PAPER

MUSE-XAE: MUtational Signature Extraction with eXplainable AutoEncoder enhances tumour types classification

Corrado Pancotti,[1] Cesare Rollo,[1] Francesco Codicè,[1] Giovanni Birolo,[1] Piero Fariselli[1,*] and Tiziana Sanavia[1,*]

[1]Department of Medical Science, University of Torino, Via Santena 19, 10126, Torino, Italy
[*] tiziana.sanavia@unito.it; piero.fariselli@unito.it

FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

Motivation: Mutational signatures are a critical component in deciphering the genetic alterations that underlie cancer development and have become a valuable resource to understand the genomic changes during tumorigenesis. Therefore, it is essential to employ precise and accurate methods for their extraction to ensure that the underlying patterns are reliably identified and can be effectively utilized in new strategies for diagnosis, prognosis and treatment of cancer patients.
Results: We present MUSE-XAE, a novel method for mutational signature extraction from cancer genomes using an explainable autoencoder. Our approach employs a hybrid architecture consisting of a nonlinear encoder that can capture nonlinear interactions among features, and a linear decoder which ensures the interpretability of the active signatures. We evaluated and compared MUSE-XAE with other available tools on both synthetic and real cancer datasets and demonstrated that it achieves superior performance in terms of precision and sensitivity in recovering mutational signature profiles. MUSE-XAE extracts highly discriminative mutational signature profiles by enhancing the classification of primary tumour types and subtypes in real world settings. This approach could facilitate further research in this area, with neural networks playing a critical role in advancing our understanding of cancer genomics.
Availability: MUSE-XAE software is freely available at https://github.com/compbiomed-unito/MUSE-XAE
Supplementary Information: Supplementary data are available at Bioinformatics online.

Key words: Autoencoder, Mutational Signatures, Explainability, Cancer, Genomics

Introduction

Mutational signatures are patterns of somatic mutations that reflect the underlying biological processes that drive cancer development (Alexandrov et al., 2013a,b; Helleday et al., 2014). Mutational signatures have become a valuable resource for deciphering the genetic alterations underpinning cancer and for developing targeted therapies (Ma et al., 2018; Secrier et al., 2016; Schulze et al., 2015). In recent years, many methods have been developed for the extraction of mutational signatures, most of which are based on matrix factorization techniques, such as non-negative matrix factorization (NMF) (Islam et al., 2022; Blokzijl et al., 2018; Bayati et al., 2020; Vöhringer et al., 2021; Ardin et al., 2016; Degasperi et al., 2020) and its probabilistic versions (Gori and Baez-Ortega, 2018; Fischer et al., 2013; Rosales et al., 2017). These approaches have been applied to several types of cancer, identifying more than 60 distinct signatures associated with specific mutational processes (Alexandrov et al., 2020; Tate et al., 2019).

Although these techniques have proven to be highly effective for the extraction of mutational signatures, some studies have highlighted possible issues that may arise during extraction and should deserve attention and further investigation (Maura et al., 2019; Koh et al., 2021). Many of the mutational signatures found in the COSMIC catalogue currently have no known aetiology. Some are merely statistically linked to a specific process, whereas others lack any statistical association. In addition, the high degree of similarity between certain signatures may suggest the existence of non-biological, overfitted signals (Pancotti et al., 2023; Schumann et al., 2019; Lal et al., 2021).

Future extraction techniques should consider these issues, implementing constraints in the decomposition to reduce the detection of overly similar profiles. Furthermore, NMF identifies only linear interactions, which could be a limitation as it might not capture potential nonlinear dependencies within the genome that can contribute to cancer development (Wojtowicz et al., 2021; Kičiatovas et al., 2022).

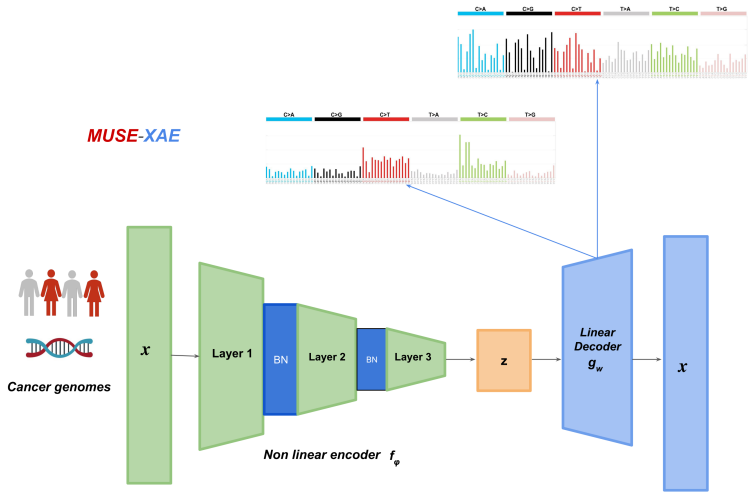

Fig. 1. MUSE-XAE schematic architecture. MUSE-XAE features a nonlinear encoder, made up of three layers that leverage a softplus activation function with batch normalization. The decoder is designed to be linear in order to enhance the interpretability.

To overcome these challenges, we present MUSE-XAE, MUtational Signature Extraction with eXplainable AutoEncoder. This model includes a nonlinear encoder and a linear decoder with a non-negative constraint and a minimum volume regularization (Miao and Qi, 2007) to detect potential nonlinear dependencies while preserving signature interpretability. Autoencoders have been successfully implemented in various domains, including genomics, to obtain compact and informative data representations. Autoencoders employing a hybrid architecture with a nonlinear encoder and a linear decoder have been applied in the context of single-cell RNA-seq and transcriptomic data (Svensson et al., 2020; Seninge et al., 2021), achieving great success due to their explainability while preserving powerful performance capabilities. However, to the best of our knowledge, no applications of this architecture have been developed so far in the context of mutational signature analysis.

