ARTIFICIAL INTELLIGENCE TOOL FOR PREDICTING END-STAGE KIDNEY DISEASE IN PATIENTS WITH IMMUNOGLOBULIN A NEPHROPATHY.

Francesco Paolo Schena, MD[1,2], Vito Walter Anelli, PhD[3], Joseph Trotta, Engr[3], Tommaso Di Noia, Engr[3], Carlo Manno, MD[1], Giovanni Tripepi, PhD[4], Graziella D'Arrigo, PhD[4], Nicholas C. Chesnaye, PhD[5], Maria Luisa Russo, PhD[6], Maria Stangou, MD[7], Aikaterini Papagianni, MD[7], Carmine Zoccali, MD[4], Vladimir Tesar, MD[8], Rosanna Coppo, MD[6].

Authors Affiliations

[1]Department of Emergency and Organ Transplant, University of Bari, Bari, Italy.

[2]Fondazione Schena, Valenzano, Bari, Italy.

[3]Department of Electrical and Information Engineering, Polytechnic of Bari, Bari, Italy.

[4]CNR-IFC, Clinical Epidemiology and Pathophysiology of Renal Diseases and Hypertension, Nefrologia-Ospedali Riuniti, 89100, Reggio Calabria, Italy.

[5]Department of Medical Informatics, Public Health Research Institute, University of Amsterdam , The Netherlands.

[6]Fondazione Ricerca Molinette, Torino, Italy.

[7]Department of Nephrology, Hippokration General Hospital, Aristotle University of Thessaloniki, Thessaloniki, Greece.

[8]Department of Nephrology, 1st Faculty of Medicine and General University Hospital, Charles University, Prague, Czech Rep.

Correspondence address

Francesco Paolo Schena, MD

paolo.schena@uniba.it

Running Head

End-stage kidney disease prediction for IgA Nephropathy

LIST

SUPPLEMENTARY METHODS

In our previous work [S1], we developed a classical neural network composed of two hidden layers that exploited the sigmoid as the activation function, and we achieved a good performance. However, in the past few years, many other activation functions have been published, which have shown excellent results in many cases studies. Moreover, the best practices have been proposed for the initialisation of the weights and different techniques to alleviate the overfitting, thus making the network more robust. Finally, several techniques have been proposed to deal with datasets that are not very large, such as in our study. We have tested all these advanced techniques to select the best approach to train our dataset.

Background Technologies

In this section, the main technologies that have been the best candidates for improving our previous neural network are described. All the depicted technologies have been tested using the stratified 10-fold cross-validation for selecting the best model. At the end of this section, we additionally described the other parameters involved in the hyper-parameters tuning phase.

Preprocessing Z score. As suggested by LeCun et al. [S2], the use of Z-score normalization could be beneficial for the learning algorithm to reach an acceptable minimum value of the cost function. The strategy consists of shifting each variable to have an average of zero and scaling it by dividing each sample with the variable variance. This strategy is called Z-score normalization and it can be summarised by the formula:

$$x_{out} = \frac{x_{in} - \mu}{\sigma}$$

ZCA Whitening. After normalization, we have applied the ZCA whitening step to decorrelate data. Since our features already have a zero mean thanks to normalization, we obtained the covariance

matrix by $\frac{X \cdot X^T}{m}$. We used U from SVD decomposition to project data in the new space. For the sake of completeness, we tested all the possible number of dimensions. Then, all the new features were whitened dividing them by the square root of the correspondent eigenvalue regularised by adding ϵ. Finally, we have used U to compute the ZCA-whitened version of X, obtaining the following formula:

$$X\_ZCA\_Whitened = U \cdot diag\left(\frac{1}{\sqrt{diag(S)+\epsilon}}\right) \cdot U^T \cdot X$$

This whitening step was not beneficial for our study and hence it was not a part of the proposed models.

Dropout. Several techniques have been proposed to avoid overfitting in neural networks models. One of the most famous and effective techniques has been described by Srivastava et al. [S3]. In this kind of models, the assumption that all the neurons were always present over the entire training procedure is dropped out. At each neuron, a probability to keep that neuron is associated. Based on the defined probability, each layer is potentially composed by a different number of neurons at each training step, and this imposes normalization of each output. This topology of modification has several effects. Indeed each neuron is more likely to operate in an independent way because of the potential lack of other neurons connected to similar weights. This could lead the neurons to learn completely different features.

Batch Normalization. This is another interesting method that has been proposed by Ioffe and Szegedy [S4]. The aim of Batch Normalization is to reduce the Internal Covariate Shift (ICS). ICS is defined as the change in the distribution of network activations due to the change in network parameters during training. It is inspired by the works [S3,S5] on faster convergence of whitened inputs. The rationale of this work is basically to whiten the outputs of each layer since it could be considered as the input of the next layer. In order to achieve this goal, for each layer, the mean and the variance of the samples batch *i* are computed for normalizing the inputs. These values are then scaled and shifted using the two parameters $\lambda = \sqrt{Var[x]}$ and $\beta = E[x]$ according to the following formula:

$$y_i \leftarrow \lambda \frac{x_i - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}} +$$

The cross-validation step showed that the adoption of Batch Normalization was not beneficial with the cross-entropy loss function, and this finding is consistent with the data reported by Clevert et al. [S6]. Nevertheless, it is the best option for our proposed novel proxy AUC cost function.

