Automatic Detection of Chronic Kidney Disease Risk: A Vision Transformer Based Approach

XXXXX[1]

[1]Business and Information Systems – University of South-Eastern Norway (USN)
Borre – Norway

XXXX

Abstract. *Chronic kidney diseases (CKD) cause millions of deaths worldwide every year. A widely used method to diagnose these diseases is renal scintigraphy. However, there's a shortage of studies applying transformer based neural network architectures to automate CKD stage prediction. More than that, there's also a lack of studies exploring the effects of using sampling algorithms on the performance of the trained models. Thus, this work aims to use a Visual Transformer encoder combined with several underlying classification algorithms and compare it against established methods applied to solve the such problem. Our results show that our approach surpasses the current state-of-the-art by almost 31%. Also, they show that the use of sampling algorithms could bring gains in accuracy up to 2%.*

1. Introduction

Chronic kidney diseases (CKD) cause millions of deaths worldwide every year. Previous researches on the topic showed that CKDs are among the diseases that cause most deaths worldwide [Vos et al. 2016]. In Norway, this statement continuous as a fact. The total CKD prevalence in Norway was 10,2% (SE 0.5): CKD stage 1 (GFR 90 ml/min per 1.73 m^2 and albuminuria), 2.7% (SE 0.3); stage 2 (GFR 60 to 89 ml/min per 1.73 m^2 and albuminuria), 3.2% (SE 0.4); stage 3 (GFR 30 to 59 ml/min per 1.73 m^2), 4.2% (SE 0.1); and stage 4 (GFR 15 to 29 ml/min per 1.73 m^2), 0.2% (SE 0.01) [Hallan et al. 2006].

A widely used method for diagnosing these diseases is renal scintigraphy. When put in contrast with the more manual approach, which is measuring the creatine elimination taking blood and urine exams over a period of 24 hours. The renal scintigraphy method shows two advantages. The first one, clearly, is that is less invasive. The second one, however, is that the renal scintigraphy exam is able to present the body activity on ROI regions [Kerl and Cook 2005].

Nevertheless, the whole process, for being very long and having a lot of variables to be analyzed, can be very error prone. Many studies, nonetheless, have addressed this issue. [Rebouças Filho et al. 2019], for example, attempted to solve the problem using computer vision methods such as LBP to extract meaningful features from a set of scintigraphic images and then pass the resulting vector to machine learning algorithms to generate a predictive model. However, this study only presents the model scores on the training set, which is a methodological flaw.

That and the statistics on CDK in Norway make crucial to develop and provide systems able to automatize the whole detection process. Even though previous research

have shown great advancement in this area, this paper improves upon them further by investigating a new way of pre-processing and encoding the scintigraphic images extracted from the patients in order to improve the prediction quality achieved. Also, we explore the use of sampling methods for dealing with the inherent unbalanced nature of the data. Even further, we improve the methodology experimentation from the previous research by using 5-fold cross-validation and macro and weighted metrics.

In this context, this paper proposes an end-to-end system that can automatically identifies the stage of CKD based on the scintigraphic images taken from the patient's kidney, informing whether a patient needs specialized care because they are at risk of developing CKD. Being more specific, this work introduces a set of computer vision based machine learning models trained to extract meaningful features from scintigraphy exams and classify them in early or later CKD stages. The proposed embedding approach on the best classifier achieved 88.38% accuracy and 88.24% f-score, showing an improvement of 18% over the accuracy reached in previous efforts by [Rebouças Filho et al. 2019]. These results sets the ground for exploring new ways of representing and processing these sorts of medical images and improving further their prediction's quality.

2. Related Works

The automated Chronic Kidney Disease prediction is an extensively studied topic in literature. The methods applied to solve this problem goes from classic machine learning methods, like support vector machines and random forests, to deep learning models like Convolutional and Long-Short Term Memory Neural Networks.