To fill this gap, this paper introduces MUSE-XAE and shows its effectiveness on different cancer datasets by comparing it to existing state-of-the-art approaches in both synthetic scenarios and real-world applications. In a comprehensive comparison with 10 other *de novo* extraction tools considering realistic synthetic scenarios, MUSE-XAE resulted the best performing model with high sensitivity and precision in recovering the true signature profiles. We then evaluated our approach in two real world settings: 2,780 cancer samples from the Pan-Cancer Analysis of Whole Genomes (PCAWG) study (icg, 2020), and 1,865 samples from another whole-genome sequencing (WGS) cohort (Islam et al., 2022). MUSE-XAE was able to extract highly discriminative signature profiles that can significantly improve the classification of tumour types and subtypes.

Materials and methods

MUSE-XAE architecture

An autoencoder is a type of neural network capable of learning a lower-dimensional representation of the data. Given an input space X, it consists of an encoder network f, represented by one or more layers, that maps the input data to a lower-dimensional latent space Z, and a decoder network g that reconstructs the input space from the latent representation. The goal of an autoencoder is to minimize the reconstruction error $\mathcal{L}(x, \hat{x})$ between the original input x and the reconstructed output $\hat{x}$. The general equations that define an autoencoder are:

$$z = f(x) = Encoder \tag{1}$$

$$\hat{x} = g(z) = g(f(x)) = Decoder \tag{2}$$

usually, both f and g represent nonlinear activation functions.

MUSE-XAE implements a hybrid architecture with a nonlinear encoder to learn a latent representation z of cancer samples, and a linear decoder with a non-negative constraint and minimum volume regularization to reconstruct the original input, such as $\hat{x} = zW^T$. Specifically, MUSE-XAE encoder f includes three hidden layers with batch normalization and a softplus activation function. The latter function offers continuous differentiability and a smoother transition from negative to positive values compared to ReLU, reducing the risk of neuron inactivation, with improved stability (Zheng et al., 2015). The decoder g is characterized by a weight matrix W with non-negativity constraint, a linear activation function that ensures interpretability, and a minimum volume regularization that helps the model find a more disentangled representation. In addition, MUSE-XAE exploits a non-negative Poisson likelihood function to take into consideration the count nature of the input data, and an early stopping criterion to avoid overfitting. Considering all the contributions, the total loss function $\mathcal{L}(x, \hat{x})$ can be written as:

$$\mathcal{L}(x, \hat{x}) = -x \log(\hat{x}) + \hat{x} + \beta \log\left(\det\left(WW^T + I\right)\right) \tag{3}$$

$$\text{subjected to} \quad W \geq 0$$

where the first two terms refer to the Poisson likelihood function, while the third term represents the logarithm of the minimum volume constraint. The β coefficient regulates the strength of the regularization. Referring to the mutational signature terminology, the latent representation z represents the cancer genome's exposures, while the decoder weight matrix W represents the mutational signatures. MUSE-XAE architecture is displayed in Fig.1.

Signatures Extraction

MUSE-XAE *de novo* extraction procedure is summarized in Algorithm 1. Training a neural network requires a substantial amount of data to exploit its capacity and fit the parameters effectively. In MUSE-XAE implementation, we used a data augmentation strategy to overcome this challenge.

Algorithm 1. MUSE-XAE De Novo Extraction procedure

Given $C \in R^{m \times 96}$ ▷ tumour catalogue matrix
Step 1: Data Bootstrapping
Obtain the augmented count matrix C_{aug}
bootstrapping each genome t times from $M(N, p)$

Step 2: Training
for k in 1..K signatures **do**
  for i in 1..n iterations **do**
    MUSE-XAE.train(C_{aug})
    C_{pred}=MUSE-XAE.predict(C)
    Collects $E_{rec} = \|C - C_{pred}\|_F$
    Collects W_{ik}

Step 3: Clustering
for each k in 1..K **do**
  K-Means clustering with matching on $\{W_{1k}, ..., W_{nk}\}$
  Obtain consensus Signatures matrix S_k

Step 4: Filtering
Given threshold th_{avg} and th_{min}
Filter solutions with
$silhouette_{\text{avg}} > th_{\text{avg}}$ and $silhouette_{\text{min}} > th_{\text{min}}$

Step 5: Optimal Solution Selection
Sort filtered solutions based on E_{rec},
The optimal solution is the one with the lowest E_{rec}

