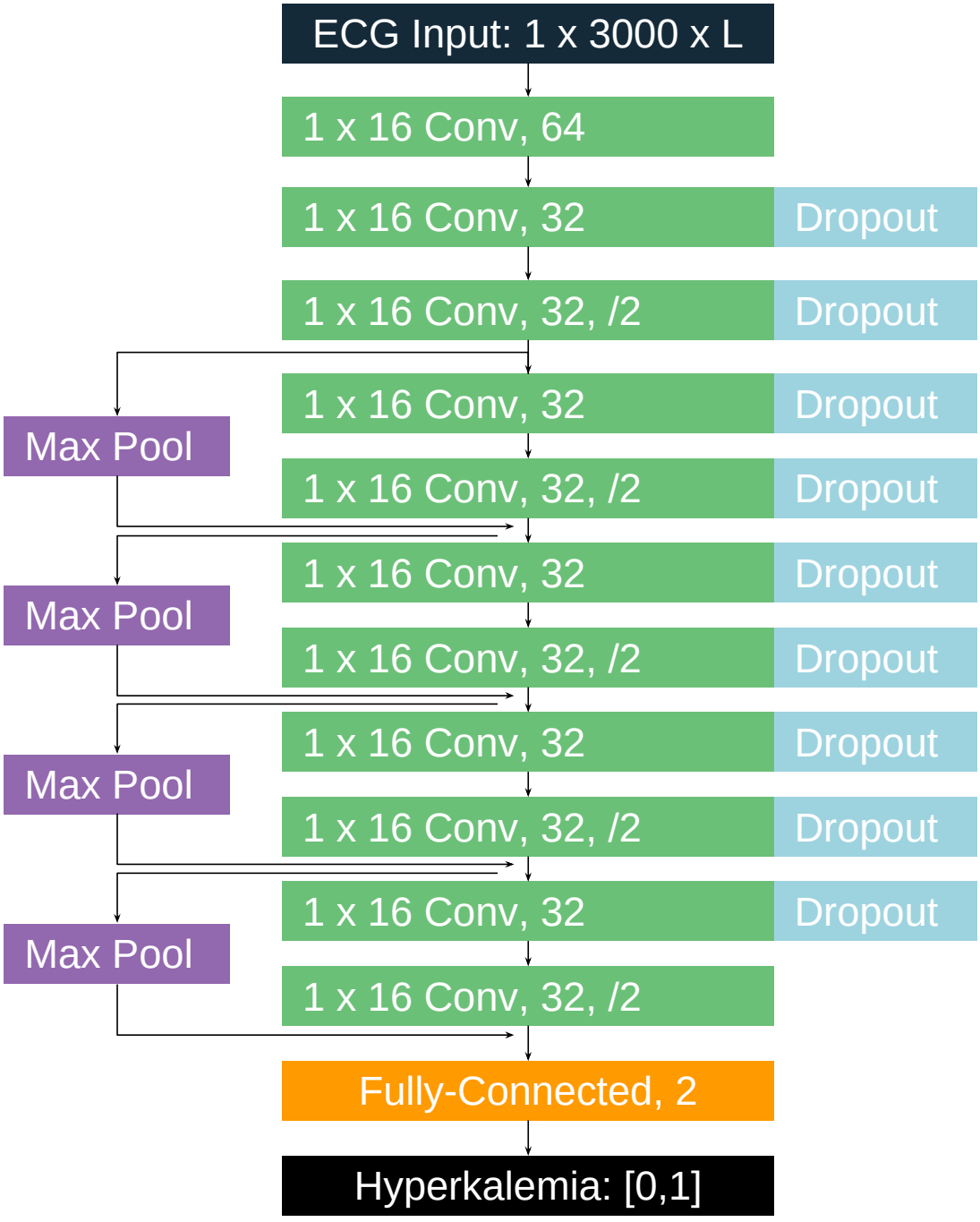
Supplementary Online Content

Galloway CD, Valys AV, Shreibati JB, et al. Development and validation of a deep deep-learning model to screen for hyperkalemia from the electrocardiogram. *JAMA Cardiol*. Published online April 3, 2019. doi:10.1001/jamacardio.2019.0640

