SUPPORTING INFORMATION

Computational Neural Networks. Most computational neural networks (CNNs) perform a similar task: an input-to-output vector mapping. The primary difference is only in the way this task is performed. Typically, only one or a small number of paradigms are needed to treat most practical problems. Multilayer feedforward CNNs trained with the backpropagation algorithm are by far the most commonly applied. This is primarily because these networks can approximate any Borel-measurable function (i.e., they are universal function approximators), have a proven history, and are easy to implement. Below we discuss how to maximize the probability of success when using backpropagation or basis-function CNNs to solve practical problems.

Supervised neural networks require training and testing data sets. Unbiased selection of these sets from the available data is best achieved by using statistical resampling methods such as the bootstrap and jackknife (leave-k-out) methods. The bootstrap method is a powerful procedure for determining the best estimator for small data sets. Bootstrap resampling is basically random sampling with replacement: a data set of N examples is randomly sampled N times to create a new data set with N examples. The new data set will have the possibility of sample repetition and a test set can be generated by comparing the new data set with the original one and selecting unique examples. Thus the bootstrap method can be used to produce a number of different training and test sets and the error estimate is taken as the average performance on the data set ensembles. In general, about 200 iterations of bootstrap estimates are needed to obtain a good representation.

Bootstrap resampling is only one of several methods that provide desirable properties in error rate estimates. Another method that works very well for neural network training and testing is the jackknife method. The general procedure is to take one (it also can be generalized to k) sample out of the available data as a test set and train on the remaining ones. This is repeated for all of the samples, producing an ensemble that is the same size as the original data set. This procedure can be used in the same fashion as described for the bootstrap but requires fewer runs.

By using statistical resampling (bootstrap or jackknife coupled with initial state sampling), a CNN can be trained on an ensemble of different training sets and the performance (cross-validation) evaluated on the compliment testing sets to determine a measure of the true performance (i.e., the average over the ensemble). This approach, called the ensemble-average

method, not only gives a better estimate of the true error (unbiased) but also reduces the size of the neural network (number of hidden nodes) over that required for any single training/testing computation and tends to smooth out (by averaging in function space instead of parameter space) the effects of over-fitting (variance reduction). An optimal CNN model for a given set of data can be obtained by performing ensemble averages over initial connection weights and training/test sets. This procedure is more computationally demanding than one consisting of a single run but provides an optimal and unbiased representation of the performance capabilities of a CNN.

Additional Figures and Tables

```
Benzene
     6 6
     -0.6760    1.2099    0.0215 C
      0.7097    1.1906   -0.0095 C
      1.3855   -0.0192   -0.0315 C
      0.6758   -1.2098   -0.0226 C
     -0.7097   -1.1901    0.0087 C
     -1.3855    0.0196    0.0308 C
     1  2  2
     1  6  1
     2  3  1
     3  4  2
     4  5  1
     5  6  2
```

Figure 1: A connection table for benzene. The first line shows that there are 6 atoms and 6 bonds in this molecule. The following 6 lines show the positions (in Cartesian coordinates) and type of each atom. The last 6 lines show how the atoms are bonded (i.e., atom 1 is bonded to atom 2 via a double (2) bond).

TABLE I: Chemicals Associated with Liver Injury

Chemicals with Acute Liver Injury as the Primary Toxic Effect

Category	Chemical Name	CAS #
Aliphatic Nitros	2-Nitropropane	79-46-9
Aromatic Amines	4,4'-Methylenedianiline	101-77-9
Aromatic Nitros	2,4,6-Trinitrotoluene	118-96-7
Chlorinated Solvents	Ethylene dichloride	107-06-2
Chlorinated Solvents	1,1,2,2-Tetrachloroethane	79-34-5
Chlorinated Solvents	Carbon tetrachloride	56-23-5
Chlorinated Solvents	Propylene dichloride	78-87-5
Halogenated Solvents	Carbon tetrabromide	558-13-4
Halogenated Solvents	Acetylene tetrabromide	79-27-6
Halogenated Solvents	Ethylene dibromide	106-93-4
Halowaxes	Hexachloronaphthalene	1335-87-1
Halowaxes	Octachloronaphthalene	2234-13-1
Halowaxes	Pentachloronaphthalene	1321-64-8
Halowaxes	Tetrachloronaphthalene	1335-88-2
Halowaxes	Trichloronaphthalene	1321-65-9
Nitrosamines	N-Nitrosodimethylamine	62-75-9
Other Solvents	Dimethylformamide	68-12-2
Other Solvents	Tetrahydrofuran	109-99-9
Other Solvents	Dimethyl acetamide	127-19-5
Other Organics	Diphenyl	92-52-4

