Predicting Mouse Lifespan-Extending Chemical Compounds with Machine Learning

Aleksey V. Belikov[1§], Caio Ribeiro[2§*], Christopher K. Farmer[3], João Pedro de Magalhães[1&*], Alex A. Freitas[2&]

[§] These authors contributed equally

[&] These authors contributed equally as senior authors

[1] Genomics of Ageing and Rejuvenation Lab, Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK.

[2] School of Computing, University of Kent, Canterbury, Kent, UK.

[3] Centre for Health Services Studies, University of Kent, Canterbury, Kent, UK.

* Corresponding Authors

Corresponding Authors Contact Details:

Name: João Pedro de Magalhães
Address: Genomics of Ageing and Rejuvenation Lab, Institute of Inflammation and Ageing, University of Birmingham, Edgbaston, Birmingham, UK
E-mail: j.p.magalhaes@bham.ac.uk

Name: Caio Ribeiro
Address: School of Computing. University of Kent. Darwin Rd, Canterbury CT2 7NF
E-mail: C.E.Ribeiro@kent.ac.uk

Abstract

Pharmacological interventions targeting the biological processes of ageing hold significant potential to extend healthspan and promote longevity. In this study, we employed machine learning to predict how likely it is for a given chemical compound to extend lifespan. We used murine lifespan data from the DrugAge database for training the models. Our most successful Random Forest classifiers were trained on the annotations of direct protein targets of compounds, such as Gene Ontology, UniProt Keywords, pathways (KEGG, Reactome, Wiki) and protein domains (InterPro), whereas models trained on gene expression (LINCS) and chemical substructures (PubChem) underperformed. Models trained on male datasets performed better than those trained on mixed-sex and female datasets, with the latter suffering from severe class imbalance due to much fewer positive-class instances. Notably, features related to G-protein coupled receptors, especially receptors for neurotransmitters, metabolic hormones and sex hormones, were identified as strong predictors of lifespan extension. We used ensemble classifiers comprised of top models to screen compounds from DrugBank, highlighting novel candidates for longevity studies. Major clusters of compounds with the highest predicted longevity-promoting effects appear to target IGF1 and insulin receptors, beta adrenergic receptors, carbonic anhydrases, dopamine and serotonin receptors, voltage-gated potassium and calcium channels, sodium-dependent dopamine, serotonin and noradrenalin transporters, muscarinic acetylcholine receptors and adenosine receptors. Our study provides an important contribution not only to the longevity pharmacology field but also informs research on the fundamental mechanisms of ageing.

Key Words

Machine learning, lifespan-extending compounds, longevity drugs, anti-ageing compounds, geroprotectors, Mus musculus

1 Introduction

The use of pharmacological interventions to promote healthspan and longevity, and to target the biological process of ageing, is currently a very active research area [1–4]. Hundreds of drug treatments have been successfully applied to significantly increase the lifespan of model organisms such as *Caenorhabditis elegans* and *Mus musculus*[2,3].

The large amounts of data related to developing longevity-promoting treatments can be analysed through the use of machine learning (ML) tools such as classification models, which are first trained on a set of instances (examples) and then are able to make predictions about previously unseen data [5]. In addition to making predictions, classification models can be analysed to give researchers insights into the prediction problem, such as patterns shared by groups of instances (chemical compounds) or ranking the predictive features by their relative importance for classifying instances. The results from machine learning experiments can inform *in vivo* experiments and help guide the development of new drugs or treatments.

Previous studies that focus on other model organisms, such as *Caenorhabditis elegans*, have applied machine learning to longevity data [6–9]. However, to the best of our knowledge, this is the first work that applies classification models from machine learning to predict whether a chemical compound promotes longevity in mice. Using *Mus musculus* as the target organism is more challenging than using simpler organisms, mainly due to their sex dichotomy. However, mice have many genes with human homologues, well-characterised strains, and relatively short reproductive cycles and lifespans. Compared to non-mammal model organisms, the insights into the biological mechanisms of ageing that apply to mice are significantly more likely to apply to other mammals such as humans [10–12]. These and other characteristics make studies in mice an expected preclinical step in drug discovery pipelines.

In order to create the datasets to train our classification models, we obtained the mouse data from the DrugAge database [13,14], including the results from the Interventions Testing Program (ITP) [15], which compiles results of peer-reviewed research on lifespan extension in model organisms. Each compound-sex combination is assigned a class label, determined by the results of studies in mice using the compound, which indicates whether it is associated with lifespan extension (positive-class) or not (negative-class).

Naturally, the efficacy of ML predictive models depends largely on the data used to train them. As longevity experiments vary in treatment regimens, mouse strains and lifespan metrics, it is more difficult for the predictive models to make correct generalizations, which are required for making accurate predictions on unseen data. One of the objectives of our study was to explore several different feature types and form a varied baseline of datasets. For this study we prepared 21 datasets for training machine learning models, using three data sources, namely STRING (protein-protein interaction networks and protein annotations [16]), LINCS (gene expression data [17]) and PubChem (chemical substructure data [18]). The data sources contain several principally different types of features that represent biological and chemical information about each compound such as the Gene Ontology, protein domains and pathway annotations of their targets. By preparing datasets that use a wide array of descriptors, we were able to explore the efficacy of orthogonal approaches for the task of predicting longevity-related compounds for mice. Section 2.2 contains a detailed description of the types of features used in our study, and methods of data collection and preparation.

We evaluated the predictive accuracy of classification models trained with each dataset, either mixing instances from male and female mice or exclusively using instances from each sex. We then selected 10 models (5 from mixed datasets and 5 from male-only datasets) to further analyse, by interpreting the features deemed most relevant for determining whether or not a compound

belongs to the positive class (promotes mice longevity) in each of those models. Models learned from female-only datasets were not analysed in detail because they led to smaller predictive accuracies than male-only datasets in general, as discussed later.

In addition to the feature importance analysis, we looked into the negative-class compounds that were most consistently classified as belonging to the positive-class (false-positive classifications) by our selected models. We believe these compounds, e.g. LY444711, putrescine, chlorpheniramine, and dehydroepiandrosterone sulphate, can be promising options for future longevity studies in mice, as although previous results did not find significant life extension in treatments with them, they share important characteristics with current successful life extension compounds.

Finally, we combined the selected models as male-only and mixed-sex ensembles of classifiers (i.e., making a final classification for each entry, based on the decisions of all members of the ensemble) and used these ensembles to classify thousands of previously unseen and unlabelled instances from an external database – DrugBank [19]. We highlighted top compounds confidently classified as members of the positive class as promising options for future longevity studies.

2 Methods

2.1 Instance definition and labelling

Each instance in our datasets represents a combination of a compound and the sex of the mice used in the experiments, based on data obtained from DrugAge build 5 [13], a database that collates results from peer-reviewed studies on the effects of chemical compounds (drugs) on the lifespan of model organisms. Its *Mus musculus* section contains data on more than a hundred different compounds.

As mouse lifespan experiments often have different results based on the sex of the animals tested [20,21], when required we created different instances for the same compound based on the sex of the mice used in the experiment. This greatly increases the number of examples available in the dataset and avoids conflating data that could lead the classification model to find incorrect patterns. However, the trade-off is that the feature values for all other features (except the sex feature) are the same among such instances, which can hinder the classifier's performance if mixed-sex data is used to train the model.

Each instance in the datasets is assigned a binary class label indicating presence (class 1, positive) or absence (class 0, negative) of substantial evidence of lifespan increase when the compound was used in treatment during the longevity experiments reported in the papers. Naturally, the labelling process depends on many variables, as do the experiments performed in the source studies. The process we used to get a consistent class label definition for each instance, based on DrugAge entries, was as follows:

First, for each peer-reviewed publication studying the effect of a compound on mice, we analysed all reported results for each compound-sex combination. If at least one dosage or treatment regimen produces a median (or mean, where the median is not available) lifespan extension that is greater than 5% compared to the control mice population and is statistically significant (with p-value ≤ 0.05), then we count that as a positive result, otherwise we count it as a negative result.

Then, the results from different publications are aggregated so that each compound-sex combination is assigned a single class label, as follows. If the number of positive results is greater or equal to the number of negative results, the instance for that compound-sex combination gets a

positive class label. Otherwise, it is labelled as the negative class (i.e., no substantial evidence of lifespan increase).

In summary, each instance in our dataset represents a combination of a compound and the sex of the mice used in the experiments (mixed-sex datasets have from 88 to 158 instances, depending on availability of data). An instance's class label is defined by the consensus of the experimental results from DrugAge's *M. musculus* section referring to that compound-sex combination, with positive labels indicating evidence of a positive association with longevity for treatments using that compound.

2.2 Dataset Preparation and Descriptions

To create our datasets, we created predictive features obtained from three source databases, namely STRING [16], LINCS [17] and PubChem [18]. The features are different types of biological and chemical descriptors of compounds and their targets, all of which have been applied to drug discovery before. In this Section we detail the data collection processes for each of these feature types.

The STRING-based datasets use binary feature values from the descriptors (i.e., Gene Ontology, protein domains, biological pathways and keywords) of the targets of each compound, i.e., the proteins a compound was designed to interact with or known to be the main mediators of its pharmacological action. We obtained the list of a compound's targets using several different sources, as detailed later in this Section. Additionally, the target descriptors are collected from STRING in two ways (functional annotation and enrichment), resulting in substantially different datasets, as described in Subsection 4.2.1.

The LINCS-based datasets have human gene expression data; consolidated numeric representations of over- and under-expression of each human gene from cell-line experiments after exposure to a compound. The consolidated values represent gene expression trends over all experiments with different configurations (cell-lines, doses and times).

Finally, the PubChem-based dataset is the only dataset of chemical descriptors used in our study. It has a set of binary variables, called Molecular Fingerprints, each one representing whether a compound's structure contains a given chemical substructure or pattern.

For the datasets with binary feature values (i.e., the STRING-based datasets and Molecular Fingerprints) there are several features with too little variation in their values across the instances, which have little to no predictive value for a classification model. Therefore, as a data cleaning step prior to training our models, we applied a simple frequency-threshold filter to remove features with fewer than 3 instances with a "1" value (or fewer than 3 instances with a "0" value). After applying the frequency-threshold filter we also removed instances that had only "0" values in all remaining features, as they could no longer be distinguished by the classifiers.

2.2.1 Datasets of STRING Descriptors of Protein Targets

We created a dataset based on target proteins associated with each compound and their descriptors, as follows. Firstly, instead of considering protein interactors listed on STITCH (a compound-protein interaction database [22], associated with STRING), which include predicted compound-protein interactions based on different levels of evidence, we have focused on identifying the specific targets of a compound, in order to create predictive features which are intuitively more biologically relevant. In order to cover the majority of compounds in DrugAge's mice data and to obtain a reliable list of targets for each of them, we combined different sources for obtaining the list of targets of each compound:

- **DrugBank** [19]: target data obtained from the DrugBank database's website, accessed 01/2024;
- **Pharos** [23]: target data obtained using the Pharos GraphQL API, Pharos v3.18.1;
- **BioGRID** [24]: target data obtained through the database's protein target table, v4.4.228;
- **ChEMBL** [25]: target data obtained using the *chembl_webresource_client* python library [26];
- **Therapeutic Target Database** [27]: target data obtained using the TTD database's website, accessed 01/2024;
- **Source publications** for the compound's DrugAge entry and other relevant publications about the compound in case only one or no targets were identified via the sources listed above

Note that all target proteins obtained are for *Homo sapiens*, not *Mus musculus*, since the above databases focus on human data. After consolidating the target lists for all compounds, we used the STRING database API (v12.0) to obtain the protein descriptors for those targets through either functional annotation (FA) or neighbour enrichment (NE). In both FA and NE, STRING was queried for each protein target individually. However, unlike simple annotations returned by FA calls, NE calls returned statistically significant enrichments of annotations for the target protein and its 10 nearest functionally associated neighbours (automatically added by STRING).

