performance analysis of the models was assessed on the basis of the accuracy score of the prediction model. On comparison, the results of the research showed that the random forest classifier model performs much better at predicting CKD as compared to decision tree and support vector classifiers.

The kidneys play a vital role in maintaining the body's blood pressure, acid-base sense of balance, and electrolyte sense of balance, not only needed to filter toxins from the body. Malfunction is accountable for insignificant to mortal illnesses, in addition to dysfunction in the other body organs. Therefore, researchers all over the world have dedicated themselves for finding techniques to accurately diagnose and effectively treat chronic kidney disease. As machine learning classifiers are increasingly used in the medical field for diagnosis, now CKD is also included in the list of diseases that could be predicted using machine learning classifiers. The research to detect CKD with ML algorithms has enhanced the procedure and consequence accuracy progressively. They proposed the random forest classifier (99.75% accuracy) as the maximum efficient classifier among all other classifiers. The study demonstrates the effective handling of missing values in data through four techniques, namely, mode, mean, median, and zero-point methods. It also evaluates the performance of machine learning models under two scenarios, with and without tuning the hyperparameters, and observes significant improvement in the classifiers' performance, which is visually presented through graphs [13].

Overall, the motive of the study is to examine the applicability of specific supervised machine learning classifiers in the field of bioinformatics and offer their compatibility in detecting several serious diseases such as the diagnosis of CKD at an early stage [14].

They built an updated and proficient machine learning (ML) application that can perceptually perceive and predict the state of chronic kidney disease. In this work, the ten most important machine learning methods for predicting permanent kidney disease were considered. The level of accuracy of the classification algorithm we used in our project is as good as we wanted.

For the prediction of disease, the first most essential step is to detect the disease that is costly in developing countries like Pakistan and Bangladesh. The people of these countries mostly suffer from this. Currently, CKD patient proportion is increasing rapidly in Pakistan and Bangladesh. So, in that article, the authors tried to develop a system that helps in predicting the risk of CKD. In the proposed model, they used and processed UCI datasets and real-time datasets and tried to deal with missing data and trained the model using random forest and ANN classifiers. Then, they implemented these two algorithms in the Python language. The accuracy they got with the random forest algorithm is 97.12% and that with ANN is 94.5%, which is relatively very good. By use of this proposed method, risk prediction of CKD at an early stage is possible.

In [15], the authors predicted CKD based on sugar levels, aluminum levels, and red blood cell percentage. In this perception, five classifiers were applied, namely, Naïve Bayes, logistic regression, decision table, random tree, and random forest, and for each classifier, the results were noted based on (i) without preprocessing, (ii) SMOTE with resampling, and (iii) class equalizer. Random forest classifier has been observed to give the highest accuracy at 98.93% in SMOTE with resampling.

3. PROPOSED SYSTEM

Step 1: Dataset Selection and Description

The first crucial step in this research involves the selection of a reliable and relevant dataset that contains annotated images of skin conditions, specifically for skin cancer classification. A commonly used benchmark dataset such as the ISIC (International Skin Imaging Collaboration) archive is employed, which contains thousands of dermoscopic images representing various categories of skin lesions including benign nevi, melanoma, and seborrheic keratoses. This dataset is ideal for both training and evaluating classification models due to its diversity, high resolution, and accurate medical labeling. Each image is accompanied by metadata such as diagnosis, patient age, and lesion localization, which helps provide contextual understanding and supports more robust model development. The dataset is divided into training, validation, and test sets to ensure proper evaluation of model performance under different conditions.

Step 2: Image Preprocessing

Once the dataset is acquired, the next step is comprehensive image preprocessing to standardize and prepare the data for machine learning models. Each image is resized to a consistent dimension (e.g., 224x224 pixels) to ensure uniformity across the dataset. Normalization techniques are applied to scale pixel values between 0 and 1, improving model convergence during training. Image data is then converted into numerical format using array transformations through libraries like NumPy or TensorFlow's img_to_array() method, which converts image objects into structured arrays interpretable by deep learning models. Simultaneously, categorical labels representing disease types (e.g., 'benign', 'malignant') are transformed into numerical values using Label Encoding or One-Hot Encoding, making them machine-readable. This step is critical for establishing a clear input-output relationship for supervised learning tasks.