[Khan et al. 2024] for example, executed a comprehensive study using several traditional machine learning algorithms like logistic regression, random forest, decision tree, k-nearest neighbor, and support vector machine, trying to perform an early prediction of CKD. Their results pointed the Support Vector Machines as the best performing model across all the scenarios tested. Along the same lines, [Revathy et al. 2019], worked with the Chronic Disease dataset within the UCI machine learning repository to predict CKD in its early stages. The authors also try to provide the best prediction framework for such a task. Contrary to [Khan et al. 2024], though, their results showed that the Random Forest Classifier was superior to the other methods, which could possibly show a dataset performance dependency in this task.

Deep learning methods, on the other hand, are also widely applied as a tool to solve this problem. As instance, [Yildiz et al. 2023] develop a hybrid model based on CNN (Convolutional Neural Network) and Long-Short Term Memory (LSTM) to be trained on a binary CKD dataset. Their approach was tested against several machine learning methods, like K-Nearest Neighbors and Logistic Regression. Although the difference between the three best models, the authors approach achieved the highest accuracy score among them.

On the other hand, [Mondol et al. 2022], explored not the use of a hybrid model, but a series of other deep learning techniques: CNN, ANN, LSTM, and their optimized versions, separately. Their study attempted to study comprehensively the performance of these algorithms. The work's results showed that, although the OCNN, OANN, and OLSTM achieved higher accuracy scores (98.75%, 96.25%, and 98,5%, respectively) among the original versions of each model, the CNN reached the highest score: 92.71%.

Although, there is a myriad of works exploring a great variety of machine learning algorithms, that variety not only is applied to the techniques. The same is true for the kind of data. [Vishnupriya et al. 2023], as instance, used 28x28 Magnetic Resonance Imaging (MRI) pictures of the kidneys to train model to predict chronic renal failure using evolutionary and deep learning algorithms. [Tissera et al. 2023], on the other hand, collected radiology images and used a CNN to analyzed them in order to provide automated and personalized CKD risk assessment. Along the same lines, [Rebouças Filho et al. 2019] and [de Alexandria et al. 2021] dealt with the processing of scintigraphy exam images to classify the stage of CDK. For this, both works used several machine learning classifiers and also employed an specific pre-processing pipeline for each set of scintigraphic images.

Although, these previous works explored deeply the use of machine learning and deep learning algorithms to predict CKD stage and risk, they didn't explore the use of transformer based models as a solution for such a task. Vision Transformers Models have shown great performance on medical image analysis tasks ([Dalmaz et al. 2022], [Manzari et al. 2023]), which raises the urge to explore its use on CKD early prediction and classification. Also, none of those studies investigated the use of sampling algorithms to improve data quality for scintigraphy datasets. Therefore, in this study we aimed to explore the use of such tools in order to provide better and more sophisticated methods for automatizing CKD detection.

3. Research Questions

In previous sections we stated that given the statistics on CKD worldwide and, more specifically in Norway, it has become crucial the development of systems that are capable of automatizing if not all the process of detecting CKD stages, at least, parts of it. Although several works have tried to use a myriad of methods trying to achieve this, using MRI, radiology images or even ultrasound images ([Vishnupriya et al. 2023], [Tissera et al. 2023], [Sudharson and Kokil 2020]), they did not experiment with transformer based architectures. Thus this paper proposes the use of ViT (Vision Transformers) feature extracting method to applied to pre-processed scintigraphy images. This raises our first question:

> **Research Question 1**
> *To what extend can transformer based neural network architectures accurately and automatically predict the CKD stage based on the certain set of patients' exams?*

They also didn't explore the use of sampling algorithms to improve data quality, which brings up our second research question:

> **Research Question 2**
> *How much the use of sampling methods can improve the algorithms' performance on the CDK stage prediction task*

4. Methodology

This section aims to describe each of the steps of the experiments conducted in this study, ranging from data collection to the training and validation of the resulting models.

4.1. Data

The dataset used in this study was the Database of Dynamic Renal Scintigraphy, a collection of anonymized patient studies and simulated data for clinical audits, inter-laboratory comparisons, development and testing of analytical methods and clinical software, as well as for teaching-learning purposes and training. This database includes five different datasets: one for adults, one for children, and three simulated datasets. The dataset used in [Rebouças Filho et al. 2019] was the DRSPRG, a series of 115 scintigraphic studies in adults. In this dataset, studies 3, 30, 33, 36, 43, 51, 103, and 109 were excluded as they were deemed inadequate for lacking image data, as we needed at least 180 images, leaving 107 studies. Each of these studies was previously classified into five stages of chronic kidney disease (CKD) based on estimated blood GFR.