eFigure 1. Deep Neural Network Architecture
eMethods. Hyperkalemia Labeling in Development Dataset, Validation Dataset Creation, and Validation Dataset Sample Size Estimation
eFigure 2. Performance for Hyperkalemia from Lead I and II of Electrocardiogram, Deep Learning Model Versus Patient Demographics
eTable 1. Association of Deep Learning Model and Patient Demographics with Hyperkalemia
eTable 2. Sensitivity Analysis: Validation Dataset Performance for Hyperkalemia from the ECG, Excluding Patients with End-Stage Renal Disease
eTable 3. Sensitivity Analysis: Validation Dataset Performance for Hyperkalemia from the ECG, no ECG-Potassium Pairs Excluded
eFigure 3. Deep Neural Network Feature Visualization

This supplementary material has been provided by the authors to give readers additional information about their work.

eFigure 1. Deep Neural Network Architecture



eMethods.

Deep Learning Model/Deep Neural Network

The DLM consisted of a convolutional deep neural network (DNN) with 11 layers, with the first 10 layers being convolutional and the last fully-connected (**eFigure 1)**. A RELU activation function was used. Skip connections, dropout, batch normalization and max pooling were all utilized to improve generalization and convergence properties. All ECGs signals were resampled to 300 Hz and processed through an existing digital filter system designed to remove noise and artifact. The DNN was designed as a binary classifier that receives a 10-second ECG signal from any number of simultaneously-acquired ECG leads and produces as output a number from 0 to 1, representing the probability that hyperkalemia was detected from the ECG. An algorithm was implemented that trains an ensemble of 7 networks independently on the same data and then generates the final output by averaging results of each network on the same input. (1) The ensemble neural network was implemented in Python and TensorFlow using standard tools (2) and trained on the development dataset.

Hyperkalemia Labeling in Development Dataset

There were several challenges posed by the use of a blood test to train a deep learning model to classify hyperkalemia at the time of an ECG recording. First, given the mean absolute time to potassium from ECG recording was 2.2 hours (SD 2.5), the patient's potassium concentration could have changed between the time of the ECG and the time blood was drawn; repeat potassium blood draws over 30-minute periods can yield mean percentage differences of 5% (SD 5%) (3), and potassium concentration may change over time due to diet, fluid status, medications and other treatments. Second, serum potassium results may also be affected by handling or processing errors of the test itself (4). With the low prevalence of hyperkalemia in the development dataset population (2% all ECGs paired with serum potassium $\geq$ 5.5 mmol/L), lab error may account for some of the elevated potassium values in the dataset.

In order to reduce error associated with serum potassium labels, for the 371,681 (23.6%) ECGs linked with more than one serum potassium, a Gaussian process model was used to estimate potassium at the time of ECG recording.

All ECGs in the development set were labeled as "hyperkalemia" or "not hyperkalemia" according to the serum potassium or, if applicable, the Gaussian process model. We defined hyperkalemia as a serum potassium value $\geq$5.5 mmol/L, as this is a commonly used cut-off to prompt treatment for hyperkalemia (5-7). Several studies have shown a rapid increase in the risk of death as serum potassium levels exceed 5.5 mmol/L (8).

Validation Dataset Creation

To validate the DLM, we identified ECG-potassium pairs in CKD patients with from Mayo Clinic's Rochester, Minnesota; Jacksonville, Florida; and Scottdale, Arizona clinics. The validation datasets were created as followed. First, for the Florida and Arizona datasets, any 12-lead ECG recorded from January 1, 2013 through March 31, 2018 was obtained. The Minnesota dataset was comprised of the previously partitioned original dataset (30%). Next, all ECGs within 4 hours before a serum potassium draw were identified. The 4-hour time frame was selected based on prior work estimating potassium from the ECG (9).

We included patients with stage 3 or greater CKD, defined as an estimated glomerular filtration rate (eGFR) < 60 ml/min/1.73 m2. We calculated eGFR by taking the average of serum creatinine values in the 12 hours prior to the ECG and up to the time of the serum potassium draw of interest, then imputing that value, along with race, age, and gender, into the CKD-EPI equation. We did not have race available for the Minnesota dataset, so that adjustment factor was not included in the eGFR calculation.