Activation Functions. After our previous work, several activation functions have been proposed and widely adopted. The former network made use of the simplest non-linear sigmoid function $g(Wx + b) = \frac{1}{1+e^{-(Wx+b)}}$ where W stands for the weights of the current layer, x is the input and b is the bias. However, because of the limited linear range, the sigmoid function showed its limits especially in deep neural networks where the vanishing gradient issue appears. One of the first attempts to address the issue was the *Rectified Linear Unit* (ReLU) [S7] with a half rectified function $R(Wx +b) = max(0,Wx +b)$. This function simply imposes $R(Wx +b)$ to be zero if $Wx +b < 0$. However, this function is not able to effectively handle negative values. A more flexible alternative to ReLU is the *Exponential Linear Unit (ELU)* proposed by Clevert et al. [S6]. The rationale behind ELU

is to avoid the mean and bias shifts in order to improve learning. The activation function formula is:

$$f(x) = \begin{cases} x, & x>0 \\ (e\verb|^|x - 1), & otherwise \end{cases}$$

where α controls the value an ELU saturates for negative net inputs. The last activation function that we decided to compare was *Scaled Exponential Linear Unit (SELU)* [S8], in which the ELU formulation is slightly modified:

$$f(x) = \begin{cases} \lambda & x, & x>0 \\ \alpha(e^x - \alpha), & otherwise \end{cases}$$

where a value of the hyperparameter λ could ensure a slope larger than one for the positive values. Again we have used cross-validation to select the activation function, and ELU resulted to be the best performing activation function for the classification task, whereas SELU was the best for the regression task.

Training algorithms. In recent years, a plethora of learning algorithms have been proposed in order to improve the different gradient descent method aspects. In the cross-validation step, several learning algorithms, such as classic Stochastic Gradient Descent, Adadelta [S9], Adagrad [S10], RMSProp [S11] and Adam [S12], have been tested with and without momentum. For the sake of conciseness, we have briefly focused on the best performing technique, Adam. The idea behind Adam is to exploit the moving average of gradients m_t and squared gradients v_t as an estimation of the 1st and 2nd moments of the gradient:

$$m_t = \beta_1 m_{t-1} + \left(1 - \beta_1\right)g_t \qquad v_t = \beta_2 v_{t-1} + \left(1 - \beta_2\right)g_t^2$$

Since the authors noted that they were biased towards zero, especially during the initial time steps, they were corrected:

$$\widehat{m_t} = \frac{m_t}{1 - \beta_1^t} \qquad \widehat{v_t} = \frac{v_t}{1 - \beta_2^t}$$

Thus they can be used to define the updated formula, similar to the Adadelta, and RMSprop learning algorithms:

$$\theta_{t+1} = \theta_t - \frac{\vee}{\sqrt{\widehat{v_t}} + \epsilon} \widehat{m_t}$$

Stopping conditions. In order to further alleviate the overfitting of the model on training data we have adopted a simple method:

(1) Train the model on the training set and compute the error on the validation set after every fifth epoch;

(2) If the new computed error on the validation set was greater than the previous one, return to the previous number of the epoch;

(3) Train the model on the training set with a number of epochs equal to an order of magnitude higher than the retrieved number, and compute the error on the validation set after every fifth epoch;

4

(4) Use the weights that the network had in the previous step at the epoch corresponding to the lowest error on the validation set.

This method has been used to take advantage of early stopping, and it has been partially inspired by the considerations on the stopping criterion published by Prechelt [S13].

Number of hidden layers, neurons, exponential decay, starting learning rate. A number of other parameters have been tuned in combination with the former ones. We have tested configurations with:

- a number of hidden layers varying from 1 to 10
- dropout, keeping the probability varying from 0.1 to 1.0
- a number of neurons varying from 5 to 300
- an exponential decay rate varying from 400 to 3200
- a starting learning rate varying from 0.2 to 0.000001

It is straightforward that optimal configuration is different for the classifier and regression model.

The *classifier model* is composed of four hidden layers with 100 neurons in each layer. Variables were obtained at the time of the kidney biopsy (age, sex, hypertension, proteinuria, serum creatinine, histological renal lesions and therapy) whereas the outcome (ESKD yes or no) represented the output data. Variables have been preprocessed by the mean of z-score normalization, and then, they have been served as the input layer. All the hyper-parameters have been tuned using 10-fold stratified cross-validation. Between each pair of layers, a batch normalization layer has been inserted to improve learning. For the sake of completeness, we obtained lower but similar results interleaving layer pairs with dropout layers with a keeping probability of 0.5. The hidden layers received data from the first layer and combined it, exploiting the interconnecting weights. Activation functions have been used to perform a transformation of the values in each neuron of the hidden layers. For this purpose, in our classifier, the non-saturating exponential linear unit (ELU) function was used [S6]. After the transformation, new information has been conveyed to the next layer. At the end of the neural network, a softmax layer transformed the obtained values in a probability distribution over the two output units. The backpropagation learning algorithm makes neural nets adaptive to the different inputs due to the error-driven weight updates. We cross-validated seven ANN models in order to find the best model (appendix methods). Different net sizes were tested in order to seek the most appropriate number of hidden nodes. Moreover, different parameter combinations were tested to build the ANN models in order to analyse the influence of risk factors on the outcome of the disease and to evaluate their importance in the predictive model. The initial hypothesis in this analysis has been that each input risk factor should be considered because it could affect the outcome of the disease.

