



HAL
open science

Combining Convolution and Involution for the Early Prediction of Chronic Kidney Disease

Hadrien Salem, Sarah Ben Othman, Marc Broucqsault, Slim Hammadi

► **To cite this version:**

Hadrien Salem, Sarah Ben Othman, Marc Broucqsault, Slim Hammadi. Combining Convolution and Involution for the Early Prediction of Chronic Kidney Disease. 24th International Conference on Computational Science (ICCS 2024), Jul 2024, Malaga, Spain. 10.1007/978-3-031-63772-8_24 . hal-04644583

HAL Id: hal-04644583

<https://hal.science/hal-04644583>

Submitted on 11 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Combining Convolution and Involution for the Early Prediction of Chronic Kidney Disease

Hadrien Salem¹, Sarah Ben Othman¹, Marc Broucqsault², and Slim Hammadi¹[0000-0002-5187-4870]

¹ CRIStAL CNRS UMR 9189, Villeneuve d’Ascq, Nord, France
{hadrien.salem, sara.ben-othman, slim.hammadi}@centralelille.fr

² Altao, Lille, Nord, France
mbroucqsault@altao.com

Abstract. Chronic Kidney Disease (CKD) is a common disease with high incidence and high risk for the patients’ health when it degrades to its most advanced stages. When detected early, it is possible to slow down the progression of the disease, leading to an increased survival rate and lighter treatment. As a consequence, many prediction models have emerged for the prediction of CKD. However, few of them manage to efficiently predict the onset of the disease months to years prior. In this paper, we propose an artificial neural network combining the strengths of convolution and involution layers in order to predict the degradation of CKD to its later stages, based on a set of 25 common laboratory analyses as well as the age and gender of the patient. Using a dataset from a French medical laboratory containing more than 400 000 patients, we show that our model achieves better performance than state-of-the-art models, with a recall of 83%, F1-score of 76%, and 97% overall accuracy. The proposed method is flexible and easily applicable to other diseases, offering encouraging perspectives in the field of early disease prediction, as well as the use of involution layers for deep learning with time series.

Keywords: Machine Learning · Chronic Kidney Disease · Disease prediction · Data processing · Big Data Analytics · Artificial Neural Networks · Convolutional Neural Networks · Involutional Neural Networks

1 Introduction

Chronic Kidney Disease (CKD) is a long-term condition corresponding to a malfunction of the kidneys. The kidneys play a crucial role in filtering waste products and excess fluids from the blood to form urine. In CKD, the kidneys gradually lose their ability to function effectively over time. This can cause a wide range of complications, such as anemia or an increased risk of cardiovascular diseases. In the most severe cases, CKD can lead to kidney failure, making heavy treatments such as dialysis or kidney transplantation necessary.

When detected early, it is generally possible to slow down the progression of the disease with lifestyle changes and medication. While it is normal for kidney

function to decrease with age, certain patients can experience an accelerated degradation and quickly evolve to advanced stages of the disease. Therefore, those patients are likely to not be properly monitored and risk more severe health consequences.

However, it is difficult to predict the speed at which one’s renal function might decrease. There lies the crux of the challenge : using seemingly-normal biological data to predict an abnormal evolution of the patients’ renal function. Risk factor calculations exist, such as the Kidney Failure Risk Equation (KFRE) [26]. However, establishing the formulae for these calculations can be difficult, and depends on the targeted population. In that regard, an alternative based on machine learning could be beneficial, as they can easily be retrained on new data. In particular, this would allow every medical laboratory to maintain their own model, catered to their own patient base.

In this paper, we propose a neural network model based on convolution and involution layers for the early prediction of a degradation in kidney function, thus helping with the monitoring of CKD for patients at risk. Our model is based solely on common laboratory tests, and its predictions could therefore be integrated seamlessly into a biological report, without requiring any additional information on the patient.

2 Related work

Disease prediction and detection has received a lot of attention in the literature, especially in recent years. The type of data and the techniques used to analyze it depend on the disease that is studied. Diabetes, for example, has been extensively researched, both for risk prognosis [16, 28] and blood glucose prediction [17]. Other diseases, such as Alzheimer’s disease [1, 8] or colorectal cancer [12] have been studied in a similar manner with various types of data: not only laboratory tests, but also imaging, and even environmental and lifestyle data [4].

In addition to ”classic” machine learning approaches, such as tree-based methods, support vector machines, or artificial neural networks, other methods such as survival analysis [21, 22, 29] and knowledge-based approaches [19, 27] have been used to establish prediction or risk-stratification models, as well as discovering semantic relationships in medical data.