In the DRSPRG dataset, each examination lasted thirty minutes, and the stored examination consists of 180 images of size 128x128 pixels captured every 10 seconds throughout the examination procedure. Both posterior and anterior images were captured as well. It was noted that 5 patient examinations did not have all 180 images, namely patients numbered 25 (CDK 1), 55 (CKD 2), 65 (CKD 2), 81 (CKD 3), and 106 (CDK 5); these examinations were removed from the experiments to maintain standardization [Rebouças Filho et al. 2019].

Patients in stage 3 of chronic kidney disease have higher mortality risks, among other comorbidities [Levey et al. 2003], such as bone diseases, in addition to being more sensitive to drug side effects [Harris and Stribling 2007]. For this reason, in [Rebouças Filho et al. 2019], they chose to work with only two classes. Class 0 comprises stages 1 and 2, and class 1 comprises stages 3, 4, and 5. Another reason for this may have been the data imbalance associated with the reduced number of samples per stage. In Figure 1, you can see the frequency of each CKD stage throughout the studies.

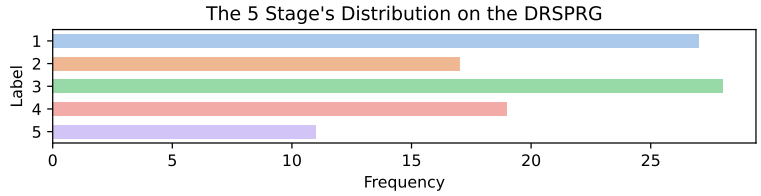

Figure 1. Frequency of each stage of CKD throughout the scintigraphic studies.

Thus, the final model will indicate whether the patient needs specialized monitoring due to possible disease progression, which is indicated by class 1, or not, represented by class 0. Therefore, as we can see in Figure 2, there are, at the end of the procedure, a distribution of 44 samples for class 0 and 58 for class 1.

4.2. Data Preparation

The dataset described in Subsection 4.1 underwent a series of processing steps to be used as input for computer vision and machine learning algorithms, which will be discussed in

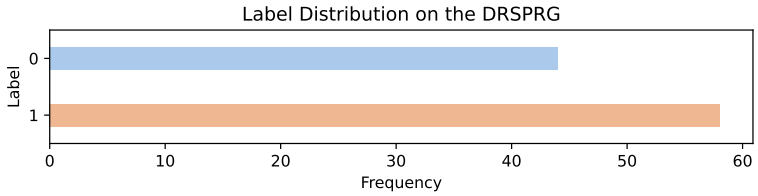

Figure 2. Distribution of the classes determined throughout the scintigraphic studies.

Subsections 4.3 and 4.4.

The processing steps applied to the dataset were heavily based on the ones used in [Rebouças Filho et al. 2019]. All 180 images of each patient examination underwent a blurring process using a mean filter. This process aims to remove noise present in the images and, as a result, smooths the image edges. In Figure 3, you can see a sample before and after blurring.

After that, the set of 180 images was divided into 15 subsets, each representing a 2-minute frame of the examination, containing 12 images each. According to [Rebouças Filho et al. 2019], this 2-minute period is a common practice in the literature, so it was reproduced in this study. Finally, after this division, an average image of all 15 subsets was generated. Equation 1 shows how it was created. In this equation, n represents the total number of images in the subset, and x represents an image instance. In Figure 4, you can observe one of the resulting images from this process. At the end of this process another mean operation was performed.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

4.3. Feature Extraction

In the data preparation step, detailed in Subsection 4.2, the studies were divided into 15 subgroups of 12 images, where each subgroup was reduced to a single image through averaging operation. After this, each patient examination was represented by a sequence of 15 scintigraphic images. However, to be suitable for a machine learning model, they still needed to undergo a feature extraction step.