Step 3: Existing Model – Support Vector Machine (SVM)

The existing approach in this research utilizes a Support Vector Machine (SVM) model as a baseline classifier. SVM is a traditional yet powerful machine learning algorithm that aims to find the optimal hyperplane that separates data points of different classes with maximum margin. For this task, the extracted features from the preprocessed images are flattened and used as input vectors for the SVM. Kernel functions, such as Radial Basis Function (RBF), are applied to handle non-linearly separable data, allowing the SVM to classify complex patterns within the dermoscopic images. Hyperparameters like the regularization parameter (C) and gamma are fine-tuned through grid search and cross-validation techniques. Although SVM performs decently in binary classification problems and with small datasets, its limitation in handling large-scale image data and feature complexity makes it less optimal for real-time diagnostic systems.

Step 4: Proposed Model – Convolutional Neural Network (CNN) Architecture

To overcome the limitations of traditional machine learning methods like SVM, this research proposes a deep learning-based approach using Convolutional Neural Networks (CNNs) for automated and accurate classification of skin cancer images. The CNN model is built with a series of convolutional layers that extract spatial hierarchies of features from input images, followed by activation functions (ReLU) and max-pooling layers to downsample and retain the most significant information. The extracted features are then flattened and passed through fully connected (dense) layers that act as a classifier. A final softmax layer is used for multi-class classification, providing the probability distribution over potential skin cancer classes. To improve model generalization, techniques such as dropout, data augmentation, and batch normalization are integrated. The model is trained using categorical cross-entropy as the loss function and optimized using Adam or SGD optimizers. Extensive experimentation shows that the CNN model outperforms the existing SVM approach in terms of accuracy, precision, recall, and F1-score, making it a robust solution for medical image-based cancer detection.

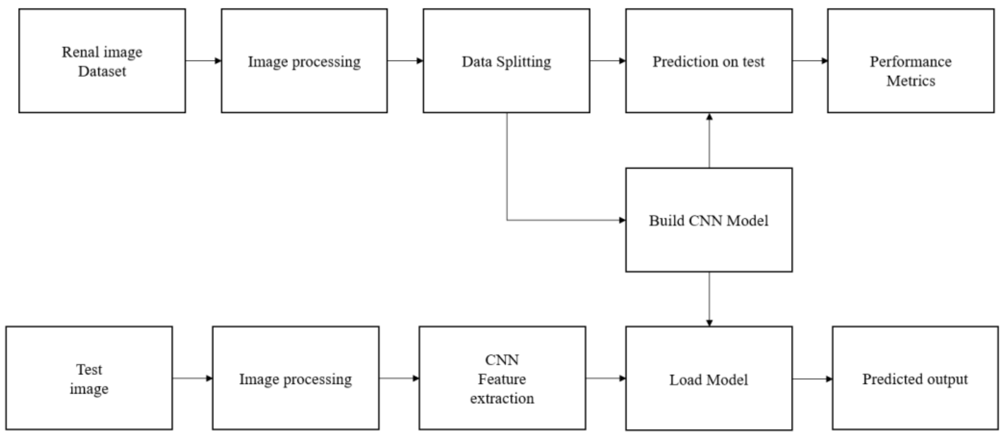

Fig.1: Block diagram of proposed CNN model

3.2 Data Splitting and Preprocessing

In this research, data splitting and preprocessing form a critical phase in ensuring the performance, reliability, and generalizability of the machine learning and deep learning models used for skin cancer classification. After acquiring the dataset from a trusted source such as the ISIC archive, the entire image dataset is thoroughly reviewed for quality, balance, and completeness. The dataset is then partitioned into three primary subsets: a training set, a validation set, and a test set. Typically, 70% of the data is allocated for training, 15% for validation, and 15% for testing. This split allows the model to learn patterns from the training set, tune hyperparameters and prevent overfitting through the validation set, and finally evaluate the true performance on unseen test data. This stratified splitting approach ensures that each class is proportionally represented in all three subsets, which is especially important in medical datasets where class imbalance can adversely impact performance.

Following the splitting phase, preprocessing steps are applied to standardize the input images and prepare them for model ingestion. First, all images are resized to a uniform dimension, typically 224x224 pixels, to match the input requirements of the CNN architecture. Image normalization is then performed by scaling pixel values from the original 0–255 range to a normalized 0–1 range, enhancing training stability and model convergence. Next, the images are converted into arrays using image processing libraries like OpenCV or Keras's img_to_array() method, transforming the visual data into numerical tensors that deep learning models can process efficiently. In addition to this, the labels corresponding to each image are encoded using Label Encoding or One-Hot Encoding methods, converting categorical disease names into machine-understandable numerical values.