ECGs with left bundle branch block (LBBB) were excluded because of the concern that the peaking of T waves and the widening of the QRS complex that occurs with hyperkalemia could be masked with a baseline LBBB.

If multiple ECGs were recorded within 4 hours of a potassium draw, the ECG closest in time to potassium was selected. Patients in the Florida or Arizona datasets who were also identified in the Minnesota dataset, were excluded.

If there was more than one ECG-potassium pair per patient, only the most chronologically recent pair was selected (for Arizona and Florida datasets); a random ECG-potassium pair was selected for Minnesota dataset.

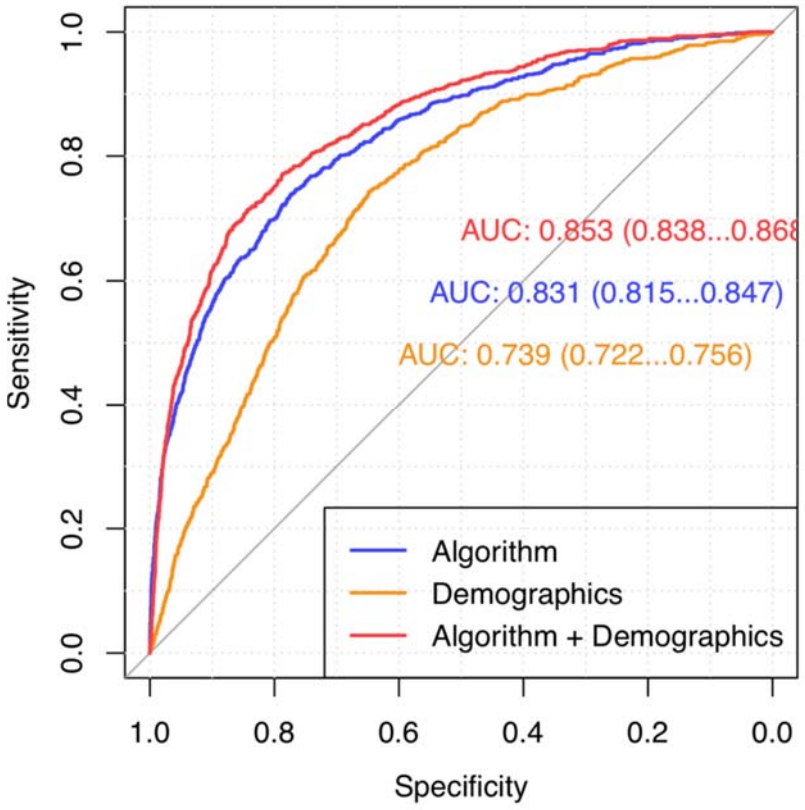
The algorithm was trained to classify serum potassium of $\geq$ 5.5 mmol/L or < 5.5 mmol/L. To avoid potentially irrelevant misclassification rates around this threshold, an analysis of patients with either $\leq$ 5.3 mmol/L or $\geq$ 5.7 mmol/L was the focus of the analysis. The exclusion of potassium levels near the threshold is similar to a

phase II biomarker design (10). The ± 0.2 mmol/L potassium lab draw error rate was estimated based upon our prior work (3).

Validation Dataset Sample Size Estimation

The Florida and Arizona validation datasets' sample size was based on desired precision of the width of the exact binomial confidence interval for sensitivity and specificity to be 10 percentage points. For the area under the curve (AUC), we set the precision to be 6 percentage points. Assuming 90% sensitivity, 85% specificity and an .90 AUC (at 3% prevalence of hyperkalemia), a target sample size of 6,133 cases was determined based on the maximal sample size needs of the AUC. The calculation was considered stratified by site. A total of 12,732 records were identified for retrieval. The final number of records reviewed was slightly lower on account of removal of cases where they were previously seen in Rochester, Minnesota (n=413).

eFigure 2. Performance for Hyperkalemia from Lead I and II of Electrocardiogram, Deep Learning Model[a] Versus Patient Demographics[b]



ROC curves were generated from the combined Arizona and Florida validation datasets.

