Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 12 May 2026
Revised 20 June 2026
Accepted 12 July 2026

Natural Resources for Human Health 2026, Vol. 6 (6s): 1333–1349

KEYWORDS:
Renal impairment, incomplete records, affine conditioning, deep ensemble, probability calibration, screening support.

MPC-Net: Detection of Chronic Kidney Disease Using Missingness-Pattern Conditioned Deep Learning

Prathap Paduvalahippe Basavaraju[1], Sumit Gupta[2], Sakshi Khullar[3], Seema H R[4], Gadhiraju Tej Varma[5], H. Nagesh Shenoy[6], Amaresh A M[7]

[1]Associate Professor, Department of Electronics & Communication Engineering, Navkis College of Engineering, Hassan, India. Email: prathap_0293@navkisce.ac.in
[2]Director, DeepCognix AI Labs, Bengaluru, Karnataka, India. Email: sumit@deepcognix.com. ORCID ID: 0009-0007-4766-3468
[3]Vivekananda Institute of Professional Studies - Technical Campus (VIPS TC) , New Delhi, India. Email: sakshikhullar2809@gmail.com
[4]Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, India. Email: seemahr@vvce.ac.in
[5]Assistant Professor, Department of IT, SRKR Engineering College, Bhimavaram, India. ORCID: 0000-0003-3477-4197. Email: tejvarma.varma@gmail.com
[6]Dept. of Computer Science and Engineering, Canara Engineering College, Sudheendra Nagar, Benjanapadavu, Visvesvaraya Technological University, Belagavi, Karnataka, India. Email: h.nagesh.shenoy@gmail.com
[7]Assistant Professor, Department of Computer Science and Engineering, JSS Science and Technology University, mysuru,570006, Karnataka, India. ORCID ID: 0000-0001-7499-6558. Email: amaresham@jssstuniv.in
*Corresponding Author: Prathap Paduvalahippe Basavaraju (prathap_0293@navkisce.ac.in)

ABSTRACT: Renal impairment is usually discovered late, because the biochemical panels capable of exposing it are ordered irregularly and returned with gaps. Standard practice fills those gaps with a column statistic and hands the repaired vector to a classifier, so a value the laboratory produced and a value the algorithm invented become indistinguishable. We take the opposite view and treat the pattern of what was left unmeasured as a second input channel. The proposed MPC-Net encodes that binary pattern through its own subnetwork and uses the resulting code to apply a feature-wise affine transform to the value pathway, so absence modulates how the recorded numbers are interpreted instead of impersonating them. The value pathway itself carries an explicit second-order interaction term, letting variable pairs contribute without an enumerated expansion, and two gated residual blocks refine the fused representation. Five independently initialised replicas vote, and a logistic recalibrator maps the pooled score onto a probability scale. Evaluation covers 600 patient records in two cohorts under five-fold partitioning repeated five times. On the first cohort MPC-Net attained 99.80 ± 0.58 per cent accuracy, 100.00 per cent area under the curve and a Brier score of 0.0013, and on the second 98.00 per cent accuracy. Permutation testing identified specific gravity and haemoglobin as the dominant evidence.

1. INTRODUCTION

Roughly one adult in ten carries some degree of permanent kidney damage, and the condition claimed on the order of 1.2 million lives in 2017, a toll growing more quickly than that of most comparable chronic diseases [1]. What makes renal failure

unusual among them is the silence of its early course. Filtration capacity can decline by two thirds before a patient notices anything worth reporting to a physician, which is why so many diagnoses are made at a point where dialysis or transplantation are the only remaining options [2]. Since symptoms arrive too late to be useful, the burden of early identification falls on biochemistry, and the practical question becomes how much diagnostic value can be extracted from the panels that primary care already orders. Predictive methods spanning classical learning through modern deep architectures have been surveyed across this area [3].

Ordering, however, is irregular. In the primary cohort examined below, only 60.8 per cent of records are free of at least one unrecorded variable, and 10.55 per cent of the individual cells carry no value at all. Nor are the holes randomly placed. Microscopy-dependent quantities such as red cell morphology go unrecorded several times more often than analyser-produced quantities such as creatinine, because one requires a technician and a slide while the other falls out of a machine. The consequence is that the emptiness in a record encodes something: which investigations the attending physician thought worth requesting, which specimens proved usable, and what the collecting site could afford. Discarding incomplete rows throws away the majority of the evidence, and substituting a column median destroys the distinction between what was measured and what was manufactured [4], [5].

There is a second gap in the literature, less often acknowledged. Reported performance on the reference renal panel has converged: numerous methods of unrelated design all announce accuracies in the high nineties [6], [7], [8]. When a metric no longer separates candidates, continuing to rank by it conveys nothing. What remains unexamined is whether the number a model attaches to its own prediction is trustworthy. If a tool declares eighty per cent confidence, roughly four of every five such patients ought to be diseased; when that correspondence fails, the output cannot inform triage and functions merely as a label [9], [10].

Our response to both gaps is a single architecture built around one idea: the missingness pattern is a modality in its own right, and should be encoded as such rather than repaired away. The specific contributions are:

1. A dedicated pattern encoder that consumes the whole binary indicator vector of a record and produces a compact code describing the shape of what is absent, in contrast to per-variable substitution schemes.
2. A conditioning mechanism in which that code emits scale and shift coefficients applied feature-wise to the value representation, so absence alters the reading of the measured values rather than standing in for them.
3. A value pathway that augments the linear projection with a factorised second-order term, admitting pairwise effects at cost linear in the number of variables.
4. An uncertainty treatment combining 5 independently seeded replicas with logistic recalibration, assessed through Brier score, negative log likelihood and expected calibration error in addition to the usual discrimination measures.
5. Relevance attribution by out-of-fold permutation testing, which measures what the fitted model actually depends on instead of assuming that some internal quantity is interpretable.

The remainder proceeds as follows. Section 2 surveys prior work organised by how each study treats incompleteness. Section 3 documents the cohorts and preprocessing. Section 4 specifies the architecture. Section 5 gives the evaluation design, Section 6 the findings, Section 7 their interpretation and limits, and Section 8 closes.

2. BACKGROUND AND RELATED WORK

Studies that repair the record first. Nearly every published renal detector begins by eliminating incompleteness. Working on the panel released by Rubini and colleagues [11], Rady and Anwar contrasted probabilistic networks against perceptrons, kernel machines and radial basis models for stage assignment, with the probabilistic variant leading overall but dropping to 96.9 per cent precisely at the middle stages where a clinician most needs help [12]. Yashfi and co-workers folded behavioural variables into the panel and obtained 97.12 per cent from a forest over a twenty-variable subset [13]. Almasoud and Ward inverted the usual objective, asking not for maximum accuracy but for minimum expenditure, and established that boosted trees over three cheap assays preserve much of the panel's discriminative content [14]. The most exhaustive sweep is that of Chittora and colleagues, who crossed seven learners with feature selection and synthetic balancing and topped out at 98.46 per cent using a linear kernel machine over the complete variable set [6]. Senan and co-workers reached comparable territory

by pairing recursive elimination with four classifiers [8], and Amaresh and Sundaram optimised the retained subset with a swarm procedure inside a screening framework [15]. Every one of these pipelines settles the question of which variables to keep before learning starts, and every one settles the question of what to do about unmeasured cells by overwriting them. None reports whether its output probabilities are calibrated.

