Deep Learning for Cancer Prognosis Prediction Using Portrait Photos by StyleGAN Embedding

Amr Hagag[1,2], Ahmed Gomaa[1,3][0009−0007−6884−2350], Dominik Kornek[1,3], Andreas Maier[2][0000−0002−9550−5284], Rainer Fietkau[1,3], Christoph Bert[1,3][0000−0002−8539−6600], Yixing Huang[1,3,][0000−0003−2627−3077]✉, and Florian Putz[1,3][0000−0003−3966−2872]

[1] Department of Radiation Oncology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
yixing.yh.huang@fau.de
[2] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
[3] Comprehensive Cancer Center Erlangen-EMN (CCC ER-EMN), Erlangen, Germany

Abstract. Survival prediction for cancer patients is critical for optimal treatment selection and patient management. Current patient survival prediction methods typically extract survival information from patients' clinical record data or biological and imaging data. In practice, experienced clinicians can have a preliminary assessment of patients' health status based on patients' observable physical appearances, which are mainly facial features. However, such assessment is highly subjective. In this work, the efficacy of objectively capturing and using prognostic information contained in conventional portrait photographs using deep learning for survival prediction purposes is investigated for the first time. A pre-trained StyleGAN2 model is fine-tuned on a custom dataset of our cancer patients' photos to empower its generator with generative ability suitable for patients' photos. The StyleGAN2 is then used to embed the photographs to its highly expressive latent space. Utilizing state-of-the-art survival analysis models and StyleGAN's latent space embeddings, this approach predicts the overall survival for single as well as pan-cancer, achieving a C-index of 0.680 in a pan-cancer analysis, showcasing the prognostic value embedded in simple 2D facial images. In addition, thanks to StyleGAN's interpretable latent space, our survival prediction model can be validated for relying on essential facial features, eliminating any biases from extraneous information like clothing or background. Moreover, our approach provides a novel health attribute obtained from StyleGAN's extracted features, allowing the modification of face photographs to either a healthier or more severe illness appearance, which has significant prognostic value for patient care and societal perception, underscoring its potential important clinical value.

Keywords: Survival prediction · cancer · StyleGAN · deep learning · latent space · face prognosis · explainable AI.

1   Introduction

Survival analysis for cancer patients is critical for optimal treatment selection and patient management. Patients with very adverse prognosis may need aggressive management and acute intervention. However, patients with good prognostic constellations similarly may benefit from specific treatment approaches like dose escalation and radical treatment of all sites in radiotherapy. So far, a vast class of survival prediction models have been investigated, ranging from conventional regression algorithms [5, 12] to the latest radiomics [4, 22] and deep learning methods [11, 13, 25, 26]. With the successful applications in various fields, nowadays deep learning has become the main trend for patient survival prediction, because of its high representation power and ability to extract essential features from various types of data. Nevertheless, all the previous methods rely on data ranging from regular clinical records to commonly used medical imaging data (e.g., MRI and CT) as well as biological data (e.g., histologic and genetic information) [23]. Within radiotherapy practice, the acquisition of patient photographs is a routine measure implemented for identity verification purposes. Nevertheless, the potential utility in predicting patient survival and overall health status based on facial features derived from such widely available 2D portrait photos remains unexplored.

Facial features observable in 2D portrait photographs, like skin colorit and facial expression can reflect underlying pathophysiologic processes like muscle wasting, lowering of muscle tone or changes in microcirculation [18]. These features therefore hold considerable potential as indicators of a patient's underlying health status. Many severe diseases will lead to evident facial changes, e.g., the most well-known diseases being genetic ones like the Down's syndrome [6]. Clinicians with extensive experience frequently utilize patients' physical appearance, as well as other related features, to estimate survival time and evaluate overall functional status. The Eastern Cooperative Oncology Group (ECOG) score is such a widely used metric for evaluating patient health status [18]. Hui et al. [7] were able to show that drooping of the nasolabial fold was a highly predictive indicator of imminent death within a three-day timeframe. Patient appearance has also been shown to be a factor in determining healthcare priority in emergency situations [1]. However, assessments made by clinicians are inherently subjective and can therefore exhibit a high degree of intra- and inter-rater variance. In this work, we aim to achieve an objective assessment of cancer patients' prognosis based on facial features using deep learning.

