Supporting Information for the paper

*Table S1. List of the studied chemicals and the values of the six descriptors selected in the best k-NN model generated for the two classes of hepatotoxicity, in addition to cross validation assignments and repeated errors. Values of the descriptors have been rounded to two decimal figures.*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ID CAS** | **NAME** | **CAS** | **ATSC8i** | **VE3\_Dzi** | **SCH-6** | **SHBint9** | **ETA\_d Epsilon\_D** | **Pubchem FP335** | **Activity Class a** | **k-NN Ext b, c** | **Full k-NN CV c** | **Repeated Errors c, d** |
| 1 | Phenylbutazone | 000050-33-9 | -16.47 | -1.49 | 0.45 | 0 | 0 | 1 | 0 | **0** | 1 |   |
| 2 | Aspirin  | 000050-78-2 | 5.98 | -2.46 | 0.08 | 0 | 0.05 | 0 | 1 | 1 | 1 |   |
| 3 | Propylthiouracil | 000051-52-5 | 2.13 | -1.21 | 0.07 | 0 | 0.09 | 0 | 1 | 1 | 1 |   |
| 4 | D\_penicillamine | 000052-67-5 | 0 | -2.70 | 0 | 0 | 0.20 | 0 | 0 | 1 | 0 |   |
| 5 | Haloperidol | 000052-86-8 | 2.49 | -6.19 | 0.24 | 2.44 | 0.02 | 0 | 0 | 0 | 1 |   |
| 6 | 2-Acetaminofluorene | 000053-96-3 | 5.69 | -2.80 | 0.37 | 0 | 0.03 | 0 | 1 | 1 | 1 |   |
| 7 | Furosemide | 000054-31-9 | -6.12 | -3.33 | 0.16 | 2.83 | 0.11 | 0 | 0 | 0 | 0 |   |
| 8 | Isoniazid | 000054-85-3 | 0 | -2.06 | 0.10 | 0 | 0.14 | 0 | 0 | 1 | 0 |   |
| 9 | N-nitrosodiethylamine | 000055-18-5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | **1** | 1 |   |
| 10 | Carbon\_tetrachloride | 000056-23-5 | 0 | -18.37 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |   |
| 11 | Chloramphenicol | 000056-75-7 | 3.81 | -4.43 | 0.08 | 5.67 | 0.09 | 0 | 1 | 1 | 1 |   |
| 12 | Phenobarbital\_sodium | 000057-30-7 | 3.10 | -4.45 | 0.15 | 0 | 0.06 | 0 | 0 | **0** | 0 |   |
| 13 | Phenytoin | 000057-41-0 | 4.53 | -67.42 | 0.47 | 0 | 0.05 | 0 | 0 | *1* | *1* |  ***+*** |
| 14 | 17a-ethynylestradiol | 000057-63-6 | -7.76 | -12.66 | 0.46 | 0 | 0.04 | 1 | 1 | 1 | 1 |   |
| 15 | Caffeine | 000058-08-2 | 0 | -3.43 | 0.24 | 0 | 0 | 0 | 0 | 0 | 1 |   |
| 16 | Methyltestosterone | 000058-18-4 | -5.99 | -7.91 | 0.46 | 0 | 0.02 | 1 | 1 | 1 | 0 |   |
| 17 | Promethazine | 000058-33-3 | -28.18 | -2.44 | 0.21 | 0 | 0 | 0 | 1 | **1** | 1 |   |
| 18 | Theophylline | 000058-55-9 | 0 | -2.90 | 0.17 | 0 | 0.05 | 0 | 1 | 0 | 1 |   |
| 19 | Puromycin aminonucleoside | 000058-60-6 | 2.40 | -3.06 | 0.51 | 1.04 | 0.10 | 0 | 1 | 0 | 0 |  - |
| 20 | Cephalothin sodium | 000058-71-9 | -5.08 | -4.29 | 0.70 | 4.14 | 0.04 | 0 | 0 | 1 | 1 |  + |