Studies that model the incompleteness. A smaller body of work treats the gaps as objects of inference. Qin and colleagues route each record to either a neighbour-based or a forest-based branch depending on a confidence test, and use the routing machinery to reconstruct absent entries prior to classification [7], which is closer in spirit to what we propose but still terminates in a repaired vector. In the statistical literature the distinction between data missing at random and missing for reasons connected to the outcome has been formalised for half a century [4], and chained-equation imputation remains the reference procedure [5]; neither, however, supplies a representation that a discriminative network can condition upon.

Architectural context. Deep models for tables have converged on embedding schemes. Categorical fields are embedded and passed through self-attention in TabTransformer [16]; TabNet applies sequential attention to choose a sparse subset per decision step [17]; FT-Transformer embeds every field and reads out through a class token [18]. All three inherit their machinery from the sequence transformer [19], and none of them represents observation status. Conditioning one representation on another by predicted scale and shift coefficients originates in visual reasoning [20] and has since become a standard fusion primitive; the factorised pairwise term in our value pathway derives from factorisation machines [21]. Elsewhere in abdominal diagnosis the field has turned to generative formulations, as in the conditional diffusion and graph-regularised segmentation of pancreatic tumours reported by Radhika and co-workers [22], and to convolutional pipelines deployed on connected-device platforms [23]; at a larger scale still, recurrent models over longitudinal renal trajectories can flag impending injury days ahead [24]. These exploit spatial or temporal regularity that a single outpatient panel does not possess.

Uncertainty. Contemporary networks are reliably overconfident, and a scalar temperature estimated on held-out data repairs most of the distortion without disturbing the induced ranking [9]; a logistic recalibrator generalises this by fitting slope and intercept together [25]. Averaging several independently initialised replicas improves both the point estimate and the spread [26]. Both devices appear in the read-out of Section 4.6.

3. MATERIALS AND PREPROCESSING

Table 1 summarises the two record sets.

Table 1. Description of the two record sets.

Property	Cohort A	Cohort B
Patients	400	200
Renal diagnosis present	250	128
Renal diagnosis absent	150	72
Fields modelled	24	24
Quantitative fields	14	14
Categorical fields	10	10
Vacancy rate (%)	10.55	4.17
Patients with a vacancy (%)	60.8	100.0
Numeric form	Continuous	Interval centre

Cohort A originates from a nephrology service in Tamil Nadu, India, was assembled over a two-month window, and is distributed through the UCI repository [11]. It comprises 400 patients, 250 carrying a confirmed renal diagnosis against 150 who do not, characterised by 24 fields, of which 14 are quantitative and 10 categorical. Unrecorded cells are flagged in the

Natural Resources for Human Health
eISSN: 2583-1194

distributed file and were retained as flags rather than repaired at load time. The overall vacancy rate is 10.55 per cent, and 60.8 per cent of patients have at least one field empty.

Cohort B follows the same clinical protocol but was curated separately, with all continuous quantities banded into ordered intervals and with filtration rate and stage supplied as extra columns. It holds 200 patients, 128 positive against 72 negative. We converted each band label to its interval centre, and each open band to its finite endpoint, so that the field set corresponds to cohort A. Diastolic pressure survives in this release only as a dichotomous flag rather than a pressure reading, so it was declared unrecorded across the whole cohort and left to the model's own absence machinery. Filtration rate and stage were excluded, as both are computed from the diagnosis and would leak it.

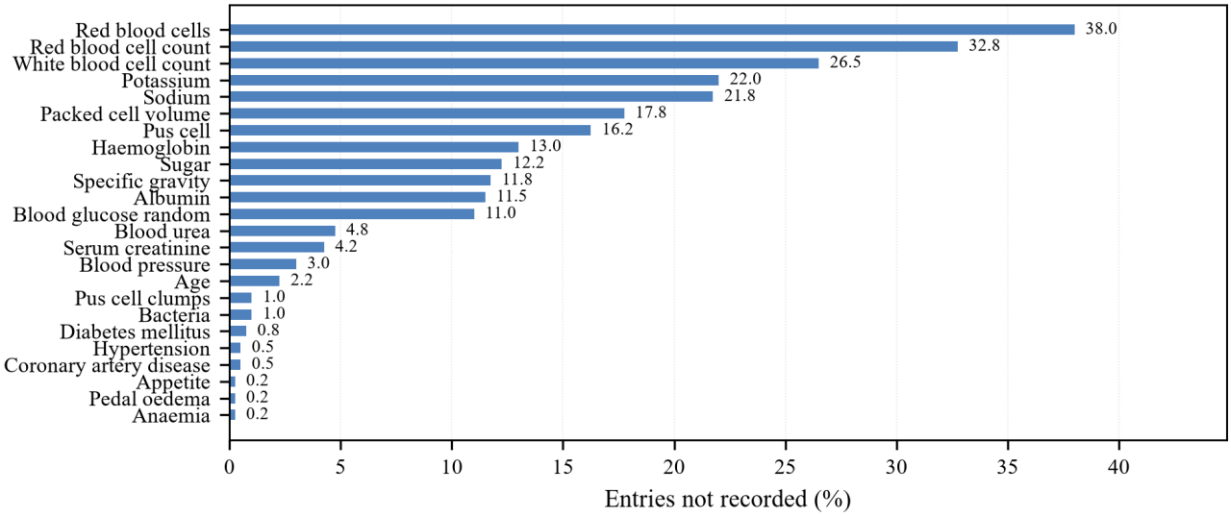

Fig. 1. Unrecorded entries by variable in cohort A.

Figure 1 makes the unevenness plain. Morphological assessment of red cells is unrecorded for over a third of patients, while age and pressure are almost universally present, and the gradient between them tracks the cost and labour of the corresponding investigation. Table 2 enumerates the fields with type, unit, cohort A mean, dispersion and vacancy rate.

Table 2. Field inventory with cohort A statistics.

Field	Type	Unit	Mean	Dispersion	Vacant (%)
Age	Numeric	years	51.48	17.15	2.25
Blood pressure	Numeric	mm Hg	76.47	13.67	3.00
Specific gravity	Numeric	-	1.02	0.01	11.75
Albumin	Numeric	0-5 scale	1.02	1.35	11.50
Sugar	Numeric	0-5 scale	0.45	1.10	12.25
Blood glucose random	Numeric	mg/dL	148.04	79.17	11.00
Blood urea	Numeric	mg/dL	57.43	50.44	4.75
Serum creatinine	Numeric	mg/dL	3.07	5.73	4.25
Sodium	Numeric	mEq/L	137.53	10.39	21.75
Potassium	Numeric	mEq/L	4.63	3.19	22.00
Haemoglobin	Numeric	g/dL	12.53	2.91	13.00

Field	Type	Unit	Mean	Dispersion	Vacant (%)
Packed cell volume	Numeric	%	38.88	8.98	17.75
White blood cell count	Numeric	cells/cmm	8406.12	2939.46	26.50
Red blood cell count	Numeric	millions/cmm	4.71	1.02	32.75
Red blood cells	Nominal	binary	0.19	0.39	38.00
Pus cell	Nominal	binary	0.23	0.42	16.25
Pus cell clumps	Nominal	binary	0.11	0.31	1.00
Bacteria	Nominal	binary	0.06	0.23	1.00
Hypertension	Nominal	binary	0.37	0.48	0.50
Diabetes mellitus	Nominal	binary	0.35	0.48	0.75
Coronary artery disease	Nominal	binary	0.09	0.28	0.50
Appetite	Nominal	binary	0.21	0.40	0.25
Pedal oedema	Nominal	binary	0.19	0.39	0.25
Anaemia	Nominal	binary	0.15	0.36	0.25

Preprocessing is deliberately minimal and is estimated inside each training partition. The indicator matrix is extracted first. Quantitative fields are then filled with the training-partition median and standardised by training-partition mean and dispersion; categorical fields are coded as indicators and filled the same way. Because the indicator matrix is taken before any filling occurs, the network retains access to the original vacancy structure even though its value channel sees a dense vector.