The new ANN models were implemented using the Tensorflow framework (https://www.tensorflow.org/api_docs/), whereas the preprocessing functions were implemented from scratch or by taking advantage of the Scikit-learn framework (http://scikit-learn.org/stable/documentation.html). The tuning of parameters has been performed using 10-fold cross-validation and early stopping on the validation set in the training cohort of IgAN patients.

The *regression model* consists of 3 layers containing 125 neurons each. The applied preprocessing operations were the same as the former model, and the only relevant design difference was the use of dropout layers with a keeping probability of 30% for the hidden layers and a keeping probability of 50% for the input layer [S3]. This choice was performed to avoid the risk of overfitting. Moreover, in this case, the best performing activation function was Scaled Exponential Linear Unit (SELU). A regression model that is able to estimate the number of years in which the patient will reach the ESKD stage (with a maximum of ten years) was designed.

SUPPLEMENTARY REFERENCES

S1. Pesce F, Diciolla M, Binetti G. et al. Clinical decision support system for end-stage kidney disease risk estimation in IgA nephropathy patients. Nephrol Dial Transplant 2015;**31**:80–6.

S2. LeCun Y, Bengio Y, Hinton G.E. 2015. Deep learning. Nature 2015;**521**,436–44. https://doi.org/10.1038/ nature14539

S3. Srivastava N, Hinton G.E., Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 2014;**15,** 1929–58.

S4. Ioffe S, Christian Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015 (JMLR Workshop and Conference Proceedings), Francis R. Bach and David M. Blei (Eds.), Vol. 37. JMLR.org, 448–456. http://jmlr.org/proceedings/papers/v37/ioffe15.html

S5. Wiesler S and Ney H. 2011. A Convergence Analysis of Log-Linear Training. In Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain., John Shawe-Taylor, Richard S. Zemel, Peter L. Bartlett, Fernando C. N. Pereira, and Kilian Q. Weinberger (Eds.). 657–665. http://papers.nips.cc/paper/4421-a-convergence-analysis-of-log-linear-training

S6. Clevert D, Unterthiner T, Hochreiter S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *CoRR* abs/1511.07289 (2015). arXiv:1511.07289 http://arxiv.org/abs/1511.07289

S7. Nair V, Hinton G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel, Johannes Fürnkranz and Thorsten Joachims (Eds.). Omnipress, 807–814. http://www.icml2010.org/papers/432.pdf

S8. Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Self-Normalizing Neural Networks. In Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, *USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 972–981. http://papers.nips.cc/paper/6698-self-normalizing-neural-networks

S9. Zeiler M.D.   ADADELTA: An Adaptive Learning Rate Method.*CoRR* abs/1212.5701(2012).

S10. Duchi J.C, Hazan E, Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. Journal of Machine Learning Research 2011;**12**:2121–59. http://dl.acm.org/citation.cfm?id=2021068

S11. Hinton G, Srivastava N, Swersky K. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. UC Berkeley Short Course 2012.

S12. Diederik P. Ba K, Ba J. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). arXiv:1412.6980 http://arxiv.org/abs/1412.6980

S13. Prechelt L. 1998. Automatic early stopping using cross validation: quantifying the criteria. Neural Networks 1998;**11**, 761– 7. https://doi.org/10.1016/S0893-6080(98)00010-0

Supplementary Table 1. Clinical and histological risk factors for ESKD in IgAN patients*

Reference	Year	Patients (N)	Race	Mean Follow-up (mos)	Outcome	eGFR	P	H	Histologic grade
						— Indipendent predictors of unfavorable renal outcome			
D'Amico et al	1986	365	C	79	ESKD		Y		Y
Beukhof et al	1986	75	C	92	ESKD	Y	Y	Y	NA
Bogenschutz et al	1990	239	C	59	ESKD	Y			Y
Rekola et al	1990	209	C	76	ESKD	Y		Y	
Alamartine et al	1991	282	C	96	Creat >135	NA	Y	Y	Y
Johnston et al	1992	220	C	65	ESKD	Y			
Ibels and Gyry	1994	121	C	61	%Change creat.		Y		Y
Katafuchi et al	1994	225	A	48	ESKD	Y	Y		Y
Koyama et al	1997	448	A	142	ESKD	Y	Y		
Haas et al	1997	244	C-others	NA	ESKD		Y	Y	Y
Frimat et al	1997	210	C	67	ESKD	Y	Y		NA
Radford et al	1997	148	C	45	ESKD	Y			Y
Bartosik et al	2001	298	C	57	Rate CCr		Y	Y	
Donadio et al	2002	91	C	69	ESKD	Y	Y		
Donadio et al	2002	63	C	20	ESKD	Y	Y		
Li et al	2002	168	A	89	ESKD	Y	Y	Y	Y
Geddes et al	2003	711	C-others	53-123	ESKD	Y	Y		NA
Manno et al	2007	437	C	108	ESKD	Y	Y		Y
Reich et al	2007	542	C-others	78	Rate CCr	Y	Y	Y	NA
Lv et al	2008	204	A	73	ESKD	Y		Y	Y
Prakash et al	2008	76	A	18	Rate CCr				Y
Prakash et al	2008	152	C	39	Rate CCr				
Hwang et al	2010	125	A	90	Rate CCr		Y	Y	
Bjornerklett et al	2012	633	C	124	ESKD	Y	Y	Y	Y
Le et al	2012	1155	A	65	ESKD	Y	Y	Y	NA
Chou et al	2012	130	A	71	ESKD		Y	Y	
Moriyama et al	2014	1012	A	95	ESKD	Y	Y	Y	Y
Li et al	2014	703	A	45	ESKD	Y	Y	Y	NA
Coppo et al	2014	1147	C-others	56	ESKD	Y	Y		Y
Park et al	2014	500	A	68	D-Scr/ESKD	Y	Y		Y
Tan M et al	2015	376	A	75	ESKD		Y		Y
Zhao et al	2016	438	A	37	Rate CCr/ESKD	Y	Y	Y	Y
Barbour et al	2016	901	C-others	>24	Rate CCr/ESKD	Y	Y	Y	Y
Iwasaki et al	2016	75	A	24	ESKD	Y			
Zhang et al	2017	538	A	51	ESKD	Y	Y	Y	Y
Deng et al	2018	988	A	48	D-Scr/ESKD	Y	Y	Y	Y
Chen et al	2018	506	A	50	ESKD	Y	Y	Y	Y