Chemicals with Liver Injury as a Secondary Toxic Effect: Elevated Liver Enzyme Levels

Category	Chemical Name	CAS #
Alcohols	Ethyl alcohol	64-17-5
Alcohols	Ethylene chlorohydrin	107-07-3
Alcohols	Methylcyclohexanol	25639-42-3
Alcohols	*n*-Propyl alcohol	71-23-8
Alcohols	Propargyl alcohol	107-19-7
Aromatic Solvents	Cumene	98-82-8
Aromatic Solvents	Ethyl benzene	100-41-4
Aromatic Solvents	*p*-tert-Butyltoluene	98-51-1
Aromatic Solvents	Toluene	108-88-3
Aromatic Solvents	Xylene isomers	1330-20-7
Chlorinated Solvents	1,1,1,2-Tetrachloroethane	630-20-6
Chlorinated Solvents	1,1,1-Trichloroethane	71-55-6
Chlorinated Solvents	1,1,2-Trichloroethane	79-00-5
Chlorinated Solvents	1,1-Dichloroethane	75-34-3
Chlorinated Solvents	1,2,3-Trichloropropane	96-18-4
Chlorinated Solvents	1,2-Dichloroethylene, all isomers	540-59-0
Chlorinated Solvents	1,3-Dichloropropene	542-75-6
Chlorinated Solvents	Chloroform	67-66-3

Table I Continued

Category	Chemical Name	CAS #
Chlorinated Solvents	Ethyl chloride	75-00-3
Chlorinated Solvents	Methyl chloride	74-87-3
Chlorinated Solvents	Methylene chloride	75-09-2
Chlorinated Solvents	Tetrachloroethylene	127-18-4
Chlorinated Solvents	Trichloroethylene	79-01-6
Chlorofluorocarbons	1,1,1,2-Tetrachloro-2,2-difluoroethane	76-11-9
Chlorofluorocarbons	1,1,2,2-Tetrachloro-1,2-difluoroethane	76-12-0
Chlorofluorocarbons	CHLOROFLUOROCARBONS	
Chlorofluorocarbons	Chloropentafluoroethane	76-15-3
Chlorofluorocarbons	Dichloromonofluoromethane	75-43-4
Chlorofluorocarbons	Dichlorotetrafluoroethane	76-14-2
Chlorofluorocarbons	Difluorodibromomethane	75-61-6
Chlorofluorocarbons	Trifluorobromomethane	75-63-8
Cyclic Hydrocarbons	Cyclohexane	110-82-7
Cyclic Hydrocarbons	Cyclopentadiene	542-92-7
Cyclic Hydrocarbons	Methylcyclohexane	108-87-2
Esters	Vinyl acetate	108-05-4
Ethers	Dichloroethyl ether	111-44-4
Ethers	Dioxane	123-91-1
Ethers	Ethyl ether	60-29-7
Glycol Ethers	2-Butoxyethanol	111-76-2
Glycol Ethers	2-Ethoxyethanol	110-80-5
Glycol Ethers	2-Methoxyethanol	109-86-4
Glycol Ethers	Diethylene glycol monobutyl ether	112-34-5
Glycol Ethers	GLYCOL ETHERS	
Halogenated Aromatic	1,2,4-Trichlorobenzene	120-82-1
Halogenated Aromatic	Chlorobenzene	108-90-7
Halogenated Aromatic	*o*-Dichlorobenzene	95-50-1
Halogenated Solvents	Bromoform	75-25-2
Halogenated Solvents	Ethyl bromide	74-96-4
Other Solvents	Carbon disulfide	75-15-0
Other Solvents	Methylal	109-87-5
Refined Petroleum	Kerosene	8008-20-6
Refined Petroleum	Petroleum distillates	8002-05-9
Refined Petroleum	Stoddard solvent	8052-41-3
Refined Petroleum	VM & P Naphtha	8032-32-4
Amines, Aliphatic	Ethylamine	75-04-7
Amines, Aromatic	4-Aminodiphenylamine	101-54-2
Amines, Aromatic	4-Dimethylaminoazobenzene	60-11-7
Amines, Aromatic	Benzidine	92-87-5
Amines, Aromatic	Dapsone	80-08-0
Amines, Aromatic	Dichloroaniline (mixed isomers)	27134-27-6
Amines, Aromatic	Dichlorobenzidine base	91-94-1
Amines, Aromatic	Diphenylamine	122-39-4
Amines, Aromatic	Monomethyl aniline	100-61-8
Amines, Aromatic	*o*-Dianisidine-based dyes	119-90-4
Amines, Aromatic	*p*-Phenylene diamine	106-50-3
Amines, Aromatic	Xylidene, mixed isomers	1300-73-8
Amines, Heterocyclic	Morpholine	110-91-8
Ethanolamines	Ethanolamine	141-43-5