Specifically, given a tumour catalogue matrix $C \in R^{m \times 96}$, where m is the number of samples and 96 is the number of mutational channels, for each cancer genome with N total number of mutations, we determined the relative mutation frequency p for each of the 96 mutational classes. Then, we generated new data points by bootstrapping cancer genomes t times through a multinomial distribution $M(N, p)$, obtaining the augmented count matrix C_{aug}. This approach has already been used by other tools to ensure the stability of a consensus signature (Alexandrov et al., 2013a; Islam et al., 2022; Alexandrov et al., 2020). Here, we repeated the process t times to increase the dataset size. Then, in order to select the optimal number of active signatures K^*, we used a revised version of the NMFk approach, originally described by (Nebgen et al., 2021) and also adopted by SigProfilerExtractor. Specifically, for each number $k = 1, ..., K$ of candidate signatures, MUSE-XAE was trained n times with different weights' initialization to guarantee a better stability. Subsequently, a custom K-Means clustering with matching and based on the cosine similarity distance was performed on the set of the decoder weight matrices $\{W_{1k}...W_{nk}\}$ to find the consensus signature matrices $\{S_k\}$. This custom clustering approach exploits the Jonker-Volgenant algorithm (Jonker and Volgenant, 1987) to solve the linear assignment problem (i.e. the matching) and to find k clusters of equal size n. Once obtained the set of clusters, whose centroids represent the signature matrices $\{S_k, k = 1...K\}$, we considered only the solutions with an average and a minimum silhouette scores above a fixed threshold, choosing as the best solution K^*, that allows the minimum reconstruction error of the original matrix of cancer samples.
A detailed description is available in Supplementary (section *MUSE-XAE De Novo Extraction procedure* and Fig. S1). The size factor t for data augmentation, the number of repetitions n for each candidate signature k and the thresholds for the average and minimum silhouette scores, i.e. th_{avg} and th_{min}, can be specified by the user. The open-source code is available at https://github.com/compbiomed-unito/MUSE-XAE.

Signatures Assignment

Once the profiles of active signatures within a set of genomes have been identified, it is necessary to understand which signatures cause mutations in a genome and in what amount, that is, we need to assign the contribution of each extracted signature to each genome. Therefore, we used a slightly modified version of MUSE-XAE for the signature extraction. Specifically, we normalized the computed consensus matrix S_k into $S_{k_{norm}}$, which is used to initialize the weights of the decoder, and then freeze them, so that the decoder is no longer trainable and only the weights of the encoder are trained.
In order to obtain a sparse representation and to avoid over-assignments of the mutational signatures, we used an L1 penalty both for the weights of the last layer of the encoder and for the output of the encoder after a ReLU activation, training the network until convergence. Our new latent representation z represents the exposure of the signatures within the genomes, i.e., the number of mutations of a certain mutational class that a signature causes within a genome. We summarized the signature assignment procedure in the Algorithm 2.

Algorithm 2. Signature Assignment Procedure

Given $C \in R^{m \times 96}$ ▷ SBS catalogue matrix
Step 1: Signature Profiles Normalization
Normalize the consensus matrix S_k from Algorithm 1.
Get $S_{k_{norm}}$

Step 2: Initialization and Freezing of Weights
Initialize and freeze decoder weights with $S_{k_{norm}}$

Step 3: Improve Sparsity
Add an L1 penalty on the weights of the last encoder layer
Add an L1 penalty on the output after activation

Step 4: Train the network and obtain Exposures
MUSE-XAE.train(C)
Exposures=MUSE-XAE.z

De Novo Extraction Scenarios

To evaluate the performance of MUSE-XAE in the context of mutational signature extraction, we considered 5 realistic synthetic scenarios (`ftp://alexandrovlab-ftp.ucsd.edu/pub/publications/Islam_et_al_SigProfilerExtractor/`), available from SigProfilerExtractor (Islam et al., 2022). Specifically:

- **Scenario 1:** 1,000 synthetic samples, modelling a subset of the pancreatic adenocarcinoma dataset from PCAWG. The 11 ground-truth signatures are based on COSMIC.
- **Scenario 2:** 1,000 synthetic tumours from flat, relatively featureless mutational signatures, including a mix of 500 synthetic renal cell carcinomas (high prevalence and

mutation load from SBS5 and SBS40) and 500 synthetic ovarian adenocarcinomas (high prevalence and mutation load from SBS3), with 11 COSMIC-based signatures.

- **Scenario 3:** 1,000 synthetic tumours from signatures with overlapping and potentially interfering profiles, mostly SBS2, SBS7a and SBS7b. The mutational load distributions were drawn from bladder transitional cell carcinoma (SBS2) and skin melanoma (SBS7a, SBS7b), with 11 COSMIC-based signatures.
- **Scenario 4:** 1,000 synthetic tumours emulating a mix of 500 synthetic renal cell carcinomas (high prevalence and mutation load from SBS5 and SBS40) and 500 synthetic ovarian adenocarcinomas (high prevalence and mutation load from SBS3). In this scenario, only 3 COSMIC-based signatures (SBS3, SBS5, SBS40) are present.
- **Scenario 5:** 2,700 synthetic samples with mutational spectra matching those in PCAWG, including 300 spectra from each of 9 different cancer types: bladder transitional cell carcinoma, oesophageal adenocarcinoma, breast adenocarcinoma, lung squamous cell carcinoma, renal cell carcinoma, ovarian adenocarcinoma, osteosarcoma, cervical and stomach adenocarcinoma. The ground-truth signatures are 21 signatures based on COSMIC.

We extracted the mutational signatures from each of these scenarios using MUSE-XAE and applied the same performance metrics as in (Islam et al., 2022). We used the Hungarian algorithm (Jonker and Volgenant, 1987) to match predicted and known signatures according to the cosine-similarity. Since the signatures in each scenario are known, an extracted signature was considered correctly identified, or a True Positive (TP), if the cosine similarity between extracted and real signatures was $\geq threshold$. If the profile of a signature is missing or the cosine similarity was $< threshold$, it was considered a False Negative (FN) or Positive (FP), respectively.
For each scenario, precision, sensitivity, and F1 score were calculated from the corresponding confusion matrices at different cosine similarity thresholds, ranging between 0.8 and 1. A description of the evaluation metrics was reported in Supplementary (section *Evaluation metrics*).