| 21 | Methotrexate | 000059-05-2 | 0.96 | -4.49 | 0.21 | 1.15 | 0.11 | 0 | 0 | 0 | 0 |   |
| 22 | Acetazolamide | 000059-66-5 | 4.13 | -3.08 | 0.14 | 0 | 0.14 | 0 | 0 | **0** | 0 |   |
| 23 | Nicotinic\_acid | 000059-67-6 | 0 | -1.69 | 0.10 | 0 | 0.06 | 0 | 0 | 1 | 1 |  + |
| 24 | Nitrofurazone | 000059-87-0 | -0.74 | -2.56 | 0.15 | 7.45 | 0.13 | 0 | 0 | 0 | 0 |   |
| 25 | Methimazole | 000060-56-0 | 0 | -2.63 | 0.24 | 0 | 0.07 | 0 | 0 | 0 | 0 |   |
| 26 | Papaverine | 000061-25-6 | 5.98 | -4.63 | 0.19 | 0 | 0 | 0 | 0 | 0 | 0 |   |
| 27 | Mefenamic\_acid | 000061-68-7 | -13.50 | -7.47 | 0.15 | 0 | 0.05 | 0 | 0 | **0** | 0 |   |
| 28 | Phenacetin | 000062-44-2 | 7.65 | -3.40 | 0.08 | 0 | 0.04 | 0 | 1 | 1 | 1 |   |
| 29 | Thioacetamide | 000062-55-5 | 0 | -0.82 | 0 | 0 | 0.19 | 0 | 1 | 0 | 0 |  - |
| 30 | Ethanol | 000064-17-5 | 0 | -1.10 | 0 | 0 | 0.16 | 0 | 0 | 0 | 0 |   |
| 31 | Tetracycline | 000064-75-5 | -12.72 | -11.61 | 0.17 | 9.75 | 0.11 | 1 | 0 | 0 | 0 |   |
| 32 | Tolbutamide | 000064-77-7 | -10.15 | -2.17 | 0.08 | 0 | 0.06 | 0 | 0 | 0 | 0 |   |
| 33 | Colchicine | 000064-86-8 | -16.08 | -6.44 | 0.05 | 0 | 0.02 | 0 | 1 | **0** | 0 |  - |
| 34 | Nitrofurantoin | 000067-20-9 | -3.98 | -7.75 | 0.41 | 7.19 | 0.04 | 0 | 0 | 0 | 1 |   |
| 35 | DL\_ethionine | 000067-21-0 | 5.95 | -1.46 | 0 | 0 | 0.14 | 0 | 1 | 0 | 0 |  - |
| 36 | Vitamin\_A | 000068-26-8 | 21.50 | -3.39 | 0.06 | 0 | 0.02 | 0 | 1 | 1 | 1 |   |
| 37 | Benzodiarone | 000068-90-6 | 2.67 | -4.54 | 0.35 | 0 | 0.02 | 0 | 1 | 1 | 1 |   |
| 38 | Chlorpromazine\_HCL | 000069-09-0 | -12.57 | -1.61 | 0.20 | 0 | 0 | 0 | 0 | **0** | 0 |   |
| 39 | Chlormezanone | 000080-77-3 | 2.85 | -3.88 | 0.13 | 0 | 0 | 0 | 1 | 1 | 1 |   |
| 40 | N-phenylanthranilic\_acid | 000091-40-7 | -0.65 | -4.32 | 0.19 | 0 | 0.06 | 0 | 0 | 1 | 0 |   |
| 41 | Coumarin  | 000091-64-5 | 0 | -3.85 | 0.15 | 0 | 0 | 0 | 1 | **1** | 1 |   |
| 42 | Chlorpropamide | 000094-20-2 | 0.20 | -2.56 | 0.08 | 0 | 0.07 | 0 | 1 | 0 | 1 |   |
| 43 | Disulfiram | 000097-77-8 | -10.36 | -54.47 | 0 | 0 | 0 | 0 | 1 | *1* | *1* |   |
| 44 | Acetaminophen | 000103-90-2 | 1.93 | -2.40 | 0.08 | 0 | 0.09 | 0 | 1 | 1 | 1 |   |
| 45 | Allyl\_Alcohol | 000107-18-6 | 0 | -0.96 | 0 | 0 | 0.12 | 0 | 1 | 1 | 1 |   |
| 46 | Bromobenzen | 000108-86-1 | 0 | -1.57 | 0.10 | 0 | 0 | 0 | 1 | **1** | 1 |   |
| 47 | Imipramine | 000113-52-0 | 3.73 | -1.51 | 0.17 | 0 | 0 | 0 | 0 | 1 | 1 |  + |
| 48 | Hexachlorobenzene | 000118-74-1 | 0 | -42.77 | 0.04 | 0 | 0 | 0 | 1 | 1 | 1 |   |