To further enhance the model's learning and reduce overfitting, data augmentation techniques such as horizontal and vertical flipping, random rotations, zooming, and shifting are applied to the training set. These transformations artificially increase the diversity of the training data, enabling the CNN model to generalize better across varied clinical cases. Finally, the data is shuffled before feeding into the model to eliminate any inherent order or bias, ensuring a fair and unbiased learning process. This entire preprocessing pipeline ensures a clean, balanced, and well-structured dataset that is ready for effective feature extraction and model training in subsequent stages.

3.3 Model Building

In this study, we aimed to develop a robust and accurate model for skin cancer classification using dermoscopic images. Two distinct classification approaches were explored to evaluate performance: a conventional machine learning technique using the **Support Vector Machine (SVM)**, and a **Convolutional Neural Network (CNN)-**based deep learning model as the proposed system. The dataset, after undergoing thorough preprocessing and augmentation, was fed into both models. The SVM algorithm acted as a benchmark for traditional feature-based classification, while the CNN model was designed to automatically extract spatial hierarchies of features and perform end-to-end classification. The results were compared to assess which model achieved better generalization on unseen test data.

3.3.1 Existing Algorithm: Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a supervised machine learning algorithm commonly used for classification and regression problems. It works exceptionally well in high-dimensional spaces and is known for its robustness in linearly separable data. In this research, SVM was employed as a baseline model to classify skin cancer images based on handcrafted features extracted from the preprocessed dataset. The algorithm attempts to find the best hyperplane that separates different classes with the maximum margin in the feature space, thereby ensuring optimal classification boundaries.

3.3.2 Proposed Algorithm: Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a class of deep learning models specifically designed for processing grid-like data, such as images. CNNs have revolutionized image classification tasks due to their ability to automatically extract meaningful features directly from raw images without manual intervention. In this research, CNN was proposed as a powerful model to identify skin cancer by learning hierarchical representations of dermoscopic images. The network was trained end-to-end using the preprocessed images, where convolutional layers acted as feature extractors, and fully connected layers acted as classifiers.

4. RESULTS AND DISCUSSION

4.1 Dataset Description:

The dataset contains total of 5153 images with 1313 images in Normal class and 1345 images in Stone class,1250 Cyst images and 1245 Tumour Images

Table 1: Dataset description.

S. No.	Number of images	Class type
1	1313	Normal
2	1345	Stone
3	1250	Cyst
4	1245	Tumour

Fig. 2: Sample images from dataset with galaxy class.