Real World datasets

In order to evaluate the performance of MUSE-XAE also in real world scenarios, we applied our method to both the Pancancer Analysis of Whole Genomes (PCAWG) dataset, including 2,780 tumour samples, and a WGS cohort of 1,865 genomes collected from various studies and including the International Cancer Genome Consortium (ICGC), as compiled in (Islam et al., 2022). For both datasets, we performed: 1) a *de Novo* extraction of mutational signatures and comparison of the profiles with those of SigProfilerExtractor and with the known signatures from COSMIC (Tate et al., 2019) and Signal (Degasperi et al., 2022) databases; 2) an evaluation of how the signatures and consequently the exposures are discriminative, performing a multi-class classification of the cancer types. Specifically, we used the exposures as new features which were fed into a Random Forest to classify both the primary sites and the cancer subtypes. Finally, we evaluated the performance in terms of balanced accuracy, Matthews Correlation Coefficient (MCC) and Kohen Kappa score in a 5-fold cross-validation setting. A description of the metrics is available in Supplementary (section *Evaluation metrics*).

Results

Data augmentation improves robustness and accuracy

We first investigated the influence of data augmentation on the extraction of mutational signatures in each of the five synthetic scenarios. Specifically, we performed *de novo* extraction with MUSE-XAE for each of the five datasets, varying the data augmentation level from 1 to 100 times the original dataset size. We repeated the extraction five times at each augmentation level to evaluate stability and accuracy. As depicted in Fig. 2, for the five datasets there is a trend where an increase in data augmentation not only enhances run-to-run stability, but also improves the correct estimation of the real number of profiles. To further assess the effects of data augmentation, we computed the average precision, sensitivity and F1 scores across the five scenarios at different thresholds of the cosine similarity between the extracted and the real profiles, ranging between 0.8 and 1. Fig. 3 shows the overall performance. Notably, the sensitivity in the signature profile detection improves with the size of the data augmentation. This confirms that the use of data augmentation is a strategy that improves the detection of signature profiles, and it can be used as an effective technique to further enhance the extraction performance.

De Novo extraction comparison in synthetic scenarios

We compared MUSE-XAE (using 100 data augmentation) with 10 state-of-the-art *de novo* signature extraction tools, considering the results reported in `ftp://alexandrovlab-ftp.ucsd.edu/pub/publications/Islam_et_al_SigProfilerExtractor/`. Precision, sensitivity and F1 score were computed in each scenario at different thresholds of the cosine similarity between the extracted and the real profiles, ranging between 0.8 and 1 for each method. Supplementary Fig. S2 shows precision, sensitivity and F1 score of the top 10 performing methods averaged across the five scenarios, while Fig. 4 and Supplementary Table S1 show the distribution of the normalized Area Under the Curve (AUC) for the performance scores. In addition, Supplementary Fig. S3 reports, for each scenario, the F1 scores at different thresholds of the cosine similarity. Observed results reveal that MUSE-XAE is, on average, the best performing method in all metrics, followed by SigProfilerExtractor and SigProfilerPCAWG.

De Novo Extraction in real world datasets

We applied MUSE-XAE for *de novo* extraction of mutational signatures in 2,780 samples from PCAWG (18 cancer primary sites and 37 cancer subtypes), and in 1,865 samples from an additional extended WGS cohort (15 cancer primary sites and 23 cancer subtypes). We used MUSE-XAE with the data augmentation strategy (i.e. 100 times the original dataset size for 100 iterations) to find stable consensus signatures. MUSE-XAE found 22 and 23 mutational signature profiles in the PCAWG and the extended WGS cohort, respectively. Their profiles are presented in Supplementary Fig. S4 and Fig. S5.
By matching the 22 profiles identified by MUSE-XAE with the 21 found by SigProfilerExtractor in the PCAWG cohort, the two methods extracted 21 highly similar profiles, showing a mean cosine similarity of 0.98, with a minimum of 0.92. (Fig. 5, left panel). On the other hand, in the extended WGS cohort, MUSE-XAE found 23 signatures, while SigProfilerExtractor 21, with a mean cosine similarity of 0.90 but a minimum of 0.36 between the 21 most similar signatures. In Fig. 5 it is possible to observe that, in the extended WGS cohort (right panel), although there are 19 out of 21 profiles with

a cosine similarity greater than 0.8, the distribution along the diagonal is lower than that one observed in PCAWG (left panel). Moreover, there are two pairs of signatures with a notably low cosine similarity, specifically 0.69 and 0.36, meaning that the two methodologies extract different signature profiles. Therefore, in general, although the two methods are fairly in agreement, MUSE-XAE seems to identify more and different profiles compared to SigProfilerExtractor.

Given that two random 96-component vectors have a cosine similarity of 0.75, and that 0.80 is commonly used as a threshold to determine if two signatures represent the same profile, we can observe from the Supplementary Table S2 that almost all MUSE-XAE profiles from the PCAWG cohort are in agreement with the known signatures from COSMIC and Signal databases. An exception is MUSE-SBSV, which shows a cosine similarity of 0.78 in both COSMIC and Signal databases, potentially indicating an incomplete extraction of the original signature. On the other hand, in the WGS-extended cohort, despite most extracted profiles align well with those in COSMIC and Signal databases (Table S3), there are three signatures (MUSE-SBSP, MUSE-SBSS, and MUSE-SBSW) showing a cosine similarity below 0.75 with the matched signatures in both databases. Therefore, we further investigated the exposures of these three signatures in the WGS-extended cohort. Notably, as shown in the Supplementary Figure S6, MUSE-SBSW is predominantly observed in Eye-Melanoma samples (32 out of 46), indicating that it could be a tumour-specific signature. To validate this hypothesis, we performed a *de novo* extraction exclusively for Eye-Melanoma samples, which revealed a strikingly similar profile, showing a pairwise cosine similarity of 0.94 with the one extracted from the pancancer analysis. Given the limited number of samples, this finding reinforces the need for a