Deep learning-based facial image analysis can reveal patient health status caused by various factors [21]. It has been reported to be effective in detecting neurological disorders [28], acromegaly [15], genetic disorders [6], and cancer [16]. However, deep learning models can be biased [24] and have poor generalizability. To ensure meaningful facial features are used, interpretability of the model is crucial. StyleGAN [8, 10] is a widely used network for face image synthesis due to its visual fidelity, diversity, and high editability over its latent space. Essential prognostic features can potentially be embedded into the latent space as well and hence it presents a promising avenue for survival analysis. A notable advantage

of leveraging the latent space for prognosis prediction is the ability to visually and qualitatively evaluate the associations between the features within the latent space and patient's prognosis.

Our contributions can be summarized as follows: a) To the best of our knowledge, this is **the first attempt** to use deep learning for cancer patient survival prediction based on 2D facial photographs; b) We propose a model that utilizes the StyleGAN expressive latent space feature representations for survival prediction. c) We introduce a **novel addition** to traditional prognostic methods by using StyleGAN embeddings alongside clinical parameters to improve survival prediction accuracy. d) To provide **interpretability and validation** to the model, a "health" attribute is extracted to allow adjusting face photographs to either a healthier or more severe illness appearance. This manipulation has significant prognostic value for patient care and societal perception, underscoring its potential important clinical value.
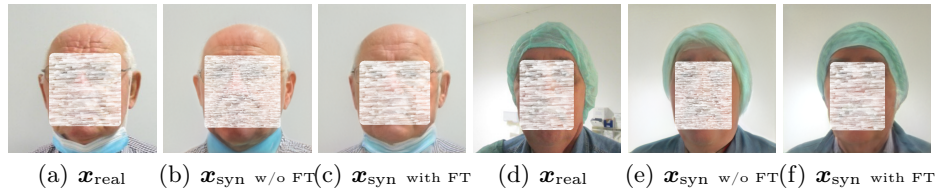
2 Material and Methods

2.1 Data

The dataset utilized in this work comprises 13,503 2D portrait photographs captured during the admission of individuals who have been diagnosed with cancer. Of this cohort, 52.5% are male and 47.5% are female. Notably, 57.3% of the dataset is right censored. In addition, the survival times for the non-censored patients range from 2 to 4,791 days. To mitigate potential biases in the photographs due to factors such as clothing and background, an automated face recognition process utilizing dlib [14] is employed to align and crop all images to a standardized size of 512×512. The clinical information about the patients is retrieved, including international classification of diseases (ICD), gender, age, number of tumors, etc. Information acquired after the photo acquisitions is excluded.

Data Anonymization Disclaimer: This work's data is anonymized to safeguard individual privacy. Facial images are synthetic, anonymized, or expert-reviewed projections, ensuring individuals cannot be identified.

2.2 Embedding Images to StyleGAN2 Latent Space

(a) x_{real} (b) $x_{\text{syn w/o FT}}$ (c) $x_{\text{syn with FT}}$ (d) x_{real} (e) $x_{\text{syn w/o FT}}$ (f) $x_{\text{syn with FT}}$

Fig. 1. Exemplary anonymized face images and their corresponding synthetic images generated by StyleGAN without (w/o) and with fine-tuning (FT).

To accurately capture patients' facial features from a given photo, a generative model should be able to accurately reproduce these expressive features into a generated image. A StyleGAN2 model pretrained on the FFHQ dataset [9] is fine-tuned using adaptive discriminator augmentation (ADA) [8] with our custom dataset of cancer patients' face images. This step is necessary to increase StyleGAN's generation ability for the special facial features in cancer patients and auxiliary medical equipment such as masks, breathing tubes and gauze. Fig. 1 shows the difference between synthesised images with and without fine-tuning. Without fine-tuning, the pretrained model fails to generate the mask and the bouffant cap in Fig. 1(b) and Fig. 1(e), respectively.