*Estimated by multivariate Cox regression analysis

Abbreviations: N (number); A (Asian); C (Caucasian); mos (months); eGFR (estimated glomerular filtration rate); P (proteinuria); H (hypertension); Y (yes); ESKD (end-stage renal disease); CCr (clearance creatinine), NA (not assessed).

References Table S1

D'Amico G, Minetti L, Ponticelli C et al. Prognostic indicators in idiopathic IgA mesangial nephropathy. *Q J Med.* 1986;228:363-378.

Beukhof JR, Kardaun O, Schaafsma W et al. Toward individual prognosis of IgA nephropathy. *Kidney Int.* 1986;29:549-556.

Bogenschütz O, Bohle A, Batz C et al. IgA nephritis: on the importance of morphological and clinical parameters in the long-term prognosis of 239 patients. *Am J Nephrol.* 1990;10:137-147.

Rekola S, Bergstrand A, Bucht H. Development of hypertension in IgA nephropathy as a marker of a poor prognosis. *Am J Nephrol.* 1990;10:290-295.

Alamartine E, Sabatier JC, Guerin C et al. Prognostic factors in mesangial IgA glomerulonephritis: An extensive study with univariate and multivariate analysis. *Am J Kidney Dis.* 1991;18:12-19.

Johnston PA, Brown JS, Braumholtz DA et al. Clinicopathological correlations and long-term follow-up of 253 United Kingdom patients with IgA nephropathy. A report from the MRC glomerulonephritis Registry. Q *J Med*. 1992,84:619-627.

Ibels LS, Gyry AZ. IgA nephropathy: Analysis of the natural history, important factors in the progression of renal disease, and a review of the literature. *Medicine.* 1994;73:79-102.

Katafuchi R, Oh Y, Hori K et al. An important role of glomerular segmental lesions on progression of IgA nephropathy: A multivariate analysis. *Clin Nephrol.* 1994;41:191-198.

Koyama A, Igarashi M, Kobayashi M et al. Natural history and risk factors for immunoglobulin A nephropathy in Japan. *Am J Kidney Dis.* 1997;29:526-532.

Haas M: Histologic subclassification of IgA nephropathy: A clinicopathologic study of 244 cases. *Am J Kidney Dis*. 1997;29:829-842.

Frimat L, Briançon S, Hestin D et al. IgA nephropathy: prognostic classification of end-stage renal failure. L'Association des Néphrologues de l'Est. *Nephrol Dial Transplant* 1997;12:2569-2575.

Radford Jr MG, Donadio Jr JV, Bergstralh EJ et al. Predicting renal outcome in IgA nephropathy. *J Am Soc Nephrol* . 1997;8:199-207.

Bartosik LP, Lajoie G, Sugar L et al. Predicting progression in IgA nephropathy. *Am J Kidney Dis.* 2001;38:728-735.

Donadio JV, Bergstralh EJ, Grande JP et al. Proteinuria patterns and their association with subsequent End-stage kidney disease in IgA nephropathy. *Nephrol Dial Transplant.* 2002;17:1197-1203.

Li PKT,   Ho KKL, Szeto CC et al. Prognostic indicators of IgA nephropathy in the Chinese--clinical and pathological perspectives*. Nephrol Dial Transplant .* 2002;17:64-69.

Geddes CC, Rauta V, Gronhagen-Riska C et al. A tricontinental view of IgA nephropathy. *Nephrol Dial Transplant.* 2003;18:1541-1548,

Manno C, Strippoli GFM, D'Altri C et al. A novel simpler histological classification for renal survival in IgA nephropathy: a retrospective study. *Am J Kidney Dis Actions.* 2007;49:763-775.

Reich HN, Troyanov S, Scholey JW et al. Remission of Proteinuria Improves Prognosis in IgA Nephropathy. *JASN*. 2007;18:3177-3183.

Lv J, Zhang H, Zhou Y et al. Natural history of immunoglobulin A nephropathy and predictive factors of prognosis: a long-term follow up of 204 cases in China. *Nephrology (Carlton).* 2008;13:242-246.

Prakash S, Kanjanabuch T, Austin PC et al. Continental variations in IgA nephropathy among Asians. *Clin Nephrol Actions.* 2008;70:377-384.