The STRING API response collates annotations for the input proteins from various sources, related to their structure and function. We selected a subset of the feature types included in the API's response, excluding those that were not relevant for our classification problem of drug discovery for longevity studies. The following feature types were used to create the binary predictive features in the Targets datasets:

- **Gene Ontology (GO) Terms**: These are widely used in similar drug discovery applications, as broad descriptors of genes' structure and functions. We included all three categories of GO terms [28]: Biological Processes, Molecular Functions and Cellular Components.
- **Protein Domains**: These features also represent information about the structure and function of proteins. The domains represented in the created dataset are from InterPro [29]. Pfam and SMART (Simple Modular Architecture Research Tool) protein domains are also present in the STRING API responses, but these were not included, the former because they were incorporated by InterPro and the latter because there were too few annotations for them.
- **Pathways**: These features contain biological pathway information from Reactome [30], WikiPathways [31] and KEGG [32]. Note that KEGG pathway features are not present in the FA datasets due to KEGG license restrictions keeping them from being returned in that call.
- **Uniprot Keywords**: These features represent a hierarchical set of keywords used to describe a protein entry on Uniprot [33].

In total we have 8 feature types in the Targets datasets, resulting in a considerably large number of features in each of them: 8,228 features in the FA dataset and 6,459 features in the NE dataset, after applying the minimum-threshold filter (these values refer to the full datasets, with 143 male and female mice instances). In order to assess the individual predictive performance of each feature type, we also did experiments with datasets prepared with single feature types.

2.2.2 Datasets of LINCS Gene Expression

The Library of Integrated Network-Based Cellular Signatures (LINCS, v1.1) [17] dataset stores, among other types of data, *Homo sapiens* gene expression signatures. We created two datasets using the LINCS data, one including all genes (12,328 genes) and another using a subset of landmark genes

called L_1000 (978 genes) [34]. The latter refers to the genes for which the gene expression values were directly measured by the experiments, rather than inferred by the database designers.

For each gene expression experiment result stored in the LINCS database, the features are numeric values representing the (measured or estimated) expression of each human gene in cells belonging to a given cell line, after exposure to a dose of a given compound for a specified amount of time. Often dozens of experiments are performed for each compound, with different doses, times and cell lines. Therefore, a single compound can be associated with up to hundreds of entries on LINCS, with different gene expression values in each entry.

In order to use the gene expression values from LINCS in our machine learning context where each instance refers to a compound-sex combination, we need to have a single value for each gene (i.e., one row per compound), which broadly represents its gene expression after exposure to the compound. Hence, for each gene, we consolidated its expression value for each compound through a two-step process, as follows.

First, we selected the gene expression from the median dose over all entries in each cell line as the signature value for that cell line. Then, we calculated the median gene expression value over all cell lines' signatures values. This is repeated for all genes in the dataset so that, after performing these steps, each gene has a single expression value for each compound. The numeric values vary roughly between -2 and +2, with positive values indicating an over-expression of the gene and vice-versa.

Naturally, this data consolidation process incurs some data loss through reducing the granularity of the LINCS data, and the validity of the aggregated values depend on many factors such as the standard deviation of the gene expression values measured over different experiments. On the other hand, the LINCS datasets have numerical values for their features, which may be advantageous for decision tree-based algorithms (like the algorithm used in this work) because the algorithm has more options for partitioning the data based on a feature's values.

2.2.3 Dataset of PubChem Molecular Fingerprints

Studies that apply machine learning for drug discovery problems often use chemical substructures of compounds as predictive features, either by themselves or in combination with biological data [7–9]. In our study, we created datasets where the binary features represent the absence/presence of 880 chemical substructures in a given compound. Importantly, the feature values for chemical data are independent from the model organism and, in the case of the substructures used in our study, easily obtainable (Molecular Fingerprints are available for any compound present in PubChem).

The Molecular Fingerprints substructures may represent an element count (e.g., compound has ≥ 8 hydrogen atoms), a type of ring system (e.g., compound has saturated or aromatic carbon-only ring size 3), atom pairings, neighbourhoods and bonding (e.g., compound has Si – Cl bond), and other such patterns. The data was obtained from the PubChem Substructure Fingerprint data, using the *pubchempy* python library (v1.0.4).

2.3 Experimental Setup

We trained Random Forest (RF) classifiers [35] using each of the datasets produced for this study, using the same experimental setup described in this Section. Ensembles of classifiers based on decision trees (like RFs) tend to outperform more complex algorithms such as deep learning when applied to tabular data for binary classification [36], and RFs also have the advantage of being much faster than deep learning algorithms in general. In addition, although RFs are not as interpretable as single decision trees, they still retain some degree of interpretability, mainly through feature importance metrics that take into account the contents of the trees in the forest (i.e. the model's internal mechanisms).

For all experiments we performed a 5-fold cross-validation process for each RF classifier and ran 10 different experiments for each dataset (varying the random seed parameter, resulting in different training/test data divisions and different RFs). The rationale for repeated experiments was to obtain a more robust estimation of the models' predictive accuracy, as our datasets have few instances and, therefore, there is more stochastic variation on each models' predictive accuracy estimation.

In Section 3 we report the average values over all folds (5 folds over 10 cross-validation runs) of two predictive accuracy metrics: the Area Under the Receiver Operating Characteristic Curve (AUC) [37] and the Geometric Mean of Sensitivity and Specificity (GMean). Both metrics vary between 0 and 1, with 1 representing a perfect classifier. The AUC is widely used as a general representation of a binary classification model's performance, and it has the advantage of providing a measurement that is independent of the threshold used for the class-label decision. When used by itself the AUC may result in over-optimistic evaluations of models that favour the majority class, so we also used the GMean as a conservative metric that assigns equal weight to both class labels and considered both metrics when evaluating our RF models.

The RFs were trained using the *sklearn* implementation (v1.4.0) [38], with 500 random trees and the default values of all other hyperparameters. The datasets created for this research have a class-imbalance issue (between 22% and 47% positive-class instances, and so between 78% and 53% negative-class instances, depending on the dataset). Although none of the imbalances are very strong, this can still lead to models biased in favour of the majority class (the negative class, representing compounds without a known association with longevity). Therefore, in each RF, we increased the weight of the minority class instances in the training set until a balance was reached, meaning the classifier penalized misclassifications of minority class instances more severely. This was preferable to the common alternative method of undersampling majority class instances, as class-weight adjustment does not reduce the number of examples in the dataset.

After training and evaluating our RF models, we selected a group of 10 models (5 from mixed-sex datasets and 5 from male-only datasets) that had the best predictive performance, to analyse how they made their predictions and to identify compounds with potential for lifespan extension in mice.

First, we ranked all features used for each of the selected models, based on each feature's ability to discriminate between the positive-label and negative-label examples, by using the default feature importance measure in sklearn (the average class-impurity decrease measure [39]). Then, the top features in that ranking are identified as the most important features for that model.

Next, we used the selected models to identify promising compounds for future studies of mice longevity, in two ways. The first was identifying the consistent false-positive classifications (i.e., over half of the top models classified the compound as positive whilst the compound is annotated in the data as negative) in the labelled DrugAge data used in these experiments. The second way was using the selected models as ensembles of RFs to classify unlabelled data from an external dataset and identifying the most confident positive-class classifications by these ensembles. In order to discuss the over 200 compounds identified in the external dataset analysis, we used the UMAP (Uniform Manifold Approximation and Projection) dimension reduction technique [40] to represent each data point (compound) in a two-dimensional space based on their similarities. Then, we grouped these data points using the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm [41] to find cohesive clusters of similar compounds. Each of these analyses is discussed in Section 3.

3 Results and Discussion

3.1 Predictive Accuracy Results

The goal of our study was to train Random Forest classifiers on the murine lifespan data from DrugAge so that they can be employed to classify compounds previously not tested in mice into potentially lifespan-extending or not. To this aim, we generated male-only, female-only and mixed-sex datasets, as there are substantial sex-related differences in lifespan responses to various compounds.

The average AUC and average GMean results from 10 runs of 5-fold cross-validation experiments for mixed-sex datasets are presented in Tables 1 and 2, and the results for male-only datasets are presented in Tables 3 and 4. Tables 2 and 4 have per-category results for each type of Targets dataset (Functional Annotation, FA, or Neighbour Enrichment, NE, as defined in Section 2.2.1).

Note that we also ran experiments with female-only datasets, but the models resulting from these experiments were not valid (i.e., they were overfitted to the majority class, and thus unable to make reliable predictions for minority-class compounds), thus we are not reporting these results here. We believe that we did not have sufficient female-only instances of the positive class to train models that are able to make meaningful predictions for positive-class compounds – notable only around 20% of the female instances belonged to the positive class, whilst around 40% of the male instances belonged to the positive class.

In order to select models to be analysed in depth in this Section, we first set a minimum threshold of GMean, as this metric is a good way to determine whether a model has been overfitted towards the majority-class. Using a secondary criterion to disregard overfitted results was required because some models achieved relatively high AUC values (e.g., 0.692 for FA Biological Process) but tended to skew predictions towards the majority (negative) class, which made their AUC results overoptimistic, and this is reflected in their low GMean results (in this case, 0.317).

For the mixed-sex datasets, which had models with lower predictive accuracy overall, the minimum GMean threshold was set to 0.4. For the male-only datasets we set it to 0.5, as these models performed better overall. Then, we selected the best 5 models that passed the GMean threshold from each group of experiments (mixed-sex datasets and male-only datasets), based on their AUC results. The 10 selected models have their results highlighted in boldface in Tables 1-4.

Table 1 – Average AUC and GMean values from models trained on each full mixed-sex dataset

Dataset (mixed sex)	AUC	GMean
Targets FA (Functional Annotation STRING descriptors)	0.678	0.327
Targets NE (Neighbour Enrichment STRING descriptors)	0.636	0.392
LINCS gene expression (all genes)	0.564	0.135
LINCS gene expression (L_1000 landmark genes)	0.634	0.161
PubChem Molecular Fingerprints	**0.637**	**0.472**

Table 2 – Average AUC and GMean values from models trained on each individual feature category of Targets mixed-sex datasets

Dataset (mixed-sex)	AUC (Targets FA)	AUC (Targets NE)	GMean (Targets FA)	GMean (Targets NE)
GO Terms – Biological Process	0.692	0.621	0.317	0.37
GO Terms – Molecular Function	0.6	0.618	0.334	0.338
GO Terms – Cellular Component	0.629	**0.653**	0.364	**0.474**
Uniprot Keywords	0.614	0.598	0.306	0.37
InterPro Protein Domains	0.589	0.555	0.483	0.468
Wiki Pathways	0.662	**0.653**	0.364	**0.446**
Reactome Pathways	0.601	**0.648**	0.354	**0.522**
KEGG Pathways	N /A	**0.629**	N /A	**0.451**

Table 3 – Average AUC and GMean values from models trained on each full male-only dataset

Dataset (male-only)	AUC	GMean
Targets FA (Functional Annotation STRING descriptors)	0.626	0.472
Targets NE (Neighbour Enrichment STRING descriptors)	**0.63**	**0.6**
LINCS gene expression (all genes)	0.557	0.469
LINCS gene expression (L_1000 landmark genes)	0.568	0.518
PubChem Molecular Fingerprints	0.636	0.497

Table 4 – Average AUC and GMean values from models trained on each individual feature category of male-only Targets datasets

Dataset (male-only)	AUC (Targets FA)	AUC (Targets NE)	GMean (Targets FA)	GMean (Targets NE)
GO Terms – Biological Process	0.657	**0.634**	0.498	**0.594**
GO Terms – Molecular Function	0.548	0.636	0.471	0.498
GO Terms – Cellular Component	0.571	**0.657**	0.465	**0.528**
Uniprot Keywords	0.578	0.597	0.43	0.475
InterPro Protein Domains	**0.663**	0.559	**0.534**	0.488
Wiki Pathways	0.565	0.617	0.38	0.55
Reactome Pathways	0.548	0.627	0.43	0.577
KEGG Pathways	N /A	**0.647**	N /A	**0.615**

Overall, the top models have similar average AUC (between 0.629 and 0.663) values. However, their GMean results show that the male-only models are stronger than the mixed-sex models, likely because the female mice instances are harder to classify correctly due to the lack of female positive-class instances, making the problem more difficult in mixed-sex data (as mentioned earlier, ~20% of the female-mice instances in our data are positive, compared to ~40% of the male-mice instances). The similar AUC values between models that have such different GMeans can be explained by a tendency of AUC to reward correct classifications of the majority class, so conservative models that tend to classify instances as negative (a safer classification, as negative instances represent a larger proportion of the data) are achieving good AUC results. An example of this issue is the GO Terms – Biological Process FA model in the mixed-sex data, which had the highest AUC value overall (0.692)

but a very low GMean (0.317). Such models overfit to the majority class, and this is reflected by metrics such as the GMean, which is why we consider both AUC and Gmean when evaluating our models.