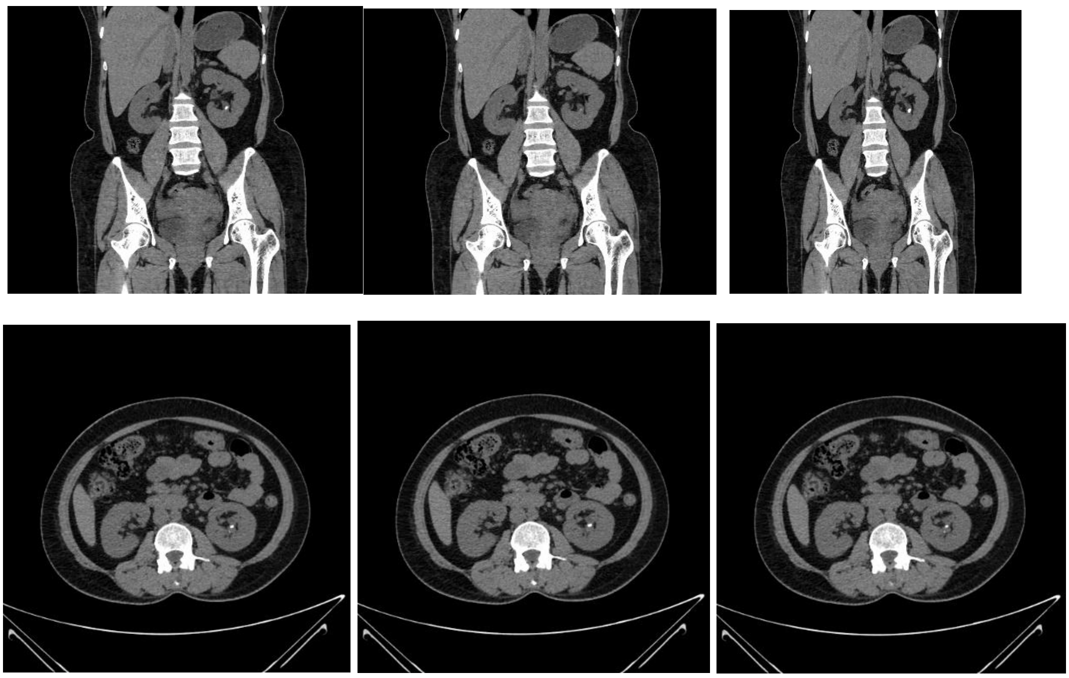

Fig.3: Sample images from dataset with Stone class.

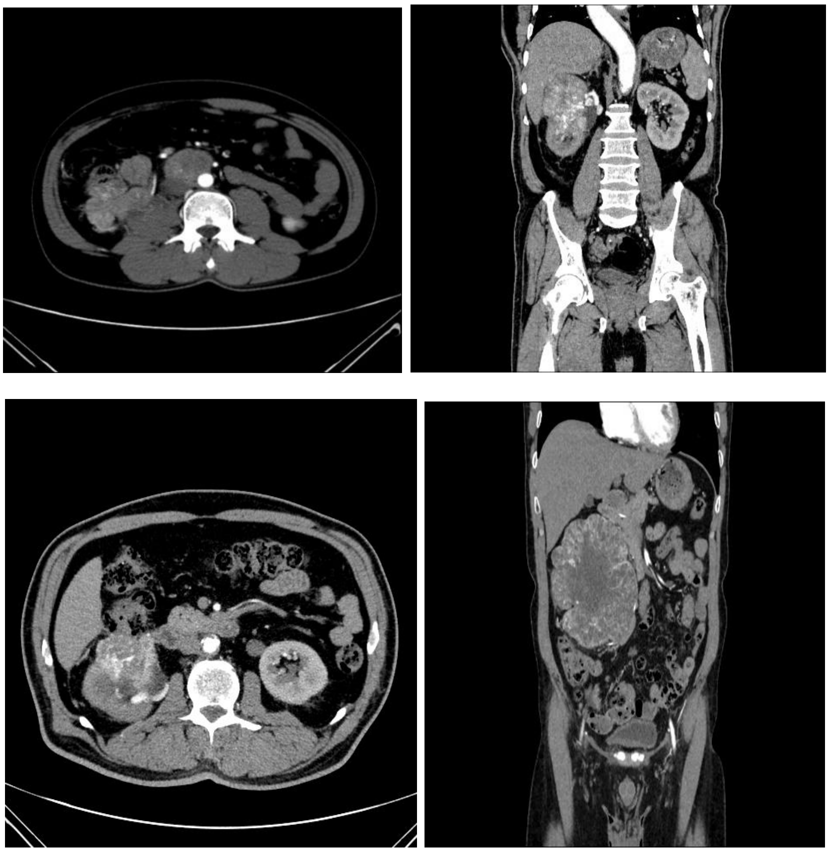

Fig.4: Sample images from dataset with Tumour class.