The feature extraction step aims to transform the data so that it can be used by learning algorithms. In this study, we aimed to use a great variety of image descriptor algorithms to encode them and make them suitable for machine learning training. Having said that, Local Binary Patterns (LBP), Gray-Level Co-Occurrence Matrix (GLCM), Zernike's Moments (ZN), and Visual Transformer were used to extract significant features from each of the mean image taken from the 15 images representing the patient's scintigraphy examination. As this algorithm requires the images to be in a non-RGB state, a binarization process was performed.

First, all images were converted to grayscale. Then, the binarization process was applied. The method used for this was Otsu's Thresholding, which can automatically identify the minimum and maximum thresholds for image binarization. In Figure 5, you

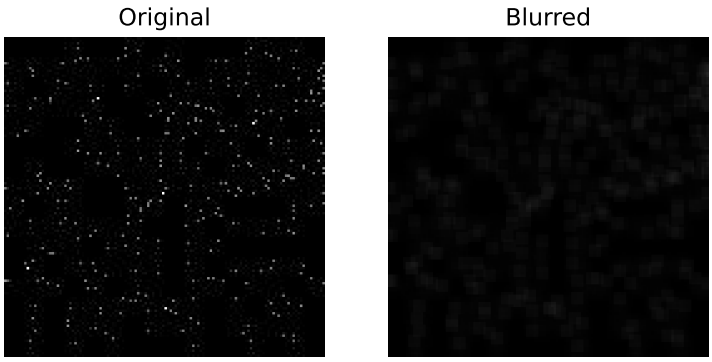

Figure 3. Scintigraphy image before and after the blurring.

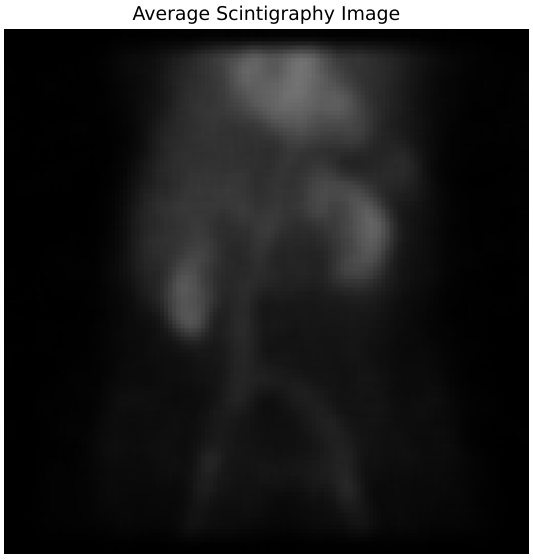

Figure 4. Scintigraphy image after the averaging procedure.

can observe the final result for one of the images. After this procedure, the image was ready for the encoder algorithms.

For each resulting image, a vector of 108 floats characterizing the texture of the image using LBP was generated. The values of radius and sampling pixels for the LBP algorithm were 1 and 106, respectively. These values were used to try to approximate the

Figure 5. Mean Binarized Scintigraphic Image

number of features that were extracted according to [Rebouças Filho et al. 2019] in their experiments, as the authors did not clarify which configuration was used.

To the GLCM, though, the resulting vector was made of only four float numbers. These numbers represented the contrast, correlation, energy, and homogeneity, respectively. This algorithm took as hyper-parameters a list of distances containing only 1 as an element, a list of angles with only 0 as an element, a integer levels set as 256, a boolean property named symmetric as true, and, finally, another boolean property, called normalized, set, also, as true.

The ViT technique, on the other hand, resulted in a vector of 768 dimensions. For achieving this result we used the base version of a pre-trained model, the `google/vit-base-patch16-224`. This model was trained on a myriad of internet images using a transformer architecture. Its goal is to leverage both the power of transformer architecture and transfer learning.

Finally, the ZN method was also used to encode the images. Its configuration was set as follows: radios of 21 and degree of 8. This configuration results in a vector's size of 15 float numbers. The purpose of this algorithm is to be used to extract shape and texture features from images. Zernike's Moments are particularly useful due to their rotation invariance, meaning the moments remain the same even if the image is rotated.