The LINCS datasets did not yield good models in any of the experiments, which we believe is due to the higher complexity of the data (numerical values of gene expression rather than binary values), which makes the small number of available examples take a more significant toll on the classifier's performance. Another explanation could be that gene expression averaged across many cell lines, dosages and timepoints is a poor representation of longevity effects of tested compounds.

In the mixed-sex datasets, the Molecular Fingerprints model (based on chemical substructures in the compounds' compositions) was selected as one of the 5 best models, whilst all other selected models belong to the Targets category (based on STRING data of protein target annotations).

In the Targets datasets, the NE mode of dataset creation, which includes annotations (GO Terms, protein domains, pathways and keywords) common to each target's neighbourhood, achieved better results than the FA mode, which includes only the annotations from the input targets themselves. 8 out of 9 top models in Targets category used the NE mode, with the exception of the FA InterPro dataset in male-only model. The NE features refer to annotations that are common to a group of proteins, which makes them more selective compared to FA annotations. This may be the reason for this disparity between the success of NE and FA models. It is also worth noting that the selected male-only models included the NE dataset combining all 8 feature categories (GO annotations, biological pathways, protein domains and UniProt keywords), with a relatively high average GMean of 0.6, the second best GMean result over all experiments. This shows that even feature categories that did not generate a top model on their own had value when considered in combination with other categories. Moreover, GO Function and Uniprot Keywords have not been selected for any of the top models, likely indicating their low informativeness for the task of predicting longevity drugs.

3.2 Analysis of Feature Importance in the Best Predictive Models

Tables 5-14 show, for each of our 10 selected models, which 10 features had the highest importance when labelling a compound as positive (predicted to increase lifespan in mice) or negative class.

Notably, for all mixed-sex dataset models, sex was the most important feature, likely because the compounds in our dataset had negative results when tested in female mice twice more often than in male mice, thus the sex variable is highly predictive.

In all tables we included a Class Tendency column indicating whether features clearly increased the chance of a certain class label. The class tendency of each binary feature F was calculated using two metrics we called Effectiveness (ratio of instances of the positive class (C=1) among instances that have the feature present (F=1)) and Dispensability (ratio of instances of the positive class among instances that have the feature absent (F=0)). The result of the division between those metrics indicates an increase/decrease of likelihood of a positive-class classification when using the feature's value to split the data. We set cut-off values to determine whether the feature's value results in a tendency towards each class, as shown in Equation 1.

$$Effectiveness(F) = \frac{\#(C = 1|F = 1)}{\#(F = 1)}$$

$$Dispensability(F) = \frac{\#(C = 1|F = 0)}{\#(F = 0)}$$

$$Tendency(F) = \frac{Effectiveness}{Dispensability} \begin{cases} \leq 0.66 \rightarrow Neg.\,Class\,Tendency \\ \geq 1.5 \rightarrow Pos.\,Class\,Tendency \end{cases}$$

Equation 1: Calculation of class tendency for each feature F in a dataset

As an example of this tendency calculation, consider the 'GO:0005834' feature in the NE Gene Ontology Components model (third-ranked feature in Table 5): out of the 32 instances with value '1' for this feature, 18 are of the positive class (Effectiveness = 18/32 = 56.25%); out of the 102 instances with value '0' for this feature, 24 are of the positive class (Dispensability = 24/102 = 23.53%). Therefore, the class tendency of 'GO:0005834' is 2.39 (56.25/23.53), which indicates a positive-class tendency for this feature.

3.2.1 Selected models from mixed-sex datasets

Table 5 – Most relevant features in the NE Gene Ontology Components model

Feature	Description	Class Tendency
sex	Sex of the mice used in the experiment	F: Negative class
GO:0016323	Basolateral plasma membrane	Unclear
GO:0005834	Heterotrimeric G-protein complex	Positive class
GO:0005886	Plasma membrane	Positive class
GO:1904813	ficolin-1-rich granule lumen	Negative class
GO:0005829	Cytosol	Negative class
GO:0005759	Mitochondrial matrix	Negative class
GO:0000785	Chromatin	Negative class
GO:0120025	Plasma membrane bounded cell projection	Positive class
GO:0005667	Transcription regulator complex	Negative class

From Table 5 we can see that compounds which target proteins located at the plasma membrane or, more specifically, plasma membrane-bounded cell projections, have a high chance of extending murine lifespan. One important group of such membrane-localised proteins is heterotrimeric G-protein complexes, which are essential signalling molecules functioning as molecular switches to transmit signals from cell surface receptors (which are thus called G protein-coupled receptors) to various intracellular effectors. They are involved in numerous physiological processes, including sensory perception, immune responses, and regulation of mood and metabolism. Interestingly, *GNAQ, GNA11* and *GNAS* - genes that encode different alpha subunits of heterotrimeric G-protein complexes – have recently been predicted to be some of the strongest cancer drivers [42], so inhibiting them with chemical compounds might extend murine lifespan by delaying cancer development. On the other hand, compounds which target proteins located in the cytosol, mitochondrial matrix, chromatin, transcription regulator complex or ficolin-1-rich granule lumen have much lower chances of extending murine lifespan. Ficolin-1 is a crucial protein in the innate immune system, primarily stored in granules within neutrophils. These granules release ficolin-1 into the extracellular environment in response to stimuli, where it binds to carbohydrate structures on pathogens,

apoptotic cells, and other particles, thereby activating the lectin pathway of complement activation. Please note that as only top 10 features are listed for each model, the lists of features discussed here and below are not comprehensive or exhaustive. The female (F) value of the feature Sex is also associated with the negative class, as in the other tables in this Section.

Table 6 – Most relevant features in the NE Wiki Pathways model

Feature	Description	Class Tendency
sex	Sex of the mice used in the experiment	F: Negative class
WP536	Calcium regulation in cardiac cells	Positive class
WP3929	Chemokine signalling pathway	Positive class
WP4583	Biomarkers for urea cycle disorders	Negative class
WP399	Wnt signalling pathway and pluripotency	Negative class
WP5046	NAD metabolism in oncogene-induced senescence and mitochondrial dysfunction-associated senescence	Negative class
WP3594	Circadian rhythm genes	Negative class
WP4313	Ferroptosis	Negative class
WP4788	Autosomal recessive osteopetrosis pathways	Unclear
WP5200	Dravet syndrome	Positive class

Table 6 demonstrates that compounds interacting with proteins involved in calcium regulation in cardiac cells, a chemokine signalling pathway or Dravet syndrome (caused by a loss of function of the voltage-gated sodium channel Nav1.1, affecting the excitability of neurons, particularly inhibitory interneurons) are likely to extend murine lifespan. On the other hand, compounds affecting urea cycle disorders, Wnt signalling pathway and pluripotency, NAD metabolism in oncogene-induced senescence and mitochondrial dysfunction-associated senescence, circadian rhythms, or ferroptosis, are less likely to do so.

Table 7 – Most relevant features in the NE Reactome Pathways model

Feature	Description	Class Tendency
Sex	Sex of the mice used in the experiment	F: Negative class
HSA-9634597	GPER1 signalling	Positive class
HSA-420092	Glucagon-type ligand receptors	Positive class
HSA-1430728	Metabolism	Unclear
HSA-381676	Glucagon-like Peptide-1 (GLP1) regulates insulin secretion	Positive class
HSA-163359	Glucagon signalling in metabolic regulation	Positive class
HSA-5576891	Cardiac conduction	Positive class
HSA-9009391	Extra-nuclear oestrogen signalling	Positive class
HSA-1296071	Potassium Channels	Positive class
HSA-3247509	Chromatin modifying enzymes	Negative class

Table 7 shows that compounds affecting extra-nuclear oestrogen signalling and G protein-coupled estrogen receptor 1 (GPER1) in particular, glucagon signalling in metabolic regulation, glucagon-type ligand receptors and glucagon-like Peptide-1 (GLP1)-mediated regulation of insulin secretion, as well as cardiac conduction and potassium channels, are predicted to extend lifespan with high

probability. Note that both GPER1 and GLP1R are G protein-coupled receptors, making the results of this model in agreement with the results of NE Gene Ontology Components model (Table 5) which selected heterotrimeric G-protein complexes as one of its top features. The "Cardiac conduction" feature is consistent with "Calcium regulation in cardiac cells" feature from the NE Wiki Pathways model (Table 6). On the other hand, compounds interacting with chromatin-modifying enzymes are less likely to extend murine lifespan according to the predictions. Notably, this is also consistent with the NE Gene Ontology Components model which labelled "chromatin" and "transcription regulator complex" as negative class features (Table 5).

Table 8 – Most relevant features in the NE KEGG Pathways model

Feature	Description	Class Tendency
Sex	Sex of the mice used in the experiment	F: Negative class
hsa04929	GnRH secretion	Positive class
hsa05152	Tuberculosis	Negative class
hsa05202	Transcriptional misregulation in cancer	Negative class
hsa04062	Chemokine signalling pathway	Positive class
hsa05017	Spinocerebellar ataxia	Positive class
hsa04022	cGMP-PKG signalling pathway	Positive class
hsa01200	Carbon metabolism	Negative class
hsa05134	Legionellosis	Negative class
hsa05146	Amoebiasis	Unclear

The results in Table 8 suggest that compounds interacting with proteins involved in gonadotropin-releasing hormone (GnRH) secretion, chemokine signalling pathway, cGMP-PKG signalling pathway or spinocerebellar ataxia are promising candidate lifespan-extending compounds. On the other hand, compounds targeting proteins involved in tuberculosis, transcriptional misregulation in cancer, carbon metabolism or legionellosis are unlikely to be effective for lifespan extension. Please note that chemokine signalling pathway has already been selected as a top feature in the NE Wiki Pathways model (Table 6)

Table 9 – Most relevant features in the Molecular Fingerprints model

Feature	Description	Class Tendency
Sex	Sex of the mice used in the experiment	F: Negative class
≥ 8 H	8 or more hydrogen atoms	Negative class
≥ 1 N	At least one nitrogen atom	Positive class
N(~C)(~C)(~H)	At least one nitrogen atom with two carbon atoms and one hydrogen atom as nearest neighbours, regardless of bond order	Positive class
N-H	At least one bonded pair of nitrogen and hydrogen atoms	Positive class
C-C-N-C-C	At least one simple SMARTS pattern: C-C-N-C-C	Positive class
O=C-C=C-[#1]	At least one simple SMARTS pattern: O=C-C=C-[#1]	Negative class
O=C-C-C	At least one simple SMARTS pattern: O=C-C-C	Negative class
C=O	At least one pair of carbon and oxygen atoms with a double bond	Negative class
C-C-C-O-[#1]	At least one simple SMARTS pattern: C-C-C-O-[#1]	Negative class

Table 9 demonstrates that compounds are more likely to extend lifespan if they have less than 8 hydrogen atoms, at least one nitrogen atom, preferably with two carbon atoms and one hydrogen atom as nearest neighbours, or even better as a C-C-N-C-C pattern. On the other hand, pairs of carbon and oxygen atoms with a double bond, as well as O=C-C=C-[#1], O=C-C-C and C-C-C-O-[#1] patterns, should be avoided. According to these criteria, amines (methylamine, dimethylamine), azoles (pyrazole, imidazole), pyridines, pyrimidines and aminopyridines are promising. However, many of these compounds are known to be neurotoxic, hepatotoxic and/or carcinogenic. Imidazole derivatives, pyrimidines and aminopyridines are less toxic and should be investigated as lifespan extending compounds.