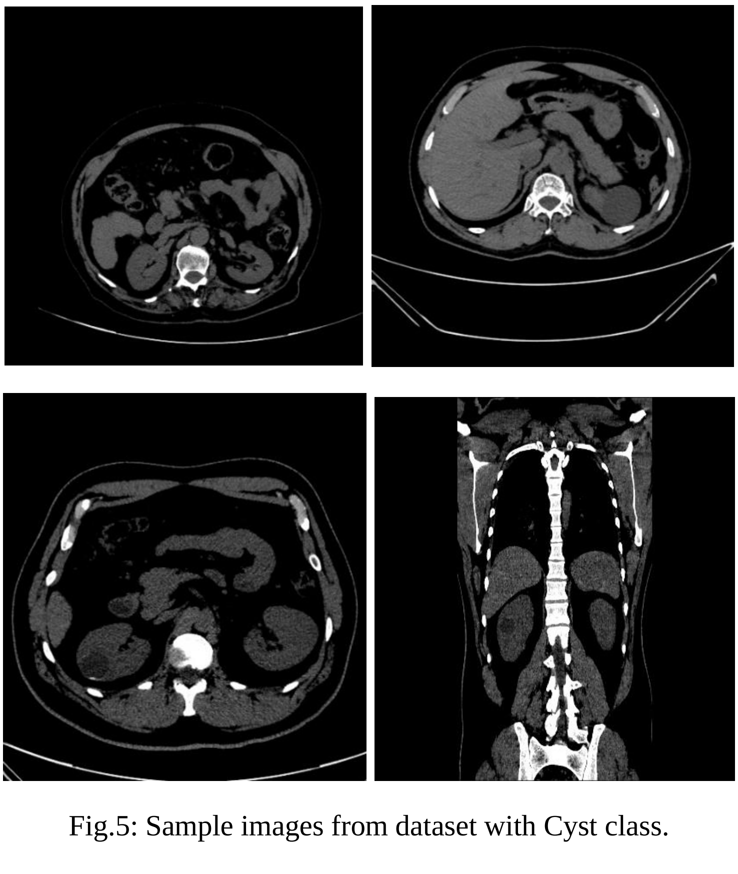

Fig.5: Sample images from dataset with Cyst class.

4.2 Results and description:

Fig.6: Metrics of Existing SVM

Figure 6 show that he existing Support Vector Classifier (SVC) model demonstrates exceptional performance in classifying kidney structural anomalies, achieving an outstanding accuracy of 99.92%. It maintains a high level of reliability with a precision of 99.93%, indicating its effectiveness in correctly identifying positive cases such as cysts, stones, or tumors. The recall score of 99.83% reflects its strong capability to detect nearly all true positive instances, minimizing false negatives.

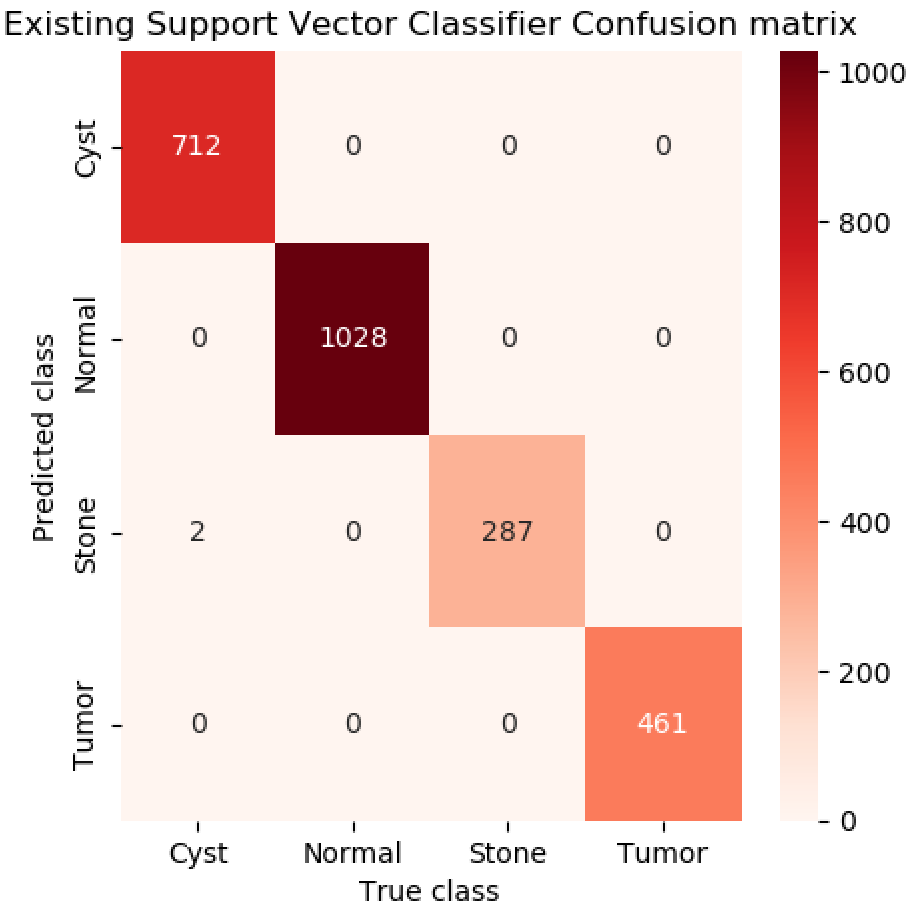

Fig.7: CM of Existing support vector classifier

Figure 7 shows that confusion matrix for an existing Support Vector Classifier (SVC). The matrix reveals the model's performance across four classes: Cyst, Normal, Stone, and Tumor. For the Cyst class, it correctly predicted 712 instances and misclassified 0. Similarly, for the Normal class, it correctly predicted 1028 instances with no misclassifications. The Stone class had 2 instances incorrectly predicted as Cyst and 287 correctly predicted.