4. THE PROPOSED MPC-NET

4.1 Design rationale

Two channels describe a patient record: the numbers that were produced, and the layout of the ones that were not. A design that repairs the second channel into the first collapses them and loses the layout. A design that gives each field its own token and its own substitute vector preserves the layout only field by field, and cannot represent the shape of the vacancy as a whole, namely that a particular record lacks the entire microscopy block, say, rather than one arbitrary field. MPC-Net keeps the two channels separate, encodes the vacancy layout jointly, and lets it govern the reading of the value channel. Figure 2 shows the arrangement.

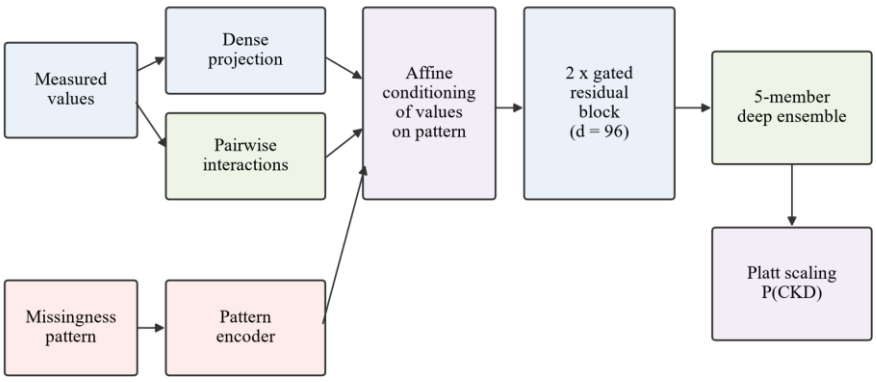

Fig. 2. Structure of the proposed MPC-Net.

Natural Resources for Human Health
eISSN: 2583-1194

4.2 Value pathway

Write $x \in \mathbb{R}^F$ for the filled and standardised record over $F = 24$ fields. A dense projection supplies the first-order representation, $h_1 = \phi(W_v x + b_v)$ with ϕ the Gaussian error linear unit and $W_v \in \mathbb{R}^{d \times F}$, $d = 96$. Renal biochemistry is, however, read in combinations rather than singly: a creatinine value is interpreted against haemoglobin, an albumin reading against specific gravity. An explicit pairwise expansion would cost $O(F^2)$ terms, so we adopt the factorised form, assigning each field a rank-k factor $v_j \in \mathbb{R}^k$, $k = 8$, and computing

$$z = 1/2 \left[(\sum_{j=1}^{F} v_j \, x_j)^2 - \sum_{j=1}^{F} v_j^2 \, x_j^2 \right] \tag{1}$$

elementwise in the factor space [21]. This is algebraically equal to $\sum_{i<j} \langle v_i, v_j \rangle x_i x_j$ yet costs $O(Fk)$. A linear map lifts z into the representation width, giving $h = h_1 + W_z z$.

4.3 Pattern encoder

Let $m \in \{0,1\}^F$ carry unity wherever a field went unrecorded. Rather than attaching a substitute to each empty field, the whole indicator vector is passed through a two-layer subnetwork,

$$c = \phi(U_2 \, \phi(U_1 m + e_1) + e_2) \in \mathbb{R}^{d_p} \tag{2}$$

with $d_p = 48$. Because c is a function of the entire vector, it can represent joint structure: a code for records lacking the full microscopy block differs from a code for records lacking one field in isolation, even when the two share a vacancy count.

4.4 Conditioning the values on the pattern

The vacancy code does not enter as an additional feature. It instead emits a scale and a shift applied elementwise to the value representation,

$$\tilde{h} = (1 + \gamma(c)) \odot h + \beta(c) \tag{3}$$

where γ and β are linear maps into $\mathbb{R}^d$ initialised at zero, so training begins from the identity transform and departs from it only where the vacancy layout proves informative [20]. This is the mechanical heart of the model. A dimension of h that summarises, say, anaemia-related evidence can be amplified in records whose haematological block is complete and suppressed in those where it is not, and the suppression is decided by the layout as a whole rather than by any single flag.

4.5 Gated residual refinement

Two identical blocks refine $\tilde{h}$. Each normalises its input [27], computes a candidate update through a widening and narrowing pair of projections, and admits that update through a learned elementwise gate,

$$h \leftarrow h + \delta\big(\sigma\big(W_g u\big) \odot W_2 \, \phi(W_1 u)\big), \quad u = \mathrm{LN}(h) \tag{4}$$

with σ the logistic function and δ dropout at 0.15. The gate lets a block decline to modify dimensions it has nothing to add to, which stabilises fitting on a set of this size. A final normalisation and a scalar projection yield the logit. One replica holds 110 017 trainable parameters.

4.6 Ensemble and recalibration

Two devices convert that logit into a usable probability. First, 5 replicas are trained from independent initialisations and independent shuffling, and their logits are averaged; unlike snapshots drawn from one trajectory, independent replicas explore separate optima and their disagreement is informative [26]. Second, the pooled logit $\bar{\eta}$ is mapped through a logistic recalibrator whose slope a and intercept b are estimated on a partition held out of fitting [25], [9],

$$\hat{p} = \big(1 + \exp\big(-(a \, \bar{\eta} + b)\big)\big)^{-1} \tag{5}$$

Fitting slope and intercept jointly, rather than a temperature alone, allows correction of a systematic offset as well as of over-sharpness, which matters when the classes are unbalanced.

4.7 Optimisation

Fitting minimises binary cross-entropy against targets smoothed by 0.06 [28], using AdamW with decoupled decay [29] on a cosine-annealed rate [30] and gradients clipped to unit norm. Table 3 records every setting. All were fixed in advance and shared by both cohorts.

Table 3. Architecture and optimisation settings.

Setting	Value
Representation width d	96
Pattern code width	48
Interaction rank k	8
Gated residual blocks	2
Dropout	0.15
Parameters per replica	110 017
Ensemble members	5
Parameters in total	550 085
Optimiser	AdamW
Learning rate	3 x 10^-3
Weight decay	5 x 10^-3
Schedule	Cosine annealing
Epochs	90
Batch size	32
Label smoothing	0.06
Gradient clipping	1.0
Recalibration split	18% of fitting partition

5. EXPERIMENTAL DESIGN

Partitioning. Each cohort was assessed by stratified five-fold partitioning repeated five times under distinct random states, producing 25 disjoint fit-and-test cycles per method. Medians, scaling constants and the recalibration split were all derived within the fitting partition and applied unaltered to the withheld one, so no test record influences any fitted quantity. A further eighteen per cent of each fitting partition was withdrawn for recalibration and excluded from parameter updates.

Reference methods. Nine comparators were exercised on identical partitions with identical preprocessing: logistic regression, a decision tree, seven-neighbour classification, a radial-basis kernel machine, a random forest [31], extremely randomised trees [32], adaptive boosting [33], gradient-boosted trees [34] and a two-layer perceptron. Each receives median-filled values without any indicator channel, matching what the renal literature does [12], [6], [8].

Measures. Accuracy, precision, recall, specificity and F1 are computed at the half threshold, together with area under the receiver curve, the Matthews coefficient [35], Cohen’s kappa [36], the Brier score [37], negative log likelihood and expected calibration error over ten equal-width bins. The correlation coefficient and kappa are included because both remain meaningful under the class ratios present here; the three probabilistic measures are included because discrimination alone no longer distinguishes candidates on this panel.