3.2.2 Selected models from male-only datasets

Table 10 – Most relevant features in the FA InterPro Domains model

Feature	Description	Class Tendency
IPR017452	GPCR, rhodopsin-like, 7TM	Positive class
IPR000276	G protein-coupled receptor, rhodopsin-like	Positive class
IPR036291	NAD(P)-binding domain superfamily	Negative class
IPR003193	ADP-ribosyl cyclase (CD38/157)	Unclear
IPR002233	Adrenoceptor family	Positive class
IPR016024	Armadillo-type fold	Positive class
IPR014729	Rossmann-like alpha/beta/alpha sandwich fold	Negative class
IPR029071	Ubiquitin-like domain superfamily	Positive class
IPR018490	Cyclic nucleotide-binding domain superfamily	Positive class
IPR000595	Cyclic nucleotide-binding domain	Positive class

Table 10 demonstrates that compounds interacting with rhodopsin-like G protein-coupled receptors (GPCRs), including adrenoceptors, as well as proteins containing Armadillo-type folds, ubiquitin-like domains, or cyclic nucleotide-binding domains, are promising for lifespan extension. GPCRs are essential in signal transduction, influencing pathways related to metabolism, stress response, and hormonal regulation, all of which are vital for maintaining homeostasis. Again, please note that this model is consistent with several models described above in nominating G protein-related features as the top predictive ones. Armadillo-type fold proteins, such as those involved in the Wnt signalling pathway, contribute to cellular proliferation, differentiation, and maintaining tissue homeostasis through their role in cell adhesion and cytoskeletal integrity. This is consistent with the NE Wiki Pathways model, where “Wnt signalling pathway and pluripotency” wad identified as a top predictive feature (Table 6). Ubiquitin-like domains are crucial for proteostasis, aiding in protein quality control and cellular stress responses. Cyclic nucleotide-binding domains, involved in binding cAMP and cGMP, regulate metabolic processes and enhance cellular resistance to oxidative stress. This is consistent with the NE KEGG Pathways model where “cGMP-PKG signalling pathway” was identified as a top predictive feature (Table 8). On the other hand, compounds targeting proteins with NAD(P)-binding domains or Rossmann-like alpha/beta/alpha sandwich folds are less promising for lifespan extension. These domains are primarily involved in redox reactions and basic metabolic processes, which, while essential for cellular function, do not seem to directly influence the regulatory pathways specifically linked to longevity.

Table 11 – Most relevant features in the NE Gene Ontology Process model

Feature	Description	Class Tendency
GO:0003008	System process	Positive class
GO:0007191	Adenylate cyclase-activating dopamine receptor signalling pathway	Positive class
GO:0042391	Regulation of membrane potential	Positive class
GO:0019752	Carboxylic acid metabolic process	Negative class
GO:0007189	Adenylate cyclase-activating G protein-coupled receptor signalling pathway	Positive class
GO:0007186	G protein-coupled receptor signalling pathway	Positive class
GO:0001503	Ossification	Positive class
GO:0032879	Regulation of localization	Positive class
GO:0043648	Dicarboxylic acid metabolic process	Negative class
GO:0007212	Dopamine receptor signalling pathway	Positive class

The top feature in Table 11 is the very broad “System process” GO term, which indicates that compounds which target multicellular organismal processes carried out by organs or tissues within an organ system, such as immune, circulatory, nervous, etc., have the most potential to extend lifespan. Specific processes worth targeting for lifespan extension include G protein-coupled receptor signalling pathways, specifically adenylate cyclase-activating ones, in particular the dopamine receptor pathway, as well as regulation of membrane potential, ossification and “regulation of localization”. The “Regulation of membrane potential” feature is consistent with the “Plasma membrane” feature from the NE Gene Ontology Components models (Table 5, Table 12) and the “Potassium channels” feature from the NE Reactome Pathways model (Table 7). “Regulation of localization” is a GO term that describes “any process that modulates the frequency, rate or extent of any process in which a cell, a substance, or a cellular entity is transported to, or maintained in, a specific location”. On the other hand, targeting carboxylic and dicarboxylic acid metabolism is unlikely to result in lifespan extension. These chemicals are important intermediates in various metabolic pathways, including the Krebs cycle, which is essential for ATP production, cellular respiration and overall metabolic health. This is consistent with NE Gene Ontology Components models (Table 5, Table 12), where the mitochondrial matrix (where the Krebs cycle occurs) was identified as a negative class feature.

Table 12 – Most relevant features in the NE Gene Ontology Components model

Feature	Description	Class Tendency
GO:0005834	Heterotrimeric G-protein complex	Positive class
GO:0005886	Plasma membrane	Positive class
GO:0045177	Apical part of cell	Positive class
GO:0005829	Cytosol	Negative class
GO:0016323	Basolateral plasma membrane	Unclear
GO:0005739	Mitochondrion	Negative class
GO:0005667	Transcription regulator complex	Negative class
GO:1904813	ficolin-1-rich granule lumen	Negative class
GO:0005759	Mitochondrial matrix	Negative class
GO:0090575	RNA polymerase II transcription regulator complex	Unclear

Table 12 shows results from the NE Gene Ontology Components model applied to a male-only dataset and is highly similar to Table 5 which shows results from the same model but applied to a mixed-sex dataset. According to these results, compounds interacting with proteins localised at the plasma membrane and/or apical part of the cell, especially the proteins comprising the heterotrimeric G-protein complexes, are likely to extend lifespan. Conversely, compounds interacting with proteins localised in the cytosol, mitochondrion, especially mitochondrial matrix, ficolin-1-rich granule lumen, or transcription regulator complex are unlikely to do so.

Table 13 – Most relevant features in the NE KEGG Pathways model

Feature	Description	Class Tendency
hsa04929	GnRH secretion	Positive class
hsa04911	Insulin secretion	Positive class
hsa04022	cGMP-PKG signalling pathway	Positive class
hsa05146	Amoebiasis	Unclear
hsa05152	Tuberculosis	Negative class
hsa05032	Morphine addiction	Positive class
hsa04742	Taste transduction	Positive class
hsa04914	Progesterone-mediated oocyte maturation	Positive class
hsa04713	Circadian entrainment	Positive class
hsa04640	Hematopoietic cell lineage	Negative class

Table 13 shows results from the NE KEGG Pathways model applied to a male-only dataset and is similar to Table 8 which shows results from the same model but applied to a mixed-sex dataset. According to these results, compounds interacting with proteins involved in gonadotropin-releasing hormone (GnRH) secretion, progesterone-mediated oocyte maturation, insulin secretion, cGMP-PKG signalling pathway, taste transduction, circadian entrainment and morphine addiction are promising candidate lifespan-extending compounds. It is peculiar that the NE Wiki Pathways model classified "Circadian rhythm genes" as a negative class feature (Table 6), which seems to be at odds with the current model (Table 13) classifying "Circadian entrainment" as a positive class feature. Morphine addiction was selected likely because it is mediated by dopamine, and thus this result is in agreement with the "Dopamine receptor signalling pathway" feature from the NE Gene Ontology Process model (Table 11). Interestingly, dopamine can also mediate addiction to food, while taste transduction can affect the palatability of food, and circadian entrainment can affect feeding times. Altogether, this triad of positive class features may indicate that some longevity drugs affect the feeding behaviour of mice and thus cause indirect (behavioural) caloric restriction, which is known to strongly extend lifespan. For example, it has been shown that dietary restriction response and lifespan extension can be achieved through the pharmacological masking of a sensory pathway that signals the presence of food [43]. On the other hand, compounds targeting proteins involved in tuberculosis and hematopoietic cell lineage are unlikely to be effective for lifespan extension.

Table 14 – Most relevant features in the NE All categories model

Feature	Category	Description	Class Tendency
HSA-381676	Reactome	Glucagon-like Peptide-1 (GLP1) regulates insulin secretion	Positive class
HSA-163359	Reactome	Glucagon signalling in metabolic regulation	Positive class
HSA-420092	Reactome	Glucagon-type ligand receptors	Positive class
GO:0003008	GO Process	System process	Positive class
HSA-9634597	Reactome	GPER1 signalling	Positive class
hsa04929	KEGG	GnRH secretion	Positive class
hsa04911	KEGG	Insulin secretion	Positive class
KW-0496	UniProt Keyword	Mitochondrion	Negative class
GO:0005834	GO Component	Heterotrimeric G-protein complex	Positive class
HSA-9660821	Reactome	ADORA2B mediated anti-inflammatory cytokines production	Positive class

Finally, the NE All categories model (Table 14) was able to select features from all categories, and thus most of them have already appeared as top features in single-category models. We can see that these features are mostly related to systemic processes, such as hormone secretion and signalling mediated via G-protein coupled receptors. The top hormones involved are Glucagon-like Peptide-1 (GLP1), glucagon, insulin, oestrogen (signalling via G protein-coupled estrogen receptor 1 (GPER1)), gonadotropin-releasing hormone (GnRH) and adenosine (signalling via adenosine A2B receptor (ADORA2B)). Notably, all of them signal via G-protein coupled receptors. These results lend support to the hormonal theory of ageing. As three of these hormones are involved in glucose metabolism, this is consistent with the occurrence and detrimental effects of diabetes and metabolic syndrome with age, as well as the effectiveness of caloric restriction and antidiabetic drugs such as acarbose and possibly metformin. Two other hormones are sex hormones, which aligns with sex-specific differences observed in human and animal lifespan and in the effectiveness of lifespan-extending compounds. The natural ligand for adenosine receptor A2B (ADORA2B) is extracellular adenosine, formed from the reduction of ATP by ENTDPases. ATP enters the extracellular space in response to tissue injury and apoptosis amongst other stress factors and has chemotactic and excitatory effects. The reduction of ATP to adenosine is thought to be a regulatory mechanism by which the synthesis of anti-inflammatory cytokines is induced [44]. Thus, the "ADORA2B mediated anti-inflammatory cytokines production" feature, together with the "Chemokine signalling pathway" feature from NE Wiki Pathways (Table 6) and NE KEGG Pathways (Table 8) models, lend support to the inflammageing theory and highlights the importance of reducing chronic inflammation. Although the UniProt Keywords model did not perform as well as other models that we selected, the identification of the UniProt Keyword "Mitochondrion" feature as a negative class one in the current combined model (Table 14) is consistent with the results of the NE Gene Ontology Components models (Table 5, Table 12), which also nominated mitochondrion and mitochondrial matrix as negative class features. These results suggest a critical look is required at the mitochondrial theory of ageing.