Photo Embedding: To project images from the image space to the latent space, for each photo $\boldsymbol{x}_{\mathrm{real}}$, an optimization process searches via gradient decent for a 512-dimensional latent space vector $\boldsymbol{z}$ if input to the StyleGAN2 generator, will produce a corresponding synthetic image $\boldsymbol{x}_{\mathrm{syn}}$ similar to the original image $\boldsymbol{x}_{\mathrm{real}}$. This similarity is defined as Euclidean distance in the feature space of a pretrained VGG-16,

$$\boldsymbol{z} = \arg\min_{\boldsymbol{z}} \phi\left(\boldsymbol{x}_{\mathrm{real}}, \boldsymbol{x}_{\mathrm{syn}}\right) = \arg\min_{\boldsymbol{z}} \phi\left(\boldsymbol{x}_{\mathrm{real}}, G(\boldsymbol{z})\right), \tag{1}$$

where G is the generator of the fine-tuned StyleGAN2 model, and $G(\boldsymbol{z})$ is the synthetic image $\boldsymbol{x}_{\mathrm{syn}}$, and ϕ is the loss function. In such a way, each face image is associated with one optimized highly expressive 512-dimensional latent vector.

2.3 Survival Analysis Models

Survival analysis models are used to predict patients' survival by estimating the hazard function, which is the instantaneous rate at which the event of interest (e.g., death) occurs at a given time, and allows for the inclusion of multiple covariates to predict the risk of the event. The most widely used survial model in practice is the Cox Proportional Hazard (CoxPH) [17] model based on conventional clinical features. In the CoxPH model, the hazard function has the following form [11],

$$\lambda(t|\boldsymbol{f}) = \lambda_0(t) \cdot e^{h(\boldsymbol{f})} = \lambda_0(t) \cdot e^{\boldsymbol{w}^{\top}\boldsymbol{f}}, \tag{2}$$

where $\lambda_0(t)$ is a baseline hazard function, $\boldsymbol{f}$ is the input feature vector, $h(\boldsymbol{f})$ is the log-risk function, and $\boldsymbol{w}$ is the coefficient vector to optimize.

To learn a more complex, non-linear relationship between the features for the hazard estimation, deep learning networks have been proposed, with DeepSurv [11, 29] being the state-of-the-art. The DeepSurv method applies a multi-layer perceptron (MLP) to estimate the log-risk function $h(\boldsymbol{f})$ of the CoxPH model. To extract features from medical images, the MLP network can be replaced by a convolutional neural network (CNN) [6, 29]. In this work, a ResNet-18 and a VGG-16 network are investigated as examples of the CNN with image data. As CNN-extracted features are highly abstract, in this work, we propose to integrate the StyleGAN network to the DeepSurv framework for the prediction of continuous survival time, i.e., utilizing the expressive StyleGAN photo embedding $\boldsymbol{z}$ as the input features of the CoxPH and DeepSurv models.
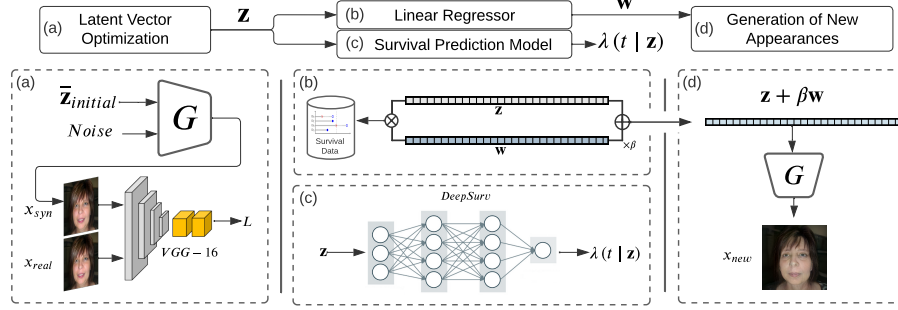
Fig. 2. The workflow of our proposed facial prognosis method: (a) Optimization of z using StyleGAN generator fine-tuned on custom cancer patient data; (b) CoxPH regression to get w for health attribute construction and latent space manipulation; (c) DeepSurv for survival prediction; (d) Generation of new appearance.