Chronic Kidney Disease (CKD) in particular is one of the conditions that has received the most attention in the literature, because of its high incidence. Machine learning tasks often focus on the detection of CKD using Electronic Health Records (EHRs) and laboratory results like creatinine or eGFR (estimated Glomerular Filtration Rate) [11, 23], using various methods such as Support Vector Machine, Random Forest, XGBoost, Logistic Regression, Neural networks, AdaBoost, etc. These models usually achieve very good accuracy (sometimes over 99%), but it could be argued that this classification task is easy, because CKD itself is diagnosed using eGFR (refer to Section 3.1 for a more detailed explanation). In other words, if eGFR is part of the learning features, the

use of machine learning for the diagnosis of CKD at one point in time can be questionable.

In order to prevent complications in CKD, it is much more interesting to be able to predict the onset or the progression of the disease months or years in advance. A lot less research can be found in the literature on the subject. Bernardini et al. [3] propose a semi supervised learning for short term prediction of ckd stage using 2.5 years of history, but is quite restrictive on the patients that the model can apply to, since all fields must have been observed in the patients' history. Zhao et al. [30] use a combination of genetic and non genetic features to study risk factors in CKD for patients with a 5-year followup, but their dataset contains only a small sample of patients with similar characteristics. Chuah et al. [6] predict the progression of patients to end-stage CKD within 2 years and compare their results against those obtained by clinicians. While they obtain, 93.9% global accuracy, they have a high false negative rate leading to 60% recall. Finally, Razavian et al. [24] propose a prediction model for 133 diseases, including end-stage CKD, based on a set of 18 common laboratory tests. Using ensemble algorithms based on convolutional neural networks (CNN) and long short-term memory networks (LSTM), they manage to predict the onset of these conditions up to 15 months prior. They obtain 92% AUC for end-stage CKD, but do not present other performance metrics. Besides, if their model shows good AUC on certain diseases, simultaneously predicting such a high number of diseases leads to poorer performance on many of them.

While recurrent networks (such as Gated Recurrent Units [7] or Long Short-Term Memory networks [9]) have been shown to provide good results for time series classification, including in the domain of disease prediction [24], we would like to explore the use of new types of layers and their ability to efficiently extract meaningful features in a patient's biological history. In particular, to the best of our knowledge, no works in the literature use involution layers [14] to perform disease prediction based on biomarkers. This type of layer, designed as an "inversion" of convolutional layers, has given good results in the literature on other tasks, in particular for image recognition. However, just like CNNs, they can be used to perform predictions using time series. Liang et al. [15] even show that combining involution and convolution can provide better results than using one or the other separately.

In this paper, our focus will be the prediction of the progression of CKD using a model combining convolutional and involucional layers. In a population of more than 400 000 patients, we will detect those who move to a critical stage of the disease within the next year.

Our contribution is as follows:

- We propose a combined convolutional and involucional neural network model to perform the early prediction of the progression of CKD. Involutional networks in particular have not, to the best of our knowledge, been used for disease prediction in the literature. This type of architecture is novel in this field, and opens up new perspectives to build neural networks that are bet-

ter able to capture meaningful features when studying longitudinal biological data.

- We use a set of features consisting exclusively of the patient’s age and gender, and 25 commonly-measured laboratory analyses. These markers do not only include biomarkers related to renal functions, such as eGFR or creatinine, but also extends to other groups, including the lipid panel and blood counts. This is an alternative to the most common approach consisting in finding targeted biomarkers, which is difficult and disease-dependent. Using common biomarkers makes our framework more flexible and easily applicable than most in the literature.
- We analyze the performances of our model on a real dataset originating from a French medical laboratory, with over 400 000 patients. Our model provides good results despite large amounts of missing data, which matches real-life use cases. We compare our results with those obtained by state-of-the-art models, and show that the combination of convolution and involution provides better results in terms of accuracy, precision and F1-score, than either of the networks used individually.

3 Methods

In this section, we will define precisely the machine learning problem we are trying to solve, before describing in detail the model.

3.1 Problem definition

In this study, our goal is the early prediction (months to years prior) of the progression of Chronic Kidney Disease (CKD) to an advanced stage of the disease.

In order to classify the gravity of this disease, physicians use the estimated Glomerular Filtration Rate (eGFR), which quantifies the filtering capacity of the kidneys. It is expressed in mL/min/1.73m² and computed with the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation [13]:

$$eGFR = 141 \times \min\left(\frac{Cr}{\kappa}, 1\right)^\alpha \times \max\left(\frac{Cr}{\kappa}, 1\right)^{-1.209} \times 0.993^{Age} \quad (1)$$

In the above equation, $\kappa = 0.9$ for males and $\kappa = 0.7$ for females, $\alpha = -0.411$ for males and $\alpha = -0.329$ for females, and Cr is the level of serum creatinine expressed in $\mu\text{mol/L}$.