4.4. Model Selection

Once the data went through the preparation and feature extraction processes, detailed in Subsections 4.2 and 4.3, the experiment entered the model selection phase. In this stage, the aim is to develop predictive models using machine learning techniques.

Thus, the first step was to define our experimental ground by setting the how the dataset (Subsection 4.1) would be used in the training and evaluating steps. For this we decided to go for the use of a 5-fold cross-validation technique. This would ensure a

good variety of training and testing sets while at the same time providing a more light computational burden to run more heavy algorithms like ViT, which needs GPU to be executed smoothly.

The experiments were then conducted combining the Random Forest, Support Vector Machines, Extra Trees, Decision Tree and Multi-Layer Perceptron learning algorithms to the ViT, ZN, GLCM, and LBP image descriptors. The image descriptors acted as encoders, being responsible to transform the scintigraphy images into meaningful feature vectors that could be later used by algorithms to fit the models that would then predict the target patient's CKD stage.

Additionally, it's worth noting that all experiments in [Rebouças Filho et al. 2019] were conducted on the training set. This is not the most appropriate way to validate a predictive model, as it doesn't allow for detecting numerous issues such as over-fitting. It can be assumed that a holdout set was not produced. Therefore, in this study, it was decided to improve the experimental methodology by adopting a 5-fold cross-validation method to evaluate the trained models. By doing such a thing we aim at approximate the model to a more realistic scenario and detect issues such as over-fitting.

4.5. Results Evaluation

In [Rebouças Filho et al. 2019], the authors adopted a binary classification approach to solve the problem of chronic kidney disease detection. When working with binary classification, the following measures are commonly used: accuracy, precision, recall, and F1 score [Kowsari et al. 2019]. Accuracy, as defined by [Hossin and Sulaiman 2015], measures the proportion of correct positive and negative classifications relative to the total evaluated samples. Precision is a statistical metric that indicates the percentage of samples correctly classified as positive out of all those that were classified as positive. Recall shows the percentage of positive samples that were correctly classified. Finally, the F1 score is the harmonic mean between these last two metrics, providing an overall idea of the model's performance. Formulas 3 and 4 show how precision, recall, and F1 score were calculated, respectively. It's also important to mention that we are taking in consideration the macro and weighted versions of each metric, so we can take a better look searching for things like over or under-fitting.

$$\textit{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN} \tag{3}$$

$$\textit{F-measure} = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

5. Results

The experiments conducted in this study had two distinct objectives. The first was to determine whether, and to which extent, it was possible to use transformer based neural networks to automatically detect if a patient needs specialized monitoring due to the risk of CKD progression. The second objective was to investigate the effects of applying a

sampling algorithm to improve the target dataset's balance. We wanted to see whether it would improve the algorithm's performance and, if the answer was yes, to which degree. As a side contribution we provided with this study a comparative analysis of the performance of different methods in CKD prediction using k-fold cross-validation to bring reliable results. In this section, therefore, the achieved results are presented.

The algorithms described in Section 4, training pipelines for SVM, Random Forest, Decision Tree, Multi-Layer Perceptron, and Extra Trees were implemented, executed, and evaluated using, as encoding techniques the following methods: Local Binary Patterns (LBP), Vision Transformer (ViT), Gray-Level Co-Occurrence Matrix (GLCM), and Zernike's Moments. Table 1 shows the scores obtained by each one of those combinations of algorithm and encoder technique in each one of the metrics outlined in the Subsection 4.5 for the unbalanced dataset.

Looking at this table we can see that the best score overall was achieved by the ViT encoder alongside the MLP algorithm. It reached about 87.3% accuracy. When we compare this result with the approach proposed by [Rebouças Filho et al. 2019], it's very clear that our approach surpass theirs by a large degree, almost 31%. Other methods like combining RF with GLCM achieved good results and only get behind the main approach by a margin of about 2%. Showing that GLCM proved itself as a very performative way of extracting meaningful features from scintigraphic exam set of images. The cannot be said about the LBP and ZM encoder techniques.