Inference. Paired comparisons use the Wilcoxon signed-rank statistic over matched partition scores [38], the recommended instrument for two learners across common splits when normality cannot be assumed [39].

Attribution. Field relevance is estimated by permutation: within each withheld partition, one column is shuffled, the loss in area under the curve is recorded, and the procedure is repeated five times per column and averaged. The estimate concerns the fitted predictor rather than any internal activation, and so cannot be inflated by reading structure into a quantity that has none.

Software and cost. Networks are built in PyTorch [40], comparators in scikit-learn [41] and XGBoost [34]. Everything runs on the processor. The full cohort A protocol, ensembles and permutation testing included, occupies 1148 s. A single replica classifies one record in 1.00 ms and the full ensemble in 4.99 ms.

6. RESULTS AND ANALYSIS

6.1 Detection quality

Table 4 gives cohort A. MPC-Net records 99.80 ± 0.58 per cent accuracy, 99.68 per cent recall, 100.00 per cent specificity, 99.84 per cent F1 and 100.00 per cent area under the curve, with a Matthews coefficient of 0.9959 and a kappa of 0.9958. Across 2000 withheld decisions the tally is 4 missed cases and 0 false alarms. The asymmetry favours recall, which is the correct priority here: a false alarm costs a confirmatory assay, whereas a missed case forfeits the interval during which decline can still be slowed.

Table 4. Partitioned performance on cohort A.

Method	Accuracy	Precision	Recall	Specificity	F1	AUC	MCC	Kappa
MPC-Net	99.80 ± 0.58	100.00 ± 0.00	99.68 ± 0.93	100.00 ± 0.00	99.84 ± 0.47	100.00 ± 0.00	0.9959	0.9958
Logistic regression	99.50 ± 0.71	100.00 ± 0.00	99.20 ± 1.13	100.00 ± 0.00	99.60 ± 0.57	100.00 ± 0.00	0.9896	0.9894
Decision tree	97.15 ± 1.25	98.33 ± 1.55	97.12 ± 1.88	97.20 ± 2.61	97.70 ± 1.02	97.16 ± 1.33	0.9402	0.9395
k-nearest neighbours	96.25 ± 2.32	100.00 ± 0.00	94.00 ± 3.71	100.00 ± 0.00	96.87 ± 2.01	99.57 ± 0.66	0.9258	0.9221
Support vector machine	99.90 ± 0.49	100.00 ± 0.00	99.84 ± 0.78	100.00 ± 0.00	99.92 ± 0.40	100.00 ± 0.00	0.9979	0.9979
Random forest	99.20 ± 0.86	99.14 ± 1.12	99.60 ± 0.80	98.53 ± 1.90	99.36 ± 0.68	99.97 ± 0.06	0.9831	0.9829
Extra trees	99.90 ± 0.34	100.00 ± 0.00	99.84 ± 0.54	100.00 ± 0.00	99.92 ± 0.27	100.00 ± 0.00	0.9979	0.9979
AdaBoost	99.20 ± 0.86	99.68 ± 0.73	99.04 ± 1.15	99.47 ± 1.22	99.36 ± 0.69	99.99 ± 0.03	0.9832	0.9830
XGBoost	98.65 ± 0.99	98.90 ± 1.12	98.96 ± 1.51	98.13 ± 1.90	98.92 ± 0.81	99.90 ± 0.13	0.9716	0.9712
Multilayer perceptron	99.25 ± 1.00	99.92 ± 0.39	98.88 ± 1.51	99.87 ± 0.65	99.39 ± 0.82	99.99 ± 0.03	0.9844	0.9841

Among the comparators the leader is the support vector machine at 99.90 per cent and the laggard the k-nearest neighbours at 96.25 per cent. Two comparators therefore edge the proposal on accuracy by a tenth of a point, a margin an order of magnitude smaller than the partition-to-partition dispersion of either, and one that reverses between repetitions. The compression at the top end is exactly the convergence described in Section 1, and it is the reason the probabilistic analysis of Section 6.3 does the real work of separation.

Table 5 repeats the exercise on cohort B, where quantitative fields survive only as interval centres and one field is unrecorded throughout. MPC-Net returns 98.00 ± 2.00 per cent accuracy and 99.96 per cent area under the curve. Here the ordering changes: the random forest leads at 99.70 per cent, and the proposal does not hold the top position. The reason is visible in the cohort description. Cohort B has been banded, so every quantitative field arrives as an interval centre, and one field is unrecorded for every patient rather than for some, which leaves the pattern encoder with an almost constant input and nothing to condition on. A model whose distinguishing mechanism is the variation in vacancy layout has no such variation to exploit here, and it degrades to an ordinary gated network. The cohort is retained because that degradation is informative, not because it flatters the design.

Table 5. Partitioned performance on cohort B.

Method	Accuracy	Precision	Recall	Specificity	F1	AUC	MCC	Kappa
MPC-Net	98.00 ± 2.00	99.55 ± 1.22	97.32 ± 3.31	99.16 ± 2.27	98.38 ± 1.66	99.96 ± 0.13	0.9591	0.9575
Logistic regression	98.60 ± 2.01	99.39 ± 1.40	98.41 ± 2.98	98.90 ± 2.53	98.87 ± 1.66	99.99 ± 0.05	0.9712	0.9703
Decision tree	96.90 ± 1.91	97.74 ± 2.13	97.46 ± 2.96	95.77 ± 4.00	97.55 ± 1.54	96.61 ± 2.02	0.9345	0.9330
k-nearest neighbours	96.50 ± 3.24	100.00 ± 0.00	94.50 ± 5.17	100.00 ± 0.00	97.10 ± 2.88	99.83 ± 0.45	0.9309	0.9269
Support vector machine	99.30 ± 1.50	99.55 ± 1.22	99.36 ± 2.17	99.16 ± 2.27	99.44 ± 1.23	100.00 ± 0.00	0.9856	0.9851
Random forest	99.70 ± 0.81	99.55 ± 1.22	100.00 ± 0.00	99.16 ± 2.27	99.77 ± 0.62	99.99 ± 0.05	0.9935	0.9934
Extra trees	99.50 ± 1.00	100.00 ± 0.00	99.21 ± 1.58	100.00 ± 0.00	99.60 ± 0.80	100.00 ± 0.00	0.9896	0.9894
AdaBoost	97.10 ± 2.62	98.44 ± 2.22	97.02 ± 3.00	97.24 ± 3.88	97.70 ± 2.09	99.58 ± 0.94	0.9388	0.9377
XGBoost	97.60 ± 2.40	98.17 ± 2.18	98.12 ± 2.51	96.63 ± 4.05	98.12 ± 1.87	99.51 ± 0.90	0.9486	0.9479
Multilayer perceptron	98.50 ± 2.00	99.70 ± 1.02	97.94 ± 3.21	99.45 ± 1.87	98.78 ± 1.67	100.00 ± 0.00	0.9695	0.9683