3.3 Identifying the Most Promising Novel Compounds for Lifespan Extension

After identifying the models with the best predictive accuracy in the mixed-sex and male-only datasets in Section 3.1, in this Section we use their predictions to highlight compounds that show some promise of lifespan-extension capabilities. The predictions of our classification models are

based on the chemical composition of a compound in the case of the Molecular Fingerprints model, and on the STRING annotations of target proteins for all other selected models.

A classifier predicts an instance as part of the positive class based on feature patterns that indicate its similarity to training instances of that class, so the compounds identified in this analysis have some commonalities with previously successful compounds. Naturally, this does not guarantee that the compounds would work similarly. For example, they may interact differently with the target (e.g., activate it whilst the successful instances inhibit it), and this would not be reflected in our binary-feature datasets.

3.3.1 Identifying recurrent false-positive classifications in top models

In order to identify novel compounds that may extend murine lifespan, we first considered the false-positive classifications in our own models' predictions, as these indicate that the model identified positive-class patterns in the compound's data. Table 15 shows six compounds that were frequently classified as longevity-related but are part of the negative class in our datasets (i.e., there is currently no sufficient evidence to claim that these compounds extend the median lifespan of mice). It includes the PubChem ID of each compound, and the ratio of times the compound was classified as a false-positive over the number of top-model datasets it was included in (Putrescine and Chlorpheniramine are included only in mixed-sex datasets, as they only have examples from female mice studies).

These compounds have been reported as unsuccessful for lifespan extension in mice (see our positive-class definition in Section 2.1), but they are consistently being labelled as positive by our most accurate models, which shows they have some similarity with positive-class examples and may warrant further investigation. Note that although the negative-class compounds have failed to extend lifespan in the existing studies, it is possible that a different treatment regimen (e.g. a different dosage, route of administration, starting age, duration) or a different mouse strain could lead to positive results. Interestingly, LY444711 slightly improves median survival in mice, although the increase does not reach statistical significance. What is remarkable is that it significantly improves maximal lifespan [45], although we do not use that parameter for training the models. Another drug in our false-positive list, dehydroepiandrosterone sulphate, has been shown to be ineffective in two mouse studies, but its levels are dramatically reduced with age in humans [46], perhaps it is one of the strongest blood biomarkers of ageing), so it is still possible that the right concentration or treatment regimen has not been achieved, given that this is a hormone with complex and precisely regulated secretion patterns [47].

Table 15 – Negative-class compounds consistently classified as false-positive by the selected models

Compound	PubChem ID	False-positive Frequency
LY444711 (ghrelin agonist)	91229598	10/10 (100%)
Putrescine	1045	5/5 (100%)
Chlorpheniramine	2725	4/5 (80%)
Dehydroepiandrosterone sulphate	5881	7/10 (70%)
2-Mercaptoethylamine Hydrochloride	6058	7/10 (70%)
Taxifolin	439533	6/10 (60%)

3.3.2 Labelling unseen data from an external dataset

For this analysis, we created two classifier ensembles using the selected (most accurate) classification models from each type of dataset: 5 from mixed-sex datasets and 5 from male-only datasets. Each model outputs a positive-class prediction probability between 0 and 1, and the ensemble's final probability is the average over all valid models' probabilities (i.e., in cases where a model has no data for a compound, it is not included in the ensemble's result calculation).

We used these ensembles to classify a large number of existing drugs and compounds that have never been tested for their lifespan extension effects in mice. We downloaded the full DrugBank [19] dataset of Drug Target Identifiers (version 5.1.12, released on 2024-03-14) and used it to create feature datasets from STRING annotations and enrichments for these targets. After applying our ensembles to these datasets, we selected all novel compounds that were predicted by the ensembles to have at least 75% positive-class probability, which resulted in a list of 57 compounds from the mixed-sex ensemble and 272 compounds from the male-only ensemble. The male-only ensemble predicted more high-confidence compounds likely because those models had higher discriminatory power (Gmean). Some compounds confidently classified as belonging to the positive class by these ensembles are very similar to previously successful compounds (i.e., sharing targets with positive-class instances in the training data), whereas other compounds might represent previously unexplored research directions for future longevity studies in mice. Compounds with 100% feature similarity to the ones in the training dataset were removed.

As mentioned in Section 2.3, we used the UMAP dimensionality reduction technique to visualise the selected compounds in a two-dimensional space based on measuring Jaccard distances (a similarity metric for binary data that disregards '0' matches) between the compounds. We adjusted the UMAP parameters to emphasise the global structure as opposed to local structure (i.e. set neg_sample_rate and n_neighhbours equal to the number of compounds minus one). Then, we used the DBSCAN algorithm to identify clusters of compounds, based on their Euclidean distance within the UMAP projection. We fine-tuned the DBSCAN eps parameter to keep the lowest possible number of groupings while preserving the distinct sets of top targets in each cluster, resulting in 4 clusters out of the 57 compounds nominated by the mixed-sex ensemble and 7 clusters out of the 272 compounds selected by the male-only ensemble. Figure 1 shows the UMAP projections of the most confident positive-class predictions of the mixed-sex and male-only ensembles, as well as the positive class likelihoods of these compounds and their clusters as determined by DBSCAN.

3.3.2.1. Mixed-sex ensemble predictions

Table 16 shows the clusters of compounds with positive-class likelihood ≥75% as estimated by the ensemble created from mixed-sex dataset models, and each cluster's most frequent protein targets are shown in Table 17.

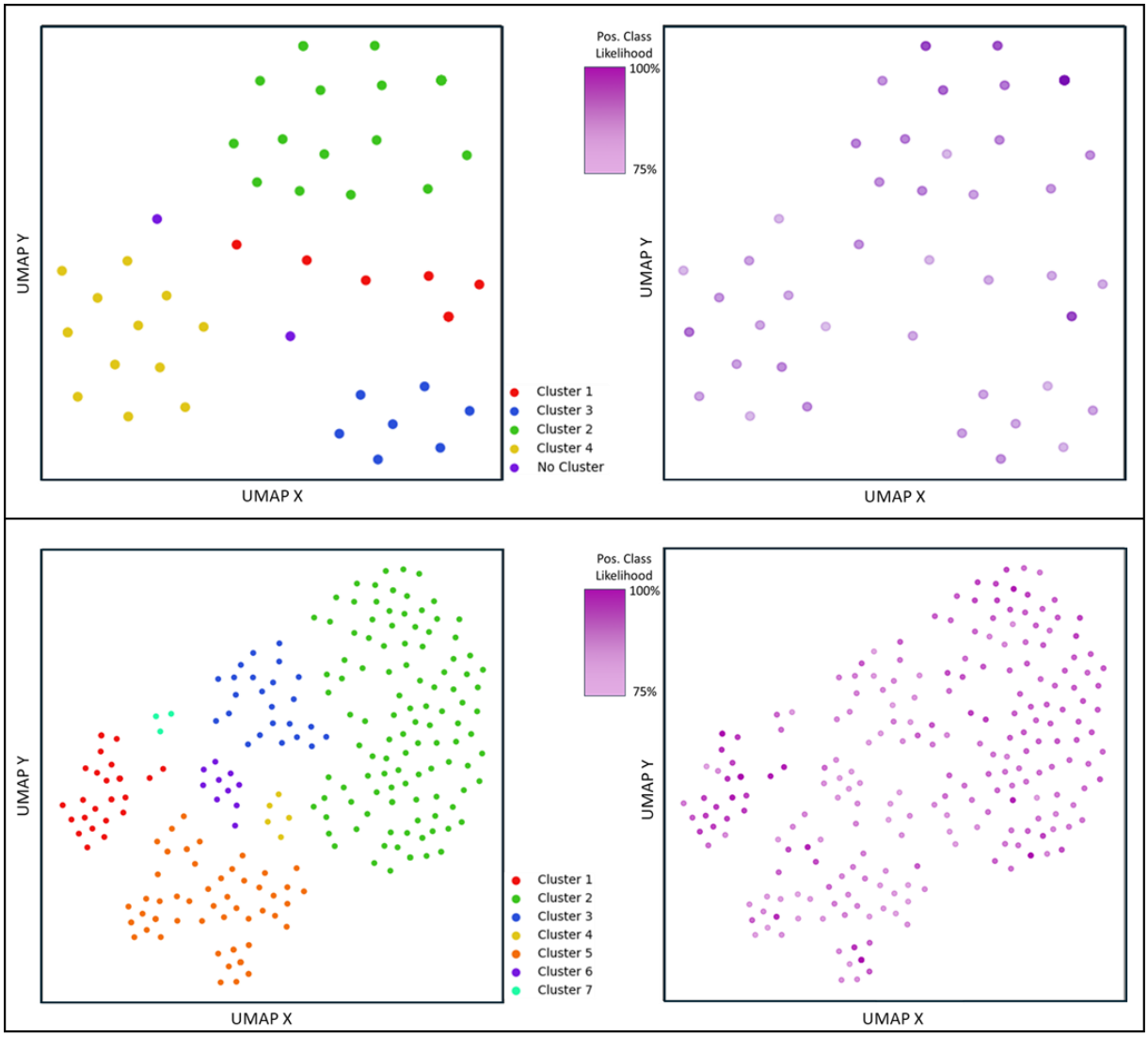

Figure 1 –Clustering (left) and positive-class likelihood (right) of DrugBank compounds with positive-class likelihood ≥75% predicted by the mixed-sex ensemble (top row) or by the male-only ensemble (bottom row)

Table 16 – Clusters of DrugBank compounds with positive-class likelihood >75% as estimated by the ensemble of mixed-sex datasets models. Potential lifespan-extending compounds are highlighted in bold.