Algorithm	Encoder	Acc	Prec_m	Prec_w	Recall_m	Recall_w	F_{1_m}	F_{1_w}
SVM	**LBP**	0.568571	0.284286	0.323612	0.500000	0.568571	0.362390	0.412363
RF	**LBP**	0.666667	0.664177	0.667619	0.656439	0.666667	0.658317	0.665348
DT	**LBP**	0.518095	0.515859	0.527201	0.516667	0.518095	0.511955	0.518355
ET	**LBP**	0.685714	0.689899	0.688747	0.668308	0.685714	0.670416	0.679244
MLP	**LBP**	0.568571	0.284286	0.323612	0.500000	0.568571	0.362390	0.412363
SVM	**ViT**	0.843810	0.849817	0.846845	0.833207	0.843810	0.837917	0.841896
RF	**ViT**	0.844286	0.859490	0.854757	0.831187	0.844286	0.835902	0.840606
DT	**ViT**	0.744762	0.768294	0.766052	0.727652	0.744762	0.726649	0.735612
ET	**ViT**	0.824286	0.826175	0.825288	0.815783	0.824286	0.819191	0.823099
MLP	**ViT**	**0.873333**	**0.877762**	**0.882808**	**0.876136**	**0.873333**	**0.872374**	**0.873579**
SVM	**GLCM**	0.636190	0.658634	0.648157	0.586667	0.636190	0.538751	0.566398
RF	**GLCM**	0.853333	0.854584	0.853919	0.846338	0.853333	0.849124	0.852368
DT	**GLCM**	0.695714	0.701159	0.708766	0.696843	0.695714	0.691025	0.694571
ET	**GLCM**	0.813810	0.817399	0.825315	0.815657	0.813810	0.810462	0.813344
MLP	**GLCM**	0.814762	0.817696	0.819869	0.808001	0.814762	0.808788	0.813446
SVM	**ZM**	0.677143	0.698456	0.688646	0.630556	0.677143	0.583385	0.608364
RF	**ZM**	0.606667	0.602564	0.611859	0.598106	0.606667	0.588826	0.597986
DT	**ZM**	0.567619	0.558524	0.568663	0.560606	0.567619	0.551727	0.560473
ET	**ZM**	0.598571	0.587493	0.596665	0.589141	0.598571	0.585366	0.594524
MLP	**ZM**	0.658571	0.682061	0.678034	0.633207	0.658571	0.615922	0.634436

Table 1. Metric scores for several classifiers and encoders tested for CKD prediction against the unbalanced dataset.

The Table 2, on the other way, shows the scores for the encoders and algorithms combinations when trained against the balanced dataset. As explained before the chosen sampling algorithm was the Random Over-Sampling, because we wanted to raise the

number of samples in the under-sampled class and avoid over-fitting and improving generalization. For this scenario, looking at the table, we can get to the conclusion that the best performing method was the SVM using the ViT as the encoder technique. This result sets the ViT as the best way to extract meaningful features from scintigraphic exam images. When we compare this approach to the previous winning one, MLP with ViT, we see that it surpass it by a margin of 2%. That could possibly mean that the SVM has a great sensibility to unbalanced data and it improves to a great extent once this is solved somehow. The approach proposed by [Rebouças Filho et al. 2019] also improved, but is still way behind our approach, by a distance of almost 18%.