Following the guidelines provided by the Kidney Disease - Improving Global Outcomes (KDIGO) organization in their 2022 clinical practice guidelines for diabetes management in CKD [25], we consider the five following stages of CKD:

- **Stage 1** (normal or high kidney function): $eGFR \geq 90 \text{ mL / min / } 1.73 \text{ m}^2$
- **Stage 2** (mildly decreased kidney function): $60 \leq eGFR < 90 \text{ mL / min / } 1.73 \text{ m}^2$
- **Stage 3** (mildly to moderately decreased kidney function): $30 \leq eGFR < 60 \text{ mL / min / } 1.73 \text{ m}^2$

- **Stage 4** (severely decreased kidney function): $15 \leq \text{eGFR} < 30 \text{ mL} / \text{min} / 1.73 \text{ m}^2$
- **Stage 5** (kidney failure): $\text{eGFR} < 15 \text{ mL} / \text{min} / 1.73 \text{ m}^2$

In the above classification, stage 1 corresponds to the lowest stage of progression of the disease, while stage 5 corresponds to end-stage CKD and usually either requires dialysis or kidney transplant for the patient. Stages 2 and 3 are intermediate stages: while stage 2 is almost never concerning, stage 3 can sometimes indicate a more severely decreased kidney function depending on other parameters, such as age or albuminuria levels.

In this study, we will consider stages 4 and 5 to be the "positive" cases. This differs from other works in the literature, where only kidney failure (stage 5) is predicted [2, 23]. Indeed, from a medical point of view, it is interesting to predict earlier stages of the disease to start monitoring and treating the patient [25].

3.2 Model

With these stages established, we can define the machine learning task as follows: we select a cohort of patients, and try to determine whether they will progress to stage 4 or 5 during a certain prediction window, using a set of laboratory analyses (biomarkers) over time. We are thus treating a binary classification problem.

The difficulty of modeling the patients' history is that there is variability for both dimensions of the input: the length of the patients' history can vary because they do not necessarily have the same amount of records, and the analyses they take at each given point in time can also vary. To address this issue, similarly to the works of Razavian et al. [24], we use a sliding window framework: each patient has a variable-length, continuously-valued history of laboratory analyses X . For each of those matrices, at each time point t , we select the B months before t to form the input $X_{t-B:t}$. Data is marked as missing for the months where no analyses were taken.

We then look at the value of eGFR in a prediction window P after a "gap" G in order to label the input matrix. The gap ensures that clinical tests realized just before the prediction window are not used by the model and guarantees an "early" prediction. Data is labeled using the different stages defined in Section 3.1. For a label to be attributed, two measurements separated by at least 90 days must be observed within one of the stages' ranges. This precaution, recommended by KDIGO2022 [25], ensures that the renal disease is indeed chronic, and not the result of a temporary phenomenon. It also helps reduce noise in the dataset.

In Fig. 1, we show an example. Each square on the timeline represents a month, and the white square corresponds to the observation date t . The parameters in this example are $B = 6$, $G = 2$ and $P = 5$. This means that using six months of biological history (in light blue), the model would predict whether this patient reaches stage 4 or stage 5 CKD within a 5-month prediction window (in dark blue) after $t+G$. The gap (in dark grey) is arbitrarily set to two months in this illustration. While it is theoretically possible to set any value for B , G

and P , all the results in this article were obtained with the parameters listed in 4.1.

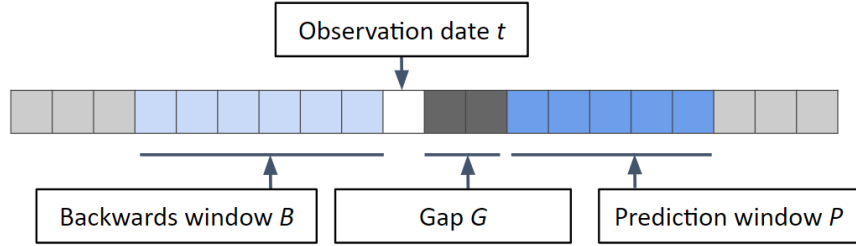


Fig. 1: Time model of the patients' biological history

3.3 Cohort selection

Fig. 2 shows our patient selection process. Out of 888 961 patients with at least one eGFR measurement in our dataset, we select those who have at least three eGFR measurements, since our approach relies on studying the history of biomarkers. For each of the 468 933 patients we have left, we are able to create several matrices using the sliding window framework described in section 3.2. Each matrix is labeled using the criteria defined in 3.1. We recall that stages 4 and 5 constitute the positive case (labeled 1).