| 49 | Griseofulvin | 000126-07-8 | -3.05 | -5.99 | 0.32 | 0 | 0 | 1 | 0 | 0 | 0 |   |
| 50 | Trimethadione | 000127-48-0 | 0 | -1.91 | 0.34 | 0 | 0 | 0 | 1 | 1 | 1 |   |
| 51 | Thioridazine | 000130-61-0 | -34.69 | -2.77 | 0.28 | 0 | 0 | 0 | 0 | 1 | 1 |  + |
| 52 | Methapyrilene | 000135-23-9 | -24.22 | -5.83 | 0.20 | 0 | 0 | 0 | 1 | 1 | 0 |   |
| 53 | Fluphenazine | 000146-56-5 | -40.93 | -2.12 | 0.28 | 0 | 0.02 | 0 | 0 | **0** | 0 |   |
| 54 | Carbamazepine | 000298-46-4 | 5.28 | -64.15 | 0.17 | 0 | 0.05 | 0 | 0 | *1* | *1* |  ***+*** |
| 55 | Chlormadinone | 000302-22-7 | 1.02 | -6.42 | 0.43 | 0 | 0.02 | 1 | 0 | 0 | 0 |   |
| 56 | Iproniazid | 000305-33-9 | -15.70 | -2.10 | 0.10 | 0 | 0.07 | 0 | 0 | 0 | 0 |   |
| 57 | Monocrotaline | 000315-22-0 | 0.56 | -13.88 | 0.38 | 0 | 0.05 | 1 | 1 | 1 | 0 |   |
| 58 | Allopurinol | 000315-30-0 | 0 | -3.82 | 0.22 | 0 | 0.11 | 0 | 0 | 0 | 0 |   |
| 59 | Triamterene | 000396-01-0 | -5.46 | -3.33 | 0.21 | 0 | 0.13 | 0 | 0 | **0** | 0 |   |
| 60 | Diazepam | 000439-14-5 | 0.63 | -3.55 | 0.17 | 0 | 0 | 0 | 1 | **1** | 1 |   |
| 61 | Ethionamide | 000536-33-4 | 5.48 | -4.23 | 0.08 | 0 | 0.08 | 0 | 1 | 1 | 1 |   |
| 62 | Amitriptyline | 000549-18-8 | 6.79 | -1.78 | 0.17 | 0 | 0 | 0 | 0 | 0 | 0 |   |
| 63 | naphtyl\_isothiocyanate | 000551-06-4 | -2.57 | -3.21 | 0.15 | 0 | 0 | 0 | 1 | 1 | 1 |   |
| 64 | Sulfasalazine | 000599-79-1 | -11.12 | -13.48 | 0.25 | 0 | 0.06 | 0 | 1 | 0 | 0 |  - |
| 65 | Clofibrate | 000637-07-0 | -6.70 | -3.42 | 0.08 | 0 | 0 | 0 | 0 | 1 | 1 |  + |
| 66 | Moxisylyte | 000964-52-3 | 28.57 | -2.65 | 0.06 | 0 | 0 | 1 | 0 | 0 | 0 |   |
| 67 | Sodium\_Valproate | 001069-66-5 | 4.64 | -34.34 | 0 | 0 | 0.05 | 1 | 0 | 0 | 0 |   |
| 68 | Ethambutol | 001070-11-7 | -11.85 | -50.86 | 0 | 10.27 | 0.13 | 0 | 1 | *1* | *1* |   |
| 69 | Bucetin | 001083-57-4 | 0.80 | -3.56 | 0.08 | 3.02 | 0.06 | 0 | 1 | **1** | 1 |   |
| 70 | Metformin | 001115-70-4 | 0 | -3.02 | 0 | 0 | 0.22 | 0 | 0 | **0** | 0 |   |
| 71 | Adapin | 001229-29-4 | -7.98 | -1.68 | 0.17 | 0 | 0 | 0 | 0 | 1 | 0 |   |
| 72 | Erythromycin | 001264-62-6 | 14.97 | -13.95 | 0.10 | 10.44 | 0.05 | 1 | 0 | 0 | 0 |   |
| 73 | Gentamicin sulfate | 001405-41-0 | -10.96 | -11.20 | 0.16 | 9.85 | 0.16 | 0 | 0 | 0 | 0 |   |
| 74 | Tacrine | 001684-40-8 | 3.62 | -2.21 | 0.21 | 0 | 0.05 | 0 | 0 | 0 | 0 |   |
| 75 | Tiopronin | 001953-02-2 | 2.15 | -5.32 | 0 | 0 | 0.16 | 0 | 0 | **0** | 0 |   |
| 76 | Pemoline | 002152-34-3 | 2.40 | -4.52 | 0.35 | 0 | 0.08 | 0 | 0 | 0 | 0 |   |