2.4 Latent Space Manipulation

In this work, we propose two tests to visually and qualitatively evaluate the semantic meaning of our photo embedding as well as the prognostic value it holds for the task of survival prediction. Each of the tests can be conducted by simple manipulations of the latent vector z.

Health Attribute Manipulation: Leveraging the linearity of the CoxPH regression model, we fit the 512-dimensional vectors z to the known survival data,

$$h(\boldsymbol{z}) = \sum_{i=0}^{N-1} \boldsymbol{w}_i \boldsymbol{z}_i, \tag{3}$$

where w is the vector of linear coefficients, and $h(z)$ is the log-risk function which is fed to the CoxPH model. By manipulating z in all the dimensions jointly with a multiplier of w,

$$\hat{\boldsymbol{z}} = \boldsymbol{z} + \beta \boldsymbol{w}, \tag{4}$$

the dimensions in z with high importance are modified to a large degree, while those with low importance stay almost unchanged. By changing β, the StyleGAN generator is able to generate new appearances of this patient from $\hat{z}$, with a healthier or more severe illness appearance. This vector w serves as a health attribute.

Age Manipulation: To further validate the quality of our facial features vectors and ensure that our predictive model is focusing on complex survival related aspects and not only a single aspect like the age, we use a simple linear regression to map the feature vectors z to the known age of the patients. Similar to (Eqn. (4)), w here would serve as a simple age attribute.

Table 1. Survival prediction results of different methods using DeepSurv (DS) and CoxPH models. C-index and Brier score (B-Score) are calculated for all models.

| Input data	Method	C-index	B-Score	Input data	Method	C-index	B-Score
Embeddings	DS+MLP	0.680	0.200	Clinic. Data	CoxPH	0.690	0.151
Clinic. Data	DS+MLP	0.729	0.168	Photos	DS+ResNet	0.510	0.471
Combined	DS+MLP	0.787	0.137	Photos	DS+VGG	0.470	0.471

3   Results and Discussion

3.1 Pan-Cancer Analysis for Survival Prediction Results

The survival prediction results are displayed in Table 1. Our approach for continuous survival time prediction using DeepSurv based on StyleGAN photo embeddings achieved for pan-cancer analysis a mean concordance index (C-index) of 0.680 and a mean Brier Score of 0.200. To put this into perspective, the model's survival predictions are evidently higher than chance. This reflects that the model uses key survival features from relevant prognostic facial attributes embedded in the latent vectors. As also shown in Table 1, the DeepSurv model using patients' portraits fed directly through a CNN achieved a C-index of 0.51 and 0.47 using the ResNet-18 and the VGG-16 networks respectively. In this case, the CNNs failed to extract the needed prognostic information directly from the portrait photos. It is worth noting that the comparison methods (i.e., DS+ResNet and DS+VGG with input photo data) used end-to-end training instead of two-step training as our proposed method (i.e., DS+MLP with input combined embedding and clinical data).

As also shown in Table 1, survival prediction models based on conventional clinical features achieved a C-index of 0.690 and 0.729 using CoxPH and Deep-Surv models respectively, which is comparable to what was achieved by the embeddings of simple 2D portraits. Combining clinical features with photo embeddings using late fusion boosted the C-index to 0.787 highlighting again the complementary prognostic value of StyleGAN's embeddings.