Algorithm	Encoder	Acc	Prec_m	Prec_w	Recall_m	Recall_w	F_{1_m}	F_{1_w}
SVM	**LBP**	0.707619	0.708901	0.712014	0.699747	0.707619	0.700581	0.706195
RF	**LBP**	0.637619	0.650803	0.650510	0.624495	0.637619	0.625034	0.632801
DT	**LBP**	0.518095	0.519722	0.528147	0.517424	0.518095	0.513604	0.518145
ET	**LBP**	0.695714	0.699693	0.698331	0.677399	0.695714	0.679612	0.688707
MLP	**LBP**	0.431429	0.215714	0.186469	0.500000	0.431429	0.301281	0.260296
SVM	**ViT**	**0.883810**	**0.889824**	**0.893847**	**0.885985**	**0.883810**	**0.882480**	**0.883919**
RF	**ViT**	0.854286	0.859582	0.858376	0.845707	0.854286	0.849006	0.852973
DT	**ViT**	0.803333	0.828999	0.826245	0.789394	0.803333	0.786868	0.794011
ET	**ViT**	0.824286	0.832967	0.829597	0.813724	0.824286	0.817684	0.822229
MLP	**ViT**	0.863810	0.869580	0.875795	0.867803	0.863810	0.862895	0.864098
SVM	**GLCM**	0.684286	0.683668	0.685719	0.667677	0.684286	0.661464	0.671620
RF	**GLCM**	0.843333	0.841443	0.843804	0.833810	0.843333	0.835802	0.842560
DT	**GLCM**	0.774762	0.717738	0.724775	0.714899	0.774762	0.709733	0.713203
ET	**GLCM**	0.813810	0.811792	0.815214	0.809470	0.813810	0.809443	0.813323
MLP	**GLCM**	0.814762	0.817504	0.823824	0.814286	0.814762	0.812211	0.813840
SVM	**ZM**	0.658095	0.677552	0.671919	0.617929	0.658095	0.594644	0.615864
RF	**ZM**	0.567619	0.573917	0.585211	0.570833	0.567619	0.562863	0.566782
DT	**ZM**	0.558095	0.540815	0.549243	0.535510	0.558095	0.532899	0.547917
ET	**ZM**	0.558571	0.557498	0.569188	0.557702	0.558571	0.551018	0.557313
MLP	**ZM**	0.658095	0.658824	0.658819	0.629672	0.658095	0.621163	0.636620

Table 2. Metric scores for the several classifiers and encoders tested for CKD prediction against the balanced dataset.

Comparing the results from both tables, we can get to at least two conclusions. The first is that ViT is the best way of encoding medical scintigraphic exam images. The second, on the other hand, is that using sampling algorithms could bring great improvements to the methods, and specifically the use of Random Over-Sampling could bring up to almost 2% improvement in the results. When observing the accuracy scores of all the models, this becomes even more evident. With this in mind, it is not unreasonable to conclude that not only is ViT with SVM is the best method among those analyzed, but also that it is possible to improve the performance of those methods for detecting the presence of CKD in patients using sampling techniques.

6. Discussion

The first research question presented in this work in Section 3 had the goal of investigating the possibility of using transformers based architectures to predict CKD stages using scintigraphy images and explore to which degree this would be effective compared to other established methods in literature. The experimental results show that using the

Visual Transformer (ViT), which attends such criteria, could not only perform the given task and give accurate predictions but also that its use brings high gains of performance when compared with other approaches used in previous research efforts.

On the other hand, our second research question sought to raise the question of whether the use of sampling algorithms being applied to improve data quality by reducing the gap in numbers between the major class in the dataset and the minor one. Our experimental results proved that it could in fact bring gains in performance in the overall performance of the methods tested against the results dataset. When using Random Over-Sampling, more specifically, this improvement could be up to 2%, but this could vary depending on the level of unbalance in the data and also the number of samples for each class evaluated.

7. Conclusion

This work aimed at developing an automatic detection system for chronic kidney diseases (CKD). Experimental studies are presented to demonstrate how the proposed approach successfully classifies patients into two groups: one for patients without risk and another for patients at risk of developing CKD, hence requiring specialized monitoring. The experimental results showed that our approach surpass the state-of-the-art for CKD prediction on scintigraphic exam images, beating the previous efforts made by [Rebouças Filho et al. 2019]. They also proved the benefit of using sampling algorithms to improve data quality by reducing the dataset's unbalance and diminishing the chances of over-fitting. For future work, it is intended to experiment with hybrid methods that make use of transformer based architectures and also increase the dataset by adding synthetic data.