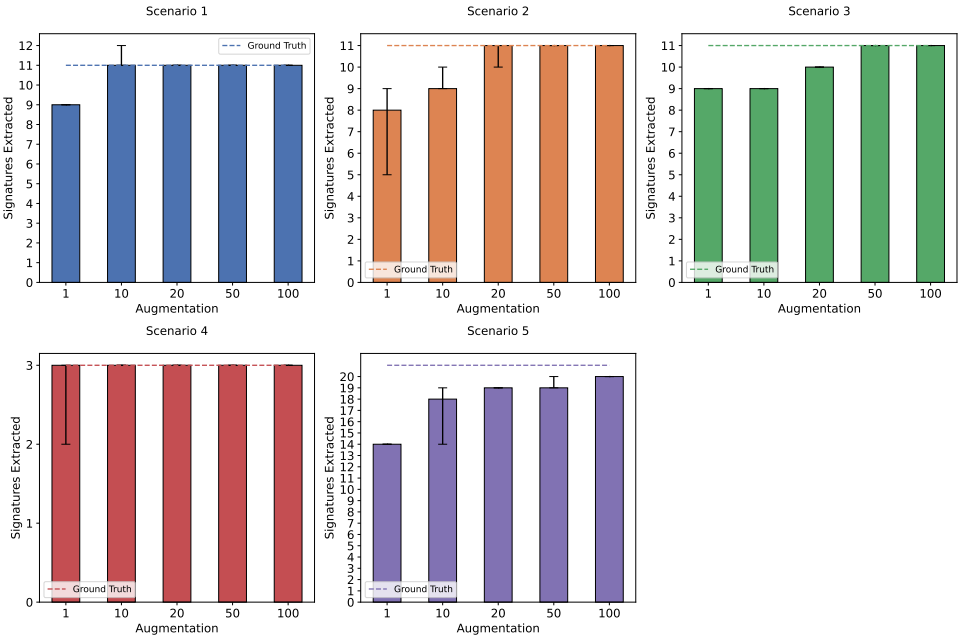

Fig. 2. Sensitivity analysis of data augmentation for each of the five synthetic scenario. Each bar represents the average number of extracted signatures over 5 repetitions. The dashed line represents the ground-truth, while the error bar represents the range between minimum and maximum.

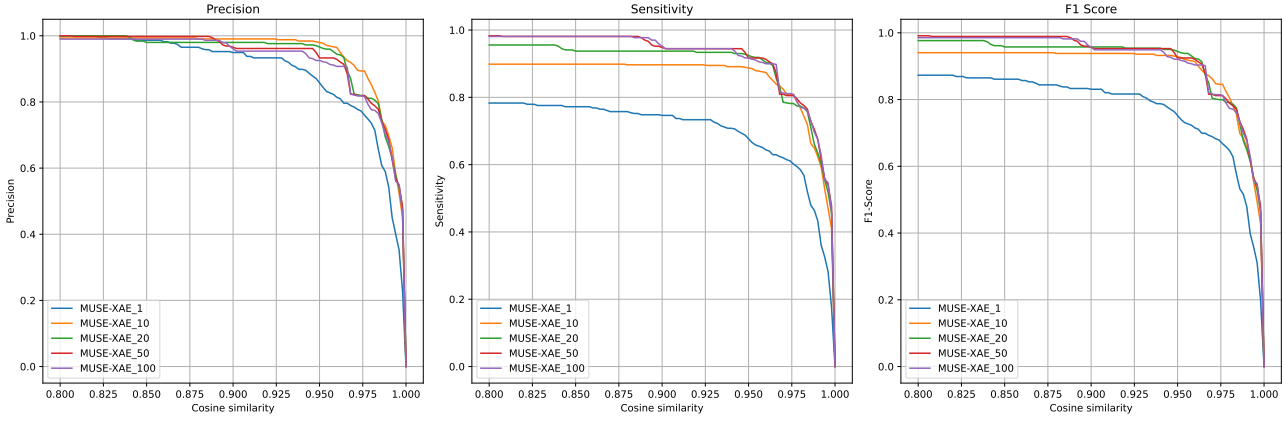

Fig. 3. Average precision, sensitivity and F1 scores in the five scenarios considering MUSE-XAE at different level of data augmentation and at varying thresholds of the cosine similarity, ranging between 0.8 and 1.

Table 1. Matthews correlation coefficient, Kappa score, and Balanced accuracy metrics calculated by 5-fold cross-validation in PCAWG and Extended WGS cohorts for primary tumour type classification. Metrics are reported as mean and standard deviation.

Model	Dataset	Matthews correlation	Cohen kappa score	Balance Accuracy
MUSE-XAE	**PCAWG cohort**	**0.75(0.02)**	**0.75(0.02)**	**0.65(0.02)**
SigProfilerExtractor	PCAWG cohort	0.71(0.01)	0.71(0.01)	0.61(0.02)
MUSE-XAE	**Extended cohort**	**0.73(0.01)**	**0.72(0.01)**	**0.65(0.01)**
SigProfilerExtractor	Extended cohort	0.70(0.02)	0.69(0.02)	0.61(0.02)

Table 2. Matthews correlation coefficient, Kappa score, and Balanced accuracy metrics calculated by 5-fold cross-validation in PCAWG and WGS-extended cohorts for tumour subtype classification. Metrics are reported as mean and standard deviation.