Table I Continued

Category	Chemical Name	CAS #
Hydrazines	1,2-Diphenylhydrazine	122-66-7
Hydrazines	Hydrazine	302-01-2
Hydrazines	Phenylhydrazine	100-63-0
Nitriles	Acetonitrile	75-05-8
Nitriles	Isobutyronitrile	78-82-0
Nitriles	Malononitrile	109-77-3
Nitros, Aliphatic	1,1-Dichloro-1-nitroethane	594-72-9
Nitros, Aliphatic	1-Nitropropane	108-03-2
Nitros, Aliphatic	Nitroethane	79-24-3
Nitros, Aliphatic	Nitromethane	75-52-5
Nitros, Aromatic	Dinitolmide	148-01-6
Nitros, Aromatic	Dinitrotoluene, all isomers	25321-14-6
Nitros, Aromatic	*m*-Dinitrobenzene	99-65-0
Nitros, Aromatic	*m*-Nitroaniline	99-09-2
Nitros, Aromatic	Nitrobenzene	98-95-3
Nitros, Aromatic	*o*-Dinitrobenzene	528-29-0
Nitros, Aromatic	*p*-Dinitrobenzene	100-25-4
Nitros, Aromatic	Picric acid	88-89-1
Nitros, Aromatic	*p*-Nitroaniline	100-01-6
Nitros, Aromatic	*p*-Nitrochlorobenzene	100-00-5
Nitros, Aromatic	Tetryl	479-45-8
Nitrosamines	NITROSAMINES	
Nitrosamines	N-Nitrosodi-n-propylamine	621-64-7
Other Nitrogen Compounds	Cyclonite	121-82-4
Pyridines	Pyridine	110-86-1

Table I shows chemicals that are associated with liver injury. The neural network model was able to classify each chemical as primary or secondary with 100 % accuracy.

Table II: Drugs and their Mechanism of Action[33]