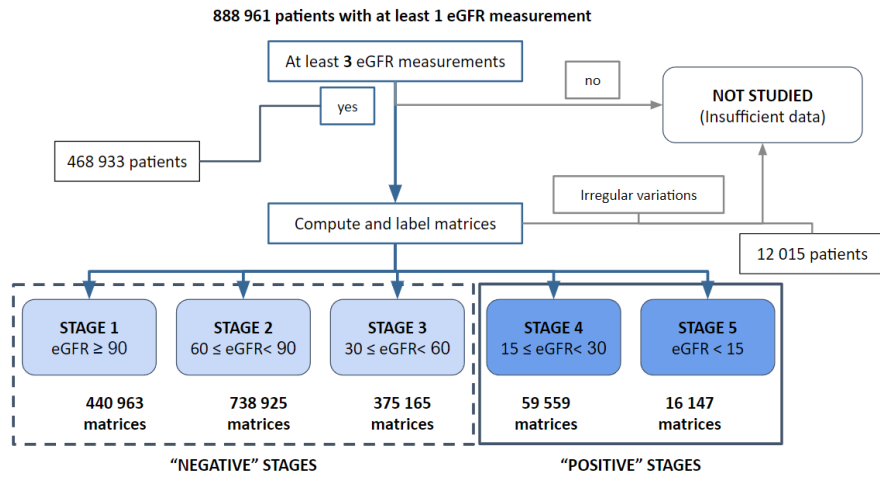


Fig. 2: Patient selection flowchart

3.4 Feature engineering

The abundance of missing data is one of the main challenges in dealing with medical data - especially so with laboratory data, because different patients are prescribed different analyses (or sets of analyses). However, there are some biomarkers that are very commonly measured compared to others, and form a "baseline" of analyses upon which other biomarkers can be added in specific cases. Targeting these baseline biomarkers is fundamental to our study, since we aim to detect abnormalities in common laboratory data. As a consequence, we select all features that are present for a significant amount of patients, in this case 25%. Out of the 76 available biomarkers available in our dataset, we are thus left with 25 commonly-measured laboratory analyses, and two demographic features (age and gender). It is important to note that those markers are not all traditionally associated with the renal function, as we use for example the lipid panel and blood counts. This means that our model can be used routinely, without necessarily suspecting CKD and prescribing specific tests for it, which is a strong advantage in the context of disease screening.

The full list of features is provided in Table 1. Preliminary analyses have shown that the results given with this subset of analyses are equivalent to those obtained using all biomarkers, and are being computed faster thanks to the reduced dimensionality.

3.5 Pre-processing

Management of missing data By using a set of common biomarkers, we have already limited the amount of missing data. However, due to the way we create matrices as presented in Section 3.2, there remains a lot of gaps in the matrices, because biomarkers are almost never measured every month. As a consequence, we need a strategy for the management of this missing data. We impute missing values for a feature based on the mean value for that column within the patient's history. We decide not to compute the mean based on the whole dataset, so that the imputed values remain close to the patient's, therefore being more meaningful.

Management of class imbalance The classes we have defined are heavily imbalanced in the dataset, since patients at stages 4 and 5 a minority in the dataset (as seen in Fig. 2). As such, positive samples represent about 5% of the total number of samples.

Several strategies have been employed in the literature to deal with class imbalance, such as over-sampling the minority class, under-sampling the majority class, Synthetic Minority Over-sampling Technique (SMOTE) [5] and its variants, or the definition of imbalance-aware cost functions such as Balanced Binary Cross-Entropy (BBCE) [18].

In order to avoid discarding the majority of the dataset, we decide to use an oversampling technique to address class imbalance. In particular, we use random oversampling, which provides the best results in our preliminary tests.

3.6 Machine learning model

Neural network model Convolution layers are popular in artificial neural networks, notably in the fields of image recognition, image segmentation, and computer vision. They rely on the learning of filters, or kernels, which are able to capture abstract patterns in data. When dealing with images, these patterns can be local edges or textures. However, convolution layers and convolutional neural networks (CNN) have been shown to also be able to tackle time series, including in disease prediction [24]. Fig. 3a shows a simple example of convolution, where a 2x2 kernel was learned by the network and applied to the data. The output is obtained by convolution product between the data matrix and the kernel, and represented on the right of the figure.

Involution layers were introduced by Li et al. [14] as an "inversion" of convolution, in that they are spatial-specific and channel-agnostic. In particular, it means that instead of learning a filter and applying it to the entire data matrix, an involution layer learns a unique function that will create different filters for each element of the matrix (pixels for an image, biomarker values in our case). This process has a twofold advantage: involucional networks require less parameters in general, since only one function needs to be learned to generate the filters; and they are by nature better at capturing local phenomena. Fig. 3b shows an example of involution. The network learned a kernel generation function ϕ , which is used to generate a 3x3 kernel from the centermost element of the data matrix. This kernel is then applied to the matrix, which results in the output depicted on the right.