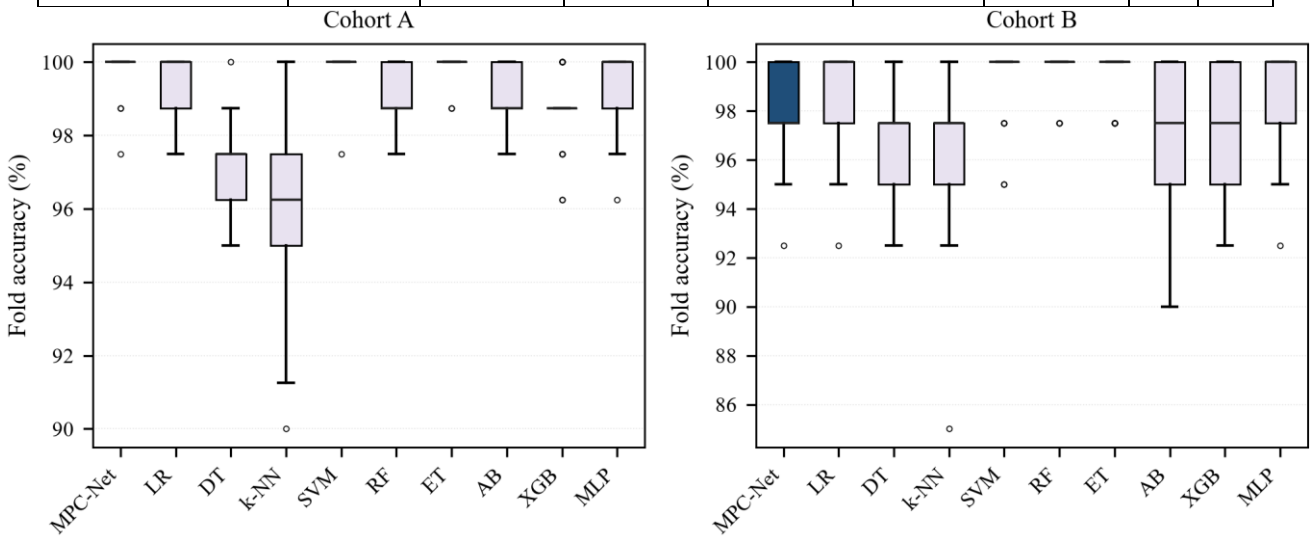

Fig. 3. Fold accuracy spread across both cohorts.

Figure 3 displays the whole distribution of partition scores rather than its centre. What a deployed screening tool requires is a short box as much as a high median, since the box height reports how much the answer depends on which patients happen to fall in the test partition.

Natural Resources for Human Health
eISSN: 2583-1194

6.2 Discrimination behaviour

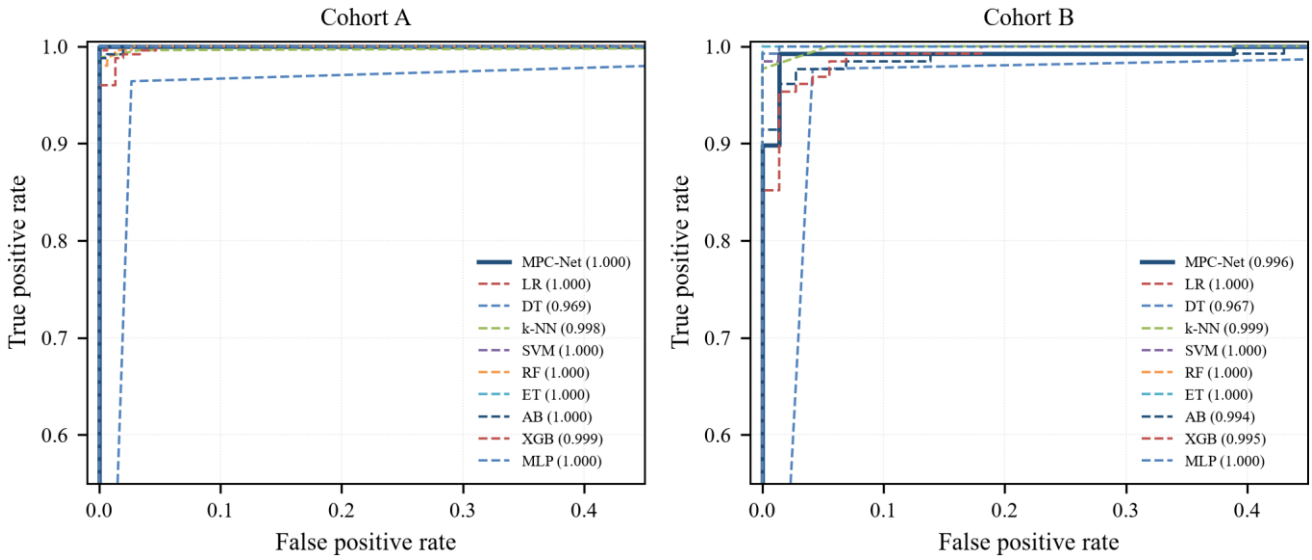

Fig. 4. Receiver curves from withheld predictions.

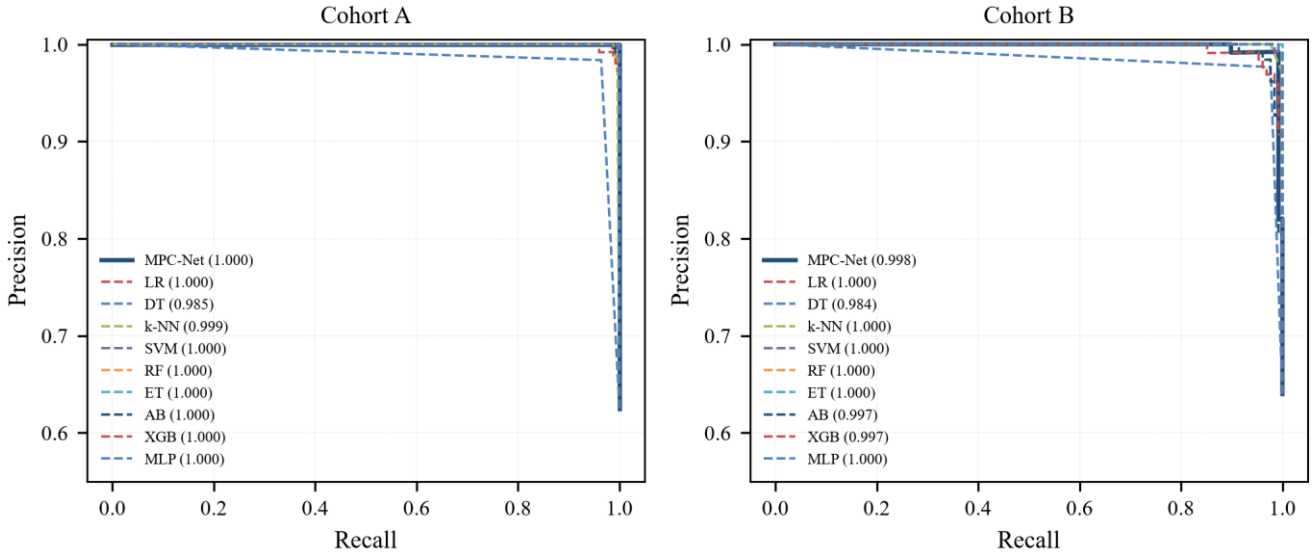

Fig. 5. Precision against recall for every method.

Figure 4 and Figure 5 are constructed from withheld predictions only, so each plotted point corresponds to a record the relevant model had not seen. Axes are cropped to the region where the traces actually diverge. The proposed model sustains sensitivity deep into the low-false-alarm zone, which is where a screening instrument is obliged to operate: at any fixed budget for unnecessary follow-up, the share of genuine cases still captured is what decides whether the instrument is worth deploying.