Mechanism of Action*	Drug	Mechanism of Action*	Drug
A2	Mitomycin	Db	Cyanomorpholinodoxorubicin
A2	Porfiromycin	Db	Hycanthone
A6	Carmustine (BCNU)	Db	Morpholino-adriamycin
A6	Chlorozotocin	Db	N-N-Dibenzyl-daunomycin
A6	Clomesone	Db	Pyrazoloacridine
A6	Lomustine (CCNU)	Di	5-6-Dihydro-5-azacytidine
A6	Mitozolamide	Di	alpha-2'-Deoxythioguanosine
A6	PCNU	Di	Azacytidine
A6	Semustine (MeCCNU)	Di	beta-2'-Deoxythioguanosine
A7	Asaley	Di	Thioguanine
A7	Busulfan	Df	Aminopterin
A7	Carboplatin	Df	Aminopterin-derivative
A7	Chlorambucil	Df	Aminopterin-derivative
A7	Cisplatin	Df	an-antifol
A7	Cyclodisone	Df	an-antifol
A7	Diaminocyclohexyl-Pt-II	Df	Baker's-soluble-antifolate
A7	Dianhydrogalactitol	Df	Methotrexate
A7	Diaziridinylbenzoquinone	Df	Methotrexate-derivative
A7	Fluorodopan	Df	Trimetrexate
A7	Hepsulfam	Dr	Guanazole
A7	Iproplatin	Dr	Hydroxyurea
A7	Mechlorethamine	Dr	Pyrazoloimidazole
A7	Melphalan	Ds	Aphidicolin-glycinate
A7	Piperazine mustard	Ds	Cyclocytidine
A7	Piperazinedione	Ds	Cytarabine (araC)
A7	Pipobroman	Ds	Floxuridine (FUdR)
A7	Spiromustine	Ds	Fluorouracil (5FU)
A7	Teroxirone	Ds	Ftorafur
A7	Tetraplatin	Ds	Thiopurine (6MP)
A7	Thiotepa	Rs	Acivicin
A7	Triethylenemelamine	Rs	Dichloroallyl-lawsone
A7	Uracil mustard	Rs	DUP785 (brequinar)
A7	Yoshi-864	Rs	L-Alanosine
T1	Camptothecin	Rs	N-phosphonoacetyl-L-aspartic-acid
T1	Camptothecin,7-Cl	Rs	Pyrazofurin
T1	Camptothecin,9-MeO	Tu	Colchicine

Table II continued

Mechanism of Action*	Drug	Mechanism of Action*	Drug
T1	Camptothecin,9-NH2 (RS)	Tu	Colchicine-derivative
T1	Camptothecin,9-NH2 (S)	Tu	Dolastatin-10
T1	Camptothecin,10-OH	Tu	Halichondrin B
T1	Camptothecin,11-formyl (RS)	Tu	Maytansine
T1	Camptothecin,11-HOMe (RS)	Tu	Trityl-cysteine
T1	Camptothecin,20-ester (S)	Tu	Vinblastine-sulfate
T1	Camptothecin,20-ester (S)	Tu	Vincristine-sulfate
T1	Camptothecin,20-ester (S)	Tu	Taxol (Paclitaxel)
T1	Camptothecin,20-ester (S)	Tu	Taxol analog
T2	Amonafide	Tu	Taxol analog
T2	Amsacrine	Tu	Taxol analog
T2	Anthrapyrazole-derivative	Tu	Taxol analog
T2	Bisantrene	Tu	Taxol analog
T2	Daunorubicin	Tu	Taxol analog
T2	Deoxydoxorubicin	Tu	Taxol analog
T2	Doxorubicin	Tu	Taxol analog
T2	Etoposide	Tu	Taxol analog
T2	Menogaril	Tu	Taxol analog
T2	Mitoxantrone	Tu	Taxol analog
T2	Oxanthrazole (piroxantrone)	P90	Geldanamycin
T2	Teniposide	Uk	3-Hydropicolinaldehyde-thiosemicarbazone
T2	Zorubicin (Rubidazone)	Uk	5-Hydroxypicolinaldehyde-thiosemicarbazone
Pi	L-Asparaginase	Uk	Inosine-glycodialdehyde

***Alkylating agents: A2, A7=alkylating at N-2, N-7 position of guanine, respectively; A6, alkylating at O-6 position of guanine; T1, topoisomerase II inhibitor; T2, topoisomerase II inhibitor; Db, DNA binder; Di, DNA incorporation; Df, antifols; Dr, ribonucleeotide reductase inhibitor; Ds, DNA synthesis inhibitor; Rs, RNA synthesis inhibitor; Tu, tubulin-active antimitotic agents; Pi, protein synthesis inhibitor; P90, hsp90 binder; Uk, unknown**