Cluster	Number of compounds	Positive class likelihood Min-Max (Mean)	Compounds
1	8	76.4% to 81.5% (mean: 78.3%)	**BMS-754807**, Insulin peglispro, Linsitinib, Mecasermin rinfabate, **N-[2-(2-iodo-5-methoxy-1H-indol-3-yl)ethyl]acetamide**, Primaquine, Somatrem, **XL765**
2	27	75.1% to 90% (mean: 81.8%)	**Acebutolol, Alprenolol, Atenolol, Befunolol, Betaxolol, Bethanidine, Bevantolol, Bisoprolol, Carteolol, Celiprolol, Cryptenamine**, DL-Methylephedrine, Isoetharine, Isoprenaline, **Levobunolol**, Mephentermine, **Metipranolol, Nadolol, Penbutolol,** Phenylpropanolamine, **Pindolol,** Pirbuterol, **Propafenone, Propranolol,** Racepinephrine, **Sotalol, Timolol**
3	7	76% to 81.1% (mean: 78.5%)	**4-Methylimidazole, Bendroflumethiazide, Brinzolamide, Chlorothiazide, Dorzolamide, Methyclothiazide, n-{2-[4-(aminosulfonyl)phenyl]ethyl}acetamide**
4	13	75.2% to 82. 1% (mean: 78.6%)	JNJ-37822681, Cinnarizine, **Dihydro-alpha-ergocryptine**, Domperidone, Norclozapine, **Piribedil, Quinagolide**, Rolicyclidine, **Sarizotan**, Sulpiride, **Sumanirole**, Tetrabenazine, Tetrahydropalmatine
No Cluster	2	0.762 to 0.782 (mean: 0.772)	Alpha-Benzyl-Aminobenzyl-Phosphonic Acid, LI-301

Table 17 – Frequent (≥ 33%) targets of clusters of DrugBank compounds with positive-class likelihood ≥ 75% as estimated by the ensemble of mixed-sex datasets models

Cluster	Target frequency	Target gene name	Target full name
1	63%	IGF1R	Insulin-like growth factor 1 receptor
	50%	INSR	Insulin receptor
2	100%	ADRB1	Beta-1 adrenergic receptor
	96%	ADRB2	Beta-2 adrenergic receptor
	37%	ADRB3	Beta-3 adrenergic receptor
3	100%	CAH2	Carbonic anhydrase 2
	86%	CAH1	Carbonic anhydrase 1
	57%	CAH4	Carbonic anhydrase 4
4	100%	DRD2	D(2) dopamine receptor
	54%	DRD3	D(3) dopamine receptor

Cluster 1 is a diverse group that includes compounds targeting primarily Insulin-like growth factor 1 receptor and Insulin receptor. Some of them are activators (agonists) of these receptors (Insulin peglispro, Mecasermin rinfabate), while others are inhibitors (antagonists) (BMS-754807, Linsitinib). There are also compounds activating Melatonin receptors (N-[2-(2-iodo-5-methoxy-1H-indol-3-yl)ethyl]acetamide), activating Growth hormone receptor (Somatrem) and inhibiting PI3K/mTOR (XL765). As noted above, our models cannot distinguish between activators and inhibitors of the same target, because this information is absent for most compounds in DrugBank and other target databases which we used for training the models. Thus, while these compounds are predicted to

have strong effects on murine lifespan, the sign of this effect can be either positive or negative, depending on whether the compound affects the target in the same way as the lifespan-extending molecules in the positive training dataset or in the opposite way. In our positive training dataset, we had *inhibitors* of IGF-1 receptor (L2-Cmu [48]), PI3K (Alpelisib [49]), mTOR (Rapamycin [50,51]) and an *activator* of Melatonin receptors (melatonin itself [52,53]). Thus, we can predict that BMS-754807, Linsitinib, XL765 and N-[2-(2-iodo-5-methoxy-1H-indol-3-yl)ethyl]acetamide are likely to increase murine lifespan, whereas Insulin peglispro, Mecasermin rinfabate and Somatrem are likely to shorten it.

Cluster 2 consists of drugs targeting Beta-adrenergic receptors. Most of them are antagonists (drug names ending in -olol, and additionally Bethanidine, Cryptenamine, Propafenone and Sotalol), but some are agonists. Because we had antagonists of Beta-adrenergic receptors (Metoprolol, Nebivolol [54]) in our positive training dataset, we predict that antagonists of these receptors will likely extend murine lifespan, whereas the agonists will likely shorten it. Beta-blockers decrease heart rate, thus somewhat supporting the "fixed number of heartbeats per lifespan" hypothesis [55]. They also decrease cardiac output. Interestingly, in a recent study of caloric restriction in genetically heterogeneous female mice, cardiac output was negatively correlated with lifespan [56]. This is also consistent with top predictive features "Calcium regulation in cardiac cells" (Table 6) and "Cardiac conduction" (Table 7) from mixed-sex models.

Cluster 3 compounds are primarily inhibitors of Carbonic anhydrases. In our positive training dataset, we had Butylated hydroxytoluene [57], which is also an inhibitor of carbonic anhydrases. Thus, we predict that most compounds in this cluster are likely to extend murine lifespan. Interestingly, an increase of tissue-specific carbonic anhydrases in mitochondria from middle-aged mouse brain and skeletal muscle has been documented [58]. Moreover, nematodes *C. elegans* exposed to CAH2 have a dose-related shorter lifespan suggesting that high CAH2 levels are life-limiting [58].

Cluster 4 compounds target mostly Dopamine receptors. Most of them inhibit these receptors, but some compounds activate them (Dihydro-alpha-ergocryptine, Piribedil, Quinagolide, Sarizotan and Sumanirole). In our positive training dataset, we had Levodopa [59] which is a Dopamine receptor agonist precursor. Thus, we predict that compounds which activate Dopamine receptors are likely to extend murine lifespan, whereas inhibitors of these receptors will likely shorten it.

3.3.2.2. Male-only ensemble predictions

Table 18 shows the clusters of compounds with positive-class likelihood > 75% as estimated by the ensemble created from mixed-sex dataset models, and each cluster's most frequent protein targets are shown in Table 19.

Table 18 – Clusters of DrugBank compounds with positive-class likelihood ≥75% nominated by the ensemble of male-only datasets models. Potential lifespan-extending compounds are highlighted in bold.

Cluster	Number of compounds	Positive class likelihood Min-Max (Mean)	Compounds
1	37	75.3% to 97.2% (mean: 90.6%)	2-hydroxymethyl-6-octylsulfanyl-tetrahydro-pyran-3,4,5-triol, **Acebutolol, Alprenolol, Arotinolol, Atenolol, Befunolol, Betaxolol, Bethanidine, Bevantolol, Bisoprolol, Bopindolol, Carteolol, Celiprolol, Cryptenamine,** Dipivefrin, DL-Methylephedrine, Droxidopa, Ephedrine, Isoetharine, Isoprenaline, **Labetalol, Levobunolol,** Mephentermine, **Metipranolol,** Mirabegron, **Nadolol,** Norepinephrine, **Penbutolol,** Phenoxybenzamine, Phenylpropanolamine, **Pindolol,** Pirbuterol, **Propafenone, Propranolol,** Racepinephrine, **Sotalol, Timolol**
2	119	76.2% to 91.2% (mean: 83.9%)	Acepromazine, Alizapride, **Amantadine**, Amisulpride, Amitifadine, Amitriptyline, Amoxapine, Amphetamine, Aniracetam, **Apomorphine**, Aripiprazole, AS-8112, Asenapine, Bicifadine, Bifeprunox, BL-1020, Blonanserin, Brasofensine, Brexpiprazole, **Bromocriptine**, Bromopride, Buspirone, **Cabergoline**, Cariprazine, Carphenazine, Chlorprothixene, Cinnarizine, Clozapine, Desipramine, **Dihydro-alpha-ergocryptine**, Dihydroergocornine, Dihydroergocristine, **Dihydroergotamin**e, Domperidone, Dopamine, Doxepin, Droperidol, Ephedra sinica root, Ergoloid mesylate, Ergotamine, Escitalopram, Etoperidone, Fluoxetine, Flupentixol, Fluspirilene, Haloperidol, Hydrocodone, Iloperidone, Imipramine, JNJ-37822681, Ketamine, Ketobemidone, **Lisuride**, Loxapine, Lumateperone, Lurasidone, Maprotiline, Melperone, Memantine, Meperidine, Mesoridazine, Methadone, Methotrimeprazine, Metoclopramide, Mianserin, Minaprine, Molindone, Naltrexone, Nefazodone, Norclozapine, Nortriptyline, Ocaperidone, Olanzapine, Ondansetron, Orphenadrine, Paliperidone, Pardoprunox, Paroxetine, Pergolide, Perospirone, Pipamperone, Pipotiazine, **Piribedil**, **Pramipexole**, Prochlorperazine, Promazine, Propiomazine, Quetiapine, **Quinagolide**, Raclopride, Remoxipride, Risperidone, Rolicyclidine, **Ropinirole**, **Rotigotine**, Sarizotan, Sertindole, Setiptiline, Sulpiride, **Sumanirole**, Tetrahydrocannabivarin, Tetrahydropalmatine, Thiethylperazine, Thioproperazine, Thioridazine, Thiothixene, Tianeptine, Tramadol, Triflupromazine, Trimipramine, Viloxazine, Vortioxetine, YKP-1358, Yohimbine, Ziprasidone, Zotepine, Zuclopenthixol
3	24	75.1% to 87.2% (mean: 80.1%)	Agmatine, **Azimilide**, **Bepridil**, Cocaine, **Dofetilide**, **Enflurane**, Flunarizine, Gabapentin, **Halothane**, Isavuconazole, **Isoflurane**, Lamotrigine, Loperamide, Nicardipine, Pentoxyverine, Perphenazine, Pimozide, Promethazine, Ranolazine, Ritodrine, **Topiramate**, Trifluoperazine, Trimebutine, Verapamil
4	5	75% to 78.5% (mean: 76.6%)	**4-Methoxyamphetamine**, **Benzatropine**, Benzphetamine, **Diphenylpyraline**, **Metamfetamine**
5	74	75% to 89.3% (mean: 78.3%)	Aclidinium, ALKS 27, Apraclonidine, Aranidipine, Astemizole, Atropine, Benzquinamide, Betahistine, Bethanechol, Brompheniramine, Butriptyline, Captodiame, Cinitapride, Cirazoline, Cisapride, Clonidine, Cyproheptadine, Darifenacin, Disopyramide, Dosulepin, Dotarizine, Doxazosin, Epicept NP-1, Epinastine, Esmirtazapine, Fenfluramine, Fenoldopam, Fesoterodine, Flibanserin, Glycopyrronium, Homatropine, Homatropine methylbromide, HY10275, Hydroxyzine, Indigotindisulfonic acid, Lofexidine, Loratadine, Lorpiprazole, Methacholine, Methantheline, Methyldopa, Metixene, Mirtazapine, Moxisylyte, Naphazoline, OBE101, OPC-28326, Oxymetazoline, Phentolamine, Pilocarpine, Pimavanserin, Pirlindole, Pizotifen, Prazosin, Propiverine, Quinidine, Revefenacin, Rociverine, Scopolamine, Silodosin, Solifenacin, Tegaserod, Terazosin, Terfenadine, Tetryzoline, Thonzylamine, Tiotropium, Tizanidine, Tolazoline, Tolterodine, Trazodone, Trihexyphenidyl, Umeclidinium, Xylometazoline
6	10	75.9% to 83.9% (mean: 78.3%)	Adenosine, Defibrotide, Doxofylline, Dyphylline, Enprofylline, **Etomidate**, Istradefylline, **Midazolam**, Pentoxifylline, Theobromine
7	3	76.4% to 84.9% (mean: 81.7%)	Clenbuterol, Epinephrine, **XL765**

Table 19 – Frequent (≥ 33%) targets of clusters of DrugBank compounds with positive-class likelihood ≥ 75% as estimated by the ensemble of male-only datasets models