Natural Resources for Human Health
eISSN: 2583-1194

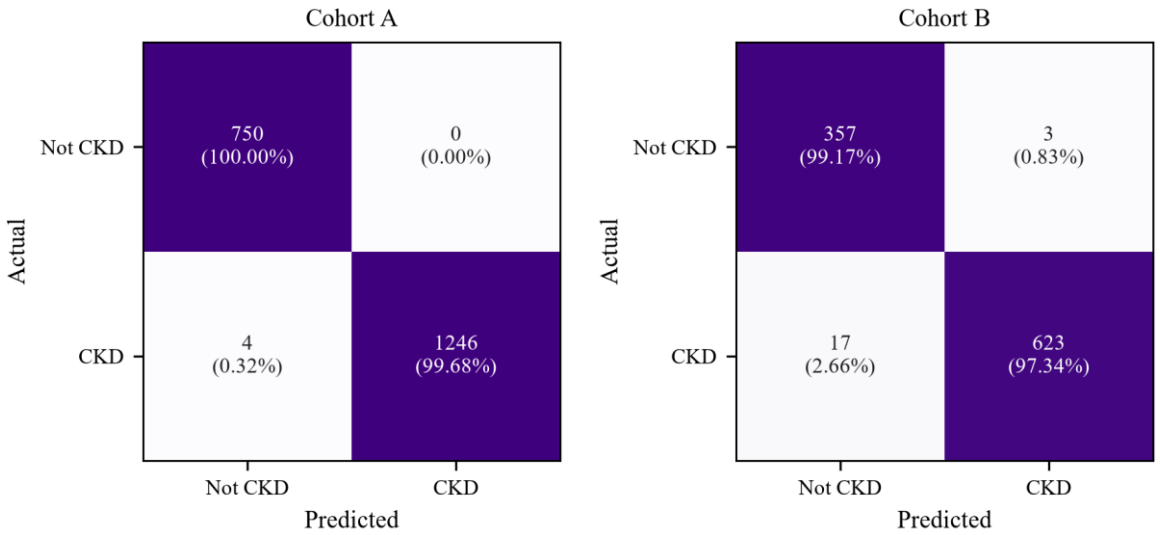

Fig. 6. Pooled confusion matrices for MPC-Net.

Figure 6 pools the confusion counts. Cohort A yields 4 errors in 2000 decisions, split 4 missed against 0 spurious; cohort B yields 17 and 3 across 1000.

6.3 Probability quality

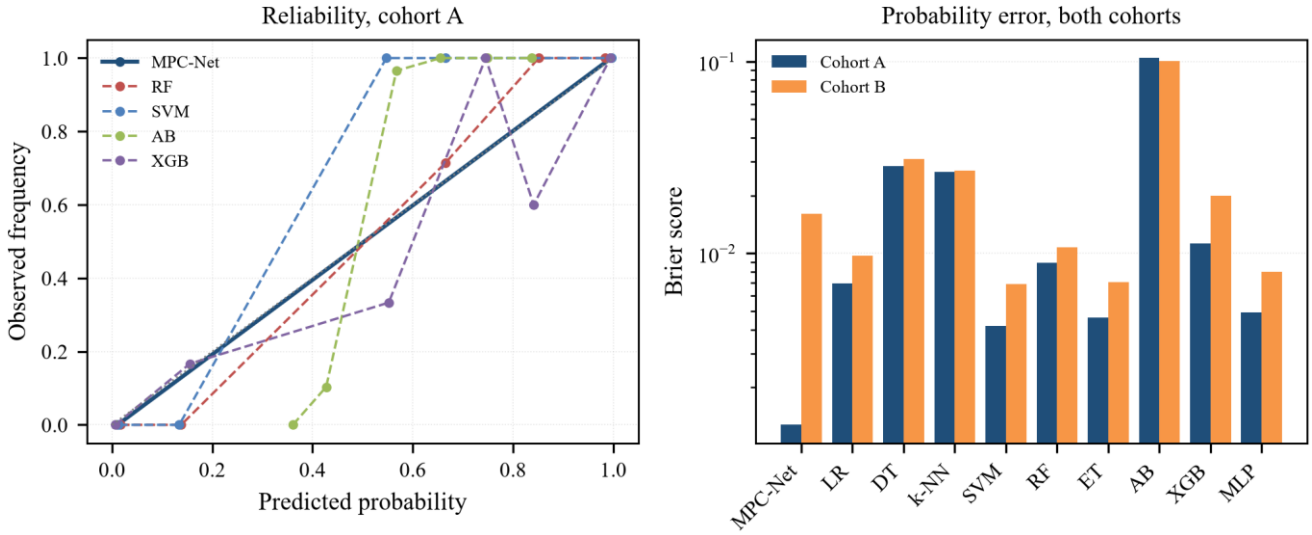

Fig. 7. Reliability traces and Brier scores compared.

Table 6 sets out Brier score, expected calibration error and negative log likelihood for every method on both cohorts, and Figure 7 draws the reliability traces. MPC-Net returns a Brier score of 0.0013 on cohort A with expected calibration error 0.0075 and negative log likelihood 0.0113; the nearest comparator, the support vector machine, sits at 0.0042, a factor of 3.3 higher in squared probability error. The advantage does not carry over to cohort B, where MPC-Net records 0.0161 against 0.0069 for the support vector machine and several comparators calibrate better. This is reported rather than omitted, and it is consistent with the account just given: the mechanism that sharpens the probabilities depends on vacancy structure that cohort B does not contain. The reliability trace of the proposal follows the diagonal, meaning that among patients assigned a given probability, close to that proportion are genuinely affected. Boosted ensembles show the opposite signature, crowding their outputs toward the extremes and reporting confidence the observed frequencies do not support [10], and the mid-range they distort is precisely where the ambiguous patient sits.

Table 6. Probability quality on both cohorts.

Method	Brier (A)	ECE (A)	NLL (A)	Brier (B)	ECE (B)
MPC-Net	0.0013	0.0075	0.0113	0.0161	0.0145
Logistic regression	0.0070	0.0268	0.0335	0.0097	0.0328
Decision tree	0.0285	0.0325	0.4594	0.0310	0.0300
k-nearest neighbours	0.0266	0.0404	0.1442	0.0269	0.0536
Support vector machine	0.0042	0.0189	0.0209	0.0069	0.0213
Random forest	0.0089	0.0342	0.0475	0.0107	0.0423
Extra trees	0.0046	0.0247	0.0287	0.0070	0.0286
AdaBoost	0.1041	0.3005	0.3787	0.1003	0.2577
XGBoost	0.0113	0.0164	0.0412	0.0200	0.0127
Multilayer perceptron	0.0049	0.0101	0.0176	0.0080	0.0175

6.4 Significance of the differences

Table 7 lists Wilcoxon signed-rank outcomes over the 25 matched partitions of cohort A, for accuracy, area under the curve and Brier score. Of the 9 comparators, 6 are beaten on accuracy at the five per cent level and 9 on Brier score. The disparity between those two counts is itself the result worth reporting: on a panel this separable many methods are indistinguishable in their decisions, yet clearly distinguishable in the confidence they attach to them.

Table 7. Signed-rank outcomes against each comparator.