| 77 | Hydroxyzine | 002192-20-3 | 2.83 | -1.83 | 0.27 | 1.66 | 0.02 | 0 | 0 | 0 | 1 |   |
| 78 | 2-Bromoethylamine | 002576-47-8 | 0 | -1.35 | 0 | 0 | 0.17 | 0 | 0 | 0 | 0 |   |
| 79 | Benzbromarone | 003562-84-3 | 2.72 | -4.79 | 0.35 | 0 | 0.02 | 0 | 1 | 1 | 1 |   |
| 80 | Disopyramide | 003737-09-5 | 11.84 | -2.81 | 0.20 | 0 | 0.04 | 0 | 1 | 0 | 0 |  - |
| 81 | Ajmaline | 004360-12-7 | 9.02 | -4.10 | 0.71 | 0 | 0.04 | 1 | 0 | 0 | 0 |   |
| 82 | Mexiletine | 005370-01-4 | 21.90 | -1.49 | 0.07 | 0 | 0.07 | 0 | 1 | 1 | 1 |   |
| 83 | Cyclophosphamide | 006055-19-2 | 3.42 | -2.79 | 0.09 | 0 | 0.04 | 0 | 1 | 1 | 1 |   |
| 84 | Quinidine | 006591-63-5 | -4.56 | -3.90 | 0.33 | 0 | 0.02 | 1 | 0 | **0** | 0 |   |
| 85 | Perhexiline | 006724-53-4 | -3.51 | -6.27 | 0.31 | 0 | 0.02 | 1 | 0 | 0 | 0 |   |
| 86 | Chlorpheniramine | 007054-11-7 | -14.55 | -3.65 | 0.19 | 0 | 0 | 1 | 0 | 0 | 0 |   |
| 87 | Gibenclamide | 010238-21-8 | 15.28 | -11.24 | 0.25 | 0 | 0.05 | 0 | 0 | 0 | 0 |   |
| 88 | Lomustine | 013010-47-4 | -1.38 | -3.18 | 0.10 | 0 | 0.04 | 0 | 1 | **1** | 1 |   |
| 89 | Rifampicin | 013292-46-1 | 39.69 | -7.87 | 0.42 | 0.33 | 0.05 | 1 | 0 | 0 | 0 |   |
| 90 | Flutamide | 013311-84-7 | -58.17 | -2.22 | 0.07 | 0 | 0.04 | 1 | 1 | 1 | 0 |   |
| 91 | Diclofenac\_Sodium | 015307-79-6 | -7.66 | -3.70 | 0.15 | 0 | 0.05 | 0 | 0 | **0** | 1 |   |
| 92 | Sulpiride | 015676-16-1 | -7.00 | -3.25 | 0.23 | 1.25 | 0.07 | 0 | 0 | 0 | 0 |   |
| 93 | Ibuprofen | 015687-27-1 | -3.50 | -3.10 | 0.08 | 0 | 0.03 | 1 | 0 | 0 | 0 |   |
| 94 | Danazol | 017230-88-5 | 16.88 | -6.62 | 0.60 | 0 | 0.02 | 1 | 0 | 0 | 0 |   |
| 95 | Chlomipramine | 017321-77-6 | 1.33 | -1.57 | 0.15 | 0 | 0 | 0 | 1 | 1 | 1 |   |
| 96 | Amiodarone | 019774-82-4 | -44.42 | -3.68 | 0.35 | 0 | 0 | 0 | 0 | 1 | 1 |  + |
| 97 | Bendazac | 020187-55-7 | -7.99 | -6.81 | 0.41 | 0 | 0.02 | 0 | 1 | 1 | 1 |   |
| 98 | Nifedipine | 021829-25-4 | 3.94 | -7.35 | 0.13 | 0 | 0.02 | 1 | 0 | **0** | 0 |   |
| 99 | Dantrolene | 024868-20-0 | -2.91 | -8.02 | 0.50 | 0 | 0.03 | 0 | 1 | 1 | 1 |   |
| 100 | Doxorubicin | 025316-40-9 | -33.95 | -5.16 | 0.25 | 30.69 | 0.10 | 0 | 0 | *0* | *0* |   |
| 101 | Gemfibrozil | 025812-30-0 | -5.29 | -3.21 | 0.07 | 0 | 0.03 | 0 | 1 | **1** | 1 |   |
| 102 | Triazolam | 028911-01-5 | -6.35 | -3.06 | 0.37 | 0 | 0 | 0 | 0 | 1 | 1 |  + |
| 103 | Labetalol | 032780-64-6 | 10.33 | -3.71 | 0.17 | 0 | 0.09 | 0 | 0 | 0 | 0 |   |
| 104 | Diltiazem | 033286-22-5 | -49.63 | -6.17 | 0.17 | 0 | 0 | 0 | 1 | 0 | 0 |  - |