3.2 Subgroup Analysis for Survival Prediction Results

Intriguingly, as shown in Fig 3, the developed pan-cancer model showed good prognostic accuracy for nearly all cancer sub-types, suggesting that prognostically-relevant facial features are present in most malignancies. However, there were also notable differences in observed prognostic accuracy. High prognostic accuracy was observed for endocrine malignancies (including small-cell cancers), metastatic cancer of unknown primary (CUP), urothelial cancer, breast cancer and melanoma, with C-indices of 0.786, 0.738, 0.723, 0.697 and 0.678, respectively, whereas notably lower accuracy was observed for lower and upper gastrointestinal cancers, prostate cancer, hematologic malignancies, sarcoma and especially renal cancer, with C-indices of 0.609, 0.595, 0.639, 0.638, 0.631 and 0.52, respectively. From a clinical perspective, this pattern could be derived from differences in tumor biology, as hematologic malignances as well as sarcomas have a different origin and a distinct biology from carcinomas [2]. In addition,
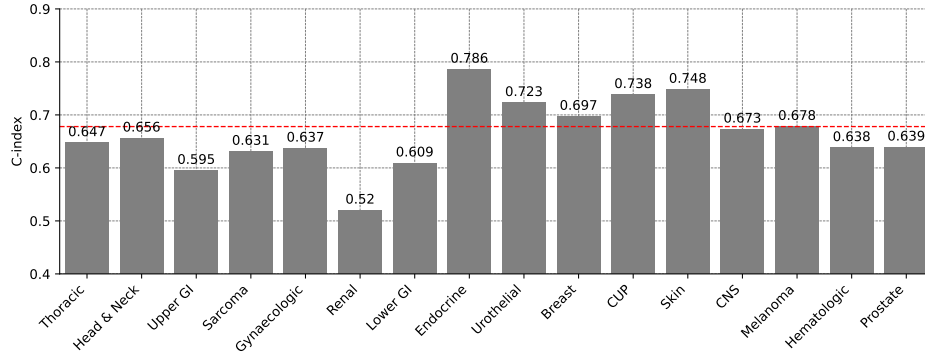
Fig. 3. A subgroup analysis on 16 cancer types with corresponding single C-indices. The dashed line represents the average C-index of the pan-cancer analysis.
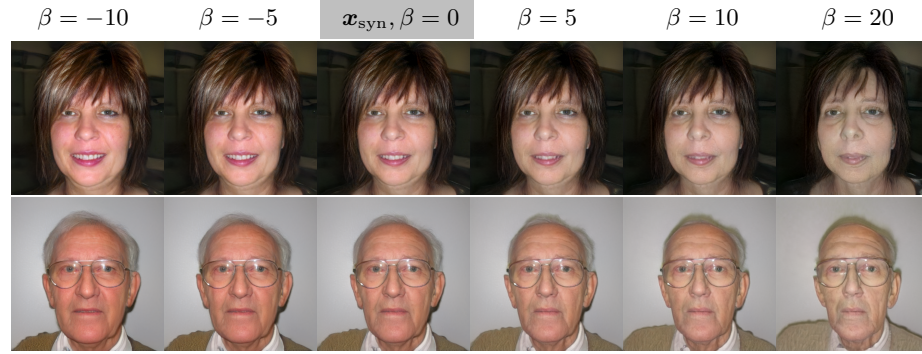

Fig. 4. Manipulating the health attribute with different β values for regenerated patients' images. The synthesized image without manipulation is at $\beta = 0$ (third column).

prostate cancer, colorectal cancer, and renal cancer are slowly proliferating tumors with markedly slower disease progression, lower tendency for metastasis and good systemic treatment options that typically result in patients showing less severe systemic signs of the disease [3].

3.3 Model Interpretability with Health Attribute

Fig. 4 demonstrates the synergetic effect of manipulating patient's images in the latent space using the health attribute (Eqn. (4)). In Fig. 4 the appearances of the patients are controlled by β. With a negative value for β, the patients appear to be healthier with a reddish skin colorit, full cheeks, a smiling expression and a strong tone of the facial muscles. In contrast, with a positive value for β, there is a loss of masticatory muscles and subcutaneous fat with a pale skin colorit. Moreover, the tone of the facial muscles is reduced with drooping of the nasolabial fold, an overall more relaxed appearance of facial features and an increased "tippiness" of the nose. These features are consistent with long-

$\beta = -10$ $\beta = -5$ $\boldsymbol{x}_{\mathrm{syn}}, \beta = 0$ $\beta = 5$ $\beta = 10$ $\beta = 20$
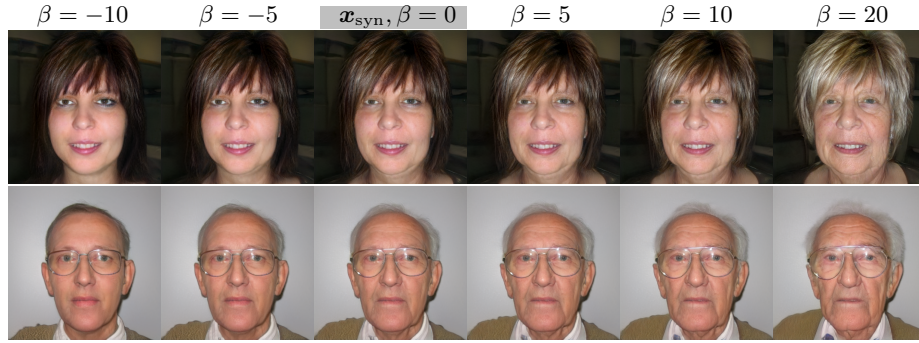
Fig. 5. Exemplary patients' images generated by manipulating the age attribute.