Model	Dataset	Matthews correlation	Cohen kappa score	Balance Accuracy
MUSE-XAE	**PCAWG cohort**	**0.73(0.01)**	**0.73(0.01)**	**0.58(0.01)**
SigProfilerExtractor	PCAWG cohort	0.67(0.02)	0.67(0.02)	0.52(0.02)
MUSE-XAE	**Extended cohort**	**0.68(0.03)**	**0.67(0.03)**	**0.50(0.03)**
SigProfilerExtractor	Extended cohort	0.66(0.02)	0.66(0.02)	0.50(0.03)

comprehensive examination of this profile, focusing on its origin and validation in an external cohort. Such an in-depth investigation, however, exceeds the objectives of our study, and it will be a focus of our future research.

MUSE-XAE enhances tumor classification

Considering *de novo* extraction of mutational signatures on real cancer datasets, although COSMIC and Signal databases can be used as reference for the extracted profiles, there is no actual ground truth to calculate the evaluation metrics. Therefore, to thoroughly evaluate the performance of MUSE-XAE, we examined the exposures of mutational signatures, i.e. the latent representation z of tumour samples, both qualitatively and quantitatively. While acknowledging that tumours of the same type may demonstrate a degree of heterogeneity, we assumed that these exposures, representing the mutations caused by a signature within a particular sample, could serve as a key discriminant between different tumour types and subtypes. Fig. 6 shows the t-distributed stochastic neighbour embedding (t-SNE) of the latent representations (exposures), coloured by the primary tumour types for both PCAWG and the extended WGS cohorts. The t-SNE of exposures displays a clear grouping pattern in both datasets, which provides compelling evidence in support of this hypothesis and indicates a coherent relationship between signatures exposures and tumour types.

To quantitatively assess this hypothesis, we implemented a Random Forest Classifier which considers the signature exposures as input features to classify both primary types and tumour subtypes. This classifier was applied to both MUSE-XAE and SigProfilerExtractor exposures using a balanced 5-fold cross-validation approach. To properly train the Random Forest in both datasets, we removed tumour types with less than 10 counts, i.e. the tumour subtypes Myeloid-MDS (n=4), Breast-DCIS (n=4) and Cervix-AdenoCA (n=2) in the PCAWG dataset, while in the extended WGS dataset we excluded Blood-CMDI (n=9), Sarcoma (n=3), and Bone-cancer (n=2). A complete description of the Random Forest implementation is reported in Section 3 of the Supplementary Materials. Classification performance metrics for primary tumour types and subtypes in PCAWG and the extended WGS cohorts are reported in Tables 1 and 2, respectively.

In both classification tasks, MUSE-XAE outperformed SigProfilerExtractor across all metrics, suggesting that the exposures and the corresponding signature profiles generated by MUSE-XAE are more discriminative and capable of accurately identifying tumour types. MUSE-XAE particularly outperformed SigProfilerExtractor in the classification of primary tumour types (Table 1), and it discriminates tumour subtypes much better, notably in the PCAWG cohort (Table 2). Supplementary Figures S7-S10 display the confusion matrices of MUSE-XAE for both primary tumour types and subtypes in PCAWG and the extended WGS cohorts. It is worth noticing that MUSE-XAE and SigProfilerExtractor both struggle with the classification of some tumour types. We investigated the possible reason behind considering Breast Cancer as a case study in Supplementary (section *Case study: Breast cancer*).

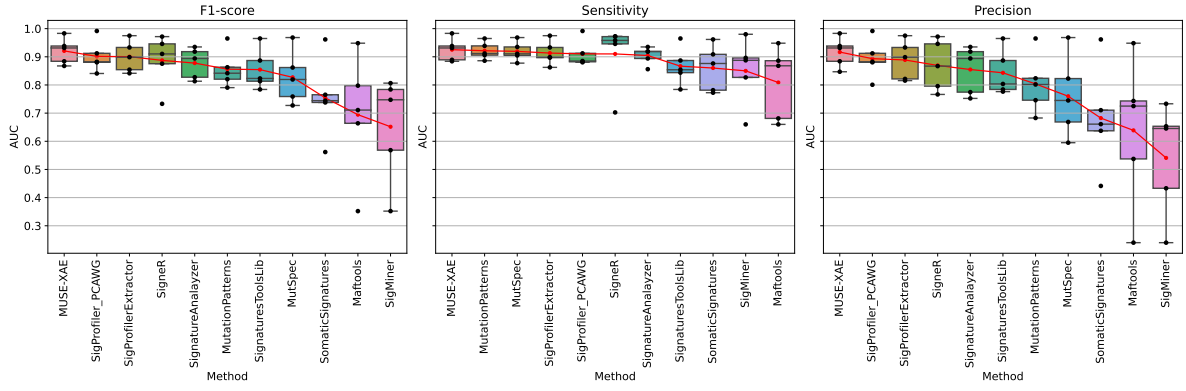

Fig. 4. Normalized AUC distribution for F1-score, Sensitivity and Precision across the five scenarios for all the tested methods. The methods are ordered according to the average (red dots).