a. Algorithm refers to deep learning model to predict hyperkalemia.
b. Patient demographic variables to predict hyperkalemia include age, race, gender, body mass index, and estimated glomerular filtration rate.

eTable 1. Association of Deep Learning Model[a] and Patient Demographics with Hyperkalemia[b]

Independent Variables (continuous or categorical)	Logistic Regression Coefficient, Univariate Analysis	P value	Logistic Regression Coefficient, Multivariable Analysis	P value
DLM output (continuous variable)	5.31	< 0.0001	4.78	<0.0001
Age (continuous)			0.0137	<0.0001
White/Not White (categorical)			0.1659	0.15
Male/Not Male (categorical)			0.2552	0.003
BMI (continuous)			-0.0037	0.34
eGFR (continuous variable)			-0.0285	<0.0001

Abbreviations: DLM, deep learning model; BMI, body mass index; eGFR, estimated glomerular filtration rate

a. Model using ECG leads I and II, at 90% sensitivity operating point.
b. Hyperkalemia defined as serum potassium of ≥ 5.5 mmol/L.

eTable 2. Sensitivity Analysis: Validation Dataset Performance for Hyperkalemia from the ECG, Excluding Patients with End-Stage Renal Disease

	2-lead ECG[a]		4-lead ECG[b]	
	Sensitivity= Specificity	High Sensitivity[c]	Sensitivity= Specificity	High Sensitivity
Validation Dataset	**Value (95% CI)**	**Value (95% CI)**	**Value (95% CI)**	**Value (95% CI)**
Minnesota (n=48,119)				
AUC	0.863 (0.849-0.876)		0.883 (0.869-0.895)	
Sensitivity, %	75.8 (72.7-78.6)	87.2 (84.8-89.4)	76.6 (73.6-79.4)	86.2 (83.6-88.4)
Specificity, %	82.0 (81.7-82.4)	63.8 (63.4-64.3)	85.0 (84.7-85.3)	70.8 (70.4-71.2)
NPV, %	99.5 (99.4-99.5)	99.6 (99.6-99.7)	99.5 (99.4-99.6)	99.7 (99.6-99.7)
PPV, %	6.9 (6.4-7.5)	4.1 (3.8-4.4)	8.3 (7.7-8.9)	5.0 (4.6-5.3)
Site 1 (n=5,283)				
AUC	0.864 (0.837-0.890)		0.890 (0.865-0.915)	
Sensitivity, %	79.5 (73.2-84.9)	91.5 (86.7-95.0)	82.0 (76.0-87.1)	91.0 (86.1-94.6)
Specificity, %	76.6 (75.4-77.7)	56.0 (54.6-57.3)	79.1 (77.9-80.2)	62.4 (61.1-63.8)
NPV, %	99.0 (98.6-99.3)	99.4 (99.1-99.7)	99.1 (98.8-99.4)	99.4 (99.1-99.7)
PPV, %	11.8 (10.1-13.6)	7.6 (6.5-8.7)	13.4 (11.5-15.4)	8.7 (7.5-10.0)
Site 2 (n=5,215)				
AUC	0.850 (0.822-0.879)		0.884 (0.859-0.908)	
Sensitivity, %	76.1 (69.2-82.1)	87.8 (82.1-92.2)	82.8 (76.5-88.0)	92.2 (87.3-95.7)
Specificity, %	76.4 (75.2-77.6)	55.8 (54.5-57.2)	78.4 (77.2-79.5)	61.4 (60.0-62.8)
NPV, %	98.9 (98.5-99.2)	99.2 (98.8-99.5)	99.2 (98.9-99.5)	99.5 (99.2-99.8)
PPV, %	10.3 (8.7-12.1)	6.6 (5.7-7.7)	12.0 (10.3-14.0)	7.9 (6.8-9.1)

Abbreviations: AUC, area under the receiver operating characteristic curve; NPV, negative predictive value; PPV, positive predictive value.