Liang et al. [15] show that combining the respective strengths of convolution and involution lead to increased performance on image classification. Similarly, we decide to create a neural network architecture combining convolution and involution for the prediction of CKD progression. Both types of layers can be adapted for use with time series by considering a two-dimensional input, where each column corresponds to a biomarker, and each row corresponds to a specific date. In other words, each row represents a biological record for the patient at a given date.

The architecture of our neural network model is presented in Fig. 4. We use a convolutional layer with twelve 1x3 kernels. ReLU activation is used for non-linearity, and followed by a batch normalization layer [10]. After this, we append two blocks constituted of an involution layer with 3x3 kernels and ReLU activation, and a MaxPooling layer. The output of that layer is flattened and fully connected to 64 neural units, before the output layer provides the binary prediction. Binary cross-entropy is used to compute the loss, and the whole model is implemented using the Keras Python library. Hyperparameters have been fine-tuned by using a sample of the data.

In order to evaluate the performance of this model, which we will call **combined model**, we define two other models:

- In the first one, we replace each involution block by another convolution block, similar to the one defined above (convolution layer, batchnorm and

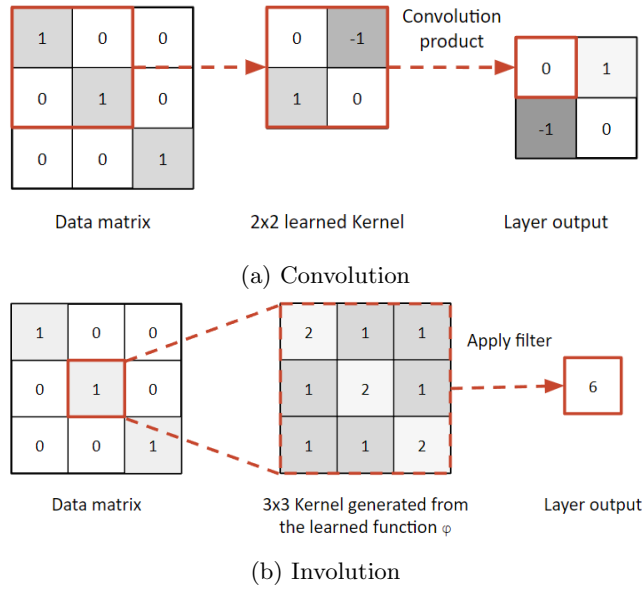


Fig. 3: Schematic illustrations for convolution and involution

ReLU activation). This model will be referred to as the **convolution-only network**.

- In the second one, we remove the convolution block, leaving only the involution blocks. This model will be referred to as the **involution-only network**.

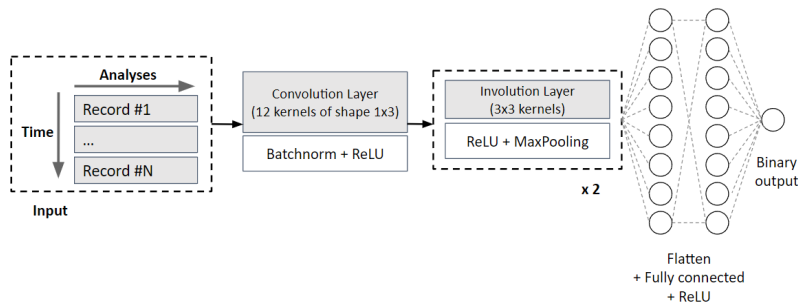


Fig. 4: Architecture of our neural network

Model evaluation Accuracy, which is defined as the proportion samples classified correctly, is not enough to assess the performance of a model with high

class imbalance. If the proportion of samples in the majority class is significantly higher than in the minority class, then good accuracy can be achieved by only predicting the majority class. This must be avoided in real-life applications, especially when it comes to disease screening.

For this reason, we will also consider other performance metrics that take into account the quality of prediction on both the majority and minority class:

- **Recall** (also called sensitivity or true positive rate) measures the proportion of actual positive samples that are correctly identified by a classifier. It is computed as $\frac{TP}{TP+FN}$, where TP is the number of true positives (positive samples classified correctly) and FN is the number of false negatives (positive samples classified incorrectly). A high recall is desirable in our context of disease screening.
- **Precision** measures the accuracy of the model on positive samples. It is computed as $\frac{TP}{TP+FP}$. A high precision indicates a low number of false positives, which is generally preferred.
- **F1-score** is the harmonic mean of sensitivity and precision. As such, it gives a balanced measure of the classifier’s performance. It is computed as $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$. The F1-score gives equal weight to precision and recall, making it a popular metric to study imbalanced datasets.
- **Area under ROC curve (AUC)**. The Receiver Operating Characteristic curve plots the true positive rate against the false positive rate at different classification thresholds. A higher area under this curve indicates better performance.