standing clinical descriptions of prognostically relevant changes in facial appearance [7, 27]. Most importantly, the change is mainly located in the head region instead of their clothes or the background. The model is also able to reflect patients' mental status to some degree by extracting their facial expressions.

This editability of patients' photos to have either a healthier appearance or prognosis of a more severe illness has important value for patient care. For cancer patients, the quality of life is as important as the prevention, treatment and management of their illness. The quality of life relies on not only physical but also mental well-being [19, 20]. Editing their photos to have a healthier appearance and sharing such optimized photos to their social networks can convey a more positive and optimistic message. For healthcare professionals, this could enable illustrating a patients' appearance in a healthy state or a state of illness progression, which could improve empathy and caretaking.

An observation that could not be neglected however is the similarity between progressing in age and worse prognostic features, as age is a common risk indicator. While one can assume an underlying correlation between age and general health related facial features, we addressed a potential bias of the model to only rely on age features for the prognosis prediction task. As shown in Fig. 4, it is clearly observable that when editing the patient's image based on the health attribute, distinct facial features are affected compared to when editing based on the age attribute, as shown in Fig. 5.

Based on the feature importance indicated by the regression model's coefficients $\boldsymbol{w}$ in Eqn.(4), changes in the embedding vector elements with corresponding higher coefficients have a more observable impact on the attributes of the generated images. This demonstrates that, as in Fig. 4, the StyleGAN regression model is indeed using meaningful facial features instead of bias or random information for survival prediction.

4   Conclusion

This work investigates on prognosis prediction of cancer patients using portrait photos. Our work demonstrates that the latent features extracted by StyleGAN contain highly relevant prognostic information, and an effective regression model can be optimized for survival prediction. Such a model using facial features has good interpretability and can achieve comparable survival prediction accuracy to a state-of-the-art model using clinical record data. The constructed health attribute allows photo editing into a healthier appearance or prognosis of a more severe illness, which has important potential value for patient care.

Disclosure of Interests. The authors have no conflict of interest to declare for this work.

Bibliography

[1] Bagnis, A., Caffo, E., Cipolli, C., De Palma, A., Farina, G., Mattarozzi, K.: Judging health care priority in emergency situations: patient facial appearance matters. Social Science & Medicine **260**, 113180 (2020)

[2] Berman, J.: Modern classification of neoplasms: reconciling differences between morphologic and molecular approaches. BMC cancer **5**, 1–12 (2005)

[3] Capitanio, U., Montorsi, F.: Renal cancer. The Lancet **387**(10021), 894–906 (2016)

[4] Chen, J., Lu, S., Mao, Y., Tan, L., Li, G., Gao, Y., Tan, P., Huang, D., Zhang, X., Qiu, Y., et al.: An mri-based radiomics-clinical nomogram for the overall survival prediction in patients with hypopharyngeal squamous cell carcinoma: a multi-cohort study. European Radiology pp. 1–10 (2022)

[5] Emura, T., Chen, Y.H., Chen, H.Y.: Survival prediction based on compound covariate under cox proportional hazard models. PLoS One **7**(10), e47627 (2012)

[6] Gurovich, Y., Hanani, Y., Bar, O., Nadav, G., Fleischer, N., Gelbman, D., Basel-Salmon, L., Krawitz, P.M., Kamphausen, S.B., Zenker, M., et al.: Identifying facial phenotypes of genetic disorders using deep learning. Nature medicine **25**(1), 60–64 (2019)