End-stage renal disease defined as estimated GFR < 15 ml/min/1.73 m2.

a. 2-lead ECG using leads I and II.
b. 4-lead ECG using leads I, II, V3, and V5.
c. Operating point at sensitivity of 90%, from the ROC from training on the development dataset.

eTable 3. Sensitivity Analysis: Validation Dataset Performance for Hyperkalemia from the ECG, no ECG-Potassium Pairs Excluded

	2-lead ECG[a]		4-lead ECG[b]	
	Sensitivity= Specificity	High Sensitivity[c]	Sensitivity= Specificity	High Sensitivity
Validation Dataset	**Value (95% CI)**	**Value (95% CI)**	**Value (95% CI)**	**Value (95% CI)**
Minnesota (n=51,247)				
AUC	0.837 (0.828-0.846)		0.858 (0.848-0.867)	
Sensitivity, %	70.7 (68.7-72.6)	85.0 (83.4-86.5)	72.8 (70.9-74.7)	83.6 (82.0-85.2)
Specificity, %	80.9 (80.5-81.2)	62.7 (62.3-63.2)	83.7 (83.4-84.0)	69.5 (69.1-69.9)
NPV, %	98.5 (98.3-98.6)	99.0 (98.9-99.1)	98.6 (98.5-98.7)	99.0 (98.9-99.1)
PPV, %	13.8 (13.2-14.5)	9.0 (8.6-9.4)	16.3 (15.5-17.0)	10.6 (10.2-11.1)
Florida (n=6,242)				
AUC	0.840 (0.820-0.860)		0.860 (0.840-0.880)	
Sensitivity, %	77.5 (73.2-81.4)	90.3 (87.1-92.9)	80.1 (76.0-83.8)	88.9 (85.5-91.7)
Specificity, %	74.5 (73.4-75.7)	54.1 (52.8-55.4)	76.3 (75.2-77.4)	59.8 (58.5-61.1)
NPV, %	97.9 (97.4-98.3)	98.7 (98.3-99.1)	98.1 (97.7-98.5)	98.7 (98.2-99.0)
PPV, %	18.1 (16.3-19.9)	12.5 (11.3-13.7)	19.7 (17.8-21.6)	13.8 (12.5-15.2)
Arizona (n=6,077)				
AUC	0.816 (0.794-0.839)		0.840 (0.819-0.861)	
Sensitivity, %	73.2 (68.5-77.5)	86.0 (82.1-89.3)	77.3 (72.8-81.4)	87.2 (83.5-90.4)
Specificity, %	74.6 (73.5-75.8)	54.4 (53.1-55.7)	76.3 (75.4-77.6)	59.7 (58.4-60.9)
NPV, %	97.6 (97.1-98.0)	98.3 (97.7-98.7)	98.0 (97.5-98.4)	98.5 (98.1-98.9)
PPV, %	16.6 (14.9-18.4)	11.5 (10.4-12.7)	18.4 (16.6-20.3)	13.0 (11.7-14.3)

Abbreviations: AUC, area under the receiver operating characteristic curve; NPV, negative predictive value; PPV, positive predictive value.

Given inherent error in potassium values and concern for misclassification, original analysis excluded ECG-potassium data if serum potassium was > 5.3 mmol/L or < 5.7 mmol/L.

a. 2-lead ECG using leads I and II.

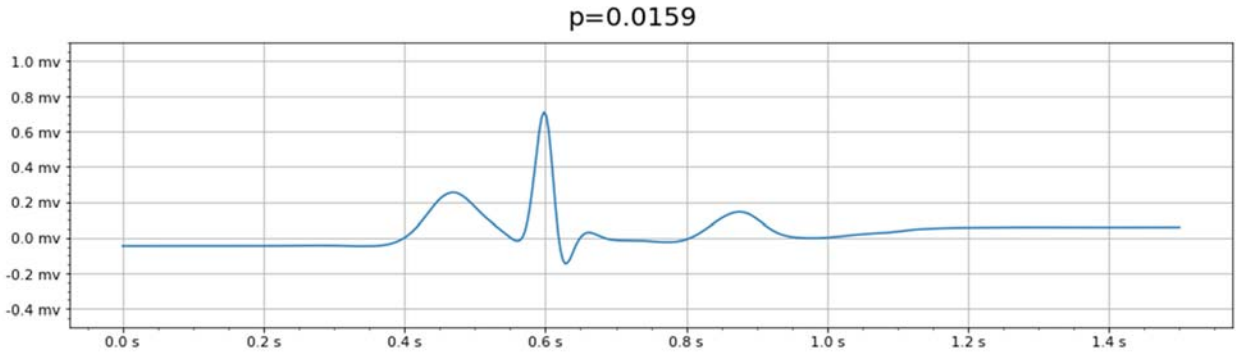
b. 4-lead ECG using leads I, II, V3, and V5.

c. Operating point at sensitivity of 90%, from the ROC from training on the development dataset.