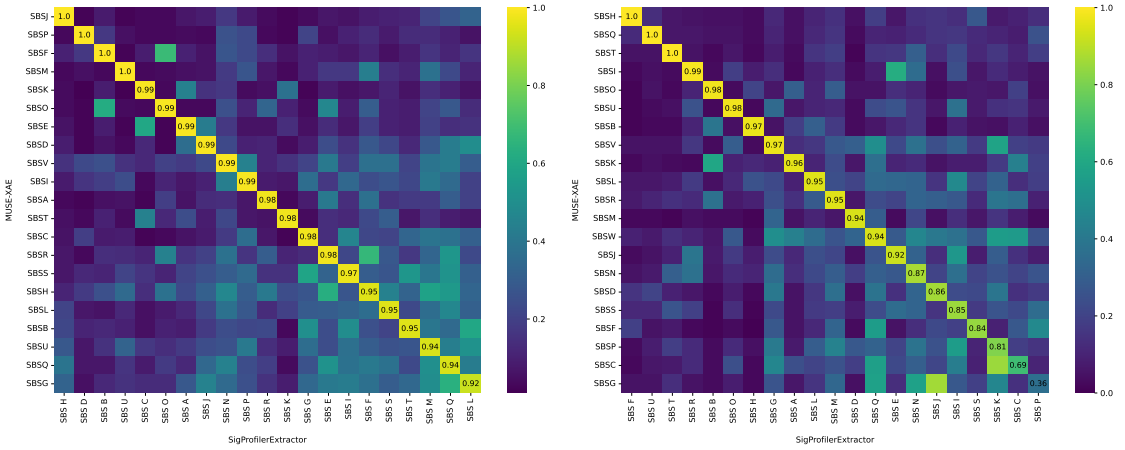

Fig. 5. Cosine similarity heatmap between the most similar signatures extracted by MUSE-XAE and SigProfilerExtractor for PCAWG and WGS-extended cohort.

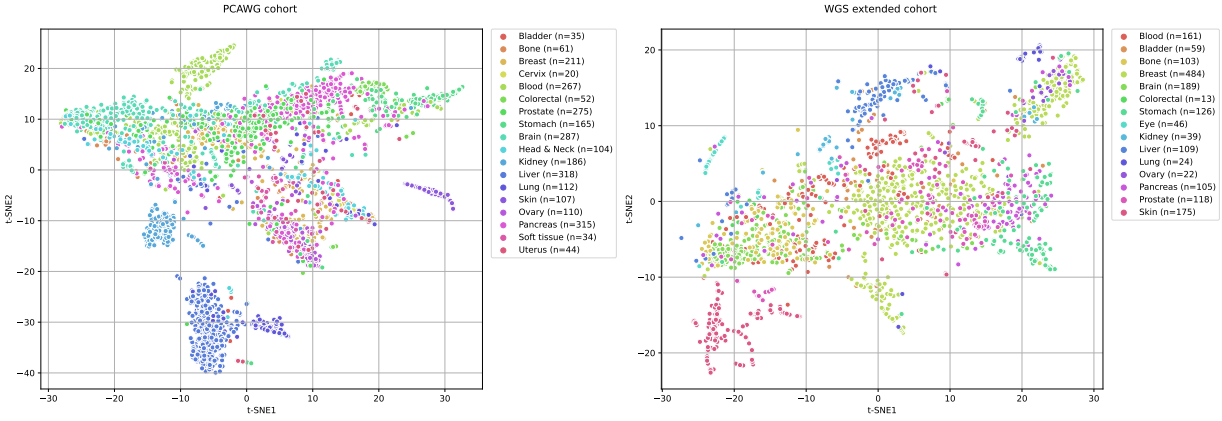

Fig. 6. t-SNE representation of the latent representation in PCAWG and the WGS-extended cohorts, post-hoc coloured by primary tumour sites.

Discussion

This study introduces MUSE-XAE, a novel method for mutational signature extraction based on an explainable autoencoder. MUSE-XAE combines a nonlinear encoder with a linear decoder by adding a non-negative constraint and a minimum volume regularization. Our method demonstrated high accuracy in the *de novo* extraction of mutational signatures, proven through a sensitivity analysis and a comprehensive comparison with 10 other available tools. In particular, MUSE-XAE resulted as the best performing and the most robust method in different realistic synthetic scenarios, with an average F1-AUC of 0.92. In addition, MUSE-XAE identified 22 mutational signature profiles in the PCAWG cohort and 23 mutational signatures in the extended WGS cohort, showing a strong agreement with the known signatures from both COSMIC v3.4 and Signal databases. Notably, in the extended WGS cohort, we found a novel candidate signature specific to Eye-Melanoma. This finding will need to be further investigated and validated in an independent cohort. A detailed investigation of mutational signature exposures revealed that MUSE-XAE profiles are capable of enhancing primary tumour type and subtype classifications. Indeed, the classification performance based on the signature exposures showed MCCs around 0.70 in predicting primary types and tumour subtypes in both PCAWG and the extended WGS cohorts.

MUSE-XAE opens up new possibilities for the development of interpretable neural network-based models for mutational signature extraction, which can leverage the increasing amount of available data and their scalability for larger datasets. Our architecture, given its extreme flexibility, can be used to build more sophisticated models which could integrate the profile of somatic mutations with other clinical and genomic information, potentially improving the extraction of mutational signatures.

Competing interests

No competing interest is declared.

Author contributions statement

C.P., P.F. and T.S. conceived the experiment(s), C.P. conducted the experiment(s), C.P., C.R., F.C. and G.B. analysed the results. All authors wrote and reviewed the manuscript.

Acknowledgments

C.P. was supported by a AIRC fellowship for Italy. The authors acknowledge Prof. Anders Krogh for his valuable advice and insights. The authors thank the Italian Ministry for Education, University and Research under the programme "Ricerca Locale ex-60%" and PNRR M4C2 HPC – 1.4 "CENTRI NAZIONALI"- Spoke 8 for fellowship support. In addition, the authors thank the European Union's Horizon 2020 projects GenoMed4All (Grant Agreement ID: 101017549) and Brainteaser (Grant Agreement ID: 101017598).