Hwang HS, Kim BS, Shin YS et al Predictors for progression in immunoglobulin A nephropathy with significant proteinuria *Nephrol* 2010;15:236-241

Bjørneklett R, Vikse BE, Bostad L et al. Long-term risk of ESKD in IgAN; validation of Japanese prognostic model in a Norwegian cohort. *Nephrol Dial Transplant.* 2012;27:1485-1491.

Le W, Liang S, Hu Y et al. Long-term renal survival and related risk factors in patients with IgA nephropathy: results from a cohort of 1155 cases in a Chinese adult population. *Nephrol Dial Transplant.* 2012;27:1479-1485.

Chou YH, Lien YC, Hu FC et al. Clinical outcomes and predictors for ESKD and mortality in primary GN. *Clin J Am Soc Nephrol.* 2012;7:1401-1408.

Moriyama T, Tanaka K, Iwasaki C et al. Prognosis in IgA nephropathy: 30-year analysis of 1,012 patients at a single center in Japan. *PLoS One.* 2014; 21:9: e91756.

Li X, Liu Y, Lv J et al. Progression of IgA nephropathy under current therapy regimen in a Chinese population. *Clin J Am Soc Nephrol.* 2014; 9:484-489.

Coppo R, Troyanov S, Bellur S et al. Validation of the Oxford classification of IgA nephropathy in cohorts with different presentations and treatments. *Kidney Int.* 2014;86:828-836.

Park KS, Han SH, Kie JH et al. Comparison of the Haas and the Oxford classifications for prediction of renal outcome in patients with IgA nephropathy. *Human Pathology*2014;45236-243

Tan M, Li W, Zou G, Cong Zhang et al. Clinicopathological features and outcomes of IgA nephropathy with hematuria and/or minimal proteinuria. *Kidney Blood Press Res Actions.* 2015;40:200-206.

Zhao YF, Zhu L, Liu LJ et al. Measures of Urinary Protein and Albumin in the Prediction of Progression of IgA Nephropathy. *Clin J Am Soc Nephrol*. 2016;11:947-955.

Barbour SJ, Espino-Hernandez G, Reich HN et al. The MEST score provides earlier risk prediction in IgA nephropathy. *Kidney Int*. 2016;89:167–175.

Iwasaki C, Moriyama T, Tanaka K et al. Effect of hematuria on the outcome of immunoglobulin A nephropathy with proteinuria. *J Nephropathol*. 2016;5:72–78.

Zhang W , Zhou Q, Hong L et al. Clinical outcomes of IgA nephropathy patients with different proportions of crescents . *Medicine*. 2017;96:e6190.

Deng W, Tan X, Zhou Q et al. Gender-related differences in clinicopathological characteristics and renal outcomes of Chinese patients with IgA nephropathy. *BMC Nephrology.* 2018;19:31.

Chen D, Liu J, Duan S et al. Clinicopathological Features to Predict Progression of IgA Nephropathy with Mild Proteinuria. *Kidney Blood Press Res*. 2018;43:318–328.

Supplementary Table 2. Scoring systems for predicting ESKD in IgAN patients

Reference	Year	Patients(N)	Race	Outcome ESKD	Variables	Risk score		
						Grade	value(range)	
Wakai et al	2006	2269	A	HD	age, sex, hypertension, proteinuria, hematuria, serum total protein, serum creatinine, histologic grade	1 2 3 4 5	0.0 - 0.9 1.0 - 4.9 5.0 - 19.9 20.0 - 49.9 50.0 - 100	
Goto et al	2009	2283	A	HD	age, sex, hypetension, proteinuria, hematuria, hypoalbuminemia, eGFR, histologic grade	1 2 3 4	0 - 4.9 50 - 19.9 20.0 - 49.9 50.0 - 100	
Okonogi et al	2011	116	A	HD	proteinuria, eGFR	1 2 3 4	<1 <1 ≥1 ≥1	≥64 <64 ≥64 <64
Berthoux et al	2011	332	C	HD eGFR<10 ml/min/1.73 m2	hypertension, proteinuria >1 g, histologic grade (global optical score, GOS)	0 1 3		
Xie et al	2012	619	A	HD, TX	Hb, serum albumin, hypertension,eGFR	1 2 3	2 5 10	
Tanaka et al	2013	698	A	HD, TX	eGFR, proteinuria (PU), Oxford histologic grade (HG)	PU eGFR HG	0,4,6,9 0,1,4,9 0,2,4,6,10	
Knoop et al	2015	1134	C	HD,TX	age, hypertension,eGFR, proteinuria, histologic grade	1 2 3 4 5 6	1.0 1.4 - 2.4 2.9 - 7.3 7.8 - 17.9 22.0 - 53.2 57.0 - 111.6	
Barbour et al.	2019	2781	A-C	HD,TX, eGFR<15 ml reduction of eGFR below 50% of the value at KB	age, race, hypertension, eGFR, proteinuria,histological grade and therapy	Probability ESKD 1 - 100		

Abbreviations: N(number); eGFR (estimated glomerular filtration rate); Hb (hemoglobin); CRF (chronic renal failure); CKD (chronic kidney disease); C (Caucasian); A (Asian); TX (renal transplant); KB (kidney biopsy)

References Table S2

Wakai K, Kawamura T, Endoh M et al. A scoring system to predict renal outcome in IgA nephropathy: from a nationwide prospective study. *Nephrol Dial Transplant*. 2006;21:2800-2808.