Cluster	Target frequency	Target gene name	Target full name
1	97%	ADRB1	Beta-1 adrenergic receptor
	95%	ADRB2	Beta-2 adrenergic receptor
	38%	ADRB3	Beta-3 adrenergic receptor
2	84%	DRD2	D(2) dopamine receptor
	59%	5HT2A	5-hydroxytryptamine receptor 2A
	47%	5HT1A	5-hydroxytryptamine receptor 1A
	45%	ADA1A	Alpha-1A adrenergic receptor
	42%	5HT2C	5-hydroxytryptamine receptor 2C
	42%	DRD1	D(1A) dopamine receptor
	39%	DRD3	D(3) dopamine receptor
	39%	ADA2A	Alpha-2A adrenergic receptor
	38%	ADA1B	Alpha-1B adrenergic receptor
	34%	ADA2C	Alpha-2C adrenergic receptor
	33%	ADA2B	Alpha-2B adrenergic receptor
3	42%	KCNH2	Potassium voltage-gated channel subfamily H member 2
	38%	CAC1C	Voltage-dependent L-type calcium channel subunit alpha-1C
	38%	CALM1	Calmodulin-1
	33%	GBRA1	Gamma-aminobutyric acid receptor subunit alpha-1
	33%	CAC1H	Voltage-dependent T-type calcium channel subunit alpha-1H
4	100%	SC6A3	Sodium-dependent dopamine transporter
	60%	SC6A4	Sodium-dependent serotonin transporter
	60%	VMAT2	Synaptic vesicular amine transporter
	60%	ADA2A	Alpha-2A adrenergic receptor
	40%	HRH1	Histamine H1 receptor
	40%	SC6A2	Sodium-dependent noradrenaline transporter
	40%	ADA1A	Alpha-1A adrenergic receptor
	40%	AOFB	Amine oxidase [flavin-containing] B
	40%	AOFA	Amine oxidase [flavin-containing] A
5	41%	ACM3	Muscarinic acetylcholine receptor M3
	41%	ACM1	Muscarinic acetylcholine receptor M1
	41%	ACM2	Muscarinic acetylcholine receptor M2
	39%	ADA1A	Alpha-1A adrenergic receptor
	36%	ACM4	Muscarinic acetylcholine receptor M4
	36%	ACM5	Muscarinic acetylcholine receptor M5
	35%	ADA2A	Alpha-2A adrenergic receptor
6	90%	AA2AR	Adenosine receptor A2a
	70%	AA1R	Adenosine receptor A1
	40%	GBRA1	Gamma-aminobutyric acid receptor subunit alpha-1
7	67%	ADRB1	Beta-1 adrenergic receptor
	67%	ADRB2	Beta-2 adrenergic receptor
	67%	TNFA	Tumor necrosis factor

Like Cluster 2 in the mixed-sex classifier results, Cluster 1 in the male-only classifier results consists of drugs targeting Beta-adrenergic receptors. This is consistent with a top predictive feature "Adrenoceptor family" from the male-only FA InterPro Domains model (Table 10). Most of these drugs are antagonists (drug names ending in -olol, and additionally Bethanidine, Cryptenamine, Labetalol, Propafenone and Sotalol), but some are agonists. Because we had antagonists of Beta-

adrenergic receptors (Metoprolol, Nebivolol [54]) in our positive training dataset, we predict that antagonists of these receptors will likely extend murine lifespan, whereas the agonists will likely shorten it.

Like Cluster 4 in the mixed-sex classifier results, Cluster 2 in the male-only classifier results consists of drugs targeting primarily Dopamine receptors; but in addition to that, 5-hydroxytryptamine (Serotonin) receptors and Alpha-adrenergic receptors. However, because we only had a Dopamine receptor agonist precursor (Levodopa [59]) in our positive training dataset, but no interactors of Serotonin or Alfa-adrenergic receptors, it is likely that Dopamine receptors are driving the effects and clustering of this group of compounds. This is also consistent with a top predictive feature "(Adenylate cyclase-activating) dopamine receptor signalling pathway" from the male-only NE Gene Ontology Process model (Table 11). Drugs from this cluster that act as Dopamine receptor agonists and thus likely to extend murine lifespan are: Amantadine, Apomorphine, Bromocriptine, Cabergoline, Dihydro-alpha-ergocryptine, Dihydroergotamine, Lisuride, Piribedil, Pramipexole, Quinagolide, Ropinirole, Rotigotine and Sumanirole.

Compounds in Cluster 3 are targeting voltage-dependent potassium and calcium channels, as well as calmodulin and GABA receptor. Several compounds from our positive training dataset (berberine [60], chloroquine [61], nebivolol [54]) are listed as interactors of voltage-dependent potassium channels in PHAROS, most likely acting as inhibitors. This is consistent with a top predictive feature "Regulation of membrane potential" from the male-only NE Gene Ontology Process model (Table 11). Additionally, another compound from our positive training dataset, melatonin [52,53], is listed as an interactor of calmodulin in DrugBank, and appears to be an inhibitor [62]. Moreover, Taurine [63], another compound from our positive training dataset, is listed in DrugBank as an agonist of GABA receptors. Thus, we predict that compounds inhibiting voltage-dependent potassium channels (Azimilide, Bepridil, Dofetilide) or calmodulin (Bepridil), or activating GABA receptors (Halothane, Isoflurane, Topiramate, Enflurane), are likely to extend murine lifespan.

Cluster 4 compounds target Sodium-dependent dopamine, serotonin and noradrenalin transporters, Synaptic vesicular amine transporter, Alpha-adrenergic receptors, Histamine H1 receptor and Amine oxidases. Interestingly, we only had a Histamine H1 receptor antagonist (Meclizine [64]) and an Amine oxidase B inhibitor (L-deprenyl/Selegiline [65]) in our positive training dataset, but no interactors of Sodium-dependent dopamine, serotonin and noradrenalin transporters or Alpha-adrenergic receptors. This indicates that compounds in this cluster are not very selective and have multiple targets simultaneously. Nevertheless, we expect drugs inhibiting Histamine H1 receptor (Benzatropine, Diphenylpyraline) or Amine oxidase B (4-Methoxyamphetamine, Metamfetamine) to extend murine lifespan unless the other targets of these drugs will lead to lifespan shortening. Interestingly, it has been shown that harmol, which simultaneously modulates Amine oxidase B and GABA-A receptor, induces mitophagy and AMPK pathway activation, and improves glucose tolerance, liver steatosis and insulin sensitivity in pre-diabetic male mice, delays frailty onset, improves glycemia, exercise performance and strength in two-year-old male and female mice, as well as extends the lifespan of *C. elegans* and female *D. melanogaster* [66].

Cluster 5 compounds target predominantly Muscarinic acetylcholine receptors and also Alpha-adrenergic receptors. Surprisingly, we did not have any compounds in our positive training dataset that target any of these receptors. Likely, compounds targeting these receptors were prioritised via feature enrichment, especially the NE method. This highlights the ability of our models to go beyond

reiterating the targets from the training dataset. Potentially, drugs targeting Muscarinic acetylcholine receptors represent an entirely novel class of lifespan-extending compounds that have not been previously tested in mice. Interestingly, the lack of M3 muscarinic acetylcholine receptors greatly ameliorated impairments in glucose homeostasis and insulin sensitivity in various forms of experimentally or genetically induced obesity in mice [67]. Moreover, there is a loss of M1 muscarinic acetylcholine receptors in Alzheimer's disease human brain tissue [68].

Compounds in Cluster 6 target Adenosine receptors and a GABA receptor. As mentioned above, Taurine [63] is the only compound from our positive training dataset that interacts with GABA receptors. There are no compounds in our positive training dataset that interact with Adenosine receptors, so similarly to Muscarinic acetylcholine receptors, they were likely prioritised via feature enrichment. This is consistent with a top predictive feature "ADORA2B mediated anti-inflammatory cytokines production" from the male-only NE All categories model (Table 14). Thus, it could be another novel class of lifespan-extending compounds. Adenosine is an immunosuppressive metabolite produced at high levels within the tumour microenvironment. Importantly, adenosine signalling through the A2a receptor expressed on immune cells potently dampens immune responses in inflamed tissues [69]. Interestingly, compounds targeting Muscarinic acetylcholine receptors, Alpha-adrenergic receptors and Adenosine receptors act as bronchodilators, vasodilators and smooth muscle relaxants. In any case, based on Taurine effects, we predict GABA receptor agonists (Etomidate, Midazolam) to prolong murine lifespan.

Finally, there are only three compounds in Cluster 7: Clenbuterol, Epinephrine and XL765. Clenbuterol and Epinephrine are both agonists of Beta-adrenergic receptors, and because our positive training dataset includes antagonists of Beta-adrenergic receptors (Metoprolol, Nebivolol [54]), Clenbuterol and Epinephrine are more likely to shorten murine lifespan rather than increase it. XL765 has also been selected in mixed-sex Cluster 1 and is a dual PI3K/mTOR inhibitor which is predicted to increase murine lifespan by analogy with Alpelisib [49] and Rapamycin [50,51].

Overall, and surprisingly, most predicted longevity drug targets are related to the nervous system function (e.g. various neurotransmitter receptors and transporters and voltage-gated ion channels). Interestingly, it has been shown that extended longevity in humans is associated with a distinct transcriptome signature in the cerebral cortex that is characterized by downregulation of genes related to neural excitation and synaptic function [70]. Recent studies in model organisms demonstrate that the aging process is frequently modified by an organism's ability to perceive and respond to changes in its environment. Many well-studied pathways that influence aging involve sensory cells, frequently neurons, that signal to peripheral tissues and promote survival during the presence of stress. Importantly, this activation of stress response pathways is often sufficient to improve health and longevity even in the absence of stress [71].

4 Conclusions

In this work we report the results of training classification models to predict whether a given chemical compound promotes longevity in mice. Each instance used to train our models was labelled using murine lifespan data from DrugAge, which reflects the current state-of-the-art literature on animal longevity studies. We created datasets using various feature types to describe these compounds, with the most successful models resulting from direct protein target annotations (GO terms, pathways, protein domains and UniProt keywords). Notably, features related to G-protein coupled receptors, especially receptors for neurotransmitters, metabolic hormones and sex hormones, were identified as strong predictors of lifespan extension.

We used the top-performing models to identify compounds with potential for murine lifespan extension, by highlighting consistent false-positive classifications and by creating ensembles to classify over 5000 unseen, unlabelled instances from DrugBank. We clustered the most confident positive-class predictions from the unlabelled data analysis based on their feature similarity, using the DBSCAN clustering algorithm, after applying UMAP dimensionality reduction. Major clusters of prioritised compounds target receptors to IGF1, insulin, adrenaline, noradrenaline, dopamine, serotonin, acetylcholine and adenosine, sodium-dependent dopamine, serotonin and noradrenalin transporters, voltage-gated potassium and calcium channels, as well as carbonic anhydrases.

There are several limitations of our work. First, the number of lifespan experiments performed on *M. musculus* is very small compared to *C. elegans* or *D. melanogaster,* due to their much higher cost and duration, which limits the number of instances available for training the classifiers. This could mean that some important classes of longevity drugs are missing altogether, although our results indicate that our models can nominate novel classes of drugs which were absent in the training dataset, presumably based on shared feature patterns between compounds of different classes. A related limitation is that there are even fewer successful lifespan-extending experiments with female mice, although both sexes are usually tested simultaneously. This could be explained by inherent biological differences in ageing or resistance mechanisms between males and females, as well as potential biases in the choice of compounds for testing. Lack of positive examples limited our ability to train successful classifiers for females, however, we were able to train moderately successful mixed-sex classifiers. Third, for the majority of compounds there is no information in DrugBank and other databases on whether they are inhibiting or activating their targets, which prevented us from constructing datasets and models that can discriminate this property. This led to prediction of multiple compounds with potentially opposing effects on lifespan. However, we were able to at least partially rectify this problem by manually comparing the predicted compounds with those in the training dataset.

In conclusion, this work provides a promising methodology for the preclinical discovery of lifespan-extending compounds in *Mus musculus*, with broader implications for human longevity research. By providing both *in silico* screening tools and biological insights into ageing mechanisms, we pave the way for the development of novel therapeutics targeting ageing and age-related diseases. Future studies should focus on validating the top predictions in animal models and exploring the translatability of these findings to humans.

5 Acknowledgements

This project was funded by a research grant from the UK's Biotechnology and Biological Sciences Research Council (BBSRC), grant reference numbers BB/V007971/1 and BB/V010123/1, to AAF and JPM.

6 Author Contributions

JPM and AAF conceived the overall project. CR and AAF designed the machine learning methodology. AVB and JPM designed the target selection and feature enrichment methodologies. AVB and CR created the datasets. CR implemented all machine learning algorithms and ran all the computational experiments. All authors analysed and discussed the results. The manuscript was written mainly by AVB and CR, but all authors contributed to writing and revising the manuscript.