Comparator	Accuracy gain	p (accuracy)	p (AUC)	p (Brier)
Logistic regression	+0.30	0.0702	1.0000	< 0.0001
Decision tree	+2.65	< 0.0001	< 0.0001	< 0.0001
k-nearest neighbours	+3.55	< 0.0001	0.0025	< 0.0001
Support vector machine	-0.10	0.7136	1.0000	< 0.0001
Random forest	+0.60	0.0116	0.0086	< 0.0001
Extra trees	-0.10	0.7136	1.0000	< 0.0001
AdaBoost	+0.60	0.0190	0.0987	< 0.0001
XGBoost	+1.15	0.0001	0.0014	< 0.0001
Multilayer perceptron	+0.55	0.0214	0.4658	< 0.0001

Natural Resources for Human Health
eISSN: 2583-1194

6.5 What the model relies on

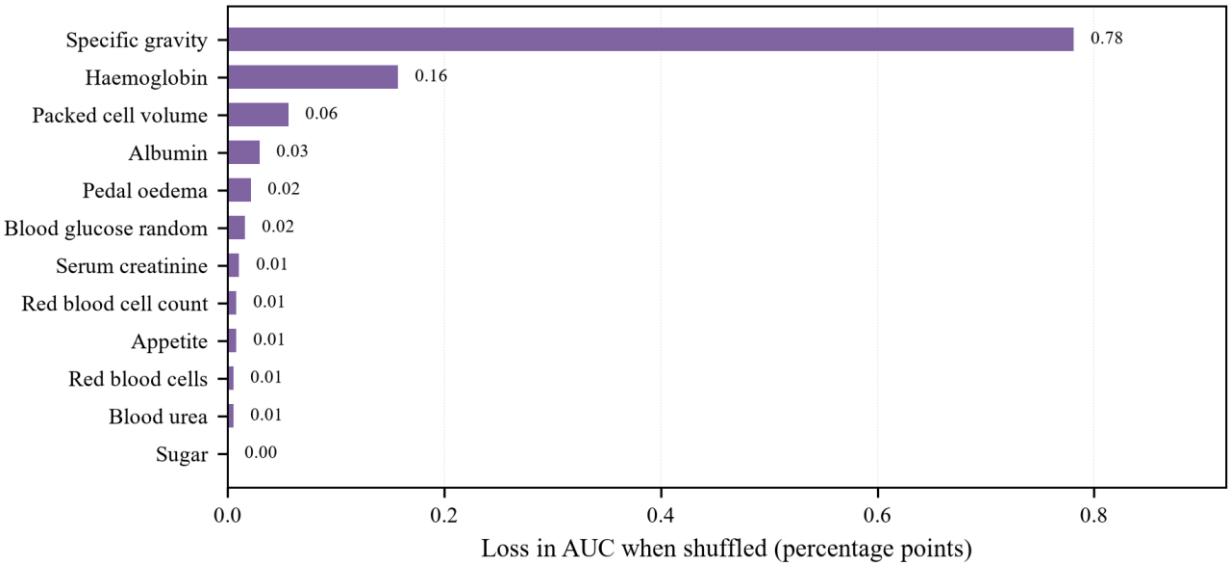

Fig. 8. Permutation relevance of the leading variables.

Figure 8 and Table 8 rank the panel by the loss in area under the curve incurred when a single column is shuffled in the withheld partition. The heaviest is specific gravity, costing 0.781 percentage points, followed by haemoglobin, packed cell volume, albumin and pedal oedema; the leading five account for 95.1 per cent of all measured degradation, and 7 of 24 fields register any degradation at all.

The ordering agrees with renal physiology rather than contradicting it. Diminished concentrating capacity, visible as a falling specific gravity, protein leakage into urine, and the anaemia that follows impaired erythropoietin synthesis together constitute the earliest dependable laboratory evidence of failing filtration, and albuminuria underpins the staging criteria themselves [2], [42]. Nothing in the objective encoded that expectation; it was recovered from the diagnostic label alone. It is worth stating what this attribution does and does not establish: because permutation acts on the fitted predictor, a field can score low either because it carries little information or because a correlated partner already carries it, and the two cases are not distinguished here.

Table 8. Fields ranked by permutation relevance.

Field	AUC loss (points)	Share (%)
Specific gravity	0.781	71.12
Haemoglobin	0.157	14.32
Packed cell volume	0.056	5.10
Albumin	0.029	2.67
Pedal oedema	0.021	1.94
Blood glucose random	0.016	1.46
Serum creatinine	0.011	0.97
Red blood cell count	0.008	0.73
Appetite	0.008	0.73
Red blood cells	0.005	0.49

6.6 Standing against published results

Table 9 positions the proposal beside prior studies on the same panel, each entered at the best figure that study reported. Rady and Anwar obtain 96.90 per cent from a probabilistic network at the intermediate stage [12], Yashfi and co-workers 97.12 per cent from a forest over a reduced field set [13], and Chittora and colleagues 98.46 per cent from a linear kernel machine over the full set [6]. MPC-Net exceeds all three. Two further contrasts do not appear in the accuracy column: none of those studies reports any measure of probability quality, and all of them dispose of unrecorded entries before fitting begins instead of representing them.

Table 9. Accuracy beside previously published results.

Study	Year	Approach	Accuracy (%)
Rady and Anwar [12]	2019	Probabilistic neural network	96.90
Yashfi et al. [13]	2020	Random forest, reduced field set	97.12
Chittora et al. [6]	2021	Linear kernel machine	98.46
Proposed MPC-Net	2026	Pattern-conditioned network	99.80

7. DISCUSSION

Vacancy layout is usable signal. Encoding what was not measured as its own channel, and letting it govern the interpretation of what was, proves compatible with the best figures obtainable on this panel while keeping every incomplete record in play. With 60.8 per cent of cohort A patients missing something, deletion would forfeit most of the evidence, and substitution would erase the boundary between measurement and invention. Conditioning preserves that boundary without paying for a separate token stream.

Confidence, not accuracy, is where methods still differ. On cohort A the leading comparators cluster inside a fraction of a point of one another on accuracy, a margin smaller than the partition-to-partition variation. Their probabilities are not similarly clustered, and the gap in Brier score spans a multiplicative factor. The qualification established in Sections 6.1 and 6.3 stands: this separation is a property of a cohort whose vacancy layout genuinely varies between patients, and it disappears on a cohort where it does not. Since a threshold-based triage policy is only as sound as the probabilities it thresholds, this is the quantity that should be reported and is not.

Attribution should be measured, not assumed. Permutation testing operates on the fitted predictor and produces a ranking whose spread can be inspected. Internal activations offer no such guarantee: a quantity may be perfectly uniform and still be presented as an explanation by taking its maximum. The ranking obtained here is informative because its spread was checked, not because it came from inside the network [43].

Limitations. The evidence base is the principal constraint. Four hundred records drawn from one service over eight weeks cannot stand in for the demographic and analytical variability of a screening population, and the cohort's near-perfect separability means margins of a tenth of a point carry no weight. Results are cross-validated, not obtained on a prospectively gathered external sample, and that distinction should be preserved when reading them, as the TRIPOD recommendations urge [44]. Both cohorts are single-timepoint, so what is detected is prevalent disease; forecasting decline is the harder problem and demands longitudinal records of the kind exploited elsewhere [24]. Diagnoses carry whatever error the originating service's process introduced. Finally, the vacancy layout observed here reflects one site's ordering habits, and whether a pattern encoder fitted on those habits transfers to a service that orders differently can only be settled by prospective multi-site data.

8. CONCLUSION AND FUTURE WORK

We have described MPC-Net, a detector for chronic kidney disease that reads a routine biochemical panel together with the layout of whatever that panel failed to record. The missingness indicator vector is encoded on its own and converted into scale and shift coefficients that condition the value representation feature-wise, so absence changes the interpretation of the measured numbers rather than masquerading as one of them; a factorised second-order term admits pairwise effects cheaply,

and gated residual blocks refine the fused code before 5 independently seeded replicas vote and a logistic recalibrator fixes the probability scale. Under five-fold partitioning repeated five times the model returned 99.80 per cent accuracy on the first cohort and 98.00 per cent on the second, with areas under the curve of 100.00 and 99.96 per cent, surpassing the accuracies published for comparable studies on this panel. On that first cohort it also produced the most trustworthy probabilities of the 10 methods examined, reducing squared probability error by a factor of 3.3 against the closest comparator, at a cost of 550.1 thousand parameters in total and 4.99 ms per patient on one processor core. On the second cohort, whose vacancy layout is effectively constant, the advantage disappears and tree ensembles lead, which locates the benefit of the design precisely where its mechanism applies. Permutation testing recovered an evidence ordering consistent with nephrological practice without being told to. The obvious continuations are prospective validation across several collecting sites, whose differing ordering habits would test the pattern encoder directly, and extension to repeated panels so that decline rather than presence becomes the quantity predicted.