4 Results and discussion

4.1 Experimental setup

The experiments were realized on a dataset containing nearly 900 000 patients that had at least one eGFR measurement in their biological history, which we use to create our training and test datasets as described in Section 2. The data was gathered between 2010 and 2022, strictly anonymized at its source and was hosted on a highly secured server. While the full dataset is not publicly available, we decide to use it because it contains a large amount of day-to-day laboratory results, as opposed to hospital data, which may be biased towards positive cases. Code and anonymized data samples are provided at <https://github.com/CristalOSL/ckd-prediction>.

Regarding the technical setup, all treatments are realized on the Windows 10 operating system with an Intel Xeon Gold 6126 CPU (2.60GHz) and 128 GB RAM. We use Python 3.9 as our programming language. Data is manipulated with the Dask library (2023.2.0) for parallel computing, and Pandas (1.5.3) for lighter treatments. Data imbalance is managed using Imbalanced-learn (0.10.1), and the neural networks are implemented using Keras (2.14.0).

The set of features we have used for this study is detailed in Table 1. For all biomarkers, we have also provided the associated Logical Observation Identifiers Names and Codes (LOINC) standardized code [20].

Table 1: Features used in the study

Feature	LOINC
Age	-
Gender	-
Creatinine	2160-0
Estimated Glomerular Filtration Rate	62238-1
Mean Corpuscular Hemoglobin Concentration	28540-3
Red blood cell distribution width	30385-9
Platelets in plasma	26516-5
Leukocytes	26464-8
Hemoglobin	718-7
Neutrophils	30448-5
Lymphocytes	26474-7
Monocytes	26484-6
Hematocrit	20570-8
Erythrocytes	26453-1
Mean Corpuscular Hemoglobin	28539-5
Mean Corpuscular Volume	30428-7
Cholesterol	2093-3
Triglyceride	2571-8
Cholesterol in HDL	2085-9
Cholesterol non HDL	43396-1
Glucose	2345-7
Potassium	2823-3
Alanine aminotransferase	1742-6
Aspartate aminotransferase	1920-8
Sodium	2951-2
Chloride	2075-0
Prostate Specific Antigen	2857-1

In the following simulations, the model described in 3.2 is parameterized as follows: we use a backwards window of $B = 12$, a prediction window of $P = 12$ and a gap of $G = 3$. In other words, we perform the prediction of the evolution of CKD towards stage 4 or 5 within the year after a three-month gap, using a year of biological history.

4.2 Experimental results

The first part of Table 2 shows the results obtained in our prediction task, comparing the three models described in 3.6: the involution-only model, the convolution-only model, and the model combining both convolution and involution layers. The performance metrics are those that were presented in 3.6: accuracy (Acc.), area under ROC curve (AUC), precision (Prec.), recall (Rec.) and F1-score (F1). A threshold of 0.5 was used to compute the latter. In order to compare these results with those obtained in the literature, the lower part of Table 2 displays these performance metrics for different models. Unavailable

figures are indicated by a dash. Since all models do not perform the exact same prediction task, the "Positive" column indicates which stages are considered to be the positive samples for each individual model. Besides, all of these studies have been performed on different datasets, making the comparison difficult. Because of these differences, Table 2 should be interpreted as an overview of state-of-the-art models and their respective performance.

Overall, we observe that our models achieve good performance compared to the literature: we obtain the best overall accuracy, recall and F1-score, as well as competitive AUC and precision. As explained in 3.6, obtaining a high recall is particularly important in our context of disease screening, since it indicates that few positive cases are missed. Our high F1-score also indicates a good balance between the detection of positive cases and the accuracy of those predictions.

If we compare the combined model to the involution- and convolution-only models, we notice that it achieves the best performance in terms of overall accuracy, precision and f1-score, while remaining comparable in terms of AUC and recall. Precision in particular is significantly higher, being 20% above the involution-only model in that regard. We conclude that this combined model strikes a good balance between the performance metrics (as indicated by the high F1-score), and manages to detect most positive cases while keeping a relatively low proportion of false positives. This shows that the combination of convolution and involution layers is able to better capture patterns within the patients' biological history in order to predict the onset of CKD.