7 Conflicts of Interest

JPM is CSO of YouthBio Therapeutics, an advisor/consultant for the BOLD Longevity Growth Fund, 199 Biotechnologies, and NOVOS, and the founder of Magellan Science Ltd, a company providing consulting services in longevity science. The other authors declare no conflict of interest.

8 Data Availability

The datasets used in the experiments will be made freely available on the web when the paper is published.

9 References

1. Kennedy, B. K. *et al.* Geroscience: linking aging to chronic disease. *Cell* **159**, 709–713 (2014).
2. Partridge, L., Fuentealba, M. & Kennedy, B. K. The quest to slow ageing through drug discovery. *Nat Rev Drug Discov* **19**, 513–532 (2020).
3. de Magalhães, J. P. Longevity pharmacology comes of age. *Drug Discov Today* **26**, 1559–1562 (2021).
4. Zhang, B., Trapp, A., Kerepesi, C. & Gladyshev, V. N. Emerging rejuvenation strategies—Reducing the biological age. *Aging Cell* **21**, e13538 (2022).
5. Jiang, T., Gradus, J. L. & Rosellini, A. J. Supervised Machine Learning: A Brief Primer. *Behav Ther* **51**, 675–687 (2020).
6. Barardo, D. G. *et al.* Machine learning for predicting lifespan-extending chemical compounds. *Aging (Albany NY)* **9**, 1721–1737 (2017).
7. Manoochehri, H. E., Kadiyala, S. S., Birjandtalab, J. & Nourani, M. Feature Selection to Predict Compound's Effect on Aging. in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* 419–427 (Association for Computing Machinery, New York, NY, USA, 2018). doi:10.1145/3233547.3233597.
8. Kapsiani, S. & Howlin, B. J. Random forest classification for predicting lifespan-extending chemical compounds. *Sci Rep* **11**, 13812 (2021).
9. Yarmolenko, M. & Howlin, B. Extreme Gradient Boosting algorithm classification for predicting lifespan-extending chemical compounds. Preprint at https://doi.org/10.21203/rs.3.rs-2199002/v1 (2022).
10. Canales, C. P. & Walz, K. The Mouse, a Model Organism for Biomedical Research. in *Cellular and Animal Models in Human Genomics Research* 119–140 (Elsevier, 2019). doi:10.1016/B978-0-12-816573-7.00006-7.
11. Flurkey, K., Mcurrer, J. & Harrison, D. Mouse Models in Aging Research. in *The Mouse in Biomedical Research* vol. III 637–672 (Elsevier, 2007).

12. Vanhooren, V. & Libert, C. The mouse as a model organism in aging research: usefulness, pitfalls and possibilities. *Ageing Res Rev* **12**, 8–21 (2013).
13. Barardo, D. *et al.* The DrugAge database of aging-related drugs. *Aging Cell* **16**, 594–597 (2017).
14. de Magalhães, J. P. *et al.* Human Ageing Genomic Resources: updates on key databases in ageing research. *Nucleic Acids Research* **52**, D900–D908 (2024).
15. Nadon, N. L. *et al.* Design of aging intervention studies: the NIA interventions testing program. *AGE* **30**, 187–199 (2008).
16. Szklarczyk, D. *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research* **51**, D638–D646 (2023).
17. Stathias, V. *et al.* LINCS Data Portal 2.0: next generation access point for perturbation-response signatures. *Nucleic Acids Research* **48**, D431–D439 (2020).
18. Kim, S. *et al.* PubChem 2023 update. *Nucleic Acids Res* **51**, D1373–D1380 (2023).
19. Knox, C. *et al.* DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Res* **52**, D1265–D1275 (2024).
20. Austad, S. N. & Bartke, A. Sex Differences in Longevity and in Responses to Anti-Aging Interventions: A Mini-Review. *Gerontology* **62**, 40–46 (2015).
21. Yuan, R. *et al.* Genetic differences and longevity-related phenotypes influence lifespan and lifespan variation in a sex-specific manner in mice. *Aging Cell* **19**, e13263 (2020).
22. Szklarczyk, D. *et al.* STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic Acids Res* **44**, D380–D384 (2016).
23. Kelleher, K. J. *et al.* Pharos 2023: an integrated resource for the understudied human proteome. *Nucleic Acids Res* **51**, D1405–D1416 (2023).
24. Oughtred, R. *et al.* The BIOGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Science* **30**, 187–200 (2021).

25. Zdrazil, B. *et al.* The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research* **52**, D1180–D1192 (2024).
26. Davies, M. *et al.* ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res* **43**, W612-620 (2015).
27. Zhou, Y. *et al.* TTD: *Therapeutic Target Database* describing target druggability information. *Nucleic Acids Research* **52**, D1465–D1477 (2024).
28. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29 (2000).
29. Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucleic Acids Research* **51**, D418–D427 (2023).
30. Milacic, M. *et al.* The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Research* **52**, D672–D678 (2024).
31. Agrawal, A. *et al.* WikiPathways 2024: next generation pathway database. *Nucleic Acids Res* **52**, D679–D689 (2024).
32. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* **51**, D587–D592 (2023).
33. The UniProt Consortium *et al.* UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **51**, D523–D531 (2023).
34. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452.e17 (2017).
35. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).
36. Grinsztajn, L., Oyallon, E. & Varoquaux, G. Why do tree-based models still outperform deep learning on tabular data? Preprint at http://arxiv.org/abs/2207.08815 (2022).
37. Flach, P. A. ROC Analysis. in *Encyclopedia of Machine Learning* (eds. Sammut, C. & Webb, G. I.) 869–875 (Springer US, Boston, MA, 2011). doi:10.1007/978-0-387-30164-8_733.
38. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

39. Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. *Classification And Regression Trees*. (Routledge, 2017). doi:10.1201/9781315139470.

40. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at http://arxiv.org/abs/1802.03426 (2020).

41. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* 226–231 (AAAI Press, Portland, Oregon, 1996).

42. Belikov, A. V., Vyatkin, A. D. & Leonov, S. V. Novel Driver Strength Index highlights important cancer genes in TCGA PanCanAtlas patients. *PeerJ* **10**, e13860 (2022).

43. Lucanic, M. *et al.* Chemical activation of a food deprivation signal extends lifespan. *Aging Cell* **15**, 832–841 (2016).

44. Cekic, C. & Linden, J. Purinergic regulation of the immune system. *Nat Rev Immunol* **16**, 177–192 (2016).

45. Kaiser, K. A. *et al.* The effect of a pharmaceutical ghrelin agonist on lifespan in C57BL/6J male mice: A controlled experiment. *Aging Cell* **22**, e13787 (2023).

46. Terekhova, M. *et al.* Single-cell atlas of healthy human blood unveils age-related loss of NKG2C+GZMB–CD8+ memory T cells and accumulation of type 2 memory T cells. *Immunity* **57**, 188–192 (2024).

47. Ceresini, G. *et al.* Evaluation of the circadian profiles of serum dehydroepiandrosterone (DHEA), cortisol, and cortisol/DHEA molar ratio after a single oral administration of DHEA in elderly subjects. *Metabolism* **49**, 548–551 (2000).

48. Mao, K. *et al.* Late-life targeting of the IGF-1 receptor improves healthspan and lifespan in female mice. *Nat Commun* **9**, 2394 (2018).

49. Hedges, C. P. *et al.* Dietary supplementation of clinically utilized PI3K p110α inhibitor extends the lifespan of male and female mice. *Nat Aging* **3**, 162–172 (2023).

50. Miller, R. A. *et al.* Rapamycin-mediated lifespan increase in mice is dose and sex dependent and metabolically distinct from dietary restriction. *Aging Cell* **13**, 468–477 (2014).

51. Strong, R. *et al.* Rapamycin-mediated mouse lifespan extension: Late-life dosage regimes with sex-specific effects. *Aging Cell* **19**, e13269 (2020).

52. Pierpaoli, W. & Regelson, W. Pineal control of aging: effect of melatonin and pineal grafting on aging mice. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 787–791 (1994).

53. Anisimov, V. N. *et al.* Melatonin increases both life span and tumor incidence in female CBA mice. *J Gerontol A Biol Sci Med Sci* **56**, B311-323 (2001).

54. Spindler, S. R. *et al.* β1-Adrenergic receptor blockade extends the life span of Drosophila and long-lived mice. *AGE* **35**, 2099–2109 (2013).

55. Levine, H. J. Rest heart rate and life expectancy. *J Am Coll Cardiol* **30**, 1104–1106 (1997).

56. Di Francesco, A. *et al.* Dietary restriction impacts health and lifespan of genetically diverse mice. *Nature* **634**, 684–692 (2024).

57. Clapp, N. K., Satterfield, L. C. & Bowles, N. D. Effects of the antioxidant butylated hydroxytoluene (BHT) on mortality in BALB/c mice. *J Gerontol* **34**, 497–501 (1979).

58. Pollard, A., Shephard, F., Freed, J., Liddell, S. & Chakrabarti, L. Mitochondrial proteomic profiling reveals increased carbonic anhydrase II in aging and neurodegeneration. *Aging* **8**, 2425–2436 (2016).

59. Cotzias, G. C., Miller, S. T., Tang, L. C. & Papavasiliou, P. S. Levodopa, fertility, and longevity. *Science* **196**, 549–551 (1977).

60. Dang, Y. *et al.* Berberine ameliorates cellular senescence and extends the lifespan of mice via regulating p16 and cyclin protein expression. *Aging Cell* **19**, e13060 (2020).

61. Doeppner, T. R. *et al.* Long-term treatment with chloroquine increases lifespan in middle-aged male mice possibly via autophagy modulation, proteasome inhibition and glycogen metabolism. *Aging (Albany NY)* **14**, 4195–4210 (2022).

62. del Río, B. *et al.* Melatonin, an endogenous-specific inhibitor of estrogen receptor alpha via calmodulin. *J Biol Chem* **279**, 38294–38302 (2004).

63. Singh, P. *et al.* Taurine deficiency as a driver of aging. *Science* **380**, eabn9257 (2023).

64. Harrison, D. E. *et al.* Astaxanthin and meclizine extend lifespan in UM-HET3 male mice; fisetin, SG1002 (hydrogen sulfide donor), dimethyl fumarate, mycophenolic acid, and 4-phenylbutyrate do not significantly affect lifespan in either sex at the doses and schedules used. *GeroScience* **46**, 795–816 (2023).

65. Archer, J. R. & Harrison, D. E. L-deprenyl treatment in aged mice slightly increases life spans, and greatly reduces fecundity by aged males. *J Gerontol A Biol Sci Med Sci* **51**, B448-453 (1996).

66. Costa-Machado, L. F. *et al.* Peripheral modulation of antidepressant targets MAO-B and GABAAR by harmol induces mitohormesis and delays aging in preclinical models. *Nat Commun* **14**, 2779 (2023).

67. Gautam, D. *et al.* Beneficial metabolic effects of M3 muscarinic acetylcholine receptor deficiency. *Cell Metab* **4**, 363–375 (2006).

68. Yi, J. H. *et al.* M1 muscarinic acetylcholine receptor dysfunction in moderate Alzheimer's disease pathology. *Brain Communications* **2**, fcaa058 (2020).

69. Leone, R. D. & Emens, L. A. Targeting adenosine for cancer immunotherapy. *J Immunother Cancer* **6**, 57 (2018).

70. Zullo, J. M. *et al.* Regulation of lifespan by neural excitation and REST. *Nature* **574**, 359–364 (2019).

71. Miller, H. A., Dean, E. S., Pletcher, S. D. & Leiser, S. F. Cell non-autonomous regulation of health and longevity. *eLife* **9**, e62659 (2020).