[7] Hui, D., Hess, K., Santos, R.d., Chisholm, G., Bruera, E.: A diagnostic model for impending death in cancer patients: preliminary report. Cancer **121**(21), 3914–3921 (2015)

[8] Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T.: Training generative adversarial networks with limited data. Advances in neural information processing systems **33**, 12104–12114 (2020)

[9] Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proc. CVPR. pp. 4401–4410 (2019)

[10] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of stylegan. In: Proc. CVPR. pp. 8110–8119 (2020)

[11] Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. BMC medical research methodology **18**(1), 1–12 (2018)

[12] Kidd, A.C., McGettrick, M., Tsim, S., Halligan, D.L., Bylesjo, M., Blyth, K.G.: Survival prediction in mesothelioma using a scalable lasso regression model: instructions for use and initial performance using clinical predictors. BMJ open respiratory research **5**(1), e000240 (2018)

[13] Kim, D.W., Lee, S., Kwon, S., Nam, W., Cha, I.H., Kim, H.J.: Deep learning-based survival prediction of oral cancer patients. Scientific reports **9**(1), 1–10 (2019)

[14] King, D.E.: Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research **10**, 1755–1758 (2009)

[15] Kong, X., Gong, S., Su, L., Howard, N., Kong, Y.: Automatic detection of acromegaly from facial photographs using machine learning methods. EBioMedicine **27**, 94–102 (2018)

[16] Liang, B., Yang, N., He, G., Huang, P., Yang, Y.: Identification of the facial features of patients with cancer: a deep learning–based pilot study. Journal of Medical Internet Research **22**(4), e17234 (2020)

[17] Lin, H., Zelterman, D.: Modeling survival data: extending the cox model (2002)

[18] Oken, M.M., Creech, R.H., Tormey, D.C., Horton, J., Davis, T.E., McFadden, E.T., Carbone, P.P.: Toxicity and response criteria of the eastern cooperative oncology group. Am. J. Clin. Oncol. **5**(6), 649–656 (1982)

[19] de Oliveira, W.A.: Quality of life, facial appearance and self-esteem in patients with orthodontic treatment. Revista Mexicana de Ortodoncia **5**(3), 138–139 (2017)

[20] Rankin, M., Borah, G.L.: Perceived functional impact of abnormal facial appearance. Plastic and reconstructive surgery **111**(7), 2140–2146 (2003)

[21] Su, Z., Liang, B., Shi, F., Gelfond, J., Šegalo, S., Wang, J., Jia, P., Hao, X.: Deep learning-based facial image analysis in medical research: a systematic review protocol. BMJ open **11**(11), e047549 (2021)

[22] van Timmeren, J.E., Leijenaar, R.T., van Elmpt, W., Reymen, B., Oberije, C., Monshouwer, R., Bussink, J., Brink, C., Hansen, O., Lambin, P.: Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam ct images. Radiotherapy and Oncology **123**(3), 363–369 (2017)

[23] Vale-Silva, L.A., Rohr, K.: Long-term cancer survival prediction using multimodal deep learning. Scientific Reports **11**(1), 1–12 (2021)

[24] Wallis, D., Buvat, I.: Clever hans effect found in a widely used brain tumour mri dataset. Medical Image Analysis **77**, 102368 (2022)

[25] Wang, D., Jing, Z., He, K., Garmire, L.X.: Cox-nnet v2. 0: improved neural-network-based survival prediction extended to large-scale emr data. Bioinformatics **37**(17), 2772–2774 (2021)

[26] Wankhede, D.S., Rangasamy, S.: Dynamic architecture based deep learning approach for glioblastoma brain tumor survival prediction. Neuroscience Informatics p. 100062 (2022)

[27] Withington, E., Lonie, I., Chadwick, J., Mann, W.N., Lloyd, G., et al.: Hippocratic writings. Penguin UK (2005)

[28] Yolcu, G., Oztel, I., Kazan, S., Oz, C., Palaniappan, K., Lever, T.E., Bunyak, F.: Deep learning-based facial expression recognition for monitoring neurological disorders. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1652–1657. IEEE (2017)

[29] Zhu, X., Yao, J., Huang, J.: Deep convolutional neural network for survival analysis with pathological images. In: 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 544–547. IEEE (2016)