REFERENCES

[1] GBD Chronic Kidney Disease Collaboration, "Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," The Lancet, vol. 395, no. 10225, pp. 709–733, Feb. 2020.

[2] Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group, "KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease," Kidney International Supplements, vol. 3, no. 1, pp. 1–150, Jan. 2013.

[3] L. Rajesh, P. MeganaSanthoshi, L. K. Moorthy, A. F. J. Alshammari, G. T. Varma, and A. A. M., "Artificial intelligence driven healthcare: A survey of machine learning, deep learning and intelligent predictive methods," Natural Resources for Human Health, vol. 6, no. 2s, 2026.

[4] M. Anand, P. Chaya, J. Vishwesh, M. V. Neethi, A. Taranum, and R. P. Kumar, "An Integrated Deep Learning Framework for Diabetic Retinopathy: Channel and Spatial Attention U-Net for Lesion Segmentation and CNN-Based Fundus Image Denoising with Ensemble Feature Classification," International Journal of Drug delivery Technology, vol. 19, no. 7s, pp. 156–165, 2026. doi: 10.25258/ijddt.16.7s.19.

[5] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," Journal of Statistical Software, vol. 45, no. 3, pp. 1–67, Dec. 2011.

[6] P. Chittora et al., "Prediction of chronic kidney disease—a machine learning perspective," IEEE Access, vol. 9, pp. 17312–17334, 2021.

[7] J. Qin, L. Chen, Y. Liu, C. Liu, C. Feng, and B. Chen, "A machine learning methodology for diagnosing chronic kidney disease," IEEE Access, vol. 8, pp. 20991–21002, 2020.

[8] E. M. Senan et al., "Diagnosis of chronic kidney disease using effective classification algorithms and recursive feature elimination techniques," Journal of Healthcare Engineering, vol. 2021, art. 1004767, Jun. 2021.

[9] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in Proc. 34th Int. Conf. Machine Learning, Sydney, Australia, Aug. 2017, pp. 1321–1330.

[10] Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in Proc. 22nd Int. Conf. Machine Learning, Bonn, Germany, Aug. 2005, pp. 625–632.

[11] L. J. Rubini, P. Soundarapandian, and P. Eswaran, "Chronic kidney disease data set," UCI Machine Learning Repository, University of California, Irvine, CA, USA, 2015.

[12] E. H. A. Rady and A. S. Anwar, "Prediction of kidney disease stages using data mining algorithms," Informatics in Medicine Unlocked, vol. 15, art. 100178, 2019.

[13] S. Y. Yashfi et al., "Risk prediction of chronic kidney disease using machine learning algorithms with lifestyle factors," in Proc. 11th Int. Conf. Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, Jul. 2020, pp. 1–5.

[14] M. Almasoud and T. E. Ward, "Detection of chronic kidney disease using machine learning algorithms with least number of predictors," International Journal of Advanced Computer Science and Applications, vol. 10, no. 8, pp. 89–96, 2019.

[15] M. Amaresh and A. M. Sundaram, "An intelligent healthcare framework for early chronic kidney disease screening using swarm-optimized feature selection," Natural Resources for Human Health, vol. 6, no. 1s, pp. 312–332, 2026.

[16] X. Huang, A. Khetan, M. Cvitkovic, and Z. Karnin, "TabTransformer: Tabular data modeling using contextual embeddings," arXiv:2012.06678, Dec. 2020.
[17] S. Ö. Arık and T. Pfister, "TabNet: Attentive interpretable tabular learning," in Proc. AAAI Conf. Artificial Intelligence, vol. 35, no. 8, May 2021, pp. 6679–6687.
[18] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko, "Revisiting deep learning models for tabular data," in Advances in Neural Information Processing Systems 34, Dec. 2021, pp. 18932–18943.
[19] Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.
[20] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. Courville, "FiLM: Visual reasoning with a general conditioning layer," in Proc. AAAI Conf. Artificial Intelligence, vol. 32, no. 1, New Orleans, LA, USA, Feb. 2018, pp. 3942–3951.
[21] S. Rendle, "Factorization machines," in Proc. IEEE Int. Conf. Data Mining, Sydney, Australia, Dec. 2010, pp. 995–1000.
[22] K. R. Radhika, S. Gupta, M. J. Oli, D. Sharma, H. Vinutha, and H. N. Shenoy, "Pancreatic tumor segmentation using conditional diffusion and graph regularization on abdominal computed tomography images," Journal of Intelligent Decision Making and Information Science, vol. 3, no. 1s, pp. 1123–1143, doi: 10.59543/jidmis.v3.489.
[23] G. Chen et al., "Prediction of chronic kidney disease using adaptive hybridized deep convolutional neural network on the Internet of Medical Things platform," IEEE Access, vol. 8, pp. 100497–100508, 2020.
[24] N. Tomasev et al., "A clinically applicable approach to continuous prediction of future acute kidney injury," Nature, vol. 572, no. 7767, pp. 116–119, Aug. 2019.
[25] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in Advances in Large Margin Classifiers, A. J. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, Eds. Cambridge, MA, USA: MIT Press, 1999, pp. 61–74.
[26] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in Advances in Neural Information Processing Systems 30, Dec. 2017, pp. 6402–6413.
[27] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," arXiv:1607.06450, Jul. 2016.
[28] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826.
[29] Loshchilov and F. Hutter, "Decoupled weight decay regularization," in Proc. 7th Int. Conf. Learning Representations, New Orleans, LA, USA, May 2019.
[30] Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in Proc. 5th Int. Conf. Learning Representations, Toulon, France, Apr. 2017.
[31] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[32] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Machine Learning, vol. 63, no. 1, pp. 3–42, Apr. 2006.
[33] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119–139, Aug. 1997.
[34] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785–794.
[35] B. W. Matthews, "Comparison of the predicted and observed secondary structure of T4 phage lysozyme," Biochimica et Biophysica Acta, vol. 405, no. 2, pp. 442–451, Oct. 1975.
[36] Cohen, "A coefficient of agreement for nominal scales," Educational and Psychological Measurement, vol. 20, no. 1, pp. 37–46, Apr. 1960.
[37] G. W. Brier, "Verification of forecasts expressed in terms of probability," Monthly Weather Review, vol. 78, no. 1, pp. 1–3, Jan. 1950.
[38] F. Wilcoxon, "Individual comparisons by ranking methods," Biometrics Bulletin, vol. 1, no. 6, pp. 80–83, Dec. 1945.
[39] Demšar, "Statistical comparisons of classifiers over multiple data sets," Journal of Machine Learning Research, vol. 7, pp. 1–30, Jan. 2006.
[40] Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Advances in Neural Information Processing Systems 32, Dec. 2019, pp. 8024–8035.

[41] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, Oct. 2011.
[42] S. Levey et al., "A new equation to estimate glomerular filtration rate," Annals of Internal Medicine, vol. 150, no. 9, pp. 604–612, May 2009.
[43] Z. C. Lipton, "The mythos of model interpretability," Communications of the ACM, vol. 61, no. 10, pp. 36–43, Sep. 2018.
[44] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement," BMC Medicine, vol. 13, art. 1, Jan. 2015.