Table 2: Experimental results and overview of results obtained for CKD prediction in state-of-the-art models

Model	Positive	Acc.	AUC	Prec.	Rec.	F1
Involution-only model	Stages 4 & 5	0.967	0.911	0.610	0.850	0.710
Convolution-only model	Stages 4 & 5	0.952	0.908	0.500	0.860	0.630
Combined model	Stages 4 & 5	0.976	0.905	0.710	0.830	0.760
Razavian et al. ensemble model [24]	Stage 5	-	0.920	-	-	-
Razavian et al. ensemble model [24]	Stage 3	-	0.864	-	-	-
Chuah et al. XGBoost [6]	Stage 5	0.939	-	0.750	0.600	0.670
Bernardini et al. SS-MTL model [3]	Stages 2 to 5	0.746	0.811	0.657	0.731	0.665
Zhao et al. logistic regression model [30]	Stage 5	-	0.894	-	0.827	-

5 Conclusion

In this paper, we have proposed a combined convolutional and involutorial neural network for the prediction of stage 4 and stage 5 CKD a year in advance, using a year of biological history. Our model includes 25 very commonly-measured biomarkers, as well as age and gender, thus being adapted for practical use by

medical laboratories. By testing our model on a large dataset containing laboratory analyses for over 400 000 patients, we have demonstrated that its performance was an improvement over state-of-the-art models in terms of accuracy, recall and F1-score, and offers more balanced results than models using only involution or convolution. More generally the use of such networks with time series should be further explored.

A future avenue of research is the generalization of this model to other diseases, as it only uses common laboratory results and makes no assumptions on the patients, meaning it could be easily transferable to conditions like diabetes or even prostate cancer. However, the generalization to other diseases poses new challenges, such as the selection of an optimal set of analyses allowing for maximum performance, and the selection of the best model parameters with regards to the backwards window and prediction window, since different diseases may evolve at different speeds. One could also explore the use of automated machine learning to adapt the model's hyperparameters to the various diseases that would be studied, and the possibility to perform transfer learning between them.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abuhmed, T., El-Sappagh, S., Alonso, J.M.: Robust hybrid deep learning models for alzheimer's progression detection. *Knowledge-Based Systems* **213**, 106688 (2021)
2. Belur Nagaraj, S., Pena, M.J., Ju, W., Heerspink, H.L., Consortium, B.D.: Machine-learning-based early prediction of end-stage renal disease in patients with diabetic kidney disease using clinical trials data. *Diabetes, Obesity and Metabolism* **22**(12), 2479–2486 (2020)
3. Bernardini, M., Romeo, L., Frontoni, E., Amini, M.R.: A semi-supervised multi-task learning approach for predicting short-term kidney disease evolution. *IEEE Journal of Biomedical and Health Informatics* **25**(10), 3983–3994 (2021)
4. Chang, V., Ganatra, M.A., Hall, K., Golightly, L., Xu, Q.A.: An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators. *Healthcare Analytics* **2**, 100118 (2022)
5. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research* **16**, 321–357 (2002)
6. Chuah, A., Walters, G., Christiadi, D., Karpe, K., Kennard, A., Singer, R., Taulikar, G., Ge, W., Suominen, H., Andrews, T.D., et al.: Machine learning improves upon clinicians' prediction of end stage kidney disease. *Frontiers in medicine* **9**, 837232 (2022)
7. Dey, R., Salem, F.M.: Gate-variants of gated recurrent unit (gru) neural networks. In: 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS). pp. 1597–1600. IEEE (2017)
8. El-Sappagh, S., Ali, F., Abuhmed, T., Singh, J., Alonso, J.M.: Automatic detection of alzheimer's disease progression: An efficient information fusion approach with heterogeneous ensemble classifiers. *Neurocomputing* **512**, 203–224 (2022)