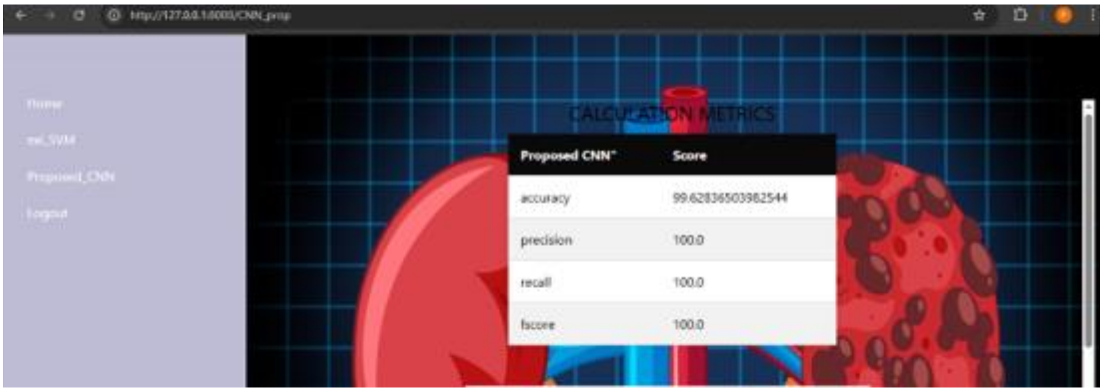

Fig.8: Metrics of CNN

Figure 8 shows that the performance metrics of a proposed Convolutional Neural Network (CNN). The CNN achieved an accuracy of approximately 99.63%, indicating a high overall correctness in its predictions. Furthermore, the model demonstrated perfect precision, recall, and F1-score, all registering at 100.0%.