Goto M, Wakai K, Kawamura T et al. A scoring system to predict renal outcome in IgA nephropathy: a nationwide 10-year prospective cohort study. *Nephrol Dial Transplant.* 2009;24:3068-3074.

Okonogi H, Utsunomiya Y, Miyazaki Y et al. A predictive clinical grading system for immunoglobulin A nephropathy by combining proteinuria and estimated glomerular filtration rate. *Nephron Clin Pract.* 2011;118:c292-300.

Berthoux F, Mohey H, Laurent B et al. Predicting the risk for dialysis or death in IgA nephropathy. *J Am Soc Nephrol.* 2011;22:752-761.

Xie J, Kiryluk K, Wang W et al. Predicting progression of IgA nephropathy: new clinical progression risk score. *PLoS One*.2012;7:e38904.

Tanaka S, Ninomiya T, Katafu chi R et al. Development and validation of a prediction rule using the Oxford classification in IgA nephropathy. *Clin J Am Soc Nephrol*. 2013;8:2082-2090.

Knoop T, Vågane AM, Vikse BE et al. Addition of eGFR and age improves the prognostic absolute renal risk-model in 1,134 Norwegian patients with IgA nephropathy. *Am J Nephrol*. 2015;41:210-219.

Barbour SJ, Coppo R, Zhang H et al. Evaluating a New International Risk-Prediction Tool in IgA Nephropathy. *JAMA Intern Med.* 2019;179:942-952.

Supplementary Table 3. Baseline characteristics of the IgAN patients included in the training and test set of the study cohort. Therapy, follow-up and clinical outcome are shown.

	Training set	Test set
Patients (n)	758	190
Age at biopsy (years)	40.5 ± 14.1	40.8+13.8
Sex (M/F)	541/217	144/46
Systolic blood pressure (mm Hg)	134.2 ± 18.8	133.7 ± 17.8
Diastolic blood pressure (mm Hg)	83.5 ± 10.8	82.1 ± 11.0
Mean arterial pressure (mm Hg)	100.4 ± 12.4	99.3 ± 12.3
Hypertension, n (%)	229 (30.2)	58 (30.5)
Serum creatinine (mg/dL)	1.20 (0.95-1.67)	1.20 (0.99-1.78)
eGFR MDRD (ml/min/1.73 m^2)	67.30 (45.54-90.72)	66.93 (42.32-87.97)
KDOQI stage		
I > 90 ml/min/1.73 m^2 , n (%)	194 (25.6)	42 (22.1)
II 60-89 ml/min/1.73 m^2, n (%)	244 (32.2)	66 (34.7)
III 30-59 ml/min/1.73 m^2 , n (%)	245 (32.3)	61 (32.1)
IV 15-29 ml/min/1.73 m^2 , n (%)	63 (8.3)	18 (9.5)
V <15 ml/min/1.73 m^2 , n (%)	12 (1.6)	3 (1.6)
Proteinuria (g/day)	1.30 (0.60-2.50)	1.35 (0.60-2.50)
Mild <1 g/day, n (%)	288 (38.0)	71 (37.4)
Moderate 1-3 g/day, n (%)	350 (46.2)	81 (42.6)
Severe >3 g/day, n (%)	120 (15.8)	38 (20.0)
Renal Biopsy		
M1, n/n (%)	249/758 (32.8)	58/190 (30.5)
E1, n/n (%)	81/758 (10.7)	27/190 (14.2)
S1, n/n (%)	575/758 (75.9)	135/190 (71.1)
T1, n/n (%)	150/758 (19.8)	44/190 (23.2)
T2, n/n (%)	38/758 (5.0)	6/190 (3.2)
C1, n/n (%)	68/758 (9.0)	18/190 (9.4)
Therapy		
RASB, n (%)	463 (61.1)	114 (60.0)
Corticosteroids/Cytotoxic agents, n (%)	200 (26.4)	58 (30.5)
Follow-up (months)	89.0 (50.0-130.2)	89.0 (50.0-149.0)
Year-man	6253	1700
Clinical Outcome		
ESKD or eGFR<15 ml/min/1.73m^2, n(%)	165 (21.8)	45 (23.7)

Data are expressed as mean ± SD, median (interquartile range), absolute and percent frequency.

Supplementary Table 4. Baseline characteristics of the IgAN patients of the study cohort divided in two different subsets based on the mean follow-up. Therapy, follow-up and clinical outcome are shown.