9. Graves, A., Graves, A.: Long short-term memory. Supervised sequence labelling with recurrent neural networks pp. 37–45 (2012)
10. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. pp. 448–456. pmlr (2015)
11. Islam, M.A., Majumder, M.Z.H., Hussein, M.A.: Chronic kidney disease prediction based on machine learning algorithms. *Journal of Pathology Informatics* **14**, 100189 (2023)
12. Kinar, Y., Kalkstein, N., Akiva, P., Levin, B., Half, E.E., Goldshtein, I., Chodick, G., Shalev, V.: Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study. *Journal of the American Medical Informatics Association* **23**(5), 879–890 (2016)
13. Levey, A.S., Stevens, L.A., Schmid, C.H., Zhang, Y., Castro III, A.F., Feldman, H.I., Kusek, J.W., Eggers, P., Van Lente, F., Greene, T., et al.: A new equation to estimate glomerular filtration rate. *Annals of internal medicine* **150**(9), 604–612 (2009)
14. Li, D., Hu, J., Wang, C., Li, X., She, Q., Zhu, L., Zhang, T., Chen, Q.: Involution: Inverting the inherence of convolution for visual recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12321–12330 (2021)
15. Liang, G., Wang, H.: I-cnet: leveraging involution and convolution for image classification. *IEEE Access* **10**, 2077–2082 (2021)
16. López Ibáñez, B., Vinas, R., Torrent-Fontbona, F., Fernández-Real Lemos, J.M.: Handling missing phenotype data with random forests for diabetes risk prognosis. In: © López, B., Herrero, P., Martín, C.(eds).(2016). AID: Artificial Intelligence for Diabetes: 1st ECAI Workshop on Artificial intelligence for Diabetes at the 22nd European Conference on Artificial Intelligence (ECAI 2016): 30 August 2016, The Hague, Holland: Proceedings, p. 39-42. European Conference on Artificial Intelligence (ECAI) (2016)
17. Martinsson, J., Schliep, A., Eliasson, B., Meijner, C., Persson, S., Mogren, O.: Automatic blood glucose prediction with confidence using recurrent neural networks. In: 3rd International Workshop on Knowledge Discovery in Healthcare Data, KDH@ IJCAI-ECAI 2018, 13 July 2018. pp. 64–68 (2018)
18. Martínez-Agüero, S., Soguero-Ruiz, C., Alonso-Moral, J.M., Mora-Jiménez, I., Álvarez Rodríguez, J., Marques, A.G.: Interpretable clinical time-series modeling with intelligent feature selection for early prediction of antimicrobial multidrug resistance. *Future Generation Computer Systems* **133**, 68–83 (2022). <https://doi.org/https://doi.org/10.1016/j.future.2022.02.021>, <https://www.sciencedirect.com/science/article/pii/S0167739X22000644>
19. Mayer, T., Cabrio, E., Villata, S.: Transformer-based argument mining for health-care applications. In: ECAI 2020, pp. 2108–2115. IOS Press (2020)
20. McDonald, C.J., Huff, S.M., Suico, J.G., Hill, G., Leavelle, D., Aller, R., Forrey, A., Mercer, K., DeMoor, G., Hook, J., et al.: Loinc, a universal standard for identifying laboratory observations: a 5-year update. *Clinical chemistry* **49**(4), 624–633 (2003)
21. Mirabnahrzham, G., Ma, D., Beaulac, C., Lee, S., Popuri, K., Lee, H., Cao, J., Galvin, J.E., Wang, L., Beg, M.F., et al.: Predicting time-to-conversion for dementia of alzheimer’s type using multi-modal deep survival analysis. *Neurobiology of Aging* **121**, 139–156 (2023)

22. Moncada-Torres, A., van Maaren, M.C., Hendriks, M.P., Siesling, S., Geleijnse, G.: Explainable machine learning can outperform cox regression predictions and provide insights in breast cancer survival. *Scientific reports* **11**(1), 6968 (2021)
23. Raju, N.G., Lakshmi, K.P., Praharshitha, K.G., Likhitha, C.: Prediction of chronic kidney disease (ckd) using data science. In: 2019 International Conference on Intelligent Computing and Control Systems (ICCS). pp. 642–647. IEEE (2019)
24. Razavian, N., Marcus, J., Sontag, D.: Multi-task prediction of disease onsets from longitudinal laboratory tests. In: Machine learning for healthcare conference. pp. 73–100. PMLR (2016)
25. Rossing, P., Caramori, M.L., Chan, J.C., Heerspink, H.J., Hurst, C., Khunti, K., Liew, A., Michos, E.D., Navaneethan, S.D., Olowu, W.A., et al.: Kdigo 2022 clinical practice guideline for diabetes management in chronic kidney disease. *Kidney International* **102**(5), S1–S127 (2022)
26. Tangri, N., Stevens, L.A., Griffith, J., Tighiouart, H., Djurdjev, O., Naimark, D., Levin, A., Levey, A.S.: A predictive model for progression of chronic kidney disease to kidney failure. *Jama* **305**(15), 1553–1559 (2011)
27. Tymoshenko, K., Somasundaran, S., Prabhakaran, V., Shet, V.: Relation mining in the biomedical domain using entity-level semantics. In: ECAI 2012, pp. 780–785. IOS Press (2012)
28. Xie, Z., Nikolayeva, O., Luo, J., Li, D.: Building risk prediction models for type 2 diabetes using machine learning techniques. *Preventing chronic disease* **16** (2019)
29. Yansari, R.T., Mirzarezaee, M., Sadeghi, M., Araabi, B.N.: A new survival analysis model in adjuvant tamoxifen-treated breast cancer patients using manifold-based semi-supervised learning. *Journal of Computational Science* **61**, 101645 (2022)
30. Zhao, J., Zhang, Y., Qiu, J., Zhang, X., Wei, F., Feng, J., Chen, C., Zhang, K., Feng, S., Li, W.D.: An early prediction model for chronic kidney disease. *Scientific reports* **12**(1), 2765 (2022)