CNN-based Classification of I-123 ioflupane dopamine transporter SPECT brain images to support the diagnosis of Parkinson's disease with Decision Confidence Estimation

Master Thesis

Master of Science in Applied Computer Science

Aleksej Kucerenko

September 28, 2023

Supervisor:

1st: Prof. Dr. Christian Ledig
2nd: Dr. Ralph Buchert, Universitätsklinikum Hamburg-Eppendorf

Chair of Explainable Machine Learning
Faculty of Information Systems and Applied Computer Sciences
Otto-Friedrich-University Bamberg

Abstract

Short summary of your thesis (max. 1 page) . . .

Abstract

Kurze Zusammenfassung Ihrer Abschlussarbeit (max. 1 Seite) ...

Acknowledgements

If you want to thank anyone (optional) ...

Contents

List of Figures v

List of Tables vi

List of Acronyms vii

1 Introduction 1

2 Background 2

2.1 SPECT imaging and SBR 2

2.2 Random Forest 2

2.3 CNN 2

3 Methods 2

3.1 SBR 2

3.2 CNN - Majority 2

3.3 CNN - Random 2

4 Data Sources and Preprocessing 2

4.1 SPECT dataset 2

4.2 External datasets 2

5 Evaluation 2

6 Discussion 2

7 Conclusion 2

A Appendix 3

Bibliography 4

List of Figures

List of Tables

List of Acronyms

AI Artificial Intelligence

Notation

This section provides a concise reference describing notation as used in the book by Goodfellow et al. (2016). If you are unfamiliar with any of the corresponding mathematical concepts, Goodfellow et al. (2016) describe most of these ideas in chapters 2–4.

Numbers and Arrays

a	A scalar (integer or real)
$\boldsymbol{a}$	A vector
$\boldsymbol{A}$	A matrix
$\mathsf{\textbf{A}}$	A tensor
$\boldsymbol{I}_n$	Identity matrix with n rows and n columns
$\boldsymbol{I}$	Identity matrix with dimensionality implied by context
$\boldsymbol{e}^{(i)}$	Standard basis vector $[0, \dots, 0, 1, 0, \dots, 0]$ with a 1 at position i
$\text{diag}(\boldsymbol{a})$	A square, diagonal matrix with diagonal entries given by $\boldsymbol{a}$
a	A scalar random variable
$\mathbf{a}$	A vector-valued random variable
$\mathbf{A}$	A matrix-valued random variable

Sets and Graphs

$\mathbb{A}$	A set
$\mathbb{R}$	The set of real numbers
$\{0, 1\}$	The set containing 0 and 1
$\{0, 1, \dots, n\}$	The set of all integers between 0 and n
$[a, b]$	The real interval including a and b
$(a, b]$	The real interval excluding a but including b
$\mathbb{A} \backslash \mathbb{B}$	Set subtraction, i.e., the set containing the elements of $\mathbb{A}$ that are not in $\mathbb{B}$
$\mathcal{G}$	A graph
$Pa_{\mathcal{G}}(\mathrm{x}_i)$	The parents of x_i in $\mathcal{G}$

Indexing

a_i	Element i of vector $\boldsymbol{a}$, with indexing starting at 1
a_{-i}	All elements of vector $\boldsymbol{a}$ except for element i
$A_{i,j}$	Element i, j of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{i,:}$	Row i of matrix $\boldsymbol{A}$
$\boldsymbol{A}_{:,i}$	Column i of matrix $\boldsymbol{A}$
$\mathsf{A}_{i,j,k}$	Element (i, j, k) of a 3-D tensor $\mathbf{\mathsf{A}}$
$\mathbf{\mathsf{A}}_{:,:,i}$	2-D slice of a 3-D tensor
a_i	Element i of the random vector $\mathbf{a}$

Linear Algebra Operations

$\boldsymbol{A}^\top$	Transpose of matrix $\boldsymbol{A}$
$\boldsymbol{A}^+$	Moore-Penrose pseudoinverse of $\boldsymbol{A}$
$\boldsymbol{A} \odot \boldsymbol{B}$	Element-wise (Hadamard) product of $\boldsymbol{A}$ and $\boldsymbol{B}$
$\det(\boldsymbol{A})$	Determinant of $\boldsymbol{A}$