References

Pan-cancer analysis of whole genomes. *Nature*, 578(7793):82–93, 2020.

L.B. Alexandrov, S. Nik-Zainal, D.C. Wedge, et al. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep*, 3(1):246–259, 2013a.

L.B. Alexandrov, S. Nik-Zainal, D.C. Wedge, et al. Signatures of mutational processes in human cancer. *Nature*, 500(7463): 415–421, 2013b.

L.B. Alexandrov, J. Kim, N.J. Haradhvala, et al. The repertoire of mutational signatures in human cancer. *Nature*, 578 (7793):94–101, 2020.

M. Ardin, V. Cahais, X. Castells, et al. Mutspec: a galaxy toolbox for streamlined analyses of somatic mutation spectra in human and mouse cancer genomes. *BMC Bioinformatics*, 17:1–10, 2016.

M. Bayati, H.R. Rabiee, M. Mehrbod, et al. Cancersign: a user-friendly and robust tool for identification and classification of mutational signatures and patterns in cancer genomes. *Sci Rep*, 10(1):1286, 2020.

F. Blokzijl, R. Janssen, R. van Boxtel, et al. Mutationalpatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med*, 10:1–11, 2018.

A. Degasperi, X. Zou, T.D. Amarante, et al. Substitution mutational signatures in whole-genome–sequenced cancers in the uk population. *Science*, 376(6591):abl9283, 2022.

Andrea Degasperi, T.D. Amarante, J. Czarnecki, et al. A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies. *Nat Cancer*, 1(2):249–263, 2020.

A. Fischer, C.J.R. Illingworth, P.J. Campbell, et al. Emu: probabilistic inference of mutational processes and their localization in the cancer genome. *Genome Biol*, 14(4):1–10, 2013.

K. Gori and A. Baez-Ortega. sigfit: flexible bayesian inference of mutational signatures. *bioRxiv*, page 372896, 2018.

T. Helleday, S. Eshtad, and S. Nik-Zainal. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet*, 15(9):585–598, 2014.

S.M.A. Islam, M. Díaz-Gay, Y. Wu, et al. Uncovering novel mutational signatures by de novo extraction with sigprofilerextractor. *Cell Genom*, 2(11):100179, 2022.

R. Jonker and T. Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. volume 38, pages 325–340. Springer, 1987.

D. Kičiatovas, Q. Guo, M. Kailas, et al. Identification of multiplicatively acting modulatory mutational signatures in cancer. *BMC Bioinformatics*, 23(1):522, 2022.

G. Koh, A. Degasperi, X. Zou, et al. Mutational signatures: emerging concepts, caveats and clinical applications. *Nat Rev Cancer*, 21(10):619–637, 2021.

A. Lal, K. Liu, R. Tibshirani, et al. De novo mutational signature discovery in tumor genomes using sparsesignatures. *PLoS Comput Biol*, 17(6):e1009119, 2021.

J. Ma, J. Setton, N.Y. Lee, et al. The therapeutic significance of mutational signatures from dna repair deficiency in cancer. *Nat Commun*, 9(1):3292, 2018.

F. Maura, A. Degasperi, F. Nadeu, et al. A practical guide for mutational signature analysis in hematological malignancies. *Nat Commun*, 10(1):2969, 2019.

L. Miao and H. Qi. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Trans Geosci Remote Sens*, 45(3):765–777, 2007.

B.T. Nebgen, R. Vangara, M. Hombrados-Herrera, et al. A neural network for determination of latent dimensionality in non-negative matrix factorization. *Machine Learning: Science and Technology*, 2(2):025012, 2021.

C. Pancotti, C. Rollo, G. Birolo, et al. Unravelling the instability of mutational signatures extraction via archetypal analysis. *Front Genet*, 13:3618, 2023.

R.A. Rosales, R.D. Drummond, R. Valieris, et al. signer: an empirical bayesian approach to mutational signature discovery. *Bioinformatics*, 33(1):8–16, 2017.

K. Schulze, S. Imbeaud, E. Letouzé, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet*, 47 (5):505–511, 2015.

F. Schumann, E. Blanc, C. Messerschmidt, et al. Sigspack, a package for cancer mutational signatures. *BMC Bioinformatics*, 20(1):1–9, 2019.

M. Secrier, X. Li, N. de Silva, et al. Mutational signatures in esophageal adenocarcinoma define etiologically distinct subgroups with therapeutic relevance. *Nat Genet*, 48(10): 1131–1141, 2016.

L. Seninge, I. Anastopoulos, H. Ding, et al. Vega is an interpretable generative model for inferring biological network activity in single-cell transcriptomics. *Nature Commun*, 12(1):5684, 2021.

Valentine Svensson, Adam Gayoso, Nir Yosef, and Lior Pachter. Interpretable factor models of single-cell rna-seq via variational autoencoders. *Bioinformatics*, 36(11):3418–3421, 2020.

J.G. Tate, S. Bamford, H.C. Jubb, et al. Cosmic: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*, 47(D1): D941–D947, 2019.

H. Vöhringer, A. van Hoeck, E. Cuppen, et al. Learning mutational signatures and their multidimensional genomic properties with tensorsignatures. *Nat Commun*, 12(1):3628, 2021.

D. Wojtowicz, J. Hoinka, B. Amgalan, et al. Repairsig: Deconvolution of dna damage and repair contributions to the mutational landscape of cancer. *Cell Syst*, 12(10):994–1003, 2021.

H. Zheng, Z. Yang, W. Liu, et al. Improving deep neural networks using softplus units. In *2015 International joint conference on neural networks (IJCNN)*, pages 1–4. IEEE, 2015.