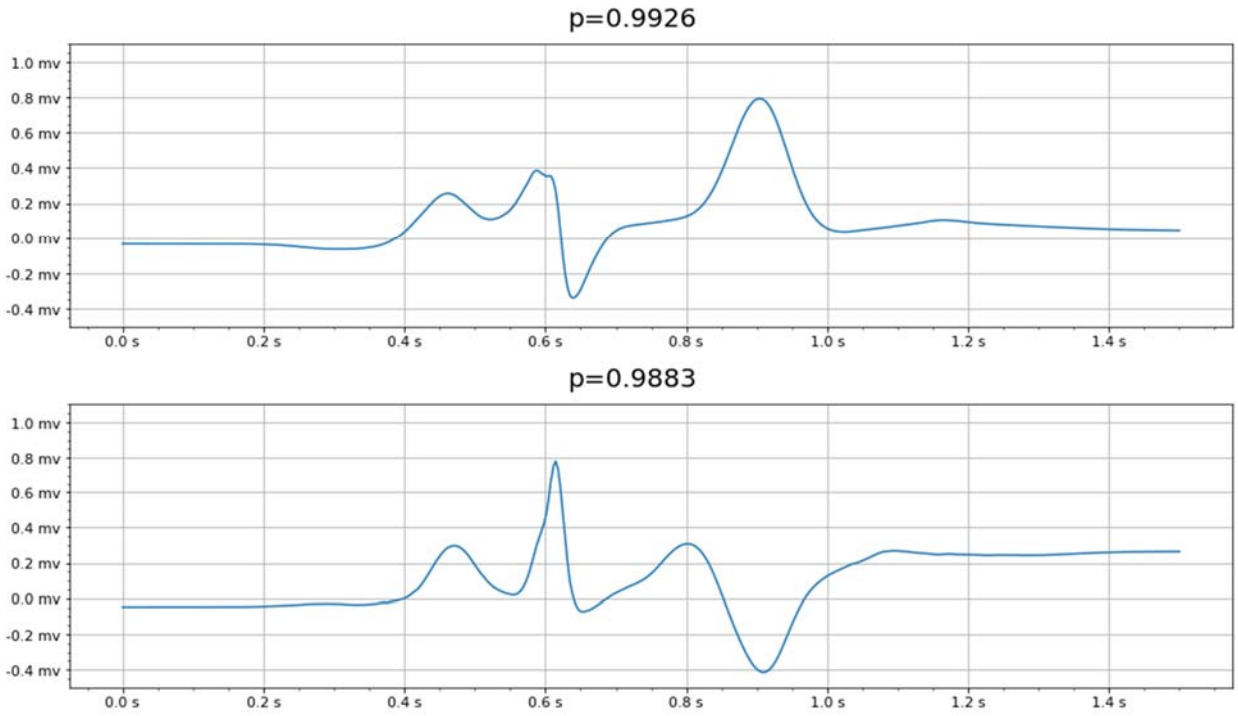
eFigure 3. Deep Neural Network Feature Visualization