Calculus

$\dfrac{dy}{dx}$	Derivative of y with respect to x
$\dfrac{\partial y}{\partial x}$	Partial derivative of y with respect to x
$\nabla_{\boldsymbol{x}} y$	Gradient of y with respect to $\boldsymbol{x}$
$\nabla_{\boldsymbol{X}} y$	Matrix derivatives of y with respect to $\boldsymbol{X}$
$\nabla_{\mathbf{\mathsf{X}}} y$	Tensor containing derivatives of y with respect to $\mathbf{\mathsf{X}}$
$\dfrac{\partial f}{\partial \boldsymbol{x}}$	Jacobian matrix $\boldsymbol{J} \in \mathbb{R}^{m \times n}$ of $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$
$\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})$ or $\boldsymbol{H}(f)(\boldsymbol{x})$	The Hessian matrix of f at input point $\boldsymbol{x}$
$\displaystyle\int f(\boldsymbol{x}) d\boldsymbol{x}$	Definite integral over the entire domain of $\boldsymbol{x}$
$\displaystyle\int_{\mathbb{S}} f(\boldsymbol{x}) d\boldsymbol{x}$	Definite integral with respect to $\boldsymbol{x}$ over the set $\mathbb{S}$

Probability and Information Theory

$\mathrm{a}\perp\mathrm{b}$	The random variables a and b are independent
$\mathrm{a}\perp\mathrm{b} \mid \mathrm{c}$	They are conditionally independent given c
$P(\mathrm{a})$	A probability distribution over a discrete variable
$p(\mathrm{a})$	A probability distribution over a continuous variable, or over a variable whose type has not been specified
$\mathrm{a} \sim P$	Random variable a has distribution P
$\mathbb{E}_{\mathrm{x}\sim P}[f(x)]$ or $\mathbb{E}f(x)$	Expectation of $f(x)$ with respect to $P(\mathrm{x})$
$\mathrm{Var}(f(x))$	Variance of $f(x)$ under $P(\mathrm{x})$
$\mathrm{Cov}(f(x), g(x))$	Covariance of $f(x)$ and $g(x)$ under $P(\mathrm{x})$
$H(\mathrm{x})$	Shannon entropy of the random variable x
$D_{\mathrm{KL}}(P\Vert Q)$	Kullback-Leibler divergence of P and Q
$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$	Gaussian distribution over $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$

Functions

$f : \mathbb{A} \rightarrow \mathbb{B}$	The function f with domain $\mathbb{A}$ and range $\mathbb{B}$
$f \circ g$	Composition of the functions f and g
$f(\boldsymbol{x}; \boldsymbol{\theta})$	A function of $\boldsymbol{x}$ parametrized by $\boldsymbol{\theta}$. (Sometimes we write $f(\boldsymbol{x})$ and omit the argument $\boldsymbol{\theta}$ to lighten notation)
$\log x$	Natural logarithm of x
$\sigma(x)$	Logistic sigmoid, $\dfrac{1}{1+\exp(-x)}$
$\zeta(x)$	Softplus, $\log(1+\exp(x))$
$\lvert\lvert\boldsymbol{x}\rvert\rvert_p$	L^p norm of $\boldsymbol{x}$
$\lvert\lvert\boldsymbol{x}\rvert\rvert$	L^2 norm of $\boldsymbol{x}$
x^+	Positive part of x, i.e., $\max(0, x)$
$\mathbf{1}_{\mathrm{condition}}$	is 1 if the condition is true, 0 otherwise

Sometimes we use a function f whose argument is a scalar but apply it to a vector, matrix, or tensor: $f(\boldsymbol{x})$, $f(\boldsymbol{X})$, or $f(\mathsf{X})$. This denotes the application of f to the array element-wise. For example, if $\mathsf{C} = \sigma(\mathsf{X})$, then $C_{i,j,k} = \sigma(X_{i,j,k})$ for all valid values of i, j and k.