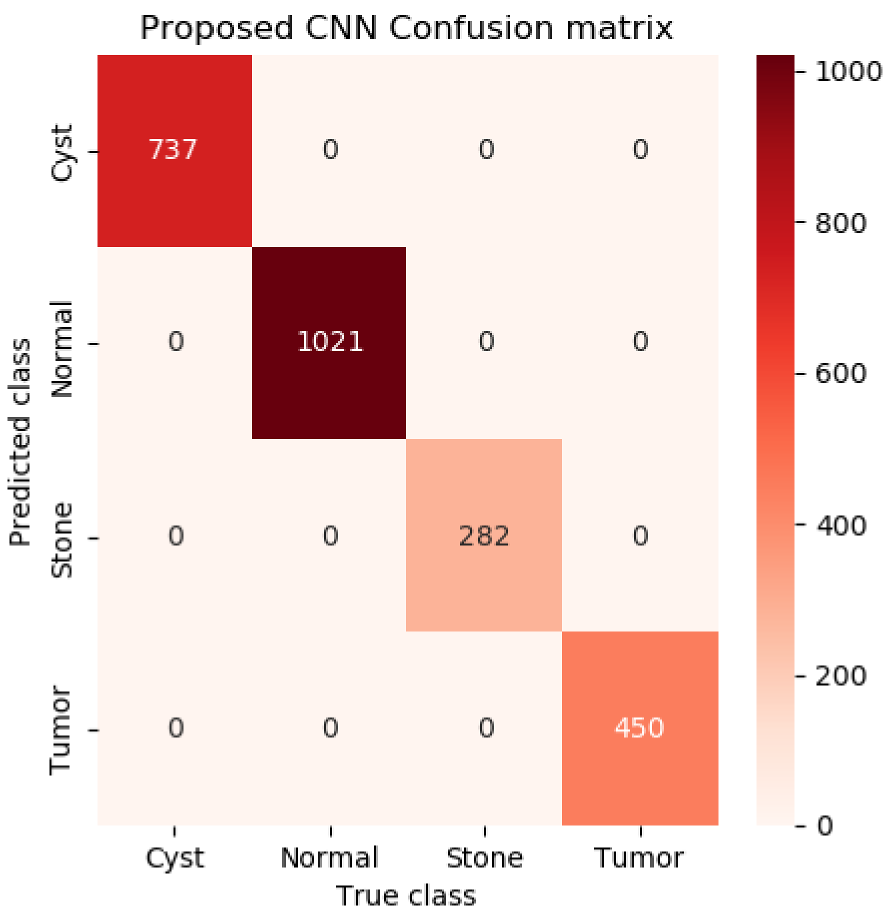

Fig.9: CM of the Proposed Algorithm

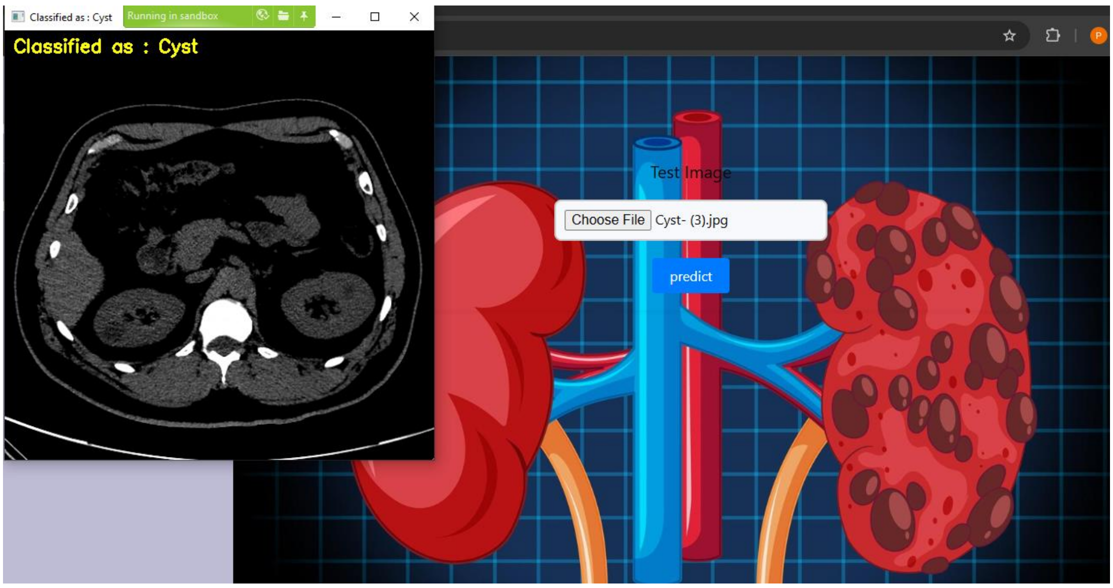

Fig.10: Predicted as Cyst

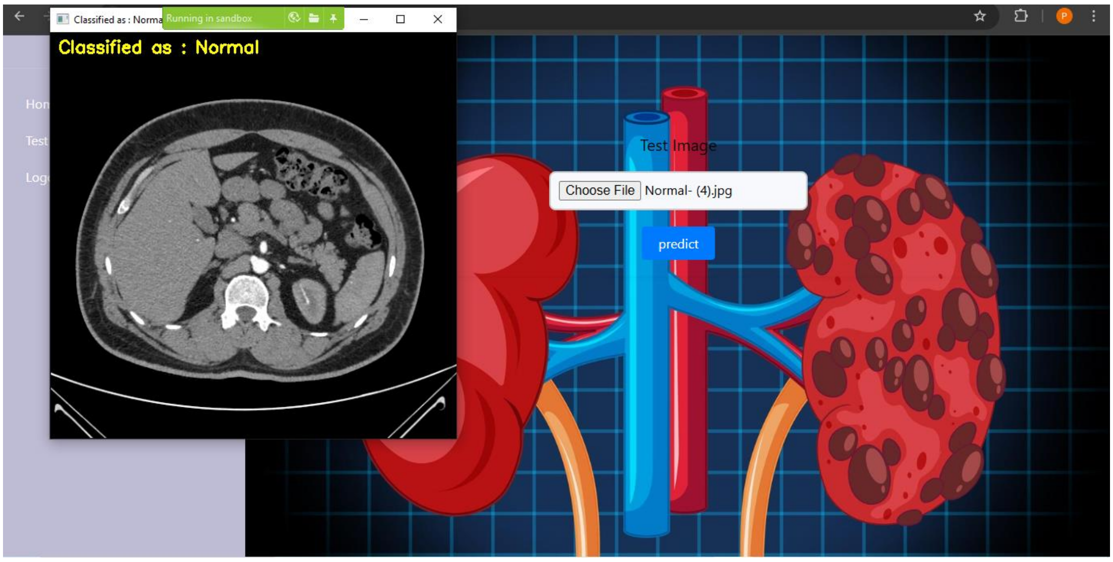

Fig.11: Predicted as Normal

5. CONCLUSION

The proposed system for deciphering anomalies in kidney structures using machine learning and deep learning techniques has demonstrated remarkable effectiveness in identifying and classifying conditions such as cysts, stones, tumors, and normal tissues. The existing Support Vector Classifier (SVC) model showed excellent performance, with an impressive accuracy of 99.92%, precision of 99.93%, recall of 99.83%, and F1-score of 99.88%, confirming its ability to perform accurate, balanced classification across multiple renal conditions. The confusion matrix further validated its strength, showing near-perfect classification, particularly for the Cyst and Normal categories, and minimal misclassifications for the Stone class. In parallel, the proposed Convolutional Neural Network (CNN) model also achieved commendable performance, with an accuracy of 99.63%, and perfect scores (100%) in precision, recall, and F1-score. This highlights the model's robust generalization and superior performance, especially in dealing with complex image-based medical data. The CNN-based model correctly classified all test instances as shown in its confusion matrix, suggesting its high reliability for real-time or automated diagnostic applications.

REFERENCES

[1] V. Jha, G. Garcia-Garcia, K. Iseki et al., “Chronic kidney disease: global dimension and perspectives,” The Lancet, vol. 382, no. 9888, pp. 260–272, 2013.

[2] R. Ruiz-Arenas, “A summary of worldwide national activities in chronic kidney disease (CKD) testing, the electronic journal of the international federation of,” Clinical Chemistry and Laboratory Medicine, vol. 28, no. 4, pp. 302–314, 2017.

[3] Thedailystar, “Over 35,000 develop kidney failure in Bangladesh every year,” 2019, https://www.thedailystar.net/city/news/18m-kidney-patients-bangladesh-every-year-1703665.

[4] Prothomalo, “Women more affected by kidney diseases,” 2018, https://en.prothomalo.com/bangladesh/Womenmore-affected-by-kidney-diseases.

[5] Scottish Intercollegiate Guidelines Network (Sign), Diagnosis and Management of Chronic Kidney Disease: A National Clinical Guideline, SIGN, Victoria, Australia, 2008.