References

Dalmaz, O., Yurt, M., and Çukur, T. (2022). Resvit: residual vision transformers for multimodal medical image synthesis. *IEEE Transactions on Medical Imaging*, 41(10):2598–2614.

de Alexandria, A. R., Ferreira, M. C., Ohata, E. F., Cavalcante, T. D. S., Da Mota, F. A. X., Nogueira, I. C., Albuquerque, V. H. C., Gondim, V. J. T., and Neto, E. C. (2021). Automated classification of dynamic renal scintigraphy exams to determine the stage of chronic kidney disease: an investigation. In *2021 3rd International Conference on Research and Academic Community Services (ICRACOS)*, pages 305–310. IEEE.

Hallan, S. I., Coresh, J., Astor, B. C., Arne, A., Powe, N. R., Romundstad, S., Hallan, H. A., Lydersen, S., Holmen, J., et al. (2006). International comparison of the relationship of chronic kidney disease prevalence and esrd risk. *Journal of the American society of nephrology*, 17(8):2275–2284.

Harris, K. and Stribling, B. (2007). Automated estimated gfr reporting: A new tool to promote safer prescribing in patients with chronic kidney disease? *Therapeutics and clinical risk management*, 3(5):969–972.

Hossin, M. and Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, 5(2):1.

Kerl, M. E. and Cook, C. R. (2005). Glomerular filtration rate and renal scintigraphy. *Clinical techniques in small animal practice*, 20(1):31–38.

Khan, N., Raza, M. A., Mirjat, N. H., Balouch, N., Abbas, G., Yousef, A., and Touti, E. (2024). Unveiling the predictive power: a comprehensive study of machine learning model for anticipating chronic kidney disease. *Frontiers in Artificial Intelligence*, 6:1339988.

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., and Brown, D. (2019). Text classification algorithms: A survey. *Information*, 10(4):150.

Levey, A. S., Coresh, J., Balk, E., Kausz, A. T., Levin, A., Steffes, M. W., Hogg, R. J., Perrone, R. D., Lau, J., and Eknoyan, G. (2003). National kidney foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Annals of internal medicine*, 139(2):137–147.

Manzari, O. N., Ahmadabadi, H., Kashiani, H., Shokouhi, S. B., and Ayatollahi, A. (2023). Medvit: a robust vision transformer for generalized medical image classification. *Computers in Biology and Medicine*, 157:106791.

Mondol, C., Shamrat, F. J. M., Hasan, M. R., Alam, S., Ghosh, P., Tasnim, Z., Ahmed, K., Bui, F. M., and Ibrahim, S. M. (2022). Early prediction of chronic kidney disease: A comprehensive performance analysis of deep learning models. *Algorithms*, 15(9):308.

Rebouças Filho, P. P., da Silva, S. P. P., Almeida, J. S., Ohata, E. F., Alves, S. S. A., and Silva, F. d. S. H. (2019). An approach to classify chronic kidney diseases using scintigraphy images. In *Anais Estendidos do XXXII Conference on Graphics, Patterns and Images*, pages 156–159. SBC.

Revathy, S., Bharathi, B., Jeyanthi, P., and Ramesh, M. (2019). Chronic kidney disease prediction using machine learning models. *International Journal of Engineering and Advanced Technology*, 9(1):6364–6367.

Sudharson, S. and Kokil, P. (2020). An ensemble of deep neural networks for kidney ultrasound image classification. *Computer Methods and Programs in Biomedicine*, 197:105709.

Tissera, W., Rathnayake, S., Marasinghe, M., Isurika, W., Perera, J., and Samarawila, D. (2023). Kidnify–elevating chronic kidney disease management with machine learning and iot through a mobile application. *International Research Journal of Innovations in Engineering and Technology*, 7(11):254.

Vishnupriya, B., Abitha, R., Harshithaa, A., Agalya, M., and Gugapriya, M. (2023). A perlustration on the optimistic prognosis of chronic renal failure using evolutionary machine learning and deep learning techniques. In *2023 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–9. IEEE.

Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., Carter, A., Casey, D. C., Charlson, F. J., Chen, A. Z., et al. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015. *The lancet*, 388(10053):1545–1602.

Yildiz, E., Cengil, E., Yildirim, M., and Bingol, H. (2023). Diagnosis of chronic kidney disease based on cnn and lstm. *Acadlore transactions on ai and machine learning*, 2(2):66–74.