| 105 | Etoposide | 033419-42-0 | -46.38 | -5.83 | 0.64 | 12.03 | 0.04 | 1 | 0 | 1 | 0 |   |
| 106 | Sulindac | 038194-50-2 | 0.35 | -3.69 | 0.40 | 0 | 0.02 | 0 | 1 | 1 | 1 |   |
| 107 | Methyldopa | 041372-08-1 | 8.10 | -3.25 | 0.07 | 0 | 0.16 | 0 | 0 | **0** | 0 |   |
| 108 | Fenofibrate | 049562-28-9 | -22.87 | -6.90 | 0.17 | 0 | 0 | 0 | 1 | 1 | 1 |   |
| 109 | WY-14643 | 050892-23-4 | -2.19 | -2.98 | 0.14 | 0 | 0.05 | 0 | 1 | 0 | 1 |   |
| 110 | Cimetidine | 051481-61-9 | 5.86 | -7.53 | 0.20 | 2.94 | 0.08 | 0 | 0 | 0 | 0 |   |
| 111 | Nimesulide | 051803-78-2 | -4.10 | -4.63 | 0.17 | 0 | 0.03 | 0 | 1 | **1** | 1 |   |
| 112 | Ticlopidine | 053885-35-1 | 1.24 | -15.05 | 0.32 | 0 | 0 | 0 | 1 | 1 | 0 |   |
| 113 | Tamoxifen | 054965-24-1 | -8.45 | -2.20 | 0.29 | 0 | 0 | 0 | 0 | 0 | 0 |   |
| 114 | Acarbose | 056180-94-0 | -20.62 | -31.00 | 0.18 | 56.02 | 0.17 | 0 | 0 | *0* | *0* |   |
| 115 | Captopril | 062571-86-2 | 5.17 | -2.42 | 0.14 | 0 | 0.08 | 1 | 0 | 0 | 0 |   |
| 116 | Ketoconazole | 065277-42-1 | 10.75 | -4.49 | 0.54 | 0 | 0 | 0 | 0 | **1** | 1 |  + |
| 117 | Ranitidine | 066357-59-3 | -6.69 | -11.52 | 0.17 | 0.78 | 0.05 | 0 | 0 | 1 | 0 |   |
| 118 | Lornoxicam | 070374-39-9 | -8.77 | -3.72 | 0.33 | 0 | 0.05 | 0 | 0 | 0 | 0 |   |
| 119 | Omeprazole | 073590-58-6 | -12.73 | -7.43 | 0.32 | 0 | 0.02 | 0 | 0 | 0 | 0 |   |
| 120 | Enalapril | 076095-16-4 | -23.07 | -6.00 | 0.24 | 10.20 | 0.04 | 0 | 0 | 0 | 1 |   |
| 121 | Famotidine | 076824-35-6 | -23.50 | -4.71 | 0.17 | 0 | 0.19 | 0 | 0 | **0** | 0 |   |
| 122 | Terbinafine | 078628-80-5 | 51.19 | -2.46 | 0.15 | 0 | 0 | 0 | 1 | 1 | 0 |   |
| 123 | Simvastatin | 079902-63-9 | -62.09 | -5.38 | 0.18 | 2.41 | 0.02 | 1 | 1 | *1* | *1* |   |
| 124 | Ciprofloxacin | 093107-08-5 | -4.17 | -5.02 | 0.39 | 3.86 | 0.05 | 0 | 0 | 0 | 0 |   |
| a 0 = Not Hepatotoxic; 1= Hepatotoxic. |
| b Prediction set chemicals are indicated in bold. |
| c Chemicals outside the Applicability Domain defined by PCA and AD by Standardised Descriptors method (Roy et al., Chemom Intell Lab Sys 2015) are reported in italic. |
| d "+" = overestimations; "-" = underestimations. |

*Figure S1. ROC plot. Comparison of results from different validation approaches calculated for the-NN full model.*



*Figure S2. PCA analysis performed on the modelling descriptors: distribution of the input variables in the PC1-PC2 space. The length of the segments originated in the centre of the graph is the weight (loading) of the respective variable in the PCA graph.*