Mean Follow-up	5 years	10 years
Patients (n)	150	83
Age at biopsy (years)	41.2 $\pm$ 13.8	40.5+14.6
Sex (M/F)	110/40	65/18
Systolic blood pressure (mm Hg)	135.8 $\pm$ 19.2	137.9 $\pm$ 20.1
Diastolic blood pressure (mm Hg)	83.7 $\pm$ 12.1	84.9 $\pm$ 13.1
Mean arterial pressure (mm Hg)	101.1 $\pm$ 13.6	102.6 $\pm$ 14.5
Hypertension, n (%)	50 (33.3)	31 (37.3)
Serum creatinine (mg/dL)	1.30 (1.00-1.90)	1.32 (1.10-2.00)
eGFR MDRD (ml/min/1.73 m^2)	60.11 (40.81-85.12)	59.65 (40.25-78.02)
KDOQI stage		
I > 90 ml/min/1.73 m^2 , n (%)	32 (21.3)	15 (18.0)
II 60-89 ml/min/1.73 m^2, n (%)	43 (28.7)	23 (27.7)
III 30-59 ml/min/1.73 m^2 , n (%)	56 (37.3)	32 (38.6)
IV 15-29 ml/min/1.73 m^2 , n (%)	19 (12.7)	13 (15.7)
V <15 ml/min/1.73 m^2 , n (%)	0 (0.0)	0 (0)
Proteinuria (g/day)	1.60 (0.70-2.85)	1.90 (1.20-2.80)
Mild <1 g/day, n (%)	45 (30.0)	16 (19.3)
Moderate 1-3 g/day, n (%)	71 (47.3)	48 (57.8)
Severe >3 g/day, n (%)	34 (22.7)	19 (22.9)
Renal Biopsy		
M1, n/n (%)	59/150 (39.3)	35/83 (42.2)
E1, n/n (%)	17/150 (11.3)	9/83 (10.8)
S1, n/n (%)	112/150 (74.7)	69/83 (83.1)
T1, n/n (%)	35/150 (23.3)	24/83 (28.9)
T2, n/n (%)	6/150 (4.0)	5/83 (6.0)
C1, n/n (%)	18/150 (12.0)	10/83 (12.0)
Therapy after renal biopsy		
RASB, n (%)	90(60.0)	46 (55.5)
Corticosteroids/Cytotoxic agents, n (%)	40 (26.6)	20 (24.1)
Clinical Outcome		
ESKD or eGFR<15 ml/min/1.73m^2, n(%)	40 (26.7)	40 (48.2)

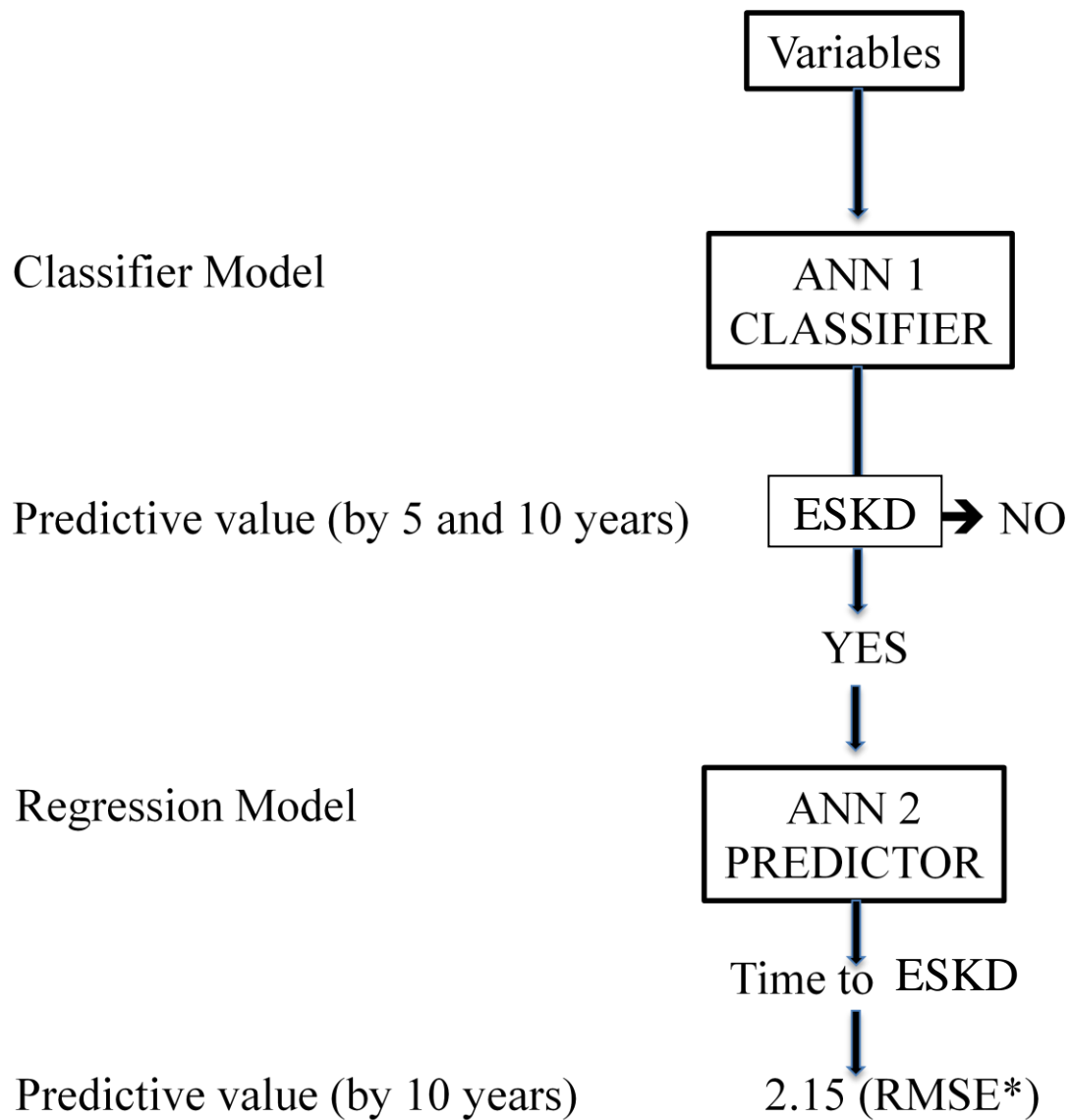Data are expressed as mean ± SD, median (interquartile range), absolute and percent frequency.