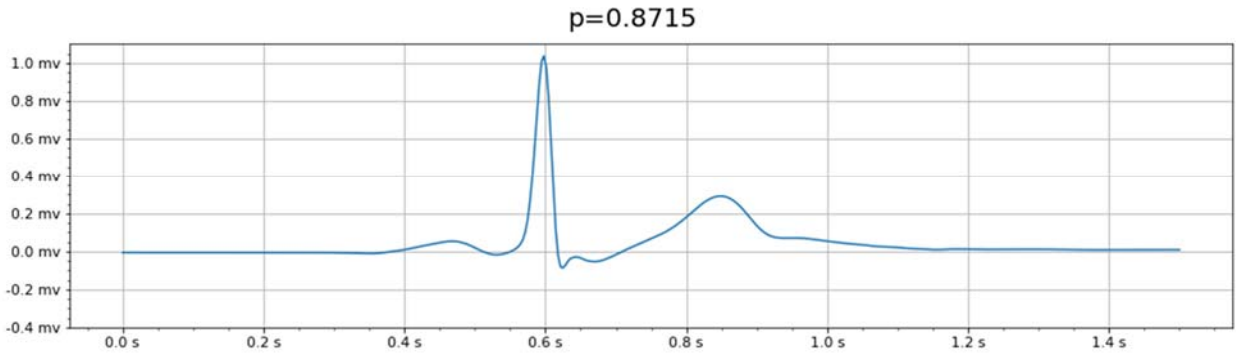
To try to understand how the deep neural network (DNN) builds up its understanding of images for hyperkalemia over many layers, we performed feature visualization on the average ECG beat.(11, 12) Since the DNN is differentiable end-to-end, we can ask "what changes in the input X would cause the probability p to increase"? By repeatedly modifying the input X to increase the probability p, we can gradually modify an input ECG in a way that causes the network to output progressively higher values of p. This enables us to visualize the changes in ECG morphology which result in a hyperkalemia prediction.

Here is average ECG beat with low probability (p) of hyperkalemia:



There are multiple morphologies the neural network detects that indicate high probability of hyperkalemia:





In the first two examples, we see expected changes in QRS width, and T wave amplitude and morphology. However, the third average ECG beat does not appear to have the morphology changes humans expect to see.

This last example underscores the value of deep learning and the limits of human visualization in the detection of hyperkalemia from the ECG: there are ECG features that the network identifies as important, that we cannot easily comprehend.

References

1. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Processing Syst. 2012:1097-105.
2. Abadi M., Agarwal A, Barham P., Brevdo E., Chen Z., Citro C., et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. 2016 [Available from: https://arxiv.org/abs/1603.04467.
3. Friedman PA, Scott CG, Bailey K, Baumann NA, Albert D, Attia ZI, et al. Errors of Classification With Potassium Blood Testing: The Variability and Repeatability of Critical Clinical Tests. Mayo Clin Proc. 2018;93(5):566-72.
4. Asirvatham JR, Moses V, Bjornson L. Errors in potassium measurement: a laboratory perspective for the clinician. N Am J Med Sci. 2013;5(4):255-9.
5. Tran HA. Extreme hyperkalemia. South Med J. 2005;98(7):729-32.
6. Rossignol P, Dobre D, McMurray JJ, Swedberg K, Krum H, van Veldhuisen DJ, et al. Incidence, determinants, and prognostic significance of hyperkalemia and worsening renal function in patients with heart failure receiving the mineralocorticoid receptor antagonist eplerenone or placebo in addition to optimal medical therapy: results from the Eplerenone in Mild Patients Hospitalization and Survival Study in Heart Failure (EMPHASIS-HF). Circ Heart Fail. 2014;7(1):51-8.
7. Yusuf AA, Hu Y, Singh B, Menoyo JA, Wetmore JB. Serum Potassium Levels and Mortality in Hemodialysis Patients: A Retrospective Cohort Study. Am J Nephrol. 2016;44(3):179-86.
8. Grodzinsky A, Goyal A, Gosch K, McCullough PA, Fonarow GC, Mebazaa A, et al. Prevalence and Prognosis of Hyperkalemia in Patients with Acute Myocardial Infarction. Am J Med. 2016;129(8):858-65.
9. Velagapudi V, O'Horo JC, Vellanki A, Baker SP, Pidikiti R, Stoff JS, et al. Computer-assisted image processing 12 lead ECG model to diagnose hyperkalemia. J Electrocardiol. 2017;50(1):131-8.
10. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst. 2001;93(14):1054-61.
11. Erhan D, Courville A, Bengio Y. Visualizing higher-layer features of a deep network. Technical report 1341. . University of Montreal; 2009.
12. Olah C, Mordvintsev A, Schubert L. Feature visualization. Distill. 2017 [Available from: https://distill.pub/2017/feature-visualization/.