[6] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart disease prediction using hybrid machine learning model," in Proceedings of the 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, January 2021.
[7] P. Ghosh, F. M. Javed Mehedi Shamrat, S. Shultana, S. Afrin, A. A. Anjum, and A. A. Khan, "Optimization of prediction method of chronic kidney disease using machine learning algorithm," in Proceedings of the 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Bangkok, Thailand, November 2020.
[8] K. R. A. Padmanaban and G. Parthiban, "Applying machine learning techniques for predicting the risk of chronic kidney disease," Indian Journal of Science and Technology, vol. 9, no. 29, 2016.

[9] Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, S. Suwannawach, and N.

Ninchawee, "Predictive analytics for chronic kidney disease using machine learning techniques," in Proceedings of the 2016 Management and Innovation Technology International Conference (MITicon), Bang-San, Thailand, October 2016.

[10] G.-S. Fu, Y. Levin-Schwartz, Q.-H. Lin, and D. Zhang, "Machine learning for medical \

imaging," Journal of healthcare engineering, vol. 2019, pp. 1-2, 2019.

[11] R. Devika, S. V. Avilala, and V. Subramaniyaswamy, "Comparative study of classifier for

chronic kidney disease prediction using naive Bayes, KNN and random forest," in Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, March 2019

[12] S. Revathy, B. Bharathi, P. Jeyanthi, and M. Ramesh, "Chronic kidney disease prediction using

machine learning models," International Journal of Engineering and Advanced Technology, vol. 9, no. 1, pp. 6364–6367, 2019.

[13] S. A. Rabby, R. Mamata, M. A. Laboni, Ohidujjaman, and S. Abujar, "Machine learning

applied to kidney disease prediction: comparison study," in Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, July 2019.

[14] M. Nishat, F. Faisal, R. Dip et al., "A comprehensive analysis on detecting chronic kidney

disease by employing machine learning algorithms," EAI Endorsed Transactions on Pervasive Health and Technology, vol. 7, Article ID 170671, 2018.

[15] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia, and J. Gutierrez, "A

comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease," in Proceedings of the 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, Greece, July 2017.