Datasets and Distributions

p_{data}	The data generating distribution
$\hat{p}_{\text{data}}$	The empirical distribution defined by the training set
$\mathbb{X}$	A set of training examples
$\boldsymbol{x}^{(i)}$	The i-th example (input) from a dataset
$y^{(i)}$ or $\boldsymbol{y}^{(i)}$	The target associated with $\boldsymbol{x}^{(i)}$ for supervised learning
$\boldsymbol{X}$	The $m \times n$ matrix with input example $\boldsymbol{x}^{(i)}$ in row $\boldsymbol{X}_{i,:}$

1 Introduction

There is a variety of quality metrics which can be computed to evaluate the performance of a binary classifier given a certain decision boundary or inconclusive range. In practice, the area under curve (AUC) for a receiver operating characteristic (ROC) curve is a commonly used metric for assessing the overall performance of a binary classifier. Alternatively, the AUC value for the precision-recall curve can be computed as a quality metric.

For a binary classification problem in the medical domain, it is of particular interest to train both the most confident and accurate classifier. A classifier with high decision confidence applied in clinical practice would require less manual inspection by the physician which reduces both effort and costs. Therefore one practically important optimization problem is to minimize the number of cases predicted within the inconclusive range by the classifier while maximizing its performance.

This thesis contributes a model-agnostic and robust evaluation metric for the diagnostic decision confidence of a classifier as well as an implementation for producing these evaluation results. We define the decision confidence to be maximized where the performance of the classifier on consensus test cases is maximized over a broad scale of determined inconclusive ranges. Therefore, as the benchmark method, the specific binding ratio (SBR) of 123I-FP-CIT in the putamen was employed and compared to conventional classification methods and convolutional neural network (CNN) approaches. The different models are trained and tested on subsets of a DAT-SPECT dataset consisting of 1740 slices of volumetric DAT-SPECT images. Data augmentation is applied to the DAT-SPECT dataset to increase the heterogeneity of the training and testing data. Additionally, the methods are evaluated on the Parkinson's Progression Markers Initiative (PPMI) dataset ZITAT and the multiple-pinhole (MPH) dataset ZITAT. This study hypothesizes that the state-of-the-art CNN approaches outperform both the benchmark SBR method and conventional classification methods when using the evaluation metric proposed in this work. - more stable and robust; more generalizable to external datasets Randomness "eintrainierbar" bei CNN random

The following research questions are addressed:

- Can a human interpretable and model-agnostic evaluation metric be constructed that allows to evaluate and compare the decision confidence of binary classification models?

  how to automatically define inconclusive ranges given requirements w.r.t. accuracy/performance?

  How to train a method which replicates the uncertainty w.r.t. label of the data

- How do different binary classification models trained on DAT-SPECT imaging data perform on diverse testing data regarding this metric?

2 Background

2.1 SPECT imaging and SBR

2.2 Random Forest

2.3 CNN

3 Methods

3.1 SBR

3.2 CNN - Majority

3.3 CNN - Random

4 Data Sources and Preprocessing

4.1 SPECT dataset

4.2 External datasets

5 Evaluation

Some more of your text. For citations, use the command `\citep{lecun2015deep}` which produces (LeCun et al., 2015) or `\cite{lecun2015deep}` which produces LeCun et al. (2015).

6 Discussion

7 Conclusion

A Appendix

If needed for supplementary material, such as detailed description of data collection, tables, or figures.

Bibliography

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521 (7553):436–444, 2015.

Declaration of Authorship

Ich erkläre hiermit gemäß §9 Abs. 12 APO, dass ich die vorstehende Abschlussarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Des Weiteren erkläre ich, dass die digitale Fassung der gedruckten Ausfertigung der Abschlussarbeit ausnahmslos in Inhalt und Wortlaut entspricht und zur Kenntnis genommen wurde, dass diese digitale Fassung einer durch Software unterstützten, anonymisierten Prüfung auf Plagiate unterzogen werden kann.

Place, Date	Signature