Supplementary Table 5. Comparisons between two models for prediction of ESKD risk

	Barbour et al, 2019	Our model
eGFR	CKD-EPI	MDRD 186
MEST classification	without C	with C
Composite endpoint	permanent decrease of eGFR below 50% of the value at time of kidney biopsy; eGFR <15ml/min/1.73 m^2; dialysis; transplantation	eGFR< 15 ml/min/1.73 m^2; dialysis; transplantation
Prediction	60-80 months	60-120 months
Model	Cox proportional hazard linear model	Artificial Neural Network Classifier and Regression Model
Tool name	llgAN-PT *	IgAN CDSS**

*   International lgAN Prediction Model (mobile app and website)

** IgAN Clinical Decision Support System (mobile app and website)

Supplementary Figure 1. The structure of the Artificial Neural Network (ANN) tool.

Classifier Model

ANN 1 CLASSIFIER

Predictive value (by 5 and 10 years)

ESKD → NO

YES

Regression Model

ANN 2 PREDICTOR

Time to ESKD

Predictive value (by 10 years)

2.15 (RMSE*)

Variables

*RMSE (Root Mean Square Error on Test Set).

Supplementary Figure 2 App screen of the Immunoglobulin A nephropathy Clinical Decision Support System (IgAN CDSS).
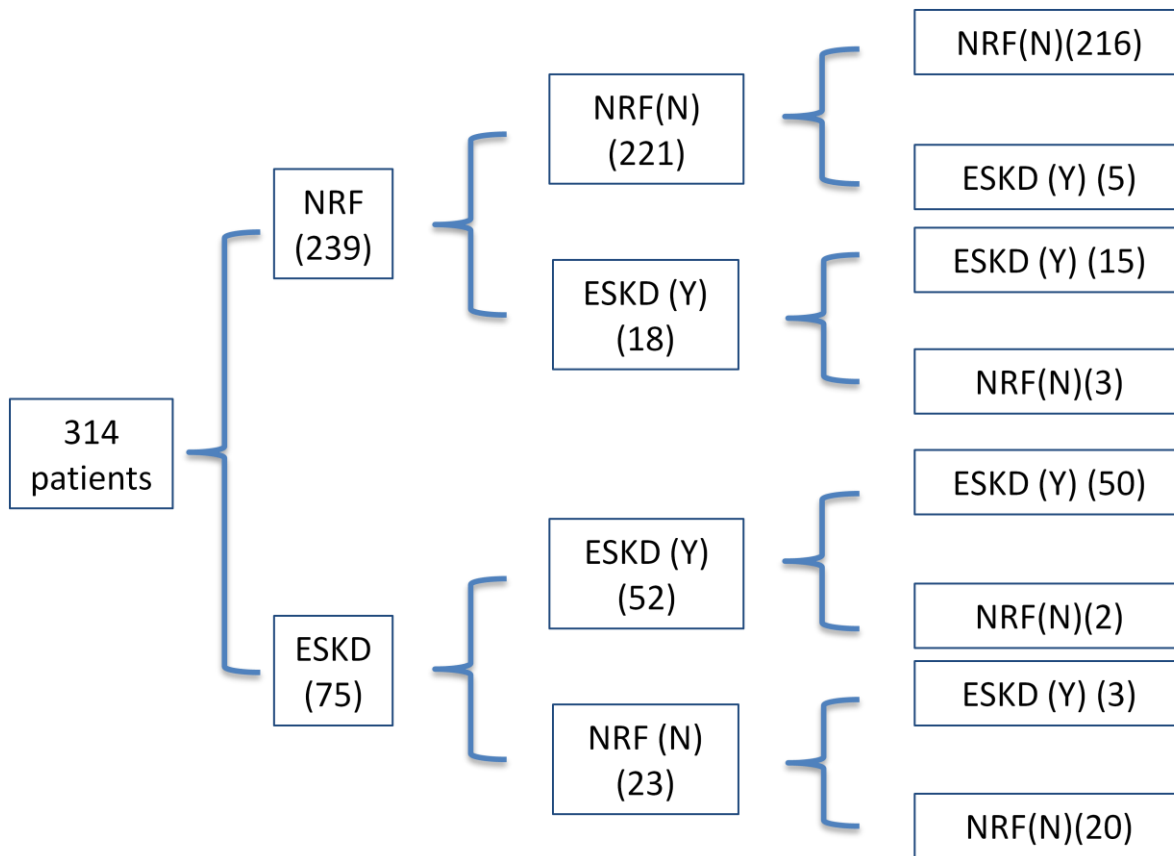
Supplementay Figure 3. IgAN patients of the Study Cohort treated with RASBs. The tool has been re-set after 12 months from the kidney biopsy

The ANN tool predicted no deterioration of the renal function (NRF) in 630/738 cases (85.3%) and ESKD in 143/210 cases (68%). Therefore, the prediction was concordant in 773 patients (81.5%) and discordant in 175 cases (18.5%).